

**UNIVERSITE VICTOR SEGALEN BORDEAUX 2**

**U.F.R. "Sciences et Modélisation"**

**COURS de STATISTIQUE MATHEMATIQUE**

**Modèles, Méthodes, Applications**

**à**

**l'usage des étudiants de DEUG, Licence et Master**

**M. Nikulin  
V. Bagdonavičius  
C. Huber  
V. Nikoulina**

**BORDEAUX  
2004/2005**



# Table des matières

<b>0</b>	<b>LOIS USUELLES. APPROXIMATIONS.</b>	<b>11</b>
0.1	Lois discrètes. Approximations normale et de Poisson. Théorème limite central . . . . .	11
0.2	Approximations normales et de Poisson . . . . .	14
0.3	Lois continues. Liaisons entre des lois . . . . .	15
0.4	Epreuves de Bernoulli et marches aléatoires. . . . .	22
0.5	Représentation d'une suite d'épreuves de Bernoulli indépendante . . . . .	22
0.6	Probabilités associées à une marche aléatoire reliant 2 points du treillis $S$ .	23
0.7	Frontière absorbante . . . . .	24
0.8	Marches aléatoires et distributions discrètes . . . . .	24
<b>1</b>	<b>QUELQUES PROBLÈMES CLASSIQUES DE LA STATISTIQUE MATHÉMATIQUE.</b>	<b>31</b>
1.1	Problèmes d'estimation et de comparaison des probabilités de succès. . . .	31
1.2	Modèle probabiliste de l'erreur de mesure. . . . .	41
1.3	Méthode de Monte-Carlo. . . . .	50
<b>2</b>	<b>ELEMENTS DE LA THEORIE DE L'ESTIMATION PONCTUELLE.</b>	<b>55</b>
2.1	Modèle statistique. Fonction de vraisemblance. . . . .	55
2.2	Statistique. Échantillon. Loi empirique. . . . .	56
2.3	Estimateur ponctuel. Consistance. Estimateur invariant . . . . .	62
2.4	Fonction de perte, fonction de risque. . . . .	64
2.5	Statistiques exhaustives, nécessaires, minimales et complètes. . . . .	65
2.6	Information de Fisher. Inégalité de Rao-Cramer-Fréchet. Théorème de Rao-Blackwell-Kolmogorov. . . . .	69
2.7	Méthode des moments. . . . .	78
2.8	Méthode des moindres carrés. Modèle de Gauss de la théorie des erreurs. .	81
2.9	Régions, intervalles, limites de confiance. . . . .	86
2.10	Méthode de Bolshev de construction des limites de confiance. . . . .	88
2.11	Théorème de Fisher. . . . .	92
2.12	Intervalle de confiance pour la moyenne d'une loi normale . . . . .	100
2.13	Intervalle de confiance pour la variance d'une loi normale . . . . .	105
2.14	Intervalle de confiance pour la différence des moyennes de deux lois normales	112
2.15	Intervalle de confiance pour le quotient des variances de deux lois normales.	117
2.16	La loi de Thompson. . . . .	119
2.17	Méthode du maximum de vraisemblance. . . . .	121
2.18	Propriétés asymptotiques du rapport de vraisemblance . . . . .	132

2.19	Decomposition orthogonale de Fisher . . . . .	151
2.20	Modèle d'analyse des variances à 2 facteurs. . . . .	154
2.21	Modèle exponentiel. Analyse statistique. . . . .	163
<b>3</b>	<b>ELEMENTS DE LA STATISTIQUE NON PARAMETRIQUE.</b>	<b>169</b>
3.1	La loi empirique. . . . .	169
3.2	Médiane de la loi empirique. . . . .	180
3.3	Théorème de Kolmogorov. . . . .	183
3.3.1	Transformation de Smirnov. Test de type de Kolmogorov-Smirnov pour des lois discrètes. . . . .	184
3.4	Tests de Kolmogorov et Smirnov pour un échantillon. . . . .	186
3.5	Test de Kolmogorov-Smirnov pour deux échantillons. . . . .	189
3.6	Test $\omega^2$ de Cramer-von Mises et statistiques associées de Lehmann, Gini, Downton, Moran-Greenwood et Sherman. . . . .	190
3.7	Les statistiques de Kolmogorov et Gihman. . . . .	195
3.8	Test des signes. . . . .	197
3.9	Test de Wilcoxon. . . . .	200
3.10	Estimation non paramétrique de la densité. Histogramme. Estimateur de Rosenblatt. Le noyau de Parzen. . . . .	204
<b>4</b>	<b>TESTS STATISTIQUES.</b>	<b>207</b>
4.1	Principe des tests. . . . .	207
4.2	Test de Neyman-Pearson. . . . .	209
4.3	Loi multinomiale et test du chi-deux de Pearson. . . . .	214
4.4	Théorème de Fisher. . . . .	220
4.5	Théorème de Chernoff-Lehmann. . . . .	224
4.6	Test du chi-deux pour une loi logistique. . . . .	225
4.7	Test du chi-deux dans un problème d'homogénéité. . . . .	228
4.8	Test du $\chi^2$ d'homogénéité pour des lois multinomiales. . . . .	233
4.9	Test du $\chi^2$ pour l'indépendance dans une table de contingence. . . . .	236
4.10	Test du Chauvenet pour la détection des observations aberrantes. . . . .	241
<b>5</b>	<b>REGRESSION</b>	<b>243</b>
5.1	Régression linéaire . . . . .	243
5.1.1	Modèle de la régression linéaire . . . . .	243
5.1.2	Codage des covariables . . . . .	244
5.1.3	Interprétation des coefficients $\beta$ . . . . .	245
5.1.4	Modèle avec interactions . . . . .	245
5.1.5	Estimateurs des moindres carrés . . . . .	246
5.1.6	Propriétés des estimateurs . . . . .	247
5.1.7	Décomposition des sommes de carrés . . . . .	250
5.1.8	Le coefficient de détermination. . . . .	252
5.1.9	Régression linéaire simple . . . . .	253
5.1.10	Régression normale . . . . .	254
5.1.11	Estimateurs du maximum de vraisemblance . . . . .	255
5.1.12	Lois des estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ . . . . .	255
5.1.13	Test de l'hypothèse $H_0 : \beta_{k+1} = \dots = \beta_m = 0$ . . . . .	257
5.1.14	Les coefficients empiriques de la corrélation partielles . . . . .	260

5.1.15	Intervalle de confiance pour les coefficients $\beta$ et leur combinaisons linéaires . . . . .	261
5.1.16	Intervalle de confiance pour les valeurs de la fonction de régression $m(x)$ . . . . .	262
5.1.17	Prédiction de la nouvelle observation . . . . .	263
5.1.18	Analyse des résidus . . . . .	263
5.2	Annexe . . . . .	266
5.3	Régression logistique . . . . .	274
5.3.1	Estimation . . . . .	276
<b>6</b>	<b>ELEMENTS D'ANALYSE DES DONNEES CENSUREES ET TRONQUEES.</b>	<b>281</b>
6.1	Distribution de survie. . . . .	281
6.2	Risque de panne ou taux de défaillance. . . . .	284
6.3	Modèles paramétriques de survie. . . . .	289
6.4	Modèles nonparamétriques . . . . .	298
6.5	Types de censure. . . . .	304
6.6	Troncature. . . . .	313
6.7	Estimateur de Kaplan-Meier. . . . .	316
6.8	Modèle de Cox. . . . .	321
6.9	Sur l'estimation semiparamétrique pour le modèle de Cox . . . . .	323
6.10	Processus de comptage et l'estimation non paramétrique . . . . .	328
6.11	Estimation dans des expériences accélérées . . . . .	336
6.11.1	Modèles de vie accélérée . . . . .	336
6.11.2	Estimation paramétrique . . . . .	341
6.11.3	Estimation semiparamétrique . . . . .	350
<b>7</b>	<b>INFERENCE BAYESIENNE</b>	<b>357</b>
7.1	La règle Bayésienne . . . . .	357
7.2	Estimation ponctuelle . . . . .	359
7.3	Approche bayésienne empirique . . . . .	370
7.4	Exemple . . . . .	370
7.4.1	La loi beta et ses propriétés . . . . .	370
7.5	Résultats principaux. . . . .	371
7.6	Aproximations . . . . .	373
<b>8</b>	<b>EXERCICES.</b>	<b>375</b>
<b>9</b>	<b>SOLUTIONS.</b>	<b>383</b>



# AVANT PROPOS

Ce fascicule est destiné tout d'abord aux étudiants de

## **l'UFR "Sciences et Modélisation"**

(ancienne l'UFR MI2S) de l'Université Victor Segalen Bordeaux 2, qui veulent apprendre les notions fondamentales de la statistiques mathématiques. Le contenu de ce fascicule est une synthèse des des cours de statistique que j'ai donné à l'Université Bordeaux 2, l'Université Bordeaux 1 et l'Univrsité Bordeaux 4 dans les années 1992-2002. Il est supposé que les étudiants aient la connaissance avec des notions fondamentales de la théorie de probabilité pour apprendre la première partie de cours et de la théorie des processus stochastiques pour la deuxième partie, exposées par exemple, dans le fascicule

"*Calcul des Probabilités et Introduction aux Processus Aléatoires*", 2000/2001, UFR MI2S, (V.Bagdonavičius, V.Nikoulina et M.Nikulin). Il y a une corrélation forte positive entre ces deux cours.

Il faut remarquer qu'à la base de cet ouvrage se trouvent les mêmes idées statistiques qui étaient exposées dans les deux polycopies de C.Huber et M.Nikulin :

"*Transformations des variables aléatoires. Applications au choix et à la réduction d'un modèle statistique*", (1991), UFR "Etudes Médicales et Biologiques", Université Paris 5, et "*Applications Statistiques des Transformations des Variables Aléatoires*", (1993), UFR MI2S, Université Bordeaux 2.

Pour traiter bien les données, c'est-à-dire pour mener à bien les estimations et les tests classiques, paramétriques ou non paramétriques, on transforme les observations brutes en calculant des statistiques bien choisies qui doivent avoir les propriétés suivantes :

1. *Perdre le moins d'information possible, éventuellement pas du tout et c'est le cas des statistiques exhaustives, tout en réduisant au minimum le volume initial des observations.*

2. *Etre calculable ou avoir une bonne approximation. Par exemple s'il s'agit d'un estimateur obtenu par la méthode de maximum de vraisemblance, il se peut que l'on ne puisse en obtenir aisément qu'une valeur approchée au premier pas à partir d'un estimateur moins bon.*

3. *Leurs lois doivent être, soit connues explicitement, soit admettre une bonne approximation. Bonne voulant dire à la fois simple à calculer et ayant une bonne vitesse de convergence vers la vraie valeur.*

Ce qui suit donne, grâce à des transformations appropriées des observations, des statistiques qui ont ces propriétés et aussi de bonnes approximations des lois usuelles et permet ainsi de n'utiliser essentiellement que deux tables : celle de la loi normale standard et celle des lois gamma (ou chi-deux). Des exemples illustrent l'application de ces méthodes, qui donnent des approximations meilleures ( vitesse de convergence plus rapide) que les approximations usuelles.

Ces techniques sont très utiles pour tous les statisticiens qui travaillent sur des pro-

blèmes concrets, en particulier pour les ingénieurs, mais aussi, et c'est moins connu, dans les domaines de la médecine, de la biologie et de la sociologie.

De plus cette approche nous permet de considérer "*les transformations des variables aléatoires*" comme le synonyme d'une partie de "*la statistique mathématique*", qui est basée sur la théorie de la probabilité. Ce point de vue sur le rôle des transformations des variables aléatoires dans la statistique a été exprimé très nettement par Professeur L.N. Bolshev dans ces articles, voir, par exemple, (1959), (1963) etc.

Dans cette optique C.Huber, T.Smith and M.Nikulin ont préparé le manuscrit "*Introduction to the Theory of Statistical Inference*",(1992), Département of Mathematics and Statistics, Queen's University, Kingston, Canada. Ce manuscrit a été largement utilisé pour créer la base du cours de la statistique que j'ai donné à Queen's University en 1991-1992, ainsi que les cours de statistiques donnés au sein de l'UFR MI2S à l'Université Bordeaux 2.

Il faut noter que pour préparer le cours actuel nous avons utilisé aussi les livres suivants :

V. Bagdonavičius & M.Nikulin, "*Accelerated Life Models*", **2002**,  
Chapman&Hall/CRC : Boca Raton,

C.Huber, "*Statistique au PCEM*",**1992**, Masson, Paris,

V.Voinov & M.Nikulin, "*Unbiased Estimators and Their Applications. Vol.1 : Univariate Case*" **1993**, Kluwer Academic Publishers, Dordrecht),

V.Voinov & M.Nikulin, "*Unbiased Estimators and Their Applications. Vol.2 : Multivariate Case*", **1996**, Kluwer Academic Publishers, Dordrecht,

P.E.Greenwood & M.Nikulin, "*A Guide to Chi-Squared Testing*", **1996**, John Wiley and Sons, New-York,

*Encyclopaedia of Mathematics*, **1994**, (Editor : M.Hasewinkel), Kluwer Academic Publishers, **v. 1-10**,

*Probability & Mathematical Statistics : Encyclopaedia*, **1999**, (Ed. : Yu.V.Prokhorov), Big Russian Encyclopaedia,Moscow,

d'où était tiré la plupart des exemples, définitions, remarques, exercices et démonstrations des résultats à caractère théorique pour construire les cours de statistique que nous avons donné à l'Université Bordeaux 2 (DEUG, Licence et Maîtrise de la filière MASS, DESS et DEA de la filière Sciences Cognitive à l'UFR MI2S, DESS de Statistique Appliquée aux Sciences Sociales et de Santé de l'ISPED. Ce cours est lié avec d'autres cours de statistiques donnés à l'Université Bordeaux 2 ( les UFR's STAPS, Sciences de la Vie, Sciences Pharmaceutiques, l'ISPED) et peut-être bien utilisé comme le support de base dans l'enseignement des cours de statistiques de niveau de DESS et DEA orientés vers le milieu biomédicale, ainsi que pour les sciences sociales et économiques. En particulier, il est bien adapté pour le DESS "*Statistique Appliquée aux Sciences Sociales et de la Santé*" et DEA d'*Epidémiologie (Option Biostatistique)* à l'Institut de Santé Publique, d'Epidémiologie et de Développement. Cet ouvrage est très lié avec notre ouvrage précédent "*Statistique mathématique : Théorie, Méthodes and Applications*", (2000/2001).

Dans ces cours nous avons essayé d'exposer les idées et les notions fondamentales de la statistique mathématique en termes de définitions, exemples et remarques et d'introduire les techniques des transformations des données et les méthodes statistiques que l'on utilise souvent dans les applications. Tout cela ensemble permet d'apprendre les bases fondamentales de la statistique mathématique, d'apprendre à travailler avec des logiciels et des tables statistiques, de construire des modèles probabilistes et de faire des inférences statistiques, et par conséquent, à être prêt de travailler dans les différents domaines d'applications des

modèles et méthodes de la statistique mathématique. Il est évident que ce cours de statistique reflète des intérêts statistiques des auteurs et que nous avons traité plus profondément les thèmes qui sont proches aux thèmes de recherches, développés au sein du Laboratoire "Statistique Mathématiques et ses Applications" de l'Université Bordeaux 2. Il faut noter que parallèlement à l'Université Bordeaux 2 on fait d'autres cours de statistiques, qui sont plus appliqués et où on considère des méthodes d'analyse des données, de la statistique multivariée, de l'analyse des régressions et surtout de l'analyse de survie dans le cadre des cours de statistiques de l'ISPED.

Vu l'importance d'applications des modèles semiparamétriques avec des covariables dépendant du temps dans l'analyse de survie, en fiabilité, dans l'économie etc., nous avons mis quelques résultats récents, liés avec la théorie des épreuves accélérées. Plus d'informations on peut voir, par exemple, dans nos monographies avec V.Bagdonavičius "*Semiparametric Models in Accelerated Life Testing*", (1995), et "*Additive and Multiplicative Semiparametric Models in Accelerated Life Testing and Survival Analysis*", (1998).

A la fin il faut ajouter que nos cours de statistiques sont accompagnés des travaux pratiques en Statistiques avec l'utilisation de SPSS.

Je remercie mes collègues des Universités Bordeaux 1, 2 et 4, de l'Université Paris 5, et tous les participants au Séminaire Statistique des Universités de Bordeaux et du Séminaire Européen "*Mathematical Methods in Survival Analysis and Reliability*", avec lesquels nous avons discuté sur les problèmes d'enseignement de la statistique. Les discussions ont été très intéressantes et très utiles pour nous, et surtout avec A.Alioum,, Ch.Bulot, D.Commenges, V.Couallier, L.Gerville-Réache, H.Lauter, M.Mesbah, J.Poix, V.Solev, V.Voinov.

Mikhail Nikouline



# Chapitre 0

## LOIS USUELLES. APPROXIMATIONS.

### 0.1 Lois discrètes. Approximations normale et de Poisson. Théorème limite central

Ici nous allons exposer des lois probabilistes que l'on utilise souvent en applications statistiques, des liaisons entre elles et des approximations utiles. Plus d'information à ce sujet on peut trouver dans les publications de L.Bolshev (1963), C.Huber et M.Nikulin (1993), où, en particulier, est exposée la théorie des transformations asymptotique de Pearson, développée par L.Bolshev, voir aussi, L.Bolshev et N.Smirnov (1968), M.Nikulin (1984), Bagdonavicius et Nikulin (2002).

**Définition 1.** On dit qu'une variable aléatoire discrète  $X$  suit la loi de Bernoulli de paramètre  $p$ ,  $p \in [0, 1]$ , si  $X$  ne prend que 2 valeurs 1 et 0 avec les probabilités

$$p = \mathbf{P}\{X = 1\} \quad \text{et} \quad q = 1 - p = \mathbf{P}\{X = 0\},$$

i.e.

$$\mathbf{P}\{X = x\} = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}. \quad (1)$$

Il est clair que

$$\mathbf{E}X = p, \quad \mathbf{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = pq \leq \frac{1}{4}.$$

On remarque que

$$\frac{\mathbf{Var} X}{\mathbf{E}X} = q < 1.$$

**Définition 2.** Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes et qui suivent la même loi de Bernoulli (1) de paramètre  $p$ . Dans ce cas on dit que la statistique

$$\mu_n = \sum_{i=1}^n X_i$$

suit la loi binomiale  $B(n, p)$  de paramètres  $n$  et  $p$ ,  $0 \leq p \leq 1$ , et on note  $\mu_n \sim B(n, p)$ .

Il est facile de montrer que

$$\mathbf{P}\{\mu_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}, \quad (2)$$

$$\mathbf{E}\mu_n = np, \quad \mathbf{Var}\mu_n = np(1-p) = npq.$$

La fonction de répartition de  $\mu_n$  est

$$\begin{aligned} \mathbf{P}\{\mu_n \leq m\} &= \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} = 1 - I_p(m+1, n-m) = \\ &I_{1-p}(n-m, m+1), \quad 0 \leq m \leq n, \end{aligned} \quad (3)$$

où

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x u^{a-1} (1-u)^{b-1} du, \quad 0 < u < 1, \quad (4)$$

est la *fonction Béta incomplète de Euler* ( $a > 0, b > 0$ ),

$$B(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du \quad (5)$$

la *fonction Béta de Euler*.

**Exemple 1.** Soit  $X_1, \dots, X_n$  une suite de variables aléatoires, qui suivent la même loi de Bernoulli de paramètre  $p = 0.5$  :

$$\mathbf{P}\{X_i = 1\} = \mathbf{P}\{X_i = 0\} = 0.5.$$

Notons

$$S_n = X_1 + \dots + X_n \quad \text{et} \quad \tau = \min\{k : S_k > a\},$$

où  $a$  est une constante positive.

Construisons des variables aléatoires

$$Y_n = S_{\tau+n} - S_{\tau+(n-1)}, \quad n = 1, 2, \dots$$

Il est facile de montrer que  $Y_1, Y_2, \dots, Y_n, \dots$  forment une suite de variables aléatoires indépendantes, ayant la même loi de Bernoulli de paramètre  $p = 0.5$  :

$$\mathbf{P}\{Y_n = 1\} = \mathbf{P}\{Y_n = 0\} = 0.5.$$

**Définition 3.** On dit qu'une variable aléatoire  $X$  suit la loi uniforme discrète sur l'ensemble  $\{1, 2, \dots, N\}$ , si

$$\mathbf{P}\{X = k\} = \frac{1}{N}, \quad \forall k \in \{1, 2, \dots, N\}.$$

Il est facile de montrer que

$$\mathbf{E}X = \frac{N+1}{2}, \quad \mathbf{Var}X = \frac{N^2-1}{12}.$$

**Définition 4.** On dit que la variable aléatoire discrète  $X$  suit la loi géométrique de paramètre  $p$ ,  $0 < p < 1$ , si

$$\mathbf{P}\{X = k\} = p(1-p)^k, \quad \forall k \in \{0, 1, 2, \dots\}.$$

On peut montrer que

$$\mathbf{E}X = \frac{1-p}{p}, \quad \mathbf{Var}X = \frac{1-p}{p^2},$$

et la fonction de répartition de  $X$  est

$$\mathbf{P}\{X \leq n\} = \sum_{k=0}^n p(1-p)^k = 1 - \mathbf{P}\{X \geq n+1\} =$$

$$1 - I_{1-p}(n+1, 1) = I_p(1, n+1), \quad n \in \{0, 1, \dots\}.$$

On remarque que

$$\frac{\mathbf{Var}X}{\mathbf{E}X} = \frac{1}{p} > 1.$$

**Définition 5.** On dit que la variable aléatoire discrète  $X$  suit la loi de Poisson de paramètre  $\lambda$ ,  $\lambda > 0$ , si

$$\mathbf{P}\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \{0, 1, 2, \dots\}.$$

Il est facile de montrer que

$$\mathbf{E}X = \mathbf{Var}X = \lambda,$$

et donc

$$\frac{\mathbf{Var}X}{\mathbf{E}X} = 1.$$

La fonction de répartition de  $X$  est

$$\mathbf{P}\{X \leq m\} = \sum_{k=0}^m \frac{\lambda^k}{k!} e^{-\lambda} = 1 - I_\lambda(m+1),$$

où

$$I_x(f) = \frac{1}{\Gamma(f)} \int_0^x t^{f-1} e^{-t} dt, \quad x > 0,$$

est la fonction Gamma incomplète de Euler avec  $f$  degrés de liberté,  $f > 0$ .

Pour les calculs très approximatifs quand les valeurs de  $\lambda$  sont assez grandes on peut utiliser l'approximation normale simple :

$$\mathbf{P}\{X \leq m\} = \Phi\left(\frac{m+0.5-\lambda}{\sqrt{\lambda}}\right) + O\left(\frac{1}{\sqrt{\lambda}}\right), \quad \lambda \rightarrow \infty.$$

## 0.2 Approximations normales et de Poisson

**Théorème Limite Central de Moivre-Laplace.** Soit  $\{X_n\}_{n=1}^\infty$  une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre  $p$ ,  $0 < p < 1$  :

$$\mathbf{P}\{X_i = 1\} = p, \quad \mathbf{P}\{X_i = 0\} = q = 1 - p,$$

$$\mu_n = X_1 + \dots + X_n, \quad F_n(x) = \mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leq x\right\}, \quad x \in \mathbb{R}^1.$$

Alors, uniformément par rapport à  $x$ ,  $x \in \mathbb{R}^1$ ,

$$F_n(x) \rightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad n \rightarrow \infty.$$

Du théorème limite central il suit que pour les grands valeurs de  $n$

$$\mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leq x\right\} \approx \Phi(x).$$

Souvent on utilise cette approximation avec la correction de continuité 0.5 :

$$\mathbf{P}\left\{\frac{\mu_n - np + 0.5}{\sqrt{npq}} \leq x\right\} \approx \Phi(x),$$

voir, par exemple, Greenwood & Nikulin (1996).

**Théorème de Poisson.**

Soit  $\{\mu_n\}$  une suite de variables binomiales,  $\mu_n \sim B(n, p_n)$ ,  $0 < p_n < 1$ , telle que

$$np_n \rightarrow \lambda, \quad \text{quand } n \rightarrow \infty, \quad \text{où } \lambda > 0.$$

Alors

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\mu_n = m\} = \frac{\lambda^m}{m!} e^{-\lambda}.$$

En pratique cela signifie que pour  $n$  “grand” et  $p$  “petit” on obtient l’approximation de Poisson de la loi binomiale  $B(n, p)$  par une loi de Poisson de paramètre  $\lambda = np$  :

$$\mathbf{P}\{\mu_n = m\} \approx \frac{\lambda^m}{m!} e^{-\lambda}.$$

On peut montrer (J.L. Hodges et L. Le Cam, 1968) que

$$\sup_x \left| \sum_{m=0}^x \binom{n}{m} p^m (1-p)^{n-m} - \sum_{m=0}^x \frac{\lambda^m}{m!} e^{-\lambda} \right| \leq \frac{C}{\sqrt{n}}, \quad \text{avec } C \leq 3\sqrt{\lambda}.$$

**Théorème Limite Central de Lévy.**

Soit  $\{X_n\}_{n=1}^\infty$  une suite de variables aléatoires indépendantes de même loi telle que

$$\mathbf{E}X_i = \mu \quad \text{et} \quad \mathbf{Var} X_i = \sigma^2$$

existent. Notons  $S_n = X_1 + \dots + X_n$ . Alors, uniformément par rapport à  $x \in \mathbf{R}^1$

$$\mathbf{P} \left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right\} \rightarrow \Phi(x), \quad n \rightarrow \infty.$$

**Corrolaire 1.** Dans les conditions du Théorème de Lévy on a : quelque soit  $\varepsilon > 0$

$$\begin{aligned} \mathbf{P} \left\{ \left| \frac{1}{n} \sum_{j=1}^n X_j - \mu \right| \geq \varepsilon \right\} &= \mathbf{P} \left\{ \left| \frac{S_n - n\mu}{\sigma\sqrt{n}} \right| > \frac{\varepsilon\sqrt{n}}{\sigma} \right\} \\ &\approx 2\Phi \left( -\frac{\varepsilon\sqrt{n}}{\sigma} \right). \end{aligned}$$

Par exemple, si  $\varepsilon = 3\sigma/\sqrt{n}$ , alors

$$\mathbf{P} \left\{ \left| \frac{1}{n} \sum_{j=1}^n X_j - \mu \right| \leq \varepsilon \right\} \approx 0.997,$$

si  $\varepsilon = 2\sigma/\sqrt{n}$ , alors

$$\mathbf{P} \left\{ \left| \frac{1}{n} \sum_{j=1}^n X_j - \mu \right| \leq \varepsilon \right\} \approx 0.965.$$

### 0.3 Lois continues. Liaisons entre des lois

**Définition 1.** On dit qu'une variable aléatoire  $U$  suit la loi uniforme sur  $[a, b]$ , si la densité de probabilité de  $U$  est donnée par la formule :

$$f(x; a, b) = \frac{1}{b-a} \mathbf{1}_{[a, b]}(x), \quad x \in \mathbf{R}^1.$$

La fonction de répartition de  $U$  est

$$F(x; a, b) = \mathbf{P}\{U \leq x\} = \frac{x-a}{b-a} \mathbf{1}_{[a, b]}(x) + \mathbf{1}_{]b, +\infty[}(x), \quad x \in \mathbf{R}^1.$$

Il est facile de vérifier que

$$\mathbf{E}U = \frac{a+b}{2}, \quad \mathbf{Var}U = \frac{(b-a)^2}{12}.$$

**Remarque 1.** Soit  $X$  une variable aléatoire continue. Notons  $F(x)$  sa fonction de répartition. Il est facile de vérifier que la variable aléatoire  $U = F(X)$  suit la loi uniforme sur  $[0, 1]$ . Souvent on dit que pour obtenir  $U$  on a appliquée la transformation de Smirnov.

**Définition 2.** On dit qu'une variable aléatoire  $Z$  suit la loi normale standard  $N(0, 1)$  ou réduite, si la densité de probabilité  $\varphi(x)$  de  $Z$  est donnée par la formule

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbf{R}^1. \quad (1)$$

La fonction de répartition correspondante joue un rôle important dans la suite. Aussi lui donne-t-on un nom particulier, on l'appelle  $\Phi$  :

$$\Phi(x) = \mathbf{P}\{Z \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad x \in \mathbf{R}^1. \quad (2)$$

De (2) on déduit que

$$\Phi(x) + \Phi(-x) \equiv 1, \quad x \in \mathbf{R}^1. \quad (3)$$

Soit  $x$  un nombre quelconque fixé et soit

$$p = \Phi(x), \quad 0 < p < 1. \quad (4)$$

Si nous notons  $\Psi(y) = \Phi^{-1}(y)$  la fonction inverse de  $y = \Phi(x)$ ,  $0 < y < 1$ , de (3) et (4) il résulte que

$$\Phi[\Psi(p)] \equiv p \quad \text{et} \quad \Phi[\Psi(1-p)] \equiv 1-p \quad (5)$$

pour tout  $p$ ,  $0 < p < 1$ . De plus comme

$$\Phi(-x) = 1 - \Phi(x) = 1 - p \quad \text{et} \quad -x = \Psi(1-p),$$

quand  $x = \Psi(p)$ , on en déduit que

$$\Psi(p) + \Psi(1-p) \equiv 0, \quad 0 < p < 1. \quad (6)$$

Il est connu que  $\mathbf{E}Z = 0$ ,  $\mathbf{Var}Z = 1$ .

Soit  $X = \sigma Z + \mu$ , où  $Z \sim N(0, 1)$ ,  $|\mu| < \infty$ ,  $\sigma > 0$ . Dans ce cas on dit que  $X$  suit la loi normale  $N(\mu, \sigma^2)$  de paramètres

$$\mu = \mathbf{E}X \quad \text{et} \quad \sigma^2 = \mathbf{Var}X. \quad (7)$$

La densité de  $X$  est

$$\frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbf{R}^1, \quad (8)$$

et la fonction de répartition est

$$\mathbf{P}\{X \leq x\} = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbf{R}^1. \quad (9)$$

**Définition 3.** On dit qu'une variable aléatoire  $\chi_f^2$  suit la loi de chi-deux à  $f$  degrés de liberté,  $f > 0$ , si sa densité de probabilité est donnée par la formule

$$q_f(x) = \frac{1}{2^{\frac{f}{2}} \Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-x/2} \mathbf{1}_{]0, \infty[}(x), \quad x \in \mathbf{R}^1, \quad (10)$$

où

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt, \quad a > 0 \quad (11)$$

est la fonction Gamma de Euler.

Nous allons noter  $Q_f(x) = \mathbf{P}\{\chi_f^2 \leq x\}$  la fonction de répartition de  $\chi_f^2$ . Par des calculs directs il est facile de montrer que

$$\mathbf{E}\chi_f^2 = f \quad \text{et} \quad \mathbf{Var}\chi_f^2 = 2f. \quad (12)$$

Cette définition de la loi du chi-deux n'est pas constructive. Pour construire une variable aléatoire  $\chi_n^2$ ,  $n \in \mathbb{N}^*$ , il suffit de prendre  $n$  variables aléatoires indépendantes  $Z_1, \dots, Z_n$ , qui suivent la même loi normale standard  $N(0, 1)$  et construire la statistique

$$Z_1^2 + \dots + Z_n^2.$$

On peut montrer que  $\mathbf{P}\{Z_1^2 + \dots + Z_n^2 \leq x\} = Q_n(x)$ , i.e.,

$$Z_1^2 + \dots + Z_n^2 = \chi_n^2 \quad (13)$$

suit la loi de chi-deux à  $n$  degrés de liberté. Souvent (13) on prend pour la définition de  $\chi_n^2$ . Nous allons suivre aussi cette tradition.

D'après le Théorème Limite Central il résulte que si  $n$  est assez grand alors on a l'*approximation normale* :

$$\mathbf{P}\left\{\frac{\chi_n^2 - n}{\sqrt{2n}} \leq x\right\} = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right).$$

On utilise aussi souvent pour la loi du  $\chi^2$  l'*approximation normale de Fisher*, d'après laquelle

$$\mathbf{P}\{\sqrt{2\chi_n^2} - \sqrt{2n-1} \leq x\} = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty.$$

Les meilleurs résultats donne l'approximation normale de Wilson-Hilferty :

$$\mathbf{P}\{\chi_n^2 \leq x\} = \Phi\left[\left(\sqrt[3]{\frac{x}{n}} - 1 + \frac{2}{9n}\right) \sqrt{\frac{9n}{2}}\right] + O\left(\frac{1}{n}\right), \quad n \rightarrow \infty.$$

**Définition 4.** On dit qu'une variable aléatoire  $\gamma_f$  suit la loi Gamma à  $f$  degrés de liberté ( $f > 0$ ), si pour tout  $x > 0$

$$\mathbf{P}\{\gamma_f \leq x\} = I_x(f), \quad (14)$$

où

$$I_x(f) = \frac{1}{\Gamma(f)} \int_0^x t^{f-1} e^{-t} dt \quad (15)$$

est la fonction Gamma incomplète de Euler.

Il est facile de vérifier que

$$\frac{1}{2}\chi_{2f}^2 = \gamma_f. \quad (16)$$

En effet,  $\forall x > 0$  on a

$$\mathbf{P}\left\{\frac{1}{2}\chi_{2f}^2 \leq x\right\} = \mathbf{P}\{\chi_{2f}^2 \leq 2x\} = Q_{2f}(2x) = \frac{1}{2^f \Gamma(f)} \int_0^{2x} t^{f-1} e^{-t/2} dt.$$

En faisant le changement de variable  $t = 2u$ , on trouve que

$$\mathbf{P}\left\{\frac{1}{2}\chi_{2f}^2 \leq x\right\} = \frac{1}{\Gamma(f)} \int_0^x u^{f-1} e^{-u} du = I_x(f) = \mathbf{P}\{\gamma_f \leq x\},$$

où  $\gamma_f$  est une variable aléatoire qui suit la loi gamma à  $f$  degrés de liberté. En utilisant la relation (16) on trouve que

$$\mathbf{E}\gamma_f = \mathbf{E}\frac{1}{2}\chi_{2f}^2 = f, \quad \mathbf{Var}\gamma_f = \mathbf{Var}\frac{1}{2}\chi_{2f}^2 = \frac{1}{4}\mathbf{Var}\chi_{2f}^2 = f.$$

Si  $f = 1$ , alors de (14) on déduit

$$\mathbf{P}\{\gamma_1 \leq x\} = \int_0^x e^{-t} dt = 1 - e^{-x}, \quad x > 0, \quad (17)$$

c'est-à-dire que la variable aléatoire  $\gamma_1$  suit la loi *exponentielle standard*. De cette propriété et de (16) on tire que  $\frac{1}{2}\chi_2^2$  suit la loi exponentielle standard aussi.

**Théorème 1** Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes, qui suivent la même loi exponentielle (17). Alors leur somme suit la loi gamma à  $n$  degrés de liberté, i.e.

$$X_1 + \dots + X_n = \gamma_n. \quad (18)$$

**Remarque 2.** Soit  $X$  une variable aléatoire qui suit la loi de Poisson de paramètre  $\lambda$ ,  $\lambda > 0$ . Il est facile de montrer que pour tout  $m \in \mathbb{N}$

$$\begin{aligned} \mathbf{P}\{X \leq m\} &= \mathbf{P}\{\gamma_{m+1} \geq \lambda\} = \mathbf{P}\{\chi_{2m+2}^2 \geq 2\lambda\} = \\ &= 1 - \mathbf{P}\{\chi_{2m+2}^2 \leq 2\lambda\} = 1 - Q_{2m+2}(2\lambda). \end{aligned} \quad (19)$$

En effet, soit  $\gamma_m$  une variable aléatoire qui suit la loi gamma de paramètre  $m$ . Dans ce cas la fonction de survie de  $\gamma_m$  est

$$\begin{aligned} \mathbf{P}\{\gamma_m \geq \lambda\} &= \frac{1}{\Gamma(m)} \int_{\lambda}^{\infty} x^{m-1} e^{-x} dx = \\ &= \frac{1}{\Gamma(m+1)} \int_{\lambda}^{\infty} e^{-x} dx^m = \mathbf{P}\{\gamma_{m+1} \geq \lambda\} - \frac{1}{\Gamma(m+1)} e^{-\lambda} \lambda^m, \end{aligned}$$

i.e. on a reçu que

$$\mathbf{P}\{\gamma_{m+1} \geq \lambda\} = \mathbf{P}\{\gamma_m \geq \lambda\} + \frac{\lambda^m}{m!} e^{-\lambda},$$

d'où par la récurrence il résulte que pour tout  $m \in \{0, 1, 2, \dots\}$

$$\mathbf{P}\{X \leq m\} = \sum_{k=0}^m \frac{\lambda^k}{k!} e^{-\lambda} = \mathbf{P}\{\gamma_{m+1} \geq \lambda\} = \frac{1}{\Gamma(m+1)} \int_{\lambda}^{\infty} x^m e^{-x} dx.$$

Supposons maintenant que  $\lambda$  est grand (en pratique cela signifie que  $\lambda \geq 25$ ). Comme

$$\mathbf{E}X = \mathbf{Var}X = \lambda$$

de l'inégalité de Tchebyshev il suit que nous pouvons compter que

$$m - \lambda = o(\lambda), \quad \lambda \rightarrow \infty,$$

parce que pour chaque  $m$ , qui ne vérifie pas cette condition, la probabilité  $\mathbf{P}\{X \leq m\}$  coïncide pratiquement avec 0 ou avec 1. De l'autre côté, de la relation (19) et de l'approximation normale pour la loi du chi-deux on obtient l'approximation normale de Bolshev (1963), d'après laquelle

$$\mathbf{P}\{X \leq m\} = 1 - \mathbf{P}\left\{ \frac{\chi_{2m+2}^2 - (2m+2)}{\sqrt{4m+4}} \leq \frac{2\lambda - 2m - 2}{\sqrt{4m+4}} \right\} =$$

$$1 - \Phi\left(\frac{\lambda - m - 1}{\sqrt{m+1}}\right) + O\left(\frac{1}{\sqrt{\lambda}}\right) = \Phi\left(\frac{m - \lambda + 1}{\sqrt{m+1}}\right) + O\left(\frac{1}{\sqrt{\lambda}}\right), \quad \lambda \rightarrow \infty.$$

On remarque que en utilisant l'approximation normale de Fisher pour la loi de chi-deux on obtient facilement une autre approximation normale de Bolshev :

$$\begin{aligned} \mathbf{P}\{X \leq m\} &= \mathbf{P}\{\chi_{2m+2}^2 \geq 2\lambda\} \approx 1 - \Phi(\sqrt{4\lambda} - \sqrt{4m+3}) = \\ &\Phi(\sqrt{4m+3} - 2\sqrt{\lambda}) = \Phi(\sqrt{4(m+0.5)+1} - 2\sqrt{\lambda}), \quad \lambda \rightarrow \infty. \end{aligned}$$

Le nombre 0.5 dans la dernière formule peut être considéré comme la *correction de continuité* dans cette approximation.

En pratique cela signifie que

$$\mathbf{P}\{X \leq m\} \approx \Phi\left(\sqrt{4m+1} - 2\sqrt{\lambda}\right) \approx \Phi(2\sqrt{m} - 2\sqrt{\lambda}), \quad \lambda \rightarrow \infty,$$

i.e., si  $\lambda \geq 25$ , alors la statistique  $\sqrt{4X+1}$  suit approximativement la loi normale  $N(2\sqrt{\lambda}, 1)$ . Les meilleurs résultats on obtient en utilisant l'approximation de Wilson-Hilferty, voir, par exemple, Bolshev (1963), Huber et Nikulin (1993), Nikulin (1984), d'après laquelle

$$\mathbf{P}\{X \leq m\} = \mathbf{P}\{\chi_{2m+2}^2 \geq 2\lambda\} \approx \Phi\left[3\sqrt{m+1}\left(1 - \sqrt[3]{\frac{\lambda}{m+1}} - \frac{4}{9(m+1)}\right)\right].$$

**Définition 5.** On dit que la variable aléatoire  $\beta = \beta_{a,b}$  suit la loi Béta de paramètres  $a$  et  $b$  ( $a > 0, b > 0$ ), si la densité de  $\beta$  est

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{]0,1[}(x), \quad (20)$$

où

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (21)$$

est la fonction Béta de Euler.

En notant

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt \quad (22)$$

la fonction incomplète Béta de Euler, on voit que

$$\mathbf{P}\{\beta \leq x\} = I_x(a, b), \quad 0 < x < 1, \quad (23)$$

et

$$\mathbf{P}\{\beta > x\} = 1 - I_x(a, b) = I_{1-x}(b, a), \quad 0 < x < 1.$$

Il est facile de vérifier que

$$\mathbf{E}\beta = \frac{a}{a+b}, \quad \mathbf{Var}\beta = \frac{ab}{(a+b)^2(a+b+1)}. \quad (24)$$

**Remarque 4.** Soit  $\mu_n$  une variable aléatoire Binomiale de paramètres  $n$  et  $p$ . Il est facile de montrer que pour  $m = 0, 1, \dots, n$

$$\mathbf{P}\{\mu_n \leq m\} = \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} = I_{1-p}(n-m, m+1). \quad (25)$$

**Remarque 5.** Soit  $\gamma_n$  et  $\gamma_m$  sont indépendantes. Il est utile de remarquer que les statistiques

$$\beta = \frac{\gamma_n}{\gamma_n + \gamma_m} \quad \gamma_{n+m} = \gamma_n + \gamma_m$$

sont indépendantes,  $\beta$  suit la loi bêta de paramètres  $a = n$  et  $b = m$ ,  $\gamma_{n+m}$  suit la loi gamma à  $n + m$  degrés de liberté.

**Définition 6.** Soit

$$\chi_m^2 = \frac{1}{2}\gamma_{\frac{m}{2}} \quad \text{et} \quad \chi_n^2 = \frac{1}{2}\gamma_{\frac{n}{2}}$$

indépendantes. Dans ce cas on dit que la statistique

$$F_{m,n} = \frac{\frac{1}{m}\chi_m^2}{\frac{1}{n}\chi_n^2} = \frac{n\gamma_{m/2}}{m\gamma_{n/2}} = \frac{1}{F_{n,m}} \quad (26)$$

la loi de Fisher à  $n$  et  $m$  degrés de liberté ( $m > 0$ ,  $n > 0$ ).

La fonction de répartition de  $F_{m,n}$  est

$$\mathbf{P}\{F_{m,n} \leq x\} = I_{\frac{mx}{n+mx}}\left(\frac{m}{2}, \frac{n}{2}\right), \quad x > 0. \quad (27)$$

On peut montrer que si  $n > 2$ , alors

$$\mathbf{E}F_{m,n} = \frac{n}{n-2}$$

et si  $n > 4$ , alors

$$\mathbf{Var} F_{m,n} = \frac{2n^2(n+m+2)}{m(n-2)^2(n-4)}.$$

Posant

$$F_{m,\infty} = \frac{1}{m}\chi_m^2,$$

on en tire l'approximation de Fisher, d'après laquelle pour tout  $m$  fixé

$$\mathbf{P}\{F_{m,n} \leq x\} = \mathbf{P}\{\chi_m^2 \leq mx\} + O\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty.$$

Si  $m = 1$ , on en déduit que

$$\mathbf{P}\{F_{1,\infty} \leq x\} = \mathbf{P}\{\chi_1^2 \leq x\} = 2\Phi(\sqrt{x}) - 1.$$

Cette relation nous permet de calculer les valeurs de  $\Phi(x)$  en utilisant les tables statistiques de la loi  $F$ . La relation suivante

$$F_{1,n} = \frac{\chi_1^2}{\frac{1}{n}\chi_n^2} = t_n^2 \quad (28)$$

nous montre que  $F_{1,n}$  représente le carré de la variable aléatoire  $t_n$  de Student à  $n$  degrés de liberté, d'où on tire que pour chaque  $x \in \mathbb{R}^1$

$$\mathbf{P}\{F_{1,n} \leq x^2\} = \mathbf{P}\{t_n^2 \leq x^2\} = I_{\frac{x^2}{n+x^2}}\left(\frac{1}{2}, \frac{n}{2}\right) = 2S_n(|x|) - 1, \quad (29)$$

où

$$S_n(x) = \mathbf{P}\{t_n \leq x\} = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^x \left(1 + \frac{u^2}{n}\right)^{-\frac{n+1}{2}} du \quad (30)$$

est la fonction de répartition de la variable aléatoire  $t_n$  de Student à  $n$  degrés de liberté. La variable aléatoire  $t_n$  peut être construite par la façon suivante.

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon normale,  $\mathbb{X}_i \sim N(\mu, \sigma^2)$ . On construit deux statistiques

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

représentant les meilleurs estimateurs sans biais pour  $\mu$  et  $\sigma^2$ . Alors la variable aléatoire

$$t_n = \sqrt{n-1} \frac{\bar{X}_n - \mu}{S_n} \quad (31)$$

suit la loi de Student à  $n$  degrés de liberté :

$$\mathbf{P}\{t_n \leq x\} = S_n(x), \quad x \in \mathbb{R}^1.$$

De (28) on tire que, si  $n \rightarrow \infty$ , alors, puisque

$$\frac{1}{n} \chi_n^2 \xrightarrow{\mathbf{P}} 1, \quad (32)$$

on a

$$S_n(x) = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right), \quad x \in \mathbb{R}^1,$$

i.e. pour les grandes valeurs de  $n$  la loi de Student est approximée par la loi normale standard.

Par contre, si dans (28)-(30) on pose  $n = 1$ , on en tire que la variable aléatoire  $t_1$  suit la loi de Student à 1 degré de liberté

$$\mathbf{P}\{t_1 \leq x\} = S_1(x) = \frac{1}{\pi} \int_{-\infty}^x \frac{dt}{1+t^2}, \quad x \in \mathbb{R}^1. \quad (33)$$

Cette loi est plus connue sous le nom de la loi standard de Cauchy ou tout simplement de Cauchy. Cette loi nous donne un très simple exemple d'une variable aléatoire  $t_1$ , dont l'espérance mathématique n'existe pas. Un autre exemple intéressant lié avec la loi de Cauchy est le suivant.

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon de la loi de Cauchy de densité

$$\frac{1}{\pi[1+(x-\mu)^2]}, \quad x \in \mathbb{R}^1,$$

avec le paramètre de translation  $\mu$ ,  $|\mu| < \infty$ . Dans ce cas la statistique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

suit la même loi de Cauchy que  $X_i$  et donc  $\bar{X}_n$  ne converge pas en probabilité vers  $\mu$ .

**Exercices 1.** Soit  $X$  suit la loi standard de Cauchy . Montrer que les statistiques

$$\frac{1}{X}, \quad \frac{2X}{1-X^2}, \quad \frac{3X-X^2}{1-3X^2}$$

suivent la même loi de Cauchy.

**Exercices 2.** Soient  $X$  et  $Y$  deux variables aléatoires standards normales indépendantes. Trouver la loi de  $Z = X/Y$ .

**Exercices 3.** Soit  $\mathbb{X} = (X_1, \dots, X_n)$  un échantillon,

$$\mathbf{P}\{X_i = k\} = \frac{1}{k!}e^{-1}, \quad k \in \mathbb{N},$$

i.e.  $X_i$  suit la loi de Poisson de paramètre  $\lambda = 1$ . Considérons la statistique

$$S_n = X_1 + \dots + X_n, \quad n = 1, 2, \dots$$

1. Montrer que  $S_n$  suit la loi de Poisson de paramètre  $\lambda = n$  :

$$\mathbf{P}\{S_n = k\} = \frac{n^k}{k!}e^{-n}, \quad k \in \mathbb{N},$$

en particulier

$$p_n = \mathbf{P}\{S_n = n\} = \frac{n^n}{n!}e^{-n}, \quad n \in \mathbb{N}^*.$$

2. En utilisant le théorème limite central montrer que

$$p_n \approx \Phi\left(\frac{1}{2\sqrt{n}}\right) - \Phi\left(-\frac{1}{2\sqrt{n}}\right) \approx \frac{1}{\sqrt{n}}\varphi(0) = \frac{1}{\sqrt{2\pi n}}, \quad (n \rightarrow \infty),$$

où  $\Phi(\cdot)$  est la fonction de répartition de la loi normale standard,  $\varphi(x) = \Phi'(x)$ .

3. En utilisant 1) et 2) obtenir la formule asymptotique de Stirling

$$n! \approx \sqrt{2\pi n}n^n e^{-n}, \quad (n \rightarrow \infty).$$

## 0.4 Épreuves de Bernoulli et marches aléatoires.

## 0.5 Représentation d'une suite d'épreuves de Bernoulli indépendante

Considérons une suite d'épreuves de Bernoulli indépendantes avec la probabilité de succès  $p$  ( $0 < p < 1$ ).

On peut représenter l'ensemble des résultats possibles de cette expérience à l'aide de la *marche aléatoire* d'une particule se déplaçant sur un treillis  $S$  dans le plan  $(xOy)$

$$S = \{(x, y); x \in \mathbb{N}; y \in \mathbb{N}\}. \quad (\text{voir fig. 1})$$

Donc, un résultat de l'expérience sera représenté par un *chemin* dans le *treillis*  $S$ . Si, après une épreuve, la particule se trouve au point de coordonnées  $(x, y)$ , après l'épreuve suivante elle se trouvera soit au point  $(x, y + 1)$  avec la probabilité  $p$  s'il y a eu succès, soit au point  $(x + 1, y)$  avec la probabilité  $q = 1 - p$  s'il y a eu échec parce qu'il n'y a pas d'autre possibilité.

Nous supposons que le point de départ de la particule est l'origine des axes  $O(0, 0)$ . Soit  $A_0, A_1, \dots, A_n, \dots$  la suite des points obtenus à l'issue de l'expérience,  $A_0 = O(0, 0)$ . Un chemin dans  $S$  peut être représenté par une ligne brisée reliant ces points (fig. 1).

On peut associer à cette expérience la suite  $X_1, X_2, \dots, X_n, \dots$  des variables aléatoires indépendantes de Bernoulli,

$$X_i = \begin{cases} 1, & \text{s'il y a eu succès à la } i\text{-ème épreuve,} \\ 0, & \text{s'il y a eu échec à la } i\text{-ème épreuve.} \end{cases}$$

Ces variables aléatoires sont indépendantes par construction et

$$P\{X_i = 1\} = p \quad \text{et} \quad P\{X_i = 0\} = q.$$

## 0.6 Probabilités associées à une marche aléatoire reliant 2 points du treillis $S$

Soient  $A_x$  et  $A_Y$  les points de  $S$  dont les coordonnées sont  $(x, y)$  et  $(X, Y)$  respectivement ( $0 \leq x \leq X; 0 \leq y \leq Y$ ).

Un chemin reliant  $A_x$  à  $A_X$  comporte  $(X - x)$  déplacements horizontaux et  $(Y - y)$  déplacements verticaux, chaque combinaison différente définissant un chemin différent ; le nombre de chemins possibles reliant  $A_x$  à  $A_X$  sera donc :

$$\binom{X - x + Y - y}{X - x} = \binom{X - x + Y - y}{Y - y}. \quad (1)$$

Il est évident que chacun de ces chemins a la même probabilité de réalisation égale à

$$p^{Y-y}(1-p)^{X-x}, \quad (2)$$

donc la probabilité d'arriver au point  $A_X$  en étant parti du point  $A_x$  est

$$\binom{X - x + Y - y}{X - x} p^{Y-y}(1-p)^{X-x}. \quad (3)$$

En particulier, si on part de l'origine  $A_0$ , la probabilité d'arriver en  $A_X$  est

$$\binom{X + Y}{X} p^Y(1-p)^X. \quad (4)$$

**Remarque 1.** De façon évidente, on déduit des formules précédentes que le nombre de chemins possibles pour aller de  $A_x(x, y)$  à  $A_U(u, v)$  en passant par  $A_X(X, Y)$  est égal au produit du nombre de chemins allant de  $A_x$  à  $A_X$  par le nombre de chemins allant de  $A_U$  à  $A_X$ .

## 0.7 Frontière absorbante

Nous allons nous intéresser aux expériences pour lesquelles la réalisation de la marche aléatoire est limitée (avec la probabilité 1) par une *frontière absorbante*  $B$  ( $B \subset S$ ). Cela signifie que l'expérience s'arrête **dès que** la particule a atteint la frontière. Un point  $b \in B$  est appelé *point limite* ou *point frontière*. Si un chemin atteint ce point, il s'arrête. On dit que  $b$  est une réalisation de la statistique *temps d'arrêt*.

Nous verrons plus tard que pour certaines expériences, la seule connaissance des coordonnées du point de la frontière où le chemin s'arrête nous permet d'estimer de la meilleure façon le paramètre  $p$  lorsque celui-ci est inconnu.

La frontière  $B$  est généralement définie par une équation de la forme  $y = f(x)$ . Nous allons étudier différentes frontières et leur associer des variables aléatoires connues.

## 0.8 Marches aléatoires et distributions discrètes

### Loi de Bernoulli (fig. 2)

Considérons une marche aléatoire à 1 pas dans un treillis limité par la frontière  $B$  donné par l'équation :

$$x + y = 1.$$

Dans ce cas il existe seulement 2 points limites. Si nous considérons la variable aléatoire  $X$  qui prend la valeur 1 lorsque le chemin se termine en  $A_1(0, 1)$  et la valeur 0 lorsqu'il se termine en  $A'_1(0, 1)$  nous obtenons :

$$P\{X = 1\} = p \quad \text{et} \quad P\{X = 0\} = 1 - p, \quad 0 < p < 1.$$

La variable  $X$  suit une distribution de Bernoulli de paramètre  $p$  :  $X \sim B(1, p) = B(p)$ .  
 $X$  représente le résultat d'une unique épreuve de Bernoulli.

On peut par exemple associer à cette épreuve un contrôle de qualité : on contrôle un article dans une production et on lui affecte la note 1 s'il est déféctueux, 0 s'il est bon.

### Loi Binomiale (fig. 3)

Considérons une marche aléatoire dans le treillis  $S$  commençant à l'origine et limitée par la frontière  $B$  d'équation  $x + y = n$  (le nombre de points frontières est  $n + 1$ ). Cette marche comporte  $n$  pas. Nous pouvons associer à cette marche  $n$  variables aléatoires de Bernoulli indépendantes de paramètres  $p$  :  $X_1, X_2, \dots, X_n$ .

Considérons la statistique :

$$T_n = \sum_{i=1}^n X_i.$$

Elle représente le nombre de succès au cours des  $n$  épreuves ou bien le nombre d'articles déféctueux dans un échantillon de taille  $n$  si on s'intéresse à un problème de contrôle de qualité.

Pour tout  $k = 0, 1, \dots, n$  l'événement  $\{T_n = k\}$  est équivalent à une marche aléatoire se terminant au point  $b$  de  $B$  de coordonnées  $(n - k, k)$ . Par suite d'après (4)

$$P\{T_n = k\} = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n,$$

et donc la loi de  $T_n$  est une loi binomiale de paramètres  $n$  et  $p$ ,  $T_n \sim B(n, p)$ .

#### Loi géométrique (fig. 4)

Supposons maintenant que la frontière  $B$  a pour équation  $y = 1$ . Cela signifie que la marche aléatoire s'arrête dès qu'on a obtenu le premier succès. Les points limites sont dans ce cas les points de coordonnées  $(x, 1)$ ,  $x \in \mathbb{N}$ , et la probabilité d'arriver au point  $(x, 1)$  par un chemin issu de l'origine est

$$p(1-p)^x.$$

Nous pouvons associer à cette marche la variable aléatoire  $Z$  : "rang du premier succès" ou "rang du premier article défectueux" rencontré dans le lot.

L'événement  $\{Z = k\}$ ,  $k \in \mathbb{N}^*$ , est équivalent à une marche aléatoire se terminant au point de  $B$  de coordonnées  $(k-1, 1)$  et par suite

$$P\{Z = k\} = p(1-p)^{k-1}.$$

On dit que  $Z$  suit la loi géométrique de paramètre  $p$  :  $Z \sim G(p)$ .

On peut montrer que

$$EZ = \frac{1}{p} \quad \text{et} \quad \text{Var}Z = \frac{1-p}{p^2}.$$

#### Loi binomiale négative (fig. 5)

On choisit la frontière  $B$  donné par l'équation  $y = r$ . Cela signifie que l'expérience cesse dès qu'on a obtenu le  $r$ -ème succès. Si la marche considérée comporte  $k$  étapes,

$r \leq k$   $k \in \mathbb{N}$ , on doit avoir  $k - r$  déplacements horizontaux et  $r$  déplacements verticaux mais le dernier pas est obligatoirement un déplacement vertical : le point  $(k - r, r)$  n'est accessible qu'à partir du point  $(k - r, r - 1)$  et ce passage se fait avec la probabilité  $p$ . Considérons la statistique  $S_r$ , rang du  $r$ -ème succès.

Alors

$$P\{S_r = k\} = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} p, \quad k = r, r+1, \dots$$

On dit que  $S_r$  suit la loi binomiale négative de paramètres  $r$  et  $p$ ,  $S_r \sim NB(r, p)$ .

### Remarques

1. Si  $r = 1$ , on retrouve la loi géométrique de paramètre  $p$  :  $G(p)$ .
2. Soient  $Z_1, Z_2, \dots, Z_r$   $r$  variables aléatoires indépendantes de même loi géométrique de paramètre  $p$   $Z_i \sim G(p)$ . Alors la statistique

$$S_r = \sum_{i=1}^r Z_i$$

suit de façon évidente la loi binomiale négative de paramètres  $r$  et  $p$  et on en déduit que

$$ES_r = \frac{r}{p} \quad \text{et} \quad \text{Var}S_r = \frac{r(1-p)}{p^2}.$$

3. De la même façon, on constate que si  $Z_1, \dots, Z_n$  sont  $n$  variables aléatoires indépendantes,  $Z_i \sim NB(r_i, p)$ , alors la statistique :

$$U_n = \sum_{i=1}^n Z_i$$

suit la loi binomiale négative de paramètres  $r = \sum_{i=1}^n r_i$  et  $p$ .

### Loi de Polya (fig. 6)

On choisit la frontière  $B$  donnée par l'équation  $y = x + r$ ,  $r \in N^*$ . Cela signifie qu'on arrête l'expérience dès que le nombre de succès est supérieur de  $r$  au nombre d'échecs (ou que le nombre d'articles défectueux dépasse de  $r$  le nombre d'articles bons).

Une marche issue de l'origine  $O$  et s'arrêtant au point frontière de coordonnées  $(k, r+k)$ ,  $k \in N$ , comporte donc  $(k, k+r)$  étapes mais le point  $(k+r, k)$  n'est accessible qu'à partir du point  $M(k, k+r-1)$  par un chemin qui ne doit pas avoir encore rencontré la frontière. Le nombre de chemins allant de  $O$  à  $M$  et qui touchent ou coupent la frontière peut être calculé de la façon suivante : lorsque le chemin touche la frontière  $B$  pour la première fois on prend son symétrique par rapport à  $B$  : c'est un chemin qui arrive au point  $M'(k-1, k+r)$  (symétrique de  $M$  par rapport à  $B$ ). Le nombre de chemins reliant  $O$  à  $M'$  est égale à

$$\binom{2k+r-1}{k-1}$$

et le nombre de chemins reliant  $O$  à  $M$  est égale à

$$\binom{2k+r-1}{k},$$

d'où on déduit donc que le nombre de réalisations possibles de la marche considérée est égale à

$$\binom{2k+r-1}{k} - \binom{2k+r-1}{k-1} = \frac{(2k+r-1)!}{k!(k+r)!} (k+r-k) = \frac{r}{2k+r} \binom{2k+r}{k}.$$

Si nous associons à cette marche la variable  $V_r$  : rang de l'épreuve pour laquelle le nombre de succès est pour la première fois supérieur de  $r$  au nombre d'échecs, alors l'événement  $\{V_r = v\}$  est équivalent à une marche partant de l'origine et comportant  $v$  étapes :

$v - r/2$  déplacements horizontaux et  $v - r/2$  déplacements verticaux.

De façon évidente on doit avoir  $v \geq r$  et  $v - r \in 2N$ , c'est-à-dire  $v = 2k + r$ ,  $k \in N$ .

Dans ce cas, pour  $r > 0$  on a :

$$P\{V_r = v\} = P\{V_r = 2k + r\} = \frac{r}{2k+r} \binom{2k+r}{k} p^{k+r} (1-p)^k.$$

Examinons le cas  $r = 0$ . Nous devons dans ce cas considérer les chemins partant non plus de l'origine  $O$  mais du point  $A_1(1, 0)$ .

Un raisonnement analogue du précédent nous montre alors que

$$P\{V_0 = 2k\} = \left[ \binom{2k-2}{k-1} - \binom{2k-2}{k} \right] [p(1-p)]^k = 2(k-1) \binom{2k-1}{k} [p(1-p)]^k.$$

**Loi hypergéométrique (fig. 7)**

Soient  $N$  et  $M$  deux entiers positifs fixés et  $0 \leq M \leq N$ .

Considérons une marche aléatoire dans le treillis  $S$  limitée par la frontière  $B : x + y = N$ . Nous nous intéressons plus particulièrement à la marche aléatoire partant de l'origine et atteignant le point  $B$  de coordonnées  $(N - M, M)$ . Soit

$$T_n = \sum_{i=1}^n X_i, \quad \text{où } X_i \sim B(p),$$

les  $X_i$  étant indépendantes, et donc  $T_n \sim B(n, p)$ . Nous savons que  $T_N = M$  et il est intéressant de savoir comment cette information influe sur la distribution de la statistique  $T_n, n < N$ . C'est-à-dire que, sachant que la marche a atteint le point  $(N - M, M)$ , nous allons évaluer la probabilité pour qu'après  $n$  pas elle soit à un point donné de la frontière

$$\beta : x + y = n.$$

Nous cherchons donc la probabilité :

$$P\{T_n = k | T_N = M\} = \frac{P\{T_n = k; T_N = M\}}{P\{T_N = M\}},$$

où

$$\text{Max}(0, n + M - N) \leq k \leq \text{Min}(n, M).$$

On sait que :

$$\begin{aligned} P\{T_n = k; T_N = M\} &= \binom{n}{k} \binom{N-n}{M-k} p^k (1-p)^{n-k} \cdot p^{N-k} (1-p)^{N-n} = \\ &= \binom{n}{k} \binom{N-n}{M-k} p^M (1-p)^{N-M} \end{aligned}$$

et

$$P\{T_N = M\} = \binom{N}{M} p^M (1-p)^{N-M}.$$

Par suite, la probabilité cherchée est égale à

$$P\{T_n = k | T_N = M\} = \frac{\binom{n}{k} \binom{N-n}{M-k}}{\binom{N}{M}} = \frac{\binom{N-M}{n-k} \binom{M}{k}}{\binom{N}{n}},$$

où

$$1 \leq n \leq N, \quad 1 \leq M \leq N, \quad \text{Max}(0, n + M - N) \leq k \leq \text{Min}(n, M).$$

Cette loi conditionnelle de  $T_n$  est la loi hypergéométrique  $H(N, M, n)$  de paramètres  $N, M$  et  $n$ . On peut remarquer qu'elle ne dépend pas du paramètre  $p$ .

On peut montrer que si  $X$  suit une loi  $H(N, M, n)$ , alors

$$\mathbf{E}X = \frac{nM}{N} \quad \text{et} \quad \mathbf{Var}X = \frac{n(N-n)M(N-M)}{N^2(N-1)}.$$



# Chapitre 1

## QUELQUES PROBLÈMES CLASSIQUES DE LA STATISTIQUE MATHÉMATIQUE.

### 1.1 Problèmes d'estimation et de comparaison des probabilités de succès.

#### Exemple 1. Estimation de la probabilité dans le schéma d'expériences de Bernoulli.

On a coutume de considérer l'hypothèse  $H_0 : p = 0.5$  selon laquelle la probabilité de la naissance d'un garçon est la même que celle d'une fille. On possède beaucoup de données statistiques pour sa vérification. Nous utiliserons ceux qui ont été données sur la Suisse : entre 1871 et 1900 naquirent en Suisse  $n = 2644757$  enfants et parmi eux

$$\mu_n = 1359671 \quad \text{garçons et} \quad n - \mu_n = 1285086 \quad \text{filles.}$$

Est-ce que ces données confirment l'hypothèse  $H_0 : p = 0.5$  ?

Nommons *succès* (!) la naissance d'un garçon et posons la question autrement en utilisant le schéma d'expériences de Bernoulli avec la probabilité de succès  $p$ . L'hypothèse  $H_0 : p = 0.5$  concorde-t-elle avec le fait que dans la série de  $n = 2644757$  expériences la fréquence de "succès" soit égale à

$$\frac{\mu_n}{n} = \frac{1359671}{2644757} = 0.5141?$$

Il est évident que si au lieu de l'hypothèse  $H_0 : p = 0.5$  on avait pris une autre hypothèse  $H_1 : p = 0.1$ , par exemple, alors cette hypothèse  $H_1$  serait rejetée par tous comme une hypothèse peu probable (ou même impossible). La question est : sur quoi est basée cette décision ?

La réponse peut être donnée puisqu'on sait que l'estimateur

$$\hat{p}_n = \frac{\mu_n}{n}$$

de la probabilité  $p$ ,  $p \in ]0, 1[$ , est basé sur la statistique  $\mu_n$  qui suit une loi binomiale  $B(n, p)$

$$\mathbf{P}\{\mu_n = k|p\} = \mathbf{P}_p\{\mu_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

d'où on tire que

$$\mathbf{E}_p \mu_n = np, \quad \mathbf{Var} \mu_n = np(1-p),$$

et par conséquent pour tout  $p \in ]0, 1[$

$$\mathbf{E}_p \frac{\mu_n}{n} = p \quad \text{et} \quad \mathbf{Var} \frac{\mu_n}{n} = \frac{p(1-p)}{n}.$$

De l'inégalité de Tchebyshev il suit que pour tout  $\varepsilon > 0$

$$\mathbf{P}_p\{|\hat{p}_n - p| > \varepsilon\} \rightarrow 0, \quad \text{quand } n \rightarrow \infty. \quad (1)$$

Nous disons que  $\{\hat{p}_n\}$  est une suite *consistante* (*cohérente*) d'estimateurs sans biais du paramètre  $p$ , puisque

$$\mathbf{E}_p \hat{p}_n = p \quad \text{et} \quad \hat{p}_n \xrightarrow{\mathbf{P}_p} p.$$

La relation (1) on peut préciser, notamment, pour tout  $\lambda > 0$  on a :

$$\mathbf{P}_p\{|\hat{p}_n - p| < \lambda \sqrt{\frac{p(1-p)}{n}}\} \geq 1 - \frac{1}{\lambda^2}.$$

En particulier, si  $\lambda = 2$ , on en tire que

$$\mathbf{P}_p\{|\hat{p}_n - p| < \frac{1}{\sqrt{n}}\} \geq 0.75.$$

En utilisant l'approximation normale, basée sur le théorème limite central de de Moivre-Laplace, on a

$$\lim_{n \rightarrow \infty} \mathbf{P}_p \left\{ \frac{\frac{\mu_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq x \right\} = \Phi(x) \quad \text{pour tout } x \in \mathbf{R}^1, \quad (2)$$

où

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad -\infty < x < \infty. \quad (3)$$

En prenant  $\alpha$  assez petit,  $0 < \alpha < 0.5$ ), (on va appeler ce nombre  $\alpha$  le *niveau de signification*, on peut affirmer, par exemple, que

$$\mathbf{P}_p \left\{ -\bar{x}_{\alpha/2} \leq \sqrt{\frac{n}{p(1-p)}} \left( \frac{\mu_n}{n} - p \right) \leq \bar{x}_{\alpha/2} \right\} \approx 1 - \alpha, \quad (4)$$

où le nombre  $\bar{x}_{\alpha/2}$  est donné par

$$\Phi(\bar{x}_{\alpha/2}) = 1 - \alpha/2. \quad (5)$$

La quantité  $\bar{x}_{\alpha/2}$  s'appelle *quantile supérieur de niveau  $\alpha/2$*  de la loi normale standard.

Par exemple,

$\bar{x}_{\alpha/2} = 3$  est le quantile supérieur de niveau  $\alpha/2 = 0.00135$ ,

$$\mathbf{P}_p \left\{ \left| \sqrt{\frac{n}{p(1-p)}} \left( \frac{\mu_n}{n} - p \right) \right| > 3 \right\} \approx 0.0027 = \alpha,$$

tandis que

le quantile  $\bar{x}_{\alpha/2} = 4$  correspond déjà à  $\alpha/2 = 0.00003167 (= 0.0000)$ ,

d'où on tire que

$$\mathbf{P} \left\{ \left| \sqrt{\frac{n}{p(1-p)}} \left( \frac{\mu_n}{n} - p \right) \right| > 4 \right\} \approx 0.000063,$$

(en pratique cette probabilité = 0.000) et

$$\mathbf{P} \left\{ \left| \sqrt{\frac{n}{p(1-p)}} \left( \frac{\mu_n}{n} - p \right) \right| \leq 4 \right\} \approx 0.999937$$

(en pratique cette probabilité = 1).

Revenons à nos données et à l'hypothèse  $H_0$ . L'hypothèse  $H_0$  suppose que  $p = 0.5$  et donc sous  $H_0$  on a :

$$\sqrt{\frac{n}{p(1-p)}} \left( \frac{\mu_n}{n} - p \right) = 2\sqrt{n} \left( \frac{\mu_n}{n} - \frac{1}{2} \right).$$

Tout d'abord on remarque qu'il y a 3 contrehypothèses naturelles pour  $H_0$  :

$$H_1 : p \neq 0.5, \quad H_1^+ : p > 0.5, \quad H_1^- : p < 0.5$$

qui sont en concurrence avec  $H_0$ . Il est naturel de dire que l'intervalle

$$S = [-\bar{x}_{\alpha/2}, \bar{x}_{\alpha/2}] \subset \mathbf{R}^1$$

représente l'ensemble des valeurs de la statistique

$$T_n = T(\mu_n) = 2\sqrt{n} \left( \frac{\mu_n}{n} - \frac{1}{2} \right),$$

qui sont favorable à l'hypothèse  $H_0$ , tandis que l'ensemble

$$K = \mathbf{R}^1 \setminus S = K_1^- \cup K_1^+ = ]-\infty, -\bar{x}_{\alpha/2}[ \cup ]\bar{x}_{\alpha/2}, \infty[,$$

appelé la *région critique* pour  $H_0$ , représente l'ensemble des valeurs de la statistique  $T_n$ , qui sont favorable à  $H_1$ . Par contre, l'ensemble  $S$  s'appelle la *région d'acceptation* de l'hypothèse  $H_0$ .

On remarque que

$$\mathbf{P}\{T_n \in S | H_0\} \approx 1 - \alpha, \quad \mathbf{P}\{T_n \in K | H_0\} \approx \alpha.$$

Il est clair que l'événement

$$\{T_n \in K_1^-\} \subset \{T_n \in K\}$$

est favorable à  $H_1^-$ , et l'événement

$$\{T_n \in K_1^+\} \subset \{T_n \in K\}$$

est favorable à  $H_1^+$ , et que

$$\mathbf{P}\{T_n \in K_1^- | H_0\} = \mathbf{P}\{T_n \in K_1^+ | H_0\} \approx \frac{\alpha}{2}.$$

Dans notre cas pour les données de Suisse nous constatons que

$$T_n = T(\mu_n) = 2\sqrt{n} \left( \frac{\mu_n}{n} - \frac{1}{2} \right) = \sqrt{\frac{2644757}{0.5 \cdot 0.5}} (0.5141 - 0.5) = 45.86 > 4,$$

i.e. l'événement  $\{T_n > 4\}$  est apparu. La valeur observée de  $T_n$  est très supérieure à la *valeur critique*  $\bar{x}_{\alpha/2} = 4$ , correspondant au niveau de signification  $\alpha/2 = 0.00003167$ , qui est égal pratiquement à 0, et donc ce phénomène doit être considéré comme impossible sous l'hypothèse  $H_0 : p = 0.5$ . Que devons nous faire ? Il faut évidemment rejeter l'hypothèse  $H_0 : p = 0.5$  en faveur de  $H_1$ , puisque  $T_n \in K$ . Nous disons que l'hypothèse  $H_0$  ne concorde pas avec les données observées. En plus comme dans l'expérience on a observé l'événement  $\{T_n \in K_1^+\}$ , il est raisonnable d'accepter l'hypothèse  $H_1^+$ . Comme estimateur de la valeur inconnue de  $p$  sous l'hypothèse  $H_1^+$  il est recommandé de prendre  $\hat{p}_n = 0.514$ .

Enfin de (4) on tire que

$$\mathbf{P}\left\{ \frac{\mu_n}{n} - \bar{x}_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{\mu_n}{n} + \bar{x}_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right\} \approx 1 - \alpha,$$

c'est-à-dire pour les grandes valeurs de  $n$  on obtient l'*intervalle de confiance* pour  $p$  avec le coefficient de confiance  $P \approx 1 - \alpha$  :

$$\mathbf{P}\left\{ \frac{\mu_n}{n} - \bar{x}_{\alpha/2} \frac{1}{2\sqrt{n}} \leq p \leq \frac{\mu_n}{n} + \bar{x}_{\alpha/2} \frac{1}{2\sqrt{n}} \right\} \approx 1 - \alpha \quad (= 0.9973 \quad \text{si} \quad \bar{x}_{\alpha/2} = 3).$$

Si, par exemple,

$$\frac{\alpha}{2} = 0.00135 \quad \text{i.e.} \quad \alpha = 0.0027,$$

dans ce cas  $\bar{x}_{\alpha/2} = 3$  et d'après nos données on obtient la réalisation de l'intervalle de confiance

$$0.5141 - 0.0003\bar{x}_{\alpha/2} \leq p \leq 0.5141 + 0.0003\bar{x}_{\alpha/2},$$

i.e.

$$0.5132 \leq p \leq 0.5150.$$

**Remarque 1.** On remarque que

$$\Phi(0) = 0.500000, \quad \Phi(1) = 0.841345, \quad \Phi(1.6) = 0.945201, \quad \Phi(2) = 0.97725,$$

$$\Phi(2.6) = 0.995339, \quad \Phi(3) = 0.998650, \quad \Phi(4) = 0.999968,$$

où  $\Phi(x)$  est donnée par (3), i.e.

$$0 = \bar{x}_{0.5}, \quad 1 = \bar{x}_{0.158655}, \quad 1.6 = \bar{x}_{0.054799}, \quad 2 = \bar{x}_{0.02275}, \quad \dots$$

**Exemple 2.** K. Pearson a jeté une pièce symétrique  $n = 24000$  fois et a observé

$$\mu_n = 12012$$

succès. On sait que

$$\hat{p}_n = \frac{\mu_n}{n}$$

est un bon estimateur pour la probabilité de succès  $p = 0.5$  (on a supposé que la pièce est symétrique c'est-à-dire l'hypothèse  $H_0 : p = 0.5$ ). Dans notre cas  $\hat{p}_n = 0.5005$ . Nous savons que

$$\mathbf{E}\hat{p}_n = 0.5 \quad \text{et} \quad \mathbf{Var}\hat{p}_n = \frac{1}{4n}.$$

En étudiant le résultat de l'expérience de K. Pearson, nous pouvons constater que la statistique  $\mu_n$  a pris une valeur très proche de sa moyenne  $\mathbf{E}\mu_n = np = 12000$ . Est-ce vraisemblable ou non ? On note que sous l'hypothèse  $H_0 : p = 0.5$  on a

$$\mathbf{Var}\mu_n = np(1-p) = \frac{n}{4},$$

et comme l'écart-type de  $\mu_n$  est

$$\sqrt{\mathbf{Var}\mu_n} = \sqrt{np(1-p)} = 77.5,$$

on pourrait donner plusieurs raisons à l'apparition de l'événement

$$\left\{ \left| \mu_n - \frac{n}{2} \right| > 77.5 \right\} = \left\{ |\mu_n - 12000| > 77.5 \right\}$$

Mais dans son expérience K. Pearson a obtenu

$$|\mu_n - 12000| = 12 \ll 77.5.$$

On pourrait penser que c'est trop beau pour être vrai. Quelle est donc la probabilité d'observer l'événement  $\left\{ \left| \mu_n - \frac{n}{2} \right| \leq 12 \right\}$  sous l'hypothèse  $H_0$  ?

On a

$$\begin{aligned} \mathbf{P}\left\{ \left| \mu_n - \frac{n}{2} \right| \leq 12 \mid H_0 \right\} &= \mathbf{P}\left\{ \frac{\left| \mu_n - \frac{n}{2} \right|}{\sqrt{np(1-p)}} \leq \frac{12}{77.5} \mid H_0 \right\} \approx \\ &\approx \Phi(0.155) - \Phi(-0.155) \approx 0.124 = \frac{1}{8}. \end{aligned}$$

Il est évident que cet événement est bien probable, donc K. Pearson pouvait observer ce résultat.

**Exemple 3.** Supposons que nous avons un générateur de nombres aléatoires et que ce générateur nous fournit les "nombres aléatoires"  $x_1, x_2, \dots, x_n$  qu'on peut considérer (hypothèse  $H_0$ ) comme des réalisations de variables aléatoires indépendantes

$$X_1, X_2, \dots, X_n,$$

ayant chacune la distribution discrète uniforme sur l'ensemble  $S = \{0, 1, \dots, 9\}$  i.e.,

$$\mathbf{P}\{X_j = i \mid H_0\} = 0.1, \quad i \in S. \quad (6)$$

Considérons maintenant un échantillon  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  de taille  $n = 10\,000$ , associé au générateur de nombres aléatoires mentionné précédemment. Nous désirons tester l'hypothèse  $H_0$  que l'échantillon  $\mathbb{X}$  est issu de la distribution uniforme (1) si dans notre échantillon on a observé seulement 4999 fois  $x_i$  ne dépassant pas 4. Quel niveau de signification doit on avoir pour rejeter  $H_0$  ?

**Solution.** Soit

$$\mu_n = \#\{X_i \leq 4\}. \quad (7)$$

On remarque que

$$\mathbf{P}\{X_i \leq 4 | H_0\} = 0.5.$$

D'après nos données :

$$\hat{p}_n = \frac{\mu_n}{n} = \frac{4999}{10\,000}$$

qui est très voisin de 0.5. Par ailleurs, sous l'hypothèse  $H_0$ , la statistique  $\mu_n$  suit une distribution binomiale  $B(n, p)$  de paramètres  $n = 10\,000$ ,  $p = 0.5$  et donc sous  $H_0$

$$\mathbf{E}\mu_n = np = 5000 \quad \text{and} \quad \mathbf{Var}\mu_n = np(1-p) = 2500. \quad (8)$$

D'où pour tout  $x = 1, 2, \dots$ , d'après le théorème de de Moivre-Laplace, nous avons (avec la correction de continuité de 0.5)

$$\begin{aligned} \mathbf{P}\left\{ \left| \mu_n - np \right| \leq x \mid H_0 \right\} &= \mathbf{P}\left\{ \frac{n}{2} - x \leq \mu_n \leq \frac{n}{2} + x \mid H_0 \right\} \approx \\ &\Phi\left(\frac{0.5n + x + 0.5 - 0.5n}{\sqrt{n \cdot 0.5 \cdot 0.5}}\right) - \Phi\left(\frac{0.5n - x - 0.5 - 0.5n}{\sqrt{n \cdot 0.5 \cdot 0.5}}\right) = 2\Phi\left(\frac{2x + 1}{\sqrt{n}}\right) - 1. \end{aligned} \quad (9)$$

Notons  $\alpha$  le niveau de signification du test ( $0 < \alpha < 0.5$ ) avec la *région critique* :

$$\left\{ \left| \mu_n - \frac{n}{2} \right| \leq \bar{x}_{\alpha/2} \right\} = \left\{ \frac{n}{2} - \bar{x}_{\alpha/2} \leq \mu_n \leq \frac{n}{2} + \bar{x}_{\alpha/2} \right\}. \quad (10)$$

Alors, à la valeur critique  $\bar{x}_{\alpha/2}$ , correspond le niveau de signification  $\alpha$  :

$$\alpha \approx 2\Phi\left(\frac{2\bar{x}_{\alpha/2} + 1}{\sqrt{n}}\right) - 1, \quad (n = 10000). \quad (11)$$

En particulier, si  $\bar{x}_{\alpha/2} = 1$ , alors

$$\alpha \approx 2\Phi\left(\frac{3}{\sqrt{n}}\right) - 1 = 2\Phi(0.03) - 1 = 2 \cdot 0.512 - 1 = 0.024.$$

*Inférence statistique* : d'après le test statistique, basé sur la région critique :

$$\{ \left| \mu_n - 5000 \right| \leq 1 \},$$

l'hypothèse  $H_0$  sera rejetée avec le niveau de signification  $\alpha \approx 0.025$ , puisque

$$\mathbf{P}\{ \left| \mu_n - 5000 \right| \leq 1 | H_0 \} \approx 0.024 < \alpha = 0.025.$$

(Voir, aussi, Cuadras C., Nikulin (1993)).

**Exemple 4. Le problème du Chevalier de Méré.** D'abord on considère l'épreuve suivante : on jette 4 fois un dé.

Soit  $A$  l'événement :

$$A = \{ \text{obtenir au moins une fois le 1 au cours de cette expérience} \}.$$

Considérons ensuite la deuxième expérience qui consiste à jeter 24 fois 2 dés.  
Soit  $B$  l'événement :

$$B = \{\text{obtenir au moins une fois le } (1,1) \text{ au cours de cette expérience}\}.$$

Le Chevalier de Méré ayant supposé que

$$p_1 = \mathbf{P}(A) < p_2 = \mathbf{P}(B)$$

avait misé sur  $B$ . Avait-il raison ?

On remarque que

$$p_1 = \mathbf{P}(A) = 1 - \left(\frac{5}{6}\right)^4 = 0.5177,$$

$$p_2 = \mathbf{P}(B) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914.$$

Mais Méré ne pouvait pas faire ces calculs. Par contre, il aurait pu faire une expérience pour résoudre ce problème par des méthodes statistiques, basées sur la loi des grands nombres.

Soient  $\mu_n^{(1)} = \mu_n(A)$  et  $\mu_n^{(2)} = \mu_n(B)$  les résultats de la modélisation de ces expériences lorsqu'on les a répété  $n = 25, 50, 100, 250$  fois chacune.

$n$	25	50	100	250
$\mu_n(A)$	18	27	52	121
$\mu_n(B)$	14	24	47	126

Ici  $\mu_n(A)$  et  $\mu_n(B)$  représentent les nombres de succès dans la première et la seconde expériences respectivement.

D'après la loi des grands nombres

$$\hat{p}_{1n} = \frac{\mu_n^{(1)}}{n} \xrightarrow{P} p_1 \quad \hat{p}_{2n} = \frac{\mu_n^{(2)}}{n} \xrightarrow{P} p_2, \quad (n \rightarrow \infty),$$

c'est-à-dire il y a la consistance de deux suites d'estimateurs  $\{\hat{p}_{1n}\}$  et  $\{\hat{p}_{2n}\}$  de paramètres  $p_1$  et  $p_2$ . En plus on sait que

$$\mathbf{E}\hat{p}_{1n} = \frac{1}{n}\mathbf{E}\mu_n^{(1)} = p_1, \quad \mathbf{E}\hat{p}_{2n} = \frac{1}{n}\mathbf{E}\mu_n^{(2)} = p_2,$$

donc pour tout  $n \in \mathbb{N}^*$   $\hat{p}_{1n}$  et  $\hat{p}_{2n}$  sont les estimateurs sans biais pour  $p_1$  et  $p_2$  respectivement. Enfin, on remarque, que quand  $n \rightarrow \infty$

$$\mathbf{Var} \hat{p}_{1n} = \frac{p_1(1-p_1)}{n} \rightarrow 0, \quad \mathbf{Var} \hat{p}_{2n} = \frac{p_2(1-p_2)}{n} \rightarrow 0.$$

En utilisant les résultats de modélisation du jeu on obtient une nouvelle table

$n$	25	50	100	250
$\frac{\mu_n^{(1)}}{n}$	0.72	0.54	0.52	0.484
$\frac{\mu_n^{(2)}}{n}$	0.56	0.48	0.47	0.504

Il faut noter que bien que  $p_1$  soit supérieur à  $p_2$  l'expérience nous donne ici

$$\mu_n^{(1)} = 121 < \mu_n^{(2)} = 126 \quad \text{pour } n = 250,$$

et donc

$$\frac{\mu_n^{(1)}}{n} = 0.484 < \frac{\mu_n^{(2)}}{n} = 0.504 \quad \text{pour } n = 250.$$

Si on arrête "le jeu" à  $n = 250$ , on aura une conclusion erronée que  $p_1 < p_2$ . On va évaluer

$$\mathbf{P}\{\mu_n^{(1)} < \mu_n^{(2)}\}$$

la probabilité d'événement  $\{\mu_n^{(1)} < \mu_n^{(2)}\}$ . Notons

$$X_n = \frac{\mu_n^{(1)} - np_1}{\sqrt{np_1(1-p_1)}}, \quad Y_n = \frac{\mu_n^{(2)} - np_2}{\sqrt{np_2(1-p_2)}}, \quad n \in \mathbb{N}^*.$$

Pour tout  $n$  les variables aléatoires  $X_n$  et  $Y_n$  sont indépendantes, et

$$\mathbf{E}X_n = \mathbf{E}Y_n = 0, \quad \mathbf{Var} X_n = \mathbf{Var} Y_n = 1.$$

En plus, du théorème de de Moivre-Laplace il suit que pour tout  $x \in \mathbb{R}^1$

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X_n \leq x\} = \lim_{n \rightarrow \infty} \mathbf{P}\{Y_n \leq x\} = \Phi(x),$$

où

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

De ce résultat il suit que

$$\frac{X_n - Y_n}{\sqrt{\mathbf{Var}(X_n - Y_n)}} = \frac{(\mu_n^{(1)} - \mu_n^{(2)}) + n(p_2 - p_1)}{\sqrt{np_1(1-p_1) + np_2(1-p_2)}}$$

est aussi asymptotiquement normale quand  $n \rightarrow \infty$ ,

$$\mathbf{P}\left\{\frac{X_n - Y_n}{\sqrt{\mathbf{Var}(X_n - Y_n)}} \leq x\right\} \approx \Phi(x), \quad x \in \mathbb{R}^1.$$

Maintenant nous sommes capable d'évaluer la probabilité de l'événement  $\{\mu_n^{(1)} < \mu_n^{(2)}\}$ .  
En effet,

$$\begin{aligned} \mathbf{P}\{\mu_n^{(1)} < \mu_n^{(2)}\} &= \mathbf{P}\{\mu_n^{(1)} - \mu_n^{(2)} < 0\} = \\ \mathbf{P}\left\{\frac{\mu_n^{(1)} - \mu_n^{(2)} + n(p_2 - p_1)}{\sqrt{np_1(1-p_1) + np_2(1-p_2)}} < \frac{\sqrt{n}(p_2 - p_1)}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}\right\} &\approx \\ \approx \Phi\left(\frac{\sqrt{n}(p_2 - p_1)}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}\right) &\rightarrow 0, \quad n \rightarrow \infty, \quad \text{si } p_2 < p_1. \end{aligned}$$

On remarque qu'en utilisant les tables statistiques on peut calculer cette probabilité pour

$n = 25, 50, 100, 250$  et  $1000$  et pour  $p_1 = 0.5177$  et  $p_2 = 0.4914$  :

$n$	25	50	100	250	1000
$\mathbf{P}\{\mu_n^{(1)} < \mu_n^{(2)}\}$	0.42	0.39	0.35	0.18	0.12

On constate que même pour  $n$  assez grand ( $n = 1000$ ) on a 12 pour cent de chances de faire une conclusion erronée, et on comprend le trouble du Chevalier.

**Exemple 5. Comparaison de deux probabilités.** On veut comparer la qualité de production de deux usines qui produisent le même article. Soit  $p_1$  (respectivement  $p_2$ ) la probabilité qu'un article de la 1<sup>ère</sup> usine (respectivement de la 2<sup>ème</sup>) soit défectueux. Pour effectuer le contrôle on a prélevé  $n_1$  articles dans la première usine et  $n_2$  articles de la seconde. Soit  $\mu_{n_1}$  (respectivement  $\mu_{n_2}$ ) le nombre d'articles défectueux pour la première (respectivement pour la seconde) usine. Supposons que nous voulions tester l'hypothèse d'homogénéité

$$H_0 : p_1 = p_2 = p, \quad p \in ]0, 1[.$$

Sous l'hypothèse  $H_0$  on a

$$\begin{aligned} \mathbf{E} \frac{\mu_{n_1}}{n_1} &= \mathbf{E} \frac{\mu_{n_2}}{n_2} = p, \\ \mathbf{Var} \frac{\mu_{n_1}}{n_1} &= \frac{p(1-p)}{n_1} \rightarrow 0, \quad (n_1 \rightarrow \infty), \\ \mathbf{Var} \frac{\mu_{n_2}}{n_2} &= \frac{p(1-p)}{n_2} \rightarrow 0, \quad (n_2 \rightarrow \infty). \end{aligned}$$

Donc, sous l'hypothèse  $H_0$  on a deux suites consistantes  $\{\hat{p}_{1n}\}$  et  $\{\hat{p}_{2n}\}$  d'estimateurs sans biais pour le paramètre  $p$ . On remarque que quels que soient  $n_1$  et  $n_2$  les estimateurs  $\hat{p}_{1n}$  et  $\hat{p}_{2n}$  sont indépendants.

En général, même si l'hypothèse  $H_0$  est vraie, dans l'expérience on observe l'événement

$$\left\{ \frac{\mu_{n_1}}{n_1} \neq \frac{\mu_{n_2}}{n_2} \right\}.$$

Il est évident que pour tester  $H_0$  contre l'alternative  $H_1 : p_1 \neq p_2$  il est raisonnable d'utiliser la statistique

$$\left| \frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2} \right|$$

comme l'estimateur de  $|p_1 - p_2|$  et rejeter  $H_0$  si

$$\left| \frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2} \right| \geq c_\alpha,$$

où il faut choisir la valeur critique  $c_\alpha$  de façon que

$$\mathbf{P} \left\{ \left| \frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2} \right| \geq c_\alpha \mid H_0 \right\} \geq \alpha, \quad 0 < \alpha < 0.5.$$

Par contre, si

$$\left| \frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2} \right| < c_\alpha,$$

on accepte  $H_0$ . On remarque que

$$\mathbf{P} \left\{ \left| \frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2} \right| < c_\alpha \mid H_0 \right\} \geq 1 - \alpha.$$

Comment trouver la valeur critique  $c_\alpha$ , correspondant au niveau de signification  $\alpha$ ? Pour  $n_1$  et  $n_2$  suffisamment grands on peut s'attendre à ce que la variable aléatoire

$$\frac{\frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2}}{\sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

soit approximativement normale, puisque

$$\lim_{\min(n_1, n_2) \rightarrow \infty} \mathbf{P} \left\{ \frac{\frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2}}{\sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq x \mid H_0 \right\} = \Phi(x).$$

Donc, en choisissant  $c_\alpha = \bar{x}_{\alpha/2}$  on a

$$\mathbf{P} \left\{ \left| \frac{\frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2}}{\sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \right| \geq \bar{x}_{\alpha/2} \mid H_0 \right\} \approx \alpha,$$

et, par conséquent, on rejette  $H_0$  en faveur de  $H_1$ , si

$$\frac{\left| \frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2} \right|}{\sqrt{\frac{\mu_n}{n} \left( 1 - \frac{\mu_n}{n} \right) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \geq \bar{x}_{\alpha/2},$$

où

$$\frac{\mu_n}{n} = \frac{\mu_{n_1} + \mu_{n_2}}{n_1 + n_2} = \hat{p}_n$$

est le meilleur estimateur sans biais pour  $p$  sous l'hypothèse  $H_0$ .

Il est évident que

$$\mathbf{P} \left\{ \left| \frac{\frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2}}{\sqrt{\frac{\mu_n}{n} \left( 1 - \frac{\mu_n}{n} \right) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \right| > \bar{x}_{\alpha/2} \mid H_0 \right\} \approx \alpha,$$

quand  $n_1$  et  $n_2$  sont suffisamment grands.

**Remarque 1.** Il est clair que si nous voulons tester l'hypothèse  $H_0 : p_1 = p_2$  contre l'hypothèse unilatérale  $H_1^+ : p_1 > p_2$ , dans ce cas il faut choisir  $c_\alpha = \bar{x}_\alpha$  et rejeter  $H_0$  si

$$\frac{\frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2}}{\sqrt{\frac{\mu_n}{n} \left( 1 - \frac{\mu_n}{n} \right) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \geq \bar{x}_\alpha,$$

où  $\Phi(\bar{x}_\alpha) = 1 - \alpha$ . Le niveau de ce test unilatéral  $\approx \alpha$ .

**Remarque 2.** Si nous voulons tester  $H_0 : p_1 = p_2$  contre l'alternative  $H_1^- : p_1 < p_2$ , qui est unilatérale, il faut rejeter  $H_0$  si

$$\frac{\frac{\mu_{n_1}}{n_1} - \frac{\mu_{n_2}}{n_2}}{\sqrt{\frac{\mu_n}{n} \left(1 - \frac{\mu_n}{n}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < -\bar{x}_\alpha.$$

Le niveau de ce test unilatéral  $\approx \alpha$ .

## 1.2 Modèle probabiliste de l'erreur de mesure.

Tout résultat d'observation provenant de quelque façon que ce soit de mesures engendre des erreurs d'origines diverses.

Les erreurs se divisent en trois groupes : *erreurs grossières, erreurs systématiques et erreurs aléatoires.*

### Les erreurs grossières :

Les erreurs grossières sont souvent appelées en statistique *observations aberrantes* (aberrations) ; elles proviennent de mauvais calculs, de lectures incorrectes sur l'appareil de mesure etc ... ; cela induit donc une donnée erronée. En général ces résultats de mesures qui contiennent des erreurs grossières diffèrent sensiblement des autres résultats et sont ainsi faciles à identifier.

### Les erreurs systématiques

Les erreurs systématiques surestiment ou sousestiment toujours les résultats de mesures, et sont dues à différentes raisons (mauvaise installation de l'équipement, effet de l'environnement, etc ...). Elles affectent systématiquement toutes les mesures et les altèrent dans une seule direction.

### Les erreurs aléatoires :

Les erreurs aléatoires ont un effet imprévisible sur les mesures, à la fois en surestimant certaines et en sousestimant d'autres résultats.

Considérons maintenant le modèle probabiliste (appelé le modèle de l'erreur de mesure) utilisé dans la pratique, lorsque nous avons à mesurer une certaine quantité  $\mu$ . Selon ce modèle, tout résultat de l'expérience destinée à estimer la quantité inconnue  $\mu$ , sera considéré comme la réalisation d'une variable aléatoire  $X$ . Dans ce cas, la variable aléatoire :

$$\delta = X - \mu \tag{1.1}$$

est appelée *erreur de mesure* ou *erreur vraie*.

De (1) il s'ensuit que

$$X = \mu + \delta, \quad (1.2)$$

et puisque  $\mu$  est une constante, on en tire

$$\mathbf{E}X = \mu + \mathbf{E}\delta \quad \text{et} \quad \mathbf{Var}X = \mathbf{Var}\delta. \quad (1.3)$$

Notons

$$b = \mathbf{E}\delta \quad \text{et} \quad \sigma^2 = \mathbf{Var}\delta \quad (1.4)$$

l'espérance mathématique et la variance de l'erreur vraie  $\delta$ .

Alors on a

$$X = \mu + b + (\delta - b). \quad (1.5)$$

La quantité  $b = \mathbf{E}\delta$  est appelée *erreur systématique* ou *biais* de la procédure de mesure.

La variable aléatoire

$$\xi = \delta - b \quad (1.6)$$

est appelée *erreur aléatoire* de la procédure de mesure. De (2), (5) et (6) il s'ensuit que la variable aléatoire  $X$  peut être représentée par la façon suivante

$$X = \mu + b + \xi, \quad (1.7)$$

où

$$\mathbf{E}\xi = 0 \quad \text{et} \quad \mathbf{Var}\xi = \sigma^2 \quad (1.8)$$

Nous obtenons donc pour notre modèle :

$$\mathbf{E}X = \mu + b, \quad \mathbf{Var}X = \sigma^2. \quad (9)$$

Souvent on dit que  $\sigma^2$  est la *précision* de la méthode ou de l'instrument qu'on utilise pour faire les mesures. Traditionnellement, en statistique mathématique on dit que  $X$  est un *estimateur sans biais* de  $\mu + b$ .

Si le biais  $b = 0$ , alors  $X$  est un estimateur sans biais de  $\mu$ .

Nous avons maintenant une décomposition très intéressante (7) de la variable aléatoire  $X$  dont nous utiliserons la réalisation pour estimer la quantité inconnue  $\mu$ .

Selon notre modèle, l'observation  $X$  est la somme de la vraie (mais inconnue) valeur  $\mu$ , du biais  $b$  qui est la valeur de l'erreur systématique de l'instrument de mesure et de l'erreur aléatoire  $\xi$ , qui satisfait (8) et dont la variance donne donc la mesure de l'imprécision et décrit la dispersion ou la variation des données si nous avons besoin de plusieurs mesures.

De façon évidente, la mesure parfaite serait celle pour laquelle  $b = 0$  et  $\sigma^2 = 0$  mais on ne peut l'obtenir dans la pratique. Par contre, on peut organiser l'expérience de façon à avoir  $b = 0$  et en même temps à minimiser  $\sigma^2$ , c'est-à-dire à augmenter la précision des mesures ou de l'appareil qu'on utilise pour obtenir ces mesures.

Si  $b = 0$ , alors  $\mathbf{E}X = \mu$  ce qui signifie l'absence d'erreur systématique. Dans ce cas  $\delta$  représente l'erreur aléatoire et nous dirons comme nous l'avons vu plus haut que  $X$  est un estimateur sans biais pour  $\mu$ .

Pour estimer la taille de l'erreur de mesure  $\delta = X - \mu$  d'un estimateur  $X$  d'une quantité inconnue  $\mu$ , on utilise souvent *l'erreur quadratique moyenne (le risque quadratique)* ou *l'erreur absolue moyenne (le risque absolu)* qui sont respectivement définies par

$$\mathbf{E}(X - \mu)^2 \quad \text{et} \quad \mathbf{E}|X - \mu|. \quad (10)$$

Dans notre modèle nous utiliserons *l'erreur quadratique moyenne* pour caractériser la performance de l'estimateur  $X$  de  $\mu$ . Dans ce cas, de (10), on déduit :

$$\mathbf{E}(X - \mu)^2 = \mathbf{E}[(X - \mathbf{E}X) + (\mathbf{E}X - \mu)]^2 = \mathbf{E}(X - \mathbf{E}X)^2 + b^2 = \sigma^2 + b^2.$$

Nous avons donc montré que l'erreur quadratique moyenne peut se décomposer en la somme  $b^2 + \sigma^2$  du carré du biais  $b$  de la procédure de mesure et de la variance  $\sigma^2$  de l'erreur aléatoire  $\xi$ .

**Remarque 1.** Souvent dans la pratique, le coefficient

$$k = \frac{1}{\sqrt{2(\sigma^2 + b^2)}}$$

est appelé *précision de l'estimateur  $X$* .

Dans le cas d'absence d'erreur systématique ( $b = 0$ )

$$k = \frac{1}{\sqrt{2\sigma^2}} = \frac{1}{\sigma\sqrt{2}}.$$

Lorsque *la déviation standard*  $\sigma$  et le biais  $b$  sont petits, nous avons une haute précision et dans ce cas l'erreur quadratique moyenne est petite ; d'où une erreur quadratique moyenne petite signifie une précision plus grande.

**Exemple 1.** Supposons que l'on cherche à déterminer le poids  $\mu_1$  d'un objet à l'aide d'une balance. On utilise *un modèle Gaussien* pour l'erreur de mesure en représentant le résultat d'une mesure comme la réalisation de la variable aléatoire

$$X = \mu_1 + \delta, \quad (14)$$

où  $\delta$  est l'erreur de mesure,  $\delta \sim N(0, \sigma^2)$ , et  $\sigma^2$  ne dépend pas de  $\mu_1$ . Il est évident que si  $\sigma^2$  est connu et que nous voulons avoir une précision  $\sigma^2/N$ , alors nous devons faire  $N$  mesures et prendre comme estimateur  $\hat{\mu}_1$  de  $\mu_1$ , la réalisation de la statistique :

$$\hat{\mu}_1 = \bar{X}_N = \frac{1}{N}(X_1 + X_2 + \dots + X_N), \quad (15)$$

moyenne des  $N$  mesures. De (14) il s'ensuit que

$$\bar{X}_N \sim N(\mu_1, \frac{\sigma^2}{N}). \quad (16)$$

Supposons maintenant que nous voulions déterminer les poids  $\mu_1$  et  $\mu_2$  de deux objets. De combien de mesures avons nous besoin pour obtenir des estimateurs  $\hat{\mu}_1$  et  $\hat{\mu}_2$  pour  $\mu_1$  et  $\mu_2$  respectivement, chacun avec la précision  $\sigma^2/N$  ? Il est évident qu'on peut peser chaque objet  $N$  fois et de cette façon obtenir les estimateurs

$$\hat{\mu}_1 = \frac{1}{N}(X_{11} + X_{12} + \dots + X_{1N})$$

et

$$\hat{\mu}_2 = \frac{1}{N}(X_{21} + X_{22} + \dots + X_{2N}) \quad (17)$$

pour  $\mu_1$  et  $\mu_2$ . Puisque

$$\hat{\mu}_1 \sim N(\mu_1, \frac{\sigma^2}{N}) \text{ et } \hat{\mu}_2 \sim N(\mu_2, \frac{\sigma^2}{N}), \quad (18)$$

notre but est atteint mais au prix de  $2N$  mesures.

Nous allons maintenant montrer comment on peut obtenir la même précision avec seulement  $N$  mesures.

On peut remarquer qu'avec une balance et 2 objets, on peut faire plusieurs choses :

- 1) on peut déterminer le poids de chaque objet séparément.
- 2) on peut les peser tous les 2 ensemble ;
- 3) on peut déterminer la différence entre les 2.

En tenant compte de cette remarque, on peut représenter aussi les résultats de ces mesures :

$$\begin{aligned} X_{1i} &= \mu_1 + \delta_{1i}, & i &= 1, 2, \dots, n_1, \\ X_{2i} &= \mu_2 + \delta_{2i}, & i &= 1, 2, \dots, n_2, \\ X_{3i} &= \mu_1 + \mu_2 + \delta_{3i}, & i &= 1, 2, \dots, n_3, \\ X_{4i} &= \mu_1 - \mu_2 + \delta_{4i}, & i &= 1, 2, \dots, n_4, \end{aligned}$$

où  $\{\delta_{ki}\}$  sont des variables aléatoires indépendantes identiquement distribuées :

$$\delta_{ki} \sim N(0, \sigma^2), \quad i = 1, \dots, n_k, \quad k = 1, 2, 3, 4. \quad (19)$$

Par symétrie, il est naturel de prendre

$$n_1 = n_2, \quad n_3 = n_4.$$

Il est évident que les statistiques

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}, \quad \bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}, \quad \bar{X}_3 = \frac{1}{n_3} \sum_{i=1}^{n_3} X_{3i}, \quad \bar{X}_4 = \frac{1}{n_4} \sum_{i=1}^{n_4} X_{4i}, \quad (20)$$

sont indépendantes et

$$\bar{X}_1 \sim N(\mu_1, \frac{\sigma^2}{n_1}), \quad \bar{X}_2 \sim N(\mu_2, \frac{\sigma^2}{n_1}), \quad (n_1 = n_2)$$

et

$$\bar{X}_3 \sim N(\mu_1 + \mu_2, \frac{\sigma^2}{n_3}), \bar{X}_4 \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{n_3}), \quad (n_3 = n_4) \quad (21)$$

d'où on déduit que

$$\hat{\mu}_1 = \bar{X}_1 \text{ et } \hat{\mu}_2 = \bar{X}_2$$

sont des estimateurs sans biais pour  $\mu_1$  et  $\mu_2$  ayant chacun pour précision  $\sigma^2/n_1$ . Construisons les statistiques

$$\mu_1^* = \frac{1}{2}(\bar{X}_3 + \bar{X}_4) \text{ et } \mu_2^* = \frac{1}{2}(\bar{X}_3 - \bar{X}_4). \quad (22)$$

Il est clair que

$$\mu_1^* \sim N(\mu_1, \frac{\sigma^2}{2n_3}) \text{ et } \mu_2^* \sim N(\mu_2, \frac{\sigma^2}{2n_3}), \quad (23)$$

d'où il s'ensuit que  $\mu_1^*$  et  $\mu_2^*$  sont aussi des estimateurs sans biais de  $\mu_1$  et  $\mu_2$ . De l'autre côté, on peut remarquer que si  $n_1 = n_3$ , alors la variance de  $\mu_1^*$  est 2 fois plus petite que la variance de  $\hat{\mu}_1$ . De même pour  $\mu_2^*$  et  $\hat{\mu}_2$ . En posant  $n_1 = N/2$ , notre but est atteint :

$$\mathbf{Var}\mu_2^* = \frac{1}{2}\mathbf{Var}\hat{\mu}_2. \quad (24)$$

**Exemple 2.** (suite). Supposons maintenant que l'on a 3 objets dont on veut déterminer les poids, en les pesant sur une balance non calibrée. Dans ce cas, les mesures pour ces trois objets peuvent être représentés de la façon suivante :

$$X_1 = \mu_1 + b + \delta_1, X_2 = \mu_2 + b + \delta_2, X_3 = \mu_3 + b + \delta_3, \quad (25)$$

respectivement, où  $b$  est l'erreur systématique ou le biais (supposé inconnu) de la procédure de mesure due au fait que la balance n'est pas calibrée et  $\delta_i$  est l'erreur aléatoire,  $\delta_i \sim N(0, \sigma^2)$ . Puisque

$$\mathbf{E}X_i = \mu_i + b, \quad (26)$$

pour estimer  $\mu_i$ , nous avons besoin du biais. Cela demande une lecture sans aucun objet sur la balance, c'est-à-dire qu'on obtient

$$X_4 = b + \delta_4, \delta_4 \sim N(0, \sigma^2). \quad (27)$$

Puisque

$$\mathbf{E}X_4 = b, \quad (28)$$

on peut utiliser  $X_4$  comme estimateur de  $b$ .

Considérons les statistiques

$$\hat{\mu}_i = X_i - X_4, i = 1, 2, 3. \quad (29)$$

Puisque toutes les mesures sont indépendantes, on peut dire que  $\delta_1, \delta_2, \delta_3, \delta_4$  sont des variables aléatoires i.i.d.,

$$\delta_i \sim N(0, \sigma^2), i = 1, \dots, 4,$$

et puisque

$$\hat{\mu}_i = X_i - X_4 = \mu_i + b + \delta_i - b - \delta_4 = \mu_i + \delta_i - \delta_4 \quad (30)$$

des propriétés de  $\delta_1, \delta_2, \delta_3, \delta_4$ , on déduit que

$$\hat{\mu}_i \sim N(\mu_i, 2\sigma^2), \quad i = 1, 2, 3. \quad (31)$$

Puisque

$$\mathbf{E}\hat{\mu}_i = \mu_i, \quad (32)$$

on peut dire que  $\hat{\mu}_i$  est un estimateur sans biais pour  $\mu_i$ . On remarque que

$$\mathbf{Var}\hat{\mu}_i = 2\sigma^2, \quad i = 1, 2, 3. \quad (33)$$

On peut représenter notre expérience à l'aide de la *matrice d'expérience*

$$\Sigma_1 = \left\| \begin{array}{cccc} \mu_1 & \mu_2 & \mu_3 & b \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right\|.$$

Considérons une autre représentation d'expérience donnée par la matrice :

$$\Sigma_2 = \left\| \begin{array}{cccc} \mu_1 & \mu_2 & \mu_3 & b \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array} \right\|.$$

Dans cette expérience les 3 premières mesures sont comme précédemment (dans  $\Sigma_1$ ) mais la quatrième détermine le poids des 3 articles ensemble, c'est à dire :

$$X_4 = \mu_1 + \mu_2 + \mu_3 + b + \delta_4.$$

Il est évident que :

$$X_4 \sim N(\mu_1 + \mu_2 + \mu_3 + b, \sigma^2),$$

$$\mathbf{E}X_4 = \mu_1 + \mu_2 + \mu_3 + b, \quad \mathbf{Var}X_4 = \mathbf{Var}\delta_4 = \sigma^2.$$

Considérons maintenant les statistiques

$$Y_1 = X_1 + X_4 - X_2 - X_3, \quad Y_2 = X_2 + X_4 - X_1 - X_3, \quad Y_3 = X_3 + X_4 - X_1 - X_2.$$

Alors :

$$\mathbf{E}Y_1 = 2\mu_1, \quad \mathbf{E}Y_2 = 2\mu_2, \quad \mathbf{E}Y_3 = 2\mu_3,$$

d'où on déduit que

$$\mu_i^* = \frac{1}{2}Y_i, \quad i = 1, 2, 3$$

sont des estimateurs sans biais pour  $\mu_1, \mu_2, \mu_3$  respectivement, c'est à dire

$$\mathbf{E}\mu_i^* = \mu_i, \quad i = 1, 2, 3.$$

De plus les variables aléatoires  $\delta_1, \delta_2, \delta_3, \delta_4$  sont indépendantes,  $\delta_i \sim N(0, \sigma^2)$ , d'où nous obtenons

$$\mathbf{Var}\mu_i^* = \frac{1}{4}\mathbf{Var}Y_i = \frac{4\sigma^2}{4} = \sigma^2.$$

Ainsi, si nous organisons l'expérience selon la matrice  $\Sigma_2$ , nous pouvons obtenir les mêmes résultats qu'avec une balance calibrée sans erreur systématique.

Enfin on remarque que si, par exemple, il nous faut déterminer les poids  $\mu_1, \dots, \mu_4$  de 4 objets et que la balance est calibrée, alors dans ce cas au lieu d'utiliser le plan avec la matrice

$$\Sigma_3 = \left\| \begin{array}{cccc} \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right\|,$$

d'après lequel on a

$$X_i = \hat{\mu}_i \sim N(\mu_i, \sigma^2),$$

il est évident qu'il est mieux de choisir le plan avec la matrice

$$\Sigma_4 = \left\| \begin{array}{cccc} \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{array} \right\|.$$

Dans ce cas on obtient les estimateurs

$$\hat{\mu}_i \sim N(\mu_i, \frac{\sigma^2}{2}), i = 1, \dots, 4.$$

**Exemple 3.** Supposons que nous observons un objet A qui se déplace uniformément avec une vitesse constante et inconnue  $\theta$ ,  $\theta > 0$ . Soit  $s(t)$  la distance parcourue par cet objet A entre les temps  $t = 0$  et  $t$ ,  $t > 0$ . En supposant que  $s(0) = 0$ , on a

$$s(t) = \theta t, \text{ pour tout } t \geq 0.$$

Pour estimer  $\theta$  on mesure les distances

$$s_1 = s(t_1), s_2 = s(t_2), \dots, s_n = s(t_n)$$

aux moments  $t_1 < t_2 < \dots < t_n$ , on suppose que  $s_0 = s(0) = 0$ .

Par ailleurs on sait que la précision de mesure de  $s_i$  est égale à  $k_i\sigma^2$ , où les constantes  $k_i$  sont données,  $i = 1, \dots, n$ ;  $\sigma^2 > 0$ . Dans ces conditions on propose souvent comme valeur expérimentale pour  $\theta$  le nombre

$$\hat{\theta} = \sum_{i=1}^n \hat{c}_i s_i,$$

où

$$\hat{c}_i = \frac{t_i}{\alpha k_i} \text{ et } \alpha = \sum_{i=1}^n \frac{t_i^2}{k_i}.$$

On remarque que les coefficients  $\hat{c}_i$  sont choisis de façon que

$$\hat{\mathbf{c}}^T \mathbf{t} = 1, \text{ où } \hat{\mathbf{c}} = (\hat{c}_1, \dots, \hat{c}_n)^T \text{ et } \mathbf{t} = (t_1, \dots, t_n)^T.$$

Construire un modèle probabiliste permettant de donner des explications raisonnables sur l'origine et l'optimalité dans un certain sens de cette procédure d'estimation.

**Solution.** Supposons que  $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$  est une réalisation d'un vecteur aléatoire  $\mathbf{S} = (S_1, \dots, S_n)^T$  dont les coordonnées  $S_i$  sont des variables aléatoires indépendantes telles que

$$\mathbf{E}S_i = \theta t_i \text{ et } \sigma_i^2 = \mathbf{Var}S_i = k_i \sigma^2 = \sigma_i^2, i = 1, \dots, n.$$

Dans ce cas nous pouvons dire que

$$S_i = \theta t_i + \delta_i, i = 1, \dots, n;$$

$$\mathbf{E}\delta_i = 0, \mathbf{Var}\delta_i = k_i \sigma^2 = \sigma_i^2, i = 1, \dots, n.$$

Nous supposons que  $t_i$  et  $k_i$  sont donnés, mais le paramètre  $\theta$  et la variance  $\sigma^2$  ne sont pas connus. Notre but est de montrer que  $\hat{\theta}$  est une réalisation du meilleur estimateur (de variance minimale) sans biais dans la classe  $\Delta_\theta$  de tous les estimateurs linéaires sans biais  $\theta_n^*$  pour  $\theta$  :

$$\Delta_\theta = \{\theta_n^* = \theta_n^*(\mathbf{S}) : \theta_n^* = \sum_{i=1}^n c_i S_i, \mathbf{E}_\theta \theta_n^* = \theta\}.$$

Pour montrer cela nous considérons en plus la classe

$$\Delta = \{\theta_n^* = \theta_n^*(\mathbf{S}) : \theta_n^* = \sum_{i=1}^n c_i S_i\}$$

de toutes les statistiques linéaires. Il est évident que  $\Delta_\theta \subset \Delta$ .

Soit  $\theta_n^*$  une statistique linéaire,  $\theta_n^* \in \Delta$ . Puisque

$$\mathbf{E}_\theta \theta_n^* = \sum_{i=1}^n c_i \mathbf{E}S_i = \sum_{i=1}^n c_i \theta t_i = \theta \sum_{i=1}^n c_i t_i,$$

on en tire que  $\theta_n^* \in \Delta_\theta$  si et seulement si

$$\sum_{i=1}^n c_i t_i = \mathbf{c}^T \mathbf{t} = 1.$$

Comme on l'a déjà remarqué, le choix des coefficients  $\hat{c}_i$  a été fait de façon à satisfaire cette condition, et donc la statistique

$$\hat{\theta}_n = \sum_{i=1}^n \hat{c}_i S_i$$

appartient à notre classe  $\Delta_\theta$  des estimateurs linéaires sans biais. Montrons que  $\hat{\theta}_n$  a la variance minimale dans la classe  $\Delta_\theta$  :

$$\mathbf{Var}\hat{\theta}_n = \min_{\theta_n^* \in \Delta_\theta} \mathbf{Var}\theta_n^*.$$

Pour tout  $\theta_n^* \in \Delta_\theta$  on a :

$$\mathbf{Var}\theta_n^* = \sum_{i=1}^n c_i^2 \mathbf{Var}S_i = \sigma^2 \sum_{i=1}^n k_i c_i^2.$$

Il nous faut construire l'estimateur  $\hat{\theta}_n, \hat{\theta}_n \in \Delta_\theta$ , tel que

$$\mathbf{Var}\hat{\theta}_n = \min_{\theta_n^* \in \Delta_\theta} \mathbf{Var}\theta_n^*.$$

Cela signifie qu'il nous faut minimiser la fonction

$$\sum_{i=1}^n k_i c_i^2$$

à condition que

$$\mathbf{c}^T \mathbf{t} = \sum_{i=1}^n c_i t_i = 1.$$

En utilisant la méthode de Lagrange nous pouvons trouver ce minimum lié. Soit  $\lambda$  un multiplicateur de Lagrange. Nous voulons minimiser la fonction de Lagrange

$$\Phi(\mathbf{c}, \lambda) = \sum c_i^2 k_i - 2\lambda(\sum c_i t_i - 1),$$

donc il nous faut résoudre l'équation

$$\text{grad}\Phi(\mathbf{c}, \lambda) = \mathbf{0}, \mathbf{0} \in \mathbf{R}^{r+1},$$

ce qui est équivalent à résoudre le système de  $n + 1$  équations

$$\frac{\partial \Phi(\mathbf{c}, \lambda)}{\partial c_i} = 2c_i k_i - 2\lambda t_i = 0, i = 1, 2, \dots, n,$$

et

$$\frac{\partial \Phi(\mathbf{c}, \lambda)}{\partial \lambda} = \sum c_i t_i - 1 = 0.$$

On trouve que

$$c_i = \lambda t_i / k_i, i = 1, \dots, n.$$

Pour trouver  $\lambda$  il faut mettre les valeurs trouvées de  $c_i$  dans la dernière équation du système, d'où on obtient que

$$\lambda = \frac{1}{\sum_{i=1}^n \frac{t_i^2}{k_i}},$$

et donc

$$\hat{c}_i = \frac{\frac{t_i}{k_i}}{\sum_{i=1}^n \frac{t_i^2}{k_i}} = \frac{t_i}{\alpha k_i}, i = 1, \dots, n.$$

Ces valeurs de  $c_i$  nous donnent justement l'estimateur  $\hat{\theta}_n$  sans biais,  $\hat{\theta}_n \in \Delta_\theta$ , dont la variance est minimale :

$$\hat{\theta}_n = \sum_{i=1}^n \hat{c}_i S_i = \sum_{i=1}^n \frac{t_i}{\alpha k_i} S_i.$$

Puisque les statistiques  $S_i$  sont indépendantes, par des calculs directs on trouve que

$$\begin{aligned}\mathbf{Var}\hat{\theta}_n &= \mathbf{Var} \sum_{i=1}^n \hat{c}_i S_i = \sum_{i=1}^n (\hat{c}_i)^2 \mathbf{Var} S_i = \sigma^2 \sum_{i=1}^n k_i (\hat{c}_i)^2 = \\ &= \sigma^2 \alpha^{-2} \sum_{i=1}^n \frac{t_i^2}{k_i} = \sigma^2 \left( \sum_{i=1}^n \frac{t_i^2}{k_i} \right)^{-1} = \frac{1}{\alpha} \sigma^2.\end{aligned}$$

### 1.3 Méthode de Monte-Carlo.

Considérons le problème d'évaluation d'un intégrale multidimensionnelle

$$I_n = \int_0^1 \cdots \int_0^1 f_n(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{K_n} f(\mathbf{x}) d\mathbf{x}, \quad (1)$$

où

$$\begin{aligned}\mathbf{x} &= (x_1, \dots, x_n)^T \in K_n = [0, 1] \times [0, 1] \times \dots \times [0, 1] = [0, 1]^n, \\ \{f_n(\cdot)\} &\text{ est une suite de fonctions données, } f_n(\cdot) : K_n \rightarrow \mathbf{R}^1, \quad n \in \mathbf{N}.\end{aligned}$$

Il est connu que le problème d'évaluation d'intégrales de ce type devient compliqué avec l'augmentation de  $n$ .

Supposons que nous pouvons construire un échantillon  $\mathbb{X}_1 = (X_{11}, \dots, X_{1n})^T$  de taille  $n$ , formé des variables aléatoires indépendantes suivant la même loi uniforme  $\mathcal{U}([0, 1])$  sur  $[0, 1]$ . Dans ce cas le vecteur  $\mathbb{X}_1$  suit une loi uniforme  $\mathcal{U}(K_n)$  sur le cube  $K_n$ .

Supposons en plus que nous pouvons construire un échantillon  $\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_N)^T$  de taille  $N$  quelque soit  $N \in \mathbf{N}$  des vecteurs aléatoires indépendants, ayant la même loi uniforme  $\mathcal{U}(K_n)$  sur le cube  $K_n$ , c'est-à-dire nous pouvons construire  $nN$  variables aléatoires indépendantes  $X_{ij}$  uniformément distribuées sur  $[0, 1]$ . On remarque que de la construction des variables aléatoires  $X_{ij}$  il suit que

$$\mathbf{E}f_n(\mathbb{X}_i) = \int_{K_n} f_n(\mathbf{x}) d\mathbf{x} = I_n, \quad (2)$$

i.e. la valeur numérique de l'intégrale n'est que la moyenne  $\mathbf{E}f_n(\mathbb{X}_i)$  de la variable aléatoire  $f_n(\mathbb{X}_i)$ . Dans ce cas pour estimer la moyenne  $\mathbf{E}f_n(\mathbb{X}_i) = I_n$  nous avons la possibilité d'utiliser la loi faible des grands nombres de Bernoulli d'après laquelle

$$\frac{1}{N} \sum_{i=1}^N f_n(\mathbb{X}_i) \xrightarrow{\mathbf{P}} I_n, \quad N \rightarrow \infty, \quad (3)$$

i.e. pour tout  $\varepsilon > 0$

$$\mathbf{P}\left\{ \left| \frac{1}{N} \sum_{i=1}^N f_n(\mathbb{X}_i) - I_n \right| > \varepsilon \right\} \rightarrow 0, \quad \text{si } N \rightarrow \infty, \quad (4)$$

ou

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{i=1}^N f_n(\mathbb{X}_i) - I_n\right| \leq \varepsilon\right\} \rightarrow 1, \quad \text{si } N \rightarrow \infty, \quad (5)$$

d'où on tire que pour les grandes valeurs de  $N$  avec une probabilité proche à 1 on a

$$I_n \approx \frac{1}{N}\sum_{i=1}^N f_n(\mathbb{X}_i) \quad (6)$$

De (4) et du Théorème Limite Central on tire que pour les grandes valeurs de  $N$

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{i=1}^N f_n(\mathbb{X}_i) - I_n\right| \geq \varepsilon\right\} \approx 2\Phi\left(-\frac{\varepsilon\sqrt{N}}{\sigma_n}\right) \quad (7)$$

et donc

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{i=1}^N f_n(\mathbb{X}_i) - I_n\right| \leq \varepsilon\right\} \approx 1 - 2\Phi\left(-\frac{\varepsilon\sqrt{N}}{\sigma_n}\right)$$

où

$$\sigma_n^2 = \mathbf{Var} f_n(\mathbb{X}_i) = \mathbf{E}[f_n(\mathbb{X}_i) - I_n]^2 = \int_{K_n} [f_n(x) - I_n]^2 dx \quad (8)$$

est la variance de  $f_n(\mathbb{X}_i)$ . (On suppose que  $\mathbf{Var} f_n(\mathbb{X}_i)$  existe). Donc si nous voulons que la probabilité dans (7) soit proche à 0.997, par exemple, il faut choisir  $\varepsilon$  de façon que

$$\frac{\varepsilon\sqrt{N}}{\sigma_n} = 3$$

i.e.

$$\varepsilon = \frac{3\sigma_n}{\sqrt{N}},$$

d'où on tire que la précision  $\varepsilon$  d'approximation de  $I_n$ , donnée par (7), est de l'ordre de  $N^{-1/2}$ . Il est important de noter que la précision de l'approximation ne dépend que de la variance  $\sigma_n^2$  de  $f_n(\mathbb{X}_i)$ . Donc pour évaluer l'intégrale  $I_n$  avec la précision  $\frac{3\sigma_n}{\sqrt{N}}$  il suffit de modeliser  $N$  vecteurs aléatoires  $\mathbb{X}_i$  et calculer  $N$  valeurs  $f_n(\mathbb{X}_i)$ .

Comparons ce résultat avec la méthode classique du calcul des intégrales en utilisant une approximation par les sommes.

Si  $n = 1$ , la méthode de Simpson avec  $N$  noeuds d'interpolation donne (pour une fonction  $f$  régulière) la précision  $\frac{1}{N^4}$ . Mais pour  $n > 1$  l'utilisation de cette méthode pour chacune des variables même seulement avec 10 noeuds d'interpolation exige  $10^n$  calculs des valeurs de la fonction  $f_n(x) = f_n(x_1, \dots, x_n)$ . Alors avec augmentation de  $n$  le calcul de l'intégral  $I_n$  par cette méthode devient pratiquement impossible à cause de cumulation des erreurs de calcul. Méthode de Monte-Carlo dans les mêmes conditions exige  $nN$  modelisations des variables aléatoires  $X_{ij}$  et  $N$  calculs des valeurs de la fonction  $f_n(\mathbb{X}_i)$  au lieu de  $10^n$  dans la méthode de Simpson. Il est clair que pour  $n$  grand la méthode de Monte-Carlo est uniquement possible. Mais il est raisonnable bien sûr de trouver un estimateur supérieur de la variance  $\sigma_n^2$ .

**Exercice 1.** Soit  $f$  une fonction continue périodique de période  $T = 1$  sur  $\mathbf{R}^1$  :

$$f(x+T) = f(x), \quad x \in \mathbf{R}^1.$$

Considérons une suite des variables aléatoires indépendantes  $\{X_n\}$ , uniformément distribuées sur  $[0, 1]$ ,  $X_i \sim U([0, 1])$ . Montrer que

$$\frac{1}{n} \sum_{k=1}^n f(x + X_k) \xrightarrow{\mathbf{P}} \int_0^1 f(x) dx.$$

**Exercice 2.** Soit  $f$  continue sur  $[0, 1]$ . Montrer que

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 f\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) dx_1 dx_2 \cdots dx_n = f\left(\frac{1}{2}\right).$$

**Exercice 3.** Calculer

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 \cos^{2m} \frac{\pi}{2n} (x_1 + x_2 + \cdots + x_n) dx_1 dx_2 \cdots dx_n, m \in \mathbf{N}.$$

**Exercice 4.** Soient  $g$  continue et bornée sur  $\mathbf{R}^1$  et

$$\{X_n\} \xrightarrow{\mathbf{P}} X.$$

Montrer que

a)  $\lim_{n \rightarrow \infty} \mathbf{E}g(X_n) = \mathbf{E}g(X)$  ( la suite du théorème de Lebesgue) ;

b)  $\lim_{n \rightarrow \infty} \mathbf{E}|g(X_n) - g(X)|^r = 0, \quad r > 0.$

**Exercice 5.** Soit  $f$  continue sur  $[0, 1]$ . Montrer que

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 f(\sqrt[n]{x_1 \cdots x_n}) dx_1 \cdots dx_n = f\left(\frac{1}{e}\right).$$

**Exercice 6.** Soient  $f$  et  $g$  continues sur  $[0, 1]$  et telles que pour tout  $x \in ]0, 1[$

$$0 \leq f(x) < cg(x), \quad c > 0.$$

Montrer que

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 \frac{f(x_1) + \cdots + f(x_n)}{g(x_1) + \cdots + g(x_n)} dx_1 \cdots dx_n = \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}.$$

**Exercice 7.** Montrer que

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 \frac{x_1^2 + \cdots + x_n^2}{x_1 + \cdots + x_n} dx_1 \cdots dx_n = \frac{2}{3}.$$

**Exercice 8.** Soit  $f$  telle que  $f''$  est continue sur  $[0, 1]$ . Montrer que

$$\lim_{n \rightarrow \infty} n \int_0^1 \cdots \int_0^1 \left[ f\left(\frac{x_1 + \cdots + x_n}{n}\right) - f\left(\frac{1}{2}\right) \right] dx_1 \cdots dx_n = \frac{f''\left(\frac{1}{2}\right)}{24}.$$

**Exercice 9.** Montrer que

a)

$$\lim_{n \rightarrow \infty} \int_{0 \leq x_i \leq 1, x_1^2 + \cdots + x_n^2 \leq \sqrt{n}} \cdots \int dx_1 \cdots dx_n = 0;$$

b)

$$\lim_{n \rightarrow \infty} \int \dots \int_{0 \leq x_i \leq 1, x_1^2 + \dots + x_n^2 \leq \frac{n}{4}} dx_1 \dots dx_n = 0;$$

c)

$$\lim_{n \rightarrow \infty} \int \dots \int_{0 \leq x_i \leq 1, x_1^2 + \dots + x_n^2 \leq \frac{n}{2}} dx_1 \dots dx_n = 1.$$

**Exercice 10.** Calculer

a)

$$\lim_{n \rightarrow \infty} \int \dots \int_{\{x_1^2 + \dots + x_n^2 \leq n\}} f(x_1) \dots f(x_n) dx_1 \dots dx_n;$$

b)

$$\lim_{n \rightarrow \infty} \int \dots \int_{\{\sum_{k=1}^n x_k^2 \leq an\}} f(x_1) \dots f(x_n) dx_1 \dots dx_n \quad (a < \sigma^2);$$

c)

$$\lim_{n \rightarrow \infty} \int \dots \int_{\{\sum_{k=1}^n x_k^2 \leq an\}} f(x_1) \dots f(x_n) dx_1 \dots dx_n \quad (a > \sigma^2);$$

si  $f$  satisfait aux condition

$$1 = \int_{-\infty}^{\infty} f(x) dx, \quad \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx < \infty.$$

**Exercice 11.** On dit qu'une suite de nombres  $\{a_n\}$ ,  $n \in \mathbf{N}^*$ ,  $a_n \in [0, 1]$  est uniformément distribuée au sens de H.Weyl sur  $[0, 1]$  si pour toute fonction continue  $f$ , intégrable sur  $[0, 1]$  au sens de Riemann

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(a_k) = \int_0^1 f(x) dx.$$

Soit  $\{X_n\}$  une suite de variables aléatoires indépendantes uniformément distribuées sur  $[0, 1]$ . Montrer que avec probabilité égale à 1  $\{X_n\}$  est uniformément distribuée au sens de Weyl sur  $[0, 1]$ .

**Remark 1.** On rappelle que

$$\int_a^b f(x) dx = \int_{u(a)}^{u(b)} \frac{f(u^{-1}(t))}{u'(u^{-1}(t))} dt,$$

en faisant le changement de variables  $t = u(x)$ . En statistique on utilise souvent les transformations en choisissant :

$$u(x) = e^{-x}, \quad u(x) = 1/x, \quad u(x) = x/(1+x).$$



# Chapitre 2

## ELEMENTS DE LA THEORIE DE L'ESTIMATION PONCTUELLE.

### 2.1 Modèle statistique. Fonction de vraisemblance.

Soient  $(\Omega, \mathcal{A}, \mathbf{P})$  un espace probabilisé et  $(\mathbf{R}^n, \mathcal{B}_n)$  un espace borélien.

**Définition 1.** Une application

$$\mathbf{X} = \mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))^T : \Omega \rightarrow \mathbf{R}^n$$

de l'ensemble  $\Omega = \{\omega\}$  de tous les événements élémentaires dans  $\mathbf{R}^n$  est appelée un *vecteur aléatoire* si

$$\mathbf{X}^{-1}(B) \in \mathcal{A}, \text{ pour tout } B \in \mathcal{B}_n. \blacksquare \quad (1)$$

**Définition 2.** Soit  $\mathbf{P}_\mathbf{X}$  une mesure sur  $(\mathbf{R}^n, \mathcal{B}_n)$ , déterminée par la formule suivante :

$$\mathbf{P}_\mathbf{X}(B) = \mathbf{P}\{\omega : \mathbf{X}(\omega) \in B\} = \mathbf{P}\{\mathbf{X}^{-1}(B)\} = \mathbf{P}\{\mathbf{X} \in B\}. \quad (2)$$

La mesure  $\mathbf{P}_\mathbf{X}$ , déterminée sur la  $\sigma$ -algèbre borélienne  $\mathcal{B}_n$  par l'égalité (2), s'appelle *la distribution* (la répartition) *de  $\mathbf{X}$  dans  $\mathbf{R}^n$* . ■

Supposons que la distribution  $\mathbf{P}_\mathbf{X}$  de  $\mathbf{X}$  appartienne à une famille

$$\mathcal{P} = \{\mathbf{P}_\theta, \quad \theta \in \Theta\}.$$

**Définition 3.** On appelle *modèle statistique* le triplet  $(\mathbf{R}^n, \mathcal{B}_n, \mathcal{P})$ .

Souvent au lieu de  $(\mathbf{R}^n, \mathcal{B}_n, \mathcal{P})$  on écrit  $(\mathbf{R}^n, \mathcal{B}_n, \mathbf{P}_\theta, \quad \theta \in \Theta)$  pour indiquer l'espace des paramètres  $\Theta$ .

**Définition 4.** Un modèle  $(\mathbf{R}^n, \mathcal{B}_n, \mathbf{P}_\theta, \theta \in \Theta)$  est dit *dominé* par une mesure  $\sigma$ -finie  $\mu$  dans  $\mathbf{R}^n$ , si la famille  $\mathcal{P} = \{\mathbf{P}_\theta, \quad \theta \in \Theta\}$  est *absolument continue* par rapport à  $\mu$  :

$$\mathbf{P}_\theta \ll \mu, \quad \forall \theta \in \Theta.$$

Autrement dit, le modèle  $(\mathbf{R}^n, \mathcal{B}_n, \mathbf{P}_\theta, \theta \in \Theta)$  est dominé par  $\mu$ , si pour tout  $\theta \in \Theta$  il existe une fonction non négative  $\mathcal{B}_n$ -mesurable  $p(\mathbf{x}; \theta)$  telle que

$$\mathbf{P}_\theta(B) = \int_B p(\mathbf{x}; \theta) d\mu(\mathbf{x})$$

pour tout  $B \in \mathcal{B}_n$ . La fonction  $p(\mathbf{x}; \theta) = p_\theta(\mathbf{x})$  est appelée *la dérivée de Radon-Nikodym* de la mesure  $\mathbf{P}_\theta$  par rapport à la  $\sigma$ -mesure  $\mu$ , et on note souvent

$$p(\mathbf{x}; \theta) = \frac{d\mathbf{P}_\theta}{d\mu}(\mathbf{x}) \quad \text{ou} \quad d\mathbf{P}_\theta(\mathbf{x}) = p(\mathbf{x}; \theta)d\mu(\mathbf{x}).$$

Considérons le modèle :

$$H_0 : \mathbf{X} \sim p(\mathbf{x}; \theta), \theta \in \Theta, \mathbf{x} \in \mathbf{R}^n,$$

d'après lequel la densité d'un vecteur aléatoire  $\mathbf{X} = \mathbf{X}(\omega)$  de dimension  $n$  appartient à une famille des densités

$$\{p(\mathbf{x}; \theta), \theta \in \Theta\}, \mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n.$$

**Définition 5.** Si  $\Theta$  est un ensemble  $\Theta$  de  $\mathbf{R}^m$ , on dit que le modèle  $H_0$  est *paramétrique*, sinon le modèle  $H_0$  s'appelle *non paramétrique*.

**Définition 6.** La variable aléatoire

$$L(\theta) = L(\mathbf{X}, \theta) = p(\mathbf{X}; \theta), \quad \theta \in \Theta \subset \mathbf{R}^m, \quad (3)$$

est appelée *la fonction de vraisemblance de  $\mathbf{X}$* . ■

**Remarque 1.** On appelle  $L(\theta)$  ainsi car la fonction de vraisemblance  $L(\theta)$ , sachant la réalisation  $\mathbf{x}$  du vecteur aléatoire  $\mathbf{X}$ , nous permet de comparer les paramètres  $\theta_1 \in \Theta$  et  $\theta_2 \in \Theta$ . Si

$$L(\theta_1) > L(\theta_2),$$

il est plus probable que  $\mathbf{X} = \mathbf{x}$  pour  $\theta = \theta_1$ . ■

Avec cette optique il est très naturel de considérer

$$\hat{\theta}_n = \hat{\theta}_n(\mathbf{X}) = \arg_{\theta} \max \mathbf{L}(\theta), \quad \text{i.e.} \quad \mathbf{L}(\hat{\theta}_n) = \max_{\theta \in \Theta} \mathbf{L}(\theta),$$

comme un estimateur de  $\theta$ , appelé *l'estimateur de maximum de vraisemblance*.

## 2.2 Statistique. Échantillon. Loi empirique.

**Définition 1.** Soit  $\mathbf{T} = \mathbf{T}(\mathbf{x})$  une application de  $(\mathbf{R}^n, \mathcal{B}_n)$  dans un espace  $\mathbf{E}$  muni d'une  $\sigma$ -algèbre borélienne  $\mathcal{E}$ ,  $\mathbf{T} : \mathbf{R}^n \rightarrow \mathbf{E}$ . On dit que  $\mathbf{T}$  est une *application borélienne* si pour tout ensemble borélien  $B$  de l'espace  $(\mathbf{E}, \mathcal{E})$ ,  $B \in \mathcal{E}$ ,  $\mathbf{T}^{-1}(B)$  est un ensemble borélien dans  $(\mathbf{R}^n, \mathcal{B}_n)$ , i.e.

$$\{\mathbf{x} : \mathbf{T}(\mathbf{x}) \in B\} = \mathbf{T}^{-1}(B) \in \mathcal{B}_n, \text{ pour tout } B \in \mathcal{E}. \blacksquare$$

**Définition 2.** Soient  $\mathbf{X} = \mathbf{X}(\omega)$  un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $\mathbf{X} : \Omega \rightarrow \mathbf{R}^n$ , et  $\mathbf{T}(\mathbf{x})$ , une application borélienne de  $\mathbf{R}^n$  dans un espace mesurable  $(\mathbf{E}, \mathcal{E})$ ,

$$\mathbf{T} : \mathbf{R}^n \rightarrow \mathbf{E}.$$

Dans ce cas on dit que  $\mathbf{T}(\mathbf{X}) = \mathbf{T}(\mathbf{X}(\omega))$  est une *statistique* et l'application  $\mathbf{T}$  elle-même s'appelle une *fonction de décision*.■

En d'autres termes n'importe quelle transformation du vecteur d'observations  $\mathbf{X}$  ne dépendant pas du paramètre inconnu  $\theta$  est une statistique.

**Définition 3.** Soit  $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))^T$  un vecteur aléatoire. Considérons un modèle  $H_0$  d'après lequel les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes et suivent la même loi. Dans ce cas on dit que  $\mathbf{X}$  est un *échantillon de taille  $n$*  et on écrit  $\mathbb{X}$  au lieu de  $\mathbf{X}$ .■

**Remarque 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon de taille  $n$ ,  $\mathbb{X} : \Omega \rightarrow \mathbf{R}^n$ . Considérons un modèle paramétrique

$$H_0 : \mathbb{X} \sim p(\mathbf{x}; \theta), \theta \in \Theta, \mathbf{x} \in \mathbf{R}^n.$$

Soit  $f(x_i; \theta)$  la densité de  $X_i : \mathbf{R}^1 \times \Theta \rightarrow \mathbf{R}^1$ . Dans ce cas pour tout  $\mathbf{x} \in \mathbf{R}^n$

$$p(\mathbf{x}; \theta) = \prod_{i=1}^n f(x_i; \theta), \theta \in \Theta,$$

et la fonction de vraisemblance de l'échantillon  $\mathbb{X}$  est

$$L(\theta) = p(\mathbb{X}; \theta) = \prod_{i=1}^n f(X_i; \theta), \theta \in \Theta. \blacksquare$$

**Exemple 1. Statistiques d'ordre. Vecteur des rangs.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $\mathbb{X} \in \mathcal{X} \subset \mathbf{R}^n$ . A toute réalisation  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathcal{X}$  de  $\mathbb{X}$  on peut associer le vecteur  $\mathbf{x}^{(n)} = (x_{(1)}, \dots, x_{(n)})^T$  obtenu en ordonnant les  $x_i$  par ordre croissant

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

La statistique correspondante  $\mathbf{X}^{(n)} = (X_{(1)}, \dots, X_{(n)})^T$  est appelée *le vecteur des statistiques d'ordre* et  $X_{(i)}$  est la  $i$ -ème statistique d'ordre dans  $\mathbf{A} \subset \mathbf{R}^n$  :

$$\mathbf{A} = \{\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbf{R}^n : x_1 \leq x_2 \leq \dots \leq x_n\}.$$

Si de plus on associe à  $\mathbb{X}$  le *vecteur*  $\mathbf{R} = (R_1, \dots, R_n)^T$  des rangs  $R_i$  des  $X_i$  ( $i = 1, \dots, n$ ), dans  $\mathbf{X}^{(n)}$ , avec

$$R_i = \sum_{j=1}^n \mathbf{1}_{\{X_j \leq X_i\}}$$

et on suppose que

$$\mathbf{P}\{X_{(1)} < X_{(2)} < \dots < X_{(n)}\} = 1,$$

alors dans ce cas la correspondance entre  $\mathbb{X}$  et la statistique  $(\mathbf{X}^{(n)}, \mathbf{R})$  est bijective. En général,  $\mathbf{R}$  est à valeurs dans l'ensemble  $\sigma_n$  des permutations des  $n$  premiers entiers, avec répétition car il peut y avoir des *ex aequo* parmi les composantes de  $\mathbb{X}$ . Cependant, si la probabilité pour qu'au moins deux des composantes de  $\mathbb{X}$  soient égales est nulle,  $\mathbf{R}$  est à valeurs dans l'ensemble  $\sigma_n$  des permutations de  $\{1, 2, \dots, n\}$ . Cela se produit en particulier si la loi de  $\mathbb{X}$  admet une densité  $p(\mathbf{x})$  par rapport à la mesure de Lebesgue sur  $\mathbf{R}^n$ . Parfois, au lieu de  $X^{(n)}$  on utilise le signe  $X^{(\cdot)}$ .

La statistique  $\mathbf{J}_n = (J_1, \dots, J_n)^T$ , où

$$J_k = \sum_{j=1}^n j \mathbf{1}_{\{R_j=k\}}, \quad k = 1, 2, \dots, h,$$

est connue comme le *vecteur des antirangs*.

Soit  $F(x) = \mathbf{P}\{X_1 \leq x\}$  la fonction de répartition de  $X_1$ . Dans ce cas on a, par exemple,

$$\mathbf{P}\{X_{(n)} \leq x\} = F^n(x), \quad \mathbf{P}\{X_{(1)} \leq x\} = 1 - [1 - F(x)]^n,$$

$$\mathbf{P}\{X_{(r)} \leq x\} = n! \sum_{k=r}^n \frac{F^k(x)(1 - F(x))^{n-k}}{k!(n-k)!},$$

puisque

$$\mathbf{P}\{X_{(r)} \leq x < X_{(r+1)}\} = \frac{n!}{r!(n-r)!} (F(x))^r [1 - F(x)]^{n-r}.$$

Donc si la loi  $F$  de  $X_1$  est absolument continue, i.e. s'il existe la densité  $f(x)$  telle que

$$F(x) = \int_{-\infty}^x f(u) du, \quad x \in \mathbf{R}^1,$$

alors la loi de  $X_{(r)}$  est absolument continue aussi et sa densité est donnée par la formule

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} (F(x))^{r-1} [1 - F(x)]^{n-r}, \quad r = 1, \dots, n.$$

**Exemple 2.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon. Dans ce cas les statistiques

$$T_1 = \sum_{i=1}^n X_i, \quad T_2 = \sum_{i=1}^n X_i^2, \quad \bar{X}_n = \frac{T_1}{n}, \quad s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

$$T_3 = X_{(1)}, \quad T_4 = X_{(n)}, \quad T_5 = X_{(n)} - X_{(1)}, \quad V_n = \frac{s_n}{\bar{X}_n}$$

donnent des exemples simples de statistiques scalaires, tandis que

$$\mathbf{T} = (T_1, T_2)^T \quad \text{et} \quad \mathbf{U} = (\bar{X}_n, s_n^2)^T$$

sont deux statistiques vectorielles de dimension deux. La statistique  $V_n$  s'appelle le coefficient de variabilité,  $T_5$  est l'étendu de l'échantillon,  $T_3$  et  $T_4$  sont les statistiques extrémales.

**Exemple 3. La loi empirique.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $F(x) = \mathbf{P}\{X_i \leq x\}$  est la fonction de répartition de  $X_i$ . Ayant la réalisation  $\mathbf{x} = (x_1, \dots, x_n)^T$  de la statistique  $\mathbb{X} = (X_1, \dots, X_n)^T$ , nous pouvons construire la fonction

$$F_n(x) = F_n(x; x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{] -\infty, x]}(x_i) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{] -\infty, x]}(x_{(i)}), \quad x \in \mathbf{R}^1,$$

dont la valeur  $F_n(x)$  en n'importe quel point  $x$ ,  $x \in \mathbf{R}^1$ , représente la réalisation de la statistique

$$\mathbb{F}_n(x) = \mathbb{F}_n(x; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{] -\infty, x]}(X_i) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{] -\infty, x]}(X_{(i)}),$$

calculée au point choisi  $x$ .

Par construction, la fonction  $F_n(x)$ ,  $x \in \mathbf{R}^1$ , a toutes les propriétés d'une fonction de répartition, car elle est croissante de 0 à 1 et continue à droite, et pour cette raison nous pouvons introduire une variable aléatoire discrète, disons  $X$ , dont la loi conditionnelle, conditionnée par  $\mathbb{X} = x$ , est donnée par la fonction  $F_n(x)$ , c'est-à-dire

$$F_n(x) = \mathbf{P}\{X \leq x \mid \mathbb{X} = x\} = \mathbf{P}\{X \leq x \mid X_1 = x_1, \dots, X_n = x_n\}, \quad x \in \mathbf{R}^1,$$

et par conséquent

$$\mathbb{F}_n(x) = \mathbf{P}\{X \leq x \mid \mathbb{X}\}, \quad x \in \mathbf{R}^1.$$

Cette formule détermine la fonction de répartition *aléatoire* et, par tradition, on l'appelle *la fonction de répartition empirique*. Par conséquent, la loi conditionnelle de la variable aléatoire  $X$ , conditionnée par  $\mathbb{X}$ , s'appelle *la loi empirique*. La loi empirique est la loi discrète de  $X$  telle que

$$\mathbf{P}\{X = X_i \mid \mathbb{X}\} = \frac{1}{n}$$

pour tout  $i = 1, 2, \dots, n$  et  $\mathbb{F}_n(x)$  est la fonction de répartition de cette loi.

Les statistiques  $\bar{X}_n$  et  $s_n^2$  représentent la moyenne et la variance de la loi empirique. Par définition la statistique

$$\hat{x}_P = X_{([nP]+1)}$$

représente  $P$ -quantile de la loi empirique, et par conséquent,  $\hat{x}_{0.5} = X_{([\frac{n}{2}]+1)}$  est la médiane de la loi empirique.

**Remarque 2.** Soit  $X = (X_1, \dots, X_n)^T$  un vecteur aléatoire,  $X \in \mathbf{R}^n$ , dont la densité est  $p_X(x)$ ,  $x = (x_1, \dots, x_n)^T$ .

Considérons une statistique  $Y = f(X)$ , où  $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$  est une application dérivable.

Notons

$$y = f(x), \quad \text{i.e.} \quad y = (y_1, \dots, y_n)^T, \quad \text{où} \quad y_j = f_j(x), \quad x \in \mathbf{R}^n.$$

Le Jacobien de  $f$  est une application

$$Df : \mathbf{R}^n \rightarrow \mathbf{R}^1,$$

donnée par la formule :

$$Df(x) = \det \left\| \frac{\partial f_j(x)}{\partial x_i} \right\|,$$

i.e.  $Df(x)$  est le déterminant de la matrice Jacobienne.

Si  $Df(x) \neq 0$  au voisinage d'un point  $x$ ,  $x \in \mathbf{R}^n$ , dans ce cas  $f^{-1}(y)$  existe au voisinage du point  $y = f(x)$  avec

$$Df^{-1}(f(x))Df(x) = 1, \quad (1)$$

ou

$$Df^{-1}(y)Df(x) = 1, \quad y = f(x).$$

Si  $f^{-1}$  existe, alors d'après une propriété connue en analyse, pour toute fonction intégrable  $\varphi$  de  $\mathbf{R}^n$  on a

$$\int_A \varphi(y) dy = \int_{f^{-1}(A)} \varphi(f(x)) |Df(x)| dx \quad (2)$$

pour tout  $A$ , borelien de  $\mathbf{R}^n$ . C'est la formule de changement de variables dans une intégrale.

**Lemme 1.** Soient  $Y = f(X)$  et  $p_X(x)$  la densité de  $X$ ,  $X \in \mathbf{R}^n$ , où  $f$  est telle que  $f^{-1}$  existe. Dans ce cas la densité  $p_Y(y)$  de la statistique  $Y$  est donnée par la formule

$$p_Y(y) = p_X(f^{-1}(y))|Df^{-1}(y)|. \quad (3)$$

**Démonstration.** D'après (2) pour tout  $B$  borélien,  $B \in \mathcal{B}_n$ , on a :

$$\begin{aligned} \mathbf{P}\{Y \in B\} &= \int_B p_Y(y) dy = \mathbf{P}\{X \in f^{-1}(B)\} = \\ &= \int_{f^{-1}(B)} p_X(x) dx = \int_B p_X(f^{-1}(y))|Df^{-1}(y)| dy, \end{aligned}$$

et donc

$$p_Y(y) = p_X(f^{-1}(y))|Df^{-1}(y)| \quad (4)$$

et vice-versa

$$p_X(x) = p_Y(f(x))|Df(x)|. \quad (5)$$

**Théorème 1.** Soit  $\mathbf{X} = (X_1, \dots, X_n)^T$ . Dans ce cas la densité de  $X_1$  est

$$p_{X_1}(x_1) = \int_{\mathbf{R}^{n-1}} p_X(x_1, \dots, x_n) dx_2 \cdots dx_n.$$

**Démonstration.** Pour tout  $A$  borélien dans  $\mathbf{R}^1$ ,  $A \in \mathcal{B}$ , on a

$$\begin{aligned} \mathbf{P}\{X_1 \in A\} &= \mathbf{P}\{X_1 \in A, -\infty < X_2 < +\infty, \dots, -\infty < X_n < +\infty\} = \\ &= \int_A \int_{\mathbf{R}^{n-1}} p_X(\mathbf{x}) dx_1 \cdots dx_n = \int_A \left\{ \int_{\mathbf{R}^{n-1}} p_X(x_1, \dots, x_n) dx_2 \cdots dx_n \right\} dx_1, \end{aligned}$$

et donc

$$X_1 \sim p_{X_1}(x_1) = \int_{\mathbf{R}^{n-1}} p_X(x_1, \dots, x_n) dx_2 \cdots dx_n.$$

**Exemple 4.** Soit  $\mathbf{X} = (X_1, X_2)^T$ ,  $Y_1 = X_1 + X_2$ . Trouvons la densité de la statistique  $Y_1$ .  
Considérons la statistique  $\mathbf{Y} = (Y_1, Y_2)^T = f(\mathbf{X})$ , où

$$Y_1 = X_1 + X_2 = f_1(\mathbf{X}), \quad Y_2 = f_2(\mathbf{X}) = X_2,$$

i.e.

$$\begin{aligned} f(\mathbf{x}) &= (y_1, y_2) = (f_1(\mathbf{x}), f_2(\mathbf{x}))^T, \\ f_1(\mathbf{x}) &= x_1 + x_2, \quad f_2(\mathbf{x}) = x_2. \end{aligned}$$

Dans ce cas

$$\frac{\partial f_1(\mathbf{x})}{\partial x_1} = 1, \quad \frac{\partial f_1(\mathbf{x})}{\partial x_2} = 1, \quad \frac{\partial f_2(\mathbf{x})}{\partial x_1} = 0, \quad \frac{\partial f_2(\mathbf{x})}{\partial x_2} = 1$$

et donc

$$Df(\mathbf{x}) = \det \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1, \quad Df^{-1}(\mathbf{y}) = 1,$$

où  $\mathbf{x} = (x_1, x_2)^T = f^{-1}(\mathbf{y})$  est donnée par les formules :

$$x_1 = f_1^{-1}(\mathbf{y}) = y_1 - y_2,$$

$$x_2 = f_2^{-1}(\mathbf{y}) = y_2,$$

et donc

$$\frac{\partial f_1^{-1}(\mathbf{y})}{\partial y_1} = 1, \quad \frac{\partial f_1^{-1}(\mathbf{y})}{\partial y_2} = -1, \quad \frac{\partial f_2^{-1}(\mathbf{y})}{\partial y_1} = 0, \quad \frac{\partial f_2^{-1}(\mathbf{y})}{\partial y_2} = 1,$$

$$Df^{-1}(\mathbf{y}) = \det \begin{vmatrix} \frac{\partial f_j^{-1}(\mathbf{y})}{\partial y_i} \end{vmatrix} = 1.$$

D'après (4) on a

$$p_{\mathbf{Y}}(\mathbf{y}) = p_{\mathbf{X}}(f^{-1}(\mathbf{y})) |Df^{-1}(\mathbf{y})| = p_{\mathbf{X}}(y_1 - y_2, y_2) \quad (6)$$

et, par conséquent, on en déduit que (avec l'aide du Théorème 1)

$$p_{Y_1}(y_1) = \int_{-\infty}^{\infty} p_{\mathbf{Y}}(\mathbf{y}) dy_2 = \int_{-\infty}^{\infty} p_{\mathbf{X}}(y_1 - y_2, y_2) dy_2. \quad (7)$$

**Théorème 2.** Si la densité  $p_{\mathbf{X}}(\mathbf{x})$  du vecteur  $\mathbf{X} \in \mathbf{R}^n$  est présentée par la formule

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n p_{X_i}(x_i),$$

où  $p_{X_i}(x_i)$  est la densité de  $X_i$ , dans ce cas les variables aléatoires  $X_1, X_2, \dots, X_n$  sont indépendantes.

**Démonstration.** Soient  $A_{i_1}, A_{i_2}, \dots, A_{i_k}$  des ensembles boréliens dans  $\mathbf{R}^1$ . Dans ce cas

$$\mathbf{P}\{X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_k} \in A_{i_k}\} = \mathbf{P}\{X_{i_j} \in A_{i_j}, j = 1, \dots, k; X_i \in \mathbf{R}^1, i \neq j\} =$$

$$\int_{A_{i_1}} \int_{A_{i_2}} \cdots \int_{A_{i_k}} \int_{\mathbf{R}^{n-k}} p_{X_{i_1}}(x_{i_1}) \cdots p_{X_{i_k}}(x_{i_k}) dx_{i_1} \cdots dx_{i_k} \prod_{i \neq j} p_{X_i}(x_i) dx_i =$$

$$\prod_{j=1}^k \int_{A_{i_j}} p(x_{i_j}) dx_{i_j} = \prod_{j=1}^k \mathbf{P}\{X_{i_j} \in A_{i_j}\}.$$

**Remarque 3.** Soit  $\mathbf{X} = (X_1, X_2)^T$  un vecteur aléatoire, dont les composantes sont indépendantes. Dans ce cas

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2), \quad (8)$$

et donc la densité de la statistique  $Y_1 = X_1 + X_2$  est donnée par la formule

$$p_{Y_1}(y_1) = \int p_{Y_1}(y_1 - y_2) p_{X_2}(y_2) dy_2 = \int p_{X_1}(y_2) p_{X_2}(y_1 - y_2) dy_2. \quad (9)$$

En effet, de (7) on trouve que

$$p_{Y_1}(y_1) = \int_{-\infty}^{\infty} p_{\mathbf{X}}(y_1 - y_2, y_2) dy_2 = \int_{-\infty}^{\infty} p_{X_1}(y_1 - y_2) p_{X_2}(y_2) dy_2$$

(on a utilisé l'indépendance de  $X_1$  et  $X_2$  et (8)).

## 2.3 Estimateur ponctuel. Consistance. Estimateur invariant

Considérons le modèle paramétrique  $H_0$  d'après lequel

$$X \sim p(\mathbf{x}; \theta), \quad x \in \mathbf{R}^n, \quad \theta \in \Theta \subset \mathbf{R}^m, \quad 1 \leq m \leq n.$$

**Définition 1.** Soit  $\mathbf{T} = \mathbf{T}(\mathbf{X})$  une statistique telle que

$$\mathbf{T} : \mathbf{R}^n \rightarrow \Theta \subset \mathbf{R}^m, \quad m \leq n.$$

Dans ce cas la statistique  $\mathbf{T}$  s'appelle *un estimateur statistique ponctuel* ou, tout simplement, *un estimateur* pour  $\theta = (\theta_1, \dots, \theta_m)^T$ . ■

Si la vraie valeur du paramètre  $\theta$  est inconnue, alors la réalisation

$$\theta^* = \mathbf{T}(\mathbf{x}), \quad \theta^* \in \Theta \subset \mathbf{R}^m,$$

de l'estimateur  $\mathbf{T}(\mathbf{X})$  est considérée comme une approximation expérimentale pour  $\theta$ ,

$$\theta \cong \theta^* = \mathbf{T}(\mathbf{x}).$$

On dit que c'est *l'estimation ponctuelle* de  $\theta$ .

**Remarque 1.** Parfois, pour souligner qu'on travaille avec un vecteur d'observations  $\mathbf{X}$  d'ordre  $n$ , on écrit  $\theta_n^*$  au lieu de  $\theta^*$ . ■

**Définition 2.** On appelle *biais* de l'estimateur  $\mathbf{T} = \mathbf{T}(\mathbf{X})$  de  $\theta, \theta \in \Theta \subset \mathbf{R}^m$ , la fonction  $\mathbf{b} : \Theta \rightarrow \mathbf{R}^m$ ,

$$\mathbf{b}(\theta) = \mathbf{E}_{\theta}(\mathbf{T} - \theta), \quad \theta \in \Theta.$$

Si

$$\mathbf{b}(\theta) \equiv \mathbf{0}_m, \quad \theta \in \Theta$$

on dit que l'estimateur  $\mathbf{T}$  est *sans biais*.

**Remarque 2.** Soient  $\theta_n^*$  et  $\bar{\theta}_n$  deux estimateurs scalaires sans biais pour  $\theta \in \Theta \subset \mathbf{R}^1$  :

$$\mathbf{E}_{\theta} \theta_n^* = \mathbf{E}_{\theta} \bar{\theta}_n, \quad \theta \in \Theta.$$

Dans ce cas  $\tilde{\theta}_n = \mathbf{E}_{\theta}(\theta_n^* | \bar{\theta}_n)$  est aussi un estimateur sans biais pour  $\theta$  :

$$\mathbf{E}_{\theta} \tilde{\theta}_n = \mathbf{E}_{\theta} \{ \mathbf{E}_{\theta}(\theta_n^* | \bar{\theta}_n) \} \equiv \theta, \quad \theta \in \Theta.$$

Supposons  $\mathbf{Var}_{\theta} \bar{\theta}_n$  et  $\mathbf{Var}_{\theta} \theta_n^*$  existent. Alors, comme

$$\mathbf{Var}_{\theta} \theta_n^* = \mathbf{E}_{\theta} \{ \mathbf{Var}_{\theta}(\theta_n^* | \bar{\theta}_n) \} + \mathbf{Var}_{\theta} \{ \mathbf{E}_{\theta}(\theta_n^* | \bar{\theta}_n) \},$$

et

$$\mathbf{E}_\theta\{\mathbf{Var}_\theta(\theta_n^*|\bar{\theta}_n)\} \geq 0,$$

on en tire que

$$\mathbf{Var}_\theta\tilde{\theta}_n = \mathbf{Var}_\theta\{\mathbf{E}_\theta(\theta_n^*|\bar{\theta}_n)\} \leq \mathbf{Var}_\theta\theta_n^*.$$

Il est évident que par symétrie on obtient également que

$$\mathbf{Var}_\theta\tilde{\theta}_n \leq \mathbf{Var}_\theta\bar{\theta}_n.$$

**Définition 3.**  $\{\mathbf{T}_n\}$  est une suite d'estimateurs *asymptotiquement sans biais* pour le paramètre  $\theta$ ,  $\theta \in \Theta$ , si pour tout  $\theta \in \Theta$

$$\mathbf{b}_n(\theta) = \mathbf{E}_\theta(\mathbf{T}_n - \theta) \rightarrow \mathbf{0}_m,$$

lorsque  $n \rightarrow \infty$ .

**Définition 4.** Soit  $\{\theta_n^*\}$  une suite d'estimateurs ponctuels pour  $\theta$ ,  $\theta_n^* = \theta_n^*(\mathbf{X})$ . On dit que  $\{\theta_n^*\}$  est une suite *consistante* ou *cohérente* pour  $\theta$ , si  $\{\theta_n^*\}$  converge en probabilité vers  $\theta$ , i.e. si pour tout  $\varepsilon > 0$

$$\mathbf{P}_\theta\{\|\theta_n^* - \theta\| > \varepsilon\} \rightarrow 0, \text{ quand } n \rightarrow \infty.$$

**Critère de consistance.** Soit  $T_n$  une suite d'estimateurs *asymptotiquement sans biais* pour le paramètre scalaire  $\theta$ ,  $\theta \in \Theta \subset \mathbf{R}^1$ , telle que  $\mathbf{Var}_\theta T_n \rightarrow 0$ , lorsque  $n \rightarrow \infty$ . Alors  $\theta_n^* \xrightarrow{\mathbf{P}} \theta$ .

En effet, de l'inégalité de Tchebychev, on tire que pour tout  $\varepsilon > 0$

$$\mathbf{P}_\theta\{|T_n - \theta| > \varepsilon\} \leq \frac{\mathbf{E}_\theta(T_n - \theta)^2}{\varepsilon^2} = \frac{\mathbf{Var}_\theta T_n}{\varepsilon^2} + \frac{b_n^2(\theta)}{\varepsilon^2} \rightarrow 0,$$

lorsque  $n \rightarrow \infty$ , puisque  $b_n(\theta) = \mathbf{E}_\theta T_n \rightarrow 0$  quand  $n \rightarrow \infty$ , et donc la suite  $\{T_n\}$  est consistante.

**Définition 5.** Soit  $\theta_n^* = \theta_n^*(X_1, \dots, X_n)$  un estimateur de paramètre  $\theta$ . On dit que  $\theta_n^*$  est invariant par rapport au paramètre de translation  $C$ ,  $C \in \mathbf{R}^1$ , si

$$\theta_n^*(X_1 + C, X_2 + C, \dots, X_n + C) = \theta_n^*(X_1, \dots, X_n).$$

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i \sim N(\mu, \sigma^2)$ . Dans ce cas la statistique  $\mu_n^*(X_1, \dots, X_n) = \bar{X}_n$  n'est pas un estimateur invariant pour  $\mu$  par rapport au paramètre de translation  $C$ , parce que

$$\mu_n^*(X_1 + C, \dots, X_n + C) = C + \bar{X}_n \neq \mu_n^*(X_1, \dots, X_n).$$

Par contre la statistique

$$s_n^2 = s_n^2(X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2$$

est un estimateur invariant pour  $\sigma^2$  par rapport au paramètre de translation  $C$ , parce que

$$s_n^2(X_1 + C, \dots, X_n + C) = s_n^2(X_1, \dots, X_n).$$

## 2.4 Fonction de perte, fonction de risque.

**Définition 1. (Fonction de perte).** Soit  $T = T(\mathbb{X}) : \mathbf{R}^n \rightarrow \Theta$  un estimateur ponctuel du paramètre  $\theta$ ,  $\theta \in \Theta \subset \mathbf{R}^1$ . Toute fonction non négative  $l(t, \theta) : \Theta \times \Theta \rightarrow \mathbf{R}_+^1$  convexe en  $t$  est appelée *fonction de perte* de l'estimateur  $T$ .

Les fonctions de perte servent à mesurer la qualité d'un estimateur ; cela suppose donc que la valeur observée  $l(t, \theta)$  de la fonction  $l(T(\mathbb{X}), \theta)$ , représente la *perte* pour chaque  $\theta$  qui résulte de l'utilisation de la valeur de  $T$  au lieu de  $\theta$ . Il est naturel de supposer que  $l(\theta, \theta) = 0$ . On utilise le plus souvent la fonction

$$l(T(\mathbb{X}), \theta) = (T(\mathbb{X}) - \theta)^2, \quad \theta \in \Theta,$$

comme fonction de perte (*fonction de perte quadratique*). Mais on peut aussi prendre

$$l(T(\mathbb{X}), \theta) = |T(\mathbb{X}) - \theta|, \quad l(T(\mathbb{X}), \theta) = \left(1 - \frac{T(\mathbb{X})}{\theta}\right)^2$$

ou

$$l(T(\mathbb{X}), \theta) = \frac{T}{\theta} - \ln\left(\frac{T}{\theta}\right) - 1.$$

Il est intéressant aussi d'utiliser des fonctions convexes et de choisir

$$l(T(\mathbb{X}), \theta) = g(T(\mathbb{X}) - \theta), \quad \theta \in \Theta,$$

où  $g$  est une fonction convexe non négative.

**Définition 2. (Fonction de risque).** On appelle *fonction de risque* ou *risque de l'estimateur*  $T$  par rapport à la fonction de perte  $l$  l'espérance mathématique de la fonction de perte

$$R_l(T, \theta) = \mathbf{E}_\theta\{l(T, \theta)\}, \quad \theta \in \Theta.$$

Cette fonction représente manifestement la perte moyenne lorsqu'on utilise l'estimateur  $T(\mathbb{X})$  quand la vraie valeur du paramètre est  $\theta$ .

Par exemple, si

$$l(T, \theta) = \begin{cases} 1, & |T - \theta| \geq \varepsilon, \\ 0, & |T - \theta| < \varepsilon, \end{cases}$$

alors la fonction de risque est

$$R_l(T, \theta) = \mathbf{E}_\theta\{l(T, \theta)\} = \mathbf{P}_\theta\{|T - \theta| \geq \varepsilon\}, \quad \theta \in \Theta.$$

Cette définition nous permet d'introduire une *relation d'ordre partiel* sur les estimateurs de  $\theta$ . Il est logique d'admettre que l'estimateur  $T_1$  est préférable à l'estimateur  $T_2$  par rapport à la fonction de perte  $l$ , si

$$R_l(T_1, \theta) \leq R_l(T_2, \theta), \quad \theta \in \Theta.$$

**Remarque 1.** Lorsque la fonction de perte choisie est la fonction de perte quadratique, le risque-associé est appelé *risque quadratique*. Dans le cas d'un estimateur sans biais, le risque quadratique correspond à la variance de l'estimateur.

**Remarque 2.** Dans le cas où le paramètre  $\theta = (\theta_1, \dots, \theta_m)^T$  est un élément de  $\Theta \subset \mathbf{R}^m$ , alors les produits sont des produits scalaires et les variances sont des matrices de covariance.

Plus d'information et des exemples on peut trouver, par exemple, dans Voinov& Nikulin (1993), (1996).

## 2.5 Statistiques exhaustives, nécessaires, minimales et complètes.

Considérons le modèle

$$H_0 : \mathbf{X} \sim p(\mathbf{x}; \theta), \theta \in \Theta \subset \mathbf{R}^m, \mathbf{x} \in \mathbf{R}^n,$$

où la densité du vecteur  $\mathbf{X} = \mathbf{X}(\omega)$  de dimension  $n$ ,  $\mathbf{X} : \Omega \rightarrow \mathbf{R}^n$ , appartient à une famille des densités

$$\{p(\mathbf{x}; \theta), \theta \in \Theta\}, \mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n.$$

**Définition 1. Exhaustivité.** On dit qu'une statistique

$$\mathbf{T} = \mathbf{T}(\mathbf{X}), \quad \mathbf{T} : \mathbf{R}^n \rightarrow \mathbf{R}^k, \quad m \leq k \leq n,$$

est *exhaustive pour le paramètre  $\theta$*  ou *pour la famille des densités*

$$\{p(\mathbf{x}; \theta), \theta \in \Theta\},$$

si la loi conditionnelle de  $\mathbf{X}$ , sachant  $\mathbf{T}$ ,

$$\mathbf{P}_\theta\{\mathbf{X} \leq \mathbf{x} \mid \mathbf{T} = \mathbf{t}\}$$

ne dépend pas de  $\theta$ , i.e.

$$\mathbf{P}_\theta\{\mathbf{X} \leq \mathbf{x} \mid \mathbf{T} = \mathbf{t}\} = \mathbf{P}\{\mathbf{X} \leq \mathbf{x} \mid \mathbf{T} = \mathbf{t}\}.$$

■

**Remarque 1.** Le fait que la loi conditionnelle de  $\mathbf{X}$ , sachant  $\mathbf{T}$ , ne dépende pas de  $\theta$  signifie que  $\mathbf{T}$  contient toute l'information sur le paramètre inconnu  $\theta$ .■

**Remarque 2.** En pratique, il est très difficile de répondre à la question s'il existe une statistique exhaustive ou non en utilisant cette définition. Mais, ce qui est plus ennuyeux c'est que cette définition ne donne aucune méthode pour construire des statistiques exhaustives. Il est donc très important d'avoir un critère simple qui permettrait de trouver des statistiques exhaustives.

**Théorème. (Critère de factorisation de Neyman-Fisher).**

Une statistique  $\mathbf{T} = \mathbf{T}(\mathbf{X})$  est exhaustive pour  $\theta$  si et seulement si la fonction de vraisemblance  $L(\theta)$  de  $\mathbf{X}$  peut être factorisée de la façon suivante :

$$L(\theta) = g(\mathbf{T}; \theta)W(\mathbf{X}), \tag{1}$$

où le premier facteur ne dépend que de  $\mathbf{T}$  et  $\theta$ , et le second ne dépend que de  $\mathbf{X}$ .

**Démonstration.** On va donner la démonstration de ce théorème dans le cas où

- i)  $X = \mathbb{X} = (X_1, \dots, X_n)^T$  est un échantillon,  
 $X_i \sim f(x_i; \theta)$ ,  $x_i \in \mathcal{X}_i = \mathcal{X} \subset \mathbf{R}^n$ ,  $\theta \in \Theta$ ;
- ii) l'espace des réalisations  $\mathcal{X}$  de  $X_i$  est fini ou infini dénombrable,

et donc la distribution de  $\mathbb{X}$  est discrète dans

$$\mathcal{X}^n = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_n = \mathcal{X} \times \mathcal{X} \times \dots \times \mathcal{X}, \quad \mathcal{X}^n \subset \mathbf{R}^n;$$

i.e.

$$\mathbb{X} \sim p(x; \theta) = \mathbf{P}_\theta\{\mathbb{X} = x\} > 0, \quad x = (x_1, \dots, x_n)^T \in \mathcal{X}^n, \quad \theta \in \Theta,$$

où

$$p(x; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

est la densité de  $\mathbb{X}$ . On suppose aussi que  $\mathcal{X}$  ne dépend pas de  $\theta$ .

Tout d'abord on démontre que si  $T = T(\mathbb{X})$  est une statistique qui vérifie (1), elle est exhaustive.

Soit  $T$  une statistique  $T : \mathcal{X}^n \rightarrow \mathcal{T}$  telle que (1) ait lieu, où  $\mathcal{T} = \{t\} \subset \mathbf{R}^k$  est l'espace des réalisations de  $T$ , i.e.

$$\mathbf{P}_\theta\{T = t\} > 0, \quad t \in \mathcal{T}.$$

Notons

$$\mathcal{X}_t = \{x = (x_1, \dots, x_n)^T : T(x) = t, \quad x \in \mathcal{X}^n \subset \mathbf{R}^n\}$$

l'orbite, correspondant à la valeur  $t$ ,  $t \in \mathcal{T}$ , de la statistique  $T$ . Il est évident que  $\mathcal{X}^n = \bigcup_{t \in \mathcal{T}} \mathcal{X}_t$ .

Comme  $\{T(\mathbb{X}) = t\} = \{\mathbb{X} \in \mathcal{X}_t\}$  on a

$$\begin{aligned} \mathbf{P}_\theta\{\mathbb{X} = x | T(\mathbb{X}) = t\} &= \frac{\mathbf{P}_\theta\{\mathbb{X} = x, T(\mathbb{X}) = t\}}{\mathbf{P}_\theta\{T(\mathbb{X}) = t\}} = \\ &= \begin{cases} \frac{\mathbf{P}_\theta\{\mathbb{X} = x\}}{\mathbf{P}_\theta\{T = t\}}, & \text{si } x \in \mathcal{X}_t, \\ 0, & \text{sinon,} \end{cases} \end{aligned}$$

car

$$\mathbf{P}_\theta\{\mathbb{X} = x, T(\mathbb{X}) = t\} = \begin{cases} \mathbf{P}_\theta\{\mathbb{X} = x\}, & \text{si } T(\mathbb{X}) = t, \\ 0, & \text{sinon.} \end{cases}$$

On remarque que d'après (1) on a

$$\mathbf{P}_\theta\{\mathbb{X} = x\} = p(x; \theta) = \begin{cases} g(t; \theta)W(x), & x \in \mathcal{X}_t, \\ 0, & \text{sinon.} \end{cases}$$

Par ailleurs pour tout  $t \in \mathcal{T}$  on a

$$\mathbf{P}_\theta\{T(\mathbb{X}) = t\} = \sum_{x \in \mathcal{X}_t} \mathbf{P}_\theta\{\mathbb{X} = x\} =$$

$$= \sum_{x \in \mathcal{X}_t} g(T(x); \theta) W(x) = \sum_{x \in \mathcal{X}_t} g(t; \theta) W(x) = g(t; \theta) \sum_{x \in \mathcal{X}_t} W(x),$$

d'où on tire que

$$\mathbf{P}_\theta\{\mathbb{X} = x | T(\mathbb{X}) = t\} = \begin{cases} \frac{W(x)}{\sum_{x \in \mathcal{X}_t} W(x)}, & x \in \mathcal{X}_t, \\ 0, & \text{sinon.} \end{cases}$$

Comme  $x$  est arbitraire,  $x \in \mathcal{X}^n$ , et  $\mathcal{X}_t$  ne dépend pas de  $\theta$ , donc

$$\mathbf{P}_\theta\{\mathbb{X} = x | T(\mathbb{X}) = t\} = p(x|t)$$

ne dépend pas de  $\theta$ , i.e.  $T$  est une statistique exhaustive.

Réciproquement, si

$$\mathbf{P}_\theta\{\mathbb{X} = x | T(\mathbb{X}) = t\} = \mathbf{P}\{\mathbb{X} = x | T(\mathbb{X}) = t\} = p(x|t)$$

ne dépend pas de  $\theta$ , alors d'après le théorème de multiplication des probabilités, on a

$$\begin{aligned} p(x; \theta) &= \mathbf{P}_\theta\{\mathbb{X} = x\} = \mathbf{P}_\theta\{\mathbb{X} = x | T(\mathbb{X}) = t\} \mathbf{P}_\theta\{T(\mathbb{X}) = t\} = \\ &= p(x|t) g(t; \theta) = g(t; \theta) W(x), \quad x \in \mathcal{X}^n = \bigcup \mathcal{X}_t, \end{aligned}$$

où  $W(x) = p(x|t) = p(x|T(x))$ , i.e. on obtient (1), et donc le théorème est démontré.

**Remarque 3.** Il faut noter que, en principe, une statistique exhaustive, comme n'importe quelle statistique, n'est pas un estimateur du paramètre inconnu. On a vu que la définition de statistique est plus large que la définition d'estimateur. Evidemment, certaines statistiques exhaustives peuvent être des estimateurs mais, en général, ce n'est pas le cas. L'idée d'utiliser des statistiques exhaustives permet de réduire les données expérimentales sans perdre l'information. Chercher des estimateurs est l'étape suivante du traitement des observations. Cela signifie que il est recommandé de chercher les estimateurs statistiques en termes des statistiques exhaustives, si elles existent.

**Définition 2.** Soit  $\mathbf{T}$  une statistique exhaustive. Dans ce cas  $\mathbf{U} = \mathbf{U}(\mathbf{T})$  est appelée une statistique *nécessaire*.

Pour que la statistique nécessaire  $U = U(T)$  soit exhaustive il suffit que  $U(\cdot)$  soit inversible.

**Définition 3.** Soit  $\mathcal{U}$  l'ensemble de toutes les statistiques exhaustives pour la famille  $\{\mathbf{P}_\theta, \theta \in \Theta\}$ . Une statistique exhaustive  $\mathbf{U}$ ,  $\mathbf{U} \in \mathcal{U}$ , est dite *minimale* si elle est nécessaire par rapport à toute autre statistique exhaustive  $\mathbf{T}$ ,  $\mathbf{T} \in \mathcal{U}$ , i.e. pour chaque  $\mathbf{T} \in \mathcal{U}$  il existe une application  $\mathbf{U} : \mathbf{U} = \mathbf{U}(\mathbf{T})$ .

On dit aussi que  $\mathbf{U}$  est une *réduction* de toute statistique exhaustive  $\mathbf{T}$  (d'où le nom de minimale). Cela signifie que  $U$  est une réduction de  $T$  si de l'égalité  $T(x_1) = T(x_2)$  il suit l'égalité  $U(x_1) = U(x_2)$ ,  $x_1, x_2 \in \mathcal{X}^n$ .

Donc, une statistique exhaustive minimale  $U$  est la statistique exhaustive la plus grossière, et donc elle "réduit" au maximum l'espace des observations sans perdre l'information sur  $\theta$ . Soit  $V = H(U)$ . Si  $H$  est inversible, c'est-à-dire  $\mathbf{H}$  est une application bijective bimeasurable, alors  $V$  est elle aussi exhaustive, sinon  $V$  n'est plus exhaustive. Si  $H$  est inversible,

$$V \sim U,$$

et dans ce sens  $U$  est unique (classe d'équivalence).

**Remarque 4.** Soient  $T = T(X)$  une statistique exhaustive,

$$L(X; \theta) = g(T; \theta) \mathbf{W}(X), \quad T : \mathcal{X}^n \rightarrow \mathcal{T},$$

et  $S = S(X)$  une autre statistique, telle que

$$S = S(X) = U(T(X)), \quad S : \mathcal{X}^n \rightarrow \mathcal{J},$$

où  $U(\cdot)$  est une fonction inversible i.e., si  $U : \mathcal{T} \rightarrow \mathcal{J}$ , alors il existe

$$R = U^{-1} : \mathcal{J} \rightarrow \mathcal{T},$$

telle que

$$T(X) = R(S) = R(S(X)).$$

On peut affirmer que  $S$  est elle aussi exhaustive ; en effet

$$\begin{aligned} L(X; \theta) &= g(T; \theta) \mathbf{W}(X) = g(R(S(X)); \theta) \mathbf{W}(X) = \\ &= g^*(S(X); \theta) \mathbf{W}(X) = g^*(S; \theta) \mathbf{W}(X). \end{aligned}$$

Nous dirons que  $T$  et  $S$  sont *équivalentes*,  $T \sim S$ , si elles sont inverses l'une de l'autre. On dit souvent aussi que  $\mathbf{W}(\mathbf{X})$  est une statistique *auxiliaire* ou *complémentaire*.

**Définition 4.** On dit que la famille de densités  $\{f(x; \theta), \theta \in \Theta \subset \mathbf{R}^m\}$  est *complète* si la seule fonction  $T, T : \mathbf{R}^1 \rightarrow \mathbf{R}^1$ , qui vérifie l'équation intégrale

$$\int_{-\infty}^{+\infty} T(x) f(x; \theta) dx = 0 \quad \text{pour tout } \theta \in \Theta$$

est telle que  $T(x) = 0$  presque partout.

**Remarque 5.** Si  $X \sim f(x; \theta), \theta \in \Theta \subset \mathbf{R}^m$ , la *complétude* de la famille  $\{f(x; \theta)\}$  signifie que le seul estimateur sans biais de 0 est une statistique  $T(X)$  qui est nulle presque partout.

**Définition 5.** Soit  $\mathbf{T} = \mathbf{T}(\mathbb{X})$  une statistique,  $\mathbf{T} : \mathbf{R}^n \rightarrow \mathbf{R}^k$ ,

$$\mathbf{T} \sim g(\mathbf{t}; \theta), \theta \in \Theta, \quad \mathbf{t} \in \mathbf{R}^k.$$

On dit que la *statistique*  $\mathbf{T}$  est *complète*, si la famille  $\{g(\mathbf{t}; \theta)\}$  est complète.

**Remarque 6.** Pour mener à bien les estimations et les tests classiques, paramétriques ou non paramétriques, on transforme les observations brutes en calculant des statistiques bien choisies qui doivent avoir les propriétés suivantes :

1) Perdre le moins d'information possible, éventuellement pas du tout (et c'est le cas des statistiques exhaustives) tout en réduisant au minimum le volume initial des observations.

2) Être calculables ou avoir une bonne approximation. Par exemple, s'il s'agit d'un estimateur obtenu par maximum de vraisemblance, il se peut que l'on ne puisse en obtenir aisément qu'une valeur approchée au premier pas à partir d'un estimateur moins bon.

3) Leurs lois doivent être, soit connues explicitement, soit admettre une bonne approximation. Bonne voulant dire à la fois simple à calculer et ayant une bonne vitesse de convergence vers la vraie valeur. Ce qui suit donne, grâce à des transformations appropriées des observations, des statistiques qui ont ces propriétés et aussi de bonnes approximations par des lois usuelles et permet ainsi de n'utiliser essentiellement que deux tables : celle de la loi

normale standard et celle des lois gamma (ou chi-deux). Des exemples illustrent l'application de ces méthodes qui donnent des approximations meilleures (de vitesse de convergence plus rapide) que les approximations usuelles.

Ces techniques sont très utiles pour tous les statisticiens qui travaillent sur des problèmes concrets, en particulier chez les ingénieurs, mais aussi, dans les domaines de la médecine et de la biologie.

Il y a plusieurs méthodes d'estimation d'un paramètre  $\theta$ , par exemple :

- 1<sup>o</sup>. La méthode des moments ( basée sur la loi empirique) ;
- 2<sup>o</sup>. la méthode des moindres carrés (basée sur la méthode de Gauss) ;
- 3<sup>o</sup>. La méthode de minimum du chi-deux ;
- 4<sup>o</sup>. La méthode du maximum de vraisemblance, etc.

En général, ces méthodes sont différentes et par conséquent les propriétés des estimateurs obtenus par ces méthodes sont différentes.

## 2.6 Information de Fisher. Inégalité de Rao-Cramer-Fréchet. Théorème de Rao-Blackwell-Kolmogorov.

Considérons un modèle paramétrique ; on a vu qu'une statistique exhaustive conserve toute " l'information " du modèle.

Pour mesurer l'information contenue dans une statistique, Fisher a défini la quantité d'information.

Considérons la famille des densités :

$$\{f(x; \theta) : \theta \in \Theta\}, \quad x \in \mathbf{R}^1, \quad \Theta \subset \mathbf{R}^1.$$

Supposons que cette famille est régulière. C'est-à-dire :

- i) il existe  $\frac{\partial}{\partial \theta} f(x, \theta)$  pour tout  $\theta \in \Theta$  ;
- ii) l'ensemble des  $x$  pour lesquels  $f(x, \theta) = 0$  est indépendant de  $\theta$  ( le support  $\mathcal{X}$  de  $f$  ne dépend pas du paramètre  $\theta$ )
- iii) on peut dériver sous l'intégrale par rapport à  $\theta$  la quantité

$$\int_{\mathbf{R}^1} f(x, \theta) dx = \int_x f(x, \theta) dx = 1. \quad (1)$$

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un  $n$ -échantillon où

$$X_i \sim f(x_i; \theta), \quad \theta \in \Theta \subset \mathbf{R}^1, \quad x_i \in \mathbf{R}^1.$$

Alors, la quantité

$$\lambda(X_j; \theta) = \frac{\partial \ln f(X_j; \theta)}{\partial \theta} \quad (2)$$

est appelé *informant de l'observation*  $X_j$  et la quantité suivante

$$\Lambda(\mathbb{X}, \theta) = \frac{\partial}{\partial \theta} \ln L(\theta) \quad (3)$$

est appelé *informant de l'échantillon*  $\mathbb{X}$  ; ( $L(\theta)$  est la fonction de vraisemblance de  $\mathbb{X}$ ).

Puisque

$$\ln L(\theta) = \sum_{j=1}^n \ln f(X_j; \theta)$$

on en tire que

$$\Lambda(\mathbb{X}; \theta) = \sum_{j=1}^n \lambda(X_j; \theta). \quad (4)$$

**Définition 1.** On appelle *information de Fisher* dans  $\mathbb{X}$  par rapport à  $\theta$  la quantité :

$$\mathbf{I}_n(\theta) = \mathbf{Var}_\theta \Lambda(\mathbb{X}, \theta), \quad (5)$$

si elle existe.

**Remarque 1.** Puisque

$$\mathbf{E}_\theta \Lambda(\mathbb{X}; \theta) = 0, \quad \theta \in \Theta, \quad (6)$$

on a

$$\mathbf{I}_n(\theta) = \mathbf{E}_\theta \Lambda^2(\mathbb{X}, \theta). \quad (7)$$

**Remarque 2.** Si (1) peut être dérivée deux fois par rapport à  $\theta$  sous le signe d'intégration, alors on peut montrer que

$$\mathbf{I}_n(\theta) = -\mathbf{E}_\theta \frac{\partial}{\partial \theta} \Lambda(\mathbb{X}, \theta). \quad (8)$$

**Remarque 3.** Puisque

$$L(\theta) = p(\mathbb{X}; \theta) = \prod_{i=1}^n f(X_i, \theta),$$

on pourra écrire :

$$\mathbf{I}_n(\theta) = n\mathbf{i}(\theta), \quad (9)$$

où

$$\mathbf{i}(\theta) = \mathbf{E}_\theta \lambda^2(X_j; \theta) \quad (10)$$

représente l'information d'une des composantes, par exemple  $X_j$ , du vecteur  $\mathbb{X}$ . Nous en déduisons que le vecteur  $\mathbb{X}$  contient  $n$  fois plus d'information que chacune de ses composantes. On remarque que si (1) peut être dérivée deux fois par rapport à  $\theta$ , alors

$$\mathbf{i}(\theta) = -\mathbf{E}_\theta \frac{\partial}{\partial \theta} \lambda(X_1, \theta). \quad (11)$$

**L'inégalité de Rao-Cramer-Fréchet.** Si  $T = T(\mathbb{X})$  un estimateur sans biais du paramètre  $\theta$ , alors sous les conditions i)-iii)

$$\mathbf{Var}_\theta T \geq \frac{1}{\mathbf{I}_n(\theta)}, \quad \theta \in \Theta. \quad (12)$$

**Démonstration.** Soit  $\tau$  la classe de tous les estimateurs  $T = T(\mathbb{X})$  sans biais pour le paramètre  $\theta$  :

$$\tau = \{T : \mathbf{E}_\theta T \equiv \theta\}.$$

Dans ce cas pour tout  $T \in \tau$  on a

$$\mathbf{E}_\theta T = \int_{\mathcal{X}^n} T(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} \equiv \theta, \quad \theta \in \Theta,$$

et donc des conditions i)-iii) on tire que

$$\frac{\partial}{\partial \theta} \int_{\mathcal{X}^n} T(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} = \int_{\mathcal{X}^n} T(\mathbf{x}) \frac{\partial}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} \equiv 1,$$

i.e. on a

$$1 \equiv \int_{\mathcal{X}^n} T(\mathbf{x}) \frac{\partial}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} = \int_{\mathcal{X}^n} T(\mathbf{x}) \left[ \frac{\partial}{\partial \theta} \ln p(\mathbf{x}; \theta) \right] p(\mathbf{x}; \theta) d\mathbf{x} = \int_{\mathcal{X}^n} T(\mathbf{x}) \Lambda(\theta) p(\mathbf{x}; \theta) d\mathbf{x} = \mathbf{E}_\theta \{ T(\mathbb{X}) \Lambda(\theta) \}, \quad \theta \in \Theta,$$

où  $\Lambda(\theta)$  est l'informant du vecteur d'observation  $\mathbb{X}$ . Comme

$$\mathbf{E}_\theta T \equiv \theta \quad \text{et} \quad \mathbf{E}_\theta \Lambda(\theta) \equiv 0$$

nous pouvons écrire que

$$\mathbf{E}_\theta \{ T(\mathbb{X}) \Lambda(\theta) \} = \mathbf{E}_\theta \{ (T - \theta) \Lambda \} = \mathbf{Cov}_\theta(T, \Lambda) \equiv 1, \quad \theta \in \Theta,$$

et donc de cette identité et de l'inégalité de Cauchy-Schwarz-Bounjakovsky on tire que

$$1 \equiv \mathbf{Cov}_\theta^2(T, \Lambda) \leq \mathbf{Var}_\theta T \times \mathbf{Var}_\theta \Lambda = \mathbf{Var}_\theta T \mathbf{I}_n(\theta),$$

d'où on obtient l'inégalité (12), connue sous le nom d'inégalité de Rao-Cramer-Fréchet.

**Remarque 4.** Si  $T = T(\mathbb{X})$  est un estimateur sans biais de la fonction différentiable  $g(\theta)$ ,  $\theta \in \Theta$ , alors on peut montrer que dans le cas régulier :

$$\mathbf{Var}_\theta T \geq \frac{[g'(\theta)]^2}{\mathbf{I}_n(\theta)}, \quad \theta \in \Theta. \quad (13)$$

Par exemple, soit  $\mathbf{E}_\theta T = g(\theta) = \theta + b(\theta)$ , i.e.  $b(\theta)$  est le biais de l'estimateur  $T$ . Dans ce cas de (13) on tire que

$$\mathbf{Var}_\theta T \geq \frac{[1 + b'(\theta)]^2}{\mathbf{I}_n(\theta)}.$$

**Remarque 5.**  $1/\mathbf{I}_n(\theta)$  n'est plus la borne inférieure de la variance d'un estimateur avec biais.

**Définition 2.** Un estimateur sans biais  $T = T(\mathbb{X})$  du paramètre  $\theta$  sera *efficace* si

$$\mathbf{Var}_\theta T = \frac{1}{\mathbf{I}_n(\theta)}. \quad (14)$$

Un estimateur efficace est donc un estimateur sans biais pour lequel la borne inférieure de l'inégalité de Rao-Cramer Fréchet est atteinte.

**Remarque 6.** En reprenant la remarque 4 on dira de la même façon que  $T$  est un estimateur efficace de  $g(\theta)$  si

$$\mathbf{Var}_\theta T = \frac{[g'(\theta)]^2}{\mathbf{I}_n(\theta)}, \quad \theta \in \Theta. \quad (15)$$

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim f(x_i; p) = p^{x_i} (1-p)^{1-x_i}, \quad p \in ]0, 1[, \quad x_i \in \mathcal{X} = \{0, 1\},$$

i.e.  $X_i$  suit une loi de Bernoulli de paramètre  $p$ . Dans ce cas la fonction de vraisemblance est

$$L(p) = \prod_{i=1}^n f(X_i; p) = p^{\sum_{i=1}^n X_i} (1-p)^{n - \sum_{i=1}^n X_i}, \quad p \in ]0, 1[$$

et donc

$$\mu_n = \sum_{i=1}^n X_i$$

est une statistique exhaustive pour  $p$ . Il est évident que la statistique  $\mu_n$  suit la loi binomiale  $B(n, p)$ . On sait que :

$$\mathbf{E}\mu_n = np \text{ et } \mathbf{Var}\mu_n = np(1-p),$$

donc la statistique

$$\hat{p}_n = \bar{X}_n = \frac{\mu_n}{n}$$

est un estimateur sans biais pour  $p$ ,

$$\mathbf{E}\hat{p}_n = \mathbf{E}\bar{X}_n = p \text{ et } \mathbf{Var}\hat{p}_n = \frac{p(1-p)}{n}, \quad p \in ]0, 1[. \quad (16)$$

Pour montrer que  $\hat{p}_n$  est le meilleur estimateur sans biais pour  $p$ , calculons la borne inférieure dans l'inégalité de Rao-Cramer-Fréchet. Comme

$$\ln L(p) = \mu_n \ln p + (n - \mu_n) \ln(1-p),$$

de (7) et (8) on déduit que

$$\Lambda(p) = \frac{\partial}{\partial p} \ln L(p) = \frac{\mu_n}{p} - \frac{n - \mu_n}{1-p}, \quad (17)$$

d'où on tire que

$$\mathbf{I}_n(p) = \mathbf{E}\Lambda^2(p) = -\mathbf{E}\frac{\partial}{\partial p}\Lambda(p) = \frac{n}{p(1-p)}, \quad p \in ]0, 1[, \quad (18)$$

on voit donc que  $\hat{p}_n$  est un estimateur efficace, puisque

$$\mathbf{I}_n(p) = \frac{1}{\mathbf{Var}\hat{p}_n}. \quad (19)$$

On va prouver maintenant qu'il y a un seul estimateur sans biais  $\hat{p}_n$  pour  $p$ , exprimé en termes de la statistique exhaustive  $\mu_n$ , c'est-à-dire qu'on va montrer que  $\mu_n$  est une statistique exhaustive complète.

Supposons qu'il existe un autre estimateur  $p_n^* = p_n^*(\mu_n)$  sans biais pour  $p$ ,

$$\mathbf{E}_p p_n^*(\mu_n) = p.$$

Dans ce cas  $\delta(\mu_n) = \hat{p}_n - p_n^*$  est un estimateur sans biais pour 0 :

$$\mathbf{E}_p \delta(\mu_n) = \mathbf{E}_p(\hat{p}_n - p_n^*) = 0, \quad p \in ]0, 1[,$$

i.e.,

$$\sum_{m=0}^n \delta(m) \binom{n}{m} p^m (1-p)^{n-m} = 0, \quad p \in ]0, 1[,$$

d'où on tire que  $\delta(m) \equiv 0$ ,  $m \in \{0, 1, \dots, n\}$ , puisque le système des fonctions  $\{1, t, t^2, \dots, t^n, \dots\}$  forme une base complète. Puisque la statistique  $\mu_n$  est complète, on en déduit que  $\hat{p}_n$  est unique, que c'est et donc le meilleur estimateur sans biais pour  $p$  et qu'il est efficace.

Supposons qu'il nous faille estimer  $p^2$ . Comme

$$\mathbf{Var}\mu_n = \mathbf{E}\mu_n^2 - (\mathbf{E}\mu_n)^2 = np - np^2,$$

on trouve que

$$\mathbf{E}\mu^2 = np + n^2 p^2 - np^2,$$

et donc

$$\mathbf{E} \frac{\mu_n^2}{n(n-1)} = \frac{p}{n-1} + p^2.$$

Comme  $\mathbf{E}\mu_n = np$ , on obtient que la statistique

$$\frac{\mu_n(\mu_n - 1)}{n(n-1)} \tag{20}$$

est le meilleur estimateur sans biais pour  $p^2$ , puisqu'il est exprimé en termes de la statistique exhaustive complète. De la même façon on peut montrer que

$$\mathbf{E} \left\{ \frac{\mu_n(\mu_n - 1) \cdots (\mu_n - k + 1)}{n(n-1) \cdots (n-k+1)} \right\} = p^k$$

pour tous les  $k = 1, 2, \dots, n$ .

**Example 2.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim f(x_i; \theta) = \frac{\theta^{x_i}}{x_i!} e^{-\theta}, \quad x_i \in \mathcal{X} = \{0, 1, 2, \dots\}, \quad \theta > 0,$$

i.e.  $X_i$  suit une loi de Poisson de paramètre  $\theta$ .

Comme

$$L(\theta) = \prod_{i=1}^n f(X_i; \theta) = e^{-n\theta} \theta^{\sum_{i=1}^n X_i} \left( \prod_{i=1}^n X_i \right)^{-1}, \quad \theta > 0,$$

du critère de factorisation on déduit que la statistique

$$T = \sum_{i=1}^n X_i$$

est exhaustive pour  $\theta$ , et comme la famille  $\{f(x; \theta)\}$  est complète, on en déduit que  $T$  est la statistique exhaustive minimale.

On remarque que dans ce modèle la statistique

$$W(\mathbb{X}) = \left( \prod_{i=1}^n X_i \right)^{-1}$$

est auxiliaire.

Il est facile de démontrer par des calculs directs que

$$\mathbf{P}_\theta\{\mathbb{X} = x | T = t\}, \quad x = (x_1, \dots, x_n)^T \in \mathcal{X}^n,$$

ne dépend pas de  $\theta$ . En effet :

$$\begin{aligned} \mathbf{P}_\theta\{\mathbb{X} = x | T = t\} &= \frac{\mathbf{P}_\theta\{X_1 = x_1, \dots, X_n = x_n, T = t\}}{\mathbf{P}_\theta\{T = t\}} = \\ &= \begin{cases} \frac{\mathbf{P}_\theta\{\mathbb{X} = x\}}{\mathbf{P}_\theta\{T = t\}}, & \text{si } x \in \mathcal{X}_t, \\ 0, & \text{sinon.} \end{cases} \end{aligned}$$

Soit  $x \in \mathcal{X}_t = \{x : \sum x_i = t\}$ . Dans ce cas pour  $\forall t \in \mathcal{X}$

$$\begin{aligned} \mathbf{P}_\theta\{\mathbb{X} = x | T = t\} &= \frac{\mathbf{P}_\theta\{\mathbb{X} = x\}}{\mathbf{P}_\theta\{T = t\}} = \frac{\frac{\theta^{x_1}}{x_1!} e^{-\theta} \dots \frac{\theta^{x_n}}{x_n!} e^{-\theta}}{\frac{(n\theta)^t}{t!} e^{-n\theta}} = \\ &= \frac{t!}{x_1! x_2! \dots x_n!} \left(\frac{1}{n}\right)^t. \end{aligned}$$

Donc, la loi conditionnelle de  $\mathbb{X}$ , sachant  $T = t$ , est la loi multinomiale uniforme, qui ne dépend pas de  $\theta$ , quelle que soit la valeur observée  $t$  de la statistique exhaustive  $T = \sum_{i=1}^n X_i$ .

On considère maintenant le problème de l'estimation du paramètre  $\theta$ . Pour estimer  $\theta$  on appliquera la méthode du maximum de vraisemblance. Pour trouver

$$\hat{\theta}_n = \arg_\theta \max L(\theta),$$

il nous faut résoudre l'équation du maximum de vraisemblance  $\Lambda(\theta) = 0$ , puisque

$$\Lambda(\theta) = \frac{\partial}{\partial \theta} \ln L(\theta).$$

Comme

$$\ln L(\theta) = -n\theta + T \ln \theta + \ln W(\mathbb{X}),$$

on doit résoudre l'équation

$$\Lambda(\theta) = -n + \frac{T}{\theta} = 0,$$

dont la solution  $\hat{\theta}_n$  est

$$\hat{\theta}_n = \frac{1}{n} T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n.$$

Comme  $T$  suit une loi de Poisson de paramètre  $n\theta$ , on obtient immédiatement que

$$\mathbf{E}_\theta \hat{\theta}_n = \theta \quad \text{et} \quad \mathbf{Var}_\theta \hat{\theta}_n = \frac{\theta}{n},$$

i.e.  $\{\hat{\theta}_n\}$  est une suite consistante d'estimateurs sans biais du paramètre  $\theta$ . On va montrer que  $\hat{\theta}_n$  est un estimateur efficace, c'est-à-dire qu'on a l'égalité :

$$\mathbf{Var} \hat{\theta}_n = \frac{1}{\mathbf{I}_n(\theta)}.$$

En effet,

$$\mathbf{I}_n(\theta) = -\mathbf{E}_\theta \frac{\partial}{\partial \theta} \Lambda(\theta),$$

et comme

$$\frac{\partial}{\partial \theta} \Lambda(\theta) = -\frac{T}{\theta^2},$$

on trouve que l'information de Fisher sur  $\theta$  dans  $\mathbb{X}$  est égale à

$$\mathbf{I}_n(\theta) = -\mathbf{E}_\theta \frac{\partial}{\partial \theta} \Lambda(\theta) = \frac{1}{\theta^2} \mathbf{E}_\theta T = \frac{n}{\theta},$$

d'où on tire que

$$\mathbf{Var} \hat{\theta}_n = \frac{1}{\mathbf{I}_n(\theta)} = \frac{\theta}{n},$$

et donc  $\hat{\theta}_n$  est un estimateur efficace pour  $\theta$ . Comme la famille des densités

$$\left\{ \frac{\theta^x}{x!} e^{-\theta}, \quad \theta > 0 \right\}$$

est complète, on en déduit que  $\hat{\theta}_n$  est un estimateur sans biais unique dans la classe des estimateurs sans biais, exprimés en termes de la statistique exhaustive  $T$  et  $\hat{\theta}_n$  est donc le meilleur estimateur sans biais pour  $\theta$ .

**Remarque 7.** Soit  $\mathbf{X} = (X_1, \dots, X_n)^T$  un vecteur aléatoire ayant une distribution discrète dans  $\mathbf{R}^n$ . Notons  $\mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_n)^T\}$  l'espace des réalisations de  $\mathbf{X}$  dans  $\mathbf{R}^n$ , c'est-à-dire que ;

$$\mathbf{P}\{\mathbf{X} = \mathbf{x}\} = p_{\mathbf{X}}(\mathbf{x}) = p(\mathbf{x}) > 0, \quad \forall \mathbf{x} \in \mathcal{X} \subset \mathbf{R}^n$$

et

$$\sum_{\mathbf{x} \in \mathcal{X}} \mathbf{P}\{\mathbf{X} = \mathbf{x}\} = \mathbf{P}\{\mathbf{X} \in \mathcal{X}\} = 1,$$

où  $\mathcal{X}$  est fini ou infini dénombrable, puisque  $\mathbf{X}$  suit une loi discrète.

Soit  $\mathbf{T} = \mathbf{T}(\mathbf{X})$  une statistique arbitraire,  $\mathbf{T}(\mathbf{x}) : \mathcal{X} \rightarrow \tau$ , où  $\tau = \{\mathbf{t}\}$  est l'espace des réalisations de  $\mathbf{T}$ ,

$$\mathbf{P}\{\mathbf{T} = \mathbf{t}\} > 0 \quad \text{pour} \quad \forall \mathbf{t} \in \tau.$$

Pour toute valeur possible  $\mathbf{t}$  de la statistique  $\mathbf{T}$ ,  $\mathbf{t} \in \tau$ , on détermine son orbite  $\mathcal{X}_{\mathbf{t}}$  dans  $\mathcal{X}$  :

$$\mathcal{X}_{\mathbf{t}} = \{\mathbf{x} : \mathbf{T}(\mathbf{x}) = \mathbf{t}, \mathbf{x} \in \mathcal{X}\}.$$

Il est évident que  $\{\mathcal{X}_{\mathbf{t}}\}$  est une partition de  $\mathcal{X}$  :

$$\bigcup_{\mathbf{t} \in \tau} \mathcal{X}_{\mathbf{t}} = \mathcal{X} \quad \text{et} \quad \mathcal{X}_{\mathbf{t}} \cap \mathcal{X}_{\mathbf{t}'} \neq \emptyset, \quad \mathbf{t}, \mathbf{t}' \in \tau, \quad \mathbf{t} \neq \mathbf{t}'. \quad (21)$$

La loi conditionnelle de  $\mathbf{X}$  sachant que  $\mathbf{T} = \mathbf{t}$  est l'ensemble des probabilités conditionnelles  $\{p(\mathbf{x} | \mathbf{t})\}$  étant donné  $\mathbf{t}$  fixé :

$$p(\mathbf{x} | \mathbf{t}) = \mathbf{P}\{\mathbf{X} = \mathbf{x} | \mathbf{T}(\mathbf{X}) = \mathbf{t}\} = \begin{cases} \frac{\mathbf{P}\{\mathbf{X}=\mathbf{x}, \mathbf{T}(\mathbf{X})=\mathbf{t}\}}{\mathbf{P}\{\mathbf{T}(\mathbf{X})=\mathbf{t}\}} = \frac{p(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{X}_{\mathbf{t}}} p(\mathbf{x})}, & \text{si } \mathbf{x} \in \mathcal{X}_{\mathbf{t}}, \\ 0, & \text{sinon,} \end{cases} \quad (22)$$

puisque

$$\mathbf{P}\{\mathbf{X} = \mathbf{x}, \mathbf{T} = \mathbf{t}\} = \begin{cases} \mathbf{P}\{\mathbf{X} = \mathbf{x}\} = p(\mathbf{x}), & \text{si } \mathbf{x} \in \mathcal{X}_{\mathbf{t}}, \\ 0, & \text{sinon .} \end{cases}$$

La famille des probabilités (22) est finie ou infinie dénombrable, et on choisit  $\mathbf{t}$  dans (22) de façon que  $\mathbf{P}\{\mathbf{T} = \mathbf{t}\} > 0$ , i.e.  $\mathbf{t} \in \tau$ .

Soit  $\mathbf{U} = \mathbf{U}(\mathbf{X})$  une autre statistique,  $\mathbf{U}(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{U}$ , telle que  $\mathbf{EU}$  existe. D'après la définition :

$$\mathbf{EU} = \mathbf{EU}(\mathbf{X}) = \sum_{\mathbf{x} \in \mathcal{X}} \mathbf{U}(\mathbf{x})p(\mathbf{x}). \quad (23)$$

On détermine l'espérance conditionnelle  $\mathbf{E}\{\mathbf{U} \mid \mathbf{T} = \mathbf{t}\}$  sachant que  $\mathbf{T} = \mathbf{t}$  en termes de la distribution conditionnelle (22) :

$$\begin{aligned} \mathbf{E}\{\mathbf{U} \mid \mathbf{T} = \mathbf{t}\} &= \sum_{\mathbf{x} \in \mathcal{X}_{\mathbf{t}}} \mathbf{U}(\mathbf{x})p(\mathbf{x} \mid \mathbf{t}) = \\ &= \frac{\sum_{\mathbf{x} \in \mathcal{X}_{\mathbf{t}}} \mathbf{U}(\mathbf{x})p(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{X}_{\mathbf{t}}} p(\mathbf{x})}. \end{aligned} \quad (24)$$

Nous pouvons considerer  $\mathbf{E}\{\mathbf{U} \mid \mathbf{T} = \mathbf{t}\}$  comme une réalisation de la variable aléatoire  $\mathbf{E}\{\mathbf{U} \mid \mathbf{T}\}$  quand  $\mathbf{T} = \mathbf{t}$ . Il est facile de prouver que

$$\mathbf{E}\{\mathbf{E}\{\mathbf{U} \mid \mathbf{T}\}\} = \mathbf{EU}.$$

De (21) et (24) il suit que

$$\begin{aligned} \mathbf{E}\{\mathbf{E}\{\mathbf{U} \mid \mathbf{T}\}\} &= \sum_{\mathbf{t} \in \tau} \mathbf{E}\{\mathbf{U} \mid \mathbf{T} = \mathbf{t}\} \mathbf{P}\{\mathbf{T} = \mathbf{t}\} = \\ &= \sum_{\mathbf{t} \in \tau} \mathbf{E}\{\mathbf{U} \mid \mathbf{T} = \mathbf{t}\} \sum_{\mathbf{x} \in \mathcal{X}_{\mathbf{t}}} p(\mathbf{x}) = \\ &= \sum_{\mathbf{t} \in \tau} \sum_{\mathbf{x} \in \mathcal{X}_{\mathbf{t}}} \mathbf{U}(\mathbf{x})p(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{X}} \mathbf{U}(\mathbf{x})p(\mathbf{x}) = \mathbf{EU}, \end{aligned}$$

puisque  $\{\mathcal{X}_{\mathbf{t}}\}$  est une partition de  $\mathcal{X}$ . On a donc montré que

$$\mathbf{E}\{\mathbf{E}\{\mathbf{U} \mid \mathbf{T}\}\} = \mathbf{EU}, \quad (25)$$

et par conséquent pour calculer  $\mathbf{EU}$  on peut tout d'abord calculer  $\mathbf{E}\{\mathbf{U} \mid \mathbf{T} = \mathbf{t}\}$  puis  $\mathbf{E}\{\mathbf{E}\{\mathbf{U} \mid \mathbf{T}\}\}$ .

Soit  $X$  un vecteur aléatoire,  $X \in \mathbf{R}^n$ ,  $X \sim F$ ,  $F \in \mathcal{F}$ , où  $\mathcal{F} = \{F\}$  est une famille de fonctions de répartition dans  $\mathbf{R}^n$ . Soient  $\psi = \psi(X)$  et  $\phi = \phi(X)$  deux statistiques,  $\psi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ ,  $\phi : \mathbf{R}^n \rightarrow \mathbf{R}^k$ , telles que

$$\mathbf{E}\psi, \quad \mathbf{E}\phi, \quad \mathbf{E}\psi\psi^T, \quad \mathbf{E}\phi\phi^T$$

existent.

Considérons la statistique

$$\Psi = \Psi(\phi) = \mathbf{E}\{\psi \mid \phi\}, \quad \Psi \in \mathbf{R}^k.$$

Il est évident que

$$\mathbf{E}\Psi = \mathbf{E}\{\mathbf{E}\{\psi|\varphi\}\} = \mathbf{E}\psi,$$

ce qui signifie que  $\Psi$  est un estimateur sans biais de  $\mathbf{E}\psi$ .

**Théorème 1** (Rao-Blackwell-Kolmogorov)

Pour tout  $z \in \mathbf{R}^k$

$$z^T \mathbf{E}\{(\Psi - \mathbf{E}\psi)(\Psi - \mathbf{E}\psi)^T\}z \leq z^T \mathbf{E}\{(\psi - \mathbf{E}\psi)(\psi - \mathbf{E}\psi)^T\}z. \quad (26)$$

Démonstration. Notons

$$\Delta = [\psi - \Psi]^T z = [(\psi - \mathbf{E}\psi) - (\Psi - \mathbf{E}\psi)]^T z.$$

Puisque  $\mathbf{E}\Delta = 0$ , nous obtenons

$$\mathbf{Var}\Delta = \mathbf{E}\Delta^2 = \mathbf{E}\Delta^T \Delta \geq 0. \quad (27)$$

Mais

$$\begin{aligned} \Delta^2 = \Delta^T \Delta = z^T \{ & (\psi - \mathbf{E}\psi)(\psi - \mathbf{E}\psi)^T - (\psi - \mathbf{E}\psi)(\Psi - \mathbf{E}\psi)^T - \\ & - (\Psi - \mathbf{E}\psi)(\psi - \mathbf{E}\psi)^T + (\Psi - \mathbf{E}\psi)(\Psi - \mathbf{E}\psi)^T \} z, \end{aligned}$$

et par suite de (27), on peut tirer que

$$\begin{aligned} 0 \leq \mathbf{E}\Delta^2 = & z^T (\mathbf{Var}\psi)z - z^T \mathbf{Cov}(\psi, \Psi)z - \\ & - z^T \mathbf{Cov}(\Psi, \psi)z + z^T (\mathbf{Var}\Psi)z. \end{aligned} \quad (28)$$

Puisque

$$\begin{aligned} \mathbf{Cov}(\psi, \Psi) &= \mathbf{E}\{(\psi - \mathbf{E}\psi)(\Psi - \mathbf{E}\psi)^T\} = \\ &= \mathbf{E}\{\mathbf{E}\{(\psi - \mathbf{E}\psi)(\Psi - \mathbf{E}\psi)^T|\varphi\}\} = \mathbf{E}\{\mathbf{E}\{(\psi - \mathbf{E}\psi)|\varphi\}(\Psi - \mathbf{E}\psi)^T\} = \\ &= \mathbf{E}\{(\Psi - \mathbf{E}\psi)(\Psi - \mathbf{E}\psi)^T\} = \mathbf{Var}\Psi, \end{aligned} \quad (29)$$

alors de (27), (28) et (29) on déduit que

$$0 \leq \mathbf{E}\Delta^2 = z^T (\mathbf{Var}\psi)z - z^T (\mathbf{Var}\Psi)z,$$

ce qu'il nous fallait démontrer.

**Remarque 8.** Si  $\psi = \psi(X)$  est un estimateur sans biais d'une fonctionnelle  $g(F)$ ,  $F \in \mathcal{F}$ , alors

$$\Psi = \mathbf{E}\{\psi|\varphi\}$$

est aussi un estimateur sans biais pour  $g(F)$ , dont le risque quadratique n'est pas plus grand que celui de  $\psi$ . Ce théorème est très intéressant lorsque  $\varphi$  est une statistique exhaustive.

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$\mathbb{X} \sim p(x; \theta), \quad \theta \in \Theta \subset \mathbf{R}^m.$$

Supposons qu'il existe une statistique exhaustive

$$\mathbf{T} = \mathbf{T}(\mathbb{X}), \quad \mathbf{T} : \mathbf{R}^n \rightarrow \mathbf{R}^k, \quad m \leq k \leq n,$$

dont la densité est  $g(t; \theta)$ ,  $t \in \mathbf{R}^k$ . Notons  $q(x, t; \theta)$  la densité conjointe de  $\mathbb{X}$  et  $\mathbf{T}$ ,  $p(x | t)$  la densité conditionnelle de  $\mathbb{X}$  sachant  $\mathbf{T} = t$ . Dans ce cas pour tout  $x$  fixé,  $x \in \mathbf{R}^n$ ,  $p(x|T)$  est un estimateur sans biais pour  $p(x; \theta)$ . En effet,

$$\mathbf{E}p(x|T) = \int_{\mathbf{R}^k} p(x|t)g(t; \theta) dt = \int_{\mathbf{R}^k} q(x, t; \theta) dt = p(x; \theta).$$

## 2.7 Méthode des moments.

La fonction de répartition  $\mathbb{F}_n(x)$  de la loi empirique associée à un échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$  est un bon estimateur de la fonction de répartition  $F(x)$  :

$$\mathbf{E}\mathbb{F}_n(x) = F(x) = \mathbf{P}\{X_i \leq x\}, \quad x \in \mathbf{R}^1,$$

et pour tout  $\varepsilon > 0$

$$\mathbf{P}\{|\mathbb{F}_n(x) - F(x)| > \varepsilon\} \rightarrow 0, \quad x \in \mathbf{R}^1$$

lorsque  $n \rightarrow \infty$  quel que soit  $x$  fixé. En pratique cela signifie que  $\mathbb{F}_n(x) \approx F(x)$  pour tout  $x$  fixé, quand  $n$  est assez grand.

Il est donc naturel de choisir les moments

$$\alpha_m = \frac{1}{n} \sum_{i=1}^n X_i^m = \int_{-\infty}^{+\infty} x^m d\mathbb{F}_n(x)$$

de la loi empirique  $\mathbb{F}_n$  comme estimateurs des moments

$$a_m = \mathbf{E}X^m = \int_{-\infty}^{+\infty} x^m dF(x)$$

de la loi  $F$ , puisque  $\alpha_m \approx a_m$ , si  $\mathbb{F}_n(x) \approx F(x)$ .

Supposons que la fonction de répartition

$$F(x; \theta) = \mathbf{P}_\theta\{X_i \leq x\}, \quad |x| < \infty$$

dépende d'un paramètre inconnu

$$\theta = (\theta_1, \dots, \theta_s)^T \in \Theta \subset \mathbf{R}^s$$

et qu'existent les moments

$$a_r(\theta) = \int_{-\infty}^{+\infty} x^r dF(x; \theta), \quad r = 1, 2, \dots, s.$$

On cherche un estimateur du paramètre  $\theta = (\theta_1, \dots, \theta_s)^T$  en résolvant le système d'équations :

$$a_m(\theta) = \alpha_m, \quad (m = 1, 2, \dots, s) \quad (1)$$

par rapport à  $\theta_1, \dots, \theta_s$ . La solution  $\tilde{\theta}_n = (\tilde{\theta}_1, \dots, \tilde{\theta}_n)^T$  de cette équation,

$$\begin{aligned}\tilde{\theta}_1 &= \tilde{\theta}_1(\alpha_1, \dots, \alpha_s), \\ \dots \\ \tilde{\theta}_s &= \tilde{\theta}_s(\alpha_1, \dots, \alpha_s),\end{aligned}$$

s'appelle *l'estimateur par la méthode des moments* de  $\theta$ .

Si les fonctions (1) déterminent une application bijective, leurs dérivées partielles existent et sont continues et les moments  $a_k(\theta)$  ( $k = 1, 2, \dots, 2s$ ) existent ; donc les estimateurs obtenus par la méthode des moments sont cohérents et de distributions asymptotiquement normales. Des propriétés asymptotiques d'estimateurs, obtenus par la méthode des moments, seront considérées dans le chapitre III.

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon lognormale  $LN(\mu, \sigma^2)$ ,

$$X_i \sim p(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \mathbf{1}_{]0, \infty[}(x), \quad \mu \in \mathbf{R}^1, \quad \sigma^2 > 0.$$

Remarquons que  $\ln X_i$  suit une loi normale  $N(\mu, \sigma^2)$ . On peut montrer que

$$a_1 = \mathbf{E}X_1 = e^{\mu + \sigma^2/2}, \quad a_2 = \mathbf{E}X_1^2 = e^{2\mu + 2\sigma^2}.$$

D'après la méthode des moments pour estimer  $\mu$  et  $\sigma^2$  il faut résoudre le système

$$\begin{cases} e^{\mu + \sigma^2/2} = \bar{X}_n & = \alpha_1, \\ e^{2\mu + 2\sigma^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 & = \alpha_2, \end{cases}$$

ce qui est équivalent à

$$\begin{cases} \mu + \sigma^2/2 & = \ln \alpha_1, \\ \mu + 2\sigma^2 & = \ln \alpha_2, \end{cases}$$

d'où on trouve les estimateurs  $\tilde{\sigma}_n^2$  et  $\tilde{\mu}_n$  :

$$\tilde{\sigma}_n^2 = \ln \alpha_2 - \ln \alpha_1^2 = \ln \left( \frac{s_n^2}{\bar{X}_n^2} + 1 \right), \quad \tilde{\mu}_n = \ln \frac{\bar{X}_n^2}{\sqrt{s_n^2 + \bar{X}_n^2}},$$

où

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

est la variance de la loi empirique.

**Exemple 2.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim p(x; \theta) = \frac{1}{\theta} \sqrt{\frac{2}{\pi}} \exp \left\{ -\frac{x^2}{2\theta^2} \right\} \mathbf{1}_{]0, \infty[}(x), \quad x \in \mathbf{R}^1, \quad \theta \in \Theta = ]0, \infty[.$$

On peut montrer que

$$\mathbf{E}X_1 = \theta \sqrt{\frac{2}{\pi}}, \quad \mathbf{E}X_1^2 = \theta^2, \quad \mathbf{Var} X_1^2 = \theta^2 \frac{\pi - 2}{\pi}.$$

Pour estimer  $\theta$  par la méthode des moments on considère l'équation

$$\theta \sqrt{\frac{2}{\pi}} = \bar{X}_n,$$

d'où on obtient l'estimateur

$$\tilde{\theta}_n = \sqrt{\frac{\pi}{2}} \bar{X}_n.$$

Il est clair que  $\mathbf{E}\tilde{\theta}_n = \theta$ , i.e.  $\tilde{\theta}_n$  est un estimateur sans biais pour  $\theta$ , et comme

$$\mathbf{Var} \bar{X}_n = \frac{\theta^2}{n} \left(1 - \frac{2}{\pi}\right),$$

on en tire que

$$\begin{aligned} \mathbf{Var} \tilde{\theta}_n &= \frac{\pi}{2} \mathbf{Var} \bar{X}_n = \frac{\theta^2}{n} \left(\frac{\pi}{2} - 1\right) = \\ &= \frac{\theta^2}{n} \frac{\pi - 2}{2} = \frac{\pi - 2}{I_n(\theta)} > \frac{1}{I_n(\theta)}, \end{aligned}$$

où

$$I_n(\theta) = \frac{2n}{\theta^2} = -n \mathbf{E} \frac{\partial^2}{\partial \theta^2} \ln p(X_1; \theta) = n \mathbf{E} \left( \frac{3}{\theta^4} X_1^2 - \frac{1}{\theta^2} \right) = \frac{2n}{\theta^2}$$

est l'information de Fisher sur  $\theta$  dans  $\mathbb{X}$ . De la dernière inégalité on voit bien que l'estimateur  $\tilde{\theta}_n$  n'est pas efficace.

**Remarque 1.** Du théorème limite central il suit que la suite des variables aléatoires

$$\frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\theta \sqrt{\frac{\pi-2}{2}}} = \frac{\sqrt{n}(\bar{X}_n - \sqrt{\frac{2}{\pi}}\theta)}{\theta \sqrt{1 - \frac{2}{\pi}}}, \quad n = 1, 2, \dots$$

est asymptotiquement normale  $N(0, 1)$ , quand  $n \rightarrow \infty$ , i.e. pour les grandes valeurs de  $n$

$$\mathbf{P} \left\{ \frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\theta \sqrt{\frac{\pi-2}{\pi}}} \leq x \right\} \approx \Phi(x), \quad x \in \mathbf{R}^1.$$

Du théorème de Slutsky on tire que les variables aléatoires

$$\frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\tilde{\theta}_n \sqrt{\frac{\pi-2}{2}}}$$

sont asymptotiquement normales  $N(0, 1)$  aussi, i.e.

$$\mathbf{P} \left\{ \frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\tilde{\theta}_n \sqrt{\frac{\pi-2}{2}}} \leq x \right\} \approx \Phi(x), \quad x \in \mathbf{R}^1,$$

si les valeurs de  $n$  sont assez grandes.

Nous pouvons utiliser ce résultat pour estimer  $\theta$  par intervalle, puisque

$$\mathbf{P} \left\{ -\bar{x}_{\alpha/2} \leq \frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\tilde{\theta}_n \sqrt{\frac{\pi-2}{2}}} \leq \bar{x}_{\alpha/2} \right\} \approx 1 - \alpha,$$

où  $\bar{x}_{\alpha/2}$  est le quantile supérieur de niveau  $\alpha/2$  pour la loi standard normale,  $0 < \alpha < 0.5$ , d'où on tire que

$$\mathbf{P} \left\{ -\bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \leq \left( 1 - \frac{\theta}{\tilde{\theta}_n} \right) \leq \bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \right\} \approx 1 - \alpha$$

et donc

$$\mathbf{P} \left\{ \tilde{\theta}_n \left( 1 - \bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \right) \leq \theta \leq \tilde{\theta}_n \left( 1 + \bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \right) \right\} \approx 1 - \alpha,$$

si  $n$  est assez grand.

## 2.8 Méthode des moindres carrés. Modèle de Gauss de la théorie des erreurs.

Supposons qu'on cherche à mesurer une constante  $\mu$ ; pour cela on fait  $n$  mesures directes  $x_1, x_2, \dots, x_n$  de  $\mu$ , indépendantes les unes des autres, de même précision, sans erreur systématique. De chaque résultat d'expérience on tire que

$$\mu \cong x_i, i = 1, 2, \dots, n. \quad (1)$$

On obtient un système de  $n$  équations, qui sont en général incompatibles si les  $x_i$  ne sont pas tous égaux. Pour cette raison il est logique de traiter  $x_i - \mu$  comme une erreur, commise au cours de la  $i$ -ème mesure de  $\mu$ , et  $\mathbf{x} - \mu \mathbf{1}_n$  comme le vecteur des erreurs que l'on a fait au cours des  $n$  expériences organisées pour déterminer  $\mu$ ; donc

$$\|\mathbf{x} - \mu \mathbf{1}_n\|^2 = (\mathbf{x} - \mu \mathbf{1}_n)^T (\mathbf{x} - \mu \mathbf{1}_n) = \sum_{i=1}^n (x_i - \mu)^2 \quad (2)$$

représente le carré de la longueur du vecteur des erreurs  $\mathbf{x} - \mu \mathbf{1}_n$ .

Compte tenu du fait que toutes les mesures sont faites dans les mêmes conditions, et que par suite les  $n$  équations ont toutes la même validité pour l'expérimentateur, Legendre a posé le problème de la détermination d'une valeur  $\mu^*, \mu^* \in \mathbf{R}^1$ , qui est meilleure que chaque résultat individuel  $x_i$  et en même temps est en meilleur accord, dans un certain sens avec tous les  $x_i$ , c'est-à-dire, avec le système d'équations (1) qui représente les résultats d'expériences de mesures du paramètre  $\mu$ .

Pour déterminer  $\mu^*$ , Legendre a proposé le *principe des moindres carrés*, d'après lequel la valeur de  $\mu$ , la plus en accord avec l'expérience est donnée par la valeur  $\mu^*$ , qui minimise  $\|\mathbf{x} - \mu \mathbf{1}_n\|^2$ , le carré de la longueur du vecteur des erreurs  $(\mathbf{x} - \mu \mathbf{1}_n)$ , i.e.

$$(\mathbf{x} - \mu^* \mathbf{1}_n)^T (\mathbf{x} - \mu^* \mathbf{1}_n) = \min_{\mu \in \mathbf{R}^1} (\mathbf{x} - \mu \mathbf{1}_n)^T (\mathbf{x} - \mu \mathbf{1}_n). \quad (3)$$

Ceci justifie le nom de la méthode, que l'on appelle *la méthode des moindres carrés*. Par tradition on dit aussi que  $\mu^*$  est un *estimateur des moindres carrés* pour  $\mu$ .

Plus tard Gauss a donné une justification logique de la méthode des moindres carrés, en utilisant un modèle classique d'erreurs de mesures, qui est connu aujourd'hui sous le

nom de *modèle de Gauss*. D'après ce modèle le résultat  $x_i$  de la  $i$ -ème mesure représente la réalisation de la variable aléatoire

$$X_i = \mu + \delta_i, i = 1, \dots, n, \quad (4)$$

où  $\delta_i$  est l'erreur aléatoire de la  $i$ -ème mesure,

$$\delta_i \in N(0, \sigma^2), \quad (5)$$

puisque  $\mathbf{E}\delta_i = 0$  par convention (absence d'erreur systématique) et  $\sigma^2 = \mathbf{Var}\delta_i > 0$  est une constante ne dépendant pas de  $i$ , car chaque mesure a été faite avec la même précision. Gauss a proposé de considérer chaque  $\delta_i$  comme une variable aléatoire de loi normale ; en effet selon la théorie des erreurs, développée par Gauss, toute erreur  $\delta_i$  représente la somme d'un grand nombre de petites erreurs, qu'on peut supposer indépendantes ; par suite, on peut supposer que leur somme est normale, ce qui peut s'expliquer dans le cadre du théorème limite central. De plus on a l'indépendance des mesures, d'où on déduit que dans le modèle de Gauss on peut supposer que  $\delta_1, \dots, \delta_n$  sont des variables aléatoires indépendantes, et donc que  $\mathbb{X} = (X_1, \dots, X_n)^T$  est un échantillon normal ; cela signifie que  $X_1, \dots, X_n$  sont des variables aléatoires indépendantes qui suivent la même loi normale de paramètres  $\mu$  et  $\sigma^2$  :

$$X_i \in N(\mu, \sigma^2), i = 1, \dots, n; \quad (6)$$

donc dans le cadre de ce modèle le résultat de l'expérience  $\mathbf{x} = (x_1, \dots, x_n)^T$  est une *réalisation* d'un vecteur aléatoire  $\mathbb{X}$ , de loi normale de dimension  $n$  et de paramètres  $\mu\mathbf{1}_n$  et  $\sigma^2\mathbf{I}_n$ ,  $\mathbb{X} \sim N_n(\mu\mathbf{1}_n, \sigma^2\mathbf{I}_n)$ ,

$$\mathbf{E}\mathbb{X} = \mu\mathbf{1}_n \text{ et } \mathbf{Var}\mathbb{X} = \mathbf{E}(\mathbb{X} - \mu\mathbf{1}_n)(\mathbb{X} - \mu\mathbf{1}_n)^T = \sigma^2\mathbf{I}_n, \quad (7)$$

où  $\sigma^2$  est une constante positive, qui exprime la précision des mesures. Nous supposons d'abord que  $\sigma^2$  est connu.

Dans le modèle de Gauss le vecteur des observations  $\mathbb{X}$  peut se représenté comme la somme

$$\mathbb{X} = \mu\mathbf{1}_n + \delta, \quad (8)$$

d'un terme déterministe, mais inconnu, et d'un terme aléatoire

$$\delta = (\delta_1, \dots, \delta_n)^T = \mathbb{X} - \mu\mathbf{1}_n, \quad (9)$$

qui suit la loi normale de dimension  $n$  et de paramètres

$$\mathbf{E}\delta = \mathbf{0}_n = (0, \dots, 0)^T \text{ et } \mathbf{E}\delta\delta^T = \sigma^2\mathbf{I}_n, \quad (10)$$

et ce vecteur  $\delta$  est le *vecteur des erreurs aléatoires*. Le problème principal dans la *théorie des erreurs*, élaborée par Gauss, est la construction du *meilleur* (en un sens à préciser) *estimateur* de  $\mu$  en utilisant la réalisation  $\mathbf{x}$  du vecteur des observations  $\mathbb{X}$ . Pour trouver ce meilleur estimateur pour  $\mu$ , Gauss a proposé d'utiliser la *méthode du maximum de vraisemblance*, d'après laquelle la valeur qui rend maximum la *fonction de vraisemblance*  $L(\mu)$ , liée au vecteur des observations  $\mathbb{X}$ , est l'estimateur du paramètre inconnu  $\mu$ . D'après le modèle que l'on a choisi, le vecteur  $\mathbb{X}$  suit une loi normale de dimension  $n$  et de paramètres (7) ; donc la fonction de vraisemblance  $L(\mu)$  est donnée par la formule

$$\begin{aligned}
L(\mu) &= (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2}(\mathbb{X} - \mu\mathbf{1}_n)^T(\mathbb{X} - \mu\mathbf{1}_n)\right\} = \\
&= (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2}\delta^T\delta\right\} = (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2}\chi_n^2\right\}, \mu \in \mathbf{R}^1, \quad (11)
\end{aligned}$$

car

$$\delta^T\delta = \sigma^2\chi_n^2. \quad (12)$$

On remarque ici que maximiser la fonction de vraisemblance  $L(\mu)$ ,  $\mu \in \mathbf{R}^1$ , revient à minimiser la fonction  $(\mathbb{X} - \mu\mathbf{1}_n)^T(\mathbb{X} - \mu\mathbf{1}_n)$  qui représente la fonction de la formule (2), mais en d'autres termes, en termes d'observations. C'est-à-dire que dans ce cas la méthode de Legendre et la méthode de Gauss sont équivalentes. Donc  $L(\mu)$ ,  $\mu \in \mathbf{R}^1$ , atteint son maximum,  $\mathbb{X}$  étant donné, pour le point  $\mu^*$ , qui rend minimum la forme quadratique

$$(\mathbb{X} - \mu\mathbf{1}_n)^T(\mathbb{X} - \mu\mathbf{1}_n) = \delta^T\delta,$$

i.e. l'estimateur statistique  $\mu^*$  est la solution du problème extrême :

$$(\mathbb{X} - \mu^*\mathbf{1}_n)^T(\mathbb{X} - \mu^*\mathbf{1}_n) = \min_{\mu \in \mathbf{R}^1} (\mathbb{X} - \mu\mathbf{1}_n)^T(\mathbb{X} - \mu\mathbf{1}_n), \quad (13)$$

obtenue pour la valeur  $\mu = \mu^*$ , qui vérifie l'équation

$$(\mathbf{1}_n^T\mathbf{1}_n)\mu = \mathbf{1}_n^T\mathbb{X}, \quad (14)$$

d'où on tire que

$$\mu^* = \bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}\mathbf{1}_n^T\mathbb{X}. \quad (15)$$

L'estimateur statistique  $\mu^* = \bar{X}_n$  s'appelle *l'estimateur des moindres carrés* ou *estimateur de Gauss* pour  $\mu$ . On remarque que

$$\begin{aligned}
L(\mu) &= (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{n}{2\sigma^2} \left[ (\bar{X}_n - \mu)^2 + \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]\right\} = \\
&\exp\left\{-\frac{n}{\sigma^2} (\bar{X}_n - \mu)^2\right\} W(\mathbb{X});
\end{aligned}$$

donc  $\bar{X}_n$  est une statistique exhaustive pour  $\mu$ . Comme  $\bar{X}_n \in \mathbf{R}^1$  et  $\mu \in \mathbf{R}^1$ , la statistique exhaustive  $\bar{X}_n$  est minimale. De (6) et (7) il suit que  $\bar{X}_n$  suit la loi normale  $N(\mu, \sigma^2/n)$  de paramètres

$$\mathbf{E}\bar{X}_n = \mu \text{ et } \mathbf{Var}\bar{X}_n = \mathbf{E}(\bar{X}_n - \mu)^2 = \frac{\sigma^2}{n}. \quad (16)$$

Puisque la famille des loi normale  $N(\mu, \sigma^2)$  est complète, on en tire que  $\bar{X}_n$  est une statistique exhaustive minimale et complète.  $\bar{X}_n$  est un estimateur *efficace* pour  $\mu$ .

Souvent la variance  $\sigma^2$  est elle aussi inconnue ; dans ce cas outre l'estimateur  $\bar{X}_n$  pour  $\mu$  il est très important d'obtenir un estimateur statistique pour  $\sigma^2$ . Notons

$$\theta = (\mu, \sigma^2)^T, \quad \theta \in \Theta = \{\theta : |\mu| < \infty, \quad \sigma^2 > 0\} \subset \mathbf{R}^2.$$

Pour estimer  $\theta$  on considère la statistique

$$\Delta = \mathbb{X} - \mu^* \mathbf{1}_n = (\mathbb{X} - \bar{X}_n \mathbf{1}_n) = (\mathbb{X} - \mu \mathbf{1}_n) + (\mu - \mu^*) \mathbf{1}_n = \delta + (\mu - \mu^*) \mathbf{1}_n, \quad (17)$$

qui s'appelle *le vecteur des erreurs apparentes*. Il est évident que la statistique  $\Delta$  suit la loi normale  $N_n(\mathbf{0}_n, \sigma^2 \mathbf{D}_n)$ , qui est dégénérée et où

$$\mathbf{D}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T, \quad (18)$$

avec  $\mathbf{I}_n$ , matrice identité d'ordre  $n$ . On remarque que  $\mathbf{D}_n$  est une matrice idempotente, puisque

$$\text{rang} \mathbf{D}_n = n - 1 \text{ et } \mathbf{D}_n^T \mathbf{D}_n = \mathbf{D}_n \mathbf{D}_n^T = \mathbf{D}_n. \quad (19)$$

De (9) et (17) on tire l'égalité

$$\delta = (\mu^* - \mu) \mathbf{1}_n + \Delta, \quad (20)$$

que l'on appelle la *décomposition orthogonale* du vecteur des erreurs aléatoires  $\delta$  en termes de  $\mu^*$  et  $\Delta$ . On remarque que

$$\begin{aligned} \delta^T \delta &= \Delta^T \Delta + (\mu^* - \mu) \mathbf{1}_n^T \mathbf{1}_n (\mu^* - \mu) = \Delta^T \Delta + n(\bar{X}_n - \mu)^2 = \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2 = n[s_n^2 + (\bar{X}_n - \mu)^2], \end{aligned} \quad (21)$$

où la statistique  $s_n^2$  est déterminée par la formule

$$s_n^2 = \frac{1}{n} \mathbb{X}^T \mathbf{D}_n \mathbb{X} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (22)$$

Comme  $\theta = (\mu, \sigma^2)^T$ , il vaut mieux écrire que la fonction de vraisemblance de  $\mathbb{X}$  est  $L(\theta) = L(\mu, \sigma^2)$ . En utilisant (11), (13), (21) et (22), nous pouvons présenter  $L(\mu, \sigma^2)$  en termes des statistiques  $s_n^2$  et  $\bar{X}_n$  par la formule suivante

$$\begin{aligned} L(\mathbb{X}; \mu, \sigma^2) &= (\sigma \sqrt{2\pi})^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \delta^T \delta \right\} = \\ &= (\sigma \sqrt{2\pi})^{-n} \exp \left\{ -\frac{n}{2\sigma^2} [s_n^2 + (\bar{X}_n - \mu)^2] \right\}, \end{aligned} \quad (23)$$

d'où on tire que la statistique  $\mathbf{T} = (\bar{X}_n, s_n^2)^T$  est exhaustive. On peut montrer que  $\mathbf{T}$  est minimale et complète.

Pour étudier les propriétés de  $\mathbf{T}$  on considère, par exemple, la transformation linéaire  $\mathbf{Y} = \mathbf{C}\mathbb{X}$  de Helmert, déterminée par la matrice orthogonale  $\mathbf{C}$ ,

$$\mathbf{C}^T \mathbf{C} = \mathbf{C} \mathbf{C}^T = \mathbf{I}_n, \quad \mathbf{C}^T = \mathbf{C}^{-1},$$

$$\mathbf{C} = \left\| \begin{array}{ccccccc} \frac{1}{\sqrt{1 \cdot 2}} & \frac{-1}{\sqrt{1 \cdot 2}} & 0 & 0 & \dots & 0 & 0 \\ \frac{1}{\sqrt{2 \cdot 3}} & \frac{1}{\sqrt{2 \cdot 3}} & \frac{-2}{\sqrt{2 \cdot 3}} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \dots & \frac{1}{\sqrt{(n-1)n}} & \frac{-(n-1)}{\sqrt{(n-1)n}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \end{array} \right\|.$$

D'après cette transformation  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ , où

$$\begin{aligned} Y_1 &= \frac{1}{\sqrt{1.2}}(X_1 - X_2), \\ Y_2 &= \frac{1}{\sqrt{2.3}}(X_1 + X_2 - 2X_3), \\ &\vdots \\ Y_{n-1} &= \frac{1}{\sqrt{(n-1)n}}(X_1 + X_2 + \dots + X_{n-1} - (n-1)X_n), \\ Y_n &= \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n) = \sqrt{n}\bar{X}_n, \end{aligned}$$

et comme  $\mathbf{C}$  est orthogonale on a

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2; \quad (24)$$

la fonction de vraisemblance de  $\mathbf{Y}$  est donc donnée par la formule :

$$\begin{aligned} L(\mathbf{Y}; \mu, \sigma^2) &= (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2} \left[ \sum_{i=1}^{n-1} Y_i^2 + (Y_n - \mu\sqrt{n})^2 \right]\right\} = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2} (Y_n - \mu\sqrt{n})^2\right\} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} \prod_{i=1}^{n-1} \exp\left\{-\frac{1}{2\sigma^2} Y_i^2\right\}, \end{aligned}$$

puisque

$$\begin{aligned} p_Y(y) &= p_{\mathbb{X}}(\mathbf{C}^{-1}y) |\det \mathbf{C}^{-1}| = p_{\mathbb{X}}(\mathbf{C}^{-1}y) = \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{C}^T \mathbf{y} - \mu \mathbf{1}_n)^T (\mathbf{C}^T \mathbf{y} - \mu \mathbf{1}_n)\right\} = \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} (\|\mathbf{y}\|^2 - 2\mu y_n \sqrt{n} + n\mu^2)\right\} = \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \left[ \sum_{i=1}^{n-1} y_i^2 + (y_n - \mu\sqrt{n})^2 \right]\right\}, \\ \mathbf{C}\mathbf{1}_n &= (0, \dots, 0, \sqrt{n})^T \quad \text{et} \quad \mu \mathbf{y}^T \mathbf{C}\mathbf{1}_n = \mu y_n \sqrt{n}, \end{aligned}$$

d'où on tire que  $Y_1, Y_2, \dots, Y_n$  sont indépendantes et

$$Y_i \sim N(0, \sigma^2), \quad i = 1, \dots, n-1; \quad (25)$$

$$Y_n \sim N(\mu\sqrt{n}, \sigma^2). \quad (26)$$

Notons que de (24) il suit que

$$ns_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^{n-1} Y_i^2,$$

donc  $s_n^2$  ne dépend pas de  $Y_n = \sqrt{n}\bar{X}_n$ , et par conséquent,  $s_n^2$  et  $\bar{X}_n$  sont indépendantes. En plus, de (24), (25) et (26) on tire que

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{et} \quad \frac{ns_n^2}{\sigma^2} = \chi_{n-1}^2, \quad (27)$$

donc que la statistique  $s_n^2$  est distribuée comme  $\sigma^2 \chi_{n-1}^2/n$ , où  $\chi_f^2$  est une variable aléatoire qui suit la loi du chi-deux à  $f$  degrés de liberté,  $f > 0$ , i.e. pour tout  $x \geq 0$

$$\mathbf{P}\{\chi_f^2 \leq x\} = \frac{2^{-f/2}}{\Gamma\left(\frac{f}{2}\right)} \int_0^x t^{\frac{f}{2}-1} e^{-t/2} dt.$$

Comme

$$\mathbf{E}\chi_f^2 = f \text{ et } \mathbf{Var}\chi_f^2 = 2f, \quad (28)$$

de (27) et (28) on tire que

$$\mathbf{E}s_n^2 = \sigma^2 \left(1 - \frac{1}{n}\right) \text{ et } \mathbf{Var}s_n^2 = \frac{2\sigma^4(n-1)}{n^2}. \quad (29)$$

On peut vérifier que dans notre problème l'estimateur de maximum de vraisemblance  $\hat{\theta}_n$  pour  $\theta$  est

$$\hat{\theta}_n = T = (\bar{X}_n, s_n^2)^T.$$

En effet, on a

$$\begin{aligned} \ln L(\mathbb{X}; \mu, \sigma^2) &= -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} s_n^2 - \frac{n}{2\sigma^2} (\bar{X}_n - \mu)^2, \\ \frac{\partial \ln L}{\partial \mu} &= \frac{n}{\sigma^2} (\bar{X}_n - \mu) \quad \text{et} \quad \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{ns_n^2}{2\sigma^4} + \frac{n}{2\sigma^4} (\bar{X}_n - \mu)^2, \end{aligned}$$

donc pour trouver  $\hat{\mu}_n$  et  $\hat{\sigma}_n^2$ , il faut résoudre le système

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = 0, \\ \frac{\partial \ln L}{\partial \sigma^2} = 0. \end{cases}$$

De la première équation du système on tire que

$$\hat{\mu}_n = \bar{X}_n,$$

et de la deuxième on tire que

$$\hat{\sigma}_n^2 = s_n^2,$$

d'où on obtient que  $\hat{\theta}_n = (\bar{X}_n, s_n^2)^T$  est l'estimateur de maximum de vraisemblance pour  $\theta = (\mu, \sigma^2)^T$ .

D'un autre côté comme de (29) on tire que

$$S_n^2 = \frac{n}{n-1} s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (30)$$

est un estimateur sans biais pour  $\sigma^2$ .

On peut montrer que la statistique  $\tilde{\theta}_n = (\bar{X}_n, S_n^2)^T$  est le *meilleur estimateur sans biais* (au sens de minimum de variance) pour  $\theta = (\mu, \sigma^2)^T$ .

## 2.9 Régions, intervalles, limites de confiance.

Dans ce paragraphe nous allons suivre les articles de Bolshev (1965) et de Bagdonavičius, Nikoulina & Nikulin (1997).

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon, dont les réalisations  $\mathbf{x} = (x_1, \dots, x_n)^T$  appartiennent à  $\mathcal{X} \subseteq \mathbf{R}^n$ ,  $\mathbf{x} \in \mathcal{X} \subseteq \mathbf{R}^n$ ,

$$H_0 : X_i \sim f(x; \theta), \theta = (\theta_1, \dots, \theta_k)^T \in \Theta \subseteq \mathbf{R}^k.$$

On s'intéresse à un problème d'estimation de la vraie valeur  $b = b(\theta)$  d'une fonction  $b(\cdot) : \Theta \Rightarrow B \subseteq \mathbf{R}^m$  au point  $\theta$ ,  $\theta \in \Theta$ . Notons  $B^0$  l'intérieur de  $B$ .

**Définition 1.** On appelle région de confiance pour  $b = b(\theta)$  de coefficient de confiance  $\gamma$  ( $0.5 < \gamma < 1$ ) ou  $\gamma$ -région de confiance tout court, un ensemble aléatoire  $\mathbf{C}(\mathbb{X})$ ,  $\mathbf{C}(\mathbb{X}) \subseteq B \subseteq \mathbf{R}^m$ , tel que

$$\inf_{\theta \in \Theta} \mathbf{P}_{\theta} \{ \mathbf{C}(\mathbb{X}) \ni b(\theta) \} = \gamma.$$

De cette définition on tire

$$\mathbf{P}_{\theta} \{ \mathbf{C}(\mathbb{X}) \ni b(\theta) \} \geq \gamma,$$

pour tous  $\theta \in \Theta$ .

Dans le cas où  $b(\theta) \in B \subseteq \mathbf{R}^1$  la région de confiance est souvent un intervalle dans  $\mathbf{R}^1$ ,

$$\mathbf{C}(\mathbb{X}) = ]b_i(\mathbb{X}), b_s(\mathbb{X})[ \subseteq B \subseteq \mathbf{R}^1,$$

et on parle de l'intervalle de confiance du coefficient de confiance  $\gamma$  pour  $b$ , si

$$\inf_{\theta \in \Theta} \mathbf{P}_{\theta} \{ b_i(\mathbb{X}) < b < b_s(\mathbb{X}) \} = \gamma.$$

Il est évident que

$$\mathbf{P}_{\theta} \{ b_i(\mathbb{X}) < b < b_s(\mathbb{X}) \} \geq \gamma$$

pour tous  $\theta \in \Theta$ . Les statistiques  $b_i(\mathbb{X})$  et  $b_s(\mathbb{X})$  sont appelées les limites de l'intervalle de confiance  $\mathbf{C}(\mathbb{X})$ . On remarque que

$$\mathbf{P}_{\theta} \{ b_i(\mathbb{X}) \leq b_s(\mathbb{X}) \} = 1.$$

**Remarque 1.** Supposons qu'on prenne un grand nombre d'échantillons  $\mathbb{X}_1, \dots, \mathbb{X}_N$  et que chaque fois on construise un intervalle de confiance  $]b_i(\mathbb{X}_i), b_s(\mathbb{X}_i)[$  du coefficient de confiance  $\gamma$ . Soit  $]b_i(\mathbf{x}_i), b_s(\mathbf{x}_i)[$  une réalisation de  $]b_i(\mathbb{X}_i), b_s(\mathbb{X}_i)[$ ;  $i = 1, \dots, N$ . Dans ce cas la vraie valeur  $b$  sera recouverte par ces intervalles  $]b_i(\mathbf{x}_i), b_s(\mathbf{x}_i)[$  au moins dans  $100\gamma\%$  des cas. Souvent on prend  $\gamma \geq 0.9$ .

**Définition 2.** Une statistique  $b_i(\mathbb{X})$  ( $b_s(\mathbb{X})$ ) est appelée la limite inférieure (supérieure) de confiance pour  $b = b(\theta)$  de coefficient de confiance  $\gamma_1$  ( $\gamma_2$ ), si

$$\inf_{\theta \in \Theta} \mathbf{P}_{\theta} \{ b_i(\mathbb{X}) < b \} = \gamma_1 \left( \inf_{\theta \in \Theta} \mathbf{P}_{\theta} \{ b_s(\mathbb{X}) > b \} = \gamma_2 \right), 0.5 < \gamma_j < 1.$$

Les statistiques  $b_i(\mathbb{X})$  et  $b_s(\mathbb{X})$  sont appelées aussi  $\gamma_1$ - limite inférieure et  $\gamma_2$  - limite supérieure tout court. Si les coefficients de confiance de  $b_i(\mathbb{X})$  et  $b_s(\mathbb{X})$  sont égaux à  $\gamma_1$  et  $\gamma_2$  respectivement, dans ce cas  $]b_i(\mathbb{X}), b_s(\mathbb{X})[$  est l'intervalle de confiance du coefficient de confiance

$$\gamma = \gamma_1 - (1 - \gamma_2) = \gamma_1 + \gamma_2 - 1$$

pour la vraie valeur de  $b = b(\theta)$ .

**Définition 3.** Les intervalles

$$]b_i(\mathbb{X}), +\infty[ \text{ et } ]-\infty, b_s(\mathbb{X})[$$

sont appelés *intervalles de confiance supérieur et inférieur* pour  $b$ . Tous les deux sont des intervalles *unilatéraux*.

## 2.10 Méthode de Bolshev de construction des limites de confiance.

**Lemme** (Bolshev (1965)) Soit  $G(t)$  la fonction de répartition d'une variable aléatoire  $T$ . Dans ce cas pour tout  $z \in [0, 1]$

$$\mathbf{P}\{G(T) \leq z\} \leq z \leq \mathbf{P}\{G(T-0) < z\}. \quad (1)$$

Si  $T$  est continue, alors

$$\mathbf{P}\{G(T) \leq z\} = z, \quad 0 \leq z \leq 1.$$

Démonstration. On va d'abord montrer que

$$\mathbf{P}\{G(T) \leq z\} \leq z, \quad 0 \leq z \leq 1. \quad (2)$$

Si  $z = 1$ , on a  $\mathbf{P}\{G(T) \leq 1\} \leq 1$ . Fixons  $z \in [0, 1)$  et pour cette valeur de  $z$  on considère les situations différentes.

1) Il existe une solution  $y$  de l'équation  $G(y) = z$ . Notons

$$y_0 = \sup\{y : G(y) = z\}.$$

On peut avoir :

a)  $G(y_0) = z$ . Dans ce cas on a

$$\mathbf{P}\{G(T) \leq z\} \leq \mathbf{P}\{T \leq y_0\} = G(y_0) = z.$$

b)  $G(y_0) > z$ . Dans ce cas on a

$$\mathbf{P}\{G(T) \leq z\} \leq \mathbf{P}\{T < y_0\} = G(y_0 - 0) \leq z.$$

2) Il n'existe pas de solution pour l'équation  $G(y) = z$ . Mais dans ce cas il existe  $y$  tel que

$$G(y) > z \quad \text{et} \quad G(y-0) < z,$$

d'où on tire que

$$\mathbf{P}\{G(T) \leq z\} \leq \mathbf{P}\{T < y\} = G(y-0) < z.$$

Donc l'inégalité (2) est démontrée.

Démontrons maintenant la seconde inégalité dans (1) :

$$z \leq \mathbf{P}\{G(T-0) < z\}, \quad 0 \leq z \leq 1. \quad (3)$$

Considérons la statistique  $-T$ . Sa fonction de répartition est

$$G^-(y) = \mathbf{P}\{-T \leq y\} = \mathbf{P}\{T \geq -y\} = 1 - G(-y-0).$$

Appliquons l'inégalité (2) en remplaçant

$$T, z, G \quad \text{par} \quad -T, 1-z \quad \text{et} \quad G^-$$

respectivement :

$$\mathbf{P}\{G^-(-T) \leq 1-z\} \leq 1-z, \quad 0 \leq z \leq 1,$$

d'où on obtient que

$$\mathbf{P}\{1 - G(T-0) \leq 1-z\} \leq 1-z,$$

$$\mathbf{P}\{G(T-0) \geq z\} \leq 1-z,$$

$$\mathbf{P}\{G(T-0) < z\} \geq z, \quad 0 \leq z \leq 1.$$

Si  $T$  est continue, dans ce cas  $G(t-0) = G(t)$ , et donc (2) et (3) nous donnent  $\mathbf{P}\{G(T) \leq z\} = z$  pour tout  $z \in [0, 1]$ .

Le Lemme de Bolshev est démontré.

**Théoreme.** *Supposons que l'on ait une variable aléatoire  $T = T(\mathbb{X}, b)$ ,  $b \in B$ , telle que sa fonction de répartition*

$$G(t, b) = \mathbf{P}_\theta\{T \leq t\}$$

*ne dépende que de  $b$  pour tous  $t \in \mathbf{R}$  et que les fonctions*

$$I(b, x) = G(T(x, b) - 0, b) \quad \text{et} \quad S(b, x) = G(T(x, b), b)$$

*soient décroissantes et continues par rapport à  $b$  pour tout  $x$  fixé,  $x \in \mathcal{X}$ . Dans ce cas **1**) la statistique  $b_i(\mathbb{X})$ ,*

$$b_i = b_i(\mathbb{X}) = \sup\{b : I(b, \mathbb{X}) \geq \gamma, b \in B\}, \quad \text{si le supremum existe,}$$

*sinon*

$$b_i = b_i(\mathbb{X}) = \inf B$$

*est la limite inférieure de confiance pour  $b \in B^0$  du coefficient de confiance supérieur où égal à  $\gamma$ ;*

**2)** *la statistique  $b_s(\mathbb{X})$  est une limite supérieure de confiance pour  $b \in B^0$  du coefficient de confiance supérieur où égale à  $\gamma$  :*

$$b_s = b_s(\mathbb{X}) = \inf\{b : \mathbf{S}(b, \mathbb{X}) \leq 1 - \gamma, b \in B\}, \quad \text{si le infimum existe,}$$

*sinon*

$$b_s = b_s(\mathbb{X}) = \sup B,$$

3) si  $\mathbf{x}, \mathbf{x} \in \mathcal{X}$ , est telle que les fonctions  $I(b, \mathbf{x})$  et  $S(b, \mathbf{x})$  sont strictement décroissantes par rapport à  $b$ , alors  $b_i(\mathbf{x})$  et  $b_s(\mathbf{x})$  sont les racines des équations

$$I(b_i(\mathbf{x}), \mathbf{x}) = \gamma \quad \text{et} \quad S(b_s(\mathbf{x}), \mathbf{x}) = 1 - \gamma.$$

Démonstration. Notons  $\mathbf{D} = \mathbf{D}(\mathbb{X})$  l'événement suivant

$$\mathbf{D} = \{ \text{il existe } b \text{ tel que } I(b, \mathbb{X}) \geq \gamma \}.$$

Alors pour la vraie valeur  $b \in B^0$  on a

$$\begin{aligned} \mathbf{P}\{b_i < b\} &= \mathbf{P}\{(b_i < b) \cap \mathbf{D}\} + \mathbf{P}\{(b_i < b) \cap \bar{\mathbf{D}}\} = \\ &= \mathbf{P}\{(\sup b^* : I(b^*, \mathbb{X}) \geq \gamma, b^* \in B) < b\} \cap \mathbf{D}\} + \mathbf{P}\{( \inf B < b) \cap \bar{\mathbf{D}}\} = \\ &= \mathbf{P}\{(I(b, \mathbb{X}) < \gamma) \cap \mathbf{D}\} + \mathbf{P}\{\bar{\mathbf{D}}\} \geq \mathbf{P}\{(I(b, \mathbb{X}) < \gamma) \cap \mathbf{D}\} + \mathbf{P}\{(I(b, \mathbb{X}) < \gamma) \cap \bar{\mathbf{D}}\} = \\ &= \mathbf{P}\{I(b, \mathbb{X}) < \gamma\} \geq \gamma, \end{aligned}$$

d'après le Lemme de Bolshev. Le théorème est démontré.

**Remarque 1.** Si  $\theta$  est unidimensionnel, les variables aléatoires  $X_i$  sont continues et la fonction  $F(x; \theta)$  est monotone et continue en  $\theta$ , on peut prendre

$$T(\mathbb{X}; \theta) = -2 \sum_{i=1}^n \ln F(X_i; \theta).$$

D'après le lemme de Bolshev  $F(X_i; \theta) \sim \mathcal{U}(0; 1)$ , donc

$$-2 \ln F(X_i; \theta) = \chi_{2, i}^2, \quad i = 1, \dots, n,$$

et comme  $X_1, \dots, X_n$  sont indépendantes

$$T(\mathbb{X}; \theta) = \chi_{2n}^2.$$

Notons  $G_{2n}(x) = \mathbf{P}\{\chi_{2n}^2 \leq x\}$ . Alors,

$$I(\theta; \mathbb{X}) = S(\theta; \mathbb{X}) = G_{2n}(T(\mathbb{X}; \theta)) = G_{2n}\left(-2 \sum_{i=1}^n \ln F(X_i; \theta)\right).$$

Si les fonctions  $I$  et  $S$  sont strictement décroissantes (ou croissantes) en  $\theta$ , alors d'après le théorème de Bolshev on a

$$-2 \sum_{i=1}^n \ln F(X_i; \underline{\theta}) = \chi_{\gamma}^2(2n) \quad (\text{ou } \chi_{1-\gamma}^2(2n)),$$

$$-2 \sum_{i=1}^n \ln F(X_i; \bar{\theta}) = \chi_{1-\gamma}^2(2n) \quad (\text{ou } \chi_{\gamma}^2(2n)).$$

**Remarque 2.** Soit  $\{\theta_n^*\}$ ,  $n \in \mathbf{N}^*$ , une suite d'estimations,  $\theta_n^* : \mathbf{R}^n \rightarrow \Theta$ , du paramètre  $\theta = (\theta_1, \dots, \theta_m)^T \in \Theta \subset \mathbf{R}^m$ , telle que

$$\sqrt{n}(\theta_n^* - \theta) \xrightarrow{\mathcal{L}} N(\mathbf{0}_m, \Sigma), \quad n \rightarrow \infty.$$

Soit  $g$  une fonction borélienne,  $g : \mathbf{R}^m \rightarrow \mathbf{R}^1$ , différentiable en  $\theta$ . Alors

$$\sqrt{n}[g(\theta_n^*) - g(\theta)] \xrightarrow{\mathcal{L}} N(0, \mathbf{grad}_{\theta}^T \Sigma \mathbf{grad}_{\theta} g), \quad n \rightarrow \infty.$$

En particulier, si  $m = 1$ ,  $\theta_n^* : \mathbf{R}^n \rightarrow \Theta \subset \mathbf{R}^1$ , et  $g : \mathbf{R}^1 \rightarrow \mathbf{R}^1$ ,

$$\sqrt{n}[g(\theta_n^*) - g(\theta)] \xrightarrow{\mathcal{L}} N(0, \sigma^2 [g'(\theta)]^2), \quad n \rightarrow \infty,$$

alors

$$\sqrt{n}[g(\theta_n^*) - g(\theta)] \xrightarrow{\mathcal{L}} N(0, \sigma^2 [g'(\theta)]^2), \quad n \rightarrow \infty.$$

On emploie très souvent des méthodes asymptotiques pour la construction des intervalles de confiance.

**Remarque 3.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$F(x; \theta) = \mathbf{P}\{X_i \leq x\}, \quad \theta = (\theta_1, \dots, \theta_m)^T \in \Theta \subset \mathbf{R}^m.$$

Sous des conditions très générales l'estimateur de maximum de vraisemblance  $\hat{\theta}_n = (\hat{\theta}_1, \dots, \hat{\theta}_m)$  est asymptotiquement efficace et normal avec les paramètres  $\theta$  et  $\mathbf{I}(\theta)$  :

$$\hat{\theta}_n \sim AN(\theta, \mathbf{I}^{-1}(\theta)),$$

où  $\mathbf{I}(\theta)$  est la matrice d'information de Fisher de  $\mathbb{X}$ .

Soit  $b : \mathbf{R}^m \rightarrow \mathbf{R}^1$  une fonction différentiable, alors  $\hat{b}_n = b(\hat{\theta}_n) = b(\hat{\theta}_1, \dots, \hat{\theta}_m)$  un estimateur de paramètre  $b = b(\theta_1, \dots, \theta_m)$ , et  $\hat{b}_n \sim AN(b, \sigma_b^2(\theta))$ , où

$$\sigma_b^2(\theta) = [\mathbf{grad}_{\theta} b(\theta)]^T \mathbf{I}^{-1}(\theta) \mathbf{grad}_{\theta} b(\theta),$$

i.e.

$$(\hat{b}_n - b) / \sigma_b^2(\hat{\theta}_n) \sim AN(0, 1).$$

Donc on peut prendre

$$T(b, \mathbb{X}) = (\hat{b}_n - b) / \sigma_b^2(\hat{\theta}_n).$$

Les fonctions

$$\mathbf{I}(b; \mathbb{X}) = S(b; \mathbb{X}) = \Phi((\hat{b}_n - b) / \sigma_b^2(\hat{\theta}_n))$$

sont décroissantes en  $b$  et d'après le théorème de Bolshev les égalités

$$\Phi((\hat{b}_n - b) / \sigma_b^2(\hat{\theta}_n)) = \gamma, \quad \Phi((\hat{b}_n - b) / \sigma_b^2(\hat{\theta}_n)) = 1 - \gamma$$

implique

$$\underline{b} = \hat{b}_n - z_{\gamma} \sigma_b^2(\hat{\theta}_n); \quad \bar{b} = \hat{b}_n + z_{\gamma} \sigma_b^2(\hat{\theta}_n),$$

où  $z_{\gamma}$  est  $\gamma$ -quantile de la loi normale standard. On peut noter que asymptotiquement  $(\underline{b}, \bar{b})$  est le plus court intervalle de confiance de niveau donné.

## 2.11 Théorème de Fisher.

Dans ce paragraphe nous allons résumer les propriétés principales des estimateurs  $\bar{X}_n$ ,  $S_n^2$  et  $s_n^2$ .

**Théorème de Fisher.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon normal de paramètres  $\mu$  et  $\sigma^2$  :  $X_i \sim N(\mu, \sigma^2)$ . Dans ce cas la statistique  $\tilde{\theta}_n = (\bar{X}_n, S_n^2)^T$  est exhaustive minimale et complète,  $\bar{X}_n$  et  $S_n^2$  sont indépendantes,

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{(n-1)}{\sigma^2} S_n^2 = \chi_{n-1}^2,$$

et la variable aléatoire

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} = t_{n-1}$$

suit la loi de Student à  $n-1$  degré de liberté. L'estimateur  $\tilde{\theta}_n$  est le meilleur estimateur sans biais pour  $\theta = (\mu, \sigma^2)^T$ .

**Remarque 1.** On note qu'une variable aléatoire  $t_f$  suit la loi de Student à  $f$  degrés de liberté,  $f > 0$ , si pour tout  $x \in \mathbf{R}^1$

$$S_f(x) = \mathbf{P}\{t_f \leq x\} = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\sqrt{\pi f} \Gamma\left(\frac{f}{2}\right)} \int_{-\infty}^x \left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}} dt = \int_{-\infty}^x s_f(t) dt.$$

**Exemple 1.** Etudions ici quelques propriétés de la statistiques  $\tilde{\theta}_n = (\bar{X}_n, S_n^2)^T$ . On sait que  $\mathbf{E}\tilde{\theta}_n = \theta = (\mu, \sigma^2)^T$ , où  $\bar{X}_n$  et  $S_n^2$  sont les estimateurs sans biais de  $\mu$  et  $\sigma^2$  respectivement. Nous savons aussi que

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad (1)$$

par suite

$$\mathbf{E}_\theta \bar{X}_n = \mu, \quad \mathbf{Var}_\theta \bar{X}_n = \frac{\sigma^2}{n}. \quad (2)$$

D'autre part on a

$$\frac{n-1}{\sigma^2} S_n^2 = \chi_{n-1}^2, \quad (3)$$

$$\mathbf{E}_\theta S_n^2 = \sigma^2, \quad \mathbf{Var}_\theta S_n^2 = \frac{2\sigma^4}{n-1}. \quad (4)$$

$\tilde{\theta}_n$  est-il un estimateur efficace pour  $\theta = (\mu, \sigma^2)^T$  ?

La fonction de vraisemblance de  $X_j$  est :

$$L_j(\theta) = \frac{1}{\sigma} \varphi\left(\frac{X_j - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X_j - \mu)^2}{2\sigma^2}\right\} \quad (5)$$

et le vecteur informant  $\lambda_j(\theta)$  de l'observation  $X_j$  est :

$$\lambda_j(\theta) = \left(\frac{\partial \ln L_j(\theta)}{\partial \mu}, \frac{\partial \ln L_j(\theta)}{\partial \sigma^2}\right)^T = \left(\frac{X_j - \mu}{\sigma^2}, \frac{(X_j - \mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right)^T. \quad (6)$$

Nous pouvons donc en déduire l'information de Fisher  $i(\theta)$  sur  $\theta$  pour une observation  $X_j$  :

$$\begin{aligned}
i(\theta) &= \mathbf{E} \lambda_j(\theta) \lambda_j^T(\theta) = \\
&= \mathbf{E}_\theta \left\| \begin{array}{cc} \frac{(X_j - \mu)^2}{\sigma^4} & \frac{(X_j - \mu)^3}{2\sigma^6} - \frac{X_j - \mu}{2\sigma^4} \\ \frac{(X_j - \mu)^3}{2\sigma^6} - \frac{X_j - \mu}{2\sigma^4} & \frac{(X_j - \mu)^4}{4\sigma^8} - \frac{(X_j - \mu)^2}{2\sigma^6} + \frac{1}{4\sigma^4} \end{array} \right\| = \\
&= \left\| \begin{array}{cc} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{3}{4\sigma^4} - \frac{1}{2\sigma^4} + \frac{1}{4\sigma^4} \end{array} \right\| = \left\| \begin{array}{cc} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{array} \right\|. \tag{7}
\end{aligned}$$

L'information de Fisher  $I_n(\theta)$  sur  $\theta$  dans  $\mathbb{X}$  est :

$$I_n(\theta) = ni(\theta) = \left\| \begin{array}{cc} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{array} \right\|, \tag{8}$$

par suite

$$I_n^{-1}(\theta) = \left\| \begin{array}{cc} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{array} \right\|. \tag{9}$$

On doit donc avoir (d'après l'inégalité de Rao-Cramer-Frechet) pour tous les estimateurs sans biais  $\mu^*$  et  $\sigma^{*2}$  de  $\mu$  et  $\sigma^2$  :

$$\mathbf{Var}_\theta \mu^* \geq \frac{\sigma^2}{n} \quad \text{et} \quad \mathbf{Var}_\theta \sigma^{*2} \geq \frac{2\sigma^4}{n}. \tag{10}$$

On voit que l'estimateur  $\hat{\mu}_n = \bar{X}_n$  est efficace pour  $\mu$ . Par contre :

$$\mathbf{Var}_\theta S_n^2 = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}, \tag{11}$$

donc  $\sigma^{*2} = S_n^2$  n'est pas efficace pour  $\sigma^2$ , donc  $\tilde{\theta}_n = (\hat{\mu}_n, S_n^2)^T$  n'est pas un estimateur efficace du paramètre  $\theta = (\mu, \sigma^2)^T$ . Nous allons cependant montrer que  $\tilde{\theta}_n = (\hat{\mu}_n, S_n^2)^T$  est le meilleur estimateur sans biais pour  $\theta$ , parce que c'est celui de variance minimum parmi tous les estimateurs sans biais de  $\theta$ . Pour le montrer il suffit de montrer qu'il n'y a pas d'autre estimateur sans biais de  $\sigma^2$  meilleur que  $S_n^2$ . Supposons qu'on ait  $\tilde{\sigma}^2 = \tilde{\sigma}^2(\mathbb{X})$  estimateur sans biais de  $\sigma^2$ ,  $\mathbf{E}_\theta \tilde{\sigma}^2 \equiv \sigma^2$ . Soit  $\delta = \tilde{\sigma}^2 - S_n^2$ . Il est clair que

$$\mathbf{E}_\theta \delta \equiv 0, \tag{12}$$

$\delta = \delta(\mathbb{X})$  est un autre estimateur sans biais de 0. Puisque  $\tilde{\theta}_n$  est exhaustive, on peut écrire l'égalité précédente sous la forme :

$$\frac{1}{(\sqrt{2\pi}\sigma)^n} \int_{\mathbb{R}^n} \delta(\mathbb{X}) \exp \left\{ -\frac{1}{2\sigma^2} [n(\bar{X}_n - \mu)^2 + (n-1)S_n^2] \right\} dX_1 dX_2 \cdots dX_n \equiv 0. \tag{13}$$

En dérivant (13) par rapport à  $\mu$ , on a

$$\frac{1}{(\sqrt{2\pi}\sigma)^n} \int_{\mathbb{R}^n} \delta(\mathbb{X}) \exp \left\{ -\frac{1}{2\sigma^2} [n(\bar{X}_n - \mu)^2 + (n-1)S_n^2] \right\} \times \\ \times \frac{n}{\sigma^2} (\bar{X}_n - \mu) dX_1 dX_2 \cdots dX_n \equiv 0,$$

puis à nouveau en dérivant par rapport à  $\mu$ , on obtient :

$$\frac{1}{(\sqrt{2\pi}\sigma)^n} \int_{\mathbb{R}^n} \delta(\mathbb{X}) \exp \left\{ -\frac{1}{2\sigma^2} [n(\bar{X}_n - \mu)^2 + (n-1)S_n^2] \right\} \times \\ \times \left[ \frac{n^2}{\sigma^4} (\bar{X}_n - \mu)^2 - \frac{n}{\sigma^2} \right] dX_1 dX_2 \cdots dX_n \equiv 0,$$

i.e.,

$$\mathbf{E}_\theta \delta(\mathbb{X})(X_n - \mu) \equiv 0, \quad (14)$$

et donc  $\delta(\mathbb{X})$  et  $\bar{X}_n - \mu$  ne sont pas corrélées. De la même façon on peut montrer que

$$\mathbf{E}_\theta \{ \delta(\mathbb{X}) S_n^2 \} \equiv 0, \quad (15)$$

i.e.,  $\delta(\mathbb{X})$  et  $S_n^2$  ne sont pas corrélées non plus. Mais par ailleurs :

$$\tilde{\sigma}^2 = \delta + S_n^2, \quad (16)$$

d'où

$$\mathbf{Var}_\theta \tilde{\sigma}^2 = \mathbf{Var}_\theta \delta + \mathbf{Var}_\theta S_n^2 \geq \mathbf{Var}_\theta S_n^2. \quad (17)$$

Cela signifie que la variance de  $S_n^2$  est minimale dans la classe de tous les estimateurs sans biais de  $\sigma^2$ , et donc  $S_n^2$  est le meilleur estimateur de  $\sigma^2$  dans ce sens.

On peut obtenir le même résultat sur la complétude de  $\tilde{\theta}_n$  en utilisant le théorème de Lehmann-Scheffé.

**Exemple 2.** Soit  $\mathbb{X}_n = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim N(\mu, 1), \quad |\mu| < \infty,$$

i.e.  $X_i$  suit une loi normale de paramètres

$$\mu = \mathbf{E}X_i \quad \text{et} \quad 1 = \mathbf{Var} X_i.$$

Comme  $X_1, \dots, X_n$  sont indépendantes, on peut aussi dire que  $\mathbb{X}_n$  suit la loi normale de dimension  $n$  :

$$\mathbb{X}_n \sim N_n(\mu \mathbf{1}_n, I_n),$$

où  $\mathbf{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$ ,  $I_n$  est la matrice identité d'ordre  $n$ , et

$$\mathbf{E}\mathbb{X}_n = \mu \mathbf{1}_n, \quad \mathbf{Var} \mathbb{X}_n = I_n.$$

La densité de  $\mathbb{X}_n$  est

$$p_{\mathbb{X}_n}(x; \mu) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (x - \mu \mathbf{1}_n)^T (x - \mu \mathbf{1}_n) \right\} =$$

$$= \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right\}, \quad x = (x_1, \dots, x_n)^T \in \mathbf{R}^n, \quad (18)$$

et donc la fonction de vraisemblance  $L(\mu)$  de  $\mathbb{X}_n$  est

$$L(\mu) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2 \right\}, \quad \mu \in \mathbf{R}^1.$$

Considérons la statistique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \mathbf{1}_n^T \mathbb{X}_n.$$

Comme

$$L(\mu) = \frac{\sqrt{n}}{\sqrt{2\pi}} \exp \left\{ -\frac{n}{2} (\bar{X}_n - \mu)^2 \right\} \frac{1}{\sqrt{n} (2\pi)^{(n-1)/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right\}, \quad (19)$$

du critère de factorisation de Neyman-Fisher il suit que  $\bar{X}_n$  est une statistique exhaustive minimale pour  $\mu$ . Il est évident que  $\bar{X}_n \sim N(\mu, \frac{1}{n})$ . Soit

$$\begin{aligned} W_n &= (X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)^T = \mathbb{X}_n - \bar{X}_n \mathbf{1}_n = \\ &= \mathbb{X}_n - \mathbf{1}_n \bar{X}_n = \mathbb{X}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \mathbb{X}_n = (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \mathbb{X}_n = D_n \mathbb{X}_n, \end{aligned} \quad (20)$$

où

$$D_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T.$$

On note que la matrice  $D_n$  est idempotente, c.a.d. :

$$D_n^T D_n = D_n D_n^T = D_n^2 = D_n,$$

et que  $D_n \mathbf{1}_n = \mathbf{0}_n$ . La formule (20) montre que la statistique  $W_n$  est le résultat d'une transformation linéaire de  $\mathbb{X}_n$ ,  $W_n = D_n \mathbb{X}_n$ , et donc on constate que la statistique  $W_n$  suit une loi normale dans  $\mathbf{R}^n$  dont la fonction caractéristique est

$$f_{W_n}(\mathbf{t}) = \exp \left\{ -\frac{1}{2} \mathbf{t}^T \mathbf{D}_n \mathbf{t} \right\}, \quad \mathbf{t} \in \mathbf{R}^n, \quad (21)$$

puisque

$$\mathbf{E} W_n = D_n \mathbf{E} \mathbb{X}_n = \left( \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \mu \mathbf{1}_n = \mu \mathbf{1}_n - \mu \mathbf{1}_n = \mathbf{0}_n$$

et

$$\begin{aligned} \mathbf{Var} W_n &= \mathbf{E} W_n W_n^T = \mathbf{E} \{ D_n \mathbb{X}_n \mathbb{X}_n^T D_n^T \} = D_n [ \mathbf{I}_n + \mu^2 \mathbf{1}_n \mathbf{1}_n^T ] D_n^T = \\ &= D_n \mathbf{I}_n D_n^T = D_n D_n^T = D_n, \end{aligned}$$

On peut remarquer que la loi de la statistique  $W_n$  ne dépend pas de paramètre  $\mu$ . C'est la raison pour laquelle on dit que  $W_n$  est une statistique *libre*, ce qui signifie que  $W_n$  n'apporte pas d'information sur  $\mu$ . Toute information sur  $\mu$  conserve la statistique exhaustive minimale  $\bar{X}_n$ .

Nous allons montrer que les statistiques  $\bar{X}_n$  et  $W_n$  sont indépendantes. Pour cela il nous faudra étudier plus attentivement la répartition de  $W_n$ . Notons

$$W_i = X_i - \bar{X}_n, \quad i = 1, \dots, n.$$

Il est facile de vérifier que  $\det D_n = 0$ , d'où on déduit que la loi de  $W_n$  est dégénérée, ce qui explique la dépendance linéaire entre  $W_1, \dots, W_n$  :

$$\sum_{i=1}^n W_i = 0, \quad \text{donc} \quad W_n = -(W_1 + \dots + W_{n-1}).$$

Considérons maintenant la statistique  $U_{n-1} = (W_1, \dots, W_{n-1})^T$ . On remarque que

$$\mathbf{E}U_{n-1} = \mathbf{0}_{n-1},$$

et sa matrice de covariance  $B_{n-1}$  est la matrice  $D_n$  sans la dernière ligne ni la dernière colonne. Par un calcul direct, on peut montrer que

$$\det B_{n-1} = \frac{1}{n}, \quad \text{i.e.} \quad \text{rang} B_{n-1} = \text{rang} D_n = n - 1,$$

et donc avec une probabilité 1 la répartition de  $W_n$  est concentrée dans  $\mathbb{R}^{n-1}$ .

On remarque que

$$B_{n-1}^{-1} = \left\| \begin{array}{cccccc} 2 & 1 & 1 & \dots & 1 \\ 1 & 2 & 1 & \dots & 1 \\ 1 & 1 & 2 & \dots & 1 \\ \vdots & & & & \\ 1 & 1 & 1 & \dots & 2 \end{array} \right\|_{n-1, n-1}$$

et  $\det B_{n-1}^{-1} = n$ . De ces résultats il suit que la statistique  $U_{n-1} = (W_1, \dots, W_{n-1})^T$  suit une loi normale  $N_{n-1}(\mathbf{0}_{n-1}, B_{n-1})$ , dont la densité

$$pU_{n-1}(u), \quad u = (u_1, \dots, u_{n-1})^T \in \mathbb{R}^{n-1},$$

est donnée par la formule

$$\begin{aligned} pU_{n-1}(u) &= \frac{1}{\sqrt{\det B_{n-1}} (2\pi)^{(n-1)/2}} \exp \left\{ -\frac{1}{2} u^T B_{n-1}^{-1} u \right\} = \\ &= \frac{\sqrt{n}}{(2\pi)^{(n-1)/2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^{n-1} u_i^2 + \left( \sum_{i=1}^{n-1} u_i \right)^2 \right] \right\}, \quad u \in \mathbb{R}^{n-1}. \end{aligned} \quad (22)$$

Maintenant, il est facile de montrer que  $U_{n-1}$  et  $\bar{X}_n$  sont indépendantes. En effet, considérons la statistique

$$Y = (Y_1, Y_2, \dots, Y_{n-1}, Y_n)^T = C \mathbb{X}_n,$$

où

$$C = \left\| \begin{array}{cccccc} 1 - \frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & -\frac{1}{n} & 1 - \frac{1}{n} & \dots & -\frac{1}{n} & -\frac{1}{n} \\ \vdots & & & & & \\ -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \dots & 1 - \frac{1}{n} & -\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} \end{array} \right\|,$$

et donc

$$Y_n = \bar{X}_n, \quad \text{et} \quad Y_j = W_j = X_j - \bar{X}_n, \quad j = 1, \dots, n-1, \quad (23)$$

d'où il suit que

$$X = C^{-1}Y, \quad \text{où} \quad C^{-1} = \left\| \begin{array}{cccccc} 1 & 0 & 0 & \dots & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & 1 \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & 1 & 1 \\ -1 & -1 & -1 & \dots & -1 & 1 \end{array} \right\|$$

et donc

$$X_i = Y_i + Y_n, \quad i = 1, \dots, n-1,$$

et

$$X_n = nY_n - \sum_{i=1}^{n-1} X_i = Y_n - \sum_{i=1}^{n-1} Y_i.$$

Pour trouver la densité  $p_Y(y; \mu)$  de la statistique  $Y$  on remarque que et

$$J = \det C^{-1} = \det \left\| \frac{\partial x_i}{\partial y_j} \right\| = n,$$

et donc de (18) on obtient que

$$\begin{aligned} p_Y(y; \mu) &= p_{\mathbb{X}_n}(C^{-1}y; \mu) |\det C^{-1}| = \frac{\sqrt{n}}{\sqrt{2\pi}} \exp \left\{ -\frac{n}{2}(y_n - \mu)^2 \right\} \times \\ &\times \frac{\sqrt{n}}{(2\pi)^{(n-1)/2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^{n-1} y_i^2 + \left( \sum_{i=1}^{n-1} y_i \right)^2 \right] \right\}. \end{aligned} \quad (24)$$

De (19) et (24) il suit que  $\bar{X}_n$  et  $U_{n-1} = (X_1 - \bar{X}_n, \dots, X_{n-1} - \bar{X}_n)^T$  sont indépendantes. Comme

$$\mathbf{1}_n^T W_n = \sum_{i=1}^n (X_i - \bar{X}_n) = 0,$$

on tire que

$$X_n - \bar{X}_n = - \sum_{i=1}^{n-1} (X_i - \bar{X}_n),$$

i.e.  $X_n - \bar{X}_n$  est une statistique de  $U_{n-1}$ , qui est indépendante de  $\bar{X}_n$ , et donc  $\bar{X}_n$  et  $W_n = (X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)^T$  sont indépendantes.

On remarque qu'on peut obtenir le même résultat par calcul direct de la fonction caractéristique  $\varphi_V(\mathbf{t})$ ,  $\mathbf{t} \in \mathbf{R}^{n+1}$ , de la statistique

$$\begin{aligned} V &= (W_n, \bar{X}_n) = (X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n, \bar{X}_n)^T \\ \varphi_V(\mathbf{t}) &= \mathbf{E} \exp \left\{ i \left[ \sum_{i=1}^n t_i (X_i - \bar{X}_n) + t_{n+1} \bar{X}_n \right] \right\}. \end{aligned}$$

**Exemple 3.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim N(\mu, \sigma^2), \quad |\mu| < \infty, \quad \sigma^2 > 0.$$

La fonction de vraisemblance  $L(\mu, \sigma^2)$  de  $\mathbb{X}$  est

$$\begin{aligned} L(\mu, \sigma^2) &= p(\mathbb{X}; \mu, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right\} = \\ &= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n X_i^2 - 2\mu \sum_{i=1}^n X_i + n\mu^2 \right]\right\}. \end{aligned} \quad (25)$$

On voit que la statistique

$$\mathbf{T} = \mathbf{T}(\mathbb{X}) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)^T$$

est exhaustive et minimale pour  $(\mu, \sigma^2)^T$ .

Soit

$$\mathcal{X}_{\mathbf{t}} = \{\mathbf{x} : \mathbf{T}(\mathbf{x}) = \mathbf{t} = (t_1, t_2)^T, \quad \mathbf{x} \in \mathbf{R}^n\}.$$

Notons  $c_{\mathbf{t}} = c_{\mathbf{t}}(\mu, \sigma^2)$  la valeur de la densité  $p(\mathbf{x}; \mu, \sigma^2)$  sur cet ensemble. Dans ce cas la loi conditionnelle de  $\mathbb{X}$  sachant  $\mathbf{T}(\mathbb{X}) = \mathbf{t}$  est uniforme sur  $\mathcal{X}_{\mathbf{t}}$ . En effet, pour tout  $\mathbf{x} \in \mathcal{X}_{\mathbf{t}}$  on a

$$\begin{aligned} p_{\mathbb{X}}(\mathbf{x} | \mathbf{T}(\mathbb{X}) = \mathbf{t}; \mu, \sigma^2) &= \frac{p_{\mathbb{X}}(\mathbf{x}; \mu, \sigma^2)}{\int_{\mathcal{X}_{\mathbf{t}}} p_{\mathbb{X}}(\mathbf{x}; \mu, \sigma^2) d\mathbf{x}} = \\ &= \frac{c_{\mathbf{t}}}{c_{\mathbf{t}} \text{mes} \mathcal{X}_{\mathbf{t}}} = \frac{1}{\text{mes} \mathcal{X}_{\mathbf{t}}} = \text{const}. \end{aligned} \quad (26)$$

Considérons la statistique  $\mathbb{Z}_n = (Z_1, \dots, Z_n)^T$ , où

$$Z_j = \frac{X_j - \bar{X}_n}{S_n}, \quad j = 1, \dots, n, \quad (27)$$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 = \frac{1}{n-1} \mathbb{X}^T \mathbf{D}_n \mathbb{X} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (28)$$

Comme les statistiques  $\mathbf{T}$  et  $\mathbf{U} = (\bar{X}_n, S_n^2)^T$  sont équivalentes, on remarque que de (26) et (28) il suit que si  $\mathbf{U}$  est fixée, dans ce cas  $\mathbb{X}$  suit la loi uniforme sur l'intersection de deux surfaces données par les équations :

$$\frac{1}{S_n^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = n-1 \quad \text{et} \quad \frac{1}{S_n} \sum_{i=1}^n (X_i - \bar{X}_n) = 0,$$

ce qui représente la sphère de dimension  $n-1$  avec le centre au point  $\bar{X}_n \mathbf{1}_n$  et de rayon  $\sqrt{n-1} S_n$  dans  $\mathbf{R}^n$ , et par conséquent on en tire que si  $\mathbf{U}$  est fixé, la loi de  $\mathbb{Z}_n$  est uniforme sur l'intersection des deux surfaces données par les équations :

$$\sum_{i=1}^n Z_i^2 = n-1 \quad \text{et} \quad \sum_{i=1}^n Z_i = 0,$$

ce qui représente la sphère de dimension de  $n - 1$  de rayon  $\sqrt{n - 1}$  dans  $\mathbf{R}^n$  dont la surface ne dépend pas de  $\mathbf{U}$  et par conséquent, on obtient que la loi conditionnelle de  $\mathbb{Z}_n$  ne dépend pas de  $\mathbf{U} = (\bar{X}_n, S_n^2)$ , donc les statistiques  $\mathbb{Z}_n$  et  $(\bar{X}_n, S_n^2)$  sont indépendantes. Comme  $\bar{X}_n$  et  $S_n^2$  sont indépendantes il s'ensuit que les trois statistiques  $\mathbb{Z}_n, \bar{X}_n$  et  $S_n^2$  sont indépendantes.

**Exemple 4.** Supposons qu'aux moments  $t = 0, 1, \dots, n$  nous observons un objet  $A$  qui se déplace uniformément avec une vitesse constante et inconnue  $\mu$ ,  $\mu > 0$ . Soit  $s(t)$  la distance parcourue par cet objet  $A$  aux moments  $t = 0, 1, \dots, n$ ;  $n \geq 1$ . Si toutes les mesures étaient correctes on aurait

$$s(k) = \mu k, \quad \text{pour tout } k = 0, 1, 2, \dots, n,$$

(on suppose que  $s(0) = 0$ ).

Supposons que l'expérience soit organisée de manière qu'il n'y ait pas d'erreurs systématiques; il y a cependant des erreurs de mesure qui sont normales et indépendantes et qui s'accumulent à chaque moment de mesure.

En supposant que toutes les erreurs de mesure ont la même variance  $\sigma^2$ , trouvons les meilleurs estimateurs sans biais pour  $\mu$  et  $\sigma^2$ .

Tout d'abord supposons que

$$\mathbf{s} = (s_0, s_1, \dots, s_n)^T, \quad \text{où } s_0 = s(0), s_1 = s(t_1), \dots, s_n = s(t_n),$$

est une réalisation d'un vecteur aléatoire  $\mathbf{S} = (S_0, S_1, \dots, S_n)^T$  dont les coordonnées  $S_i$  selon le modèle physique sont des variables aléatoires telles que

$$S_0 = \delta_0, S_1 = \mu + \delta_1, S_2 = 2\mu + \delta_1 + \delta_2, \dots, S_n = n\mu + \delta_1 + \dots + \delta_n,$$

où toutes les erreurs de mesures  $\delta_0, \delta_1, \dots, \delta_n$  sont indépendantes et suivent la même loi normale  $N(0, \sigma^2)$ . Dans ce cas la fonction de vraisemblance du vecteur des erreurs  $\delta = (\delta_0, \delta_1, \dots, \delta_n)^T$  est

$$L(\delta; \mu, \sigma^2) = (2\pi)^{-(n+1)/2} \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=0}^n \delta_i^2 \right\}.$$

Soit

$$L_{i-1} = S_i - S_{i-1} \quad (i = 1, 2, \dots, n), \quad \text{où } S_0 = \delta_0.$$

Alors

$$\delta_i = L_{i-1} - \mu \quad \text{pour } i = 1, 2, \dots, n,$$

et la fonction de vraisemblance de la statistique  $\mathbf{S}$  est

$$\begin{aligned} L(\mathbf{S}; \mu, \sigma^2) &= (2\pi)^{-(n+1)/2} \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \delta_0^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (L_{i-1} - \mu)^2 \right\} = \\ &= (2\pi)^{-(n+1)/2} \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \delta_0^2 + \sum_{i=1}^n (L_{i-1} - \bar{L}_n)^2 + n(\bar{L}_n - \mu)^2 \right] \right\}, \end{aligned}$$

où

$$\bar{L}_n = \frac{1}{n} \sum_{i=1}^n L_{i-1},$$

on en tire donc que

$$\mathbf{U} = \left( \bar{L}_n, \delta_0^2 + \sum_{i=1}^n (L_{i-1} - \bar{L}_n)^2 \right)^T$$

est une statistique exhaustive pour  $(\mu, \sigma^2)^T$ . Il est évident, que la statistique

$$T_n = \delta_0^2 + \sum_{i=1}^n (L_{i-1} - \bar{L}_n)^2$$

est distribuée comme la variable aléatoire  $\sigma^2 \chi_n^2$ , et on en déduit que

$$\mathbf{E} \left\{ \frac{T_n}{n} \right\} = \sigma^2 \quad \text{et} \quad \mathbf{E} \bar{L}_n = \mu.$$

Comme la famille des distributions normales est complète, la statistique exhaustive  $\mathbf{U}$  est donc complète et on en déduit que

$$\bar{L}_n = \frac{1}{n} \sum_{i=1}^n L_{i-1} = \frac{1}{n} \sum_{i=1}^n (S_i - S_{i-1})$$

et

$$\frac{T_n}{n} = \frac{1}{n} \left[ S_0^2 + \sum_{i=1}^n [(S_i - S_{i-1}) - \bar{L}_n]^2 \right]$$

sont les estimateurs sans biais uniques qui s'expriment en fonction de la statistique exhaustive  $\mathbf{U}$  et par conséquent ils sont les *meilleurs estimateurs sans biais* pour  $\mu$  et  $\sigma^2$ .

## 2.12 Intervalle de confiance pour la moyenne d'une loi normale

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim N(\mu, \sigma^2), \quad |\mu| < \infty, \quad \sigma^2 > 0.$$

Considérons ici le problème d'estimation des paramètres  $\mu$  et  $\sigma^2$  par intervalles. Nous savons que la variable aléatoire

$$t_{n-1} = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} = \sqrt{n-1} \frac{\bar{X}_n - \mu}{s_n}$$

suit la loi de Student à  $f = n - 1$  degrés de liberté

$$\mathbf{P}\{t_f \leq t\} = S_f(t).$$

On donne ici quelques valeurs de  $S_f(t)$  :

$f$	3	4	6	8	16	52
$t$	2.3534	2.1318	1.9432	1.8595	2.1199	2.0024
$S_f(t)$	0.9500	0.9500	0.9500	0.9500	0.9750	0.9750

Nous pouvons trouver pour chaque  $\alpha$ ,  $0 < \alpha < 0.5$ , les valeurs  $t_{n-1}(\alpha)$  et  $\bar{t}_{n-1}(\alpha)$  telles que

$$\begin{cases} \mathbf{P}\{t_{n-1} \leq \underline{t}_{n-1}(\alpha)\} = S_{n-1}(\underline{t}_{n-1}(\alpha)) = \alpha, \\ \mathbf{P}\{t_{n-1} \leq \bar{t}_{n-1}(\alpha)\} = S_{n-1}(\bar{t}_{n-1}(\alpha)) = 1 - \alpha, \end{cases} \quad (1)$$

et donc

$$\mathbf{P}\{t_{n-1}(\alpha) \leq t_{n-1} \leq \bar{t}_{n-1}(\alpha)\} = 1 - 2\alpha. \quad (2)$$

$\bar{t}_{n-1}(\alpha)$  est souvent appelé  $\alpha$ -quantile supérieur ou  $(1 - \alpha)$ -quantile de la loi de Student avec  $f = n - 1$  degrés de liberté et  $\underline{t}_{n-1}(\alpha)$  est appelé  $\alpha$ -quantile inférieur de la loi de Student avec  $f = n - 1$  degrés de liberté. De la symétrie par rapport à zéro de la densité  $s_{n-1}(x)$  nous avons

$$\underline{t}_{n-1}(\alpha) = -\bar{t}_{n-1}(\alpha), \quad (3)$$

et donc (2) peut être présentée

$$\mathbf{P}\{-\bar{t}_{n-1}(\alpha) \leq t_{n-1} \leq \bar{t}_{n-1}(\alpha)\} = 1 - 2\alpha. \quad (4)$$

Les quantiles  $\bar{t}_f(\alpha)$  pour différentes valeurs de  $f$  et  $\alpha$  peuvent être trouvés dans des tables statistiques.

Maintenant en utilisant (4) et le Theoreme de Fisher nous pouvons construire l'intervalle de confiance ou l'estimateur par intervalle pour la moyenne  $\mu$  de la loi normale  $N(\mu, \sigma^2)$ .

Nous disons que l'intervalle aléatoire

$$l(\mathbb{X}) \leq \mu \leq L(\mathbb{X}) \quad (5)$$

est l'intervalle de confiance de niveau  $(1 - \alpha)$  ou l'estimateur par intervalle avec le coefficient de confiance  $(1 - \alpha)$  pour la moyenne inconnue  $\mu$  si

$$\mathbf{P}\{l(\mathbb{X}) \leq \mu \leq L(\mathbb{X})\} = 1 - \alpha. \quad (6)$$

Les statistiques  $l(\mathbb{X})$  et  $L(\mathbb{X})$  s'appellent *limites de confiance inférieure et supérieure* respectivement pour  $\mu$ .

Fixons  $\alpha$  ( $0 < \alpha < 0.5$ ) et choisissons les quantiles

$$\bar{t}_{n-1}(\alpha/2) \quad \text{et} \quad \underline{t}_{n-1}(\alpha/2) = -\bar{t}_{n-1}(\alpha/2),$$

alors du Théorème de Fisher et de (4) on tire que

$$\mathbf{P}\left\{-\bar{t}_{n-1}(\alpha/2) \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq \bar{t}_{n-1}(\alpha/2)\right\} = 1 - \alpha, \quad (7)$$

ou, ce qui est équivalent,

$$\mathbf{P}\left\{\bar{X}_n - \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2) \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2)\right\} = 1 - \alpha. \quad (8)$$

Donc l'intervalle aléatoire

$$\left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2) \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2)\right) \quad (9)$$

est l'intervalle de confiance de niveau  $(1 - \alpha)$  pour  $\mu$ . La limite inférieure de confiance de cet intervalle est

$$l(\mathbb{X}) = \bar{X}_n - \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2),$$

et la limite supérieure de confiance est

$$L(\mathbb{X}) = \bar{X}_n + \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2).$$

**Exercice 1.** La charge d'un électron est  $e = \mu 10^{-10}$ . Miliken a obtenu expérimentalement 58 mesures de  $\mu$ . Les résultats de Miliken sont présentés dans le tableau suivant :

4.781	4.764	4.777	4.809	4.761	4.769
4.795	4.776	4.765	4.790	4.792	4.806
4.769	4.771	4.785	4.779	4.758	4.779
4.792	4.789	4.805	4.788	4.764	4.785
4.779	4.772	4.768	4.772	4.810	4.790
4.775	4.789	4.801	4.791	4.799	4.777
4.772	4.764	4.785	4.788	4.799	4.749
4.791	4.774	4.783	4.783	4.797	4.781
4.782	4.778	4.808	4.740	4.790	
4.767	4.791	4.771	4.775	4.747	

On considère un modèle  $H_0$  où ces résultats sont traités comme des réalisations des variables aléatoires indépendantes  $X_1, X_2, \dots, X_n$  ( $n = 58$ ) qui suivent la même loi normale  $N(\mu, \sigma^2)$ .

**a.** Trouver la statistique exhaustive minimale pour  $\theta = (\mu, \sigma^2)^T$ .

**b.** Trouver l'estimateur de maximum de vraisemblance  $\hat{\theta}_n$  de  $\theta$ .

**c.** Montrer que le meilleur (le plus court) intervalle de confiance de niveau  $P = 1 - \alpha = 0.95$  pour  $\mu$ , sachant que  $\bar{t}_{0.025}(57) = 2.0025$ ,  $\bar{X}_n = 4.7808$  et  $S_n^2 = 23383 \cdot 10^{-8}$ , est

$$4.7768 < \mu < 4.7848.$$

**Exercice 2.** Soit  $x$  une réalisation observée de la somme des carrés des erreurs de mesures dans une expérience. Nous supposons que le nombre de mesures  $f$  est inconnu et que l'expérience est organisée de façon que toutes les mesures puissent être considérées comme des erreurs normales faites dans les mêmes conditions et indépendamment les unes des autres en l'absence d'erreur systématique.

a) Trouver le meilleur estimateur sans biais  $\hat{f}$  pour  $f$ .

b) Supposons que l'expérience donne  $x=407.41$ . En utilisant la distribution asymptotique de  $\hat{f}$  et l'approximation normale de Fisher construire  $\approx 0.9$ -limites de confiance pour  $f$ .

**Solution.** Le nombre  $x$  peut-être observé comme la réalisation de la variable aléatoire

$$\sum_{i=1}^f X_i^2 = \hat{f},$$

où  $\mathbb{X} = (X_1, \dots, X_f)^T$  un échantillon de taille  $f$ ,  $X_i \sim N(0, \sigma^2)$ . Ici  $X_i$  est une erreur de la  $i$ -ème mesure. Il est clair que  $\hat{f}$  suit la loi de chi-deux à  $f$  degrés de liberté, i.e.

$$\mathbf{P}\{\hat{f} \leq x\} = \mathbf{P}\left\{\sum_{i=1}^f X_i^2 \leq x\right\} = \mathbf{P}\{\chi_f^2 \leq x\} = Q_f(x), \quad x \geq 0.$$

Comme  $\mathbf{E}\chi_f^2 = f$ , la statistique  $\hat{f}$  est l'estimateur sans biais de  $f$ . On sait que la variable aléatoire

$$\sqrt{2\hat{f}} - \sqrt{2f-1} = \sqrt{2\chi_f^2} - \sqrt{2f-1}$$

est asymptotiquement normale (approximation de Fisher), quand  $f \rightarrow \infty$ , i.e. pour tout  $z$  fixé

$$\mathbf{P}\{\sqrt{2\hat{f}} - \sqrt{2f-1} \leq z\} \approx \Phi(z),$$

pour les grandes valeurs de  $f$ . De cette égalité on déduit

$$\mathbf{P}\{-1.28 \leq \sqrt{2\hat{f}} - \sqrt{2f-1} \leq 1.28\} \approx 0.8,$$

puisque  $\Phi^{-1}(0.9) = \bar{x}_{0.1} = 1.28$ , et donc on obtient l'intervalle de confiance pour  $f$

$$\mathbf{P}\left\{\frac{1}{2} + \frac{1}{2} \left(\sqrt{2\hat{f}} - 1.28\right)^2 \leq f \leq \frac{1}{2} + \frac{1}{2} \left(\sqrt{2\hat{f}} + 1.28\right)^2\right\} \approx 0.8.$$

Comme  $\hat{f} = 407.81$ ,  $\sqrt{2\hat{f}} = 28.54$ , on en tire que

$$373 \leq f \leq 445.$$

Il est utile de remarquer que pour avoir l'estimateur par intervalle de confiance (9) avec le coefficient de confiance  $1 - \alpha$  nous devons choisir les quantiles de niveau  $\alpha/2$ . Il faut remarquer encore que la longueur  $L_n$  de cette intervalle est une variable aléatoire

$$L_n = 2 \frac{S_n}{\sqrt{n}} \bar{t}_{n-1}(\alpha/2) \quad (10)$$

et puisque

$$\mathbf{E}S_n = \sqrt{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \sigma,$$

(voir, par exemple, Voinov & Nikulin (1993)), on en tire que

$$\mathbf{E}L_n = 2\sigma \bar{t}_{n-1}(\alpha/2) \sqrt{\frac{2}{n(n-1)}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}. \quad (11)$$

D'un autre côté nous savons que pour chaque  $x \in \mathbb{R}^1$

$$S_f(x) = \mathbf{P}\{t_f \leq x\} \rightarrow \Phi(x), \quad \text{quand } f \rightarrow \infty,$$

et en plus (voir, par exemple, Huber et Nikulin (1992)),

$$S_f(x) - \Phi(x) = O(1/\sqrt{f})$$

uniformément par rapport à  $x$ ,  $x \in R^1$ , et donc de (11) il suit que pour grandes valeurs de  $n$

$$\mathbf{E}L_n = \frac{2\sigma}{\sqrt{n}}\bar{x}(\alpha/2) + O\left(\frac{1}{n^{3/2}}\right) \quad (12)$$

où  $\bar{x}(\alpha/2) = \bar{x}_{\alpha/2}$  est le quantile supérieur de niveau  $\alpha/2$  de la loi standard normale. Puisque  $S_n^2$  est un estimateur sans biais de  $\sigma^2$ ,  $\mathbf{E}S_n^2 = \sigma^2$ , alors de (10) il suit que

$$\mathbf{E}L_n^2 = \frac{4\sigma^2}{n}t_{n-1}^2(\alpha/2),$$

et donc

$$\mathbf{Var}L_n = \mathbf{E}L_n^2 - (\mathbf{E}L_n)^2 = \frac{4\sigma^2}{n}t_{n-1}^2(\alpha/2) \left[1 - \frac{2}{n-1} \frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)}\right]. \quad (13)$$

Puisque

$$1 - \frac{2}{n-1} \frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} = \frac{1}{2n} + O\left(\frac{1}{n^2}\right), \quad (n \rightarrow \infty)$$

de (13) il suit que pour les grandes valeurs de  $n$

$$\mathbf{Var}L_n = \frac{2\sigma^2}{n^2}\bar{x}^2(\alpha/2) + O\left(\frac{1}{n^3}\right), \quad (14)$$

et donc on peut dire que  $L_n$  est pratiquement constante,  $L_n \approx \mathbf{E}L_n$ . En pratique cela signifie que

$$L_n = \frac{2\sigma}{\sqrt{n}}\bar{x}(\alpha/2), \quad (15)$$

quand  $n$  est assez grand.

Supposons maintenant que la variance  $\sigma^2$  est connue. Comment cette information change l'intervalle de confiance pour  $\mu$ ? Si  $\sigma^2$  est donné, dans ce cas  $\bar{X}_n$  est une statistique exhaustive pour paramètre  $\mu$  et, comme il est bien connu,  $\bar{X}_n$  est le meilleur estimateur sans biais pour  $\mu$  et suit la loi normal  $N(\mu, \sigma^2/n)$ ,  $|\mu| < \infty$ . Donc la variable aléatoire

$$Z = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

suit la loi normal standard  $N(0, 1)$ .

Il s'en suit que lorsqu'on choisit  $\bar{x}(\alpha/2)$ ,  $0 < \alpha < 0.5$ , comme le quantile supérieur de niveau  $\alpha/2$  de la loi normale standard, alors on a

$$\mathbf{P}\{-\bar{x}(\alpha/2) \leq Z \leq \bar{x}(\alpha/2)\} = 1 - \alpha$$

ou, ce qui est la même chose,

$$\mathbf{P}\{-\bar{x}(\alpha/2) \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq \bar{x}(\alpha/2)\} = 1 - \alpha,$$

d'où on obtient l'intervalle de confiance de longueur minimale avec le coefficient de confiance  $(1 - \alpha)$  pour  $\mu$ :

$$\mathbf{P}\left\{\bar{X}_n - \frac{\sigma}{\sqrt{n}}\bar{x}(\alpha/2) \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}\bar{x}(\alpha/2)\right\} = 1 - \alpha, \quad (16)$$

Par exemple, si

$$\alpha = 0.05, \quad \text{alors} \quad 1 - \alpha = 0.95, \quad \alpha/2 = 0.025, \quad \bar{x}(0.025) = 1.96$$

et donc dans ce cas particulier on obtient

$$\mathbf{P}\left\{\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right\} = 0.95, \quad (17)$$

et on dit que avec la probabilité 0.95 l'intervalle aléatoire

$$\left(\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right)$$

inclu ou couvre la vraie (mais inconnue !) valeur de  $\mu$ .

La longueur  $L_n$  de l'intervalle de confiance (16) est

$$L_n = \frac{2\sigma}{\sqrt{n}} \bar{x}(\alpha/2) \quad (18)$$

et comme on le voit de (15) il coïncide avec la longueur moyenne de l'intervalle de confiance pour  $\mu$  quand  $\sigma^2$  est inconnu et il n'est pas aléatoire !

## 2.13 Intervalle de confiance pour la variance d'une loi normale

Nous voulons maintenant construire l'intervalle de confiance de niveau  $(1 - \alpha)$  pour la variance  $\sigma^2$  de la loi normale  $N(\mu, \sigma^2)$ . Considérons d'abord le cas où  $\mu$  est aussi inconnue. Le Théorème de Fisher nous dit que

$$\frac{n-1}{\sigma^2} S_n^2 = \chi_{n-1}^2, \quad (1)$$

où

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (2)$$

est un meilleur estimateur sans biais de  $\sigma^2$ . Pour chaque  $\alpha$  fixé,  $0 < \alpha < 0.5$ , on peut trouver des tables statistiques des quantiles

$$\underline{\chi}_{n-1}^2(\alpha/2) \quad \text{et} \quad \bar{\chi}_{n-1}^2(\alpha/2)$$

tels que

$$\mathbf{P}\{\chi_{n-1}^2 \leq \underline{\chi}_{n-1}^2(\alpha/2)\} = \frac{\alpha}{2} \quad \text{et} \quad \mathbf{P}\{\chi_{n-1}^2 \leq \bar{\chi}_{n-1}^2(\alpha/2)\} = 1 - \frac{\alpha}{2}, \quad (3)$$

c'est-à-dire

$$\mathbf{P}\{\underline{\chi}_{n-1}^2(\alpha/2) \leq \chi_{n-1}^2 \leq \bar{\chi}_{n-1}^2(\alpha/2)\} = 1 - \alpha. \quad (4)$$

De (1) et (4) on a

$$\mathbf{P}\{\underline{\chi}_{n-1}^2(\alpha/2) \leq \frac{n-1}{\sigma^2} S_n^2 \leq \bar{\chi}_{n-1}^2(\alpha/2)\} = 1 - \alpha$$

et donc

$$\mathbf{P}\left\{\frac{(n-1)S_n^2}{\bar{\chi}_{n-1}^2(\alpha/2)} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\underline{\chi}_{n-1}^2(\alpha/2)}\right\} = 1 - \alpha. \quad (5)$$

Voilà pourquoi l'intervalle aléatoire

$$\frac{(n-1)S_n^2}{\bar{\chi}_{n-1}^2(\alpha/2)} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\underline{\chi}_{n-1}^2(\alpha/2)} \quad (6)$$

est appelé l'intervalle de confiance de niveau  $(1 - \alpha)$  ou l'estimateur par intervalle avec le coefficient de confiance  $(1 - \alpha)$  pour la variance  $\sigma^2$  de la loi normale  $N(\mu, \sigma^2)$  quand  $\mu$  est inconnue. La longueur  $L_n$  de cet intervalle est égale à

$$L_n = (n-1)S_n^2 \left( \frac{1}{\underline{\chi}_{n-1}^2(\alpha/2)} - \frac{1}{\bar{\chi}_{n-1}^2(\alpha/2)} \right).$$

Il faut remarquer ici qu'à l'aide de (5) on peut construire l'intervalle de confiance de niveau  $(1 - \alpha)$  pour  $\sigma$ .

Ici nous donnons quelques valeurs de la fonction de répartition  $Q_f(x)$  de  $\chi_f^2$  :

$$Q_f(x) = \mathbf{P}\{\chi_f^2 \leq x\} = \frac{1}{2^{f/2} \Gamma\left(\frac{f}{2}\right)} \int_0^x y^{f/2-1} e^{-y/2} dy, \quad x \geq 0.$$

$f$	1	1	3	4	4	4	57	57
$x$	3.844	2.706	7.815	9.488	7.779	0.711	79.572	38.027
$Q_f(x)$	0.950	0.900	0.950	0.950	0.900	0.050	0.975	0.025

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon de taille  $n = 5$ ,  $X_i \sim N(\mu, \sigma^2)$ , et  $\mu$  et  $\sigma^2$  sont inconnus. On va construire le plus court intervalle de confiance de niveau  $(1 - \alpha)$  pour  $\mu$ , quand  $\alpha = 0.1$  et

$$X_1 = 2.96, \quad X_2 = 3.07, \quad X_3 = 3.02, \quad X_4 = 2.98, \quad X_5 = 3.06.$$

D'après (10.9) l'intervalle le plus court de confiance de niveau  $(1 - \alpha)$  pour  $\mu$  est

$$\bar{X}_n - \bar{t}_{n-1} \left( \frac{\alpha}{2} \right) \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \bar{t}_{n-1} \left( \frac{\alpha}{2} \right) \frac{S_n}{\sqrt{n}}.$$

Dans notre cas

$$\bar{X}_n = \bar{X}_5 = 3.018, \quad S_n^2 = S_5^2 = 0.00232, \quad \frac{S_5^2}{5} = 0.000464, \quad \frac{S_5}{\sqrt{5}} = 0.046,$$

$$\alpha/2 = 0.05, \quad \bar{t}_{n-1} \left( \frac{\alpha}{2} \right) = \bar{t}_4(0.05) = 2.132$$

et donc le plus court intervalle pour  $\mu$

$$2.972 \leq \mu \leq 3.064.$$

Construisons maintenant l'intervalle de confiance de niveau  $(1 - \alpha)$  pour  $\sigma^2$ , si  $\alpha = 0.01$ . D'après (11.6) l'intervalle de confiance de niveau 0.90 pour  $\sigma^2$  est

$$\frac{4S_5^2}{\bar{\chi}_4^2(0.05)} \leq \sigma^2 \leq \frac{4S_5^2}{\underline{\chi}_4^2(0.05)}.$$

Puisque dans notre cas

$$S_5^2 = 0.00232, \quad \bar{\chi}_4^2(0.05) = 0.711 \quad \text{and} \quad \underline{\chi}_4^2(0.05) = 9.488$$

nous obtenons la réalisation de l'intervalle de confiance de niveau 0.9 pour  $\sigma^2$  :

$$0.00098 \leq \sigma^2 \leq 0.0131.$$

Supposons maintenant que  $\mu$  est connu et il nous faut estimer  $\sigma^2$ . Il est évident que dans ce cas la statistique

$$\tilde{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (7)$$

est le meilleur estimateur sans biais de  $\sigma^2$  :

$$\mathbf{E}\tilde{s}_n^2 = \sigma^2, \quad (8)$$

et comme  $X_1, \dots, X_n$  sont indépendants et  $(X_i - \mu)/\sigma$  suit la loi normale standard  $N(0, 1)$ , on trouve que

$$n \frac{\tilde{s}_n^2}{\sigma^2} = \chi_n^2. \quad (9)$$

Pour chaque fixé  $\alpha$ ,  $0 < \alpha < 0.5$ , on peut trouver des tables statistiques les quantiles

$$\underline{\chi}_n^2(\alpha/2) \quad \text{et} \quad \bar{\chi}_n^2(\alpha/2)$$

tels que

$$\mathbf{P}\{\chi_n^2 \leq \underline{\chi}_n^2(\alpha/2)\} = \frac{\alpha}{2} \quad \text{et} \quad \mathbf{P}\{\chi_n^2 \leq \bar{\chi}_n^2(\alpha/2)\} = 1 - \frac{\alpha}{2}, \quad (10)$$

i.e.

$$\mathbf{P}\{\underline{\chi}_n^2(\alpha/2) \leq \chi_n^2 \leq \bar{\chi}_n^2(\alpha/2)\} = 1 - \alpha. \quad (11)$$

De (9) et (11) nous obtenons

$$\mathbf{P}\{\underline{\chi}_n^2(\alpha/2) \leq \frac{n\tilde{s}_n^2}{\sigma^2} \leq \bar{\chi}_n^2(\alpha/2)\} = 1 - \alpha. \quad (12)$$

et donc

$$\mathbf{P}\left\{ \frac{n\tilde{s}_n^2}{\bar{\chi}_n^2(\alpha/2)} \leq \sigma^2 \leq \frac{n\tilde{s}_n^2}{\underline{\chi}_n^2(\alpha/2)} \right\} = 1 - \alpha. \quad (13)$$

C'est pourquoi l'intervalle aléatoire

$$\frac{ns_n^2}{\bar{\chi}_n^2(\alpha/2)} \leq \sigma^2 \leq \frac{ns_n^2}{\underline{\chi}_n^2(\alpha/2)} \quad (14)$$

est appelé l'intervalle de confiance ou l'estimateur par intervalles avec le coefficient de confiance  $1 - \alpha$  pour la variance  $\sigma^2$  de la loi normale  $N(\mu, \sigma^2)$ , quand  $\mu$  est connu.

En pratique on choisit souvent pour le coefficient de confiance  $1 - \alpha = 0.90$  ou  $0.95$ , ou  $0.99$ , ce qui correspond à  $\alpha$  égale à  $0.1$ ,  $0.05$  ou  $0.01$  respectivement.

**Exemple 2.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon de taille  $n = 201$ ,  $X_i \sim N(\mu, \sigma^2)$ , et soit

$$S_{201}^2 = \frac{1}{200} \sum_{i=1}^{201} (X_i - \bar{X}_n)^2$$

un meilleur estimateur sans biais pour  $\sigma^2$ . Il faut évaluer la probabilité

$$\mathbf{P}\{0.8\sigma^2 \leq S_{201}^2 \leq 1.2\sigma^2\}.$$

*Solution.* Comme nous savons

$$\frac{200}{\sigma^2} S_{201}^2 = \chi_{200}^2$$

et donc

$$\mathbf{P}\{0.8\sigma^2 \leq S_{201}^2 \leq 1.2\sigma^2\} = \mathbf{P}\{160 < \frac{200}{\sigma^2} S_{201}^2 < 240\} = \mathbf{P}\{160 < \chi_{200}^2 < 240\}.$$

Pour calculer cette probabilité on peut utiliser l'approximation normale simple pour la loi chi-deux, d'après laquelle pour chaque  $x \in \mathbf{R}^1$

$$\mathbf{P}\left\{\frac{\chi_f^2 - f}{\sqrt{2f}} < x\right\} \rightarrow \Phi(x), \quad \text{quand } f \rightarrow \infty,$$

et donc

$$\begin{aligned} \mathbf{P}\{160 < \chi_{200}^2 < 240\} &= \mathbf{P}\left\{-\frac{40}{20} < \frac{\chi_{200}^2 - 200}{20} < \frac{40}{20}\right\} = \\ \mathbf{P}\left\{-2 < \frac{\chi_{200}^2 - 200}{20} < 2\right\} &\approx 2\Phi(2) - 1 = 2 \cdot 0.9772 - 1 = 0.9544, \end{aligned}$$

c'est-à-dire

$$\mathbf{P}\{0.8\sigma^2 \leq S_{201}^2 \leq 1.2\sigma^2\} \approx 0.9544.$$

Meilleure approximation pour  $\mathbf{P}\{0.8\sigma^2 \leq S_{201}^2 \leq 1.2\sigma^2\}$  peut être obtenue à partir de l'approximation normale de Fisher, d'après laquelle pour chaque  $x \in \mathbf{R}^1$

$$\mathbf{P}\{\sqrt{2\chi_f^2} - \sqrt{2f-1} < x\} \rightarrow \Phi(x), \quad \text{quand } f \rightarrow \infty.$$

En utilisant cette approximation, nous avons

$$\mathbf{P}\{0.8\sigma^2 \leq S_{201}^2 \leq 1.2\sigma^2\} = \mathbf{P}\{4\sqrt{10} < \chi_{200} < 4\sqrt{15}\} =$$

$$\begin{aligned}
& \mathbf{P}\{8\sqrt{5} - 20 < \sqrt{2\chi_{200}^2} - \sqrt{400} < 4\sqrt{30} - 20\} \\
&= \mathbf{P}\{-2.112 < \sqrt{2\chi_{200}^2} - \sqrt{400} < 1.908\} \\
&\approx \Phi(1.908) + \Phi(-2.112) = 0.9718 + 0.9827 - 1 = 0.9545.
\end{aligned}$$

Il faut remarquer ici que la valeur exacte (avec 5 chiffres décimaux) est

$$\begin{aligned}
& \mathbf{P}\{0.8\sigma^2 \leq S_{201}^2 \leq 1.2\sigma^2\} = \mathbf{P}\{160 < \chi_{200}^2 < 240\} \\
&= 0.98292 - 0.02796 = 0.95496 \approx 0.9550.
\end{aligned}$$

**Exemple 3.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon de taille  $n = 16$ ,  $X_i \sim N(\mu, \sigma^2)$ . Calculons la probabilité

$$\mathbf{P}\left\{|\bar{X}_n - \mu| < \frac{3}{\sqrt{n}}S_n\right\},$$

où

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

sont les meilleurs estimateurs sans biais pour  $\mu$  et  $\sigma^2$ .

D'après le Théorème de Fisher la variable aléatoire

$$t_{n-1} = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$$

suit la loi de Student avec  $f = n - 1 = 15$  degrés de liberté et donc nous avons

$$\begin{aligned}
\mathbf{P}\left\{|\bar{X}_n - \mu| < \frac{3}{\sqrt{n}}S_n\right\} &= \mathbf{P}\{|t_{15}| < 3\} = \int_{-3}^3 s_{15}(x) dx \\
&= 2 \int_0^3 s_{15}(x) dx = 2S_{15}(3) - 1 = 0.991,
\end{aligned}$$

où  $s_{15}(x)$  est la densité de la loi de Student à 15 degré de liberté et  $S_{15}(x)$  sa fonction de répartition. On peut remarquer que si l'on utilise l'approximation normale pour l'estimation de la même probabilité, on aura

$$\mathbf{P}\left\{|\bar{X}_n - \mu| < \frac{3}{\sqrt{n}}S_n\right\} \approx 0.9973 > 0.991$$

pour chaque  $n$ .

**Exemple 4.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon, dont  $X_i \sim N(\mu, \sigma^2)$ . On va chercher, en utilisant l'approximation normale, la plus petite valeur de  $n = n(\varepsilon)$  pour lequel

$$\mathbf{P}\left\{\frac{|S_n^2 - \sigma^2|}{\sigma^2} < \varepsilon\right\} \geq 0.9,$$

quand  $\varepsilon = 0.5$  et  $\varepsilon = 0.05$ . Du Théorème de Fisher il suit que

$$\mathbf{P}\left\{\frac{|S_n^2 - \sigma^2|}{\sigma^2} < \varepsilon\right\} = \mathbf{P}\left\{(n-1)(1-\varepsilon) < (n-1)\frac{S_n^2}{\sigma^2} < (n-1)(1+\varepsilon)\right\} =$$

$$\mathbf{P}\{(n-1)(1-\varepsilon) < \chi_{n-1}^2 < (n-1)(1+\varepsilon)\}.$$

Du Théorème limite central il suit que  $\chi_f^2$  est asymptotiquement normale pour les grandes valeurs de  $f$  et donc en utilisant l'approximation normale on obtient

$$\begin{aligned} \mathbf{P}\left\{\frac{|S_n^2 - \sigma^2|}{\sigma^2} < \varepsilon\right\} &= \mathbf{P}\{(n-1)(1-\varepsilon) < \chi_{n-1}^2 < (n-1)(1+\varepsilon)\} = \\ &= \mathbf{P}\left\{-\varepsilon\sqrt{\frac{n-1}{2}} < \frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} < \varepsilon\sqrt{\frac{n-1}{2}}\right\} \approx \\ &= \Phi\left(\varepsilon\sqrt{\frac{n-1}{2}}\right) - \Phi\left(-\varepsilon\sqrt{\frac{n-1}{2}}\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n-1}{2}}\right) - 1, \end{aligned}$$

d'où il suit que

$$2\Phi\left(\varepsilon\sqrt{\frac{n-1}{2}}\right) - 1 \geq 0.9,$$

si

$$\Phi\left(\varepsilon\sqrt{\frac{n-1}{2}}\right) \geq 0.95,$$

et comme  $\Phi$  est croissante, la dernière inégalité est équivalente à la suivante :

$$\varepsilon\sqrt{\frac{n-1}{2}} \geq \Phi^{-1}(0.95) = 1.645,$$

d'où on tire que la plus petite valeur de  $n = n(\varepsilon)$  vérifie la relation suivante :

$$\varepsilon\sqrt{\frac{n-1}{2}} \approx 1.6,$$

i.e.

$$n \approx 1 + 5.2/\varepsilon^2.$$

Par exemple, si  $\varepsilon = 0.5$ , alors  $n \approx 21$ , et par calculs directs on obtient que

$$\mathbf{P}\{10 < \chi_{20}^2 < 30\} = 0.8973 < 0.9,$$

mais pour  $n = 22$  on a

$$\mathbf{P}\{10.5 < \chi_{21}^2 < 31.5\} = 0.901 > 0.9,$$

et donc pour  $\varepsilon = 0.5$  la plus petite valeur de  $n = n(0.5) = 22$ . Dans le cas  $\varepsilon = 0.05$  nous pouvons résoudre le problème asymptotiquement et nous aurons

$$n \approx 1 + 2\frac{2.6}{\varepsilon^2} = 1 + 2\frac{2.6}{0.0025} = 2080.$$

**Exemple 5.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i \sim N(\mu, \sigma^2)$ , où  $\mu$  et  $\sigma^2$  sont inconnus. Notre but est de construire un intervalle de prédiction pour une nouvelle observation  $X_{n+1}$ ,  $X_{n+1} \sim N(\mu, \sigma^2)$ , qui est indépendante de  $\mathbb{X}$ .

Comme  $\mathbb{X}$  est un échantillon normale  $N(\mu, \sigma^2)$ , nous pouvons travailler avec la statistique exhaustive minimale

$$U = (\bar{X}_n, S_n^2)^T,$$

où

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

sont les meilleurs estimateurs sans biais pour  $\mu$  et  $\sigma^2$ ,

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{n-1}{\sigma^2} S_n^2 = \chi_{n-1}^2,$$

$\bar{X}_n$  et  $S_n^2$  sont indépendantes. Puisque  $X_{n+1}$  et  $\mathbb{X}$  sont indépendants, alors  $X_{n+1}$  est indépendante de  $\bar{X}_n$  et  $S_n^2$ , et donc

$$X_{n+1} - \bar{X}_n \sim N\left(0, \frac{n+1}{n} \sigma^2\right),$$

i.e. la variable aléatoire

$$Z = \frac{X_{n+1} - \bar{X}_n}{\sigma \sqrt{1 + \frac{1}{n}}}$$

suit la loi normale standard,  $Z \sim N(0, 1)$ . Il est évident que  $Z$  est indépendante de  $S_n^2$  et donc la statistique

$$\tau = \frac{Z}{\sqrt{S_n^2/\sigma^2}} = \frac{X_{n+1} - \bar{X}_n}{S_n \sqrt{1 + \frac{1}{n}}}$$

suit la loi de Student avec  $n - 1$  degrés de liberté. c'est-à-dire

$$\mathbf{P}\{\tau \leq t\} = S_{n-1}(t).$$

Par conséquent

$$\mathbf{P}\left\{-\bar{t}_{n-1}\left(\frac{\alpha}{2}\right) \leq \frac{X_{n+1} - \bar{X}_n}{S_n \sqrt{1 + \frac{1}{n}}} \leq \bar{t}_{n-1}\left(\frac{\alpha}{2}\right)\right\} = 1 - \alpha$$

d'où il suit que

$$\mathbf{P}\left\{\bar{X}_n - S_n \sqrt{1 + \frac{1}{n}} \bar{t}_{n-1}\left(\frac{\alpha}{2}\right) \leq X_{n+1} \leq \bar{X}_n + S_n \sqrt{1 + \frac{1}{n}} \bar{t}_{n-1}\left(\frac{\alpha}{2}\right)\right\} = 1 - \alpha.$$

L'intervalle

$$\bar{X}_n - S_n \sqrt{1 + \frac{1}{n}} \bar{t}_{n-1}\left(\frac{\alpha}{2}\right) \leq X_{n+1} \leq \bar{X}_n + S_n \sqrt{1 + \frac{1}{n}} \bar{t}_{n-1}\left(\frac{\alpha}{2}\right)$$

est connu sous le nom du *plus court intervalle de prédiction* de niveau de confiance  $1 - \alpha$  pour une seule nouvelle observation  $X_{n+1}$ ,  $X_{n+1} \sim N(\mu, \sigma^2)$ .

Par exemple, supposons que  $n = 5$ , et

$$X_1 = -0.79, \quad X_2 = -0.89, \quad X_3 = 0.32, \quad X_4 = 0.50, \quad X_5 = -0.20.$$

Dans ce cas  $\bar{X}_5 = -0.212$ ,  $S_5^2 = 0.3960$ ,

$$S_n \sqrt{1 + \frac{1}{n}} = S_5 \sqrt{1 + \frac{1}{5}} = \sqrt{0.47517} = 0.689,$$

et puisque  $\bar{t}_4(0.025) = 2.776$ , l'intervalle de prédiction pour  $X_6$  est

$$-2.125 \leq X_6 \leq 1.701.$$

**Exemple 6.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i \sim N(\mu, 1)$ , où  $\mu$  est inconnu, et il nous faut construire l'intervalle de prédiction pour une nouvelle observation  $X_{n+1}$ ,  $X_{n+1} \sim N(\mu, 1)$ , qui est indépendante de  $\mathbb{X}$ . Il est clair que dans ce cas la variable aléatoire

$$Z = \frac{X_{n+1} - \bar{X}_n}{\sqrt{1 + \frac{1}{n}}}$$

suit la loi normale standard et donc

$$\mathbf{P} \left\{ \left| \frac{X_{n+1} - \bar{X}_n}{\sqrt{1 + \frac{1}{n}}} \right| < \bar{x}(\alpha/2) \right\} = 1 - \alpha,$$

où  $\bar{x}(\alpha/2)$  est  $\alpha/2$ -quantile supérieur de la loi normale standard. Par exemple, si  $\alpha = 0.05$ , alors pour les données de l'exemple 4 nous avons

$$\bar{x}(\alpha/2) = \bar{x}(0.025) = 1.96$$

et par conséquent l'intervalle de prédiction pour  $X_6$  est

$$|X_6 + 0.212| < 1.96\sqrt{1.2} = 1.96 \cdot 1.095 = 2.15,$$

ou

$$-2.36 < X_6 < 1.94.$$

## 2.14 Intervalle de confiance pour la différence des moyennes de deux lois normales

Soient  $\mathbb{X} = (X_1, \dots, X_m)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_n)^T$  deux échantillons,

$$X_i \sim N(\mu_X, \sigma_X^2), \quad Y_j \sim N(\mu_Y, \sigma_Y^2).$$

Supposons que  $\mathbb{X}$  et  $\mathbb{Y}$  sont indépendants. Notre but est d'estimer  $\mu_X - \mu_Y$ . D'abord on étudie le cas quand  $\sigma_Y^2$  et  $\sigma_X^2$  sont connues. Dans notre problème la statistique  $T = (\bar{X}_m, \bar{Y}_n)^T$  est exhaustive pour  $\mu = (\mu_X, \mu_Y)^T$ , où

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j \quad (1)$$

sont les meilleurs estimateurs sans biais pour  $\mu_X$ , et  $\mu_Y$ , et comme on le sais déjà bien

$$\bar{X}_m \sim N\left(\mu_X, \frac{\sigma_X^2}{m}\right) \quad \text{et} \quad \bar{Y}_n \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right). \quad (2)$$

Par conséquent, la statistique  $\bar{X}_m - \bar{Y}_n$  est le meilleur estimateur sans biais pour  $\mu_X - \mu_Y$  et

$$\bar{X}_m - \bar{Y}_n \sim \left(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right). \quad (3)$$

Il suit de (3) que la variable aléatoire

$$Z = \frac{\bar{X}_m - \bar{Y}_n - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \quad (4)$$

suit la loi normale standard,  $Z \sim N(0, 1)$ , et donc

$$\mathbf{P} \left\{ -\bar{x}(\alpha/2) \leq \frac{\bar{X}_m - \bar{Y}_n - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \leq \bar{x}(\alpha/2) \right\} = 1 - \alpha, \quad (5)$$

ou, ce qui est équivalent,

$$\mathbf{P} \left\{ \bar{X}_m - \bar{Y}_n - \bar{x}(\alpha/2) \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \leq \mu_X - \mu_Y \leq \bar{X}_m - \bar{Y}_n + \bar{x}(\alpha/2) \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \right\} = 1 - \alpha. \quad (6)$$

Cette formule donne le plus court intervalle de confiance de niveau  $(1 - \alpha)$  pour la différence  $\mu_X - \mu_Y$  quand les variances  $\sigma_X^2$  et  $\sigma_Y^2$  sont connues.

### §15. Intervalle de confiance pour la différence des moyennes de deux lois normales quand les variances sont inconnues.

Soient  $\mathbb{X} = (X_1, \dots, X_m)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_n)^T$  deux échantillons normales indépendants,

$$X_i \sim N(\mu_X, \sigma_X^2), \quad Y_j \sim N(\mu_Y, \sigma_Y^2), \quad (1)$$

et on s'intéresse à l'estimation de  $\mu_X - \mu_Y$ , quand  $\mu_X$  et  $\mu_Y$  sont inconnues et

$$\sigma_Y^2 = \sigma_X^2 = \sigma^2,$$

où  $\sigma^2$  est aussi inconnue. Il est évident que

$$T = (\bar{X}_m, \bar{Y}_n, S_X^2, S_Y^2)^T \quad (2)$$

est une statistique exhaustive pour  $\theta = (\mu_X, \mu_Y, \sigma^2)^T$ , où

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i \sim N\left(\mu_X, \frac{\sigma^2}{m}\right), \quad \bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j \sim N\left(\mu_Y, \frac{\sigma^2}{n}\right), \quad (3)$$

$$\frac{(m-1)S_X^2}{\sigma^2} = \chi_{m-1}^2 \quad \text{et} \quad \frac{(n-1)S_Y^2}{\sigma^2} = \chi_{n-1}^2 \quad (4)$$

sont des variables aléatoires indépendantes. La statistique  $T$  n'est pas une statistique minimale exhaustive pour  $\theta = (\mu_X, \mu_Y, \sigma^2)^T$ . Dans ce problème la statistique minimale exhaustive est

$$U = (\bar{X}_m, \bar{Y}_n, S^2)^T, \quad (5)$$

où  $S^2$  est l'estimateur de  $\sigma^2$  appelé *l'estimateur unifié sans biais* :

$$S^2 = \frac{m-1}{n+m-2} S_X^2 + \frac{n-1}{n+m-2} S_Y^2. \quad (6)$$

De (4) et (6) il suit que

$$\frac{n+m-2}{\sigma^2} S^2 = \chi_{m+n-2}^2, \quad (7)$$

et donc

$$\mathbf{E}S^2 = \sigma^2 \quad \text{et} \quad \mathbf{Var} S^2 = \frac{2\sigma^4}{m+n-2}. \quad (8)$$

Il est clair que des composantes  $\bar{X}_m, \bar{Y}_n, S^2$  de la statistique minimale exhaustive  $U$  sont des variables aléatoires indépendantes. L'estimateur unifié  $S^2$  est une *moyenne pondérée* de  $S_X^2$  et  $S_Y^2$ . On peut voir que le poids plus grand sera donné à celui des estimateurs de  $\sigma^2$  qui correspond au échantillon de taille  $\max(m, n)$ . Si  $n = m$  alors  $S^2$  est une moyenne ordinaire de  $S_X^2$  et  $S_Y^2$ . Il faut remarquer que de (6), (7) et (8) il suit que

$$\mathbf{Var} S^2 = \frac{2\sigma^4}{n+m-2} < \begin{cases} \mathbf{Var} S_X^2 = \frac{2\sigma^4}{m-1}, \\ \mathbf{Var} S_Y^2 = \frac{2\sigma^4}{n-1}, \end{cases} \quad (9)$$

et on voit que l'estimateur unifié  $S^2$  est meilleur que  $S_X^2$  ou  $S_Y^2$ .

Puisque  $\bar{X}_m$  et  $\bar{Y}_n$  sont les meilleurs estimateurs sans biais pour  $\mu_X$  et  $\mu_Y$  respectivement, on en déduit immédiatement que  $\bar{X}_m - \bar{Y}_n$  est le meilleur estimateur sans biais pour  $\mu_X - \mu_Y$ , et de (3) il suit que

$$\bar{X}_m - \bar{Y}_n \sim N\left(\mu_X - \mu_Y, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right). \quad (10)$$

Par conséquent, la variable aléatoire

$$Z = \frac{\bar{X}_m - \bar{Y}_n - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad (11)$$

suit la loi normale standard. Comme la statistique  $S$ , donnée par (6) est indépendante de  $\bar{X}_m - \bar{Y}_n$ , et grâce à la relation (7), du Théorème de Fisher il résulte que la variable aléatoire

$$\frac{\bar{X}_m - \bar{Y}_n - (\mu_X - \mu_Y)}{S \sqrt{\frac{1}{m} + \frac{1}{n}}} = t_{n+m-2} \quad (12)$$

suit la loi de Student avec  $m + n - 2$  degrés de liberté, et donc

$$\mathbf{P} \left\{ |(\mu_X - \mu_Y) - (\bar{X}_m - \bar{Y}_n)| \leq \bar{t}_{m+n-2} \left( \frac{\alpha}{2} \right) S \sqrt{\frac{1}{m} + \frac{1}{n}} \right\} = 1 - \alpha, \quad (13)$$

c'est-à-dire

$$\begin{aligned} \bar{X}_m - \bar{Y}_n - \bar{t}_{m+n-2} \left( \frac{\alpha}{2} \right) S \sqrt{\frac{1}{m} + \frac{1}{n}} &\leq \mu_X - \mu_Y \leq \\ \bar{X}_m - \bar{Y}_n + \bar{t}_{m+n-2} \left( \frac{\alpha}{2} \right) S \sqrt{\frac{1}{m} + \frac{1}{n}} & \end{aligned} \quad (14)$$

est le plus court intervalle de confiance de niveau  $(1 - \alpha)$  pour la différence  $\mu_X - \mu_Y$  de deux moyennes des lois normales possédant la même variance inconnue.

**Remarque 1.** Supposons que  $\sigma_X^2$  et  $\sigma_Y^2$  sont inconnues, mais leur quotient  $\sigma_X^2/\sigma_Y^2$  est donné, par exemple,

$$\sigma_X^2/\sigma_Y^2 = k, \quad k > 0, \quad (15)$$

et il faut construire le plus court intervalle de confiance de niveau  $1 - \alpha$  pour la différence  $\mu_X - \mu_Y$ , où  $\mu_X$  et  $\mu_Y$  sont aussi inconnues. Le cas  $k = 1$  vient d'être considéré. So l'on note  $\sigma_Y^2 = \sigma^2$ , alors  $\sigma_X^2 = k\sigma^2$  et au lieu de (3) et (4) nous aurons

$$\bar{X}_m \sim M \left( \mu_X, \frac{k\sigma^2}{m} \right) \quad \text{et} \quad \bar{Y}_n \sim N \left( \mu_Y, \frac{\sigma^2}{n} \right), \quad (16)$$

$$\frac{(m-1)S_X^2}{k\sigma^2} = \chi_{m-1}^2 \quad \text{et} \quad \frac{(n-1)S_Y^2}{\sigma^2} = \chi_{n-1}^2, \quad (17)$$

et au lieu de (10) on a

$$\bar{X}_m - \bar{Y}_n \sim N \left( \mu_X - \mu_Y, \frac{k\sigma^2}{m} + \frac{\sigma^2}{n} \right), \quad (18)$$

d'où il vient que la variable aléatoire

$$Z = \frac{\bar{X}_m - \bar{Y}_n - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{k}{m} + \frac{1}{n}}} \quad (19)$$

suit la loi normale standard. D'autre côté, puisque

$$\frac{(m-1)S_X^2}{k\sigma^2} + \frac{(n-1)S_Y^2}{\sigma^2} = \chi_{m-1}^2 + \chi_{n-1}^2 = \chi_{m+n-2}^2, \quad (20)$$

de (17) il suit que l'estimateur unifié sans biais pour  $\sigma^2$  est

$$S^2 = \frac{1}{m+n-2} \left\{ \frac{m-1}{k} S_X^2 + (n-1) S_Y^2 \right\}. \quad (21)$$

Comme

$$\frac{m+n-2}{\sigma^2} S^2 = \chi_{m+n-2}^2, \quad (22)$$

et  $S^2$  est indépendante de  $Z$ , donnée par (19), du Théorème de Fisher on déduit que la variable aléatoire

$$\frac{\bar{X}_m - \bar{Y}_n - (\mu_X - \mu_Y)}{S \sqrt{\frac{k}{m} + \frac{1}{n}}} = t_{m+n-2} \quad (23)$$

suit la loi de Student avec  $m + n - 2$  degrés de liberté, et donc

$$\mathbf{P} \left\{ |(\mu_X - \mu_Y) - (\bar{X}_m - \bar{Y}_n)| \leq \bar{t}_{m+n-2} \left( \frac{\alpha}{2} \right) S \sqrt{\frac{k}{m} + \frac{1}{n}} \right\} = 1 - \alpha, \quad (24)$$

c'est-à-dire

$$\begin{aligned} \bar{X}_m - \bar{Y}_n - \bar{t}_{m+n-2} \left( \frac{\alpha}{2} \right) S \sqrt{\frac{k}{m} + \frac{1}{n}} \leq \mu_X - \mu_Y \leq \\ \bar{X}_m - \bar{Y}_n + \bar{t}_{m+n-2} \left( \frac{\alpha}{2} \right) S \sqrt{\frac{k}{m} + \frac{1}{n}} \end{aligned} \quad (25)$$

est le plus court intervalle de confiance de niveau  $(1 - \alpha)$  pour la différence  $\mu_X - \mu_Y$  de deux moyennes des lois normales possédant le quotient donné  $k = \sigma_X^2 / \sigma_Y^2$  des variances inconnues  $\sigma_X^2$  et  $\sigma_Y^2$ .

**Exemple 1.** Pour mesurer un angle  $A$  il étaient effectuées deux expériences indépendantes. Dans le premier étaient reçues deux valeurs

$$21^0.76 \quad \text{et} \quad 20^0.98, \quad (26)$$

et dans le second il'en avait 6

$$21^0.64, \quad 21^0.54, \quad 22^0.32, \quad 20^0.56, \quad 21^0.43, \quad 21^0.07. \quad (27)$$

Nous supposons que toutes les erreurs de mesures sont des réalisations des variables aléatoires normales indépendantes, et dans le deuxième expérience on utilise un instrument de mesure dont la précision est 4 fois meilleur que celui du premier expérience. Il faut construire le plus court intervalle de confiance de niveau  $(1 - \alpha)$  pour la différence  $b_X - b_Y$  des erreurs systématiques  $b_X$  et  $b_Y$  des instruments utilisés dans le premier et second expériences ( $\alpha = 0.01$ ).

*Solution.* Suivant la théorie des erreurs de Gauss nous pouvons supposer que les données (26) représentent la réalisation d'un échantillon normale

$$\mathbb{X} = (X_1, X_2)^T, \quad X_i \sim N(\mu_X, \sigma_X^2), \quad (28)$$

et les données (27) représentent la réalisation d'un échantillon normale

$$\mathbb{Y} = (Y_1, \dots, Y_6)^T, \quad Y_j \sim N(\mu_X, \sigma^2), \quad (29)$$

où  $\sigma_X^2 = 4\sigma^2$ , car  $k = 4$ . Dans ce cas la statistique exhaustive est

$$(\bar{X}_m, S_X^2, \bar{Y}_n, S_Y^2)^T,$$

avec

$$\bar{X}_m = \bar{X}_2 = 21.37, \quad \bar{Y}_n = \bar{Y}_6 = 21.42, \quad S_X^2 = 0.3042, \quad S_Y^2 = 0.3445. \quad (30)$$

De (21) il suit que l'estimateur unifié pour  $\sigma^2$  est

$$S^2 = \frac{1}{m+n-2} \left\{ \frac{m-1}{k} S_X^2 + (n-1) S_Y^2 \right\} = \frac{1}{6} \left\{ \frac{1}{4} S_X^2 + 5 S_Y^2 \right\}. \quad (31)$$

Puisque

$$\sqrt{\frac{k}{m} + \frac{1}{n}} = \sqrt{\frac{13}{6}} \quad \text{et} \quad \bar{t}_6(0.05) = 1.943, \quad (32)$$

et comme

$$\mu_X - \mu_Y = b_X - b_Y,$$

de (25) on a

$$\mathbf{P} \left\{ |(b_X - b_Y) - (\bar{X}_m - \bar{Y}_n)| \leq \bar{t}_6(0.05)S \sqrt{\frac{1}{m} + \frac{1}{n}} \right\} = 1 - \alpha, \quad (33)$$

et donc de (30)-(32) nous obtenons que la différence systématique  $b_X - b_Y$  appartient à l'intervalle

$$|(b_X - b_Y) - (-0.05)| \leq 1.57,$$

c'est-à-dire

$$-1^{0.62} \leq b_X - b_Y \leq 1^{0.52}.$$

## 2.15 Intervalle de confiance pour le quotient des variances de deux lois normales.

Soient  $\mathbb{X} = (X_1, \dots, X_m)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_n)^T$  deux échantillons indépendantes,

$$X_i \sim N(\mu_X, \sigma_X^2) \quad \text{et} \quad Y_j \sim N(\mu_Y, \sigma_Y^2).$$

D'après le théorème de Fisher nous avons

$$\frac{(m-1)S_X^2}{\sigma_X^2} = \chi_{m-1}^2 \quad \text{et} \quad \frac{(n-1)S_Y^2}{\sigma_Y^2} = \chi_{n-1}^2 \quad (1)$$

où

$$S_X^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2 \quad \text{et} \quad S_Y^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \quad (2)$$

sont les meilleurs estimateurs sans biais pour  $\sigma_X^2$  et  $\sigma_Y^2$ , et

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i \quad \text{et} \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \quad (3)$$

sont des meilleurs estimateurs sans biais pour  $\mu_X$  et  $\mu_Y$ . Puisque les échantillons  $\mathbb{X}$  et  $\mathbb{Y}$  sont indépendantes, les statistiques  $S_X^2$  et  $S_Y^2$  sont indépendantes aussi, et donc nous obtenons

**Théorème 1.** *La variable aléatoire*

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} = F_{m-1, n-1} \quad (4)$$

suit la loi  $F$  avec  $m-1$  et  $n-1$  degrés de liberté.

Nous allons utiliser ce théorème pour construire l'intervalle de confiance pour le quotient  $\sigma_Y^2/\sigma_X^2$ . En utilisant (4) et la table de  $F$ -répartition on peut trouver deux quantiles

$$\underline{F}_{m-1,n-1}(\alpha/2) = \frac{1}{\bar{F}_{n-1,m-1}(\alpha/2)} \quad \text{et} \quad \bar{F}_{m-1,n-1}(\alpha/2)$$

tels que

$$\mathbf{P}\{\underline{F}_{m-1,n-1}(\alpha/2) \leq F_{m-1,n-1} \leq \bar{F}_{m-1,n-1}(\alpha/2)\} = 1 - \alpha. \quad (5)$$

Dans ce cas de (4) et (5) nous avons

$$\mathbf{P}\left\{\underline{F}_{m-1,n-1}(\alpha/2) \leq \frac{\sigma_Y^2 S_X^2}{\sigma_X^2 S_Y^2} \leq \bar{F}_{m-1,n-1}(\alpha/2)\right\} = 1 - \alpha \quad (6)$$

ou

$$\mathbf{P}\left\{\underline{F}_{m-1,n-1}(\alpha/2) \frac{S_Y^2}{S_X^2} \leq \frac{\sigma_Y^2}{\sigma_X^2} \leq \frac{S_Y^2}{S_X^2} \bar{F}_{m-1,n-1}(\alpha/2)\right\} = 1 - \alpha. \quad (7)$$

Puisque

$$\bar{F}_{m-1,n-1}(\alpha/2) = \frac{1}{\underline{F}_{n-1,m-1}(\alpha/2)}, \quad (8)$$

nous obtenons l'intervalle de confiance de niveau  $(1 - \alpha)$  pour le quotient  $\sigma_Y^2/\sigma_X^2$  :

$$\mathbf{P}\left\{\frac{1}{\bar{F}_{n-1,m-1}(\alpha/2)} \frac{S_Y^2}{S_X^2} \leq \frac{\sigma_Y^2}{\sigma_X^2} \leq \frac{S_Y^2}{S_X^2} \bar{F}_{m-1,n-1}(\alpha/2)\right\} = 1 - \alpha. \quad (9)$$

De (9) il suit immédiatement que l'intervalle de confiance de niveau  $(1 - \alpha)$  pour le quotient  $\sigma_X^2/\sigma_Y^2$  est

$$\mathbf{P}\left\{\frac{1}{\bar{F}_{m-1,n-1}(\alpha/2)} \frac{S_X^2}{S_Y^2} \leq \frac{\sigma_X^2}{\sigma_Y^2} \leq \frac{S_X^2}{S_Y^2} \bar{F}_{n-1,m-1}(\alpha/2)\right\} = 1 - \alpha. \quad (10)$$

Par conséquent, (9) et (10) nous donnent deux intervalles de confiance de niveau  $(1 - \alpha)$  pour  $\sigma_Y^2/\sigma_X^2$  et  $\sigma_X^2/\sigma_Y^2$  respectivement.

**Exemple 1.** Soient  $\mathbb{X} = (X_1, \dots, X_m)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_n)^T$  deux échantillons indépendantes,

$$X_i \sim N(\mu_X, \sigma_X^2) \quad \text{et} \quad Y_j \sim N(\mu_Y, \sigma_Y^2)^T.$$

Nous supposons que un experiment pour  $m = 25$  et  $n = 14$  on a obtenu

$$S_X^2 = 74 \cdot 10^{-6} \quad \text{et} \quad S_Y^2 = 20 \cdot 10^{-6}.$$

En utilisant (10) nous construisons l'intervalle de confiance de niveau  $(1 - \alpha)$  pour le quotient des variances  $\sigma_X^2/\sigma_Y^2$ . Prenons  $\alpha = 0.1$ . Puisque  $S_X^2/S_Y^2 = 3.70$ ,

$$\bar{F}_{n-1,m-1}\left(\frac{\alpha}{2}\right) = \bar{F}_{13,24}(0.05) = 2.13$$

et

$$1/\bar{F}_{m-1,n-1}\left(\frac{\alpha}{2}\right) = \frac{1}{\bar{F}_{24,13}\left(\frac{\alpha}{2}\right)} = 1/2.35 = 0.426,$$

on a que

$$1.58 < \frac{\sigma_X^2}{\sigma_Y^2} < 7.88,$$

avec le coefficient de confiance 0.9.

## 2.16 La loi de Thompson.

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon normal,  $X_i \sim N(\mu, \sigma^2)$ . Notons

$$\eta_j = \frac{X_j - \bar{X}_n}{s_n} = \sqrt{\frac{n}{n-1}} Z_j, \quad j = 1, 2, \dots, n, \quad (2.1)$$

où  $Z_j$  est donné par (8.57),

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Dans ce cas pour tout  $j$  la statistique  $\eta_j$  suit la loi de Thompson à  $n - 2$  degrés de liberté,

$$\mathbf{P}\{\eta_j \leq x\} = T_{n-2}(x) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-2}{2}\right)} \int_{-\sqrt{n-1}}^x \left(1 - \frac{t^2}{n-1}\right)^{\frac{n-4}{2}} dt, \quad (2.2)$$

pour  $|x| < \sqrt{n-1}$ .

Soit  $\tau_m$  une variable aléatoire qui suit la loi de Thompson à  $m$  degrés de liberté,

$$\mathbf{P}\{\tau_m \leq x\} = T_m(x).$$

On sait que la statistique

$$t_m = \tau_m \sqrt{\frac{m}{m+1 - \tau_m^2}} \quad (2.3)$$

suit la distribution de Student à  $m$  degrés de liberté,

$$\mathbf{P}\{t_m \leq x\} = S_m(x).$$

On voit de (3) que

$$\tau_m = t_m \sqrt{\frac{m+1}{m+t_m^2}} \quad (2.4)$$

et par conséquent il en résulte que les quantiles  $\tau(\alpha, m)$  de la loi de Thompson à  $m$  degrés de liberté (de niveau  $\alpha$ ) s'expriment en fonction des quantiles correspondants  $t(\alpha, m)$  de la loi de Student à  $m$  degrés de liberté par la formule

$$\tau(\alpha, m) = t(\alpha, m) \sqrt{\frac{m+1}{m+t^2(\alpha, m)}}. \quad (2.5)$$

On sait que si  $n \rightarrow \infty$ , alors

$$S_m(x) \rightarrow \Phi(x) \quad (2.6)$$

et par conséquent de (3) à (5) on déduit une approximation normale pour le loi de Thompson, en utilisant la liaison qui existe entre les variables aléatoires  $\tau_m$  et  $\beta = \beta_{\frac{m}{2}, \frac{m}{2}}$  :

$$\beta = \frac{\tau_m + \sqrt{m+1}}{2\sqrt{m+1}}, \quad (2.7)$$

ce qui est équivalent à

$$\mathbf{P}\{\tau_m \leq x\} = \mathbf{P}\left\{\beta_{\frac{m}{2}, \frac{m}{2}} \leq \frac{x + \sqrt{m+1}}{2\sqrt{m+1}}\right\} = I_{\frac{x+\sqrt{m+1}}{2\sqrt{m+1}}}\left(\frac{m}{2}, \frac{m}{2}\right), \quad (2.8)$$

où  $\beta_{\alpha, \beta}$  est une variable aléatoire qui suit la loi bêta de paramètres  $\alpha$  et  $\beta$ .

**Remarque 1.** (Coefficient de corrélation d'un échantillon normale dans  $R^2$ ).

Soit

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$$

un échantillon d'une loi normale de dimension 2, i.e. pour tout  $(x, y) \in R^2$

$$\mathbf{P}\{X_i \leq x, Y_i \leq y\} = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\frac{x-\mu_x}{\sigma_x}} \int_{-\infty}^{\frac{y-\mu_y}{\sigma_y}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right\} dudv,$$

où

$$\begin{aligned} \mu_x &= \mathbf{E}X_i, & \mu_y &= \mathbf{E}Y_i, & \sigma_x^2 &= \mathbf{Var} X_i, & \sigma_y^2 &= \mathbf{Var} Y_i, \\ \rho &= \frac{1}{\sigma_x \sigma_y} \mathbf{E}(X_i - \mu_x)(Y_i - \mu_y). \end{aligned}$$

On peut montrer que les statistiques

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i, & \bar{Y}_n &= \frac{1}{n} \sum_{i=1}^n Y_i, & s_x^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \\ s_y^2 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2, & \hat{\rho}_n &= \frac{s_{xy}}{s_x s_y} \end{aligned}$$

sont les estimateurs de maximum de vraisemblance pour les paramètres  $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$  et  $\rho$  respectivement, où

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n).$$

On peut montrer sous l'hypothèse  $H_0 : \rho = 0$  la densité  $p_n(r)$ ,  $n \geq 3$ , de la statistique  $\hat{\rho}_n$  est donnée par la formule :

$$p_n(r) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} (1-r^2)^{\frac{n-4}{2}}, \quad |r| < 1, \quad (2.9)$$

d'où on tire que si l'hypothèse  $H_0$  est juste, alors

$$\hat{\rho}_n^2 = \beta_{\frac{1}{2}, \frac{n-2}{2}} \quad \text{et} \quad t_{n-2} = \hat{\rho}_n \sqrt{\frac{n-2}{1-\hat{\rho}_n^2}}. \quad (2.10)$$

## 2.17 Méthode du maximum de vraisemblance.

Supposons que'on a un échantillon

$$X \sim P_{\theta}, \theta = (\theta_1, \dots, \theta_m)^T \in \Theta \subset \mathbf{R}^m$$

et que  $P_{\theta}$  est absolument continue par rapport à une mesure  $\sigma$ -finie  $\mu$ . Notons par  $f(x; \theta)$  la densité de  $X$ .

Soit

$$L(\theta) = L(X, \theta) = f(X; \theta), \quad \theta \in \Theta \subset \mathbf{R}^m,$$

la *fonction de vraisemblance* de  $\mathbf{X}$ .

On appelle  $L(X, \theta)$  ainsi car, sachant une réalisation  $x$  du vecteur aléatoire  $X$ , la valeur  $L(x, \theta) = f(x, \theta)$  de  $L(X, \theta)$  nous permet de trouver les plus vraisemblables valeurs du paramètre  $\theta$ .

En effet, soit  $V(x)$  un voisinage infiniment petit de  $x$ . Alors

$$\mathbf{P}_{\theta}(X \in V(x)) \approx f(x, \theta) \mu(V(x)) \quad (1)$$

(dans le cas discret on a une égalité). Les valeurs de  $\theta$  plus vraisemblables sont telles qui maximisent la probabilité que  $X$  prend la valeur observée  $x$  (ou prend la valeur dans un infiniment petit voisinage de  $x$ , si telles probabilités sont égales à zero), donc d'après (1) maximisent la réalisation  $L(x, \theta) = f(x, \theta)$  de la fonction de vraisemblance  $L(X, \theta)$  par rapport à  $\theta$ .

**Définition 1.** Une statistique  $\hat{\theta}_n = \hat{\theta}_n(X)$  est appelée *estimateur de maximum de vraisemblance (EMV) du paramètre  $\theta$ , si  $\mu$ -p.s.*

$$L(X, \hat{\theta}_n) = \sup_{\theta \in \Theta} L(X, \theta). \quad (2)$$

Si  $\mathbf{g} : \Theta \rightarrow \mathbf{R}^k$  est une fonction mesurable,  $k \leq m$ , alors  $\hat{\mathbf{g}}_n = \mathbf{g}(\hat{\theta}_n)$  est appelé *estimateur de maximum de vraisemblance de  $\mathbf{g} = \mathbf{g}(\theta)$* .

**Rémarque 1.** Si  $T = T(X)$  est une statistique exhaustive, alors le critère de factorisation  $L(X, \theta) = g(T(X), \theta)h(X)$  implique que l'EMV est une fonction de  $T$ .

Généralement on cherche l'EMV en maximisant la fonction  $\ln L(X, \theta)$  par rapport à  $\theta$ , car cette fonction atteint le maximum dans le même point que  $L$  et dans la plupart des cas concrets est plus simple.

Si la fonction  $\ln L(X, \theta)$  est dérivable par rapport à  $\theta$ , alors l'EMV vérifie le système d'équations de vraisemblance

$$U(\theta) = \mathbf{0},$$

où

$$U(\theta) = \left( \frac{\partial \ln L(X, \theta)}{\partial \theta} \right)^T = \left( \frac{\partial \ln L(X, \theta)}{\partial \theta_1}, \dots, \frac{\partial \ln L(X, \theta)}{\partial \theta_m} \right)^T \quad (3)$$

est la fonction score.

La forme de la fonction de vraisemblance dépend de la structure de l'échantillon.

**Exemple 1.** Si  $X = (X_1, \dots, X_n)^T$  est un échantillon simple,  $X_i \sim p(x, \theta)$ ,  $\theta \in \Theta \subset \mathbf{R}^m$ , alors

$$L(X, \theta) = \prod_{i=1}^n p(X_i, \theta), \quad \ln L(X, \theta) = \sum_{i=1}^n \ln p(X_i, \theta),$$

and

$$U(\theta) = \left( \sum_{i=1}^n \frac{\partial \ln p(X_i, \theta)}{\partial \theta} \right)^T. \quad (4)$$

**Exemple 2. Censure du premier type.** On fixe le temps  $t$  de l'expérience et on observe  $n$  sujets. Les durées de vie  $T_1, \dots, T_n$  de sujets sont des v.a. i.i.d. de la fonction de répartition  $F(t, \theta)$ ,  $\theta \in \Theta \subset \mathbf{R}^m$  et de la densité  $p(t, \theta)$  par rapport à la mesure de Lebesgue. La valeur  $t_i$  de la variable aléatoire  $T_i$  n'est pas observée, si  $t_i > t$ . Les moments  $t_{(1)} \leq \dots \leq t_{(d(t))}$  de  $d(t)$  décès, ( $d(t) \leq n$ ), sont observés pendant l'expérience, si  $d(t) > 0$ . Si  $d(t) = 0$ ,  $t_{(i)}$  ne sont pas observés. Le vecteur

$$(t_{(1)}, \dots, t_{(d(t))}, d(t))^T$$

est une réalisation d'un vecteur aléatoire

$$(T_{(1)}, \dots, T_{(D(t))}, D(t))^T.$$

Cherchons la densité de ce vecteur :

$$\begin{aligned} & f_{T_{(1)}, \dots, T_{(D(t))}, D(t)}(t_1, \dots, t_d, d) \\ &= \lim_{h_1, \dots, h_d \downarrow 0} \frac{1}{h_1 \dots h_d} \mathbf{P}\{t_1 < T_{(1)} \leq t_1 + h_1, \dots, t_d < T_{(d)} \leq t_d + h_d, D(t) = d\} \\ &= \lim_{h_1, \dots, h_d \downarrow 0} \frac{1}{h_1 \dots h_d} \mathbf{P}\{D(t_1) = 0, D(t_1 + h_1) - D(t_1) = 1, \dots, D(t_d + h_d) - D(t_d) = 1, \\ & \quad D(t) - D(t_d) = 0, D(\infty) - D(t) = n - d\} = \\ & \quad \frac{n!}{(n-d)!} [1 - F(t, \theta)]^{n-d} p(t_1, \theta) \dots p(t_d, \theta), \end{aligned}$$

si  $t_1 < t_2 < \dots < t_d$ ,  $d = 1, 2, \dots$ .

Donc la fonction de vraisemblance est

$$L(\theta) = \frac{n!}{(n-D(t))!} [1 - F(t, \theta)]^{n-D(t)} p(T_{(1)}, \theta) \dots p(T_{(D(t))}, \theta), \quad (5)$$

si  $D(t) = 1, 2, \dots$ , et

$$L(\theta) = [1 - F(t, \theta)]^n, \quad (6)$$

si  $D(t) = 0$ .

La même fonction de vraisemblance (avec une constante près) peut être obtenu différemment. Posons

$$X_i = \min(T_i, t), \quad \delta_i = \mathbf{1}_{\{T_i \leq t\}}.$$

Sachant les paires

$$(X_1, \delta_1), \dots, (X_n, \delta_n),$$

on peut trouver  $T_{(1)}, \dots, T_{(D(t))}$  : il faut ordonner les  $X_i$ , qui correspondent à  $\delta_i = 1$ . Les vecteurs aléatoires  $(X_i, \delta_i)$  sont i.i.d., donc cherchons la loi de  $(X_1, \delta_1)$ . On a

$$F_{X_1, \delta_1}(x, 1; \theta) = \mathbf{P}_\theta(X_1 \leq x, \delta_1 = 1) = \mathbf{P}_\theta(T_1 \leq x, T_1 \leq t) =$$

$$F_{T_1}(\min(x, t)) = \int_0^x p(u, \theta) \mathbf{1}_{\{u \leq t\}} du,$$

$$F_{X_1, \delta_1}(x, 0; \theta) = \mathbf{P}_\theta(X_1 \leq x, \delta_1 = 0) = \mathbf{P}_\theta(t \leq x, T_1 > t) =$$

$$\mathbf{1}_{\{t \leq x\}} (1 - F(t, \theta)).$$

Considérons la mesure  $\mu$  sur  $\mathbf{R}_+ \times \{0, 1\}$  suivante :

$$\mu([0, x] \times \{1\}) = \int_0^x \mathbf{1}_{\{u \leq t\}} du, \quad \mu([0, x] \times \{0\}) = \mathbf{1}_{\{t \leq x\}}.$$

Alors

$$F_{X_1, \delta_1}(x, k; \theta) = \int_0^x p^k(u, \theta) [1 - F(t, \theta)]^{1-k} \mu(du, k),$$

et donc la densité de  $(X_i, \delta_i)$  par rapport à  $\mu$  est

$$p_{X_i, \delta_i}(x_i, k_i; \theta) = p^{k_i}(x_i, \theta) [1 - F(t, \theta)]^{1-k_i}.$$

Donc la fonction de vraisemblance est

$$L(X_1, \delta_1, \dots, X_n, \delta_n; \theta) = \prod_{i=1}^n p^{\delta_i}(X_i, \theta) [1 - F(X_i, \theta)]^{1-\delta_i}. \quad (7)$$

Notons que cette fonction est égale à la fonction donnée par (5) et (6) à la constante près :

$$L(X_1, \delta_1, \dots, X_n, \delta_n; \theta) = \prod_{i=1}^{D(t)} p(T_{(i)}, \theta) [1 - F(t, \theta)]^{n-D(t)}, \quad \text{si } D(t) > 0$$

ou

$$L(X_1, \delta_1, \dots, X_n, \delta_n; \theta) = [1 - F(t, \theta)]^n,$$

si  $D(t) = 0$ .

Des censures de plusieurs types sont considérée dans les chapitres suivants.

**Exemple 3. (Données groupés)** Soit  $Z_n = (Z_{n1}, \dots, Z_{nN})$  vecteur aléatoire qui suit la loi multinomiale  $M_N(n, p(\theta))$ , où  $p(\theta) = (p_1(\theta), \dots, p_N(\theta))^T$ ,  $\theta \in \Theta \subset \mathbf{R}^m$ .

Par exemple, si la région  $\mathcal{X}$  des valeurs des v.a. i.i.d.  $X_i \sim F(x, \theta)$ ,  $\theta \in \Theta \subset \mathbf{R}^m$  ( $i = 1, \dots, n$ ) est divisé en  $N$  intervalles  $I_1, \dots, I_N$ , alors  $Z_{nj}$  peut être interprété comme le nombre aléatoire des  $X_i$ , qui appartiennent à  $I_j$  :

$$Z_{nj} = \sum_{i=1}^n \mathbf{1}_{\{X_i \in I_j\}} \quad \text{et} \quad p_i(\theta) = \mathbf{P}_\theta(X_i \in I_j).$$

Donc

$$\mathbf{P}_\theta(Z_n = z_n) = \mathbf{P}\{Z_{n1} = k_1, \dots, Z_{nN} = k_N\} =$$

$$\frac{n!}{k_1! \dots k_N!} p_1^{k_1}(\theta) p_2^{k_2}(\theta) \dots p_N^{k_N}(\theta).$$

Supposons que n'observe que les v.a.  $Z_{nj}$ . Alors la fonction de vraisemblance est

$$L(Z_n, \theta) = \frac{n!}{Z_{n1}! \dots Z_{nN}!} p_1^{Z_{n1}}(\theta) p_2^{Z_{n2}}(\theta) \dots p_N^{Z_{nN}}(\theta).$$

### 19. Propriétés asymptotiques des estimateurs de maximum de vraisemblance

On va démontrer que sous conditions générales des estimateurs de maximum de vraisemblance sont consistants et asymptotiquement efficaces.

Soit

$$X = (X_1, \dots, X_n),$$

un échantillon, où  $X_1, \dots, X_n$  sont des vecteurs aléatoires indépendants,

$$X_i \sim p_i(x_i, \theta), \theta \in \Theta \subset \mathbf{R}^m,$$

où  $p_i(x_i, \theta)$  est la densité du vecteur  $r_i$ -dimensionnel  $X_i$  par rapport à une mesure  $\sigma$ -fini  $\mu$ .

La fonction de vraisemblance a la forme

$$L(X, \theta) = \prod_{i=1}^n p_i(X_i, \theta).$$

On a vu que sous des conditions générales la matrice d'information de Fisher a la forme

$$I_n(\theta) = \mathbf{E}_\theta \hat{I}_n(X, \theta), \quad \text{où} \quad \hat{I}_n(X, \theta) = -\frac{\partial^2}{\partial \theta^2} \ln L(X, \theta).$$

Si  $X_1, \dots, X_n$  sont des vecteurs aléatoires i.i.d. de la même dimension  $r$  (en cas  $r = 1$  on a un échantillon simple), alors  $p_i = p$ ,  $I_n(\theta) = nI_1(\theta)$ , où

$$I_1(\theta) = \mathbf{E}_\theta \hat{I}_1(X_1, \theta), \quad \hat{I}_1(X_1, \theta) = \frac{\partial^2}{\partial \theta^2} p(X_1, \theta).$$

**Théorème.** *Supposons que les vecteurs aléatoires  $X_1, \dots, X_n$  sont i.i.d. et*

- 1)  $\Theta$  est ouvert ;
- 2) presque pour tout  $y \in \mathbf{R}^r$  la densité  $p(y, \theta)$  est deux fois continument dérivable par rapport à  $\theta$  dans un voisinage  $V_\rho = \{\theta : \|\theta - \theta_0\| \leq \rho\}$  de la vraie valeur  $\theta_0$  du paramètre  $\theta$  ;
- 3) on peut dériver deux fois sous le signe de l'intégrale :

$$\int_{\mathbf{R}^r} \frac{\partial}{\partial \theta} p(y, \theta) dy = \frac{\partial}{\partial \theta} \int_{\mathbf{R}^r} p(y, \theta_0) dy = \mathbf{0},$$

$$\int_{\mathbf{R}^r} \frac{\partial^2}{\partial \theta^2} p(y, \theta_0) dy = \frac{\partial}{\partial \theta} \int_{\mathbf{R}^r} \frac{\partial}{\partial \theta} p(y, \theta_0) dy = \mathbf{0};$$

- 4) la matrice d'information de Fisher  $I_1(\theta_0)$  est définie positive ;
- 5) il existent des fonctions non-negatives  $h$  et  $b$ , telles que pour presque tous  $y \in \mathbf{R}^r$  et tous  $\theta \in V_\rho$

$$\|\hat{\mathbf{I}}_1(y, \theta) - \hat{\mathbf{I}}_1(y, \theta_0)\| \leq h(y) b(\theta), \quad \mathbf{E}_{\theta_0} \{h(X_1)\} < \infty, \quad b(\theta_0) = 0,$$

la fonction  $b$  est continue au point  $\theta_0$ .

Alors il existe une suite des estimateurs  $\{\hat{\theta}_n\}$  telle que

$$\mathbf{P}(U(X, \hat{\theta}_n) = \mathbf{0}) \rightarrow 1, \quad \hat{\theta}_n \xrightarrow{P} \theta_0, \quad (1)$$

et

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_m(\mathbf{0}, I_1^{-1}(\theta_0)). \quad (2)$$

*Démonstration.* Soit  $c > 0$  une constante et

$$B_c^n = \{\theta : (\theta - \theta_0)^T I_n(\theta_0)(\theta - \theta_0) \leq c^2\} = \{\theta : \|I_n^{1/2}(\theta_0)(\theta - \theta_0)\| \leq c\} \quad (3)$$

un voisinage de  $\theta_0$ . Notons par  $\partial V_\rho = \{\theta : \|\theta - \theta_0\| = \rho\}$  la frontière de  $V_\rho$ . La condition 4) implique que

$$\inf_{\theta: \theta \in \partial V_\rho} (\theta - \theta_0)^T I_1(\theta_0)(\theta - \theta_0) > 0,$$

donc il existe  $N = N(\rho) > 0$  tel que  $B_c^n \cap \partial V_\rho = \emptyset$ , quand  $n > N$  et donc  $B_c^n \subset V_\rho$ . Il est évident aussi que  $B_c^n \rightarrow \theta_0$ , i.e.  $\sup_{\theta \in B_c^n} \|\theta - \theta_0\| \rightarrow 0$  quand  $n \rightarrow \infty$ .

On va montrer que

$$\mathbf{P}_{\theta_0} \left( \sup_{\theta \in \partial B_c^n} \ln L(\theta) - \ln L(\theta_0) < 0 \right) \rightarrow 1, \quad \text{quand } n \rightarrow \infty. \quad (4)$$

Pour tout  $\theta \in \partial B_c^n$  écrivons la formule de Taylor :

$$\ln L(\theta) - \ln L(\theta_0) = U^T(\theta_0)(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)^T \hat{I}_n(\theta^*)(\theta - \theta_0), \quad (5)$$

où  $\theta^* = \theta^*(X)$  est un point sur la ligne entre  $\theta$  et  $\theta_0$ .

On va montrer d'abord que

$$\frac{1}{n} \hat{I}_n(\theta^*) = \frac{1}{n} I_n(\theta_0) + o_P(\mathbf{1}). \quad (6)$$

La condition 5) implique que

$$\mathbf{E}_{\theta_0} \left\| \frac{1}{n} (\hat{I}_n(\theta^*) - \hat{I}_n(\theta_0)) \right\| \leq \mathbf{E}_{\theta_0} \left\| \hat{I}_1(\theta^*) - \hat{I}_1(\theta_0) \right\| \leq \sup_{\theta \in B_c^n} b(\theta) \mathbf{E}_{\theta_0} h(X_1) \rightarrow 0.$$

Cette convergence implique que

$$\frac{1}{n} \hat{I}_n(\theta^*) - \frac{1}{n} \hat{I}_n(\theta_0) \xrightarrow{L_1} 0 \implies \frac{1}{n} \hat{I}_n(\theta^*) - \frac{1}{n} \hat{I}_n(\theta_0) \xrightarrow{P} 0. \quad (7)$$

La loi de grands nombres implique que

$$\frac{1}{n} \hat{I}_n(\theta_0) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln p(X_i, \theta_0) \xrightarrow{P} I_1(\theta_0), \quad (8)$$

car  $\hat{I}_n(\theta_0)$  est la somme de vecteurs aléatoires i.i.d. de l'espérance  $\mathbf{0}$  et de la variance  $I_1(\theta_0)$ .  
Donc on a

$$\frac{1}{n}\hat{I}_n(\theta^*) = \frac{1}{n}\hat{I}_n(\theta_0) + o_P(\mathbf{1}) = I_1(\theta_0) + o_P(\mathbf{1}) = \frac{1}{n}I_n(\theta_0) + o_P(\mathbf{1}). \quad (9)$$

Cette égalité, l'égalité (5) et la définition de  $\partial B_c^n$  (voir (3)) impliquent que uniformément sur  $\partial B_c^n$

$$\begin{aligned} \ln L(\theta) - \ln L(\theta_0) &= U^T(\theta_0)(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)^T I_n(\theta_0)(\theta - \theta_0) + o_P(\mathbf{1}) \\ &= U^T(\theta_0)(\theta - \theta_0) - \frac{c^2}{2} + o_P(\mathbf{1}). \end{aligned} \quad (10)$$

Donc

$$\begin{aligned} &\mathbf{P}_{\theta_0} \left( \sup_{\theta \in \partial B_c^n} \ln L(\theta) - \ln L(\theta_0) < 0 \right) \geq \\ &\mathbf{P}_{\theta_0} \left( \sup_{\theta \in \partial B_c^n} U^T(\theta_0)(\theta - \theta_0) + \sup_{\theta \in \partial B_c^n} |o_P(\mathbf{1})| < \frac{c^2}{2} \right) \\ &\geq \mathbf{P}_{\theta_0} \left( \sup_{\theta \in \partial B_c^n} U^T(\theta_0)(\theta - \theta_0) < \frac{c^2}{4}, |o_P(\mathbf{1})| < \frac{c^2}{4} \right) \geq \\ &1 - \mathbf{P}_{\theta_0} \left( \sup_{\theta \in \partial B_c^n} U^T(\theta_0)(\theta - \theta_0) \geq \frac{c^2}{4} \right) - \mathbf{P}_{\theta_0} \left( |o_P(\mathbf{1})| \geq \frac{c^2}{4} \right). \end{aligned} \quad (11)$$

Notons que  $\sup_{\mu \in \mathbf{R}^m, \|\mu\|=1} a^T \mu = \|a\|$  pour tout  $a \in \mathbf{R}^m$ , donc

$$\begin{aligned} \sup_{\theta \in \partial B_c^n} U^T(\theta_0)(\theta - \theta_0) &= c \sup_{\theta \in \partial B_c^n} U^T(\theta_0) I_n^{-1/2}(\theta_0) I_n^{1/2}(\theta_0)(\theta - \theta_0)/c \\ &\leq c \sup_{\mu \in \mathbf{R}^m, \|\mu\|=1} U^T(\theta_0) I_n^{-1/2}(\theta_0) \mu = c \|U^T(\theta_0) \mathbf{I}_n^{-1/2}(\theta_0)\|. \end{aligned} \quad (12)$$

L'inégalité de Tchebyshev-Bienaimé implique que

$$\begin{aligned} &\mathbf{P}_{\theta_0} \left( \|U^T(\theta_0) I_n^{-1/2}(\theta_0)\| \geq \right. \\ &\quad \left. c/4 \right) \leq (4/c)^2 \mathbf{E}_{\theta_0} (\|U^T(\theta_0) I_n^{-1/2}(\theta_0)\|^2) \\ &= (4/c)^2 \mathbf{E}_{\theta_0} U^T(\theta_0) I_n^{-1}(\theta_0) U(\theta_0) = (4/c)^2 m. \end{aligned} \quad (13)$$

Pour tout  $\delta > 0$  on peut trouver  $c > 0$  tel que  $(4/c)^2 \leq \delta/2$ . Fixons un tel  $c$ . Alors

$$\mathbf{P}_{\theta_0} \left( \sup_{\theta \in \partial B_c^n} U^T(\theta_0)(\theta - \theta_0) \geq \frac{c^2}{4} \right) < \delta/2. \quad (14)$$

On peut trouver  $N = N(\delta) > 0$  tel que pour tous  $n \geq N$

$$\mathbf{P}_{\theta_0} \left( |o_P(\mathbf{1})| \geq \frac{c^2}{4} \right) < \delta/2. \quad (15)$$

L'inégalité (11)-(15) impliquent la convergence (4).

La fonction  $\ln L(\theta)$  est continument dérivable sur  $V_\rho \supset B_c^n$ , donc cette convergence implique qu'il existe une suite d'estimateurs  $\{\hat{\theta}_n\}$  telle que

$$\mathbf{P}_{\theta_0} (U(\hat{\theta}_n) = \mathbf{0}, (\hat{\theta}_n - \theta_0)^T I_n(\theta_0) (\hat{\theta}_n - \theta_0) \leq c^2) \rightarrow 1,$$

donc pour tout  $\varepsilon > 0$  la relation

$$\mathbf{P}_{\theta_0} (U(\hat{\theta}_n) = \mathbf{0}, \|\hat{\theta}_n - \theta_0\| \leq \varepsilon) \rightarrow 1$$

implique la suivante

$$\mathbf{P}_{\theta_0} (U(\hat{\theta}_n) = \mathbf{0}) \rightarrow 1, \quad \hat{\theta}_n \xrightarrow{P} \theta_0.$$

Démontrons la normalité asymptotique des estimateurs  $\hat{\theta}_n$ . En intégrant la gauche et la droite de l'égalité

$$\begin{aligned} \frac{\partial}{\partial t} U\{\theta_0 + t(\hat{\theta}_n - \theta_0)\} &= \\ \frac{\partial}{\partial \theta} U(\{\theta_0 + t(\hat{\theta}_n - \theta_0)\})(\hat{\theta}_n - \theta_0) &= \\ -\hat{I}_n\{\theta_0 + t(\hat{\theta}_n - \theta_0)\}(\hat{\theta}_n - \theta_0) & \end{aligned}$$

par rapport à  $t$ , on obtient

$$-U(\theta_0) = U(\hat{\theta}_n) - U(\theta_0) = -\int_0^1 \hat{I}_n(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt (\hat{\theta}_n - \theta_0). \quad (15)$$

Montrons que le deuxième integrale est asymptotiquement equivalent à  $I_n(\theta_0)$ . La condition 5) implique

$$\begin{aligned} & \frac{1}{n} \left\| \int_0^1 \hat{I}_n(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt - \hat{I}_n(\theta_0) \right\| \\ & \leq \frac{1}{n} \sum_{i=1}^n \int_0^1 \left\| \frac{\partial^2}{\partial \theta^2} \ln p(X_i, \theta_0 + t(\hat{\theta}_n - \theta_0)) - \frac{\partial^2}{\partial \theta^2} \ln p(X_i, \theta_0) \right\| dt \\ & \leq \frac{1}{n} \sum_{i=1}^n h(X_i) \int_0^1 b(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt. \end{aligned} \quad (17)$$

Le premier facteur à la droite est la moyenne de v.a. i.i.d. de l'espérance fini, donc la loi de grands nombres implique que

$$\frac{1}{n} \sum_{i=1}^n h(X_i) \xrightarrow{P} \mathbf{E}_{\theta_0} h(X_1). \quad (18)$$

Montrons que le deuxième facteur tend en probabilité vers 0. La continuité de la fonction  $b$  en  $\theta_0$  et la condition  $b(\theta_0) = 0$  impliquent que pour tout  $\varepsilon > 0$  il existe  $\Delta = \Delta(\varepsilon)$  tel que  $b(\theta) < \varepsilon$ , si  $\|\theta - \theta_0\| < \Delta$ . Si  $\|\hat{\theta}_n - \theta_0\| < \Delta$ , alors pour tout  $t \in [0, 1]$

$$b(\theta_0 + t(\hat{\theta}_n - \theta_0)) < \varepsilon \quad \Rightarrow \quad \int_0^1 b(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt < \varepsilon.$$

Donc

$$\mathbf{P}_{\theta_0} \left( \int_0^1 b(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt \geq \varepsilon \right) \leq \mathbf{P}_{\theta_0} (\|\hat{\theta}_n - \theta_0\| \geq \Delta) \rightarrow 0. \quad (19)$$

Les convergences (18) et (19) et l'inégalité (16) impliquent

$$\frac{1}{n} \int_0^1 \hat{I}_n(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt = \frac{1}{n} \hat{I}_n(\theta_0) + o_P(\mathbf{1}) = \frac{1}{n} I_n(\theta_0) + o_P(\mathbf{1}). \quad (20)$$

L'égalités (16) et (20) impliquent

$$\begin{aligned} \frac{1}{\sqrt{n}} U(\theta_0) &= \left( \frac{1}{n} \hat{I}_n(\theta_0) + o_P(\mathbf{1}) \right) \sqrt{n}(\hat{\theta}_n - \theta_0) = \\ &= (I_1(\theta_0) + o_P(\mathbf{1})) \sqrt{n}(\hat{\theta}_n - \theta_0) \end{aligned} \quad (21)$$

La v.a.  $U(\theta_0)$  est une somme de vecteurs aléatoires i.i.d. de l'espérance  $\mathbf{0}$  et de la matrice de covariance  $I_1(\theta_0)$ . Le théorème limite centrale implique que

$$\frac{1}{\sqrt{n}} U(\theta_0) \xrightarrow{d} N_m(\mathbf{0}, I_1(\theta_0)). \quad (22)$$

Cette convergence, l'égalité (21) et le théorème de Slutsky impliquent que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_m(\mathbf{0}, I_1^{-1}(\theta_0) I_1(\theta_0) I_1^{-1}(\theta_0)) = N_m(\mathbf{0}, I_1^{-1}(\theta_0)).$$

**Corollaire.** *Sous les hypothèses du Théorème*

$$(\hat{\theta}_n - \theta_0)^T \hat{I}_n(\hat{\theta}_n) (\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi_m^2. \quad (23)$$

*Démonstration.* Le résultat du théorème implique que

$$(\hat{\theta}_n - \theta_0)^T I_1(\theta_0) (\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi_m^2. \quad (24)$$

La condition 5) du Théorème implique

$$\begin{aligned} \mathbf{E}_{\theta_0} \left\| \frac{1}{n} \hat{I}_n(X, \hat{\theta}_n) - \hat{I}_n(X, \theta_0) \right\| &\leq \\ \mathbf{E}_{\theta_0} \left\| \hat{\mathbf{I}}_1(X_1, \hat{\theta}_n) - \hat{\mathbf{I}}_1(X_1, \theta_0) \right\| &\leq \mathbf{E}_{\theta_0} h(X_1) b(\hat{\theta}_n) \rightarrow 0, \end{aligned}$$

donc

$$\frac{1}{n} \hat{I}_n(\hat{\theta}_n) = I_1(\theta_0) + o_P(\mathbf{1}). \quad (25)$$

(23) et (24) impliquent (22).

**Corollaire.** *Sous les hypothèses du Théorème*

$$U^T(\theta_0) I_n^{-1}(\theta_0) U(\theta_0) \xrightarrow{d} \chi_m^2$$

et

$$U^T(\theta_0) \hat{I}_n^{-1}(\hat{\theta}_n) U(\theta_0) \xrightarrow{d} \chi_m^2. \quad (26)$$

**Corollaire.** *Si la fonction  $\mathbf{g} : \Theta \rightarrow G \subset \mathbf{R}^k$  a des dérivés partielles du premier ordre continues, les hypothèses du Théorème sont vérifiées,  $\hat{\mathbf{g}}_n = \mathbf{g}(\hat{\theta}_n)$  est l'EMV de  $\mathbf{g} = \mathbf{g}(\theta)$ , alors*

$$\sqrt{n}(\hat{\mathbf{g}}_n - \mathbf{g}_0) \xrightarrow{d} N_k(\mathbf{0}, G(\theta_0) I_1^{-1}(\theta_0) G^T(\theta_0)),$$

où  $\mathbf{g}_0$  est la vraie valeur de  $\mathbf{g}$  et

$$G(\boldsymbol{\theta}_0) = \left[ \frac{\partial g_i(\boldsymbol{\theta}_0)}{\partial \theta_j} \right]_{k \times m}.$$

Ce résultat est impliqué par la méthode delta.

**Corollaire.** Sous les hypothèses du Corollaire

$$(\hat{\mathbf{g}}_n - \mathbf{g}_0)^T \left\{ G(\hat{\boldsymbol{\theta}}_n) \hat{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n) G^T(\hat{\boldsymbol{\theta}}_n) \right\}^{-1} (\hat{\mathbf{g}}_n - \mathbf{g}_0) \xrightarrow{d} \chi_k^2.$$

*Démonstration.* Corollaire implique

$$\sqrt{n}(\hat{\mathbf{g}}_n - \mathbf{g}_0)^T \left\{ G(\boldsymbol{\theta}_0) I_n^{-1}(\boldsymbol{\theta}_0) G^T(\boldsymbol{\theta}_0) \right\}^{-1} \sqrt{n}(\hat{\mathbf{g}}_n - \mathbf{g}_0) \xrightarrow{d} \chi_k^2. \quad (27)$$

La fonction  $G$  est continue, donc

$$G(\hat{\boldsymbol{\theta}}_n) = G(\boldsymbol{\theta}_0) + o_P(\mathbf{1}). \quad (28)$$

ce qui implique le résultat.

Le cas important est  $\mathbf{g} = (\theta_{l_1}, \dots, \theta_{l_k})$ , où  $1 \leq l_1 \leq \dots \leq l_k \leq m$ . Dans ce cas  $g_{ij}(\boldsymbol{\theta}) = 1$ , si  $j = l_i$ , et  $g_{ij}(\boldsymbol{\theta}) = 0$ , sinon. Donc

$$A_{i_1 \dots i_k} = G(\hat{\boldsymbol{\theta}}_n) \hat{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n) G^T(\hat{\boldsymbol{\theta}}_n)$$

est la sous-matrice de  $\hat{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n)$  étante sur intersection de  $i_1, \dots, i_k$ -èmes lignes et  $i_1, \dots, i_k$ -èmes colognes. Donc

$$(\hat{\theta}_{l_1} - \theta_{0l_1}, \dots, \hat{\theta}_{l_k} - \theta_{0l_k})^T A_{i_1 \dots i_k}^{-1} (\hat{\theta}_{l_1} - \theta_{0l_1}, \dots, \hat{\theta}_{l_k} - \theta_{0l_k}) \xrightarrow{d} \chi_k^2. \quad (29)$$

Généralisons le théorème pour le cas, quand les vecteurs  $X_i$  ne sont pas nécessairement identiquement distribués.

**Théorème.** Supposons que

- 1)  $\Theta$  est ouvert ;
- 2) presque pour tout  $x_i \in \mathbf{R}^{r_i}$  ( $r_i \leq r$ ) la densité  $p_i(x_i, \boldsymbol{\theta})$  est deux fois continument dérivable par rapport à  $\boldsymbol{\theta}$  dans un voisinage  $V_\rho = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \rho\}$  ;
- 3) on peut dériver deux fois par rapport à  $\boldsymbol{\theta}$  sous le signe des intégrales :

$$\int_{\mathbf{R}^{r_i}} \frac{\partial}{\partial \boldsymbol{\theta}} p(x_i, \boldsymbol{\theta}) dx_i = \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbf{R}^{r_i}} p(x_i, \boldsymbol{\theta}_0) dx_i = \mathbf{0},$$

$$\int_{\mathbf{R}^{r_i}} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} p(x_i, \boldsymbol{\theta}_0) dx_i = \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbf{R}^{r_i}} \frac{\partial}{\partial \boldsymbol{\theta}} p(x_i, \boldsymbol{\theta}_0) dx_i = \mathbf{0};$$

- 4) la matrice  $\lim_{n \rightarrow \infty} \frac{1}{n} I_n(\boldsymbol{\theta}_0) = I_n(\boldsymbol{\theta}_0)$  est définie positive.

- 5) existent des fonctions non-negatives  $h_i$  et  $b$ , telles que pour presque tous  $x_i \in \mathbf{R}^{r_i}$  et tous  $\boldsymbol{\theta} \in V_\rho$

$$\left\| \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln p_i(x_i, \boldsymbol{\theta}) - \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln p_i(x_i, \boldsymbol{\theta}_0) \right\| \leq h_i(x_i) b(\boldsymbol{\theta}),$$

$$\mathbf{E}_{\theta_0} \{ \sup_i h(X_i) \} < \infty, \quad b(\theta_0) = 0,$$

la fonction  $b$  est continue en  $\theta_0$ .

6) il existe un nombre positif  $\delta > 0$ , tel que

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\delta}} \sum_{i=1}^n \mathbf{E}_{\theta_0} \left\| \frac{\partial^2}{\partial \theta^2} \ln p_i(X_i, \theta_0) \right\|^{1+\delta} = 0.$$

Alors il existe une suite des estimateurs  $\{\hat{\theta}_n\}$  telle que

$$\mathbf{P}(U(X, \hat{\theta}_n) = \mathbf{0}) \rightarrow 1, \quad \hat{\theta}_n \xrightarrow{P} \theta_0. \quad (30)$$

Supposons, de plus, que

$$7) \quad \mathbf{E}_{\theta_0} \sup_i \left\| \frac{\partial^2}{\partial \theta^2} \ln p_i(X_i, \theta_0) \right\|^{2+\delta} < \infty.$$

Alors

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_m(\mathbf{0}, I^{-1}(\theta_0)). \quad (31)$$

*Démonstration.* Soit  $B_c^n$  un voisinage de  $\theta_0$  défini par (3). De même que dans le théorème précédant la condition 4) implique que  $B_c^n \rightarrow \theta_0$  et que  $B_c^n \subset V_p$ , si  $n$  est grand.

Pour tout  $\theta \in \partial B_c^n$  écrivons le development (5). La condition 5) implique

$$\begin{aligned} & \mathbf{E}_{\theta_0} \left\| \frac{1}{n} (\hat{I}_n(\theta^*) - \hat{I}_n(\theta_0)) \right\| \leq \\ & \mathbf{E}_{\theta_0} \left\| \frac{\partial^2}{\partial \theta^2} \ln p_i(X_i, \theta^*) - \frac{\partial^2}{\partial \theta^2} \ln p_i(X_i, \theta_0) \right\| \leq \\ & \mathbf{E}_{\theta_0} \sup_i h_i(X_i) \sup_{\theta \in B_c^n} b(\theta) \rightarrow 0, \end{aligned}$$

donc la convergence (7) a lieu.

La condition 6) et la loi de grands nombres impliquent

$$\begin{aligned} & \frac{1}{n} (\hat{I}_n(X, \theta_0) - I_n(\theta_0)) = \\ & -\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial^2}{\partial \theta^2} \ln p_i(X_i, \theta_0) - \mathbf{E}_{\theta_0} \left( \frac{\partial^2}{\partial \theta^2} \ln p_i(X_i, \theta_0) \right) \right\} \xrightarrow{P} 0. \end{aligned}$$

Cette convergence et la convergence (7) impliquent

$$\frac{1}{n} \hat{I}_n(\theta^*) = \frac{1}{n} I_n(\theta_0) + o_P(\mathbf{1}).$$

Le reste de démonstration de la consistance est le même comme dans Théorème.

Démontrons la normalité asymptotique. On écrit l'égalité (16). La condition 5) implique que

$$\frac{1}{n} \left\| \int_0^1 \hat{I}_n(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt - \hat{I}_n(\theta_0) \right\| \leq$$

$$\sup_i h_i(X_i) \int_0^1 b(\theta_0 + t(\hat{\theta}_n - \theta_0)) dt \xrightarrow{P} 0.$$

Donc

$$\frac{1}{\sqrt{n}} U(\theta_0) = \left( \frac{1}{n} I_n(\theta_0) + o_p(\mathbf{1}) \right) \sqrt{n}(\hat{\theta}_n - \theta_0).$$

Notons

$$Y_i = \frac{\partial}{\partial \theta} \ln p_i(X_i, \theta_0).$$

Soit  $a \in \mathbf{R}^m \setminus \mathbf{0}$ . Alors

$$a^T U(\theta_0) = \sum_{i=1}^n a^T Y_i, \quad \mathbf{E}(a^T Y_i) = 0,$$

$$\mathbf{Var}_{\theta_0}(a^T U(\theta_0)) = a^T I_n(\theta_0) a.$$

Alors

$$\frac{a^T U(\theta_0)}{a^T I_n(\theta_0) a} \xrightarrow{d} N(0, 1),$$

si la condition de Liapunov

$$\frac{\sum_{i=1}^n \mathbf{E} |a^T Y_i|^{2+\delta}}{(a^T I_n(\theta_0) a)^{1+\delta/2}} \rightarrow 0$$

est vérifiée. Mais l'inégalité

$$\mathbf{E} |a^T Y_i|^{2+\delta} \leq \|a\|^{2+\delta} \mathbf{E} \sup_i \|Y_i\|^{2+\delta}$$

implique que

$$\frac{\sum_{i=1}^n \mathbf{E} |a^T Y_i|^{2+\delta}}{(a^T I_n(\theta_0) a)^{1+\delta/2}} \leq n^{-\delta} \frac{\|a\|^{2+\delta}}{(a^T \frac{1}{n} I_n(\theta_0) a)^{1+\delta/2}} \mathbf{E} \sup_i \|Y_i\|^{2+\delta} \rightarrow 0,$$

car l'espérance à la droite est finie d'après la condition 7), la matrice  $I(\theta_0)$  est définie positive et donc

$$a^T \frac{1}{n} I_n(\theta_0) a \rightarrow a^T I(\theta_0) a > 0,$$

d'où on tire que pour tout  $a \in \mathbf{R}^m \setminus \mathbf{0}$

$$\frac{1}{\sqrt{n}} a^T U(\theta_0) \xrightarrow{d} N_m(0, a^T I(\theta_0) a)$$

et donc

$$\frac{1}{\sqrt{n}} I^{-1}(\theta_0) U(\theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)),$$

$$\frac{1}{\sqrt{n}} \left( \frac{1}{n} I_n(\theta_0) \right)^{-1} U(\theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)),$$

d'où on tire que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) =$$

$$\frac{1}{\sqrt{n}} \left( \frac{1}{n} I_n(\theta_0) + o_P(\mathbf{1}) \right)^{-1} U(\theta_0) \xrightarrow{d} N_m(0, I^{-1}(\theta_0)).$$

## 2.18 Propriétés asymptotiques du rapport de vraisemblance

**Théorème.** *Sous les conditions du théorème on a*

$$-2 \ln \frac{L(X, \theta_0)}{L(X, \hat{\theta}_n)} \xrightarrow{d} \chi^2(m).$$

*Démonstration.* D'après la formule de Taylor

$$\begin{aligned} \ln L(X, \theta_0) - \ln L(X, \hat{\theta}_n) &= U^T(X, \hat{\theta}_n)(\theta_0 - \hat{\theta}_n) - \\ &\quad \frac{1}{2}(\hat{\theta}_n - \theta_0)^T \hat{I}_n(X, \theta^*(X))(\hat{\theta}_n - \theta_0) = \\ &\quad -\frac{1}{2} \sqrt{n}(\hat{\theta}_n - \theta_0)^T \frac{1}{n} \hat{I}_n(X, \theta^*(X)) \sqrt{n}(\hat{\theta}_n - \theta_0) \end{aligned}$$

où  $\theta^*(X)$  est un point sur la ligne entre  $\hat{\theta}_n$  et  $\theta_0$  et

$$\|\theta^*(X) - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\| \xrightarrow{P} 0,$$

donc  $\theta^*(X) \xrightarrow{P} \theta_0$ .

Comme dans la démonstration du théorème (voir) , on a

$$\frac{1}{n} \hat{I}_n(\theta^*) - \frac{1}{n} \hat{I}_n(\theta_0) \xrightarrow{P} 0. \quad (1)$$

Donc

$$\frac{1}{n} \hat{I}_n(\theta^*) = \frac{1}{n} \hat{I}_n(\theta_0) + o_P(1) = I_1(\theta_0) + o_P(1).$$

et

$$\begin{aligned} -2(\ln L(X, \theta_0) - \ln L(X, \hat{\theta}_n)) &= \\ \sqrt{n}(\hat{\theta}_n - \theta_0)^T I_1(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) &+ o_P(1). \end{aligned}$$

La convergence

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z \sim N_m(\mathbf{0}, I_1^{-1}(\theta_0))$$

implique que

$$-2(\ln L(X, \theta_0) - \ln L(X, \hat{\theta}_n)) \xrightarrow{d} Z^T I_1(\theta_0) Z \sim \chi^2(m).$$

Soit

$$\phi = (\phi_1, \phi_2) : \Theta \rightarrow \mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2 \subset \mathbf{R}^k \times \mathbf{R}^{m-k}$$

une bijection continument dérivable. Notons par  $\psi : \mathcal{G}_1 \times \mathcal{G}_2 \rightarrow \Theta$  la fonction inverse.

Soient  $\mathbf{g}_{10}$  un point dans  $\mathcal{G}_1$  et  $\Theta_0$  un sous-ensemble de  $\Theta$ , défini par

$$\Theta_0 = \{\theta : \phi_1(\theta) = \mathbf{g}_{10}\} = \{\theta : \theta = \psi(\mathbf{g}_{10}, \mathbf{g}_2), \mathbf{g}_2 \in \mathcal{G}_2\} \subset \Theta. \quad (2)$$

**Exemple 1.** Soit

$$\phi_1(\theta) = \theta_1 = (\theta_1, \dots, \theta_k), \quad \phi_2(\theta) = \theta_2 = (\theta_{k+1}, \dots, \theta_m)$$

des projection de

$$\theta = (\theta_1, \dots, \theta_m), \quad \theta_1 \in \Theta_1, \quad \theta_2 \in \Theta_2, \quad \Theta = \Theta_1 \times \Theta_2.$$

Alors  $\phi(\theta) = \theta$ ,  $\psi(\theta) = \theta$  et

$$\Theta_0 = \{\theta : \theta_1 = \theta_{10}\} = \{\theta : \theta = (\theta_{10}, \theta_2), \theta_2 \in \Theta_2\},$$

où  $g_{10} = \theta_{10}$  est une valeur de  $\theta_1$  fixée.

**Exemple 2.** Soit  $k = 1$ ,

$$\phi_1(\theta) = \ln \theta_1, \quad \phi_2(\theta) = \theta_2 = (\theta_2, \dots, \theta_m).$$

Alors

$$\phi(\theta) = (\ln \theta_1, \theta_2), \quad \psi(g_1, \theta_2) = (e^{g_1}, \theta_2)$$

et

$$\Theta_0 = \{\theta : \ln \theta_1 = g_{10}\} = \{\theta : \theta = (e^{g_{10}}, \theta_2), \theta_2 \in \Theta_2\}.$$

**Exemple 3.** Soit  $k = 1$ ,

$$\phi_1(\theta) = \theta_1 - \theta_2, \quad \phi_2(\theta) = \theta_2.$$

Alors

$$\phi(\theta) = (\theta_1 - \theta_2, \theta_2), \quad \psi(g_1, \theta_2) = (g_1 + \theta_2, \theta_2)$$

et

$$\Theta_0 = \{\theta : \theta_1 - \theta_2 = g_{10}\} = \{\theta : \theta = (g_{10} + \theta_2, \theta_2), \theta_2 \in \Theta_2\}.$$

**Exemple 4.** Soit  $k = 1$ ,

$$\phi_1(\theta) = \theta_1/\theta_2, \quad \phi_2(\theta) = \theta_2.$$

Alors

$$\phi(\theta) = (\theta_1/\theta_2, \theta_2), \quad \psi(g_1, \theta_2) = (g_1 \theta_2, \theta_2)$$

et

$$\Theta_0 = \{\theta : \theta_1/\theta_2 = g_{10}\} = \{\theta : \theta = (g_{10} \theta_2, \theta_2), \theta_2 \in \Theta_2\}.$$

**Théorème** Supposons que les conditions du Théorème précédent sont vérifiées et  $\Theta_0$  est l'ensemble défini par (2). Si  $\theta_0 \in \Theta_0$  alors

$$\begin{aligned} R(X, g_{10}) &= -2 \ln \frac{\sup_{\theta \in \Theta_0} L(X, \theta)}{\sup_{\theta \in \Theta} L(X, \theta)} \\ &= -2 \ln \frac{\sup_{\theta: \phi_1(\theta) = g_{10}} L(X, \theta)}{L(X, \hat{\theta}_n)} \xrightarrow{d} \chi^2(k), \end{aligned}$$

i.e. pour tout  $x \in \mathbf{R}$

$$\mathbf{P}_{\theta_0}(R(X, g_{10}) \leq x) \rightarrow F_{\chi_k^2}(x).$$

*Démonstration.* On a

$$\begin{aligned} \sup_{\theta \in \Theta_0} L(\theta) &= \sup_{\theta: \theta = \psi(\mathbf{g}_{10}, \mathbf{g}_2), \mathbf{g}_2 \in \mathcal{G}_2} L(\theta) \\ &= \sup_{\mathbf{g}_2: \mathbf{g}_2 \in \mathcal{G}_2} L(\psi(\mathbf{g}_{10}, \mathbf{g}_2)) = \sup_{\mathbf{g}_2: \theta_2 \in \mathcal{G}_2} L^*(\mathbf{g}_2), \end{aligned}$$

où  $L^*(\mathbf{g}_2) = L(\psi(\mathbf{g}_{10}, \mathbf{g}_2))$ . La v.a.  $L^*(X, \mathbf{g}_2)$  est la fonction de vraisemblance pour le modèle statistique

$$X \sim f^*(x, \mathbf{g}_2), \mathbf{g}_2 \in \mathcal{G}_2,$$

où  $f^*(x, \mathbf{g}_2) = f(x, \psi(\mathbf{g}_{10}, \mathbf{g}_2))$ .

La consistance de  $\hat{\theta}_n$  implique que

$$I_1(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} U(\theta_0) + o_P(1), \quad (3)$$

donc

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = I_1^{-1}(\theta_0) \frac{1}{\sqrt{n}} U(\theta_0) + o_P(1). \quad (4)$$

Ce résultat implique que

$$\begin{aligned} 2(\ln L(X, \hat{\theta}_n) - \ln L(X, \theta_0)) &= \sqrt{n}(\hat{\theta}_n - \theta_0)^T I_1(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1) \\ &= \frac{1}{\sqrt{n}} U^T(\theta_0) I_1^{-1}(\theta_0) I_1(\theta_0) I_1^{-1}(\theta_0) \frac{1}{\sqrt{n}} U(\theta_0) + o_P(1) \\ &= \frac{1}{\sqrt{n}} U^T(\theta_0) I_1^{-1}(\theta_0) \frac{1}{\sqrt{n}} U(\theta_0) + o_P(1). \end{aligned} \quad (5)$$

De même, notant

$$\tilde{\mathbf{g}}_{2n} = \tilde{\mathbf{g}}_{2n}(\mathbf{g}_{10}, X)$$

l'EMV de  $\mathbf{g}_2$  sous notre modèle, on a

$$\begin{aligned} &2(\ln L^*(X, \tilde{\mathbf{g}}_{2n}) - \ln L^*(X, \mathbf{g}_{20})) \\ &= \frac{1}{\sqrt{n}} (U^*)^T(\mathbf{g}_{20}) (I_1^*)^{-1}(\mathbf{g}_{20}) \frac{1}{\sqrt{n}} U^*(\mathbf{g}_{20}) + o_P(1). \end{aligned} \quad (6)$$

La fonction score est

$$\begin{aligned} U^*(\mathbf{g}_2) &= \frac{\partial \ln L^*(\mathbf{g}_2)}{\partial \mathbf{g}_2} = \\ &= \frac{\partial \ln L(\psi(\mathbf{g}_{10}, \mathbf{g}_2))}{\partial \mathbf{g}_2} = \frac{\partial \psi(\mathbf{g}_{10}, \mathbf{g}_2)}{\partial \mathbf{g}_2} U(\psi(\mathbf{g}_{10}, \mathbf{g}_2)) = \\ &= A(\mathbf{g}_0) U(\psi(\mathbf{g}_{10}, \mathbf{g}_2)), \end{aligned} \quad (7)$$

où

$$A(\mathbf{g}_{10}, \mathbf{g}_2) = \frac{\partial \psi(\mathbf{g}_{10}, \mathbf{g}_2)}{\partial \mathbf{g}_2}.$$

En particulier,

$$U^*(\mathbf{g}_{20}) = A(\mathbf{g}_0) U(\theta_0), \quad (8)$$

La matrice d'information de Fisher en  $\mathbf{g}_{20}$  est

$$\begin{aligned} I_1^*(\mathbf{g}_{20}) &= \mathbf{E}_{\theta_0} U^*(\mathbf{g}_{20})(U^*)^T(\mathbf{g}_{20}) = \\ & A(\mathbf{g}_0)\mathbf{E}_{\theta_0} U(\theta_0)U^T(\theta_0)A(\mathbf{g}_0)^T = A(\mathbf{g}_0)I_1(\theta_0)A(\mathbf{g}_0)^T. \end{aligned} \quad (9)$$

Les égalités (7) et (9) impliquent

$$\begin{aligned} & 2(\ln L(X, \hat{\theta}_n) - \ln L^*(X, \tilde{\mathbf{g}}_{2n})) = \\ & \frac{1}{\sqrt{n}}U^T(\theta_0)\{I_1^{-1}(\theta_0) - A^T(\mathbf{g}_0)(I_1^*)^{-1}(\mathbf{g}_{20})A(\mathbf{g}_0)\}\frac{1}{\sqrt{n}}U(\theta_0). \end{aligned} \quad (10)$$

La convergence

$$\frac{1}{\sqrt{n}}U(\theta_0) \xrightarrow{d} Z \sim N(\mathbf{0}, I_1(\theta_0))$$

implique que

$$2(\ln L(X, \hat{\theta}_n) - \ln L^*(X, \tilde{\mathbf{g}}_{2n})) \xrightarrow{d} Z^T \{I_1^{-1} - A^T(I_1^*)^{-1}A\}Z. \quad (11)$$

La v.a. limite est une forme quadratique des v.a. normales. On va utiliser le résultat (voir) qui dit que si

$$Y \sim N(\mathbf{0}, \Sigma) \text{ et } B\Sigma B = B, \text{ tr}(B\Sigma) = k,$$

alors  $Y^T B Y \sim \chi_k^2$ . Dans notre cas

$$\begin{aligned} & (I_1^{-1} - A^T(I_1^*)^{-1}A)I_1(I_1^{-1} - A^T(I_1^*)^{-1}A) = \\ & I_1^{-1} - A^T(I_1^*)^{-1}A - A^T(I_1^*)^{-1}A \\ & + A^T(I_1^*)^{-1}A I_1 A^T(I_1^*)^{-1}A = I_1^{-1} - A^T(I_1^*)^{-1}A, \end{aligned} \quad (12)$$

car  $A I_1 A^T = I_1^*$ . Le rang

$$\begin{aligned} & \text{tr}((I_1^{-1} - A^T(I_1^*)^{-1}A)I_1) = \\ & \text{tr}(E_m - A^T(I_1^*)^{-1}A I_1) = m - \text{tr}((I_1^*)^{-1}A I_1 A^T) = m - \text{tr}(E_{m-k}) = k. \end{aligned} \quad (13)$$

D'où le résultat du théorème.

**Corollaire.** Sous les hypothèses du théorème

$$U^T(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n}))\hat{I}_n^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n}))U(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) \xrightarrow{d} \chi_k^2. \quad (14)$$

*Démonstration.* Notons que

$$\begin{aligned} \frac{1}{\sqrt{n}}U(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) &= \frac{1}{\sqrt{n}}U(\Psi(\mathbf{g}_{10}, \mathbf{g}_{20})) + o_P(\mathbf{1}) = \\ & \frac{1}{\sqrt{n}}U(\theta_0) + o_P(\mathbf{1}), \end{aligned} \quad (15)$$

$$\begin{aligned} n\hat{I}_n^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) &= nI_n^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) + o_P(\mathbf{1}) \\ &= I_1^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) + o_P(\mathbf{1}) = I_1^{-1}(\theta_0) + o_P(\mathbf{1}). \end{aligned} \quad (16)$$

L'égalité  $U^*(\tilde{\mathbf{g}}_{2n}) = \mathbf{0}$ , les égalités (15) et (16) impliquent

$$\begin{aligned}
& U^T(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) \hat{I}_n^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) U(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) \\
&= \frac{1}{\sqrt{n}} U^T(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) I_1^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) \frac{1}{\sqrt{n}} U(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) - \\
& \quad \frac{1}{\sqrt{n}} U^{*T}(\tilde{\mathbf{g}}_{2n})(I_1^*)^{-1}(\tilde{\mathbf{g}}_{2n}) \frac{1}{\sqrt{n}} U^*(\tilde{\mathbf{g}}_{2n}) + o_P(\mathbf{1}) = \\
& \quad \frac{1}{\sqrt{n}} U^T(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) \{I_1^{-1}(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) - \\
& \quad A^T(\tilde{\mathbf{g}}_{2n})(I_1^*)^{-1}(\tilde{\mathbf{g}}_{2n}) A(\tilde{\mathbf{g}}_{2n})\} \frac{1}{\sqrt{n}} U(\Psi(\mathbf{g}_{10}, \tilde{\mathbf{g}}_{2n})) + o_P(\mathbf{1}) = \\
& \quad \frac{1}{\sqrt{n}} U^T(\theta_0) \{I_1^{-1}(\theta_0) - A^T(\mathbf{g}_0)(I_1^*)^{-1}(\mathbf{g}_{20}) A(\mathbf{g}_0)\} \frac{1}{\sqrt{n}} U(\theta_0) + o_P(\mathbf{1}) \xrightarrow{d} \chi_k^2.
\end{aligned}$$

Le cas particulier important est, quand

$$\mathbf{g}_1(\theta) = \theta^{(1)} = (\theta_{l_1}, \dots, \theta_{l_k}) \text{ et } \mathbf{g}_2(\theta) = \theta^{(2)} = (\theta_{s_1}, \dots, \theta_{s_{m-k}})$$

où  $(l_1, \dots, l_k, s_1, \dots, s_{m-k})$  est une permutation de  $(1, \dots, m)$ ,

$$1 \leq l_1 \leq \dots \leq l_k \leq m, \quad 1 \leq s_1 \leq \dots \leq s_{m-k} \leq m.$$

. Dans ce cas

$$A = \left[ \frac{\partial \theta}{\partial \theta^{(2)}} \right] = [a_{ij}]_{(m-k) \times m},$$

où

$$a_{ij} = \begin{cases} 1, & \text{si } (i, j) = (l, s_l) \ (l = 1, \dots, m-k), \\ 0, & \text{sinon.} \end{cases}$$

Notons que les  $s_1, \dots, s_{m-k}$  composantes de  $U(\theta_0^{(1)}, \tilde{\theta}_n^{(2)})$  sont égales à zero, car

$$\mathbf{0} = U^*(\tilde{\theta}_n^{(2)}) = AU(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}) = (U_{s_1}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}), \dots, U_{s_{m-k}}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}))^T.$$

Posons

$$U_{l_1, \dots, l_k}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}) = (U_{l_1}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}), \dots, U_{l_k}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}))^T$$

et  $A_{i_1 \dots i_k}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)})$  la sous-matrice de

$$\hat{I}_n^{-1}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)})$$

étante sur intersection de  $l_1, \dots, l_k$ -èmes lignes et  $l_1, \dots, l_k$ -èmes colognes. Donc

$$U_{l_1, \dots, l_k}^T(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}) A_{i_1 \dots i_k}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}) U_{l_1, \dots, l_k}(\theta_0^{(1)}, \tilde{\theta}_n^{(2)}) \xrightarrow{d} \chi_k^2.$$

## 21. Exemples et remarques

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon lognormale  $LN(\mu, \sigma^2)$ ,

$$X_i \sim p(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \mathbf{1}_{]0, \infty[}(x), \quad \mu \in \mathbf{R}^1, \quad \sigma^2 > 0.$$

Remarquons que  $\ln X_i$  suit une loi normale  $N(\mu, \sigma^2)$ . On peut montrer que

$$a_1 = \mathbf{E}X_1 = e^{\mu + \sigma^2/2}, \quad a_2 = \mathbf{E}X_1^2 = e^{2\mu + 2\sigma^2}.$$

D'après la méthode des moments pour estimer  $\mu$  et  $\sigma^2$  il faut résoudre le système

$$\begin{cases} e^{\mu + \sigma^2/2} = \bar{X}_n & = \alpha_1, \\ e^{2\mu + 2\sigma^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 & = \alpha_2, \end{cases}$$

ce qui est équivalent à

$$\begin{cases} \mu + \sigma^2/2 & = \ln \alpha_1, \\ \mu + 2\sigma^2 & = \ln \alpha_2, \end{cases}$$

d'où on trouve les estimateurs  $\tilde{\sigma}_n^2$  et  $\tilde{\mu}_n$  :

$$\tilde{\sigma}_n^2 = \ln \alpha_2 - \ln \alpha_1^2 = \ln \left( \frac{s_n^2}{\bar{X}_n^2} + 1 \right), \quad \tilde{\mu}_n = \ln \frac{\bar{X}_n^2}{\sqrt{s_n^2 + \bar{X}_n^2}},$$

où

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

est la variance de la loi empirique.

**Exemple 2.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim p(x; \theta) = \frac{1}{\theta} \sqrt{\frac{2}{\pi}} \exp \left\{ -\frac{x^2}{2\theta^2} \right\} \mathbf{1}_{]0, \infty[}(x), \quad x \in \mathbf{R}^1, \quad \theta \in \Theta = ]0, \infty[.$$

On peut montrer que

$$\mathbf{E}X_1 = \theta \sqrt{\frac{2}{\pi}}, \quad \mathbf{E}X_1^2 = \theta^2, \quad \mathbf{Var} X_1^2 = \theta^2 \frac{\pi - 2}{\pi}.$$

Pour estimer  $\theta$  par la méthode des moments on considère l'équation

$$\theta \sqrt{\frac{2}{\pi}} = \bar{X}_n,$$

d'où on obtient l'estimateur

$$\tilde{\theta}_n = \sqrt{\frac{\pi}{2}} \bar{X}_n.$$

Il est clair que  $\mathbf{E}\tilde{\theta}_n = \theta$ , i.e.  $\tilde{\theta}_n$  est un estimateur sans biais pour  $\theta$ , et comme

$$\mathbf{Var} \bar{X}_n = \frac{\theta^2}{n} \left( 1 - \frac{2}{\pi} \right),$$

on en tire que

$$\begin{aligned} \mathbf{Var} \tilde{\theta}_n &= \frac{\pi}{2} \mathbf{Var} \bar{X}_n = \frac{\theta^2}{n} \left( \frac{\pi}{2} - 1 \right) = \\ &= \frac{\theta^2 \pi - 2}{n} = \frac{\pi - 2}{I_n(\theta)} > \frac{1}{I_n(\theta)}, \end{aligned}$$

où

$$I_n(\theta) = \frac{2n}{\theta^2} = -n\mathbf{E} \frac{\partial^2}{\partial \theta^2} \ln p(X_1; \theta) = n\mathbf{E} \left( \frac{3}{\theta^4} X_1^2 - \frac{1}{\theta^2} \right) = \frac{2n}{\theta^2}$$

est l'information de Fisher sur  $\theta$  dans  $\mathbb{X}$ . De la dernière inégalité on voit bien que l'estimateur  $\tilde{\theta}_n$  n'est pas efficace.

**Remarque 1.** Du théorème limite central il suit que la suite des variables aléatoires

$$\frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\theta \sqrt{\frac{\pi-2}{2}}} = \frac{\sqrt{n}(\bar{X}_n - \sqrt{\frac{2}{\pi}}\theta)}{\theta \sqrt{1 - \frac{2}{\pi}}}, \quad n = 1, 2, \dots$$

est asymptotiquement normale  $N(0, 1)$ , quand  $n \rightarrow \infty$ , i.e. pour les grandes valeurs de  $n$

$$\mathbf{P} \left\{ \frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\theta \sqrt{\frac{\pi-2}{\pi}}} \leq x \right\} \approx \Phi(x), \quad x \in \mathbf{R}^1.$$

Du théorème de Slutsky on tire que les variables aléatoires

$$\frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\tilde{\theta}_n \sqrt{\frac{\pi-2}{2}}}$$

sont asymptotiquement normales  $N(0, 1)$  aussi, i.e.

$$\mathbf{P} \left\{ \frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\tilde{\theta}_n \sqrt{\frac{\pi-2}{2}}} \leq x \right\} \approx \Phi(x), \quad x \in \mathbf{R}^1,$$

si les valeurs de  $n$  sont assez grandes.

Nous pouvons utiliser ce résultat pour estimer  $\theta$  par intervalle, puisque

$$\mathbf{P} \left\{ -\bar{x}_{\alpha/2} \leq \frac{\sqrt{n}(\tilde{\theta}_n - \theta)}{\tilde{\theta}_n \sqrt{\frac{\pi-2}{2}}} \leq \bar{x}_{\alpha/2} \right\} \approx 1 - \alpha,$$

où  $\bar{x}_{\alpha/2}$  est le quantile supérieur de niveau  $\alpha/2$  pour la loi standard normale,  $0 < \alpha < 0.5$ , d'où on tire que

$$\mathbf{P} \left\{ -\bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \leq \left( 1 - \frac{\theta}{\tilde{\theta}_n} \right) \leq \bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \right\} \approx 1 - \alpha$$

et donc

$$\mathbf{P} \left\{ \tilde{\theta}_n \left( 1 - \bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \right) \leq \theta \leq \tilde{\theta}_n \left( 1 + \bar{x}_{\alpha/2} \sqrt{\frac{\pi-2}{2n}} \right) \right\} \approx 1 - \alpha,$$

si  $n$  est assez grand.

**Exemple 3.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim f(x; \theta) = \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(x), \quad \theta \in \Theta = ]0, \infty[,$$

i.e.  $X_i$  suit la loi uniforme sur  $[0, \theta]$ . Dans ce cas la fonction de vraisemblance est

$$L(\theta) = L(\mathbb{X}; \theta) = \prod_{j=1}^n \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(X_j) = \frac{1}{\theta^n} \mathbf{1}_{[0, \theta]}(X_{(n)}),$$

puisque  $\mathbf{P}\{0 \leq X_{(1)} \leq X_{(n)} \leq \theta\} = 1$ , d'où on tire que  $X_{(n)}$  est une statistique exhaustive minimale. Il est évident que  $\hat{\theta}_n = X_{(n)}$ .

Donc, pour estimer  $\theta$ , nous pouvons utiliser la statistique  $\hat{\theta}_n = X_{(n)}$  comme estimateur ponctuel.

Par ailleurs, comme  $\mathbf{E}X_i = \theta/2$  on en déduit que la statistique

$$\theta_n^* = 2\bar{X}_n = \frac{2}{n} \sum_{i=1}^n X_i$$

peut être considérée comme un autre estimateur sans biais de  $\theta$ , puisque

$$\mathbf{E}_\theta \theta_n^* = \theta.$$

On va comparer les deux estimateurs  $\hat{\theta}_n$  et  $\theta_n^*$ . Comme  $\mathbf{Var}X_i = \theta^2/12$ , il s'ensuit que

$$\mathbf{Var}\theta_n^* = \mathbf{Var}\left(\frac{2}{n} \sum_{i=1}^n X_i\right) = \frac{4}{n^2} \sum_{i=1}^n \mathbf{Var}X_i = \frac{\theta^2}{3n} = O\left(\frac{1}{n}\right) \rightarrow 0, \quad (n \rightarrow \infty),$$

et donc du critère de consistance on tire que  $\{\theta_n^*\}$  converge en probabilité vers  $\theta$ , i.e.  $\{\theta_n^*\}$  est une suite consistante d'estimateurs sans biais de  $\theta$ . De plus d'après le théorème central limite on obtient que pour tout  $x \in \mathbf{R}^1$

$$\mathbf{P}_\theta \left\{ \frac{\theta_n^* - \mathbf{E}\theta_n^*}{\sqrt{\mathbf{Var}\theta_n^*}} \leq x \right\} = \mathbf{P}_\theta \left\{ \frac{\sqrt{3n}(\theta_n^* - \theta)}{\theta} \leq x \right\} \rightarrow \Phi(x), \quad n \rightarrow \infty, \quad (1)$$

i.e.  $\{\theta_n^*\}$  est une suite d'estimateurs asymptotiquement normale de paramètres  $\theta$  et  $\theta/\sqrt{3n}$ .

Étudions maintenant la statistique  $\hat{\theta}_n = X_{(n)}$ , qui est l'estimateur de maximum de vraisemblance de  $\theta$ . Tout d'abord, on remarque que

$$\mathbf{P}_\theta\{0 \leq X_{(n)} \leq \theta\} = 1, \quad \theta > 0.$$

Pour tout  $t \in [0, \theta]$  on a

$$\mathbf{P}_\theta\{X_{(n)} \leq t\} = \mathbf{P}_\theta\{X_1 \leq t, \dots, X_n \leq t\} = \left(\frac{t}{\theta}\right)^n = G(t; \theta), \quad (2)$$

la densité  $g(t; \theta) = G'(t, \theta)$  de  $X_{(n)}$  est donc donnée par :

$$g(t; \theta) = G'(t; \theta) = \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} \mathbf{1}_{[0, \theta]}(t),$$

d'où on tire que

$$\mathbf{E}_\theta X_{(n)} = \mathbf{E}_\theta \hat{\theta}_n = \frac{n}{\theta} \int_0^\theta t \left(\frac{t}{\theta}\right)^{n-1} dt = \frac{n}{n+1} \theta,$$

$$\mathbf{E}_\theta X_{(n)}^2 = \frac{n}{\theta} \int_0^\theta t^2 \left(\frac{t}{\theta}\right)^{n-1} dt = \frac{n}{n+2} \theta^2,$$

donc

$$\begin{aligned} \mathbf{Var}_\theta X_{(n)} &= \frac{n}{n+2} \theta^2 - \frac{n^2}{(n+1)^2} \theta^2 = \\ &= \frac{n}{(n+2)(n+1)^2} \theta^2 = O\left(\frac{1}{n^2}\right) \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

On remarque que  $\{\hat{\theta}_n\}$  est une suite consistante d'estimateurs asymptotiquement sans biais du paramètre  $\theta$ , car pour tout  $n \in \mathbf{N}^*$  le biais  $b_n(\theta)$  de l'estimateur  $\hat{\theta}_n$  est

$$b_n(\theta) = \mathbf{E}_\theta(\hat{\theta}_n - \theta) = \frac{n}{n+1} \theta - \theta = -\frac{\theta}{n+1} \rightarrow 0 \quad (n \rightarrow \infty).$$

Le risque quadratique  $R(\hat{\theta}_n, \theta)$  de  $\hat{\theta}_n$  est égal à

$$R(\hat{\theta}_n, \theta) = \mathbf{Var} \hat{\theta}_n + b_n(\theta)^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$

Soit

$$\theta_n^{**} = \frac{n+1}{n} \hat{\theta}_n, \quad n \in \mathbf{N}.$$

Comme

$$\mathbf{E}_\theta \theta_n^{**} = \theta \text{ et } \mathbf{Var}_\theta \theta_n^{**} = \frac{(n+1)^2}{n^2} \mathbf{Var}_\theta \hat{\theta}_n = \frac{\theta^2}{n(n+2)} = O\left(\frac{1}{n^2}\right),$$

on voit que  $\{\theta_n^{**}\}$  est une suite consistante d'estimateurs sans biais du paramètre  $\theta$ .

Pour trouver la loi limite de  $X_{(n)} = \hat{\theta}_n$  on remarque que pour les grandes valeurs de  $n$ ,  $\mathbf{Var} X_{(n)} \asymp \frac{\theta^2}{n^2}$  et donc pour tout  $x > 0$

$$\begin{aligned} \mathbf{P}_\theta \left\{ 0 \leq \frac{\theta - X_{(n)}}{\theta/n} \leq x \right\} &= \mathbf{P}_\theta \left\{ X_{(n)} \geq \theta \left(1 - \frac{x}{n}\right) \right\} = \\ 1 - \mathbf{P}_\theta \left\{ X_{(n)} \leq \theta \left(1 - \frac{x}{n}\right) \right\} &= 1 - \left(1 - \frac{x}{n}\right)^n \rightarrow 1 - e^{-x}, \quad (n \rightarrow \infty). \end{aligned} \quad (3)$$

Choisissons un coefficient de confiance  $\mathcal{P} = 1 - \alpha$ , où  $0 < \alpha < 0.5$ , et donc  $0.5 < \mathcal{P} < 1$ , et, en utilisant (1) et (3), trouvons les deux quantiles  $\bar{x}_{\frac{\alpha}{2}}$  et  $y_\alpha$  tels que :

$$\mathbf{P}_\theta \left\{ \left| \theta_n^* - \theta \right| \leq \frac{\bar{x}_{\frac{\alpha}{2}} \theta}{\sqrt{3n}} \right\} \approx 1 - 2\Phi(-\bar{x}_{\frac{\alpha}{2}}) = \mathcal{P} = 1 - \alpha,$$

$$\mathbf{P}_\theta \left\{ \hat{\theta}_n \leq \theta \leq \frac{\hat{\theta}_n}{\left(1 - \frac{y_\alpha}{n}\right)} \right\} \approx 1 - e^{-y_\alpha} = \mathcal{P} = 1 - \alpha.$$

On a donc construit 2 intervalles de confiance de niveaux de confiance  $\approx \mathcal{P} = 1 - \alpha$  pour la valeur inconnue  $\theta$ , basés sur les estimateurs  $\theta_n^*$  et  $\hat{\theta}_n$  :

$$\theta_n^* \left(1 + \frac{\bar{x}_{\frac{\alpha}{2}}}{\sqrt{3n}}\right)^{-1} \leq \theta \leq \theta_n^* \left(1 - \frac{\bar{x}_{\frac{\alpha}{2}}}{\sqrt{3n}}\right)^{-1}$$

et

$$\hat{\theta}_n \leq \theta \leq \hat{\theta}_n \left(1 - \frac{y_\alpha}{n}\right)^{-1}$$

de longueurs

$$l_n^* = l(\theta_n^*) \approx 2\theta_n^* \bar{x}_{\frac{\alpha}{2}} / \sqrt{3n} \quad \text{et} \quad \hat{l}_n = l(\hat{\theta}_n) \approx \hat{\theta}_n y_\alpha / n$$

respectivement, d'où on tire que

$$\frac{l_n^*}{\hat{l}_n} \approx \sqrt{n} \frac{2\bar{x}_{\frac{\alpha}{2}}}{\sqrt{3}y_\alpha} \quad (n \rightarrow \infty),$$

car  $\theta_n^*/\hat{\theta}_n$  est très proche de 1 avec une grande probabilité. Par exemple, si  $\alpha = 0.05$ , soit  $\mathcal{P} = 0.95$ , on a  $\bar{x}_{\frac{\alpha}{2}} = 1.96$ ,  $y_\alpha = 2.99$  et dans ce cas

$$\frac{l_n^*}{\hat{l}_n} \approx 0.76\sqrt{n}.$$

**Remarque 2.** On voit que

$$R(\theta_n^*, \theta) = \mathbf{Var}\theta_n^* = \frac{\theta^2}{3n}, \quad R(\hat{\theta}_n, \theta) = \frac{2\theta^2}{(n+1)(n+2)},$$

$$R(\theta_n^{**}, \theta) = \mathbf{Var}\theta_n^{**} = \frac{\theta^2}{n(n+2)},$$

d'où on tire que  $\theta_n^*$  et  $\hat{\theta}_n$  sont des estimateurs inadmissibles pour  $\theta$  par rapport à la fonction de perte quadratique, puisque

$$R(\theta_n^{**}, \theta) < R(\theta_n^*, \theta), \quad \theta \in \Theta,$$

et pour tout  $n \geq 2$

$$R(\theta_n^{**}, \theta) < R(\hat{\theta}_n, \theta), \quad \theta \in \Theta.$$

**Exemple 4.** Changeons un peu le problème. Supposons que dans les conditions de l'exemple 1 on ait :

$$f(x; \theta) = \frac{1}{\theta} \mathbf{1}_{]0, \theta[}(x), \quad \theta > 0,$$

i.e.  $X_i$  suit la loi uniforme sur  $]0, \theta[$ . Alors,

$$L(\theta) = \prod_{j=1}^n \frac{1}{\theta^n} \mathbf{1}_{]0, \theta[}(X_j) = \frac{1}{\theta^n} \mathbf{1}_{]0, \theta[}(X_{(n)}), \quad \theta \in \Theta = ]0, \infty[.$$

Donc,  $X_{(n)}$  est une statistique exhaustive, mais  $L(\theta)$  n'a pas de maximum et donc, il n'existe pas de l'estimateur du maximum de vraisemblance pour  $\theta$ .

On sait d'après la définition d'un estimateur,  $\theta_n^* : \mathbf{R}^n \rightarrow \Theta$ , qu'il faut qu'il prenne ces valeurs dans  $\Theta$ , mais ici  $X_{(n)}$  n'appartient pas à  $\Theta$  ( $X_{(n)}$  est toujours plus petit que  $\theta$ ); par conséquent dans cet exemple l'estimateur de maximum de vraisemblance n'existe pas. On peut choisir  $\theta$  très proche de  $X_{(n)}$ , mais pas égal à  $X_{(n)}$ .

**Exemple 5.** Donnons maintenant un exemple de non unicité de l'estimateur de maximum de vraisemblance lié avec une loi uniforme.

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$ ,

$$H_0 : X_i \sim f(x; \theta) = \mathbf{1}_{[\theta, \theta+1]}(x), \quad \theta \in \Theta = \mathbf{R}^1.$$

La fonction de vraisemblance est

$$L(\theta) = \mathbf{1}_{[\theta, \theta+1]}(X_{(1)}) \mathbf{1}_{[\theta, \theta+1]}(X_{(n)}) = \mathbf{1}_{[X_{(n)}-1, X_{(1)}]}, \quad \theta \in \Theta = \mathbf{R}^1.$$

et donc  $T = (X_{(1)}, X_{(n)})^T$  est une statistique exhaustive minimale. On remarque que  $T \in \mathbf{R}^2$ , tandis que  $\theta \in \Theta = \mathbf{R}^1$ . N'importe quel  $\theta$  dans l'intervalle  $[X_{(n)} - 1, X_{(1)}] \subset \Theta$  peut-être considéré comme estimateur de maximum de vraisemblance ; en particulier

$$\hat{\theta}_1 = X_{(1)} \quad \text{ou} \quad \hat{\theta}_2 = X_{(n)} - 1.$$

On note que ni  $\hat{\theta}_1$  ni  $\hat{\theta}_2$  ne sont des statistiques exhaustives, mais ce sont des statistiques nécessaires.

On remarque que c'est justement en ces deux points

$$\hat{\theta}_1 = X_{(1)} \quad \text{et} \quad \hat{\theta}_2 = X_{(n)} - 1,$$

que  $L(\theta)$  a des ruptures (des sauts). Pour construire estimateur sans biais pour  $\theta$  on peut prendre, par exemple, la statistique

$$\theta_n^* = \frac{1}{2}(\hat{\theta}_1 + \hat{\theta}_2) = \frac{X_{(1)} + X_{(n)} - 1}{2}, \quad \mathbf{E}\theta_n^* = \theta. \quad (4)$$

On peut montrer que

$$\mathbf{Var}\theta_n^* = \frac{1}{2(n+1)(n+2)}.$$

**Remarque 3.** En présence d'une statistique exhaustive  $T$  pour  $\theta$  l'estimateur de maximum de vraisemblance  $\hat{\theta}_n = \hat{\theta}_n(T)$  est donc une statistique nécessaire.

**Remarque 4.** Soit  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  un échantillon, dont la réalisation observée est  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ . Notre problème est de construire une loi empirique, en utilisant le vecteur des données  $\mathbf{x}$  et le principe du maximum de vraisemblance. Comme les éléments  $X_i$  de l'échantillon  $\mathbb{X}$  sont indépendants, on peut écrire que

$$\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \bigcap_{i=1}^n \{X_i = x_i\},$$

donc

$$\mathbf{P}\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \mathbf{P}\left[\bigcap_{i=1}^n \{X_i = x_i\}\right] = \prod_{i=1}^n \mathbf{P}\{X_i = x_i\}.$$

Pour construire une loi empirique il faut choisir les probabilités

$$p_i = \mathbf{P}\{X_i = x_i\} \geq 0, \quad i = 1, 2, \dots, n,$$

telles que

$$p_1 + p_2 + \dots + p_n = 1, \quad p_i \geq 0.$$

Le principe du maximum de vraisemblance nous dit qu'il faut choisir les  $p_i$  de façon que le produit

$$\prod_{i=1}^n p_i$$

soit maximal. Comme

$$\left( \prod_{i=1}^n p_i \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n p_i \text{ et } \sum_{i=1}^n p_i \leq 1,$$

on en déduit que

$$\left( \prod_{i=1}^n p_i \right)^{1/n} \leq \frac{1}{n},$$

et donc

$$\prod_{i=1}^n p_i \leq \left( \frac{1}{n} \right)^n,$$

d'où on trouve que notre solution est

$$p_1 = p_2 = \dots = p_n = \frac{1}{n},$$

et c'est donc la loi empirique classique qui donne la meilleure solution au sens du principe de maximum de vraisemblance.

**Remarque 5.** (*Principe d'invariance de l'estimateur de maximum de vraisemblance*).

Soit  $\hat{\theta}_n$  l'estimateur de maximum de vraisemblance de  $\theta$ ,  $\theta \in \Theta \subset \mathbf{R}^n$ . Supposons que nous voulions estimer la valeur  $g(\theta)$  d'une application  $g : \Theta \rightarrow G \subset \mathbf{R}^1$ . Dans ce cas

$$\hat{g} = g(\hat{\theta}_n) \tag{5}$$

est l'estimateur de maximum de vraisemblance pour  $g(\theta)$ .

Par exemple, si

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

est l'estimateur du maximum de vraisemblance pour la variance  $\sigma^2$  de la loi normale  $N(\mu, \sigma^2)$ , quand  $\mu$  et  $\sigma^2$  sont inconnus, alors

$$s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

est l'estimateur de maximum de vraisemblance pour  $\sigma$ .

Pour démontrer (29), notons

$$\Theta_g = \{\theta : \theta \in \Theta, \quad g(\theta) = g\}, \quad g \in G,$$

i.e.  $\Theta_g$  est l'orbite de l'application  $g(\theta)$ , correspondant à une valeur  $g$  de  $g(\theta)$ . Il est évident que  $\{\Theta_g\}$  est une partition de  $\Theta$ ,

$$\bigcup_{g \in G} \Theta_g = \Theta, \quad \Theta_{g'} \cap \Theta_g = \emptyset.$$

Soit

$$L_g = \sup_{\theta \in \Theta_g} L(\theta), \quad g \in G.$$

Il est évident que

$$L(\hat{\theta}_n) = \sup_{\theta \in \Theta} L(\theta) = \sup_{g \in G} \sup_{\theta \in \Theta_g} L(\theta) = \sup_{g \in G} L_g.$$

Choisissons

$$\hat{g} = g(\hat{\theta}_n), \quad \hat{g} \in G,$$

et considérons l'orbite  $\Theta_{\hat{g}}$ ,  $\hat{\theta}_n \in \Theta_{\hat{g}}$ .

Comme pour tout  $g \in G$

$$\sup_{g \in G} L_g \geq L_g$$

et, en particulier,

$$\sup_{g \in G} L_g \geq L_{\hat{g}} = \sup_{\theta \in \Theta_{\hat{g}}} L(\theta) = L(\hat{\theta}_n),$$

on en tire que  $L(\hat{\theta}_n) = L_{\hat{g}}$ , et donc (29) est démontrée.

**Exemple 5.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon

$$X_i \sim f(x_i; \theta) = \theta^{x_i} (1 - \theta)^{1 - x_i}, \quad x_i \in \mathcal{X} = \{0, 1\}, \quad 0 < \theta < 1.$$

Supposons que nous voulions estimer  $g(\theta) = 1/\theta$ . Kolmogorov a montré que parmi les fonctions de  $\theta$ , seuls les polynômes

$$u_m(\theta) = \sum_{k=1}^m a_k \theta^k, \quad \theta \in \Theta = ]0, 1[, \quad 1 \leq m \leq n,$$

de degré  $m \leq n$ , sont estimables, c'est-à-dire peuvent être estimés à l'aide d'estimateurs sans biais en termes de la statistique exhaustive  $\mu_n = \sum_{i=1}^n X_i$ . Comme  $g(\theta) = 1/\theta$  n'est pas un polynôme, il n'existe pas d'estimateur sans biais pour  $1/\theta$ . Mais comme l'estimateur de maximum de vraisemblance  $\hat{\theta}_n = \mu_n/n$  existe pour  $\theta$ , du principe du maximum de vraisemblance on tire que

$$T_n = g(\hat{\theta}_n) = \frac{n}{\mu_n}$$

est l'estimateur de maximum de vraisemblance pour  $1/\theta$ . On remarque que  $\mathbf{E}_\theta T_n$  n'existe pas puisque

$$\mathbf{P}_\theta\{\mu_n = 0\} = (1 - \theta)^n > 0.$$

Par ailleurs, comme nous estimons  $g(\theta) = 1/\theta$ , la borne inférieure dans l'inégalité de Rao-Cramer-Fréchet est égale à

$$\frac{[g'(\theta)]^2}{\mathbf{I}_n(\theta)} = \frac{\theta(1 - \theta)}{\theta^4 n} = \frac{1 - \theta}{n\theta^3},$$

et donc

$$T_n \sim AN\left(\frac{1}{\theta}, \frac{1-\theta}{n\theta^3}\right),$$

i.e. pour tout  $x \in \mathbf{R}^1$

$$\mathbf{P}_\theta \left\{ \sqrt{\frac{n\theta^3}{1-\theta}} \left( \frac{n}{\mu_n} - \frac{1}{\theta} \right) \leq x \right\} \rightarrow \Phi(x).$$

**Exemple 6.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon normale,

$$X_i \sim N(\theta, \theta), \quad \theta \in \Theta = ]0, \infty[.$$

Considérons le problème d'estimation du paramètre  $\theta$  dans ce modèle. On remarque que

$$\theta = \mathbf{E}X_i = \mathbf{Var} X_i.$$

Dans ce cas la fonction de vraisemblance est

$$\begin{aligned} L(\theta) = L(\mathbb{X}, \theta) &= \frac{1}{(2\pi\theta)^{n/2}} \prod_{i=1}^n \exp\left\{-\frac{1}{2\theta}(X_i - \theta)^2\right\} = \\ &= \frac{1}{(2\pi\theta)^{n/2}} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n X_i^2 + \sum_{i=1}^n X_i - \frac{n\theta}{2}\right\} = \\ &= \frac{1}{(2\pi\theta)^{n/2}} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n X_i^2 - \frac{n\theta}{2}\right\} \exp\left\{\sum_{i=1}^n X_i\right\}, \end{aligned}$$

d'où on tire que la statistique

$$T_n = \sum_{i=1}^n X_i^2$$

est exhaustive et minimale pour  $\theta$ . Il est intéressant de noter que la statistique

$$\sum_{i=1}^n X_i = n\bar{X}_n$$

n'est pas une statistique exhaustive dans notre problème ! Puisque  $L(\theta) > 0$  pour tout  $\theta \in \Theta$  et

$$\lim_{\theta \downarrow 0} L(\theta) = \lim_{\theta \rightarrow \infty} L(\theta) = 0,$$

on en tire que l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  de  $\theta$  est la racine positive de l'équation du maximum de vraisemblance  $\Lambda(\theta) = 0$ , où

$$\Lambda(\theta) = \frac{\partial}{\partial \theta} \ln L(\theta) = -\frac{n}{2\theta} + \frac{T_n}{2\theta^2} - \frac{n}{2}.$$

Donc  $\hat{\theta}_n$  est la racine positive de l'équation

$$\theta^2 + \theta - \frac{1}{n}T_n = 0,$$

i.e.

$$\hat{\theta}_n = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{n}T_n}.$$

Il est facile de vérifier que  $\{\hat{\theta}_n\} \xrightarrow{\mathbf{P}} \theta$ . En effet, d'après la loi des grands nombres

$$\frac{1}{n}T_n \xrightarrow{\mathbf{P}} \mathbf{E}_\theta X_1^2 = \mathbf{Var}_\theta X_1 + (\mathbf{E}_\theta X_1)^2 = \theta + \theta^2,$$

d'où, en utilisant le théorème de Slutsky, on tire que

$$\hat{\theta}_n \xrightarrow{\mathbf{P}} -\frac{1}{2} + \sqrt{\frac{1}{4} + \theta + \theta^2} = -\frac{1}{2} + \frac{1}{2} + \theta = \theta.$$

**Remarque 6.** Soit  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  un échantillon,  $X_i$  suit la loi, dont la densité  $f(x; \theta)$  appartient à la famille  $\mathcal{F} = \{f(x; \theta)\}$ , où

$$f(x; \theta) = h(x) \exp \left\{ \sum_{k=1}^n \theta_k x^k + V(\theta) \right\}, \quad x \in \mathcal{X}, \quad (2.7)$$

$\theta = (\theta_1, \theta_2, \dots, \theta_s)^T \in \Theta \subset R^s$ ,  $\mathcal{X}$  est un ensemble borelien en  $R^1$ . La famille (6) est très riche. Par exemple, la famille des distributions normales  $N(\mu, \sigma^2)$ ,  $\theta = (\mu, \sigma^2)^T$ , appartient à  $\mathcal{F}$ , la famille des distributions de Poisson appartient à  $\mathcal{F}$  aussi etc. Comme il est connu la statistique

$$U_n = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \dots, \sum_{i=1}^n X_i^s \right)^T$$

est exhaustive pour la famille (6).

Supposons que

- 1) l'ensemble  $\mathcal{X}$  ne dépend pas de paramètre  $\theta$ ;
- 2) la matrice de Hesse

$$- \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} V(\theta) \right\|_{s \times s}$$

de la fonction  $V(\theta)$  est positivement définie sur  $\Theta$ ;

- 3) il existe le moment  $a_s = \mathbf{E}_\theta X_1^s$ .

Dans ce cas

$$-\mathbf{grad} V(\theta) = a(\theta) = (a_1(\theta), a_2(\theta), \dots, a_s(\theta))^T,$$

et donc la statistique  $T = \frac{1}{n}U_n$  est le meilleur estimateur sans biais pour  $a(\theta)$ , i.e.

$$\mathbf{E}_\theta T_n = a(\theta),$$

ce qui nous permet d'estimer  $\theta$  (trouver l'estimateur  $\theta_n^*$  par la méthode des moments de façon unique de l'équation  $T_n = a(\theta)$  dans les termes de la statistique exhaustive  $U_n$ ). De l'autre côté les conditions 1)–3) sont suffisantes (voir, par exemple, Zacks, 1971) pour l'existence de l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  :

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta), \quad \text{où} \quad L(\theta) = \prod_{i=1}^n f(X_i, \theta),$$

et pour la famille (6) l'estimateur  $\hat{\theta}_n$  est la racine unique de la même équation  $T_n = a(\theta)$ , et donc de ce fait on tire que pour la famille exponentielle (6) la méthode du maximum de vraisemblance et la méthode des moments donnent le même estimateur  $\theta_n^* = \hat{\theta}_n$  pour le paramètre  $\theta$ .

**Exemple 7.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i$  suit la loi normale  $N(\mu, \sigma^2)$ ,  $\theta = (\mu, \sigma^2)^T$ . Dans ce cas la statistique

$$\hat{\theta}_n = (\bar{X}_n, s_n^2)^T$$

est l'estimateur du maximum de vraisemblance pour  $\theta$  et elle-même nous donne l'estimateur par la méthode des moments.

**Exemple 8.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i$  suit la loi de Poisson de paramètre  $\theta$ ,  $\theta \in \Theta = ]-\infty, +\infty[$  :

$$\mathbf{P}_\theta\{X_i = k\} = \frac{\theta^k}{k!} e^{-\theta}, \quad k = 0, 1, 2, \dots$$

Dans ce cas la statistique  $\sum_{i=1}^n X_i$  est exhaustive pour le paramètre  $\theta$  et donc la moyenne  $\bar{X}_n$  de la loi empirique est le meilleur estimateur sans biais pour  $\theta$  et en même temps  $\bar{X}_n$  est l'estimateur du maximum de vraisemblance pour  $\theta$ .

**Exemple 9.** On a  $n$  expériences indépendantes de Bernoulli avec trois états possibles  $E_1, E_2, E_3$ ,  $E_1 \cup E_2 \cup E_3 = \Omega$ ,  $E_i \cap E_j = \emptyset$ , dont les probabilités sont

$$\begin{cases} \mathbf{P}(E_1) = p_1(\theta) = \theta, \\ \mathbf{P}(E_2) = p_2(\theta) = 2\theta, \\ \mathbf{P}(E_3) = p_3(\theta) = 1 - 3\theta, \end{cases}$$

où  $0 < \theta < 1/3$ . Trouver l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  pour  $\theta$ .

**Solution.** Soit  $\mathbf{v} = (v_1, v_2, v_3)^T$  le vecteur des fréquences observées,  $n = v_1 + v_2 + v_3$  - le nombre des épreuves. Comme la distribution du vecteur  $\mathbf{v}$  est trinominale des paramètres  $n$  et  $\mathbf{p} = (p_1, p_2, p_3)^T$ ,  $p_i = p_i(\theta)$ , la fonction de vraisemblance  $L(\theta)$  est

$$L(\theta) = \frac{n!}{v_1!v_2!v_3!} p_1^{v_1} p_2^{v_2} p_3^{v_3} = \frac{n!}{v_1!v_2!v_3!} \theta^{v_1} (2\theta)^{v_2} (1 - 3\theta)^{v_3}, \quad (7)$$

et donc

$$\ln L(\theta) = \text{const} + (v_1 + v_2) \ln \theta + v_3 \ln(1 - 3\theta).$$

Par conséquent l'équation de vraisemblance

$$\Lambda(\theta) = \frac{d \ln L(\theta)}{d\theta} = 0 \quad (8)$$

s'écrit de la façon suivante :

$$\Lambda(\theta) = \frac{d \ln L(\theta)}{d\theta} = \frac{v_1 + v_2}{\theta} - \frac{3v_3}{1 - 3\theta} = 0,$$

d'où l'on tire l'équation

$$\frac{v_1 + v_2}{\theta} = \frac{3v_3}{1 - 3\theta},$$

dont la racine  $\hat{\theta}_n$  est

$$\hat{\theta}_n = \frac{v_1 + v_2}{3n}.$$

On a trouvé l'estimateur du maximum de vraisemblance  $\hat{\theta}$  et donc

$$\begin{cases} \hat{p}_1 = p_1(\hat{\theta}_n) = \hat{\theta}_n, \\ \hat{p}_2 = p_2(\hat{\theta}_n) = 2\hat{\theta}_n, \\ \hat{p}_3 = p_3(\hat{\theta}_n) = 1 - 3\hat{\theta}_n, \end{cases}$$

sont les estimateurs du maximum de vraisemblance de  $p_i(\theta)$ ,  $i = 1, 2, 3$ .

En général  $p_i = p_i(\theta)$  sont des fonctions de  $\theta$  plus compliquées et dans ce cas l'équation de vraisemblance (8) n'est pas si facile à résoudre. Par exemple, dans notre cas, que l'on vient de considérer, on a

$$\Lambda(\theta) = \frac{d \ln L(\theta)}{d\theta} = v_1 \frac{p'_1(\theta)}{p_1(\theta)} + v_2 \frac{p'_2(\theta)}{p_2(\theta)} + v_3 \frac{p'_3(\theta)}{p_3(\theta)} = 0. \quad (9)$$

Comme

$$p_1(\theta) + p_2(\theta) + p_3(\theta) \equiv 1,$$

on a

$$p'_1(\theta) + p'_2(\theta) + p'_3(\theta) \equiv 0 \quad \text{et} \quad p''_1(\theta) + p''_2(\theta) + p''_3(\theta) \equiv 0,$$

et de (1) on tire que

$$\begin{aligned} & \frac{d^2}{d\theta^2} \ln L(\theta) = \\ & v_1 \left[ \frac{p''_1(\theta)}{p_1(\theta)} - \left( \frac{p'_1(\theta)}{p_1(\theta)} \right)^2 \right] + v_2 \left[ \frac{p''_2(\theta)}{p_2(\theta)} - \left( \frac{p'_2(\theta)}{p_2(\theta)} \right)^2 \right] + v_3 \left[ \frac{p''_3(\theta)}{p_3(\theta)} - \left( \frac{p'_3(\theta)}{p_3(\theta)} \right)^2 \right]. \end{aligned}$$

Pour trouver une bonne approximation de la racine  $\hat{\theta}_n$  de l'équation (9), nous pouvons appliquer la procédure suivante (*the scoring method* of Fisher). Soit

$$\hat{p}_i = \frac{v_i}{n}, \quad i = 1, 2, 3, \quad (10)$$

les estimateurs de maximum de vraisemblance pour des probabilités  $p_i(\theta)$ . Parmi ces trois équations  $p_i(\theta) = \hat{p}_i$  (par rapport à  $\theta$ ) on choisit la plus simple d'où l'on tire la solution  $\hat{\theta}_{0n}$ , que l'on peut prendre comme approximation initiale pour l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$ . Comme dans notre cas l'information de Fisher

$$I_n(\theta) = -\mathbf{E} \left\{ \frac{d^2}{d\theta^2} \ln L(\theta) \right\}$$

est égale à

$$I_n(\theta) = n \left[ \frac{(p'_1(\theta))^2}{p_1(\theta)} + \frac{(p'_2(\theta))^2}{p_2(\theta)} + \frac{(p'_3(\theta))^2}{p_3(\theta)} \right],$$

on trouve une nouvelle approximation  $\hat{\theta}_{1n}$ , qui est donnée par la formule suivante :

$$\hat{\theta}_{1n} = \hat{\theta}_{0n} + \frac{1}{I_n(\hat{\theta}_{0n})} \left. \frac{d \ln L(\theta)}{d\theta} \right|_{\theta=\hat{\theta}_{0n}}. \quad (11)$$

On peut montrer que l'estimateur  $\hat{\theta}_{1n}$  est asymptotiquement équivalent à l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$ , c'est-à-dire si  $n \rightarrow \infty$ , alors

$$\sqrt{I_n(\theta)} (\hat{\theta}_{1n} - \theta)$$

suit dans la limite la loi normale de paramètre 0 et 1,

$$\lim_{n \rightarrow \infty} \left\{ \sqrt{I_n(\theta)} (\hat{\theta}_{1n} - \theta) < x \right\} = \Phi(x).$$

Par exemple, soit  $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$ , où

$$\begin{cases} p_1 = p_1(\theta) = 2\theta, \\ p_2 = p_2(\theta) = 0.5 - 4\theta, \\ p_3 = p_3(\theta) = 0.5 + \theta, \\ p_4 = p_4(\theta) = \theta. \end{cases}$$

Il est clair que  $0 \leq \theta \leq 1/8$ . Comme la fonction de vraisemblance  $L(\theta)$  est égale à

$$\begin{aligned} L(\theta) &= \frac{n!}{v_1! v_2! v_3! v_4!} p_1^{v_1} p_2^{v_2} p_3^{v_3} p_4^{v_4} = \\ &= \frac{n!}{v_1! v_2! v_3! v_4!} (2\theta)^{v_1} (0.5 - 4\theta)^{v_2} (0.5 + \theta)^{v_3} \theta^{v_4} = \\ &= \frac{n! 2^{v_1}}{v_1! v_2! v_3! v_4!} \theta^{v_1 + v_4} (0.5 - 4\theta)^{v_2} (0.5 + \theta)^{v_3} \end{aligned}$$

et donc on trouve que la statistique  $T = (v_1 + v_4, v_2, v_3)^T$  est exhaustive pour le paramètre  $\theta$ . Supposons que  $n = 1000$  et que l'on ait observé

$$v_1 = 195, v_2 = 110, v_3 = 590, v_4 = 105.$$

Notons

$$q_1 = p_1 + p_4, q_2 = p_2, q_3 = p_3 \text{ et } \mu_1 = v_1 + v_4, \mu_2 = v_2, \mu_3 = v_3.$$

Avec ces notations la fonction de vraisemblance  $L(\theta)$  peut s'écrire de la manière suivante :

$$L(\theta) = \text{const} (3\theta)^{\mu_1} (0.5 - 4\theta)^{\mu_2} (0.5 + \theta)^{\mu_3},$$

d'où l'on déduit

$$\ln L(\theta) = \ln(\text{const}) + \mu_1 \ln \theta + \mu_2 \ln(0.5 - 4\theta) + \mu_3 \ln(0.5 + \theta),$$

$$\frac{d \ln L(\theta)}{d\theta} = \frac{\mu_1}{\theta} - \frac{4\mu_2}{0.5 - 4\theta} + \frac{\mu_3}{0.5 + \theta}$$

et donc on obtient l'équation du maximum de vraisemblance

$$\mu_1(0.5 - 4\theta)(0.5 + \theta) - 4\mu_2\theta(0.5 + \theta) + \mu_3\theta(0.5 - 4\theta) = 0,$$

qui est équivalente à la suivante :

$$160\theta^2 + 15\theta - 3 = 0,$$

dont les solutions  $\theta_1$  et  $\theta_2$  sont données par les formules suivantes :

$$\theta_1 = \frac{-15 + \sqrt{225 + 160 * 12}}{320} \quad \text{et} \quad \theta_2 = \frac{-15 - \sqrt{225 + 160 * 12}}{320}.$$

Comme  $0 < \theta < 1/8$ , on en déduit que l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  est égale à  $\theta_1$  et donc on obtient que

$$\hat{\theta}_n = \theta_1 = \frac{-15 + 46.31}{320} \cong 0.0978.$$

Comme

$$\frac{d^2}{d\theta^2} \ln L(\theta) = -\frac{\mu_1}{\theta^2} - \frac{16\mu_2}{(0.5 - 4\theta)^2} - \frac{\mu_3}{(0.5 + \theta)^2},$$

et  $\mathbf{E}\mu_i = nq_i$ , on trouve que

$$I_n(\theta) = -\mathbf{E} \left\{ \frac{d^2}{d\theta^2} \ln L(\theta) \right\} = n \left[ \frac{3\theta}{\theta^2} + \frac{16}{0.5 - 4\theta} + \frac{1}{0.5 + \theta} \right] =$$

$$n \left[ \frac{3}{\theta} + \frac{32}{1 - 8\theta} + \frac{2}{1 + 2\theta} \right].$$

Comme on l'a déjà noté la variable aléatoire

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\frac{1}{I_n(\theta)}}} = \sqrt{I_n(\theta)}(\hat{\theta}_n - \theta)$$

suit à la limite quand  $n \rightarrow \infty$  la loi normale  $N(0, 1)$ . Du théorème de Cramer on déduit que

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\frac{1}{I_n(\hat{\theta}_n)}}} = \sqrt{I_n(\hat{\theta}_n)}(\hat{\theta}_n - \theta)$$

suit aussi à la limite la loi normale  $N(0, 1)$ .

Nous pouvons aussi utiliser le scoring méthode de Fisher pour trouver un estimateur de  $\theta$ . Si on prend  $\hat{\theta}_0 \cong 0.1$  comme approximation initiale, on trouve

$$I_n(\hat{\theta}_0) = \frac{1150000}{6}$$

et donc en utilisant la formule (35)

$$\hat{\theta}_{1n} = \hat{\theta}_0 + \frac{1}{I_n(\hat{\theta}_0)} \left. \frac{d \ln L(\theta)}{d\theta} \right|_{\theta=\hat{\theta}_0},$$

on trouve que

$$\hat{\theta}_{1n} = 0.1 + \frac{6}{1150000} \left[ 300 - \frac{440}{0.1} + \frac{590}{0.6} \right] = 0.1 - 0.0022 = 0.0978 = \hat{\theta}_n.$$

Admettons que quelqu'un suppose  $\theta = 0.11$ . Avec quelle certitude peut-on affirmer d'après les données observées que  $\theta = 0.11$  ?

Comme nous le savons

$$\mathbf{P} \left\{ \sqrt{I_n(\hat{\theta}_n)} |\hat{\theta}_n - \theta| > 0.0121 \sqrt{I_n(\hat{\theta}_n)} \right\} \cong 2 \left[ 1 - \Phi \left( 0.0121 \sqrt{I_n(\hat{\theta}_n)} \right) \right] =$$

$$2 [1 - \Phi(5.297)] = 5 \cdot 10^{-7},$$

ce qui nous permet d'affirmer l'invraisemblance que  $\theta = 0.11$ .

## 2.19 Décomposition orthogonale de Fisher

Supposons que les résultats d'une expérience soient présentés par la matrice

$$\mathbf{A} = \|a_{ij}\|, \quad i \in \mathbf{I} = \{1, \dots, I\}; \quad j \in \mathbf{J} = \{1, \dots, J\}.$$

Les valeurs observées  $a_{ij}$  nous pouvons considérer comme les valeurs  $a(i, j)$  d'une fonction  $a(\cdot, \cdot)$ , déterminée sur l'ensemble  $\mathbf{I} \star \mathbf{J}$ . On peut poser une question : est ce que la fonction  $a(\cdot, \cdot)$  est constante,

$$a(i, j) = a_{ij} = \text{const} = a_{..}, \quad (1)$$

ou peut-être c'est une fonction d'une variable, par exemple  $i$ , et qui prend les valeurs  $a_i$  :

$$a(i, j) = a_{ij} = a_{..} + \alpha_i, \quad (2)$$

où

$$\alpha_i = a_{i.} - a_{..}, \quad (3)$$

ou peut-être c'est une fonction présentée comme la somme de deux fonctions d'une variable chacune

$$a(i, j) = a_{ij} = a_{i.} + a_{.j} - a_{..} = a_{..} + \alpha_i + \alpha_{.j},$$

avec  $\alpha_{.j} = (a_{.j} - a_{..})$ , ou peut-être c'est une fonction de deux variables avec une *interaction* entre les arguments  $i$  et  $j$  :

$$a(i, j) = a_{ij} = a_{..} + \alpha_i + \alpha_{.j} + \alpha_{ij}, \quad (4)$$

où

$$\alpha_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..} \quad (5)$$

Toutes ces questions sont importantes si nous voulons construire une approximation pour la fonction  $a(\cdot, \cdot)$  et suivant l'information que nous avons nous pouvons proposer la meilleur approximation dans un certain sens.

Nous pouvons toujours compter que nous avons une fonction qui est présentée par la formule (\*) et donc il nous faut faire le meilleur choix des constantes, en utilisant des données.

Si nous avons la fonction de deux variables, alors il se peut qu'il soit intéressant de l'approximer par une fonction d'une variable ou par la somme de deux fonctions d'une

variable chacune, avec ou sans interactions. On cherchera l'approximation dans le sens de moindres carrés :

$$\sum_{i=1}^I \sum_{j=1}^J (a_{ij} - \alpha)^2 \rightarrow \min. \quad (6)$$

Le premier à avoir considéré ce problème en statistique est Sir R.Fisher qui a proposé de choisir les constantes suivantes :

$$a_{i.} = \frac{1}{J} \sum_{j=1}^J a_{ij}, \quad a_{.j} = \frac{1}{I} \sum_{i=1}^I a_{ij}, \quad (7)$$

$$a_{..} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J a_{ij} = \frac{1}{J} \sum_{j=1}^J a_{.j} = \frac{1}{I} \sum_{i=1}^I a_{i.}. \quad (8)$$

Donc dans le cas où nous cherchons la meilleure approximation par la fonction d'une variable, par exemple qui ne dépend que de  $i$ , d'après Fisher il faut choisir  $\alpha = a_{i.}$ . Si nous cherchons une approximation par la somme de deux fonction d'une variable chacune sans leurs interactions, alors il nous faut choisir

$$\alpha = a_{i.} + a_{.j} - a_{..} = a_{..} + (a_{i.} - a_{..}) + (a_{.j} - a_{..}), \quad (9)$$

etc. On fait ce problème de la même façon dans le cas continue. Par exemple, on peut introduire

$$a_{..} = \frac{1}{IJ} \int_0^I \int_0^J a(i, j) di dj, \quad a_{i.} = \frac{1}{I} \int_0^J a(i, j) dj, \quad (10)$$

$i \in \mathbf{I} = [0, I], j \in \mathbf{J} = [0, J]$ .

On peut généraliser les résultats de Fisher pour le cas de l'espace de plus haute dimension. Notons

$$[a_{ij}] = (a_{11}, a_{12}, \dots, a_{1J}, a_{21}, \dots, a_{2J}, \dots, a_{I1}, \dots, a_{IJ})^T$$

le vecteur-colonne de dimension  $IJ$ ,  $[a_{ij}] \in \mathbf{R}^{IJ}$ , c'est-à-dire tous les éléments de la matrice  $\mathbf{A}$  sont présentés en forme d'un vecteur de  $\mathbf{R}^{IJ}$ , et soit  $[a_{.}]$  le vecteur de même espace  $\mathbf{R}^{IJ}$ , dont tous les éléments sont égaux à  $a_{..}$ . Nous pouvons dire que

$$[a_{.}] = a_{..} \mathbf{1}_{IJ}, \text{ où } \mathbf{1}_{IJ} = (1, 1, \dots, 1)^T \in \mathbf{R}^{IJ}. \quad (11)$$

Dans ce cas nous pouvons écrire que

$$[a_{ij}] = [a_{.}] + [a_{ij} - a_{.}], \quad \text{where } [a_{ij} - a_{.}] = [a_{ij}] - [a_{.}]. \quad (12)$$

Soit  $L_1$  est le sousespace linéaire engendré par le vecteur  $[a_{.}]$ ,  $L_1 \subset \mathbf{R}^{IJ}$ . Par les calculs directes on peut montrer, en utilisant les formules (6) et (7), que les vecteurs  $[a_{.}]$  et  $[a_{ij} - a_{.}]$  sont orthogonaux, c'est-à-dire

$$[a_{.}]^T [a_{ij} - a_{.}] = 0, \quad (13)$$

et donc le vecteur  $[a_{ij} - a_{.}] \in L_{IJ-1}$  et le sousespace

$$L_{IJ-1} = \mathbf{R}^{IJ} \ominus L_1$$

est orthogonale à  $L_1$ ,

$$\mathbf{R}^{IJ} = L_1 \oplus L_{IJ-1}, \quad (14)$$

et de cette façon on a montré que la fonction  $a_{..}$  donne la meilleure (dans le sens (6)) approximation de notre fonction  $a(i, j)$  par la constante.

Maintenant on considère le second problème : quelle fonction d'une variable, par exemple  $i$ , donne la meilleure approximation pour  $[a_{ij} - a_{..}] \in L_{IJ-1}$ . On a l'identité

$$[a_{ij} - a_{..}] = [a_{i.} - a_{..}] + [a_{ij} - a_{i.}], \quad (15)$$

d'où on déduit que si nous voulons construire une approximation qui ne dépend que de  $j$ , par exemple, alors on revient de nouveau au problème précédent, car les vecteurs

$$[a_{i.} - a_{..}] = [a_{i.}] - [a_{..}] \quad \text{et} \quad [a_{ij} - a_{i.}] = [a_{ij}] - [a_{i.}] \quad (16)$$

sont orthogonaux :

$$[a_{i.} - a_{..}]^T [a_{ij} - a_{i.}] = 0. \quad (17)$$

On note que

$$[a_{i.}] = (a_{1.}, \dots, a_{1.}, a_{2.}, \dots, a_{2.}, \dots, a_{I.}, \dots, a_{I.})^T \in \mathbf{R}^{IJ}$$

and

$$[a_{.j}] = (a_{.1}, \dots, a_{.1}, a_{.2}, \dots, a_{.2}, \dots, a_{.J}, \dots, a_{.J})^T \in \mathbf{R}^{IJ}.$$

On remarque que

$$\sum_{i=1}^I (a_{i.} - a_{..}) = 0, \quad \sum_{j=1}^J (a_{.j} - a_{..}) = 0.$$

Puisque pour tout  $i$  fixé,  $i \in \mathbf{I}$ ,

$$\sum_{j=1}^J (a_{ij} - a_{i.}) = 0, \quad (18)$$

où

$$a_{i.} = \frac{1}{J} \sum_{j=1}^J a_{ij},$$

on en déduit que

$$[a_{i.} - a_{..}] \in L_{I-1} \quad \text{et} \quad [a_{ij} - a_{i.}] \in L_{IJ-I} = L_{I(J-1)}, \quad (19)$$

et que les sousespaces  $L_{I-1}$  et  $L_{I(J-1)}$  sont orthogonaux :

$$L_{I-1} \oplus L_{I(J-1)} = L_{IJ-1}, \quad (20)$$

et que

$$L_1 \oplus L_{I-1} \oplus L_{I(J-1)} = \mathbf{R}^{IJ} \quad (21)$$

Si nous avançons plus loin de la même façon on obtient sur le pas suivant l'identité

$$[a_{ij} - a_{i.}] = [a_{.j} - a_{..}] + [a_{ij} - a_{i.} - a_{.j} + a_{..}], \quad (22)$$

où

$$[a_{.j} - a_{..}] \in L_{J-1} \quad (23)$$

et

$$[a_{.j} - a_{..}]^T [a_{ij} - a_{i.} - a_{.j} + a_{..}] = 0. \quad (24)$$

Mais comme

$$[a_{ij} - a_{i.} - a_{.j} + a_{..}] \in L_{IJ-I-J+1} = L_{(I-1)(J-1)}, \quad (25)$$

de (6), (9), (14)-(17) et (19) on déduit que

$$\mathbf{R}^{IJ} = L_1 \oplus L_{I-1} \oplus L_{J-1} \oplus L_{(I-1)(J-1)}, \quad (26)$$

c'est-à-dire on a reçu la décomposition de  $\mathbf{R}^{IJ}$  en somme directe de quatre sousespaces orthogonaux, et donc la *décomposition de Fisher* n'est que la projection du vecteur des données  $[a_{ij}] \in \mathbf{R}^{IJ}$  sur ces sousespaces. De plus nous pouvons dire que la *décomposition orthogonale de Fisher*

$$[a_{ij}] = [a_{..}] + [a_{i.} - a_{..}] + [a_{.j} - a_{..}] + [a_{ij} - a_{i.} - a_{.j} + a_{..}], \quad (i = 1, \dots, I; j = 1, \dots, J)$$

ne dépend que de  $IJ$  coefficients, et non pas de  $1 + I + J + IJ$ . En plus du *Théorème de Pythagore* on obtient l'identité suivante :

$$\|[a_{ij}]\|^2 = \|[a_{..}]\|^2 + \|[a_{i.} - a_{..}]\|^2 + \|[a_{.j} - a_{..}]\|^2 + \|[a_{ij} - a_{i.} - a_{.j} + a_{..}]\|^2,$$

d'où on tire l'*identité de Fisher* :

$$\sum_{i=1}^I \sum_{j=1}^J a_{ij}^2 = IJa_{..}^2 + J \sum_{i=1}^I (a_{i.} - a_{..})^2 + I \sum_{j=1}^J (a_{.j} - a_{..})^2 + \sum_{i=1}^I \sum_{j=1}^J (a_{ij} - a_{i.} - a_{.j} + a_{..})^2.$$

On utilise ce fait pour faire un analyse de variances.

## 2.20 Modèle d'analyse des variances à 2 facteurs.

Suposons que sous l'hypothèse  $H_0$  on a le modèle de régression d'après lequel on a  $I \times J \times K$  observation sont indépendantes

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \delta_{ijk},$$

$$i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J; \quad k = 1, 2, \dots, K,$$

où  $\mu, \alpha_i, \beta_j, \gamma_{ij}$  sont des constantes inconnues, et

$$\delta_{ijk} \sim N(0, \sigma^2).$$

On note  $\mathbf{Y} = (Y_{111}, \dots, Y_{IJK})^T$  le vecteur d'observation,  $\mathbf{Y} \in \mathbf{R}^{IJK}$ . On suppose que  $I \leq J$ . Dans le cadre de ce modèle il faut estimer les paramètres suivants :

$$\mu, \quad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_I)^T, \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_J)^T, \quad \boldsymbol{\gamma} = \|\gamma_{ij}\|_{I \times J} \quad \text{and} \quad \sigma^2.$$

On note

$$\boldsymbol{\alpha} = \frac{1}{I} \sum_{i=1}^I \alpha_i, \quad \boldsymbol{\beta} = \frac{1}{J} \sum_{j=1}^J \beta_j,$$

$$\gamma_{i.} = \frac{1}{J} \sum_{j=1}^J \gamma_{ij}, \quad (j = 1, 2, \dots, J); \quad \gamma_{.j} = \frac{1}{I} \sum_{i=1}^I \gamma_{ij}, \quad (j = 1, 2, \dots, J);$$

$$\gamma_{..} = \frac{1}{IJ} \sum_{j=1}^J \sum_{i=1}^I \gamma_{ij}.$$

On suppose sans perdre la gènèralitè que

$$\alpha_i = \beta_j = \gamma_{i.} = \gamma_{.j} = \gamma_{..} = 0, \quad (i = 1, 2, \dots, I; j = 1, 2, \dots, J).$$

Pour tout  $i$  et pour tout  $j$  on note

$$X_{ij} = Y_{ij.} = \frac{1}{K} \sum_{k=1}^K Y_{ijk} = (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{Y}_{ij},$$

où

$$\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijK})^T, \quad \mathbf{1} = \mathbf{1}_K = (1, 1, \dots, 1)^T \in \mathbf{R}^K.$$

Notons

$$\mathbf{X} = (X_{11}, \dots, X_{IJ})^T, \quad \mathbf{X} \in \mathbf{R}^{IJ},$$

où  $X_{ij} = Y_{ij.}$ , ( $i = 1, 2, \dots, I; j = 1, 2, \dots, J$ ). Il est claire que sous  $H_0$

$$X_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \delta_{ij.}, \quad \delta_{ij.} \sim N\left(0, \frac{\sigma^2}{K}\right),$$

parce que

$$\mathbf{E}Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad k = 1, 2, \dots, K,$$

et donc sous  $H_0$  pour tout  $k$  fixé la fonction de vraisemblance  $p_k(\mu, \alpha, \beta, \gamma, \sigma^2)$  du vector  $(Y_{11k}, \dots, Y_{IJK})^T$  est donnée par la formule suivante :

$$p_k(\mu, \alpha, \beta, \gamma, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{IJ/2}} \exp \left\{ - \sum_{j=1}^J \sum_{i=1}^I \frac{(Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2}{2\sigma^2} \right\}.$$

Puisque sous  $H_0$  les variables  $\delta_{ijk}$  sont indépendantes et suivent la même loi normale  $N(0, \sigma^2)$  on en tire que la fonction de vraisemblance du vector d'observations  $\mathbf{Y}$  est

$$L(\mu, \alpha, \beta, \gamma, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{IJK/2}} \exp \left\{ - \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K \frac{(Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2}{2\sigma^2} \right\} =$$

$$\frac{1}{(2\pi\sigma^2)^{IJK/2}} \exp \left\{ \frac{-SC_{int}}{2\sigma^2} \right\} \exp \left\{ \frac{-K}{2\sigma^2} \sum_{j=1}^J \sum_{i=1}^I (X_{ij} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2 \right\},$$

où

$$SC_{int} = \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K (Y_{ijk} - X_{ij})^2 = \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K (Y_{ijk} - Y_{ij.})^2.$$

On voit que la statistique  $(SC_{int}, \mathbf{X})^T$  est exhaustive pour  $(\mu, \alpha, \beta, \gamma, \sigma^2)^T$ , et que  $SC_{int}$  et  $\mathbf{X} = (X_{11}, \dots, X_{IJ})^T$  sont indépendantes. Il est évident que sous  $H_0$

$$\frac{SC_{int}}{\sigma^2} = \chi_{(K-1)IJ}^2 \quad \text{et} \quad X_{ij} \sim N\left(\mu + \alpha_i + \beta_j + \gamma_{ij}; \frac{\sigma^2}{K}\right).$$

En utilisant la d ecomposition orthogonale de Fisher on a

$$X_{ij} = X_{..} + (X_{i.} - X_{..}) + (X_{.j} - X_{..}) + (X_{ij} - X_{i.} - X_{.j} + X_{..})$$

et

$$\|\mathbf{X}\|^2 = \|[X_{ij}]\|^2 = \|[X_{..}]\|^2 + \|[X_{i.} - X_{..}]\|^2 + \|[X_{.j} - X_{..}]\|^2 + \|[X_{ij} - X_{i.} - X_{.j} + X_{..}]\|^2,$$

d'o u on tire l'identit e de Fisher Pythagore :

$$\sum_{i=1}^I \sum_{j=1}^J X_{ij}^2 = IJX_{..}^2 + J \sum_{i=1}^I (X_{i.} - X_{..})^2 + I \sum_{j=1}^J (X_{.j} - X_{..})^2 + \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_{i.} - X_{.j} + X_{..})^2.$$

Maintenant nous pouvons pr esenter  $L(\mu, \alpha, \beta, \gamma, \sigma^2)$  par la fa on suivante :

$$L(\mu, \alpha, \beta, \gamma, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{IJK/2}} \exp \left\{ \frac{-SC_{int}}{2\sigma^2} \right\} \times \\ \exp \left\{ \frac{-K}{2\sigma^2} \left[ IJ(X_{..} - \mu)^2 + J \sum_{i=1}^I (X_{i.} - X_{..} - \alpha_i)^2 + I \sum_{j=1}^J (X_{.j} - X_{..} - \beta_j)^2 + \right. \right. \\ \left. \left. \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_{i.} - X_{.j} + X_{..} - \gamma_{ij})^2 \right] \right\},$$

d'o u on obtient les meilleurs estimateurs sans biais (au sens du minimum de risk quadratique) pour  $\mu, \alpha_i, \beta_j, \gamma_{ij}$  :

$$\hat{\mu} = X_{..}, \quad \hat{\alpha}_i = X_{i.} - X_{..}, \quad \hat{\beta}_j = X_{.j} - X_{..}, \quad \hat{\gamma}_{ij} = X_{ij} - X_{i.} - X_{.j} + X_{..}$$

Pour estimer  $\sigma^2$  il faut utiliser la relation

$$\frac{SC_{int}}{\sigma^2} = \chi_{(K-1)IJ}^2,$$

qui nous donne l'estimateur

$$\hat{\sigma}^2 = \frac{1}{IJ(K-1)} SC_{int} = \frac{1}{IJ(K-1)} \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K (Y_{ijk} - X_{ij})^2.$$

Comme nous avons dit nous avons construit MVUE's (voir, Voinov and Nikulin (1996)) puisque

$$\mathbf{E}\hat{\mu} = \mathbf{E}X_{..} = \mu, \quad \mathbf{E}\hat{\alpha}_i = \alpha_i, \quad \mathbf{E}\hat{\beta}_j = \beta_j, \\ \mathbf{E}\hat{\gamma}_{ij} = \gamma_{ij}, \quad \mathbf{E}\hat{\sigma}^2 = \sigma^2.$$

On consid ere ici les sommes de carr es suivantes :

$$SC_{ent\alpha} = KJ \sum_{i=1}^I (X_{i.} - X_{..})^2, \quad SC_{ent\beta} = KI \sum_{j=1}^J (X_{.j} - X_{..})^2, \\ SC_{inter} = K \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_{i.} - X_{.j} + X_{..})^2.$$

On note aussi

$$SC_{tot} = \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K (Y_{ijk} - Y_{...})^2 = \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K (Y_{ijk} - X_{..})^2,$$

où

$$Y_{...} = X_{..} = \frac{1}{IJK} \sum_{j=1}^J \sum_{i=1}^I \sum_{k=1}^K Y_{ijk}.$$

Dans ce cas on a la relation suivante entre ces sommes de carrés :

$$SC_{tot} = SC_{int} + SC_{ent\alpha} + SC_{ent\beta} + SC_{inter}.$$

On remarque que les statistiques  $SC_{int}$ ,  $SC_{ent\alpha}$ ,  $SC_{ent\beta}$ ,  $SC_{inter}$  sont indépendantes.

On considère les trois hypothèses suivantes :

$$H_{0\alpha} : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0,$$

$$H_{0\beta} : \beta_1 = \beta_2 = \dots = \beta_J = 0,$$

$$H_{0\gamma} : \gamma_1 = \gamma_2 = \dots = \gamma_{IJ} = 0.$$

On remarque que

si  $H_{0\alpha}$  est vraie, alors

$$\frac{SC_{ent\alpha}}{\sigma^2} = \chi_{I-1}^2,$$

si  $H_{0\beta}$  est vraie, alors

$$\frac{SC_{ent\beta}}{\sigma^2} = \chi_{J-1}^2,$$

si  $H_{0\gamma}$  est vraie, alors

$$\frac{SC_{inter}}{\sigma^2} = \chi_{(I-1)(J-1)}^2.$$

Pour tester  $H_{0\alpha}$  on calcule la statistique

$$\frac{IJ(K-1)SC_{ent\alpha}}{(I-1)SC_{int}} = F_{I-1, IJ(K-1)}.$$

Pour tester  $H_{0\beta}$  on calcule la statistique

$$\frac{IJ(K-1)SC_{ent\beta}}{(J-1)SC_{int}} = F_{J-1, IJ(K-1)}.$$

Pour tester  $H_{0\gamma}$  on calcule la statistique

$$\frac{IJ(K-1)SC_{inter}}{(I-1)(J-1)SC_{int}} = F_{(I-1)(J-1), IJ(K-1)}.$$

**Exemple 1. Analyse de variance à un facteur.** On suppose que l'on a mesuré les diamètres de 5 billes. Pour chaque bille on a répété 5 fois les mesures. On considère l'hypothèse  $H_0$  selon laquelle

1) les 5 valeurs (inconnues) des diamètres de ces 5 billes sont des réalisations de 5 variables aléatoires qui sont indépendantes et suivent la même loi normale  $N(a, \sigma_o^2)$ .

2) toutes les mesures sont indépendantes, de même précision et sans biais.

3) les erreurs aléatoires de ces mesures suivent la même loi normale  $N(0, \sigma^2)$ , dont la variance  $\sigma^2$  est inconnue.

La Table nous donne les résultats suivants pour des mesures (en mm). On note  $x_{ij}$  la  $j$ -ème mesure de la bille avec le numéro  $i$ , et  $x_i$  la valeur moyenne des mesures pour ce sujet.

Numéro des mesures	SUJETS				
	1	2	3	4	5
1	12.093	11.996	12.017	12.023	11.900
2	12.097	11.995	12.012	12.026	11.893
3	12.096	11.990	12.014	12.028	11.896
4	12.094	11.991	12.017	12.028	11.899
5	12.100	11.998	12.010	12.021	11.898

Table 1

Il nous faut trouver les meilleurs estimateurs sans biais des valeurs inconnues des diamètres des billes, de  $a$ ,  $\sigma_o^2$  et  $\sigma^2$ , et aussi des surfaces des sections de ces 5 billes.

Tout d'abord il faut construire la fonction de vraisemblance. Il est clair, que nous pouvons considérer le résultat  $x_{ij}$  de la  $j$ -ème mesure de la bille  $i$  comme la réalisation d'une variable aléatoire  $X_{ij}$ , où

$$X_{ij} = a + \delta_i + \delta_{ij}, \quad i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J. \quad (1)$$

Les éléments  $\delta_1, \delta_2, \dots, \delta_I$  du vecteur  $\delta = (\delta_1, \delta_2, \dots, \delta_I)^T$  et  $\delta_{ij}$  de la matrice  $\Delta = \|\delta_{ij}\|$  sont indépendants,  $\delta_i$  suit la loi normale  $N(0, \sigma_o^2)$ ,  $\delta_{ij}$  suit la loi normale  $N(0, \sigma^2)$ , les paramètres  $a$ ,  $\sigma_o^2$  et  $\sigma^2$  sont inconnus.

Notons

$$X_i = \frac{1}{J} \sum_{j=1}^J X_{ij}, \quad X_{..} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J X_{ij} = \frac{1}{I} \sum_{i=1}^I X_i, \quad (2)$$

On remarque que dans notre cas

$$\begin{aligned} X_1. &= 12.0960, \quad X_2. = 11.9920, \quad X_3. = 12.0140, \\ X_4. &= 12.0252, \quad X_5. = 11.8972, \quad X_{..} = 12.00488. \end{aligned}$$

De plus notons

$$\delta_i = \frac{1}{J} \sum_{j=1}^J \delta_{ij}, \quad \delta_{..} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \delta_{ij}, \quad \delta_{.i} = \frac{1}{I} \sum_{i=1}^I \delta_i. \quad (3)$$

Dans ce cas, l'observation  $X_{ij}$  peut-être représentée comme

$$X_{ij} = X_{..} + (X_i - X_{..}) + (X_{ij} - X_i) \quad (4)$$

et nous pouvons remarquer que

$$\begin{aligned} X_i &= a + \delta_i + \delta_{i.}, \\ X_{..} &= a + \delta_{..} + \delta_{..}, \end{aligned}$$

et que

$$\begin{aligned} X_{ij} - X_i &= \delta_{ij} - \delta_i, \\ X_i - X_{..} &= (\delta_i - \delta_{..}) + (\delta_i + \delta_{..}), \\ X_{..} - a &= \delta_{..} + \delta_{..}. \end{aligned} \quad (5)$$

Comme toutes les variables  $\delta_1, \delta_2, \dots, \delta_I, \delta_{11}, \dots, \delta_{IJ}$  sont indépendantes et normales, on a

$$\begin{aligned} (\delta_i - \delta_{..}), \quad \delta_{..}, \quad \delta_{ij} \quad &\text{sont indépendantes,} \\ (\delta_i - \delta_{..}), \quad \delta_{..}, \quad (\delta_{ij} - \delta_i), \quad \delta_i \quad &\text{sont indépendantes,} \\ (\delta_i - \delta_{..}), \quad \delta_{..}, \quad (\delta_{ij} - \delta_i), \quad (\delta_i - \delta_{..}), \quad \delta_{..} \quad &\text{sont indépendantes} \end{aligned} \quad (6)$$

et de plus la variable aléatoire

$$\delta_{..} + \delta_{..} \quad \text{suit la loi normale} \quad N\left(0, \frac{\sigma_o^2}{I} + \frac{\sigma^2}{IJ}\right). \quad (7)$$

La variable aléatoire

$$\sum_{i=1}^I [(\delta_i - \delta_{..}) + (\delta_i - \delta_{..})]^2 = \sum_{i=1}^I [(\delta_i + \delta_i) - (\delta_{..} + \delta_{..})]^2$$

est distribuée comme

$$\left(\sigma_o^2 + \frac{1}{J}\sigma^2\right)\chi_{I-1}^2,$$

c'est-à-dire que

$$\frac{1}{\left(\sigma_o^2 + \frac{1}{J}\sigma^2\right)} \sum_{i=1}^I [(\delta_i - \delta_{..}) + (\delta_i - \delta_{..})]^2 = \chi_{I-1}^2, \quad (8)$$

et il est évident, que

$$\frac{1}{\sigma^2} \sum_{i,j} (\delta_{ij} - \delta_i)^2 = \chi_{I(J-1)}^2. \quad (9)$$

Par conséquent, de (6)-(9) on déduit que

$$\frac{(\delta_{..} + \delta_{..})^2}{\frac{\sigma_o^2}{I} + \frac{\sigma^2}{IJ}} + \frac{\sum_{i=1}^I [(\delta_i - \delta_{..}) + (\delta_i - \delta_{..})]^2}{\sigma_o^2 + \frac{1}{J}\sigma^2} + \frac{1}{\sigma^2} \sum_{i,j} (\delta_{ij} - \delta_i)^2 = \chi_{IJ}^2,$$

ce qui est équivalent à

$$\frac{(X_{..} - a)^2}{\frac{1}{I}(\sigma_o^2 + \frac{\sigma^2}{J})} + \frac{\sum_{i=1}^I (X_i - X_{..})^2}{\sigma_o^2 + \frac{\sigma^2}{J}} + \frac{\sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_{..})^2}{\sigma^2} = \chi_{IJ}^2. \quad (10)$$

On trouve maintenant la fonction de vraisemblance  $L(a, \sigma_o^2, \sigma^2)$  de notre échantillon  $(X_{11}, \dots, X_{IJ})^T$ . L'expression (10) est proportionnelle, à un terme additif près, à  $\ln L(a, \sigma_o^2, \sigma^2)$ . Pour le montrer on remarque que

$$\frac{1}{\sigma_o \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{J/2} \sigma^J} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma^2} \sum_{i=1}^J (x_{ij} - a - y_i)^2 + \frac{y_i^2}{\sigma_o^2} \right] \right\} dy_i =$$

$$\begin{aligned} & \frac{1}{\sigma_o(2\pi)^{(J+1)/2}\sigma^J} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma^2} \sum_{j=1}^J (x_{ij} - a)^2 - \frac{2y_i}{\sigma^2} \sum_{j=1}^J (x_{ij} - a) + \right. \right. \\ & \left. \left. \left( \frac{1}{\sigma_o^2} + \frac{J}{\sigma^2} \right) y_i^2 \right] \right\} dy_i = \frac{1}{\sigma_o(2\pi)^{J/2}\sigma^J \left( \frac{1}{\sigma_o^2} + \frac{J}{\sigma^2} \right)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^J (x_{ij} - a)^2 \right\} \times \\ & \int_{-\infty}^{\infty} \left\{ \exp \left[ \frac{y_i}{\sigma^2} \sum_{j=1}^J (x_{ij} - a) \right] \frac{1}{\sqrt{2\pi}} \left( \frac{1}{\sigma_o^2} + \frac{J}{\sigma^2} \right)^{1/2} \exp \left[ -1/2 \left( \frac{1}{\sigma_o^2} + \frac{J}{\sigma^2} \right) y_i^2 \right] \right\} dy_i. \quad (11) \end{aligned}$$

De l'autre côté on sait que si une variable aléatoire  $\zeta$  suit la loi normale  $N(\mathbf{E}\zeta, \mathbf{Var}\zeta)$  de paramètres  $\mathbf{E}\zeta$  et  $\mathbf{Var}\zeta$ , alors

$$\mathbf{E}e^{it\zeta} = \exp\left\{it\mathbf{E}\zeta - \frac{t^2}{2}\mathbf{Var}\zeta\right\}. \quad (12)$$

représente la fonction caractéristique de  $\zeta$ . Dans notre cas

$$\begin{aligned} \zeta &= \delta. + \delta_{..} = X_{..} - a, \\ \mathbf{E}\zeta &= 0, \quad \mathbf{Var}\zeta = \frac{1}{\frac{1}{\sigma_o^2} + \frac{J}{\sigma^2}}, \quad \text{it} = \frac{1}{\sigma^2} \sum_{j=1}^J (x_{ij} - a)^2, \end{aligned} \quad (13)$$

et par conséquent de (12)-(13) on déduit que l'intégrale en (11) est égale à

$$\exp \left\{ \frac{1}{2\sigma^4 \left( \frac{1}{\sigma_o^2} + \frac{J}{\sigma^2} \right)} \left[ \sum_{i=1}^J (x_{ij} - a) \right]^2 \right\} \quad (14)$$

et donc

$$\begin{aligned} L(a, \sigma_o^2, \sigma^2) &= \frac{1}{(2\pi)^{IJ/2} \sigma^{IJ} \sigma_o^I \left( \frac{1}{\sigma_o^2} + \frac{J}{\sigma^2} \right)^{I/2}} \times \\ & \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - a)^2 + \frac{1}{2 \left( \frac{\sigma_o^4}{\sigma^2} + J\sigma^2 \right)} \sum_{i=1}^I \left[ \sum_{j=1}^J (X_{ij} - a) \right]^2 \right\}, \end{aligned} \quad (15)$$

d'où l'on tire que

$$\begin{aligned} \ln L(a, \sigma_o^2, \sigma^2) &= \ln(const) - \frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - a)^2 + \\ & \frac{1}{2 \left( \frac{\sigma_o^4}{\sigma^2} + J\sigma^2 \right)} \sum_{i=1}^I \left[ \sum_{j=1}^J (X_{ij} - a) \right]^2. \end{aligned} \quad (16)$$

Mais par ailleurs, de (2)-(5) on déduit

$$\begin{aligned} X_{ij} - a &= (X_{..} - a) + (X_i - X_{..}) + (X_{ij} - X_i), \\ \sum_{j=1}^J (X_{ij} - a) &= J(X_{..} - a) + J(X_i - X_{..}), \end{aligned}$$

$$\sum_{i=1}^I \left[ \sum_{j=1}^J (X_{ij} - a) \right]^2 = IJ^2(X_{..} - a)^2 + J^2 \sum_{i=1}^I (X_i - X_{..})^2, \quad (17)$$

$$\begin{aligned} (X_{ij} - a)^2 &= (X_{..} - a)^2 + (X_i - X_{..})^2 + (X_{ij} - X_i)^2 + \\ &2[(X_{..} - a)(X_i - X_{..}) + (X_{..} - a)(X_{ij} - X_i) + (X_i - X_{..})(X_{ij} - X_i)], \\ \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - a)^2 &= IJ(X_{..} - a)^2 + J \sum_{i=1}^I (X_i - X_{..})^2 + \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2, \end{aligned}$$

et par conséquent de (16) et (17) il résulte que

$$\begin{aligned} -2 \ln L(a, \sigma_o^2, \sigma^2) &= -2 \ln(const) + \frac{IJ}{\sigma^2} (X_{..} - a)^2 + \frac{J}{\sigma^2} \sum_{i=1}^I (X_i - X_{..})^2 + \\ &\frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2 - \frac{IJ^2}{\sigma_o^4 + \sigma^2 J} (X_{..} - a)^2 - \frac{J^2}{\sigma_o^4 + \sigma^2 J} \sum_{i=1}^I (X_i - X_{..})^2 = \\ &-2 \ln(const) + \frac{(X_{..} - a)^2}{\frac{1}{I}(\sigma_o^2 + \frac{\sigma^2}{J})} + \frac{\sum_{i=1}^I (X_i - X_{..})^2}{\sigma_o^2 + \frac{\sigma^2}{J}} + \frac{\sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2}{\sigma^2} = \\ &= -2 \ln(const) + \chi_{IJ}^2, \end{aligned} \quad (18)$$

comme on le voit à partir de (11). De (18) on déduit que

$$T = \left( X_{..}, \sum_{i=1}^I (X_i - X_{..})^2, \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2 \right)^T \quad (19)$$

est une statistique exhaustive. Il est évident que les meilleurs estimateurs sans biais pour  $a$ ,  $\sigma^2$  et  $\sigma_o^2 + \sigma^2/J$  sont

$$\hat{a} = X_{..}, \quad \hat{\sigma}^2 = \frac{1}{I(J-1)} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2, \quad (20)$$

$$\hat{\sigma}_o^2 + \frac{\hat{\sigma}^2}{J} = \frac{1}{I-1} \sum_{i=1}^I (X_i - X_{..})^2, \quad (21)$$

et par conséquent, on trouve

$$\hat{a} = 12.00488, \quad \hat{\sigma}^2 = 0.00000918, \quad \sqrt{\hat{\sigma}^2} = 0.00303, \quad (22)$$

$$\hat{\sigma}_o^2 + \frac{\hat{\sigma}^2}{J} = 0.0051400, \quad \hat{\sigma}_o^2 = 0.0051382, \quad \sqrt{\hat{\sigma}_o^2} = 0.07168. \quad (23)$$

Comme

$$\frac{\frac{(X_{..} - a)^2}{\frac{\sigma_o^2}{I} + \frac{\sigma^2}{IJ}}}{\frac{\sum_{i=1}^I (X_i - X_{..})^2}{(I-1)(\sigma_o^2 + \frac{\sigma^2}{J})}} = \frac{I(X_{..} - a)^2}{\frac{1}{I-1} \sum_{i=1}^I (X_i - X_{..})^2} = \frac{\chi_1^2}{\frac{1}{I-1} \chi_{I-1}^2} = F_{1, I-1} = t_{I-1}^2,$$

on a

$$\mathbf{P} \left\{ \sqrt{I} \frac{|X_{..} - a|}{\sqrt{\frac{1}{I-1} \sum_{i=1}^I (X_i - X_{..})^2}} \leq \sqrt{F_{1,I-1}(P)} \right\} = P, \quad (24)$$

où  $F_{1,I-1}(P)$  est le quantile de niveau  $P$  ( $P > 0.5$ ) de la distribution  $F$  à 1 et  $I - 1$  degrés de liberté, dont on rappelle la définition :

$$\mathbf{P}\{F_{1,I-1} \leq F_{1,I-1}(P)\} = P. \quad (25)$$

Par exemple, si  $P = 0.95$ , alors  $F_{1,4}(0.95) = 7.7086$ . De (24) et (25) on déduit l'intervalle de confiance

$$|a - X_{..}| \leq \sqrt{\frac{1}{I} F_{1,I-1}(P) \frac{1}{I-1} \sum_{i=1}^I (X_i - X_{..})^2} \quad (26)$$

de coefficient de confiance  $P$ . Par conséquent, on trouve l'intervalle de confiance pour le paramètre  $a$  :

$$11.9159 < a < 12.0939$$

En continuant les calculs, liés à la table 1 des données initiales, on obtient la table suivante :

i					
$\frac{1}{4} \sum_{i=1}^J (X_{ij} - X_i)^2$	$75 * 10^{-7}$	$115 * 10^{-7}$	$95 * 10^{-7}$	$97 * 10^{-7}$	$77 * 10^{-7}$

Table 2.

De plus, on a

$$\begin{aligned} \sum_{i=1}^I (X_i - X_{..})^2 &= 0.020559808, & \frac{1}{4} \sum_{i=2}^I I (X_i - X_{..})^2 &= 0.005139952, \\ \sqrt{\frac{1}{4} \sum_{i=1}^I (X_i - X_{..})^2} &= 0.071693458557946, & \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2 &= 0.0001836, \\ \frac{1}{20} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2 &= 0.00000918, & \sqrt{\frac{1}{20} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2} &= 0.003029851481508, \\ \frac{1}{4} \sum_{i=1}^I (X_i - X_{..})^2 - \frac{1}{100} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2 &= 0.005138116, \\ \sqrt{\frac{1}{4} \sum_{i=1}^I (X_i - X_{..})^2 - \frac{1}{100} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2} &= 0.071680652898814, \\ \frac{\frac{1}{4} \sum_{i=1}^I (X_i - X_{..})^2}{\frac{1}{20} \sum_{i=1}^I \sum_{j=1}^J (X_{ij} - X_i)^2} &= 559.9076252723311. \end{aligned}$$

La surface de section de la bille de numéro  $i$  est égale à

$$\frac{\pi}{4}(a + \delta_i)^2,$$

et l'espérance de la section de n'importe quelle bille est égale à

$$\frac{\pi}{4}\mathbf{E}(a + \delta_i)^2 = \frac{\pi}{4}(a^2 + \mathbf{E}\delta_i^2) = \frac{\pi}{4}(a^2 + \sigma_o^2),$$

car  $\mathbf{E}\delta_i = 0$  et  $\mathbf{Var}\delta_i = \mathbf{E}\delta_i^2$ . Mais comme

$$\mathbf{E}X_{..}^2 = \mathbf{Var}X_{..} + (\mathbf{E}X_{..})^2 = \frac{\sigma_o^2}{I} + \frac{\sigma^2}{IJ} + a^2,$$

on obtient l'estimateur sans biais de  $\frac{\pi}{4}(a^2 + \sigma_o^2)$  :

$$\frac{\pi}{4} \left( X_{..}^2 - \frac{\hat{\sigma}_o^2}{I} - \frac{\hat{\sigma}^2}{IJ} - \hat{\sigma}_o^2 \right) =$$

$$\frac{\pi}{4} \left[ (12.00488)^2 + 0.00514 - \frac{0.00514}{5} \right] = 113.1926.$$

Par ailleurs, on a

$$a + \delta_i \cong X_i, \quad \mathbf{E}\{X_i | \delta_i\} = a + \delta_i$$

et donc

$$\mathbf{Var}\{X_i | \delta_i\} = \frac{\sigma^2}{J}, \quad \mathbf{E}\{X_i^2 | \delta_i\} = \mathbf{Var}\{X_i | \delta_i\} + (a + \delta_i)^2,$$

et donc l'estimateur sans biais pour la surface de la section de la bille de numéro  $i$  est

$$\frac{\pi}{4} \left( X_i^2 - \frac{\hat{\sigma}^2}{J} \right) = \frac{\pi}{4} (X_i^2 - 0.0000018).$$

Les valeurs numériques de ces estimateurs pour les billes de numéros 1,2,3,4,5 sont

$$114.91413, 112.32974, 113.36138, 113.57284, 111.16790$$

respectivement (la moyenne arithmétique est égale à 113.06920). Enfin, on remarque que

$$\sigma^2 < 0.007926\sigma_o^2 \quad \text{et} \quad \sigma^2 > 0.00127$$

avec le coefficient de confiance  $P = 0.98$ .

## 2.21 Modèle exponentiel. Analyse statistique.

Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon d'une loi exponentielle  $\mathcal{E}(\mu, \sigma)$ , i.e.

$$X_i \sim f(x; \theta), \quad \theta \in \Theta = \{\theta = (\mu, \sigma)^T : |\mu| < \infty, \sigma > 0\},$$

où

$$f(x; \theta) = \begin{cases} \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right), & x \geq \mu, \\ 0, & \text{sinon.} \end{cases}$$

Il est évident que

$$f(x; \theta) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) H(x-\mu), \quad (2.1)$$

$$H(x) = \begin{cases} 1, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases}$$

On sait que

$$\mathbf{E}X_i = \mu + \sigma \quad \text{et} \quad \mathbf{Var}X_i = \sigma^2. \quad (2.2)$$

Notons  $X^{(n)} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})^T$  le vecteur des statistiques d'ordre associé au vecteur de observation  $\mathbb{X}$ ,

$$\mathbf{P}\{X_{(1)} < X_{(2)} < \dots < X_{(n)}\} = 1. \quad (2.3)$$

Il est facile de montrer que  $T = (X_{(1)}, S)^T$  est une statistique exhaustive pour le paramètre  $\theta$ , où

$$X_{(1)} = \min(X_1, X_2, \dots, X_n) \quad \text{et} \quad S = \sum_{i=2}^n (X_{(i)} - X_{(1)}). \quad (2.4)$$

En effet, la fonction de vraisemblance de  $\mathbb{X}$  est

$$\begin{aligned} L(\mathbb{X}; \theta) &= \prod_{i=1}^n f(X_i; \theta) = \frac{1}{\sigma^n} \exp\left\{-\frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu)\right\} H(X_{(1)} - \mu) = \\ &= \frac{1}{\sigma^n} \exp\left\{-\frac{1}{\sigma} \sum_{i=1}^n (X_{(i)} - \mu)\right\} H(X_{(1)} - \mu). \end{aligned} \quad (2.5)$$

Comme

$$\sum_{i=1}^n X_i = \sum_{i=1}^n X_{(i)} = \sum_{i=2}^n (X_{(i)} - X_{(1)}) + nX_{(1)} = \sum_{i=2}^n (X_{(i)} - X_{(1)}) + nX_{(1)},$$

on en tire que la statistique  $\mathbf{T} = (X_{(1)}, S)^T$  est exhaustive minimale pour  $\theta = (\mu, \sigma)^T$ . Il est connu que  $X^{(n)}$  est une statistique exhaustive pour  $\theta$ , mais  $X^{(n)}$  n'est pas intéressante parce qu'elle a la même dimension  $n$  que le vecteur  $\mathbb{X}$ , c'est-à-dire que  $X^{(n)}$  ne réduit pas des données. Le vecteur

$$U = (X_{(1)}, \sum_{i=2}^n X_{(i)})^T$$

est aussi une statistique exhaustive minimale pour  $\theta$ . Il est facile de montrer que la densité de  $X_{(1)}$  est donnée par la formule

$$\frac{n}{\sigma} \exp\left\{-\frac{n}{\sigma}(x_{(1)} - \mu)\right\} H(x_{(1)} - \mu), \quad (2.6)$$

i.e.,  $X_{(1)}$  suit une loi exponentielle  $\mathcal{E}(\mu, \sigma/n)$ ,

$$\mathbf{E}X_{(1)} = \mu + \frac{\sigma}{n} \quad \text{et} \quad \mathbf{Var}X_{(1)} = \frac{\sigma^2}{n^2}. \quad (2.7)$$

Nous pouvons donc dire que la statistique  $nX_{(1)} \sim \mathcal{E}(n\mu, \sigma)$ , et de (2) et (7) on obtient que

$$\mathbf{E}\{nX_{(1)}\} = n\mu + \sigma \quad \text{et} \quad \mathbf{Var}\{nX_{(1)}\} = \sigma^2. \quad (2.8)$$

Maintenant nous allons montrer que  $X_{(1)}$  et  $S$  sont indépendantes. Tout d'abord on remarque que la densité de  $\mathbb{X}^{(\cdot)}$  est

$$\begin{aligned} g(x^{(\cdot)}; \theta) &= n! \prod_{i=1}^n f(x_{(i)}; \theta) = \frac{n!}{\sigma^n} \exp \left\{ -\frac{1}{\sigma} \sum_{i=1}^n (x_{(i)} - \mu) \right\} H(x_{(1)} - \mu) = \\ &= \frac{n}{\sigma} \exp \left\{ -\frac{n}{\sigma} (x_{(1)} - \mu) \right\} H(x_{(1)} - \mu) \frac{(n-1)!}{\sigma^{n-1}} \times \\ &\quad \times \exp \left\{ -\frac{1}{\sigma} \sum_{i=2}^n (x_{(i)} - x_{(1)}) \right\} H(x_{(2)} - x_{(1)}), \end{aligned} \quad (2.9)$$

où

$$x^{(\cdot)} = (x_{(1)}, \dots, x_{(n)})^T \in B_\mu = \{x \in \mathbb{R}^n : \mu \leq x_1 \leq x_2 \leq \dots \leq x_n\}, \quad (2.10)$$

d'où on tire que

$$\frac{(n-1)!}{\sigma^{n-1}} \exp \left\{ -\frac{1}{\sigma} \sum_{i=2}^n (x_{(i)} - x_{(1)}) \right\}, \quad x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}, \quad (2.11)$$

représente la densité conditionnelle de

$$(X_{(2)}, X_{(3)}, \dots, X_{(n)})^T \quad \text{sachant que} \quad X_{(1)} = x_{(1)}.$$

On constate que cette loi conditionnelle ne dépend pas de  $\mu$ . En plus de (4) et (9) on déduit que si la valeur  $x_{(1)}$  de la statistique  $X_{(1)}$  est fixée,  $X_{(1)} = x_{(1)}$ , alors la statistique  $(X_{(2)}, X_{(3)}, \dots, X_{(n)})^T$  représente le vecteur des statistiques d'ordre obtenu à partir d'un échantillon de dimension  $n-1$ , dont les éléments suivent la loi exponentielle

$$\frac{1}{\sigma} \exp \left\{ -\frac{x - x_{(1)}}{\sigma} \right\} H(x - x_{(1)}).$$

Maintenant on va chercher la densité conjointe  $q(y; \theta)$ ,

$$y = (y_1, \dots, y_n)^T \in B_\mu = \{x \in \mathbb{R}^n : \mu \leq y_1, 0 \leq y_2 \leq \dots \leq y_n\},$$

des statistiques

$$X_{(1)} \quad \text{et} \quad (X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(1)})^T,$$

c'est-à-dire la densité de la statistique

$$Y = (Y_1, Y_2, \dots, Y_n)^T,$$

où

$$Y_1 = X_{(1)}, \quad Y_j = X_{(j)} - X_{(1)}, \quad j = 2, \dots, n. \quad (2.12)$$

On constate que la statistique  $Y$  est le résultat d'une transformation linéaire la statistique  $X^{(n)}$  :

$$Y = BX^{(n)},$$

où

$$B = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ -1 & 0 & 0 & \dots & 1 \end{vmatrix},$$

et donc

$$X^{(n)} = B^{-1}Y,$$

où

$$B^{-1} = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \dots & 1 \end{vmatrix}.$$

Comme  $\det B = 1$ , de (9) on tire

$$\begin{aligned} q(y; \theta) &= g(B^{-1}y; \theta) |\det B^{-1}| = g(y_1, y_1 + y_2, \dots, y_1 + y_n; \theta) = \\ &= \frac{n}{\sigma} \exp \left\{ -\frac{n}{\sigma} (y_1 - \mu) \right\} H(y_1 - \mu) \frac{(n-1)!}{\sigma^{n-1}} \left\{ -\frac{1}{\sigma} \sum_{i=2}^n y_i \right\}, \quad y \in B_\mu \subset \mathbb{R}^n, \end{aligned} \quad (2.13)$$

d'où on tire que la densité conjointe de  $X_{(1)}$  et  $(X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(1)})^T$  est le produit de deux densités et donc les statistiques  $X_{(1)}$  et  $(X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(1)})^T$  sont indépendantes, d'où on tire que  $X_{(1)}$  et  $\sum_{i=2}^n (X_{(i)} - X_{(1)})$  sont indépendantes.

En plus de (13) il suit que

$$\sum_{i=2}^n (X_{(i)} - X_{(1)})$$

suit une loi gamma dont la densité est

$$\frac{1}{\sigma^{n-1} \Gamma(n-1)} y^{n-2} e^{-y/\sigma} H(y),$$

parce que

$$\frac{(n-1)!}{\sigma^{n-1}} \exp \left\{ -\frac{1}{\sigma} \sum_{i=2}^n y_i \right\}, \quad 0 \leq y_2 \leq y_3 \leq \dots \leq y_n,$$

représente la densité conjointe du vecteur des statistiques d'ordre de dimension  $(n-1)$ , associé avec une loi exponentielle

$$\frac{1}{\sigma} \exp \left\{ -\frac{1}{\sigma} y \right\} H(y),$$

i.e. avec une loi exponentielle  $\mathcal{E}(0, \sigma)$ , et donc la variable aléatoire

$$\frac{1}{\sigma} \sum_{i=2}^n Y_i = \frac{1}{\sigma} \sum_{i=2}^n (X_{(i)} - X_{(1)}) = \gamma_{n-1}$$

est distribuée comme la somme de  $(n - 1)$  variables aléatoires indépendantes, qui forment un échantillon de volume  $(n - 1)$  d'une loi exponentielle  $\mathcal{E}(0, 1)$ , i.e.,  $S$  suit une loi gamma avec  $(n - 1)$  degrés de liberté et de paramètre d'échelle  $\sigma$ .

$$S = \sum_{i=2}^n Y_i = \sum_{i=2}^n (X_{(i)} - X_{(1)}) = \sigma \gamma_{n-1}, \quad (2.14)$$

et donc

$$\mathbf{E}S = \mathbf{E}\{\sigma \gamma_{n-1}\} = (n - 1)\sigma, \quad \mathbf{Var} S = \mathbf{Var}\{\sigma \gamma_{n-1}\} = \sigma^2(n - 1). \quad (2.15)$$

Dans ce cas la statistique

$$\bar{\sigma}_n = \frac{1}{n-1} \sum_{i=2}^n (X_{(i)} - X_{(1)}) = \frac{n}{n-1} (\bar{X}_n - X_{(1)}) \quad (2.16)$$

est le meilleur estimateur sans biais pour  $\sigma$ . De (15) on tire que

$$\mathbf{Var} \bar{\sigma}_n = \frac{\sigma^2}{n-1}. \quad (2.17)$$

Enfin, en utilisant (7) et (16) nous pouvons construire le meilleur estimateur sans biais  $\bar{\mu}_n$  pour  $\mu$  :

$$\bar{\mu}_n = X_{(1)} - \frac{\bar{\sigma}_n}{n} = X_{(1)} - \frac{1}{n(n-1)} \sum_{i=2}^n (X_{(i)} - X_{(1)}) = X_{(1)} - \frac{1}{n-1} (\bar{X}_n - X_{(1)}). \quad (2.18)$$

Comme les statistique  $X_{(1)}$  et  $S$  sont indépendantes, les statistiques  $X_{(1)}$  et  $\bar{\sigma}_n$  sont aussi indépendantes et par conséquent

$$\mathbf{Var} \bar{\mu}_n = \mathbf{Var} X_{(1)} + \frac{1}{n^2} \mathbf{Var} \bar{\sigma}_n = \frac{\sigma^2}{n^2} + \frac{\sigma^2}{(n-1)n^2} = \frac{\sigma^2}{n(n-1)}. \quad (2.19)$$

**Corollaire 1.** Comme

$$\sum_{i=2}^n (X_{(i)} - X_{(1)}) = \sum_{i=2}^n Y_i = \sum_{i=2}^n (n-i-1)[X_{(i)} - X_{(i-1)}], \quad (2.20)$$

de (9) et (12) il suit que les statistiques

$$nX_{(1)}, (n-1)[X_{(2)} - X_{(1)}], \dots, (n-i-1)[X_{(i)} - X_{(i-1)}], \dots, X_{(n)} - X_{(n-1)}$$

sont indépendantes et

$$nX_{(1)} \sim \mathcal{E}(n\mu, \sigma), \quad \text{i.e.} \quad n(X_{(1)} - \mu) \approx \mathcal{E}(0, \sigma), \quad (2.21)$$

$$(n-i-1)[X_{(i)} - X_{(i-1)}] \sim \mathcal{E}(0, \sigma), \quad i = 2, 3, \dots, n. \quad (2.22)$$

Il est évident que toutes ces propriétés importantes d'un échantillon  $\mathbb{X}$  d'une loi exponentielle sont dues à l'indépendance temporelle de la distribution exponentielle (une loi exponentielle est sans mémoire).

**Remarque 1.** (Méthode des moments). Comme

$$\mathbf{E}X_i = \mu + \sigma \quad \text{et} \quad \mathbf{Var} X_i = \sigma^2$$

pour estimer  $\theta = (\mu, \sigma)^T$  nous pouvons utiliser la méthode des moments. D'après cette méthode en qualité d'estimateurs  $\tilde{\mu}_n$  et  $\tilde{\sigma}_n$  de  $\mu$  et  $\sigma$  il faut choisir la solution du système

$$\begin{cases} \mu + \sigma = \bar{X}_n, \\ \sigma^2 = s_n^2, \end{cases}$$

puisque  $\bar{X}_n$  et  $s_n^2$  sont l'espérance et la variance de la loi empirique correspondant à l'échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$ , d'où on obtient que

$$\tilde{\theta}_n = (\tilde{\mu}_n, \tilde{\sigma}_n)^T,$$

où

$$\begin{aligned} \tilde{\mu}_n &= \bar{X}_n - s_n = \bar{X}_n - \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}, \\ \tilde{\sigma}_n &= s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}. \end{aligned}$$

**Remarque 2.** (Méthode du maximum de vraisemblance). De (5) on a

$$L(\mathbb{X}; \theta) = L(\mathbb{X}; \mu, \sigma) = \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu) \right\} H(X_{(1)} - \mu),$$

d'où on tire immédiatement que

$$\hat{\mu}_n = X_{(1)}.$$

Puisque

$$\frac{\partial \ln L(\mathbb{X}; \theta)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu),$$

on en tire que  $\hat{\sigma}_n$  est la solution de l'équation

$$-\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n) = 0,$$

i.e.,

$$\hat{\sigma}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n) = \bar{X}_n - X_{(1)},$$

et donc

$$\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n)^T.$$

On remarque que les meilleurs estimateurs sans biais pour  $\mu$  et  $\sigma$  sont

$$\bar{\mu}_n = \frac{n}{n-1} \left( \hat{\mu}_n - \frac{1}{n} \bar{X}_n \right) \quad \text{et} \quad \bar{\sigma}_n = \frac{n-1}{n} \hat{\sigma}_n.$$

# Chapitre 3

## ELEMENTS DE LA STATISTIQUE NON PARAMETRIQUE.

### 3.1 La loi empirique.

Soit l'hypothèse  $H_0$  selon laquelle les éléments  $X_1, X_2, \dots, X_n$  de l'échantillon  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  suivent une loi donnée, dont la fonction de répartition est  $F(x)$ , i.e. pour tout  $x \in \mathbf{R}^1$

$$\mathbf{P}\{X_i \leq x | H_0\} = F(x), \quad i = 1, 2, \dots, n, \quad (3.1)$$

et  $X_1, X_2, \dots, X_n$  sont indépendantes. Supposons en outre que la fonction de répartition  $F(x)$ ,  $x \in \mathbf{R}^1$ , soit telle que le moment  $a_{2k}$  existe,

$$a_{2k} = \mathbf{E}X_1^{2k} = \int_{-\infty}^{+\infty} x^{2k} dF(x). \quad (3.2)$$

On sait que dans ce cas tous les moments  $a_j$ ,  $1 \leq j \leq 2k$ , existent ainsi que les moments centraux  $m_j$ ,

$$m_j = \mathbf{E}(X_1 - \mathbf{E}X_1)^j = \mathbf{E}(X_1 - a)^j, \quad j = 1, 2, \dots, 2k, \quad (3.3)$$

où  $a = a_1 = \mathbf{E}X_1$ . Notons aussi

$$\sigma^2 = \mathbf{Var}X_1 = m_2 = \mathbf{E}(X_1 - a)^2. \quad (3.4)$$

Ayant la réalisation  $\mathbf{x} = (x_1, \dots, x_n)^T$  de la statistique  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ , nous pouvons construire la fonction

$$F_n(x) = F_n(x; x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(x_i), \quad x \in \mathbf{R}^1, \quad (3.5)$$

dont la valeur  $F_n(x)$  en n'importe quel point  $x$ ,  $x \in \mathbf{R}^1$ , représente la réalisation de la statistique

$$\mathbb{F}_n(x) = \mathbb{F}_n(x; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(X_i), \quad (3.6)$$

calculée au point choisi  $x$ .

Par construction, la fonction  $F_n(x)$ ,  $x \in \mathbf{R}^1$ , a toutes les propriétés d'une fonction de répartition, car elle est croissante de 0 à 1 et continue à droite, et pour cette raison nous pouvons introduire une variable aléatoire, disons  $X$ , dont la loi conditionnelle, conditionnée par  $\mathbb{X} = \mathbf{x}$ , est donnée par la fonction  $F_n(x)$ , c'est-à-dire

$$\mathbf{P}\{X \leq x | \mathbb{X} = \mathbf{x}\} = \mathbf{P}\{X \leq x | X_1 = x_1, \dots, X_n = x_n\} = F_n(x), \quad x \in \mathbf{R}^1, \quad (3.7)$$

et par conséquent de (6) et (7) il résulte que

$$\mathbb{F}_n(x) = \mathbf{P}\{X \leq x | \mathbb{X}\}, \quad x \in \mathbf{R}^1, \quad (3.8)$$

c'est-à-dire que (8) détermine une fonction de répartition *aléatoire*, qu'on appelle *fonction de répartition empirique*. Par conséquent, la loi conditionnelle de la variable aléatoire  $X$ , conditionnée par  $\mathbb{X}$ , s'appelle *la loi empirique*. De (5)–(8) il résulte que la loi empirique est la loi discrète d'après laquelle

$$\mathbf{P}\{X = X_i | \mathbb{X}\} = \frac{1}{n} \quad \text{pour tout } i = 1, 2, \dots, n, \quad (3.9)$$

c'est-à-dire que la loi empirique affecte le même poids  $1/n$  à chaque élément  $X_i$  de l'échantillon  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ , et  $\mathbb{F}_n(x)$  est la fonction de répartition de cette loi. Soit  $\alpha_m$  le moment d'ordre  $m$  de la loi empirique. Alors de (6), (8) et (9) on déduit

$$\alpha_m = \mathbf{E}\{X^m | \mathbb{X}\} = \frac{1}{n} \sum_{i=1}^n X_i^m, \quad (3.10)$$

et, par conséquent, on obtient la moyenne  $\alpha_1$  de la loi empirique :

$$\alpha_1 = \mathbf{E}\{X | \mathbb{X}\} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n. \quad (3.11)$$

De même, la variance de la loi empirique s'exprime par la formule

$$\mathbf{E}\{(X - \alpha_1)^2 | \mathbb{X}\} = \mathbf{E}\{(X - \bar{X}_n)^2 | \mathbb{X}\} = \alpha_2 - \alpha_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = s_n^2. \quad (3.12)$$

La loi empirique (9) et sa fonction de répartition  $\mathbb{F}_n(x)$ ,  $x \in \mathbf{R}^1$ , jouent un rôle très important dans la statistique mathématique ; c'est pour cela que nous allons parler un peu plus en détail de ses propriétés et qualités.

Premièrement, on remarque que pour tout  $x$  fixé,  $x \in \mathbf{R}^1$ ,

$$\mathbf{E}\mathbf{1}_{]-\infty, x]}(X_i) = \mathbf{P}\{X_i \leq x\} = F(x), \quad (3.13)$$

c'est-à-dire que la statistique  $\mathbf{1}_{]-\infty, x]}(X_i)$  est un estimateur sans biais de  $F(x)$ . On remarque ici que  $\mathbf{1}_{]-\infty, x]}(X_i)$  est la fonction de répartition empirique construite avec une seule observation  $X_i$ . Il est facile de vérifier que

$$\mathbf{Var} \mathbf{1}_{]-\infty, x]}(X_i) = F(x)[1 - F(x)], \quad (3.14)$$

car pour tout  $x$  fixé la statistique  $\mathbf{1}_{]-\infty, x]}(X_i)$  représente la variable aléatoire de Bernoulli de paramètre  $p = F(x)$ , puisque

$$\begin{cases} \mathbf{P}\{\mathbf{1}_{]-\infty, x]}(X_i) = 1\} = \mathbf{P}\{X_i \leq x\} = F(x) = p, \\ \mathbf{P}\{\mathbf{1}_{]-\infty, x]}(X_i) = 0\} = \mathbf{P}\{X_i > x\} = 1 - F(x) = 1 - p = q. \end{cases} \quad (3.15)$$

D'autre part nous avons

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i) = \frac{1}{n} \mathbf{v}_n(x), \quad (3.16)$$

où

$$\mathbf{v}_n(x) = \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i). \quad (3.17)$$

Comme les variables aléatoires  $X_1, X_2, \dots, X_n$  sont indépendantes et suivent la même loi  $F(x)$ , i.e.  $\mathbf{P}\{X_i \leq x\} = F(x)$ , de (13)-(17) il s'ensuit que pour tout  $x$  fixé

$$\mathbf{P}\{\mathbf{v}_n(x) = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n, \quad (3.18)$$

où  $p = F(x)$ . Comme

$$\mathbf{E}\mathbf{v}_n(x) = np = nF(x), \quad \mathbf{Var} \mathbf{v}_n(x) = npq = nF(x)[1 - F(x)], \quad (3.19)$$

on a

$$\mathbf{E}\mathbb{F}_n(x) = F(x) \quad \text{et} \quad \mathbf{Var} \mathbb{F}_n(x) = \frac{1}{n} F(x)[1 - F(x)]. \quad (3.20)$$

De (20) il déduit que si  $n \rightarrow \infty$

$$\mathbf{Var} \mathbb{F}_n(x) \rightarrow 0$$

pour tout  $x$  fixé,  $x \in \mathbf{R}^1$ ; par conséquent, de l'inégalité de Tchebyshev, il résulte que pour tout  $\varepsilon > 0$

$$\mathbf{P}\{|\mathbb{F}_n(x) - F(x)| \geq \varepsilon\} \leq \frac{\mathbf{Var} \mathbb{F}_n(x)}{\varepsilon^2} = \frac{F(x)[1 - F(x)]}{\varepsilon^2} \rightarrow 0, \quad (3.21)$$

quand  $n \rightarrow \infty$ . Ainsi de (20) et (21) résulte le

**Théorème 1.** Si  $\mathbf{P}\{X_i \leq x\} = F(x)$ , alors

$$\begin{cases} 1) \quad \mathbf{E}\mathbb{F}_n(x) = F(x), \\ 2) \quad \mathbf{P}\{|\mathbb{F}_n(x) - F(x)| > \varepsilon\} \rightarrow 0, \quad \text{quand } n \rightarrow \infty, \end{cases} \quad (3.22)$$

quel que soit  $x$  fixé,  $x \in \mathbf{R}^1$ .

**Remarque 1.** Le théorème 1 nous dit que  $\{\mathbb{F}_n(x)\}_{n \in \mathbf{N}}$  est une suite cohérente d'estimateurs sans biais de  $F(x)$  pour tout  $x$  fixé,  $x \in \mathbf{R}^1$ . Cela signifie que si la taille  $n$  de

l'échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$  est grande, alors la valeur de la fonction  $F_n(x)$  en un point  $x$  la réalisation de la fonction de répartition empirique  $\mathbb{F}_n(x)$  en ce point, peut-être considérée comme une bonne approximation de la valeur  $F(x)$ . Cela veut dire que si  $F$  est inconnue, on pourra supposer que

$$F_n(x) \cong F(x) \quad (3.23)$$

pour tout  $x$  et cette approximation est d'autant meilleure que le nombre  $n$  des observations, c'est-à-dire notre information sur  $F$ , est plus grand.

**Remarque 2.** Du théorème 1 il résulte que

$$\mathbf{P}\{X \leq x\} = \mathbf{E}\mathbf{P}\{X \leq x | \mathbb{X}\} = \mathbf{E}\mathbb{F}_n(x) = F(x),$$

c'est-à-dire que la loi déconditionnée de la variable aléatoire  $X$  est la même que celle de  $X_i$ , élément de l'échantillon  $\mathbb{X}$ ,

$$\mathbf{P}\{X \leq x\} = \mathbf{P}\{X_i \leq x\} = F(x).$$

Le théorème 1 peut-être affiné en considérant la fonction de répartition empirique  $\mathbb{F}_n(x)$ ,  $x \in \mathbf{R}^1$ , dans son ensemble et non pas pour chaque  $x$  pris séparément. On va s'intéresser au maximum de l'écart entre  $\mathbb{F}_n(x)$  et  $F(x)$ , que l'on notera  $\mathbf{D}_n$  :

$$\mathbf{D}_n = \mathbf{D}_n(\mathbb{X}) = \sup_{|x| < \infty} |\mathbb{F}_n(x) - F(x)|. \quad (3.24)$$

La statistique  $\mathbf{D}_n$  s'appelle *la statistique de Kolmogorov* (1933).

**Théorème 2. (Glivenko-Cantelli)**

$$\mathbf{P}\left\{\lim_{n \rightarrow \infty} \mathbf{D}_n = 0\right\} = 1. \quad (3.25)$$

Le théorème de Glivenko-Cantelli nous dit que la suite  $\{\mathbb{F}_n(x)\}$  des fonctions de répartition empiriques converge presque sûrement vers  $F(x)$  *uniformément* par rapport à  $x$  quand  $n \rightarrow \infty$ . La réalisation

$$D_n = \sup_{|x| < \infty} |\mathbb{F}_n(x) - F(x)|$$

de la statistique de Kolmogorov  $\mathbf{D}_n$  nous donne la déviation maximale observée sur l'axe réel de la fonction de répartition empirique  $\mathbb{F}_n(x)$  et de la fonction de répartition  $F(x)$  de la variable aléatoire  $X_1$ . Du théorème de Glivenko-Cantelli il résulte que pour tout  $x$ , avec la probabilité 1, cette déviation devient plus petite que tout nombre positif  $\varepsilon$  arbitrairement petit, ce qui justifie encore une fois l'approximation (23).

**Théorème 3. (Donsker)** Si  $n \rightarrow \infty$ , alors

$$\sqrt{n}(\mathbb{F}_n(x) - F(x)) \xrightarrow{\mathcal{L}} W(x), \quad x \in \mathbf{R}^1,$$

où  $W(x)$  est un processus gaussien,  $\mathbf{E}W(x) \equiv 0$ , dont la fonction de covariance est

$$k(x, y) = F(x) \wedge F(y) - F(x)F(y), \quad (x, y) \in \mathbf{R}^1 \times \mathbf{R}^1.$$

**Théorème 4.** Si  $F(x)$  est continue, alors

$$\mathbf{P}\left\{\limsup_{n \rightarrow \infty} \left( \sqrt{\frac{2n}{\ln \ln n}} \sup_x |\mathbb{F}_n(x) - F(x)| \right) = 1\right\} = 1.$$

**Remarque 3.** Pour avoir une idée de la conduite de  $F(x)$  on construit souvent le graphe de la fonction  $F_n(x)$ , réalisation de la fonction de répartition empirique  $\mathbb{F}_n(x)$ . Pour construire le graphe de  $F_n(x)$  on utilise le vecteur

$$X^{(\cdot)} = (X_{(1)}, \dots, X_{(n)})^T$$

des statistiques d'ordre, construit à partir de l'échantillon  $\mathbb{X}$ . Soit  $x^{(\cdot)} = (x_{(1)}, \dots, x_{(n)})^T$ , la réalisation de la statistique  $X^{(\cdot)}$ . Comme on le sait le vecteur  $x^{(\cdot)}$  s'obtient à partir de  $x = (x_1, \dots, x_n)^T$  en ordonnant les  $x_i$  par ordre croissant, c'est-à-dire que l'on a

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}. \quad (3.26)$$

De (26) il résulte que les statistiques d'ordre  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  sont liées (avec la probabilité 1) par les inégalités :

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}. \quad (3.27)$$

Supposons pour l'instant qu'il n'y ait pas d'ex-aequo, ce qui a lieu avec la probabilité 1 si  $F$  n'a pas de saut. En utilisant (26), (27) de (5), (6) et (9) on obtient que

$$\mathbb{F}_n(x) = \begin{cases} 0, & \text{si } x < X_{(1)}, \\ \frac{i}{n}, & \text{si } X_{(i)} \leq x < X_{(i+1)}, \\ 1, & \text{si } x \geq X_{(n)}, \end{cases} \quad (3.28)$$

par conséquent on a

$$F_n(x) = \begin{cases} 0, & \text{si } x < x_{(1)}, \\ \frac{i}{n}, & \text{si } x_{(i)} \leq x < x_{(i+1)}, \\ 1, & \text{si } x \geq x_{(n)}. \end{cases} \quad (3.29)$$

De (29) on déduit que  $F_{(n)}(x)$  a des sauts aux points  $x_{(i)}$ . Ces sauts sont égaux à  $1/n$ . Dans le cas général,  $F$  peut avoir des sauts et donc, parmi les  $x_{(i)}$ , il peut y avoir des ex-aequo. Pour construire le graphe de  $F_n(x)$ , notons

$$\begin{cases} e_1 = x_{(1)} = \min\{x_1, x_2, \dots, x_n\}, \\ e_2 = \min\{x_{(i)} : x_{(i)} > x_{(1)} = e_1\}, \\ \vdots \\ e_j = \min\{x_{(i)} : x_{(i)} > e_{j-1}\}, \\ \vdots \\ e_k = x_{(n)} = \max\{x_1, x_2, \dots, x_n\} \end{cases} \quad (3.30)$$

les différentes valeurs prises par les  $x_i$ . Le nombre  $k$  des différentes valeurs

$$e_1 < e_2 < \dots < e_k, \quad (3.31)$$

prises par  $x_1, x_2, \dots, x_n$ , peut être strictement inférieur à  $n$  s'il y a des ex-aequo. Notons  $v_j$  la fréquence de la valeur  $e_j$ ,  $j = 1, 2, \dots, k$ . Il est évident que

$$v_1 + v_2 + \dots + v_k = n.$$

En utilisant les valeurs observées  $e_1, e_2, \dots, e_k$  et leurs fréquences  $v_1, v_2, \dots, v_k$  on peut facilement obtenir une autre représentation de la réalisation  $F_n(x)$  de la fonction de répartition empirique  $\mathbb{F}_n(x)$  en termes des fréquences  $v_1, v_2, \dots, v_k$  des valeurs  $e_1, e_2, \dots, e_k$  :

$$F_n(x) = \begin{cases} 0, & \text{si } x < e_1, \\ \frac{1}{n} \sum_{j=1}^i v_j, & \text{si } e_j \leq x < e_{i+1}, \\ 1, & \text{si } x \geq e_k. \end{cases} \quad (3.32)$$

La fonction  $F_n(x)$  est aussi appelé *la fonction cumulative*, parce que on "accumule" les fréquences  $v_1, v_2, \dots, v_k$  en partant de la plus petite valeur  $e_1 = x_{(1)}$  vers la plus grande  $e_k = x_{(n)}$ .

On voit que la fonction cumulative  $F_n(x)$  est croissante de 0 à 1, qu'elle est continue à droite et qu'elle a des sauts de hauteurs  $v_i/n$  en tout point  $e_i, i = 1, 2, \dots, k$ , tout en restant constante entre deux valeurs observées  $e_i$  et  $e_{i+1}$  consécutives.

**Remarque 4. (Loi empirique et méthode des moments)** Maintenant que nous savons que la fonction de répartition  $\mathbb{F}_n(x)$  de la loi empirique est un bon estimateur de la fonction de répartition  $F(x)$  de  $X_i$  au sens des théorèmes 1 et 2, il est très naturel de choisir les moments (10)

$$\alpha_m = \mathbf{E}\{X^m | \mathbb{X}\} = \frac{1}{n} \sum_{i=1}^n X_i^m, \quad m = 1, 2, \dots, 2k$$

de la loi empirique (9) comme estimateurs des moments  $a_m = \mathbf{E}X_1^m$  de la loi  $F$ . Comme

$$\mathbf{E}\alpha_m = \mathbf{E}\{\mathbf{E}\{X^m | \mathbb{X}\}\} = \frac{1}{n} \mathbf{E}\left\{\sum_{i=1}^n X_i^m\right\} = a_m, \quad m = 1, 2, \dots, 2k, \quad (3.33)$$

on voit que le moment  $\alpha_m$  de la loi empirique est un estimateur sans biais de  $a_m$ . On remarque ici que tous les moments  $\alpha_m, m = 1, 2, \dots$ , de la loi empirique (9) existent, tandis que la loi  $F$  n'a d'après notre hypothèse (2), que les moments  $a_1, \dots, a_{2k}$ . Si nous prenons  $m \leq k$ , alors nous pouvons calculer la variance de la statistique  $\alpha_m$ , car

$$\begin{aligned} \mathbf{Var} \alpha_m &= \mathbf{Var} \left\{ \frac{1}{n} \sum_{i=1}^n X_i^m \right\} = \frac{1}{n} \mathbf{Var} X_1^m = \\ &= \frac{1}{n} \{ \mathbf{E}X_1^{2m} - (\mathbf{E}X_1^m)^2 \} = \frac{1}{n} (a_{2m} - a_m^2). \end{aligned} \quad (3.34)$$

De cette formule on déduit que la variance,  $\mathbf{Var} \alpha_m$ , de l'estimateur  $\alpha_m$  existe si  $m \leq k$ . De plus on en déduit que  $\mathbf{Var} \alpha_m \rightarrow 0$  quand  $n \rightarrow \infty$ , et par conséquent de l'inégalité de Tchebschev il résulte que pour tout  $\varepsilon > 0$

$$\mathbf{P}\{|\alpha_m - a_m| > \varepsilon\} = \mathbf{P}\{|\alpha_m - \mathbf{E}\alpha_m| > \varepsilon\} \leq \frac{\mathbf{Var} \alpha_m}{\varepsilon^2} = \frac{a_{2m} - a_m^2}{n\varepsilon^2} \rightarrow 0, \quad (3.35)$$

quand  $n \rightarrow \infty$ . Ainsi de (33) et (35) il résulte que  $\{\alpha_m\}$  est une suite consistante (cohérente) d'estimateurs sans biais de  $a_m$  ( $m = 1, 2, \dots, k$ ). On peut remarquer que pour estimer la

précision de l'estimateur  $\alpha_m$  du moment  $a_m$  on a eu besoin d'utiliser le moment  $\alpha_{2m}$  d'ordre  $2m$ .

**Exemple 1.** Soient  $a = \mathbf{E}X_1$  et  $\sigma^2 = \mathbf{Var}X_1$  et supposons que nous voulions estimer  $a$ . Comme nous l'avons dit, nous pouvons prendre la moyenne

$$\alpha_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

de la loi empirique comme estimateur de  $a = a_1$ , moyenne de la loi  $F$ .

D'après (33) on a

$$\mathbf{E}\bar{X}_n = a = \mathbf{E}X_1$$

et de (34) on déduit

$$\mathbf{Var}\bar{X}_n = \frac{1}{n} \mathbf{Var}X_1 = \frac{\sigma^2}{n} = \frac{1}{n}(a_2 - a^2),$$

et, par conséquent, de (35) on déduit que pour tout  $\varepsilon > 0$

$$\mathbf{P}\{|\bar{X}_n - a| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon n} \rightarrow 0, \quad n \rightarrow \infty, \quad (3.36)$$

c'est-à-dire que  $\{\bar{X}_n\}$ ,  $n \in \mathbb{N}$ , est une suite consistante d'estimateurs sans biais de la moyenne  $a$  de la loi  $F$ , si  $\sigma^2 < \infty$ .

**Remarque 5. (Théorème de Khinchine.)** On peut montrer que pour que la suite  $\{\bar{X}_n\}$ ,  $n \in \mathbb{N}$ , soit cohérente il suffit que  $\mathbf{E}X_1$  existe.

**Exemple 2.** Supposons que nous voulions estimer

$$\sigma^2 = \mathbf{Var}X_1 = a_2 - a_1^2 = a_2 - a^2. \quad (3.37)$$

Comme nous l'avons dit, nous pouvons prendre la variance

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \alpha_2 - \alpha_1^2 \quad (3.38)$$

de la loi empirique comme estimateur de  $\sigma^2$ . De (38) on déduit

$$\begin{aligned} \mathbf{E}s_n^2 &= \mathbf{E}\alpha_2 - \mathbf{E}\alpha_1^2 = a_2 - [\mathbf{Var}\alpha_1 + a^2] = \\ &= a_2 - a^2 - \frac{a_2 - a^2}{n} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2, \end{aligned}$$

i.e.  $s_n^2$  est un estimateur de  $\sigma^2$  qui a un biais  $b_n$ ,

$$b_n = \mathbf{E}(s_n^2 - \sigma^2) = -\frac{\sigma^2}{n}. \quad (3.39)$$

Si nous prenons la statistique

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} s_n^2 \quad (3.40)$$

comme estimateur de  $\sigma^2$ , alors on aura un estimateur sans biais de  $\sigma^2$ , car de (40) on déduit :

$$\mathbf{E}S_n^2 = \mathbf{E}\left(\frac{n}{n-1} s_n^2\right) = \frac{n}{n-1} \mathbf{E}s_n^2 = \sigma^2. \quad (3.41)$$

Pour calculer la variance  $\mathbf{Var} s_n^2$  de la statistique  $s_n^2$ ,

$$\mathbf{Var} s_n^2 = \mathbf{E}(s_n^2)^2 - (\mathbf{E}s_n^2)^2 = \mathbf{E}(s_n^2)^2 - \left(\frac{n-1}{n} \sigma^4\right)^2, \quad (3.42)$$

il nous faut savoir calculer  $\mathbf{E}(s_n^2)^2$ . Pour faire cela on remarque que la statistique  $(X_i - \bar{X}_n)^2$  est invariante par rapport à la moyenne  $a = \mathbf{E}X_1$  de la loi  $F$ . Cela veut dire que si nous posons  $Y_i = X_i - c$  ( $i = 1, 2, \dots, n$ ), où  $c$  est un nombre arbitraire et si

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i,$$

alors on voit que

$$Y_i - \bar{Y}_n = X_i - c - (\bar{X}_n - c) = X_i - \bar{X}_n, \quad (3.43)$$

donc pour calculer  $\mathbf{E}(s_n^2)^2$  nous pouvons admettre que  $a = \mathbf{E}X_1 = 0$ . Dans ce cas  $m_j = a_j$  et nous pouvons écrire :

$$\begin{aligned} s_n^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n X_i \right)^2 = \\ &= \frac{n-1}{n^2} \sum_{i=1}^n X_i^2 - \frac{2}{n} \sum_{i < j} X_i X_j, \end{aligned} \quad (3.44)$$

d'où, comme  $\mathbf{E}X_1 = 0$  par hypothèse et de l'indépendance de  $X_i$  et  $X_j$ , on déduit

$$\begin{aligned} \mathbf{E}(s_n^2)^2 &= \mathbf{E} \left\{ \frac{(n-1)^2}{n^4} \sum_{i < j} X_i^2 X_j^2 + \frac{4}{n^4} \sum_{i < j} X_i^2 X_j^2 \right\} = \\ &= \frac{(n-1)^2}{n^3} m_4 + \frac{(n-1)^2 + 2}{n^3} (n-1) \sigma^4. \end{aligned} \quad (3.45)$$

De (42) et (45) il résulte que

$$\mathbf{Var} S_n^2 = \frac{(n-1)^2}{n^3} \left( m_4 - \frac{n-3}{n-1} \sigma^4 \right), \quad (3.46)$$

et par conséquent, on en déduit que

$$\mathbf{Var} s_n^2 \rightarrow 0 \quad \text{quand } n \rightarrow \infty.$$

De (46) il est facile de déduire la variance  $\mathbf{Var} S_n^2$  de la statistique  $S_n^2$ , qui est le meilleur estimateur sans biais de  $\sigma^2$ (41). On a

$$\mathbf{Var} S_n^2 = \mathbf{Var} \left( \frac{n}{n-1} s_n^2 \right) = \frac{n^2}{(n-1)^2} \mathbf{Var} s_n^2 = \frac{1}{n} \left( m_4 - \frac{n-3}{n-1} \sigma^4 \right), \quad (3.47)$$

et on voit que  $\mathbf{Var} S_n^2$  tend aussi vers 0 quand  $n \rightarrow \infty$ . Comme pour tout  $\varepsilon > 0$

$$\mathbf{P} \{ |S_n^2 - \sigma^2| \geq \varepsilon \} = \mathbf{P} \{ |S_n^2 - \mathbf{E}S_n^2| \geq \varepsilon \} \leq \frac{\mathbf{Var} S_n^2}{\varepsilon^2} \rightarrow 0, \quad (3.48)$$

quand  $n \rightarrow \infty$ , nous pouvons dire que  $\{S_n^2\}$  est une suite cohérente d'estimateurs sans biais de la variance  $\sigma^2$  de la loi  $F(x)$ . On remarque ici, que de (47) on déduit

$$\mathbf{Var} s_n^2 < \mathbf{Var} S_n^2,$$

i.e. le risque quadratique de  $s_n^2$  est plus petit de celui de  $S_n^2$ , mais l'estimateur  $s_n^2$  a le biais  $b_n = -\sigma^2/n$ .

Nous avons montré (35) que le moment

$$\alpha_m = \frac{1}{n} \sum_{i=1}^n X_i^m$$

d'ordre  $m$  ( $m = 1, 2, \dots, k$ ) de la loi empirique est un bon estimateur du moment

$$a_m = \mathbf{E}X_1^m = \int_{-\infty}^{\infty} x^m dF(x),$$

de la loi  $F(x)$  en ce sens que

$$\mathbf{E}\alpha_m = a_m \quad \text{et} \quad \mathbf{Var} \alpha_m = \frac{1}{n}(a_{2m} - a_m^2) \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

Que pouvons nous dire de plus ? La statistique

$$\alpha_m = \frac{1}{n} \sum_{i=1}^n X_i^m, \quad m = 1, \dots, k,$$

est la somme des variables indépendantes  $X_1^m, \dots, X_n^m$ , puisque les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes et que de plus elles suivent la même loi. En outre, nous savons que la variance

$$\mathbf{Var} X_i^m = a_{2m} - a_m^2$$

existe pour tout  $m = 1, 2, \dots, k$ . Par conséquent du théorème central limite il résulte que

$$\mathbf{P} \left\{ \frac{\alpha_m - a_m}{\sqrt{\mathbf{Var} \alpha_m}} < x \right\} = \mathbf{P} \left\{ \sqrt{n} \frac{\alpha_m - a_m}{\sqrt{a_{2m} - a_m^2}} < x \right\} \rightarrow \Phi(x) \quad \text{quand } n \rightarrow \infty, \quad (3.49)$$

c'est-à-dire que la suite  $\{\alpha_m\}_{m \in \mathbb{N}^*}$  est asymptotiquement normalement distribuée de paramètres  $a_m$  et  $(a_{2m} - a_m^2)/n$ , ce que nous écrivons de la manière suivante :

$$\alpha_m \quad \text{est} \quad AN \left( a_m, \frac{a_{2m} - a_m^2}{n} \right), \quad (3.50)$$

ou la suivante :

$$\sqrt{n} \frac{\alpha_m - a_m}{\sqrt{a_{2m} - a_m^2}} \quad \text{est} \quad AN(0, 1). \quad (3.51)$$

D'après (35) nous savons que si  $n \rightarrow \infty$  alors pour tout  $\varepsilon > 0$

$$\mathbf{P} \{ |\alpha_m - a_m| \geq \varepsilon \} \rightarrow 0. \quad (3.52)$$

En utilisant l'approximation normale (49) nous pouvons estimer la probabilité de l'événement  $\{|\alpha_m - a_m| \geq \varepsilon\}$ . On a

$$\mathbf{P}\{|\alpha_m - a_m| \geq \varepsilon\} = \mathbf{P}\left\{\frac{|\alpha_m - a_m|}{\sqrt{\mathbf{Var}\alpha_m}} \geq \frac{\varepsilon}{\sqrt{\mathbf{Var}\alpha_m}}\right\}. \quad (3.53)$$

Si  $n$  est assez grand alors de (49) et (53) il résulte que

$$\mathbf{P}\{|\alpha_m - a_m| \geq \varepsilon\} \approx 2\Phi\left\{-\frac{\varepsilon\sqrt{n}}{\sqrt{a_{2m} - a_m^2}}\right\}, \quad (3.54)$$

où

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

est la fonction de répartition de la loi normale  $N(0, 1)$ , qui satisfait l'identité suivante :

$$\Phi(x) + \Phi(-x) \equiv 1, \quad |x| < \infty. \quad (3.55)$$

Notons ici, que si  $Z$  est une variable aléatoire qui suit la loi normale  $N(0, 1)$ ,

$$\mathbf{P}\{Z \leq x\} = \Phi(x), \quad x \in \mathbb{R}^1,$$

alors de (55) il résulte que

$$\mathbf{P}\{|Z| \leq x\} = 2\Phi(x) - 1, \quad (3.56)$$

ce qui a déjà été utilisé pour obtenir (54) à partir de (49). Ainsi, de (53)-(56) il résulte que

$$\mathbf{P}\{|\alpha_m - a_m| \leq \varepsilon\} \approx 1 - 2\Phi\left\{-\frac{\varepsilon\sqrt{n}}{\sqrt{a_{2m} - a_m^2}}\right\} = 2\Phi\left\{\frac{\varepsilon\sqrt{n}}{\sqrt{a_{2m} - a_m^2}}\right\} - 1, \quad (3.57)$$

i.e. pour tout  $\varepsilon > 0$  on a

$$\mathbf{P}\{\alpha_m - \varepsilon \leq a_m \leq \alpha_m + \varepsilon\} \approx 2\Phi\left\{\frac{\varepsilon\sqrt{n}}{\sqrt{a_{2m} - a_m^2}}\right\} - 1, \quad (3.58)$$

quand  $n$  est assez grand.

Nous devons constater que nous ne pouvons pas utiliser (58) directement pour savoir avec quelle probabilité l'intervalle

$$[\alpha_m - \varepsilon; \alpha_m + \varepsilon] \quad (3.59)$$

"couvre" la valeur inconnue de  $a_m$ , ou, comme on dit, avec quelle probabilité  $a_m$  appartient à l'intervalle  $[\alpha_m - \varepsilon; \alpha_m + \varepsilon]$ , que l'on appelle *un intervalle de confiance*. Pour avoir la possibilité d'utiliser (58) pour estimer

$$\mathbf{P}\{a_m \in [\alpha_m - \varepsilon; \alpha_m + \varepsilon]\}$$

nous devons substituer aux paramètres inconnus  $a_{2m}$  et  $a_m^2$  dans la partie droite de (58) leurs estimateurs  $\alpha_{2m}$  et  $\alpha_m^2$  et de cette manière nous aurons pour  $n$  assez grand la relation suivante :

$$\mathbf{P}\{\alpha_m - \varepsilon \leq a_m \leq \alpha_m + \varepsilon\} \approx 2\Phi\left\{\frac{\varepsilon\sqrt{n}}{\sqrt{\alpha_{2m} - \alpha_m^2}}\right\} \quad (3.60)$$

Maintenant nous allons utiliser l'approximation (60) pour construire l'intervalle de confiance (59) tel que

$$\mathbf{P}\{a_m \in [\alpha_m - \varepsilon; \alpha_m + \varepsilon]\} \approx P = 1 - \alpha, \quad (3.61)$$

où la probabilité  $P = 1 - \alpha$ , appelée *le coefficient de confiance*, est choisie d'avance,  $0.5 < P < 1$ ,  $0 < \alpha < 0.5$ . Ayant choisi un coefficient de confiance  $P = 1 - \alpha$ , il nous faut résoudre l'équation

$$2\Phi\left\{\frac{\varepsilon\sqrt{n}}{\sqrt{\alpha_{2m} - \alpha_m^2}}\right\} - 1 = P = 1 - \alpha \quad (3.62)$$

pour trouver  $\varepsilon$  qui satisfait à (61).

Soit  $\alpha$  une probabilité telle que  $0 < \alpha < 0.5$ . Notons  $z_\alpha^+$  et  $z_\alpha^-$  les quantiles de seuils  $\alpha$  et  $1 - \alpha$  respectivement, c'est-à-dire que  $z_\alpha^+$  et  $z_\alpha^-$  satisfont aux relations :

$$\Phi(z_\alpha^-) = \alpha \quad \text{et} \quad \Phi(z_\alpha^+) = 1 - \alpha, \quad 0 < \alpha < 0.5.$$

De (55) il résulte que  $z_\alpha^+ = -z_\alpha^-$ . En utilisant les notations de (62) on a

$$\Phi\left\{\frac{\varepsilon\sqrt{n}}{\sqrt{\alpha_{2m} - \alpha_m^2}}\right\} = \frac{1+P}{2} = 1 - \frac{\alpha}{2} \quad (3.63)$$

d'où l'on obtient

$$x_{\alpha/2}^+ = \frac{\varepsilon\sqrt{n}}{\sqrt{\alpha_{2m} - \alpha_m^2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (3.64)$$

et par conséquent on trouve que

$$\varepsilon = \frac{x_P}{\sqrt{n}} \sqrt{\alpha_{2m} - \alpha_m^2} = \frac{1}{\sqrt{n}} x_{\alpha/2}^+ \sqrt{\alpha_{2m} - \alpha_m^2}. \quad (3.65)$$

De (60)–(62) et (65) il résulte que

$$\mathbf{P}\left\{\alpha_m - x_{\alpha/2}^+ \sqrt{\frac{\alpha_{2m} - \alpha_m^2}{n}} \leq a_m \leq \alpha_m + x_{\alpha/2}^+ \sqrt{\frac{\alpha_{2m} - \alpha_m^2}{n}}\right\} \approx P = 1 - \alpha. \quad (3.66)$$

Nous voyons qu'en utilisant les moments  $\alpha_{2m}$  et  $\alpha_m$  de la loi empirique, et le fait qu'ils sont asymptotiquement normalement distribués, nous sommes parvenus à construire pour le moment  $a_m$  un intervalle de confiance  $(\alpha_m - \varepsilon; \alpha_m + \varepsilon)$  dont le coefficient de confiance est approximativement égal à  $P = 1 - \alpha$ . Dans la table 1 ci-dessous nous donnons les valeurs de  $P = 1 - \alpha$  les plus répandues dans la pratique et les valeurs  $x_{\alpha/2}^+$  correspondantes, ce qui permet facilement de calculer  $\varepsilon$  en utilisant la formule (65).

$P = 1 - \alpha$	0.90	0.95	0.99	0.995
$x_{\alpha/2}^+$	1.644854	1.959964	2.575829	2.807034

(3.67)

Table 1.

**Exemple 3.** Soit  $m = 1$ , c'est-à-dire que nous estimons la moyenne  $a = \mathbf{E}X_1$  de la loi  $F(x)$ . Nous savons, d'après l'exemple 1, que  $\alpha_1 = \bar{X}_n$ , moyenne de la loi empirique, est un estimateur sans biais de  $a$ , en outre, nous savons d'après (36) que

$$\mathbf{P}\{|\bar{X}_n - a| \geq \varepsilon\} \rightarrow 0. \quad (3.68)$$

Maintenant, en utilisant (57), nous obtenons que

$$\mathbf{P}\{|\bar{X}_n - a| \leq \varepsilon\} \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{\mathbf{Var}\bar{X}_n}}\right) - 1 = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1, \quad (3.69)$$

car

$$\mathbf{Var}\bar{X}_n = \frac{\sigma^2}{n}, \quad \text{où } \sigma^2 = \mathbf{Var}X_1 = a_2 - a_1^2 = a_2 - a^2.$$

Dans (69) nous pouvons, en utilisant l'exemple 2, estimer le paramètre inconnu  $\sigma = \sqrt{\sigma^2}$  par la statistique  $S_n = \sqrt{S_n^2}$ , sachant que  $\mathbf{E}S_n^2 = \sigma^2$ . Dans ce cas, de (69) il résulte que

$$\mathbf{P}\{|\bar{X}_n - a| \leq \varepsilon\} \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{S_n}\right) - 1 \quad (3.70)$$

et par conséquent on obtient un analogue de (66)

$$\mathbf{P}\left\{\bar{X}_n - x_{\alpha/2}^+ \frac{S_n}{\sqrt{n}} \leq a \leq \bar{X}_n + x_{\alpha/2}^+ \frac{S_n}{\sqrt{n}}\right\} \approx P = 1 - \alpha, \quad (3.71)$$

en choisissant dans (67) le coefficient de confiance  $P = 1 - \alpha$  et le quantile  $x_{\alpha/2}^+$  de la loi normale  $N(0, 1)$ . Il est évident que dans (71) on aurait pu utiliser la statistique  $s_n$  comme estimateur de  $\sigma$  au lieu de  $S_n$ , où  $s_n^2$  est la variance de la loi empirique.

## 3.2 Médiane de la loi empirique.

1. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon et  $X_i$  suit une loi de fonction de répartition

$$F(x) = \mathbf{P}(X_i < x).$$

Notons  $X^{(\cdot)} = (X_{(1)}, \dots, X_{(n)})^T$  le vecteur des statistiques d'ordre associé au vecteur  $\mathbb{X}$ . Par définition, la médiane de la loi empirique est la statistique

$$\mu_n = \begin{cases} X_{(k+1)}, & \text{si } n = 2k + 1, \\ \frac{1}{2}(X_{(k)} + X_{(k+1)}), & \text{si } n = 2k. \end{cases}$$

On sait que si  $n$  est impair,  $n = 2k + 1$ , alors

$$\mathbf{P}\{\mu_{2k+1} < x\} = I_{F(x)}(k + 1, k + 1), \quad (3.1)$$

et on obtient

$$\mathbf{P}\{\mu_{2k+1} < x\} = S_{2k+2} \left[ \left( F(x) - \frac{1}{2} \right) \sqrt{\frac{2k+2}{F(x)[1-F(x)]}} \right], \quad (3.2)$$

où  $S_f(x)$  est la fonction de la répartition de la loi de Student à  $f$  degrés de liberté. Dans le cas où  $n$  est un nombre pair,  $n = 2k$ , la distribution de la statistique  $\mu_{2k}$  est beaucoup plus

compliquée. On remarque que d'habitude, dans la pratique, lorsque  $n$  est assez grand, on utilise le fait que

$$\mathcal{L}(\sqrt{n}(\mu_n - \mu)) \rightarrow N\left(0, \frac{1}{4f^2(\mu)}\right), \quad (3.3)$$

ou plus précisément :

$$\mathbf{P}\{2\sqrt{n}f(\mu)(\mu_n - \mu) < y\} = \Phi(y) + O\left(\frac{1}{\sqrt{n}}\right),$$

où  $\mu$  est la médiane de la loi  $F(x)$ ,  $F(\mu) = 0.5$ , et  $f(x)$  est la densité de cette loi, c'est-à-dire que  $f(x) = F'(x)$ . La précision de cette approximation normale n'est pas très bonne quand  $n$  n'est pas assez grand. Par ailleurs, il est très naturel de s'attendre à ce que la distribution de la statistique  $\mu_{2k+1}$  soit plus proche de la distribution de la statistique  $\mu_{2k}$ , et justement Bolshev (1963) a utilisé ce fait pour construire une approximation qui est meilleure que l'approximation normale (3).

Soit

$$F_n(x) = \mathbf{P}\{\mu_n < x\sqrt{2\pi t}\}, \quad (3.4)$$

où  $t = 1/(8[n/2] + 5)$ . Bolshev (1963) a démontré que

$$F_{2k}(x) - F_{2k+1}(x) = -8(\pi - 2)x\varphi(x)t^2 + O(t^3), \quad (3.5)$$

et

$$F_{2k+1}(x) = \Phi(x) + \varphi(x)\frac{3x - (2\pi - 6)x^3}{6}t + O(t^2), \quad (3.6)$$

d'où l'on peut déduire que la statistique

$$Y_n = \frac{\mu_n}{\sqrt{2\pi t}} \left[ 1 + \frac{1}{\sigma} \left( 3 - (2\pi - 6) \left( \frac{\mu_n}{\sqrt{2\pi t}} \right)^2 \right) \right]$$

est asymptotiquement normale de paramètres 0 et 1,

$$\mathbf{P}\{Y_n < y\} = \Phi(y) + O(t^2).$$

Notons  $\mu_n(P)$  le  $P$ -quantile (le quantile de niveau  $P$ ) de la distribution de la statistique  $\mu_n$  :

$$\mathbf{P}\{\mu_n < \mu_n(P)\} = P.$$

Dans ce cas de (6) on déduit que

$$\mu_{2k}(P) = \mu_{2k+1}(P)[1 + 8(\pi - 2)t^2] + O(t^2)$$

donc

$$\mu_{2k}(P) \cong \mu_{2k}^*(P), \quad (3.7)$$

où

$$\mu_{2k}^*(P) = \mu_{2k+1}(P)[1 + 8(\pi - 2)t^2].$$

La formule (7) donne une bonne approximation, même pour les petites valeurs de  $n$ . Par exemple si  $k = 1$ , alors la différence

$$D = \mathbf{P}\{\mu_2 < \mu_2^*(P)\} - P$$

prend les valeurs suivantes

$$-0.0001, \quad -0.0002, \quad 0.0000, \quad 0.0004, \quad 0.0012, \quad 0.0011, \quad 0.0000,$$

correspondant aux valeurs de  $P$

$$0.0014, \quad 0.0064, \quad 0.0228, \quad 0.0664, \quad 0.1575, \quad 0.3074, \quad 0.5000.$$

2. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon, dont la fonction de la répartition  $F(x)$  appartient à une famille  $\mathcal{F} = (F)$  de fonctions de répartition continues. Comme précédemment, on note  $f(x)$  la densité de  $F(x)$  :

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Dans ce cas, si  $\mu = \mu(F)$  est la médiane de la distribution, dont la fonction de répartition est  $F(x)$ , alors

$$\int_{-\infty}^{\mu(F)} f(x) dx = F(\mu(F)) = 0.5,$$

i.e.

$$\mathbf{P}\{X_i < \mu(F)\} = \mathbf{P}\{X_i \geq \mu(F)\} = 0.5.$$

Notre but est de construire un intervalle de confiance pour  $\mu(F)$ .

Soit  $X^{(\cdot)} = (X_{(1)}, \dots, X_{(n)})^T$  le vecteur des statistiques d'ordre, construit en utilisant l'échantillon  $\mathbb{X}$ . Dans ce cas avec la probabilité 1

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}.$$

Comme intervalle de confiance, on peut choisir

$$(X_{(i)}, X_{(j)}), \quad i < j.$$

Il est très naturel de choisir cet intervalle symétrique en posant  $j = n - i - 1$ , puisque nous nous sommes intéressés à l'obtention de conclusions statistiques qui sont indépendantes de la distribution inconnue  $F$ . De la définition des statistiques d'ordre  $X_{(1)}, \dots, X_{(n)}$  il s'ensuit que

$$\begin{aligned} & \mathbf{P}\{X_{(i)} < \mu(F) < X_{(j)} | F\} = \\ & = 1 - \mathbf{P}\{X_{(i)} \geq \mu(F) | F\} - \mathbf{P}\{X_{(j)} \leq \mu(F) | F\} = \\ & = 1 - \mathbf{P}\{F(X_{(i)}) \geq F(\mu(F)) | F\} - \mathbf{P}\{F(X_{(j)}) \leq F(\mu(F)) | F\} = \\ & = 1 - \mathbf{P}(U_{(i)} \geq 0.5) - \mathbf{P}(U_{(j)} \leq 0.5) = \\ & = 1 - \sum_{m=0}^{i-1} \binom{n}{m} \left(\frac{1}{2}\right)^n - \sum_{m=j}^n \binom{n}{m} \left(\frac{1}{2}\right)^n, \end{aligned}$$

et on voit bien que cette probabilité *ne* dépend pas de  $F$ , c'est-à-dire qu'on a obtenu une statistique "libre" comme on l'avait voulu. On note ici que comme d'habitude

$$U^{(\cdot)} = (U_{(1)}, \dots, U_{(n)})^T$$

représente le vecteur des statistiques d'ordre associé à l'échantillon  $U = (U_1, \dots, U_n)^T$  de la loi uniforme sur  $(0,1)$ .

Maintenant, considérons l'intervalle de confiance "symétrique", i.e.  $j = n - i + 1$ . Dans ce cas on a

$$\mathbf{P}\{X_{(i)} < \mu(F) < X_{(n-i+1)} | F\} = 1 - 2 \sum_{m=0}^{i-1} \binom{n}{m} \left(\frac{1}{2}\right)^n,$$

car

$$\sum_{m=j}^n \binom{n}{m} \left(\frac{1}{2}\right)^n = \sum_{m=n-i+1}^n \binom{n}{m} \left(\frac{1}{2}\right)^n = \sum_{m=0}^{i-1} \binom{n}{m} \left(\frac{1}{2}\right)^n.$$

Donc quand  $n$  est grand, du Théorème de de Moivre-Laplace on déduit que

$$\begin{aligned} \mathbf{P}\{X_{(i)} < \mu(F) < X_{(n-i+1)} | F\} &\cong 1 - 2\Phi\left(\frac{i-1-\frac{n}{2}+0.5}{0.5\sqrt{n}}\right) = \\ &= 1 - 2\Phi\left(\frac{2i-n-1}{\sqrt{n}}\right) = 2\Phi\left(\frac{n+1-2i}{\sqrt{n}}\right) - 1. \end{aligned} \quad (3.8)$$

Comment trouver le numéro  $i$  dans (8) quand le coefficient de confiance  $P$  est donné d'avance ? Pour cela il faut résoudre l'équation

$$2\Phi\left(\frac{n+1-2i}{\sqrt{n}}\right) - 1 = P \quad (3.9)$$

par rapport à  $i$  (on remarque que  $0.5 < P < 1$ ), d'où l'on obtient

$$\frac{n+1-2i}{\sqrt{n}} = \Psi\left(\frac{1+P}{2}\right),$$

où  $\Psi(z) = \Phi^{-1}(z)$ , et donc

$$i = \left[ 0.5 \left\{ n+1 - \sqrt{n} \Psi\left(\frac{1+P}{2}\right) \right\} + 1 \right],$$

où  $[a]$  dans la dernière formule est la partie entière du nombre  $a$ .

### 3.3 Théorème de Kolmogorov.

A.Kolmogorov (1933) a trouvé la distribution limite ( $n \rightarrow \infty$ ) de la statistique  $\sqrt{n}D_n$  lorsque  $F(x)$  est une fonction continue.

**Theoreme** (de Kolmogorov). *Si  $F(x)$  est continue, alors pour  $z > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\sqrt{n}D_n \leq z\} = K(z) = \sum_{j=-\infty}^{+\infty} (-1)^j \exp(-2j^2z^2). \quad (3.1)$$

On dit que  $K(z)$  est la *fonction de répartition de Kolmogorov*. Il y a des tables statistique (voir, par exemple, Smirnov (1939), Birnbaum (1952), Bolshev et Smirnov (1968), Conover (1980)) des valeurs de la fonction de Kolmogorov  $K(z)$ , mais en pratique pour

faire des calculs approximatifs, quand  $z > 2.5$ , on utilise souvent une simple approximation évidente :

$$K(z) \cong 1 - 2e^{-2z^2}.$$

Soit  $P$  un nombre fixé,  $0.5 < P < 1$ , et soit  $z_P$  le quantile de niveau  $P$  de la fonction de la répartition de Kolmogorov, c'est-à-dire  $z_P$  est la racine de l'équation  $K(z) = P$  :

$$K(z_P) = P.$$

Dans ce cas de (1) on tire que

$$\begin{aligned} & \mathbf{P} \{ \sqrt{n}D_n \leq z_P \} = \\ & = \mathbf{P} \left\{ \mathbb{F}_n(x) - \frac{1}{\sqrt{n}}z_P \leq F(x) \leq \mathbb{F}_n(x) + \frac{1}{\sqrt{n}}z_P \right\} \rightarrow K(z_P) = P, \end{aligned} \quad (3.2)$$

quand  $n \rightarrow \infty$ . C'est-à-dire que si  $n$  est grand, alors avec la probabilité  $\cong P$  les valeurs  $F(x)$  pour tout  $x$  satisfont les équations

$$F_n(x) - \frac{1}{\sqrt{n}}z_P \leq F(x) \leq F_n(x) + \frac{1}{\sqrt{n}}z_P. \quad (3.3)$$

Comme  $0 \leq F(x) \leq 1$ , la dernière relation peut être s'écrire :

$$\max \left( 0, F_n(x) - \frac{1}{\sqrt{n}}z_P \right) \leq F(x) \leq \min \left( F_n(x) + \frac{1}{\sqrt{n}}z_P, 1 \right).$$

### 3.3.1 Transformation de Smirnov. Test de type de Kolmogorov-Smirnov pour des lois discrètes.

**Transformation de Smirnov pour une distribution continue.** Soit  $X$  une variable aléatoire dont la fonction de répartition  $F(x) = \mathbf{P}\{X \leq x\}$  est continue et croissante. Dans ce cas, la statistique  $U = F(X)$  suit une loi uniforme sur  $[0, 1]$ . Pour prouver cette affirmation on remarque tout d'abord que

$$\mathbf{P}\{U \leq u\} = 0 \quad \text{pour tout } u \leq 0$$

et que

$$\mathbf{P}\{U \leq u\} = 1 \quad \text{pour tout } u \geq 1.$$

Soit  $u$  un nombre réel quelconque,  $0 < u < 1$ . Dans ce cas comme  $F(x)$  est continue et croissante on obtient

$$\begin{aligned} \mathbf{P}\{U \leq u\} &= \mathbf{P}\{F^{-1}(U) \leq F^{-1}(u)\} = \mathbf{P}\{X \leq F^{-1}(u)\} = \\ &= F(F^{-1}(u)) = u, \quad 0 < u < 1. \end{aligned}$$

**Transformation de Smirnov pour une distribution arbitraire.** Soit  $X$  une variable aléatoire quelconque et soit

$$F(x) = \mathbf{P}\{X \leq x\} \quad \text{et} \quad F_-(x) = \mathbf{P}\{X < x\}.$$

Il est évident que si  $X$  est une variable aléatoire continue

$$F(x) = F_-(x).$$

Alors on peut démontrer (voir §V.1), que

$$\mathbf{P}\{F(X) \leq z\} \leq z \leq \mathbf{P}\{F_-(X) < z\}$$

pour tout  $z \in [0, 1]$ .

**Colloraire 1.** Si la distribution de  $X$  est continue, dans ce cas

$$\mathbf{P}\{F(X) \leq z\} = \mathbf{P}\{F(X) < z\} = z, \quad z \in [0, 1].$$

**Colloraire 2.** Soit  $U$  une variable aléatoire qui suit la loi uniforme sur  $[0,1]$  et qui est indépendante de  $X$ . Dans ce cas la statistique

$$Z = F_-(X) + U [F(X) - F_-(X)]$$

suit la loi uniforme sur  $[0,1]$ ,

$$\mathbf{P}\{Z \leq z\} = z$$

pour tout les  $z \in [0, 1]$ .

**Colloraire 3.** Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires indépendantes dont les fonctions de répartition sont connues :

$$F_i(x) = \mathbf{P}\{X_i \leq x\}, \quad F_{i-}(x) = \mathbf{P}\{X_i < x\}, \quad i = 1, 2, \dots, n.$$

De plus, soient  $U_1, U_2, \dots, U_n$  des variables aléatoires indépendantes, qui suivent la même loi uniforme sur  $[0,1]$  et qui sont indépendantes de  $X_1, X_2, \dots, X_n$ . Dans ce cas, d'après de colloraire 2, les statistiques  $Z_1, Z_2, \dots, Z_n$ , où

$$Z_i = F_{i-}(X_i) + U_i [F_i(X_i) - F_{i-}(X_i)],$$

sont indépendantes et uniformément distribuées sur  $[0,1]$ .

Le colloraire 3 nous donne la possibilité de construire les tests non paramétriques de Kolmogorov, de Smirnov, d'omega-carré de Von Mises etc., dans les situations où les données  $X_1, X_2, \dots, X_n$  sont indépendantes et suivent des lois continues ou discrètes.

### Applications statistiques.

Soient  $X_1, X_2, \dots, X_m$  des variables aléatoires indépendantes et nous avons à tester l'hypothèse  $H_0$  selon laquelle

$$\mathbf{P}\{X_i = k\} = \frac{n_i!}{(n_i - k)!k!} p_i^k (1 - p_i)^{n_i - k}, \quad i = 1, 2, \dots, m,$$

où tout les  $p_i$  et  $n_1, n_2, \dots, n_m$  sont connus,  $0 < p_i < 1$ ;  $k = 0, 1, \dots, n_i$ . C'est-à-dire que, si  $H_0$  est vrai, alors  $X_i$  suit une loi binomiale  $B(n_i, p_i)$  (de paramètres  $n_i$  et  $p_i$ , et donc

$$F_i(x) = \mathbf{P}\{X_i \leq x\} = I_{1-p_i}(n_i - x, x + 1), \quad x = 0, 1, \dots, n_i,$$

et

$$F_{i-}(x) = \mathbf{P}\{X_i < x\} = \mathbf{P}\{X_i \leq x - 1\} = I_{1-p_i}(n_i - x + 1, x),$$

où  $I_x(a, b)$  définie ci-dessous est la fonction béta-incomplète d'Euler, et par conséquent pour appliquer le test de Kolmogorov, par exemple, pour tester  $H_0$  il ne reste qu'à construire d'après le corollaire 3 les statistiques

$$Z_i = I_{1-p_i}(n_i - X_i + 1, X_i) + U_i [I_{1-p_i}(n_i - X_i, X_i + 1) - I_{1-p_i}(n_i - X_i + 1, X_i)],$$

$i = 1, 2, \dots, m.$

Plus de détails on peut trouver dans Nikulin (1992), Huber et Nikulin (1993), Greenwood et Nikulin (1996).

Récemment M.Hocine a fait les études intéressantes sur le comportement de ce test et du test de type de omega-carré basées sur cette transformation de Smirnov.

### 3.4 Tests de Kolmogorov et Smirnov pour un échantillon.

Si la fonction de répartition de  $X_1$  est inconnue mais qu'on a fait l'hypothèse  $H_0$ , d'après laquelle

$$\mathbf{P}\{X_1 \leq x\} = F(x),$$

où  $F(x)$  est une fonction de répartition continue donnée, alors nous pouvons tester  $H_0$ , en utilisant le théorème de Kolmogorov. Symboliquement l'hypothèse  $H_0$  peut être présentée par la façon suivante :

$$H_0 : \mathbf{E}\mathbb{F}_n(x) \equiv F(x).$$

On détermine la statistique de Kolmogorov

$$D_n = \sup_{|x| < \infty} |\mathbb{F}_n(x) - F(x)|,$$

qui est désignée pour tester  $H_0$  contre l'hypothèse bilatérale

$$H_1 : \sup_{|x| < \infty} |\mathbf{E}\mathbb{F}_n(x) - F(x)| > 0,$$

et on considère en outre, les statistiques de Smirnov

$$D_n^+ = \sup_{|x| < \infty} (\mathbb{F}_n(x) - F(x)) \quad \text{et} \quad D_n^- = - \inf_{|x| < \infty} (\mathbb{F}_n(x) - F(x)),$$

qui sont utilisées pour tester  $H_0$  contre les alternatives unilatérales

$$H_1^+ : \sup_{|x| < \infty} (\mathbf{E}\mathbb{F}_n(x) - F(x))$$

et

$$H_1^- : - \inf_{|x| < \infty} (\mathbf{E}\mathbb{F}_n(x) - F(x))$$

respectivement.

Il est clair que  $D_n = \max(D_n^+, D_n^-)$ . En utilisant la transformation de Smirnov, on peut montrer que

$$D_n^+ = \max_{1 \leq m \leq n} \left( \frac{m}{n} - F(X_{(m)}) \right) \quad \text{et} \quad D_n^- = \max_{1 \leq m \leq n} \left( F(X_{(m)}) - \frac{m-1}{n} \right). \quad (3.1)$$

Il est clair aussi, que si  $H_0$  est vraie, alors

$$\mathbf{P} \{ D_n^+ \leq x | H_0 \} = \mathbf{P} \{ D_n^- \leq x | H_0 \}, \quad (3.2)$$

c'est-à-dire que,  $D_n^+$  et  $D_n^-$  suivent la même loi, quand  $H_0$  est vraie.

Comme a montré Smirnov (1944), pour tout  $x \in (0, 1)$

$$\mathbf{P} \{ D_n^+ \geq x | H_0 \} = \sum_{k=0}^{[n(1-x)]} \binom{n}{k} x \left( x + \frac{k}{n} \right)^{k-1} \left( 1 - x - \frac{k}{n} \right)^{n-k}, \quad (3.3)$$

$[a]$  - partie entière de  $a$ .

On peut montrer (Kolmogorov (1933), Smirnov (1944), Chernoff and Savage (1958), Bolshev (1963), Huber, Nikulin (1993)), que si  $n \rightarrow \infty$  et  $x$  appartient au domaine

$$\left\{ x : 0 < \varepsilon \leq x = O(n^{1/3}) \right\},$$

alors

$$\mathbf{P} \left\{ \frac{(6nD_n^+ + 1)^2}{18n} < x | H_0 \right\} = (1 - e^{-x}) + \frac{2x^2 - 4x - 1}{18n} e^{-x} + O\left(\frac{1}{n\sqrt{n}}\right), \quad (3.4)$$

et

$$\begin{aligned} & \mathbf{P} \left\{ \frac{(6nD_n + 1)^2}{18n} < x | H_0 \right\} = \\ & = K \left( \sqrt{\frac{x}{2}} \right) - \frac{1}{18} \sum_{k=-\infty}^{\infty} (-1)^k e^{-k^2 x} [P_k(x) + 2k^4 x - k^2] + O\left(\frac{1}{n\sqrt{n}}\right), \end{aligned} \quad (3.5)$$

où

$$\begin{aligned} P_k(x) &= \left[ k^2 - \frac{1 - (-1)^k}{2} \right] (1 - 2k^2 x) + 2k^2 x (k^2 x - 3) = \\ &= k^2 [2k^2 x^2 - 2x(k^2 + 3) + 1] + \frac{(-1)^k - 1}{2} (1 - 2k^2 x). \end{aligned}$$

Comme  $\chi_{2m}^2 = 2\gamma_m$  et

$$\mathbf{P} \{ \gamma_1 \leq x \} = 1 - e^{-x}, \quad \text{pour tout } x > 0,$$

de (4) et de (5) on déduit que pour les grandes valeurs de  $n$  la statistique

$$\frac{(6nD_n^+ + 1)^2}{9n}$$

est approximativement distribuée comme  $\chi_2^2$  et que

$$\mathbf{P} \left\{ \frac{(6nD_n + 1)^2}{18n} < x \right\} \approx K \left( \sqrt{\frac{x}{2}} \right).$$

Ces deux approximations sont déjà bonnes pour  $n \geq 20$ , les erreurs de ces approximations diminuent comme  $\frac{1}{n}$ .

Soit  $\alpha$  le niveau du test de Kolmogorov ( $0 < \alpha < 0.5$ ), basé sur la statistique  $D_n$ , et soient  $x_\alpha^+$  et  $x_\alpha$ , les valeurs critiques des tests basés sur  $D_n^+$  et  $D_n$ , i.e.

$$\mathbf{P}\{D_n^+ \geq x_\alpha^+\} = \alpha \quad \text{et} \quad \mathbf{P}\{D_n \geq x_\alpha\} = \alpha.$$

D'après le test de Kolmogorov

on rejette  $H_0$  en faveur de l'hypothèse  $H_1$  si  $D_n \geq x_\alpha$ .

De la même façon, d'après le test de Smirnov

on rejette  $H_1$  en faveur de l'hypothèse  $H_1^+$  si  $D_n^+ \geq x_\alpha$ .

On remarque que pour les petites valeurs de  $\alpha$  ( $0 < \alpha \leq 0.2$ ) il y a une liaison entre les valeurs critiques  $x_\alpha$  et  $x_{\alpha/2}^+$  :

$$x_\alpha \cong x_{\alpha/2}^+,$$

et l'erreur dans cette égalité est inférieure à 0.0005 :

$$|x_\alpha - x_{\alpha/2}^+| \leq 0.0005.$$

On peut montrer que cette erreur diminue très vite quand  $\alpha$  diminue. Par exemple, si  $\alpha \leq 0.1$ , alors

$$|x_\alpha - x_{\alpha/2}^+| \leq 0.00005.$$

Si  $n \geq 10$  et  $0.01 \leq \alpha \leq 0.2$ , pour calculer  $x_\alpha$  et  $x_{\alpha/2}^+$  il est recommandé d'utiliser les approximations de Bolshev (1963) :

$$x_\alpha \cong \sqrt{\frac{1}{2n} \left( y - \frac{2y^2 - 4y - 1}{18n} \right)}, \quad y = -\ln \frac{\alpha}{2}$$

$$x_\alpha^+ \cong \sqrt{\frac{1}{2n} \left( y - \frac{2y^2 - 4y - 1}{18n} \right)}, \quad y = -\ln \alpha.$$

On peut remarquer que si  $n$  est assez grand, alors

$$\sqrt{\frac{1}{2n} \left( y - \frac{2y^2 - 4y - 1}{18n} \right)} \cong \sqrt{\frac{y}{2n}}.$$

Dans la pratique ces formules donnent déjà de bons résultats dans le cas  $\alpha > 0.001$  pour  $n \geq 20$ .

Enfin, si

$$0.2 \leq \alpha \leq 0.3 \quad \text{et} \quad 10 \leq n \leq 50,$$

alors en prenant pour  $y$  la racine de l'équation

$$K \left( \sqrt{\frac{y}{2}} \right) = 1 - \alpha,$$

on obtient encore une approximation de Bolshev (1963)

$$x_\alpha \cong \sqrt{\frac{1}{2n} \left\{ y - \frac{1}{18n} [(2y^2 - 4y - 1) - \alpha^3(3y^2 - y + 0.5)] \right\}} - \frac{1}{6n}.$$

Dans le cas  $n \geq 100$  toutes ces approximations sont très bonnes pour calculer  $x_\alpha$  et  $x_\alpha^+$  pour tout  $\alpha$  tel que  $0.0001 \leq \alpha \leq 0.5$ .

### 3.5 Test de Kolmogorov-Smirnov pour deux échantillons.

Soient  $\mathbb{X} = (X_1, \dots, X_n)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_m)^T$  deux échantillons indépendants, et soit

$$F(x) = \mathbf{P}\{X_i < x\} \quad \text{et} \quad G = \mathbf{P}\{Y_j < y\}$$

les fonctions de répartition continues de  $X_i$  et  $Y_j$  respectivement. Nous pouvons construire deux lois empiriques, qui correspondent aux deux échantillons donnés  $\mathbb{X}$  et  $\mathbb{Y}$ . Notons  $\mathbb{F}_n(x)$  et  $\mathbb{G}_m(x)$  les fonctions de répartition de ces lois empiriques.

On utilise le test de Kolmogorov-Smirnov pour tester l'hypothèse

$$H_0 : F(x) \equiv G(x), \quad x \in \mathbf{R}^1,$$

qui peut s'écrire en fonction de  $\mathbb{F}_n$  et  $\mathbb{G}_m$  de la façon suivante :

$$H_0 : \mathbf{E}\mathbb{F}_n(x) \equiv \mathbf{E}\mathbb{G}_m(x),$$

contre l'hypothèse bilatérale

$$H_1 : \sup_{|x| < \infty} |\mathbf{E}\mathbb{G}_m(x) - \mathbf{E}\mathbb{F}_n(x)| > 0,$$

ou contre l'une de deux hypothèses unilatérales :

$$H_1^+ : \sup_{|x| < \infty} (\mathbf{E}\mathbb{G}_m(x) - \mathbf{E}\mathbb{F}_n(x)) > 0$$

ou

$$H_1^- : - \inf_{|x| < \infty} (\mathbf{E}\mathbb{G}_m(x) - \mathbf{E}\mathbb{F}_n(x)) > 0$$

respectivement. Pour tester  $H_0$  contre  $H_1$  on peut utiliser la statistique

$$D_{m,n} = \sup_{|x| < \infty} |\mathbb{G}_m(x) - \mathbb{F}_n(x)|, \quad (3.1)$$

où  $\mathbb{G}_m(x)$  et  $\mathbb{F}_n(x)$  sont les fonctions empiriques, associées à  $\mathbb{Y}$  et  $\mathbb{X}$ .

Si on teste  $H_0$  contre  $H_1^+$  ou  $H_1^-$ , on utilise les statistiques

$$D_{m,n}^+ = \sup_{|x| < \infty} (\mathbb{G}_m(x) - \mathbb{F}_n(x)) \quad \text{et} \quad D_{m,n}^- = - \inf_{|x| < \infty} (\mathbb{G}_m(x) - \mathbb{F}_n(x)). \quad (3.2)$$

Smirnov a montré (1939) que si l'hypothèse  $H_0$  est vraie, alors les statistiques  $D_{m,n}^+$ ,  $D_{n,m}^+$ ,  $D_{m,n}^-$ ,  $D_{n,m}^-$  suivent la même loi. En pratique les valeurs des statistiques (1) et (2) sont calculées d'après les formules suivantes :

$$D_{m,n}^+ = \max_{1 \leq r \leq m} \left( \frac{r}{m} - \mathbb{F}_n(Y_{(r)}) \right) = \max_{1 \leq s \leq n} \left( \mathbb{G}_m(X_{(s)}) - \frac{s-1}{n} \right),$$

$$D_{m,n}^- = \max_{1 \leq r \leq m} \left( \mathbb{F}_n(Y_{(r)}) - \frac{r-1}{m} \right) = \max_{1 \leq s \leq n} \left( \frac{s}{n} - \mathbb{G}_m(X_{(s)}) \right),$$

$$D_{m,n} = \max(D_{m,n}^+, D_{m,n}^-),$$

où  $X_{(i)}$  et  $Y_{(j)}$  sont les statistiques d'ordre, correspondant aux échantillons. On peut obtenir ces formules en utilisant la transformation de Smirnov et les propriétés des statistiques d'ordre de la loi uniforme sur  $[0, 1]$ . Smirnov (1939) a montré, que si  $\min(m, n) \rightarrow \infty$ , alors pour tout  $y$  positif

$$\lim \mathbf{P} \left\{ \sqrt{\frac{mn}{m+n}} D_{m,n}^+ < y | H_0 \right\} = 1 - e^{-2y^2},$$

$$\lim \mathbf{P} \left\{ \sqrt{\frac{mn}{m+n}} D_{m,n} < y | H_0 \right\} = K(y),$$

où  $K(z)$  est la fonction de Kolmogorov.

### 3.6 Test $\omega^2$ de Cramer-von Mises et statistiques associées de Lehmann, Gini, Downton, Moran-Greenwood et Sherman.

Souvent pour tester l'hypothèse simple

$$H_0 : \mathbf{E}\mathbb{F}_n(x) \equiv F(x), \quad |x| < \infty,$$

contre l'alternative

$$H_1 : \sup_{|x| < \infty} |\mathbf{E}\mathbb{F}_n(x) - F(x)| > 0,$$

au lieu d'utiliser le test de Kolmogorov, on construit le test  $\omega^2$  de Cramer et Von Mises, fondé sur la statistique

$$\omega^2 = \omega_n^2 = n \int_{-\infty}^{\infty} [\mathbb{F}_n(x) - F(x)]^2 dF(x).$$

La statistique  $\omega^2$  est aussi très intéressante à cause de ses liaisons avec d'autres statistiques, bien connues en statistique, par exemple, avec la statistique  $L_n$  de Lehmann, la statistique  $G$  de Gini, la statistique " $\sigma$ " de Downton, la statistique  $M_n$  de Moran et Greenwood (pour plus de détails voir, par exemple, Kendall et Stewart, Cramer, Mises). Pour démontrer ces propriétés de la statistique  $\omega^2$ , on peut l'écrire sous une autre forme, beaucoup plus pratique dans les applications :

$$\omega^2 = \sum_{i=1}^n \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n},$$

où  $\mathbb{X}^{(\cdot)} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})^T$  est le vecteur des statistiques d'ordre, associé à l'échantillon  $\mathbb{X}$ .

En fait, on a

$$\begin{aligned}\omega_n^2 &= n \int_{-\infty}^{\infty} [\mathbb{F}_n(x) - F(x)]^2 dF(x) = n \sum_{i=0}^n \int_{X(i)}^{X(i+1)} \left[ \frac{i}{n} - F(x) \right]^2 dF(x) = \\ &= \frac{n}{3} \sum_{i=0}^n \int_{X(i)}^{X(i+1)} d \left[ F(x) - \frac{i}{n} \right]^3 = \frac{n}{3} \sum_{i=0}^n \left[ F(X(i+1)) - \frac{i}{n} \right]^3 - \left[ F(X(i)) - \frac{i}{n} \right]^3.\end{aligned}$$

On suppose que :

$$\mathbf{P} \{ F(X_{(n+1)}) = 1 \} = 1 \quad \text{et} \quad \mathbf{P} \{ F(X_{(0)}) = 0 \} = 0.$$

Comme

$$F(x) - \frac{i}{n} = F(x) - \frac{i+1}{n} + \frac{1}{n}$$

alors

$$\begin{aligned}\left[ F(X_{(i+1)}) - \frac{i}{n} \right]^3 &= \left[ F(X_{(i+1)}) - \frac{i+1}{n} \right]^3 + \frac{3}{n} \left[ F(X_{(i+1)}) - \frac{i+1}{n} \right]^2 + \\ &+ \frac{3}{n^2} \left[ F(X_{(i+1)}) - \frac{i+1}{n} \right] + \frac{1}{n^3},\end{aligned}$$

on en déduit que

$$\begin{aligned}\omega_n^2 &= \frac{n}{3} \left\{ \sum_{i=1}^{n+1} \left[ F(X_{(i)}) - \frac{i}{n} \right]^3 + \frac{3}{n} \sum_{i=1}^{n+1} \left[ F(X_{(i)}) - \frac{i}{n} \right]^2 + \right. \\ &+ \left. \frac{3}{n^2} \sum_{i=1}^{n+1} \left[ F(X_{(i)}) - \frac{i}{n} \right] + \frac{n+1}{n^3} \sum_{i=1}^n \left[ F(X_{(i)}) - \frac{i}{n} \right]^3 \right\} = \\ &= \frac{n}{3} \left\{ \left( 1 - \frac{n+1}{n} \right)^3 + \frac{3}{n} \sum_{i=1}^n \left[ F(X_{(i)}) - \frac{i}{n} \right]^2 + \frac{3}{n} \left( 1 - \frac{n+1}{n} \right)^2 + \right. \\ &+ \left. \frac{3}{n} \sum_{i=1}^n \left[ F(X_{(i)}) - \frac{i}{n} \right]^2 + \frac{3}{n^2} \left( 1 - \frac{n+1}{n} \right) + \frac{n+1}{n^3} \right\} = \\ &= \frac{n}{3} \left\{ \frac{1}{n^2} + \frac{3}{n} \sum_{i=1}^n \left\{ \left[ F(X_{(i)}) - \frac{i}{n} \right]^2 + \frac{1}{n} \left[ F(X_{(i)}) - \frac{i}{n} \right] + \frac{1}{4n^2} \right\} - \frac{1}{4n^2} \right\} = \\ &= \frac{n}{3} \left\{ \frac{3}{n} \sum_{i=1}^n \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{1}{4n^2} \right\} = \\ &= \sum_{i=1}^n \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}.\end{aligned}$$

Donc si les éléments  $X_i$  de l'échantillon  $\mathbb{X}$  sont des variables continues, des propriétés de la transformation de Smirnov il suit que la statistique  $\mathbb{U} = (U_1, \dots, U_n)^T$ ,  $U_i = F(X_i)$ , représente un échantillon, où  $U_i$  suit la loi uniforme sur  $[0, 1]$ . Si nous notons  $\mathbb{U}^{(\cdot)} = (U_{(1)}, U_{(2)}, \dots, U_{(n)})^T$

le vecteur des statistiques d'ordre, associé à la statistique  $\mathbb{U}$ , alors en fonction de  $\mathbb{U}^{(\cdot)}$  la statistique  $\omega^2$  peut être présentée de façon suivante :

$$\omega^2 = \sum_{i=1}^n \left[ U_{(i)} - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}.$$

Cette présentation de la statistique  $\omega^2$  montre bien que sa distribution ne dépend pas de  $F(x)$  si  $H_0$  est vraie. Il y a des tables statistiques de la loi limite ( $n \rightarrow \infty$ ) de la statistique  $\omega^2$ , qui a été étudiée par Smirnov (1944) et T.W.Anderson et D.A.Darling (1952).

Nous allons considérer maintenant une modification  $\Omega_n^2$  de la statistique  $\omega_n^2$ , qui d'un côté est très liée avec les statistiques  $L_n$  de Lehmann,  $G$  de Gini, " $\sigma$ " de Downton et  $M_n$  de Moran et Greenwood, et d'un autre côté a une distribution asymptotique très simple sous l'hypothèse  $H_0$ , quand  $n \rightarrow \infty$ , voir, par exemple, Greenwood & Nikulin (1996).

Soit  $\Sigma^{-1}$  la matrice inverse de la matrice de covariance  $\Sigma$  du vecteur  $\mathbb{U}^{(\cdot)}$ . On peut facilement vérifier que

$$\Sigma^{-1} = \|\sigma^{ij}\|,$$

où

$$\sigma^{ij} = \begin{cases} 2(n+1)(n+2), & \text{si } i = j, \\ -(n+1)(n+2), & \text{si } |i-j| = 1, \\ 0, & \text{si } |i-j| \geq 2. \end{cases}$$

Notons  $\Omega_n^2$  la statistique

$$\Omega_n^2 = \left[ \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right]^T \Sigma^{-1} \left[ \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right],$$

que l'on peut écrire :

$$\Omega_n^2 = 2(n+1)(n+2) \left[ \sum_{i=1}^n U_{(i)}^2 - \sum_{i=1}^{n-1} U_{(i)}U_{(i+1)} - U_{(n)} + \frac{n}{2(n+1)} \right].$$

Nous savons que

$$\mathbf{E}\mathbb{U}^{(\cdot)} = \left[ \frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1} \right]^T$$

et que la matrice de covariance de  $\mathbb{U}^{(\cdot)}$  est

$$\mathbf{E} \left( \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right) \left( \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right)^T = \Sigma = \|\sigma_{ij}\|,$$

où

$$\sigma_{ij} = \sigma_{ji} = \mathbf{E} \left( U_{(i)} - \frac{i}{n+1} \right) \left( U_{(j)} - \frac{j}{n+1} \right) = \begin{cases} \frac{i(n-j-1)}{(n+1)^2(n+2)}, & \text{si } i \leq j, \\ \frac{j(n-i+1)}{(n+1)^2(n+2)}, & \text{si } i \geq j, \end{cases}$$

En utilisant ces propriétés de la statistique  $\mathbb{U}^{(\cdot)}$ , on peut montrer que

$$\mathbf{E}\omega^2 = \frac{1}{6} \quad \text{et} \quad \mathbf{Var}\omega^2 = \frac{4n-3}{180},$$

et qu'on a la représentation suivante pour la statistique  $\omega^2$  :

$$\omega^2 = L_n + \Psi_n + \frac{1}{6(n+1)},$$

où

$$L_n = \sum_{i=1}^n \left( U_{(i)} - \frac{i}{n+1} \right)^2 = \left( \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right)^T \left( \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right)$$

est la statistique de Lehmann (1973),  $\Psi_n$  étant une combinaison linéaire des statistiques d'ordre :

$$\Psi_n = \sum_{i=1}^n \frac{n-2i+1}{n(n+1)} \left( U_{(i)} - \frac{i}{n+1} \right).$$

Par des calculs directs, on peut montrer (voir, par exemple, Nikulin et Osidze (1985)), que

$$\mathbf{E}L_n = \frac{n}{6(n+1)}, \quad \mathbf{Var}L_n = \frac{n^2}{45(n+1)^2},$$

$$\mathbf{E}\Psi_n = 0, \quad \mathbf{Var}\Psi_n = \frac{(n-1)(n+3)}{180n(n+1)^3},$$

$$\mathbf{Cov}(L_n, \Psi_n) = \frac{n-1}{90(n+1)^2}, \quad \mathbf{Corr}(L_n, \Psi_n) = \sqrt{\frac{n-1}{n(n+3)}}.$$

De plus on peut facilement vérifier que  $\Psi_n$  est liée par la relation suivante

$$\Psi_n = \frac{n-1}{2(n+1)}G + \frac{n-1}{6(n+1)}$$

à la statistique  $G$  de Gini :

$$G = \frac{1}{n(n-1)} \sum_{i,j} |U_{(i)} - U_{(j)}|,$$

qui à son tour est liée à la statistique " $\sigma$ " de Downton :

$$''\sigma'' = \frac{\pi}{2}G,$$

et par conséquent on trouve que

$$\omega^2 = L_n + \frac{n-1}{2(n+1)}G + \frac{n}{6(n+1)} = L_n + \sqrt{\pi} \frac{n-1}{n+1} ''\sigma'' + \frac{n}{6(n+1)}.$$

Nous allons considérer maintenant une modification  $\Omega_n^2$  de la statistique  $\omega^2$ . Soit  $\Sigma^{-1}$  la matrice inverse de la matrice de covariance  $\Sigma$  du vecteur  $\mathbb{U}^{(\cdot)}$ . On peut facilement vérifier que

$$\Sigma^{-1} = \|\sigma^{ij}\|,$$

où

$$\sigma^{ij} = \begin{cases} 2(n+1)(n+2), & \text{si } i = j, \\ -(n+1)(n+2), & \text{si } |i-j| = 1, \\ 0, & \text{si } |i-j| \geq 2. \end{cases}$$

Notons  $\Omega_n^2$  la statistique *omega-deux généralisée*

$$\Omega_n^2 = \left[ \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right]^T \boldsymbol{\Sigma}^{-1} \left[ \mathbb{U}^{(\cdot)} - \mathbf{E}\mathbb{U}^{(\cdot)} \right],$$

que l'on peut écrire de la manière suivante :

$$\Omega_n^2 = 2(n+1)(n+2) \left[ \sum_{i=1}^n U_{(i)}^2 - \sum_{i=1}^{n-1} U_{(i)}U_{(i+1)} - U_{(n)} + \frac{n}{2(n+1)} \right].$$

En utilisant cette représentation de la statistique  $\Omega_n^2$ , on peut montrer que

$$\mathbf{E}\Omega_n^2 = n, \quad \mathbf{Var}\Omega_n^2 = \frac{4n(n+1)^2}{(n+3)(n+4)}, \quad \mathbf{E} \left[ \Omega_n^2 - n \right]^3 = \frac{16n(n+1)^2(5n-2)}{(n+3)(n+4)(n+5)(n+6)}.$$

De plus de cette dernière présentation de la statistique  $\Omega_n^2$  il suit que

$$\Omega_n^2 = (n+1)(n+2)M_n - (n+2),$$

où

$$M_n = \sum_{i=1}^n [U_{(i+1)} - U_{(i)}]^2$$

est la statistique de Moran-Greenwood (voir, par exemple, Moran (1947)). La liaison directe entre les statistique  $M_n$  et  $\Omega_n^2$  et leurs propriétés nous permet d'affirmer que

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \sqrt{\left(1 + \frac{3}{n}\right) \left(1 + \frac{3}{n+1}\right)} \frac{\Omega_n^2 - n}{2\sqrt{n+1}} < x | H_0 \right\} = \Phi(x), \quad x \in \mathbb{R}^1.$$

Donc pour tester  $H_0$  on peut utiliser la normalité asymptotique de la statistique  $\Omega_n^2$ .

Parlons maintenant de la statistique de Sherman (1950), qui est liée avec les statistiques considérées dans ce paragraphe.

Soit  $\mathbb{U} = (U_1, \dots, U_n)^T$  un échantillon, où  $U_i$  suit la loi uniforme sur  $[0, 1]$ . Comme précédemment, notons

$$\mathbb{U}^{(\cdot)} = (U_{(1)}, \dots, U_{(n)})^T \tag{3.1}$$

le vecteur des statistiques d'ordre, associé à la statistique  $\mathbb{U}$ .

Notons

$$U_{(0)} \equiv 0 \quad \text{et} \quad U_{(n+1)} \equiv 1. \tag{3.2}$$

Nous déterminons la statistique de Sherman  $s_n$  par la formule

$$s_n = \frac{1}{2} \sum_{i=1}^{n+1} \left| D_i - \frac{1}{n+1} \right|, \tag{3.3}$$

où

$$D_i = U_{(i)} - U_{(i-1)}. \tag{3.4}$$

On sait que

$$\mathbf{E}s_n = \left[ 1 - \frac{1}{n+1} \right]^{n+1} \tag{3.5}$$

et

$$\mathbf{Var}s_n = \frac{2n^{n+2} + n(n-1)^{n+2}}{(n+2)(n+1)^{n+2}} - \left[1 - \frac{1}{n+1}\right]^{2(n+1)}. \quad (3.6)$$

En utilisant ces propriétés de la statistique  $s_n$ , on déduit que

$$\mathbf{E}s_n \rightarrow \frac{1}{e} \quad \text{et} \quad \mathbf{Var}s_n \rightarrow \frac{e-1}{e^2}, \quad n \rightarrow \infty.$$

D'après le théorème limite centrale, si  $n$  est assez grand,

$$\mathbf{P}\left\{\frac{es_n - 1}{\sqrt{e-1}} \leq x\right\} = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right),$$

ce qui signifie que la statistique de Sherman est asymptotiquement normale  $N(0, 1)$ , et donc la statistique

$$X^2 = \frac{(es_n - 1)^2}{e - 1}$$

suit à la limite ( $n \rightarrow \infty$ ) la loi du chi-deux à un degré de liberté, et on peut utiliser ce résultat pour tester l'hypothèse  $H_0$  selon laquelle  $U_i$  suit une loi uniforme sur  $[0, 1]$ .

### 3.7 Les statistiques de Kolmogorov et Gihman.

Soit  $\mathbb{U} = (U_1, U_2, \dots, U_n)^T$  un échantillon,  $U_i$  suit une loi uniforme sur  $[0, 1]$ ,

$$\mathbf{P}\{U_i \leq x\} = x, \quad x \in [0, 1]. \quad (3.1)$$

Notons  $\mathbb{U}^{(\cdot)} = (U_{(1)}, \dots, U_{(n)})^T$  le vecteur des statistiques d'ordre, associé à la statistique  $\mathbb{U}$  :

$$0 \equiv U_{(0)} \leq U_{(1)} \leq \dots \leq U_{(n-1)} \leq U_{(n)} \equiv 1. \quad (3.2)$$

Soit  $\mathbb{F}_n(x)$  la fonction de répartition de la loi empirique associée à  $\mathbb{U}$  :

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[U_i \leq x]}, \quad x \in [0, 1]. \quad (3.3)$$

Il est facile de montrer (voir, par exemple §10) que pour tout  $x$  donné,  $x \in [0, 1]$ , la statistique  $n\mathbb{F}_n(x)$  suit la loi binomiale  $B(n, x)$  de paramètres  $n$  et  $x$  et par conséquent on a :

$$\mathbf{E}\mathbb{F}_n(x) = x \quad \text{et} \quad n\mathbf{Cov}(\mathbb{F}_n(x), \mathbb{F}_n(y)) = x \wedge y - xy, \quad 0 \leq x, y \leq 1; \quad (3.4)$$

$$\mathbb{F}_n(x) \rightarrow x \quad \text{avec la probabilité 1 pour tout } x \text{ quand } n \rightarrow \infty.$$

Dans la pratique il faut avoir beaucoup d'observations pour utiliser la fonction empirique  $\mathbb{F}_n(x)$ . Pour cette raison on peut raisonnablement considérer la situation avec des données groupées. Il est intéressant d'étudier la conduite de la fonction de répartition de la loi empirique  $\mathbb{G}_n(x)$ , correspondant aux données groupées.

Soit  $\mathbf{p} = (p_1, p_2, \dots, p_r, p_{r+1})^T$  un vecteur de probabilités positives,

$$p_i > 0, \quad p_1 + p_2 + \dots + p_r + p_{r+1} = 1, \quad (3.5)$$

où  $r(n) \geq 1$ . Posons  $x_0 = 0, x_{r+1} = 1$ ,

$$x_j = p_1 + p_2 + \dots + p_j, j = 1, \dots, r.$$

On obtient ainsi une partition de  $[0,1]$  en  $r + 1$  intervalles

$$[0, x_1], (x_1, x_2], \dots, (x_{r-1}, x_r], (x_r, x_{r+1}]. \quad (3.6)$$

Soit  $\mathbf{v} = (v_1, \dots, v_r, v_{r+1})^T$  le vecteur des fréquences obtenues en regroupant  $U_1, \dots, U_n$  dans les classes (6). Nous déterminons la fonction de répartition empirique  $\mathbb{G}_n(x)$  associée au vecteur  $\mathbf{v}$  par la formule :

$$\mathbb{G}_n(x) = \begin{cases} 0, & x = x_0 = 0, \\ \frac{v_1 + v_2 + \dots + v_i}{n}, & x_{i-1} < x \leq x_i, \quad i = 1, 2, 3, \dots, r+1. \end{cases} \quad (3.7)$$

Nous pouvons maintenant construire la statistique de Gihman

$$\mathbf{Z}_n = (Z_{n1}, \dots, Z_{nr})^T,$$

où

$$Z_{ni} = \sqrt{n} [\mathbb{G}_n(x_i) - x_i] = \sqrt{n} \left[ \frac{v_1 + \dots + v_i}{n} - (p_1 + \dots + p_i) \right]. \quad (3.8)$$

Il est clair que

$$\mathbf{E}\mathbf{Z}_n = (0, \dots, 0)^T = \mathbf{0}_r \quad \text{et} \quad \mathbf{E}\mathbf{Z}_n \mathbf{Z}_n^T = \mathbf{\Sigma}, \quad (3.9)$$

où

$$\mathbf{\Sigma} = \begin{pmatrix} x_1 & x_1 & x_1 & \cdots & x_1 \\ x_1 & x_2 & x_2 & \cdots & x_2 \\ x_1 & x_2 & x_3 & \cdots & x_3 \\ \vdots & \vdots & \vdots & & \vdots \\ x_1 & x_2 & x_3 & \cdots & x_r \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_r \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \dots & x_r \end{pmatrix}. \quad (3.10)$$

Nous allons étudier les propriétés asymptotiques de la statistique  $\mathbf{Z}_n$  quand  $n \rightarrow \infty$ .

a) Supposons tout d'abord que

$$r = r(n) \rightarrow \infty \quad \text{quand} \quad n \rightarrow \infty \quad (3.11)$$

de façon que la longueur maximale des intervalles (6) de groupement des données aille vers zéro assez vite, i.e., que

$$\max_{1 \leq i \leq r+1} np_i \rightarrow 0 \quad \text{si} \quad n \rightarrow \infty. \quad (3.12)$$

Notons

$$D_n^* = \max_{1 \leq i \leq r} |Z_{n,i}| \quad \text{et} \quad D_n = \sup_{0 \leq x \leq 1} \sqrt{n} |\mathbb{F}_n(x) - x|.$$

**Théorème** (Gihman, 1961). *Si  $r \rightarrow \infty$  et que (12) est vérifiée quand  $n \rightarrow \infty$ , alors les statistiques  $D_n$  et  $D_n^*$  sont asymptotiquement équivalentes :*

$$\lim_{n \rightarrow \infty} \mathbf{P}\{D_n^* \leq z\} = \lim_{n \rightarrow \infty} \mathbf{P}\{D_n \leq z\} = K(z), \quad (3.13)$$

où  $K(z)$  est la fonction de répartition de Kolmogorov,

$$K(z) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 z^2}, \quad 0 < z < \infty.$$

De ce théorème il suit que sous la condition (12) nous pouvons utiliser la statistique  $\mathbf{Z}_n$  quand  $n$  est assez grand pour construire des tests bien connus comme  $\omega^2$  de Smirnov,  $W_n^2$  de Anderson et Darling (1952) ou de Sherman (1950) etc.

b) Maintenant nous supposons que les intervalles (6) sont fixés,  $r+1 \geq 2$ . Dans ce cas de (3), (4), (8) et du théorème limite central multidimensionnel on déduit que la loi limite de  $\{\mathbf{Z}_n\}$  quand  $n \rightarrow \infty$  est la loi normale  $N(\mathbf{0}_r, \mathbf{\Sigma})$  de paramètres donnés par (9). Comme le rang de la matrice de covariance  $\mathbf{\Sigma}$  est égale à  $r$ , on en déduit qu'il existe une matrice

$$\mathbf{\Sigma}^{-1} = \|\sigma^{ij}\|$$

dont les éléments  $\sigma^{ij}$  sont donnés par la formule suivante :

$$\left\{ \begin{array}{ll} \sigma^{ij} = 0, & |i-j| \geq 2, \\ \sigma^{i,i+1} = -\frac{1}{x_{i+1}-x_i} = -\frac{1}{p_{i+1}}, & i = 1, \dots, r-1, \\ \sigma^{i,i-1} = -\frac{1}{x_i-x_{i-1}} = -\frac{1}{p_i}, & i = 1, \dots, r, \\ \sigma^{ii} = -(\sigma^{i,i-1} + \sigma^{i,i+1}) = \frac{1}{x_{i+1}-x_i} + \frac{1}{x_i-x_{i-1}}, & i = j. \end{array} \right. \quad (3.14)$$

Nous pouvons maintenant construire la statistique  $Y_n^2$  en posant

$$Y_n^2 = \mathbf{Z}_n^T \mathbf{\Sigma}^{-1} \mathbf{Z}_n.$$

Grâce à la normalité asymptotique de la statistique  $\mathbf{Z}_n$  on obtient que

$$\lim_{n \rightarrow \infty} \mathbf{P}\{Y_n^2 \leq x\} = \mathbf{P}\{\chi_r^2 \leq x\}.$$

Il est facile de vérifier que  $Y_n^2$  est la statistique classique de Pearson :

$$Y_n^2 = \sum_{i=1}^{r+1} \frac{(v_i - np_i)^2}{np_i}. \quad (3.15)$$

c) Enfin nous considérons le cas

$$r = r(n) \rightarrow -\infty \quad \text{quand} \quad n \rightarrow -\infty, \quad (3.16)$$

de façon que

$$\max_{1 \leq i \leq r+1} p_i \rightarrow 0 \quad \text{et} \quad \min_{1 \leq i \leq r+1} np_i \rightarrow \infty. \quad (3.17)$$

**Théorème** (Tumanian, 1956). *Si  $r \rightarrow \infty$  et si les conditions (7) ont lieu quand  $n \rightarrow \infty$ , alors*

$$\sup_{|x| < \infty} \left| \mathbf{P}\{Y_n^2 \geq x\} - 1 + \Phi\left(\frac{x-r}{\sqrt{2r}}\right) \right| \rightarrow 0, \quad n \rightarrow \infty. \quad (3.18)$$

### 3.8 Test des signes.

Soit  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  un échantillon. On suppose que la fonction de répartition  $F(x) = \mathbf{P}\{X_i \leq x\}$  de  $X_i$  est continue, mais inconnue.

Soit  $\mu$  la médiane inconnue, elle aussi de la loi  $F(x)$ , c'est-à-dire que

$$F(\mu) = 0.5,$$

et supposons que nous voulions tester l'hypothèse  $H_0 : \mu = \mu_0$ , où  $\mu_0$  est un nombre donné, contre l'une des trois hypothèses suivantes :

$$H_1^+ : F(\mu_0) > 0.5, \quad \text{ce qui signifie que } \mu_0 > \mu;$$

$$H_1^- : F(\mu_0) < 0.5, \quad \text{ce qui signifie que } \mu_0 < \mu;$$

$$H_1 : F(\mu_0) \neq 0.5, \quad \text{ce qui signifie que } \mu_0 \neq \mu.$$

Le test des signes est fondé sur la statistique

$$v_n = Y_1 + Y_2 + \dots + Y_n,$$

où

$$Y_i = \begin{cases} 1, & \text{si } X_i > \mu_0, \\ 0, & \text{si } X_i \leq \mu_0. \end{cases}$$

Il est évident que

$$\mathbf{P}\{Y_i = 1|H_0\} = F(\mu_0) = 1 - F(\mu_0) = 0.5,$$

i.e. sous l'hypothèse  $H_0$  la statistique  $Y_i$  suit une loi de Bernoulli de paramètre de succès  $p = 0.5$ , et par conséquent la statistique  $v_n$  sous l'hypothèse  $H_0$  suit une loi binomiale de paramètres  $n$  et  $p = 0.5$  :

$$\begin{aligned} \mathbf{P}\{v_n \leq m|H_0\} &= W(m, n) = \sum_{i=0}^m \binom{n}{i} (0.5)^n = \\ &= I_{0.5}(n - m, m + 1) = 1 - I_{0.5}(m + 1, n - m). \end{aligned} \quad (3.1)$$

Donc pour avoir le test, il faut trouver des nombres entiers  $k$  et  $K$  tels que

$$\begin{cases} W(k, n) \leq \alpha, \\ W(k + 1, n) > \alpha, \end{cases} \quad \text{et} \quad \begin{cases} W(K - 1, n) \geq 1 - \alpha, \\ W(K - 2, n) < 1 - \alpha, \end{cases} \quad (3.2)$$

où  $\alpha$  est une probabilité inférieure à 0.5,  $0 < \alpha < 0.5$ .

Il est évident que les valeurs critiques  $k = k(\alpha, n)$  et  $K = K(\alpha, n)$  sont des fonctions non décroissantes de  $n$ , et que, si la fonction  $F(x)$  est continue, alors  $k + K = n$ . Si on teste  $H_0$  contre  $H_1^+$ , alors on est obligé de rejeter  $H_0$  en faveur de  $H_1^+$ , si

$$v_n \leq k(\alpha, n), \quad (3.3)$$

et dans ce cas on a le test des signes de niveau  $\leq \alpha$ . On procède de même si on teste  $H_0$  contre  $H_1^-$ , en rejetant  $H_0$  en faveur de  $H_1^-$  si

$$v_n \geq K(\alpha, n) \quad (3.4)$$

et le niveau de ce test est  $\leq \alpha$ . Dans le cas où on teste  $H_0$  contre l'alternative  $H_1$ , on est obligé de rejeter  $H_0$  en faveur de  $H_1$ , si

$$\min(v_n, n - v_n) \leq k(\alpha, n), \quad (3.5)$$

et le niveau de ce test est  $\leq 2\alpha$ .

**Exemple 1.** Pendant le premier jour, un compteur a enregistré 20021 impulsions, tandis que le jour suivant il y en a eu seulement 19580. Peut-on dire que le second jour on a observé

une diminution de l'intensité d'arrivée des impulsions ? Pour répondre à cette question on choisit le modèle statistique d'après lequel les nombres d'impulsions observées sont des réalisations de deux variables indépendantes  $X$  et  $Y$  où  $X$  suit la loi de Poisson de paramètre  $\lambda$  ( $\lambda > 0$ ) et  $Y$  suit la loi de Poisson de paramètre  $\mu$  ( $\mu > 0$ ). Dans ce modèle il est bien naturel de considérer comme hypothèse  $H_0 : \lambda = \mu$ , et comme alternative  $H_1 : \lambda > \mu$ . Pour tester  $H_0$  contre  $H_1$  on peut utiliser le test des signes.

Si notre modèle est bon, alors pour tout  $x, y \in \{0, 1, 2, \dots\}$

$$\begin{aligned} \mathbf{P}\{X = x, Y = y\} &= \frac{\lambda^x}{x!} e^{-\lambda} \frac{\mu^y}{y!} e^{-\mu} = \\ &= \frac{(\lambda + \mu)^{x+y}}{(x+y)!} e^{-(\lambda+\mu)} \frac{(x+y)!}{x!y!} \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(1 - \frac{\lambda}{\lambda+\mu}\right)^y, \end{aligned}$$

et donc la loi conditionnelle de  $X$ , conditionnée par la somme  $X + Y = n$ , est binomiale de paramètres  $n$  et  $p = \lambda/(\lambda + \mu)$ , et par conséquent on en tire que l'hypothèse  $H_0 : \lambda = \mu$  est vraie si et seulement si la loi conditionnelle de  $X$  est binomiale de paramètres  $n$  et  $p = 0.5$  :

$$\mathbf{P}\{X = x | X + Y = n, H_0\} = \binom{n}{x} (0.5)^n,$$

et il nous faut tester l'hypothèse  $H_0 : p = 0.5$  contre une alternative  $H_1 : p > 0.5$ . On peut montrer que c'est le test des signes qui est le plus puissant dans ce problème. D'après ce test on doit rejeter  $H_0$ , si  $X \geq K = K(\alpha, n)$ , où  $n = 20021 + 19580 = 39601$ . La valeur critique  $K$  est déterminée comme étant la solution du système

$$\begin{cases} \mathbf{P}\{X \geq K | X + Y = 39601, p = 0.5\} \leq \alpha, \\ \mathbf{P}\{X \geq K - 1 | X + Y = 39601, p = 0.5\} > \alpha. \end{cases}$$

Mais d'après le théorème de de Moivre-Laplace

$$\mathbf{P}\{X \geq K | X + Y = n, p = 0.5\} \cong \Phi\left(\frac{K - 0.5n - 0.5}{\sqrt{0.25n}}\right),$$

donc

$$K = \begin{cases} K^*, & \text{si } K^* \text{ est entier,} \\ [K^* + 1], & \text{si } K^* \text{ est nonentier,} \end{cases}$$

où

$$K^* = \frac{n+1}{2} + \Psi(1-\alpha) \frac{\sqrt{n}}{2}.$$

Dans notre cas,  $\alpha = 0.05$  et

$$K^* = \frac{39602}{2} + 1.645 \frac{\sqrt{39601}}{2} = 19964.7,$$

par conséquent  $K = 19965$ . Comme

$$X = 20021 > 19965,$$

on prend l'hypothèse  $H_1$ , d'après laquelle on observe diminution d'intensité.

**Exemple 2.** Soit  $\mathbb{Z} = (Z_1, \dots, Z_n)^T$  un échantillon,  $Z_i = (X_i, Y_i)^T$  est un vecteur aléatoire à deux dimensions dont la densité  $p(x, y)$  est inconnue. Supposons que pour tout  $i$ ,  $X_i$  et  $Y_i$  soient indépendantes et qu'il faille tester l'hypothèse

$$H_0 : p(x, y) = p(y, x). \quad (3.6)$$

Comme les  $X_i$  sont indépendantes de  $Y_i$ , la condition (6) signifie que  $X_i$  et  $Y_i$  sont distribuées d'après la même loi (inconnue), et par conséquent pour tester  $H_0$  on peut construire le test des signes. En fait, soit

$$V_i = \begin{cases} 1, & \text{si } X_i - Y_i > 0, \\ 0, & \text{si } X_i - Y_i < 0, \end{cases} \quad i = 1, 2, \dots, n. \quad (3.7)$$

Sous l'hypothèse  $H_0$  la distribution de  $V_i$  est symétrique par rapport à 0, et donc si nous posons

$$v_n = V_1 + V_2 + \dots + V_n,$$

de (6) et (7) il s'ensuit que sous l'hypothèse  $H_0$  la statistique  $v_n$  est distribuée selon la loi (1) donc en utilisant (2)–(5) nous pouvons utiliser le test des signes pour tester cette hypothèse.

### 3.9 Test de Wilcoxon.

Soient  $\mathbb{X} = (X_1, \dots, X_n)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_m)^T$  deux échantillons indépendants, et soit

$$F(x) = \mathbf{P}\{X_i \leq x\} \quad \text{et} \quad G(y) = \mathbf{P}\{Y_j \leq y\}$$

les fonctions de répartition de  $X_i$  et  $Y_j$  respectivement. Le test de Wilcoxon est utilisé pour tester l'hypothèse

$$H_0 : F(x) \equiv G(x), \quad x \in \mathbf{R}^1,$$

contre l'hypothèse

$$H_- : F(x) < G(x), \quad x \in \mathbf{R}^1,$$

ou contre l'hypothèse

$$H_+ : F(x) > G(x), \quad x \in \mathbf{R}^1,$$

ou contre  $H_-$  et  $H_+$  ensemble.

Ce test est fondé sur la statistique linéaire des rangs

$$W = W_{n,m} = \sum_{i=1}^m R_i,$$

où

$$R_1 < R_2 < R_3 < \dots < R_m$$

sont les rangs des observations  $Y_1, \dots, Y_m$  dans l'échantillon unifié

$$\mathbb{Z} = (X_1, \dots, X_n, Y_1, \dots, Y_m)^T$$

de taille  $n + m$ . Pour construire le vecteur  $\mathbb{R}_Y = (R_1, R_2, \dots, R_m)^T$  des rangs des observations  $Y_j$ , il faut construire le vecteur  $\mathbb{Z}^{(\cdot)}$  des statistiques d'ordre, associé à l'échantillon  $\mathbb{Z}$ , et déterminer les numéros des positions des variables aléatoires  $Y_j$ . Si, par exemple, l'hypothèse  $H_-$  est vraie, on dit que les variables aléatoires  $Y_j$  sont stochastiquement plus grandes que les variables aléatoires  $X_i$ , ce qui signifie en pratique que les variables aléatoires  $Y_j$  ont tendance (sous l'hypothèse  $H_-$ ) à prendre des positions à l'extrémité droite du vecteur des statistiques d'ordre  $\mathbb{Z}$  et par conséquent leurs rangs  $R_i$  ont tendance à avoir de grandes valeurs, et par suite la statistique de Wilcoxon a tendance à prendre de grandes valeurs, ce que l'on utilise pour tester  $H_0$  contre  $H_-$ , en rejetant  $H_0$  en faveur de  $H_-$  quand  $W > c_\alpha$ , où  $c_\alpha$  est la valeur critique du test de Wilcoxon. On peut montrer que

$$W = U + \frac{n(n+1)}{2},$$

où

$$U = U_{m,n} = \sum_{i=1}^n \sum_{j=1}^m V_{ij}, \quad (3.1)$$

est la statistique de Mann-Whitney,

$$V_{ij} = \begin{cases} 1, & \text{si } Y_j > X_i, \\ 0, & \text{si } Y_j < X_i. \end{cases} \quad (3.2)$$

Par des calculs directs (mais pas simples !) on peut montrer que

$$\mathbf{E}\{W|H_0\} = \frac{m(N+1)}{2} \quad \text{et} \quad \mathbf{Var}\{W|H_0\} = \frac{mn(N+1)}{12},$$

où  $N = n + m$ . Les valeurs critiques  $c_\alpha$  de niveau  $\alpha$  ( $0 < \alpha < 0.5$ ) de la statistique  $W$  sont des nombres entiers, qui satisfont aux inégalités

$$\mathbf{P}\{W \leq c_\alpha | H_0\} \leq \alpha \quad \text{et} \quad \mathbf{P}\{W \leq c_\alpha + 1 | H_0\} > \alpha.$$

Pour les calculer on utilise, par exemple, les tables statistiques de Verdooren (1963) pour

$$m = 1(1)25, \quad n = m(1)25 \quad \text{et} \quad \alpha = 0.001, 0.005, 0.010, 0.025, 0.05, 0.1.$$

Comme la distribution de la statistique  $W$  est symétrique par rapport à son espérance mathématique  $\mathbf{E}W$ , pour calculer une valeur critique  $c_{1-\alpha}$ ,  $0 < \alpha < 0.5$ , on utilise la relation suivante :

$$c_{1-\alpha} = \mathbf{E}W - c_\alpha.$$

Il est évident que le couple  $(c_\alpha, c_{1-\alpha})$  nous donne les valeurs critiques du test bilatéral de Wilcoxon de niveau  $2\alpha$ , que l'on utilise pour tester  $H_0$  contre  $H_+$  et  $H_-$  à la fois.

Si l'un des deux nombres  $n$  ou  $m$  est supérieur à 25, pour calculer les valeurs critiques du test de Wilcoxon, on utilise l'approximation normale de Mann et Whitney (1947), d'après laquelle

$$\mathbf{P}\left\{\frac{W - \mathbf{E}W}{\sqrt{\mathbf{Var}W}} < w | H_0\right\} \rightarrow \Phi(w),$$

quand  $\min(m, n) \rightarrow \infty$ ,  $|w| < \infty$ .

Fix et Hodges (1955) ont donné une autre approximation, qui donne déjà de bons résultats quand  $\min(m, n) \geq 5$ . D'après cette approximation

$$\mathbf{P}\{W \leq w | H_0\} \cong \Phi(x) + \phi(x)(x^3 - 3x) \frac{N^2 + N - mn}{20mn(N+1)},$$

où

$$N = m + n \quad \text{et} \quad x = \frac{w - \mathbf{E}W + 0.5}{\sqrt{\mathbf{Var}W}}.$$

Ce résultat permet d'obtenir assez facilement des approximations normales pour des valeurs critiques  $c_\alpha$  :

$$c_\alpha \cong \left( \frac{m(N+1) - 1}{2} - \Psi(1 - \alpha) \sqrt{\frac{mn(N+1)}{12}} \right),$$

où  $[x]$  dénote la partie entière du nombre  $x$ . On remarque ici que tous ces résultats, liés avec des approximations, sont valables si parmi les  $X_i$  et  $Y_j$  il n'y a pas d'ex aequo. En principe, on ne devrait pas en avoir, puisque  $X_i$  et  $Y_j$  sont des variables aléatoires continues et par conséquent  $\mathbf{P}\{X_i = Y_j\} = 0$ . Mais à cause des erreurs d'arrondis, on obtient souvent des observations égales. Dans ce cas on attribue aux observations qui sont des ex aequo, un rang égal à la moyenne arithmétique des rangs que ces observations auraient eu avant la procédure d'arrondissement. Notons  $W^* = W_{n,m}^*$  la statistique de Wilcoxon dans ce cas. L'opération d'arrondissement ne change pas  $\mathbf{E}W$ ,  $\mathbf{E}W = \mathbf{E}W^*$ , mais elle change la variance. Par des calculs directs, on peut montrer qu'alors :

$$\mathbf{Var}W_{n,m}^* = \frac{nm}{12}(N+1) \left( 1 - \frac{\sum_{i=1}^M t_i(t_i^2 - 1)}{N(N^2 - 1)} \right),$$

où  $t_i$  est le nombre d'ex aequo dans le groupe numéro  $i$  et  $M$  est le nombre des groupes d'ex aequo.

Démonstration.

Soient  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$  des variables aléatoires continues,  $X_i$  suit une loi dont la fonction de répartition est  $F(x)$  et  $Y_j$  suit une loi dont la fonction de répartition est  $G(x)$  avec, par exemple,  $G(x) = F(x - \theta)$ . Supposons que l'on teste l'hypothèse  $H_0$ , contre l'hypothèse  $H_-$ . Donc si  $H_0$  est vraie, alors les variables aléatoires

$$X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$$

forment un échantillon

$$\mathbb{Z} = (X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)^T$$

de taille  $N = n + m$ . On remarque que

$$\mathbf{P}\{X_i = Y_j\} = 0,$$

car  $X_i$  et  $Y_j$  sont continues, mais à cause des erreurs d'arrondi on a des ex aequo.

Tout d'abord, on remarque que comme

$$W = W_{n,m} = U_{n,m} + \frac{n(n+1)}{2},$$

alors  $\mathbf{Var}W = \mathbf{Var}U_{n,m}$ .

Supposons que le vecteur  $\mathbb{Z}^{(\cdot)}$  des statistiques d'ordre ait au moins un groupe de statistiques d'ordre qui soient égales et que les rangs de ces ex aequo dans ce groupe soient

$$k+1, k+2, \dots, k+t.$$

Soit  $\mu$  le nombre des  $X_i$  de ce groupe, alors  $t - \mu$  est le nombre des  $Y_j$  parmi ces  $t$  ex-aequo. Il est clair que  $\mu$  suit la loi hypergéométrique :

$$\mathbf{P}\{\mu = x\} = \frac{\binom{n}{x} \binom{m}{t-x}}{\binom{N}{t}}.$$

Pour  $k$  et  $t$  fixés posons

$$U^* = U_{n,m}^*(\mu) = \sum_{i=1}^n \sum_{j=1}^m V_{ij}^*, \quad (3.3)$$

où

$$V_{ij}^* = \begin{cases} 1, & \text{si } X_i > Y_j, \\ 0.5, & \text{si } X_i = Y_j, \\ 0, & \text{si } X_i < Y_j. \end{cases} \quad (3.4)$$

De (1) – (4) il résulte qu'en cas de présence d'un seul groupe d'ex aequo, on a l'identité par rapport à  $\mu$  :

$$U_{n,m}^*(\mu) + U_{\mu,t-\mu} - \frac{\mu(t-\mu)}{2} \equiv W_{n,m}. \quad (3.5)$$

En cas de présence de  $M$  groupes d'ex aequo, la dernière identité peut être généralisée de la façon suivante :

$$U_{n,m}^*(\mu_1, \mu_2, \dots, \mu_M) + \sum_{i=1}^M \left( U_{\mu_i, t_i - \mu_i} - \frac{\mu_i(t_i - \mu_i)}{2} \right) \equiv U_{n,m}, \quad (3.6)$$

où  $t_i$  est le nombre d'ex aequo dans le groupe de numéro  $i$ ,  $\mu_i$  le nombre des  $X_i$  dans ce groupe. De (5) il suit que

$$\mathbf{E}\{U_{n,m}^*(\mu_1, \mu_2, \dots, \mu_M) | \mu_1, \mu_2, \dots, \mu_M\} = \frac{nm}{2}. \quad (3.7)$$

Comme la partie droite de (7) ne dépend pas de  $\mu_i$ , on en tire que

$$\mathbf{E}U_{n,m}^* = \frac{nm}{2}.$$

De la même façon, comme

$$\mathbf{Var}U_{n,m} = \frac{nm}{2}(n+m+1) = \frac{nm(N+1)}{2},$$

on obtient que

$$\mathbf{Var}\{U_{n,m}^*(\mu_1, \mu_2, \dots, \mu_M) | \mu_1, \mu_2, \dots, \mu_M\} + \sum_{i=1}^M \frac{1}{12} \mu_i(t_i - \mu_i)(t_i + 1) =$$

$$= \frac{nm}{12}(n+m+1).$$

Comme

$$\mathbf{Var} \left\{ \mathbf{E} \left\{ U_{n,m}^*(\mu_1, \mu_2, \dots, \mu_M) \mid \mu_1, \mu_2, \dots, \mu_M \right\} \right\} = \mathbf{Var} \frac{nm}{12} = 0,$$

on en tire que

$$\mathbf{Var} U_{n,m}^* = \mathbf{E} \left\{ \mathbf{Var} \left\{ U_{n,m}^* \mid \mu_1, \mu_2, \dots, \mu_M \right\} \right\},$$

donc on en déduit que

$$\mathbf{Var} \left\{ U_{n,m}^*(\mu_1, \mu_2, \dots, \mu_M) \right\} + \sum_{i=1}^M \frac{t_i + 1}{12} \mathbf{E} \left\{ \mu_i(t_i - \mu_i) \right\} = \frac{nm}{12}(n+m+1).$$

Mais

$$\mathbf{E} \left\{ \mu_i(t_i - \mu_i) \right\} = \sum_j \frac{\binom{n}{j} \binom{m}{t_i - j}}{\binom{N}{t_j}} j(t_i - j) = \frac{t_i(t_i - 1)nm}{N(N-1)},$$

donc

$$\mathbf{Var} U^* = \frac{nm}{12}(N+1) \left( 1 - \frac{\sum_{i=1}^M t_i(t_i^2 - 1)}{N(N^2 - 1)} \right) = \mathbf{Var} W^*,$$

où  $N = n + m$ .

### 3.10 Estimation non paramétrique de la densité. Histogramme. Estimateur de Rosenblatt. Le noyau de Parzen.

Le problème, que l'on désigne souvent par *estimation non paramétrique de la densité*, est le suivant :

étant donné un échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$ , issu d'une distribution continue et dont la densité  $f$  est inconnue, construire un bon estimateur de  $f$ .

Soit  $\{h_n\}$  une suite de nombres positives (*tailles de fenêtre*) telle que  $h_n > 0$ ,  $h_n \downarrow 0$ ,  $nh_n \rightarrow 0$ , quand  $n \rightarrow \infty$ . Pour tout  $n$  fixé nous pouvons construire une partition de  $\mathbf{R}^1$

$$\mathbf{R}^1 = \bigcup_{k \in \mathbb{Z}} ]kh_n, (k+1)h_n],$$

en utilisant la taille de fenêtre  $h_n$  correspondante. Pour tout  $x \in \mathbf{R}^1$  il existe un intervalle  $]kh_n, (k+1)h_n]$ , avec  $k = \left\lfloor \frac{x}{h_n} \right\rfloor$ , tel que  $x \in ]kh_n, (k+1)h_n]$  et donc nous pouvons déterminer une application aléatoire  $f_n : \mathbf{R}^1 \rightarrow \mathbf{R}_+^1$  par la formule :

$$f_n(x) = \frac{1}{nh_n} \sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(X_j), \quad x \in \mathbf{R}^1. \quad (3.1)$$

**Définition 1.** Nous disons que  $f_n(x)$ ,  $x \in \mathbf{R}^1$ , est la densité empirique, basée sur l'échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$ . Le graphe de  $f_n(x)$  s'appelle histogramme.

De (1) il suit que pour tout  $x \in ]kh_n, (k+1)h_n]$ ,  $k \in \mathbf{Z}$ , on a

$$f_n(x) = \frac{1}{nh_n} [\mathbb{F}_n((k+1)h_n) - \mathbb{F}_n(kh_n)] = \frac{v_k}{nh_n}, \quad (3.2)$$

où  $\mathbb{F}_n(x)$  est la fonction empirique, basée sur  $\mathbb{X} = (X_1, \dots, X_n)^T$ ,  $v_k$  est le nombre de  $X_j$  dans l'intervalle  $]kh_n, (k+1)h_n]$ . Souvent on dit que  $f_n(x)$  est un estimateur *non paramétrique classique* de la densité  $f(x)$ .

En 1956 M. Rosenblatt a proposé un *estimateur de type noyau*

$$f_n(x) = \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{x - X_j}{h_n}\right), \quad (3.3)$$

où  $K(\cdot)$ , un noyau, est une fonction telle que

$$\int_{-\infty}^{\infty} K(x)dx = 1 \quad \text{et} \quad \int_{-\infty}^{\infty} K^2(x)dx < \infty.$$

Le choix du noyau  $K$  dépend en général des propriétés de la densité  $f$  que l'on désire avoir. Par exemple, Parzen (1962) a proposé de choisir le noyau

$$K(x) = 0.5\mathbf{1}_{[-1,1]}(x), \quad \text{avec} \quad k = \frac{1}{2}. \quad (3.4)$$

Il est clair que si on choisit le *noyau de Parzen*, alors de (1), (2) et (4) on obtient l'estimateur  $f_n(x)$ , appelé l'estimateur naïf de  $f(x)$  :

$$f_n(x) = \frac{v_k}{2nh_n},$$

où  $v_k$  est le nombre de  $X_j$  dans l'intervalle  $]x - h_n, x + h_n]$ .

Souvent on utilise le *noyau de Epanechnikov* (1969)

$$K(x) = 0.72(1 - x^2)\mathbf{1}_{[-1,1]}(x), \quad \text{avec} \quad k = \frac{2}{3},$$

voir aussi Bartlett (1963).

On donne ici encore quelques d'autres exemples :

le *noyau de Gauss* :

$$K(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2}, \quad \text{avec} \quad k = \frac{1}{\sqrt{2\pi}},$$

le *noyau de Laplace* :

$$K(x) = \frac{1}{2}e^{-|x|}, \quad \text{avec} \quad k = \frac{1}{2},$$

le *noyau de Cauchy* :

$$K(x) = \frac{1}{\pi(1+x^2)}, \quad \text{avec} \quad k = \frac{1}{\pi},$$

le noyau de Fejer :

$$K(x) = \frac{1}{2\pi} \left( \frac{\sin \frac{x}{2}}{\frac{x}{2}} \right), \quad \text{avec } k = \frac{1}{3\pi},$$

le noyau de Tukey

$$K(x) = \frac{15}{16} (1 - x^2)^2 \mathbf{1}_{[-1,1]}(x).$$

Dans certains cas l'expression de  $K$  peut être plus compliquée. Les propriétés asymptotiques de  $f_n$  ont été bien étudiées, voir par exemple, Deheuvels (1973, 1974), Devroye et Györfi (1985), Watson et Leadbether (1963), Silverman (1986), Nikulin & Solev (2002), etc.

Il est facile de montrer que pour l'estimateur classique (1) on a

$$|\mathbf{E}f_n(x) - f(x)| \leq \omega_f(h_n),$$

où

$$\omega_f(h) = \sup_{|x-y| \leq h} |f(x) - f(y)|,$$

est le *module de continuité* de  $f$ , d'où on tire que si  $x$  est un point de continuité de  $f$ , alors

$$\mathbf{E}f_n(x) = f(x) + o(h_n), \quad n \rightarrow \infty$$

et donc de la loi de grands nombres il suit que

$$f_n(x) \xrightarrow{\mathbf{P}} f(x),$$

i.e.  $\{f_n(x)\}$  est une suite consistante d'estimateurs  $f_n(x)$  de  $f(x)$ .

De la même façon comme pour l'estimateur non paramétrique *classique* on peut démontrer, sous quelques conditions de régularité sur  $f$  et  $K$ , que pour l'estimateur de type noyau on a :

$$\mathbf{E}f_n(x) = \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{x-y}{h_n}\right) f(y) dy \rightarrow f(x), \quad \text{quand } n \rightarrow \infty,$$

$$\lim n h_n \mathbf{Var} f_n(x) = k f(x), \quad n \rightarrow \infty,$$

i.e.  $f_n(x)$  est un estimateur asymptotiquement sans biais pour  $f(x)$ , et on en tire que  $f_n(x) \xrightarrow{\mathbf{P}} f(x)$ , i.e.  $\{f_n(x)\}$  est une suite consistante d'estimateurs  $f_n(x)$  de  $f(x)$ .

Enfin on remarque que à propos du choix de la taille de la fenêtre  $h_n$  nous recommandons regarder Devroue et Györfi (1985), Bretagnolle et Huber (1979), Freedman et Diaconis (1981). Souvent pour choisir  $h_n$  on pose

$$h_n = \frac{1}{[n \int_{-\infty}^{\infty} [f^{(2)}(x)]^2 dx]^{1/5}} \left[ \frac{k}{\int_{-\infty}^{\infty} x^2 K^2(x) dx} \right]^{2/5}.$$

# Chapitre 4

## TESTS STATISTIQUES.

### 4.1 Principe des tests.

Soit  $\mathbf{X} = (X_1, \dots, X_n)^T$  un vecteur aléatoire,  $\mathbf{X} \in \mathbf{R}^n$ . Faisons l'hypothèse  $H$  sur la distribution de  $\mathbf{X}$  dans  $\mathbf{R}^n$  selon laquelle cette distribution appartient à une famille  $\mathcal{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}$  dans  $\mathbf{R}^n$ , paramétrée par  $\theta$ . On note  $H : \theta \in \Theta$  et l'ensemble  $\Theta$  est appelé *espace des paramètres*.

**Définition 1.** Soit  $\Theta_0 \subset \Theta$ . Nous appelons  $H_0 : \theta \in \Theta_0$  l'hypothèse nulle selon laquelle la distribution de  $\mathbf{X}$  appartient à la famille

$$\mathcal{P}_0 = \{\mathbf{P}_\theta, \theta \in \Theta_0\} \subset \mathcal{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}.$$

**Définition 2.** Si  $\Theta = \{\theta\}$  n'a qu'un seul élément  $\theta$ , i.e. la distribution de  $\mathbf{X}$  est  $\mathbf{P}_\theta$ , alors, on dit que l'hypothèse  $H$  est *simple*, sinon  $H$  est *composée* (ou multiple).

Soient  $\Theta_0 \subset \Theta$  et  $\Theta_1 \subset \Theta$  telles que  $\Theta_0 \cap \Theta_1 = \emptyset$ .

**Définition 3.** L'hypothèse  $H_1 : \theta \in \Theta_1$  est appelée l'*alternative* de  $H_0$ .

**Exemple 1.** Soit

$$\Theta = [\theta_0, \infty[ \subset \mathbf{R}^1, \quad \Theta_0 = \{\theta_0\}, \quad \Theta_1 = \{\theta > \theta_0\}.$$

Dans ce cas l'hypothèse  $H_0 : \theta = \theta_0$ , i.e.  $H_0 : \theta \in \Theta_0$ , est simple, et l'alternative  $H_1 : \theta > \theta_0$ , i.e.  $H_1 : \theta \in ]\theta_0, \infty[$ , est composée. De même, si

$$\Theta = ]-\infty, \theta_0], \quad \Theta_0 = \{\theta_0\}, \quad \text{et } \Theta_1 = ]-\infty, \theta_0[,$$

l'alternative  $H_1 : \theta < \theta_0$  est composée. Dans ces deux cas les alternatives  $H_1 : \theta > \theta_0$  ou  $H_1 : \theta < \theta_0$  sont *unilatérales*.

**Exemple 2.** Soit  $\Theta = ]\theta_1, \theta_2[ \subset \mathbf{R}^1$ ,  $\Theta_0 = \{\theta_0\}$ ,  $\theta_1 < \theta_0 < \theta_2$  et

$$\Theta_1 = \Theta \setminus \Theta_0 = ]\theta_1, \theta_0[ \cup ]\theta_0, \theta_2[.$$

Ici l'alternative  $H_1 : \theta \neq \theta_0$ , i.e.  $H_1 : \theta \in \Theta_1 = \Theta \setminus \{\theta_0\}$ , est *bilatérale* (et composée).

**Définition 4.** On appelle modèle statistique *paramétrique* un modèle  $(\mathbf{R}^n, \mathcal{B}_n, \mathcal{P})$  tel qu'il existe  $k \in \mathbf{N}$  :

$$\mathcal{P} = \{\mathbf{P}_\theta, \theta \in \Theta \subset \mathbf{R}^k\},$$

sinon on dit que le modèle  $(\mathbf{R}^n, \mathcal{B}_n, \mathcal{P})$  est *non paramétrique*.

**Exemple 3.** Soit  $\mathbf{X}$  un vecteur aléatoire et soit  $H_0$  l'hypothèse selon laquelle la fonction de répartition de  $\mathbf{X}$  est continue. Dans ce cas le modèle est non paramétrique.

**Exemple 4.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i$  suit une loi normale  $N(\mu, \sigma^2)$ , i.e.  $\theta = (\mu, \sigma^2)^T \in \Theta$ ,  $\Theta = \{\theta : |\mu| < \infty, \sigma^2 > 0\}$ . Comme  $\Theta \subset \mathbf{R}^2$ , on a l'exemple d'un modèle paramétrique.

Soient  $\mathbf{X} = \mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon et  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbf{R}^n$  une réalisation de  $\mathbb{X}$ , reçue dans l'expérience.

Soit  $\varphi : \mathbf{R}^n \rightarrow [0, 1]$  une application borélienne qu'on appellera *fonction critique*.

**Définition 5.** On dit qu'une fonction critique  $\varphi$  détermine *le test statistique* pour tester  $H_0 : \theta \in \Theta_0$  contre  $H_1 : \theta \in \Theta_1$  si l'on *rejette*  $H_0$  avec la probabilité  $\varphi(\mathbf{x})$  et on *rejette*  $H_1$  avec la probabilité  $1 - \varphi(\mathbf{x})$ .

**Définition 6.** La fonction

$$\beta_\varphi(\theta) = \mathbf{E}_\theta \varphi(\mathbb{X}), \quad \theta \in \Theta_0 \cup \Theta_1, \quad (1)$$

est appelée *la fonction de puissance* du test, basé sur la fonction critique  $\varphi$ .

**Définition 7.** La fonction

$$\beta_\varphi(\theta), \quad \theta \in \Theta_0$$

est appelée *le risque de première espèce*. C'est le risque de rejeter  $H_0$  à tort ; on constate que le risque de première espèce est la restriction de la fonction de puissance à  $\Theta_0$ .

**Définition 8.** La fonction

$$\beta_\varphi(\theta), \quad \theta \in \Theta_1$$

est appelée *la puissance* du test, basé sur la fonction critique  $\varphi$  ; on constate que la puissance est la restriction de la fonction de puissance  $\beta_\varphi(\theta)$  à  $\Theta_1$ .

**Définition 9.** La fonction

$$1 - \beta_\varphi(\theta) = \mathbf{E}_1(1 - \varphi(X)) = 1 - \int_{\mathcal{X}} \varphi(x) p_1(x) \mu(dx), \quad \theta \in \Theta_1$$

est appelé *le risque de deuxième espèce*. C'est le risque d'accepter  $H_0$  à tort.

Si  $\varphi$  est de la forme

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in K \subset \mathbf{R}^n, \\ 0, & \mathbf{x} \in \mathbf{R}^n \setminus K, \end{cases} \quad (2)$$

alors le test statistique, basé sur cette fonction critique, est appelé *pur* ou *non randomisé*, sinon le test est *randomisé*.

L'ensemble  $K$  est appelé *la région critique* ou *la zone de rejet* de ce test : on y rejette  $H_0$  (et on y accepte l'alternative  $H_1$ ).

L'ensemble  $\bar{K} = \mathbf{R}^n \setminus K$  est appelé *la zone d'acceptation* (de non rejet) de  $H_0$ .

Soit  $\varphi : \mathbf{R}^n \rightarrow [0, 1]$  une fonction critique.

Il est évident qu'en cas de test non randomisé :

$$\beta_\varphi(\theta) = \mathbf{P}_\theta(\mathbb{X} \in K), \quad \theta \in \Theta_0 \cup \Theta_1, \quad (3)$$

et donc  $\beta_\varphi(\theta)$  nous donne la probabilité avec laquelle  $\mathbb{X}$  tombe dans la région critique  $K$  si la vraie valeur de paramètre est  $\theta$ .

Donc dans le cas d'un test pur le risque de première espèce est la probabilité de *rejeter à tort* l'hypothèse  $H_0$  quand  $\theta \in \Theta_0$ , lorsque l'hypothèse  $H_0$  est *vraie*. Le risque de deuxième

espèce est la probabilité d'accepter l'hypothèse  $H_0$  quand  $\theta \in \Theta_1$ , lorsque l'hypothèse  $H_0$  est fausse.

Le test  $\varphi$  est bon, si les erreurs sont petites. On ne peut pas les rendre simultanément aussi petites que l'on veut, parce que, en augmentant  $K$ , l'erreur de 2-ème espèce diminue mais l'erreur de 1-ère espèce augmente et vice versa, en diminuant  $K$  l'erreur de 1-ère espèce diminue mais celle de 2-ème espèce augmente.

Soit  $H_0 : \theta \in \Theta_0$ .

Le nombre

$$\alpha = \sup_{\theta \in \Theta_0} \beta_\varphi(\theta), \quad 0 < \alpha < 1,$$

est appelé le *niveau ou le seuil de signification* du test  $\varphi$ , ce qui signifie que la probabilité de rejeter  $H_0$  à tort ne devra pas dépasser  $\alpha$ .

Le test  $\varphi$  de niveau  $\alpha$  est *sans biais*, si sa puissance est supérieure ou égale à  $\alpha$ , i.e. si

$$\beta_\varphi(\theta) \geq \alpha \quad \text{pour} \quad \forall \theta \in \Theta_1.$$

Le test  $\varphi$  est *uniformément le plus puissant* (UPP) de seuil  $\alpha$ , si pour tout autre test  $\psi$  on a

$$\begin{aligned} \beta_\varphi(\theta) &\leq \beta_\psi(\theta) \leq \alpha & \forall \theta \in \Theta_0, \\ \beta_\varphi(\theta) &\geq \beta_\psi(\theta) & \forall \theta \in \Theta_1. \end{aligned}$$

Considérons le cas de l'hypothèse  $H_0$  et de l'alternative  $H_1$  simples :

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1.$$

Dans ce cas la puissance d'un test statistique non randomisé, destiné à tester  $H_0$  contre  $H_1$ , est la probabilité de rejeter  $H_0$  quand l'alternative  $H_1$  est vraie :

$$\pi = \mathbf{P}_{\theta_1} \{ \mathbb{X} \in K \} = \beta_\varphi(\theta_1),$$

et le niveau de signification est la probabilité de rejeter  $H_0$  à tort :

$$\alpha = \mathbf{P}_{\theta_0} \{ \mathbb{X} \in K \} = \beta_\varphi(\theta_0).$$

C'est la *probabilité d'erreur de première espèce*. La probabilité  $\beta = 1 - \pi$  s'appelle la *probabilité d'erreur de deuxième espèce*.

## 4.2 Test de Neyman-Pearson.

Supposons que  $\mathcal{P} = \{ \mathbf{P}_{\theta_0}, \mathbf{P}_{\theta_1} \}$  est dominée par une mesure  $\sigma$ -finie  $\mu$  et notons  $f_0$  et  $f_1$  les densités de  $\mathbf{P}_{\theta_0}$  et  $\mathbf{P}_{\theta_1}$  par rapport à  $\mu$ .

**Lemme de Neyman-Pearson.** *Pour tout  $\alpha \in ]0, 1[$  il existe des constantes  $c_\alpha > 0$  et  $\gamma_\alpha \in [0, 1]$  telles, que le test, basé sur la fonction critique*

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{si } p_1(\mathbf{x}) > c_\alpha p_0(\mathbf{x}), \\ \gamma_\alpha, & \text{si } p_1(\mathbf{x}) = c_\alpha p_0(\mathbf{x}), \\ 0, & \text{sinon,} \end{cases}$$

a le niveau  $\alpha$  et est le plus puissant parmi tous les tests  $\psi$  tels que  $\mathbf{E}_{\theta_0}\psi(\mathbb{X}) \leq \alpha$ .

**Démonstration.**

1) On cherche des constantes  $c_\alpha$  et  $\gamma_\alpha$  telles que  $\mathbf{E}_{\theta_0}\varphi(\mathbb{X}) = \alpha$  :

$$\mathbf{E}_{\theta_0}\varphi(\mathbb{X}) = \mathbf{P}_{\theta_0}\{p_1(\mathbb{X}) > c_\alpha p_0(\mathbb{X})\} + \gamma_\alpha \mathbf{P}_{\theta_0}\{p_1(\mathbb{X}) = c_\alpha p_0(\mathbb{X})\} = \alpha. \quad (1)$$

Posons

$$F(c) = \mathbf{P}_{\theta_0}\left\{\frac{p_1(\mathbb{X})}{p_0(\mathbb{X})} \leq c\right\}, \quad c \geq 0.$$

$F$  a un sens, puisque  $p_0(\mathbb{X}) > 0$  p.s., si  $\mathbb{X} \sim p_0(x)$ .

Avec cette notation l'égalité (1) peut être écrite sous la forme

$$\mathbf{E}_{\theta_0}\varphi(\mathbb{X}) = 1 - F(c_\alpha) + \gamma_\alpha[F(c_\alpha) - F(c_\alpha - 0)] = \alpha, \quad (2)$$

puisque la fonction  $F$  est continue à droite.

a) S'il existe  $c : F(c) = 1 - \alpha$ , on peut prendre  $c_\alpha = c$ ,  $\gamma_\alpha = 0$  pour lesquelles on a l'égalité qu'il nous faut :

$$\mathbf{E}_{\theta_0}\varphi(\mathbb{X}) = \alpha.$$

b) Sinon il existe  $c$  :

$$F(c - 0) \leq 1 - \alpha < F(c). \quad (3)$$

On peut prendre  $c_\alpha = c$  et définir  $\gamma$  en résolvant l'équation

$$\alpha = 1 - F(c) + \gamma[F(c) - F(c - 0)].$$

On obtient

$$\gamma = [\alpha - 1 + F(c)]/[F(c) - F(c - 0)] = \frac{F(c) - (1 - \alpha)}{F(c) - F(c - 0)}.$$

Des inégalités (3) on tire

$$\alpha - 1 + F(c) \leq F(c) - F(c - 0) \quad \text{et} \quad F(c) + \alpha - 1 = F(c) - (1 - \alpha) > 0,$$

c'est pourquoi  $0 < \gamma \leq 1$ .

2) On montre que le test  $\varphi$  est le plus puissant. Supposons que  $\psi$  est un autre test, tel que  $\mathbf{E}_{\theta_0}\psi(\mathbb{X}) \leq \alpha$ . Alors

$$\mathbf{E}_{\theta_0}(\varphi(\mathbb{X}) - \psi(\mathbb{X})) \geq 0.$$

De la définition de  $\varphi$  on tire :

si  $p_1(x) - c_\alpha p_0(x) > 0$ , alors  $\varphi(x) = 1 \geq \psi(x)$  et donc  $\varphi(x) - \psi(x) \geq 0$  ;

si  $p_1(x) - c_\alpha p_0(x) < 0$ , alors  $\varphi(x) = 0 \leq \psi(x)$  et donc  $\varphi(x) - \psi(x) \leq 0$  ;

c'est pourquoi

$$(\varphi(x) - \psi(x))(p_1(x) - c_\alpha p_0(x)) \geq 0 \quad \forall x,$$

$$\int_x (\varphi(x) - \psi(x))(p_1(x) - c_\alpha p_0(x))\mu(dx) \geq 0$$

et

$$\int_x (\varphi(x) - \psi(x))p_1(x)\mu(dx) \geq c_\alpha \int_x (\varphi(x) - \psi(x))p_0(x)\mu(dx).$$

La dernière inégalité peut s'écrire :

$$\mathbf{E}_{\theta_1} \varphi(\mathbb{X}) - \mathbf{E}_{\theta_1} \psi(\mathbb{X}) \geq c_\alpha (\mathbf{E}_{\theta_0} \varphi(\mathbb{X}) - \mathbf{E}_{\theta_0} \psi(\mathbb{X})) \geq 0.$$

C'est pourquoi  $\mathbf{E}_{\theta_1} \varphi(\mathbb{X}) \geq \mathbf{E}_{\theta_1} \psi(\mathbb{X})$ , et le test  $\varphi$  est plus puissant que  $\psi$ .

**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$H : X_i \sim f(x; \theta) = \theta^x (1 - \theta)^{1-x}, \quad x \in \mathcal{X}, \quad \theta \in \Theta = ]0, 1[,$$

i.e. on a le modèle statistique selon lequel  $X_i$  suit la loi de Bernoulli de paramètre  $\theta$ ,  $\theta \in \Theta$ . Notre problème consiste à tester au niveau  $\alpha = 0.05$  l'hypothèse nulle  $H_0 : \theta = \theta_0$  contre l'alternative  $H_1 : \theta = \theta_1 > \theta_0$ . On remarque que dans notre modèle  $H$  il existe une statistique exhaustive minimale unidimensionnelle (scalaire)

$$\mu_n = X_1 + \dots + X_n.$$

Si l'hypothèse  $H_0 : \theta = \theta_0$  est vraie, alors :

$$\mathbf{P}_{\theta_0} \{\mu_n = x\} = \binom{n}{x} \theta_0^x (1 - \theta_0)^{n-x}, \quad x \in \mathcal{X}_0^n = \{0, 1, \dots, n\}. \quad (1)$$

Si l'alternative  $H_1$  est vraie, alors :

$$\mathbf{P}_{\theta_1} \{\mu_n = x\} = \binom{n}{x} \theta_1^x (1 - \theta_1)^{n-x}, \quad x \in \mathcal{X}_0^n = \{0, 1, \dots, n\}. \quad (2)$$

Donc le problème de tester  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta = \theta_1$  revient au problème de tester l'hypothèse que  $\mu_n$  suit la loi Binomiale  $B(n, \theta_0)$  contre l'alternative que  $\mu_n$  suit la loi binomiale  $B(n, \theta_1)$ ,  $\theta_1 > \theta_0$ .

En fonction de la statistique  $\mu_n$  le rapport de vraisemblance est

$$L(\mu_n) = \frac{\binom{n}{\mu_n} \theta_1^{\mu_n} (1 - \theta_1)^{n-\mu_n}}{\binom{n}{\mu_n} \theta_0^{\mu_n} (1 - \theta_0)^{n-\mu_n}} = \left(\frac{\theta_1}{\theta_0}\right)^{\mu_n} \left(\frac{1 - \theta_1}{1 - \theta_0}\right)^{n-\mu_n}.$$

On peut remarquer que

$$\frac{\theta_1}{\theta_0} > 1 \quad \text{et} \quad \frac{1 - \theta_1}{1 - \theta_0} < 1,$$

et donc  $L(\mu_n)$  est monotone en  $\mu_n$ , d'où on tire que le meilleur test (le test de Neyman-Pearson) de niveau  $\alpha$  pour  $H_0$  contre  $H_1$  est basé sur la statistique

$$\varphi(\mu_n) = \begin{cases} 1, & \text{si } \mu_n > c_\alpha, \\ \gamma, & \text{si } \mu_n = c_\alpha, \\ 0, & \text{sinon,} \end{cases}$$

où les constantes  $c_\alpha$  ( la valeur critique) et  $\gamma = \gamma_{0.05}$  sont telles que

$$\mathbf{E}_{\theta_0} \varphi(\mu_n) = \mathbf{P}_{\theta_0} \{\mu_n > c_\alpha\} + \gamma_{0.05} \mathbf{P}_{\theta_0} \{\mu_n = c_\alpha\} = \alpha = 0.05. \quad (3)$$

Supposons que  $n = 10$ ,  $\theta_0 = 0.25 = 1/4$ . Si l'on choisit  $K = K_5 = \{5, 6, 7, 8, 9, 10\}$ , dans ce cas on a

$$\begin{aligned} \mathbf{P}_{\theta_0}\{\mu_n \in K_5\} &= \sum_{x=5}^{10} \binom{10}{x} \theta_0^x (1 - \theta_0)^{10-x} = \\ &= \sum_{x=5}^{10} \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x} = 0.0781 > \alpha = 0.05. \end{aligned}$$

Si l'on choisit  $K = K_6 = \{6, 7, 8, 9, 10\}$ , dans ce cas on a

$$\mathbf{P}_{\theta_0}\{\mu_n \in K_6\} = \sum_{x=6}^{10} \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x} = 0.0197 < \alpha = 0.05.$$

Donc on voit que

$$\mathbf{P}_{\theta_0}\{\mu_n \in K_5\} = \mathbf{P}_{\theta_0}\{\mu_n \geq 5\} = 0.0781 \quad \text{et} \quad \mathbf{P}_{\theta_0}\{\mu_n \in K_6\} = \mathbf{P}_{\theta_0}\{\mu_n \geq 6\} = 0.0197,$$

d'où on tire que

$$\mathbf{P}_{\theta_0}\{\mu_n = 5\} = \mathbf{P}_{\theta_0}\{\mu_n \geq 5\} - \mathbf{P}_{\theta_0}\{\mu_n \geq 6\} = 0.0781 - 0.0197 = 0.0584.$$

On détermine à l'aide de (3) la probabilité  $\gamma$  :

$$\gamma(0.05) = \frac{\alpha - 0.0197}{0.0781 - 0.0197} = \frac{0.05 - 0.0197}{0.0781 - 0.0197} = 0.519,$$

et on obtient la fonction critique du meilleur test de Neyman-Pearson de niveau  $\alpha$  :

$$\varphi(\mu_n) = \begin{cases} 1, & \text{si } \mu_n \in K_6, \\ 0.519, & \text{si } \mu_n = 5, \\ 0, & \text{sinon.} \end{cases}$$

On voit que

$$\begin{aligned} \mathbf{E}_{\theta_0}\varphi(\mu_n) &= 1 \cdot \mathbf{P}_{\theta_0}\{\mu_n \in K_6\} + \gamma(0.05)\mathbf{P}_{\theta_0}\{\mu_n = 5\} = \\ &= 0.0197 + 0.519 \cdot 0.0584 = 0.050 = \alpha. \end{aligned}$$

La puissance de ce test randomisé quand  $\theta = \theta_1 = \frac{1}{2}$  est égale à

$$\begin{aligned} \pi &= \mathbf{E}_{\theta_1}\{\varphi(\mu_n)\} = \mathbf{P}_{\theta_1}\{\mu_n \in K_6\} + \gamma(0.05)\mathbf{P}_{\theta_1}\{\mu_n = 5\} = \\ &= \sum_{x=6}^{10} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} + 0.519 \binom{10}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \\ &= 0.3770 + 0.519 \cdot 0.2461 = 0.5047. \end{aligned}$$

Enfin on remarque que le risque de deuxième espèce  $\beta = 0.4953$ .

**Exemple 2.** Soit  $\mathbb{X} = (X_1, \dots, X_5)^T$  un échantillon. Trouver le plus puissant test de niveau  $\alpha = 0.1$  vérifiant l'hypothèse  $H_0 : U(-0.5; 0.5)$  contre l'alternative  $H_1 : N(0; 0.009)$ . Vérifier l'hypothèse  $H_0$  si des réalisations de  $\mathbb{X}$  sont

$$-0.114; -0.325; 0.196; -0.174; -0.460.$$

Solution. On cherche le test de Neyman-Pearson pur :

$$\varphi(\mathbb{X}) = \begin{cases} 1, & \text{si } L_1(\mathbb{X}) > cL_0(\mathbb{X}); \\ 0, & \text{sinon,} \end{cases}$$

où

$$L_0(\mathbb{X}) = \mathbf{1}\{-0.5 \leq X_{(1)} \leq X_{(5)} \leq 0.5\},$$

$$L_1(\mathbb{X}) = \frac{1}{\sigma^5(2\pi)^{5/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^5 X_i^2\right\}, \quad \sigma^2 = 0.009.$$

L'inégalité  $L_1(\mathbb{X}) > cL_0(\mathbb{X})$  est vraie si et seulement si

$$\left\{\sum_{i=1}^5 X_i^2 < k\right\} \cup \left\{X_{(1)} < -0.5\right\} \cup \left\{X_{(5)} < 0.5\right\}.$$

On cherche  $k$  de condition

$$\mathbf{P}_0 \left\{ \left\{ \sum_{i=1}^5 X_i^2 < k \right\} \cup \left\{ X_{(1)} < -0.5 \right\} \cup \left\{ X_{(5)} < 0.5 \right\} \right\} = \mathbf{P}_0 \left\{ \sum_{i=1}^5 X_i^2 < k \right\} = \alpha,$$

où

$$\int \dots \int_{\substack{\sum_{i=1}^5 X_i^2 < R^2 \\ -0.5 < X_i < 0.5}} dx_1 \dots dx_5 = 0.1, \quad (4.1)$$

où  $k = R^2$ .

Si  $R \leq 0.5$ , cette intégrale est égale à l'intégrale

$$I = \int \dots \int_{\sum_{i=1}^5 X_i^2 < R^2} dx_1 \dots dx_5. \quad (4.2)$$

Dans l'intégrale (1) on fait le changement de variables

$$\begin{aligned} x_1 &= r \cos \varphi_1 \\ x_2 &= r \sin \varphi_1 \cos \varphi_2 \\ x_3 &= r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3 \\ x_4 &= r \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \cos \varphi_4 \\ x_5 &= r \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \sin \varphi_4. \end{aligned}$$

Le Jacobien

$$J = r^4 \sin^3 \varphi_1 \sin^2 \varphi_2 \sin \varphi_3.$$

$$I = \int_0^R r^4 dr \int_0^\pi \sin^3 \varphi_1 d\varphi_1 \int_0^\pi \sin^2 \varphi_2 d\varphi_2 \int_0^\pi \sin \varphi_3 d\varphi_3 \int_0^{2\pi} d\varphi_4 = \frac{8\pi^2 R^5}{15}.$$

Si  $R = 0.5$ ,

$$I = \frac{\pi^2}{60} > \frac{9}{60} = \frac{3}{20} > 0.1,$$

donc (1) peut être vraie, si  $R < 0.5$ .

$R$  satisfait l'équation

$$\frac{8\pi^2 R^5}{15} = 0.1,$$

donc

$$R^5 = \frac{3}{16\pi^2}.$$

On rejette l'hypothèse  $H_0$  si

$$\sum_{i=1}^5 X_i^2 < \left(\frac{3}{16\pi^2}\right)^{5/2} \quad \text{où } X_{(1)} < -0.5 \quad \text{ou } X_{(5)} > 0.5.$$

Dans notre cas  $\sum_{i=1}^5 X_i^2 = 0.399$ ,  $X_{(1)} = -0.325$ ,  $X_{(5)} = 0.196$ . On a

$$0.399 > \left(\frac{3}{16\pi^2}\right)^{5/2}$$

l'hypothèse est accepté.

## 4.3 Loi multinomiale et test du chi-deux de Pearson.

### Loi multinomiale.

Considérons une suite de  $n$  épreuves indépendantes et supposons que dans chaque épreuve il ne puisse se passer qu'un seul événement parmi  $k$  possibles  $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_k$ , dont les probabilités,

$$p_1 = \mathbf{P}(\mathbf{E}_1), p_2 = \mathbf{P}(\mathbf{E}_2), \dots, p_k = \mathbf{P}(\mathbf{E}_k),$$

sont positives et  $p_1 + \dots + p_k = 1$ .

Notons  $\mathbf{p} = (p_1, \dots, p_k)^T$  et  $\mathbf{v} = (v_1, \dots, v_k)^T$ , où  $v_i$  est la fréquence de  $\mathbf{E}_i$  dans la suite d'épreuves ( $i = 1, \dots, k$ ). Il est évident que les valeurs prises par les  $v_i$  sont des valeurs entières  $n_i$ ,  $0 \leq n_i \leq n$ ,

$$n_1 + n_2 + \dots + n_k = n. \quad (1)$$

Le vecteur  $\mathbf{v}$  suit la loi multinomiale de paramètres  $n$  et  $\mathbf{p}$  :

$$\mathbf{P}\{v_1 = n_1, \dots, v_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}, \quad (2)$$

pour tout  $n_1, \dots, n_k$  entiers, satisfaisant aux conditions (1).

Par des calculs directs, on peut établir que le vecteur des espérances,  $\mathbf{E}\mathbf{v}$ , et la matrice de covariance,

$$\Sigma = \mathbf{Var}\mathbf{v} = \mathbf{E}(\mathbf{v} - \mathbf{E}\mathbf{v})(\mathbf{v} - \mathbf{E}\mathbf{v})^T,$$

du vecteur  $\mathbf{v}$  sont égaux à

$$\mathbf{E}\mathbf{v} = n\mathbf{p}, \Sigma = \mathbf{E}(\mathbf{v} - n\mathbf{p})(\mathbf{v} - n\mathbf{p})^T = n(\mathbf{P} - \mathbf{p}\mathbf{p}^T), \quad (3)$$

ou  $\mathbf{P}$  est la matrice diagonale dont les éléments sur la diagonale principale sont  $p_1, \dots, p_k$ . Il est facile de vérifier que  $\text{rang}(\Sigma) = k - 1$ , à cause de la condition (1).

**Test du chi-deux de Pearson.**

Soit  $\mathbf{1} = \mathbf{1}_k = (1, \dots, 1)^T \in \mathbf{R}^k$ . Nous pouvons écrire que

$$\mathbf{p}^T \mathbf{1}_k = 1, \mathbf{v}^T \mathbf{1}_k = k.$$

Notons

$$\tilde{\mathbf{p}} = (p_1, \dots, p_{k-1})^T, \tilde{\mathbf{v}} = (v_1, \dots, v_{k-1})^T, \tilde{\mathbf{1}} = \mathbf{1}_{k-1},$$

$\tilde{\mathbf{P}}$  est la matrice que l'on obtient à partir de la matrice  $\mathbf{P}$ , en enlevant la dernière ligne et la dernière colonne, c'est-à-dire que  $\tilde{\mathbf{P}}$  est la matrice diagonale dont les éléments de la diagonale principale sont  $p_1, \dots, p_{k-1}$ . De la même façon on obtient la matrice

$$\tilde{\Sigma} = n(\tilde{\mathbf{P}} - \tilde{\mathbf{p}}\tilde{\mathbf{p}}^T).$$

Il est facile de vérifier que  $\tilde{\mathbf{p}}^T \tilde{\mathbf{1}} = 1 - p_k$ ,  $\text{rang}(\tilde{\Sigma}) = k - 1$  et que la matrice inverse  $\tilde{\Sigma}^{-1}$  de  $\tilde{\Sigma}$  est

$$\tilde{\Sigma}^{-1} = \frac{1}{n} \left( \tilde{\mathbf{P}}^{-1} + \frac{1}{p_k} \tilde{\mathbf{1}}\tilde{\mathbf{1}}^T \right), \quad (4)$$

où  $\tilde{\mathbf{P}}^{-1}$  est la matrice inverse de  $\tilde{\mathbf{P}}$ .

Soit  $\mathbf{p}_0 = (p_{01}, p_{02}, \dots, p_{0k})^T$  un vecteur arbitraire qui satisfait la condition

$$\mathbf{p}_0^T \mathbf{1} = 1,$$

tel que tous les  $p_{0i}$  sont positifs, et supposons que le vecteur  $\mathbf{v}$  suive la loi multinomiale (2) de paramètres  $n$  et  $\mathbf{p}$ . Dans ce cas si  $n \rightarrow \infty$ , alors d'après le théorème limite central à plusieurs dimensions le vecteur  $\frac{1}{\sqrt{n}}(\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}_0)$  est asymptotiquement distribué selon la loi normale à  $(k - 1)$  dimensions de paramètres

$$(\tilde{\mathbf{p}} - \tilde{\mathbf{p}}_0) \text{ et } \tilde{\mathbf{P}} - \tilde{\mathbf{p}}\tilde{\mathbf{p}}^T = \frac{1}{n}\tilde{\Sigma}.$$

Par conséquent la forme quadratique de Pearson

$$X_n^2 = \frac{1}{n}(\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}_0)^T \left( \tilde{\mathbf{P}}^{-1} + \frac{1}{p_k} \tilde{\mathbf{1}}\tilde{\mathbf{1}}^T \right) (\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}_0) \quad (5)$$

est distribuée approximativement (quand  $n$  tend vers l'infini) comme la variable aléatoire  $\chi_{k-1}^2(\lambda_n)$ , où

$$\lambda_n = n(\tilde{\mathbf{p}} - \tilde{\mathbf{p}}_0)^T \left( \tilde{\mathbf{P}}^{-1} + \frac{1}{p_k} \tilde{\mathbf{1}}\tilde{\mathbf{1}}^T \right) (\tilde{\mathbf{p}} - \tilde{\mathbf{p}}_0). \quad (6)$$

Comme

$$(\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}_0)^T \tilde{\mathbf{P}}^{-1} (\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}_0) = \sum_{i=1}^k \frac{(v_i - np_{0i})^2}{np_i} \quad (7)$$

et

$$\tilde{\mathbf{I}}^T(\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}_0) = -(\mathbf{v}_k - np_{0k}), \quad (8)$$

la statistique de Pearson  $X_n^2$  peut s'écrire :

$$X_n^2 = \sum_{i=1}^k \frac{(\mathbf{v}_i - np_{0i})^2}{np_i}. \quad (9)$$

**Théorème 1.** Soit  $\{\mathbf{p}_n\}$  une suite de vecteurs  $\mathbf{p}_n = (p_{n1}, p_{n2}, \dots, p_{nk})^T$  tels que  $\mathbf{p}_n^T \mathbf{1} = 1$  et tous les  $p_{ni}$  soient positifs. Supposons que

$$\hat{\lambda}_n = n \sum_{i=1}^k \frac{(p_{ni} - p_{0i})^2}{p_{0i}} \rightarrow \lambda, (\lambda > 0) \quad (10)$$

quand  $n \rightarrow \infty$ . Dans ce cas la statistique de Pearson

$$X_n^2 = \sum_{i=1}^k \frac{(\mathbf{v}_i - np_{ni})^2}{np_{ni}} \quad (11)$$

suit à la limite, quand  $n \rightarrow \infty$ , la même loi que la variable aléatoire  $\chi_{k-1}^2(\lambda)$ .

Supposons que nous ayons à tester l'hypothèse  $\mathbf{H}_0 : \mathbf{p} = \mathbf{p}_0$ . Soit  $x(\alpha, k-1)$ , le quantile supérieur de niveau  $\alpha$  de la distribution du chi-deux à  $(k-1)$  degrés de liberté, c'est-à-dire que

$$\mathbf{P}\{\chi_{k-1}^2 \geq x(\alpha, k-1)\} = \alpha. \quad (12)$$

D'après le test du chi-deux de Pearson, fondé sur la statistique de Pearson  $X_n^2$ , on rejette l'hypothèse  $\mathbf{H}_0$  si

$$X_n^2 \geq c_\alpha = x(\alpha, k-1). \quad (13)$$

Le nombre  $c_\alpha$  s'appelle la valeur critique du test. De (12),(13),(6) et (9) on déduit que

$$\mathbf{P}\{X_n^2 \geq x(\alpha, k-1) \mid \mathbf{H}_0\} \rightarrow \alpha, \quad \text{quand } n \rightarrow \infty. \quad (14)$$

Par ailleurs si l'hypothèse  $\mathbf{H}_{1n} : \mathbf{p} = \mathbf{p}_n$  est vraie, alors du Théorème 1 il résulte que

$$\mathbf{P}\{X_n^2 \geq x(\alpha, k-1) \mid \mathbf{H}_{1n}\} = \mathbf{P}\{\chi_{k-1}^2(\lambda) \geq x(\alpha, k-1)\} + o(1), \quad (15)$$

si

$$n \sum_{i=1}^k \frac{(p_{ni} - p_{0i})^2}{p_{0i}} \rightarrow \lambda, \quad \text{quand } n \rightarrow \infty. \quad (16)$$

Par exemple, si

$$p_{ni} = p_{0i} + \frac{\delta_i}{\sqrt{n}}, \quad (17)$$

où

$$\delta_1 + \delta_2 + \dots + \delta_k = 0,$$

$$\lambda_n = \lambda = \sum_{i=1}^k \frac{\delta_i^2}{p_{0i}}. \quad (18)$$

La probabilité

$$\beta_n = \mathbf{P} \{X_n^2 \geq x(\alpha, k-1) \mid \mathbf{H}_{1n}\} \cong \mathbf{P} \{\chi_{k-1}^2(\lambda_n) \geq x(\alpha, k-1)\} \quad (19)$$

s'appelle la *puissance du test du chi-deux de Pearson*.

Par ailleurs, la relation

$$\mathbf{P} \{X_n^2 \leq x(\alpha, k-1) \mid \mathbf{H}_{1n}\} = 1 - \beta_n \quad (20)$$

nous donne la probabilité d'*erreur de seconde espèce* que l'on commet en prenant  $\mathbf{H}_0$  à tort parce que l'on a observé l'événement  $\{X_n^2 \leq x(\alpha, k-1)\}$ , tandis qu'en fait c'est l'hypothèse  $\mathbf{H}_{1n}$  qui est vraie. On remarque ici que plus la puissance  $\beta_n$  est grande, plus petite est la probabilité de commettre l'erreur de prendre  $\mathbf{H}_0$  à tort. Enfin, on note que pour calculer  $1 - \beta_n$  on peut utiliser l'approximation normale de la loi du chi-deux non centrale, d'après laquelle

$$1 - \beta_n = \mathbf{P} \{X_n^2 \leq x(\alpha, k-1) \mid \mathbf{H}_{1n}\} \cong \Phi \left\{ \frac{x(\alpha, k-1) - (k-1 + \lambda_n)}{\sqrt{2(k-1 + 2\lambda_n)}} \right\}, \quad (21)$$

et par conséquent on obtient

$$\beta_n \cong \Phi \left\{ \frac{k-1 + \lambda_n - x(\alpha, k-1)}{\sqrt{2(k-1 + 2\lambda_n)}} \right\}, \quad (22)$$

pourvu que  $k + \lambda_n$  soit assez grand, c'est-à-dire, en pratique, supérieur ou égal à 30.

Supposons maintenant, que  $\mathbf{H}_{1n}$  soit telle que  $\mathbf{p}_n \neq \mathbf{p}_0$  et

$$\lambda_n = n \sum_{i=1}^k \frac{(p_{ni} - p_{0i})^2}{p_{0i}} \rightarrow \infty, \quad (23)$$

quand  $n \rightarrow \infty$ . Dans ce cas, de (20) il résulte que  $(1 - \beta_n) \rightarrow 0$  et donc  $\beta_n \rightarrow 1$ , quand  $n \rightarrow \infty$ , et on dit que le test est *consistant*.

**Remarque sur la correction de continuité.**

Si  $k = 2$ , alors

$$X_n^2 = \frac{(v_1 - np_{01})^2}{np_{01}} + \frac{(v_2 - np_{02})^2}{np_{02}} = \frac{(v_1 - np_{01})^2}{np_{01}(1 - p_{01})}, \quad (24)$$

car  $v_1 + v_2 = n$ . Supposons que l'hypothèse  $\mathbf{H}_0$  soit vraie. Dans ce cas la fréquence  $v_1$  suit la loi binomiale de paramètres  $n$  et  $p_{01}$  et par conséquent du théorème de Moivre-Laplace il résulte que si  $n \rightarrow \infty$ , alors pour tout  $m$  ( $1 \leq m \leq n$ )

$$\mathbf{P} \{v_1 \leq m \mid \mathbf{H}_0\} = \Phi \left\{ \frac{m + 0.5 - np_{01}}{\sqrt{np_{01}(1 - p_{01})}} \right\} + O\left(\frac{1}{\sqrt{n}}\right), \quad (25)$$

d'où on tire

$$\begin{aligned} \mathbf{P} \{v_1 \geq m \mid \mathbf{H}_0\} &= 1 - \mathbf{P} \{v_1 \leq m-1 \mid \mathbf{H}_0\} = \\ &= \Phi \left\{ \frac{m - 0.5 - np_{01}}{\sqrt{np_{01}(1 - p_{01})}} \right\} + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad (26)$$

De (25) et (26) il résulte que si nous voulons utiliser le test du chi-deux de Pearson, fondé sur la statistique  $X_n^2$  du niveau de signification  $\cong \alpha$ , nous devons rejeter  $\mathbf{H}_0$  quand

$$\Phi \left\{ \frac{v_1 + 0.5 - np_{01}}{\sqrt{np_{01}(1-p_{01})}} \right\} \leq \frac{\alpha}{2} \quad \text{ou} \quad \Phi \left\{ -\frac{v_1 - 0.5 - np_{01}}{\sqrt{np_{01}(1-p_{01})}} \right\} \leq \frac{\alpha}{2}. \quad (27)$$

De (27) on déduit que l'on doit rejeter  $\mathbf{H}_0$  si l'un des événements

$$\frac{v_1 - np_{01}}{\sqrt{np_{01}(1-p_{01})}} \leq \Psi\left(\frac{\alpha}{2}\right) - \frac{1}{\sqrt{np_{01}(1-p_{01})}} \quad (28)$$

ou

$$\frac{v_1 - np_{01}}{\sqrt{np_{01}(1-p_{01})}} \geq -\Psi\left(\frac{\alpha}{2}\right) + \frac{1}{\sqrt{np_{01}(1-p_{01})}} \quad (29)$$

est apparu, où  $\Psi(y)$  est la fonction inverse de  $\Phi(x)$ . Donc on a montré que le test du chi-deux de Pearson à 1 degré de liberté rejette  $\mathbf{H}_0$  si

$$X_n^2 \geq \left[ \Psi\left(1 - \frac{\alpha}{2}\right) + \frac{1}{2\sqrt{np_{01}(1-p_{01})}} \right]^2 \quad (30)$$

(ici nous avons utilisé l'identité :  $\Psi(y) + \Psi(1-y) \equiv 0, y \in [0, 1]$ .)

De la formule (30) il résulte que si  $k = 2$ , alors la valeur critique  $c_\alpha$  du test du chi-deux doit être égale à

$$c_\alpha = \left[ \Psi\left(1 - \frac{\alpha}{2}\right) + \frac{1}{\sqrt{np_{01}(1-p_{01})}} \right]^2 \quad (31)$$

pour avoir le niveau du test  $\cong \alpha$ .

### Test du chi-deux pour des données de Mendel.

Dans ses expériences Mendel a observé 315 pois ronds et jaunes, 108 pois ronds et verts, 101 pois ridés et jaunes, 32 pois ridés et verts. Au total Mendel a observé 556 pois. D'après l'hypothèse  $H_0$  de Mendel les probabilités  $p_1, p_2, p_3, p_4$  d'observer un pois Rond et Jaune, un pois Rond et vert, un pois ridé et Jaune, un pois ridé et vert sont proportionnelles à 9,3,3 et 1 respectivement (voir Remarque 1). Peut-on dire que les données de Mendel sont en accord avec son hypothèse  $H_0$  ?

Notons  $n$  le nombre total des pois (dans l'expérience de Mendel  $n = 556$ ), et soit  $\mathbf{v} = (v_1, v_2, v_3, v_4)^T$  est le vecteur des fréquences des événements que Mendel a observés :

$$v_1 = 315, v_2 = 108, v_3 = 101, v_4 = 31.$$

Dans notre modèle le vecteur  $\mathbf{v}$  suit la loi multinomiale de paramètres  $n$  et  $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$ , où

$$p_1 + p_2 + p_3 + p_4 = 1, p_i > 0.$$

Si l'hypothèse de Mendel est vraie, alors

$$p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, p_4 = \frac{1}{16}.$$

Pour tester l'hypothèse de Mendel on peut utiliser le test du chi-deux de Pearson, fondé sur la statistique de Pearson

$$X_n^2 = \sum_{i=1}^4 \frac{(v_i - np_i)^2}{np_i},$$

dont la distribution (sous l'hypothèse  $H_0$ ) est proche de la distribution du chi-deux à  $f = 4 - 1 = 3$  degrés de liberté. Choisissons  $\alpha = 0.05$ . Dans ce cas la valeur critique  $c_\alpha = \chi_3^2(\alpha) = 7.81$ . Comme pour les données de Mendel

$$X_n^2 = 0.470 < \chi_3^2(0.05) = 7.81,$$

nous ne rejetons pas l'hypothèse  $H_0$ , considérant que les données de Mendel sont en bon accord avec son hypothèse.

**Remarque 1.** On croise différentes variétés de petits pois. A la première génération, on obtient les différentes catégories suivantes : Ronds et Jaunes, Ronds et verts, ridés et Jaunes, ridés et verts, sachant que :

dans les caractères de formes

le dominant est rond ( $R$ ), le récessif est ridé ( $r$ ) ;

dans les caractères de couleurs

le dominant est jaune ( $J$ ), le récessif est vert ( $v$ ).

On obtient pour la deuxième génération le tableau suivant :

<i>Parents</i>	<i>RJ</i>	<i>Rv</i>	<i>rJ</i>	<i>rv</i>
<i>RJ</i>	<i>RJ</i>	<i>RJ</i>	<i>RJ</i>	<i>RJ</i>
<i>Rv</i>	<i>RJ</i>	<i>Rv</i>	<i>RJ</i>	<i>Rv</i>
<i>rJ</i>	<i>RJ</i>	<i>RJ</i>	<i>rJ</i>	<i>rJ</i>
<i>rv</i>	<i>RJ</i>	<i>Rv</i>	<i>rJ</i>	<i>rv</i>

Soit  $RJ, Rv, rJ$  et  $rv$  dans les proportions 9,3,3,1.

On vient de considérer le test de Pearson pour le cas où les probabilités  $p_i$  sont connues, ou, comme on dit, pour des *hypothèses simples*. La situation devient un peu plus compliquée, quand les  $p_i$  sont inconnues ou dépendent d'un paramètre  $\theta$  inconnu,  $p_i = p_i(\theta)$ . Il y a des possibilités différentes pour tester  $H_0$ , dont on dit qu'elle est *composée*. Dans le paragraphe suivant nous allons parler d'une solution de Fisher et Cramer.

## 4.4 Théorème de Fisher.

### Conditions de Cramer et methode du minimum de chi-deux.

Soit  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  un échantillon ; supposons que nous voulions tester l'hypothèse  $H_0$ , selon laquelle les variables aléatoires indépendantes  $X_1, \dots, X_n$  suivent la même loi

$$\mathbf{P}\{X_i \leq x\} = F(x, \theta), \quad \theta = (\theta_1, \dots, \theta_s)^T \in \Theta \subset \mathbf{R}^s,$$

où la fonction de la répartition  $F$  est donnée, mais le paramètre  $\theta$  est inconnu. En posant  $x_0 = -\infty$  et  $x_k = \infty$ , notons  $\mathbf{v} = (v_1, \dots, v_k)^T$  le vecteur des fréquences que nous obtenons comme résultat du groupement des variables aléatoires sur les  $k$  intervalles ( $k \geq s + 2$ )

$$(x_0, x_1], (x_1, x_2], \dots, (x_{k-1}, x_k),$$

qui sont choisis d'avance. Il est évident que  $\mathbf{v}^T \mathbf{1}_k = n$ , et si l'hypothèse  $H_0$  est vraie, alors le vecteur  $\mathbf{v}$  suit la loi multinomiale de paramètres  $n$  et  $\mathbf{p}$ , où

$$\mathbf{p} = \mathbf{p}(\theta) = (p_1(\theta), p_2(\theta), \dots, p_k(\theta))^T$$

et

$$p_i(\theta) = \mathbf{P}\{X_1 \in (x_{i-1}, x_i] \mid H_0\} = \int_{x_{i-1}}^{x_i} dF(x, \theta) = \int_{x_{i-1}}^{x_i} f(x, \theta) dx,$$

où  $f(x, \theta)$  est la densité de  $F(x, \theta)$ , si elle existe. Supposons que les conditions suivantes de Cramer soient satisfaites :

1) il existe un nombre positif  $c$  ( $c > 0$ ) tel que pour tout  $i = 1, \dots, k$

$$p_i(\theta) > c, \quad \theta \in \Theta;$$

2) les fonctions  $\frac{\partial^2 p_i(\theta)}{\partial \theta_j^2}$  sont continues sur  $\Theta$  ;

3) le rang de la matrice d'information de Fisher  $\mathbf{J}(\theta) = \mathbf{B}(\theta)^T \mathbf{B}(\theta)$ ,

$$\mathbf{B} = \left\| \frac{1}{\sqrt{p_i}} \frac{\partial p_i(\theta)}{\partial \theta_j} \right\|,$$

est égal à  $s$ .

Comme le paramètre  $\theta$  est inconnu, Fisher a proposé de choisir pour estimateur de  $\theta$  le  $\tilde{\theta}_n$  qui rend minimum la variable aléatoire

$$X^2(\theta) = \sum_{i=1}^k \frac{[v_i - np_i(\theta)]^2}{np_i(\theta)}$$

i.e.

$$X^2(\tilde{\theta}_n) = \min_{\theta \in \Theta} X^2(\theta).$$

On dit que  $\tilde{\theta}_n$  est l'estimateur du minimum de chi-deux. Comme Fisher l'a prouvé (1928), si l'hypothèse  $H_0$  est vraie, alors pour tout  $x$  fixé

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X^2(\tilde{\theta}_n) \leq x\} = \mathbf{P}\{\chi_{k-s-1}^2 \leq x\}.$$

Cramer a démontré plus tard (1946) que le résultat de Fisher reste valable si au lieu de  $\tilde{\theta}_n$  on choisit l'estimateur de maximum de vraisemblance  $\theta_n^* = \theta_n^*(v_1, v_2, \dots, v_k)$ , qui rend maximum la fonction de vraisemblance :

$$l(\theta_n^*) = \sup_{\theta \in \Theta} l(\theta),$$

où

$$l(\theta) = \frac{n!}{v_1!v_2!\dots v_k!} (p_1(\theta))^{v_1} (p_2(\theta))^{v_2} \dots (p_k(\theta))^{v_k}.$$

On voit bien que l'estimateur  $\theta_n^*$  est obtenu à partir des données groupées, et, si la distribution  $F(x, \theta)$  est continue, alors la statistique  $v = (v_1, \dots, v_k)^T$  n'est pas exhaustive et par conséquent l'estimateur  $\theta_n^*$  n'est pas le meilleur, mais comme on l'a déjà dit

$$\lim_{n \rightarrow \infty} \{X^2(\theta_n^*) \leq x \mid H_0\} = \mathbf{P}\{\chi_{k-s-1}^2 \leq x\}.$$

**Exemple 1.** Il a été établi qu'au cours d'une épidémie de grippe, parmi les 2000 individus contrôlés, 181 personnes sont tombées malades une seule fois et seulement 9 personnes ont eu cette maladie deux fois. L'hypothèse  $H_0$  selon laquelle le nombre de fois où une personne tombe malade est une variable aléatoire qui suit une loi binomiale de paramètres  $p$  et  $n = 2$  ( $0 < p < 1$ ) est-elle vraisemblable ?

Soit  $X$  une variable aléatoire de loi binomiale  $B(2, p)$ , c'est-à-dire que

$$\mathbf{P}\{X = i\} = \binom{2}{i} p^i (1-p)^{2-i}, \quad i = 0, 1, 2, \quad 0 < p < 1.$$

Et soit  $v = (v_0, v_1, v_2)^T$  le vecteur des fréquences observées, où  $v_i$  est le nombre des individus qui sont tombés malades  $i$  fois,

$$v_0 + v_1 + v_2 = n = 2000, \quad v_0 = 1810, \quad v_1 = 181, \quad v_2 = 9.$$

Notons  $p$  la probabilité de tomber malade ( $0 < p < 1$ ) et soit  $l(p)$  la fonction de vraisemblance :

$$l(p) = \frac{n!}{v_0!v_1!v_2!} [(1-p)^2]^{v_0} [2p(1-p)]^{v_1} (p^2)^{v_2} = \frac{n!2^{v_1}}{v_0!v_1!v_2!} (1-p)^{2v_0+v_1} p^{2v_2+v_1}.$$

Il est facile de voir que les meilleurs estimateurs sans biais pour les probabilités

$$p_0 = p^2, \quad p_1 = p(1-p) \quad \text{et} \quad p_2 = (1-p)^2$$

sont

$$\tilde{p}_0 = \frac{(v_1 + 2v_2)(v_1 + 2v_2 - 1)}{2n(2n - 1)}, \quad \tilde{p}_1 = \frac{(v_1 + 2v_2)(v_1 + 2v_0)}{2n(2n - 1)},$$

$$\tilde{p}_2 = \frac{(v_1 + 2v_0)(v_1 + 2v_0 - 1)}{2n(2n - 1)}$$

respectivement, dont les réalisations observées sont

$$\tilde{p}_0 = \frac{199 \cdot 198}{4000 \cdot 3999} = \frac{4.9}{2000}, \tilde{p}_1 = \frac{199 \cdot 3801}{4000 \cdot 3999} = \frac{94.6}{2000},$$

$$\tilde{p}_2 = \frac{3801 \cdot 3800}{4000 \cdot 3999} = \frac{1805.9}{2000},$$

d'où l'on tire que

$$n\tilde{p}_0 = 4.9; 2n\tilde{p}_1 = 189.2; n\tilde{p}_3 = 1805.9.$$

Pour tester  $H_0$  on va utiliser le test du chi-deux, fondé sur la statistique de Pearson  $X^2$  qui dans notre cas est distribuée approximativement (si l'hypothèse  $H_0$  est vraie) comme la variable aléatoire  $\chi_f^2$  avec  $f = 3 - 1 - 1 = 1$  degrés de liberté. On a

$$X^2 = \sum_{i=0}^2 \frac{(v_i - n\tilde{p}_i)^2}{n\tilde{p}_i} =$$

$$= \frac{(1810 - 1805.9)^2}{1805.9} + \frac{(181 - 189.2)^2}{189.2} + (9 - 4.9)^2 4.9 =$$

$$= \frac{(4.1)^2}{1805.9} + \frac{(8.2)^2}{189.2} + \frac{(4.1)^2}{4.9} \cong 3.795 < \chi_1^2(0.05) = 3.841,$$

où  $\chi_1^2(0.05) = 3.841$  est le quantile du niveau 0.05 de la distribution du chi-deux à 1 degré de liberté :

$$\mathbf{P}\{\chi_1^2 > \chi_1^2(0.05)\} = 0.05.$$

Comme  $X^2$  est inférieur à la valeur critique 3.841, on ne rejette pas l'hypothèse  $H_0$ .

**Exemple 2.** Parmi 2020 familles ayant deux enfants on a enregistré 530 familles où les deux enfants sont des garçons et 473 familles où les deux enfants sont des filles, et dans les 1017 familles restantes les enfants sont de sexe différent. Peut-on dire, avec le niveau de signification  $\alpha = 0.1$ , que le nombre des garçons dans une famille de deux enfants est une variable aléatoire qui suit une loi binomiale ? Les probabilités de naissance d'un garçon et d'une fille sont-elles égales ?

Soit  $X$  une variable aléatoire qui suit la loi binomiale  $B(2, p)$ , c'est-à-dire que

$$\mathbf{P}\{X = i\} = \binom{2}{i} p^i (1-p)^{2-i}, i = 0, 1, 2, 0 < p < 1.$$

De plus soit  $\mathbf{v} = (v_0, v_1, v_2)^T$  le vecteur des fréquences observées, où  $v_i$  est le nombre de familles où il y a  $i$  garçons,  $i = 0, 1, 2$ . Dans notre cas

$$v_0 + v_1 + v_2 = n = 2020, v_1 = 1017, v_0 = 473, v_2 = 530,$$

et donc si l'hypothèse de la binomialité est vraie, alors la fonction de la vraisemblance  $l(p)$  peut s'écrire :

$$l(p) = \frac{n!}{v_0!v_1!v_2!} [(1-p)^2]^{v_0} [2p(1-p)]^{v_1} (p^2)^{v_2},$$

où  $p$  est la probabilité de naissance d'un garçon.

Comme on le sait, les meilleurs estimateurs sans biais pour les probabilités

$$p_0 = p^2, p_1 = p(1-p) \text{ et } p_2 = (1-p)^2$$

sont

$$\tilde{p}_0 = \frac{(v_1 + 2v_2)(v_1 + 2v_2 - 1)}{2n(2n-1)}, \quad \tilde{p}_1 = \frac{(v_1 + 2v_2)(v_1 + 2v_0)}{2n(2n-1)},$$

$$\tilde{p}_2 = \frac{(v_1 + 2v_0)(v_1 + 2v_0 - 1)}{2n(2n-1)}$$

respectivement, dont les réalisations observées sont

$$\tilde{p}_0 = \frac{2077 \cdot 2076}{4040 \cdot 4039}, \quad \tilde{p}_1 = \frac{2077 \cdot 1963}{4040 \cdot 4039}, \quad \tilde{p}_2 = \frac{1963 \cdot 1962}{4040 \cdot 4039}$$

d'où l'on tire que

$$n\tilde{p}_0 \cong 533.8; \quad 2n\tilde{p}_1 = 1009.4; \quad n\tilde{p}_2 = 476.8.$$

Pour tester  $H_0$  on va utiliser le test du chi-deux, fondé sur la statistique de Pearson  $X^2$  qui dans notre cas est distribuée approximativement (sous l'hypothèse  $H_0$ ) comme une variable aléatoire  $\chi_f^2$  à  $f = 3 - 1 - 1 = 1$  degrés de liberté. On a

$$X^2 = \sum_{i=0}^2 \frac{(v_i - n\tilde{p}_i)^2}{n\tilde{p}_i} =$$

$$= \frac{(473 - 476.8)^2}{476.8} + \frac{(1017 - 1009.4)^2}{1009.4} + \frac{(530 - 533.8)^2}{533.8} =$$

$$= \frac{(3.8)^2}{476.8} + \frac{(7.6)^2}{1009.4} + \frac{(3.8)^2}{533.8} < 1 < \chi_1^2(0.1) = 2.706,$$

où  $\chi_1^2(0.1) = 2.706$  est le quantile du niveau 0.1 de la distribution de chi-deux à 1 degré de liberté :

$$\mathbf{P}\{\chi_1^2 > \chi_1^2(0.1)\} = 0.10.$$

Comme  $X^2$  est inférieur à la valeur critique 2.706, on constate que les données ne sont pas en contradiction avec l'hypothèse  $H_0$ , d'après laquelle le nombre des garçons dans une famille est une réalisation d'une variable aléatoire  $X$ , qui suit la loi binomiale  $B(2, p)$ .

Si les probabilités de naissance d'un garçon et d'une fille sont égales, la probabilité  $p$  est égale à 0.5 (l'hypothèse  $H_1$ ). Dans ce cas, d'après le théorème de de Moivre-Laplace, on obtient

$$\mathbf{P}\{v_1 + v_2 \geq 2077 \mid p = 0.5\} \cong 1 - \Phi \left\{ \frac{2077 - 0.5 - \frac{4040}{2}}{\sqrt{4040 * 0.5 * 0.5}} \right\} =$$

$$= 1 - \Phi \left\{ 1132\sqrt{1010} \right\} = 1 - \Phi \left\{ \frac{113}{63.56} \right\} = 1 - \Phi(1.778) = 1 - 0.9623 = 0.0377.$$

Pour tous les niveaux  $\alpha \geq 0.04$  on est obligé de rejeter l'hypothèse  $H_1 : p = 0.5$  en faveur de l'hypothèse  $H_2 : p > 0.5$ . Comme nos calculs le montrent, le meilleur estimateur sans biais de  $p$  est

$$\tilde{p} = \frac{2077}{4040} = 0.514.$$

## 4.5 Théorème de Chernoff-Lehmann.

Soit  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  un échantillon ; supposons que nous voulions tester l'hypothèse  $H_0$ , selon laquelle les variables aléatoires indépendantes  $X_1, \dots, X_n$  suivent la même loi

$$\mathbf{P}\{X_i \leq x\} = F(x, \theta), \quad \theta = (\theta_1, \dots, \theta_s)^T \in \Theta \subset \mathbf{R}^s,$$

où la fonction de la répartition  $F$  est donnée, mais le paramètre  $\theta$  est inconnu. En posant  $x_0 = -\infty$  et  $x_k = \infty$ , notons  $\mathbf{v} = (v_1, \dots, v_k)^T$  le vecteur des fréquences que nous obtenons comme résultat du groupement des variables aléatoires sur les  $k$  intervalles ( $k > 2$ )

$$(x_0, x_1], (x_1, x_2], \dots, (x_{k-1}, x_k),$$

qui sont choisis d'avance. Si l'hypothèse  $H_0$  est vraie, alors le vecteur  $\mathbf{v}$  suit la loi multinomiale de paramètres  $n$  et  $\mathbf{p}$ , où

$$\mathbf{p} = \mathbf{p}(\theta) = (p_1(\theta), p_2(\theta), \dots, p_k(\theta))^T,$$

$$p_i(\theta) = \mathbf{P}\{X_1 \in (x_{i-1}, x_i] \mid H_0\} = \int_{x_{i-1}}^{x_i} dF(x, \theta) = \int_{x_{i-1}}^{x_i} f(x, \theta) d\mu(x),$$

où  $f(x, \theta)$  est la densité de  $F(x, \theta)$  par rapport à une mesure dominante  $\mu$ .

Supposons que la matrice d'information de Fisher existe :

$$\mathbf{I}(\theta) = \mathbf{E}\Lambda_i(\theta)\Lambda_i^T(\theta)$$

pour l'observation  $X_i$ , où

$$\Lambda_i(\theta) = \left( \frac{\partial \ln(X_i, \theta)}{\partial \theta_1}, \frac{\partial \ln(X_i, \theta)}{\partial \theta_2}, \dots, \frac{\partial \ln(X_i, \theta)}{\partial \theta_s} \right)^T,$$

et que les conditions de Cramer 1)-3) du paragraphe précédent sont satisfaites. Dans ce cas, il existe un estimateur  $\hat{\theta}_n$  de maximum de vraisemblance basé sur les données initiales,  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ , qui maximise la fonction de vraisemblance

$$L(\theta) = f(X_1, \theta)f(X_2, \theta) \cdots f(X_n, \theta) : \quad L(\hat{\theta}_n) = \sup_{\theta \in \Theta} L(\theta).$$

Sous des conditions supposées de régularité sur la famille  $\{F(x, \theta)\}$  on connaît le comportement asymptotique de la suite  $\{\hat{\theta}_n\}$ , quand  $n \rightarrow \infty$  (voir, par exemple, Barra (1971), Rao (1973)) :

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{I}^{-1}(\theta)\Lambda_i(\theta) + o_{\mathbf{p}}(\mathbf{1}_s),$$

d'où on obtient immédiatement que le vecteur  $\sqrt{n}(\hat{\theta}_n - \theta)$  a une distribution asymptotiquement normale  $N(\mathbf{0}_s, \mathbf{I}^{-1}(\theta))$ , quand  $n \rightarrow \infty$ .

### **Théorème de Lehmann et Chernoff.**

En utilisant ces propriétés de l'estimateur de maximum de vraisemblance  $\hat{\theta}_n$ , Lehmann et Chernoff ont montré (1954), que sous l'hypothèse  $H_0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \{X^2(\theta) \leq x\} = \mathbf{P} \{\chi_{k-s-1}^2 + \lambda_1(\theta)\xi_1^2 + \dots + \lambda_s(\theta)\xi_s^2 \leq x\},$$

où  $\xi_1, \xi_2, \dots, \xi_s, \chi_{k-s-1}^2$  sont des variables aléatoires indépendantes,  $\xi_i$  suit la loi normale standard  $N(0, 1)$ , et  $0 < \lambda_i(\theta) < 1$ .

**Statistique  $Y_n^2$ .**

D'après ce résultat on constate qu'en général il est impossible d'utiliser la statistique standard de Pearson  $X^2(\hat{\theta}_n)$  pour tester des hypothèses composées, lorsqu'on utilise des estimateurs de maximum de vraisemblances  $\hat{\theta}_n$  ou leurs équivalents. On peut tout de même construire un test du chi-deux pour tester des hypothèses composées. Notons  $\Sigma(\theta)$  la matrice de covariance de la distribution limite du vecteur  $\frac{1}{\sqrt{n}}(\mathbf{v} - n\mathbf{p}(\hat{\theta}_n))$ . On peut montrer (voir, par exemple, Nikulin (1973), Nikulin et Greenwood (1990), Huber (1991)), que  $\text{rang}\Sigma = k - 1$ . Notons  $\Sigma^-(\theta)$  la matrice inverse généralisée de  $\Sigma(\theta)$  et soit

$$Y_n^2 = \frac{1}{n} (\mathbf{v} - n\mathbf{p}(\hat{\theta}_n))^T \Sigma^-(\hat{\theta}_n) (\mathbf{v} - n\mathbf{p}(\hat{\theta}_n)).$$

Par des calculs directs on peut vérifier que la statistique  $Y_n^2$  est indépendante du choix de la matrice  $\Sigma^-$ . On peut utiliser la statistique  $Y_n^2$  pour tester la validité de l'hypothèse  $H_0$  selon laquelle la distribution des éléments  $X_i$  de l'échantillon  $\mathbb{X}$  suit la loi  $F(x, \theta)$ . On a en effet (voir, par exemple, Nikulin (1973), Greenwood et Nikulin (1996)) :

$$\lim_{n \rightarrow \infty} \mathbf{P} \{Y_n^2 \leq x \mid H_0\} = \mathbf{P} \{\chi_{k-1}^2 \leq x\}.$$

Pour plus de détails sur la construction des tests du chi-deux, fondés sur la statistique  $Y_n^2$ , on se reportera aux articles de Nikulin (1973), (1979), (1990), (1991), Dzharidze et Nikulin (1974), Nikulin et Voinov (1989), Greenwood et Nikulin (1996), Nikulin et Seddik-Ameur (1991). On remarque enfin, que dans les cas de l'existence de statistiques exhaustives, on peut utiliser aussi les meilleurs estimateurs sans biais pour construire un test du chi-deux fondé sur la statistique  $Y_n^2$  et en utilisant la technique exposée dans les articles que l'on vient de mentionner.

## 4.6 Test du chi-deux pour une loi logistique.

La loi "logistique", qui a reçu son nom de Berkson et Reed (1929) est souvent utilisée. (Entre autres, par Pearl et Reed (1920) pour le développement des levures, par Oliver (1964) comme modèle de données agricoles et Grizzle (1961) dans le domaine de la santé Publique, etc.)

Cette loi a une fonction de répartition dépendant de deux paramètres  $\mu$  et  $\sigma > 0$  :

$$F(x) = G\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{1 + \exp\left\{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu}{\sigma}\right)\right\}}, x \in \mathbb{R}. \quad (1)$$

Un livre vient d'être publié par Balakrishnan (1992) sur la théorie, méthodologie et applications de cette loi. Ici nous allons suivre l'article de Aguirre et Nikulin (1994).

Soit  $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$  - un échantillon et supposons que nous voulions tester l'hypothèse  $H_0$  selon laquelle

$$\mathbf{P}\{X_i < x\} = G\left(\frac{x-\mu}{\sigma}\right). \quad (2)$$

Dans cette situation nous nous proposons d'utiliser les résultats précédents pour construire un test du chi-deux.

### 1. Notations.

Soit  $g(x) = G'(x)$ , et donc  $\frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right)$  est la densité de  $X_i$  sous  $H_0$ ,

$$\frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right) = \frac{\pi \exp\left\{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right\}}{\sigma\sqrt{3} \left[1 + \exp\left\{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right\}\right]^2}. \quad (3)$$

$g$  est paire ( $g(-x) = g(x)$ ).

### 2. Estimation de $\mu$ et $\sigma$ .

Pour estimer  $\theta = (\mu, \sigma)^T$  on utilise l'estimateur  $\hat{\theta}_n = (\hat{\mu}, \hat{\sigma}^2)^T$  du maximum de vraisemblance. On sait que  $\sqrt{n}(\hat{\theta}_n - \theta)$  est asymptotiquement normal  $N(\mathbf{0}, \mathbf{I}^{-1})$ , où

$$\mathbf{I} = \frac{1}{\sigma^2} \|I_{ij}\|_{i,j=1,2}, \quad I_{11} = \int_{-\infty}^{+\infty} \left[\frac{g'(x)}{g(x)}\right]^2 g(x) dx = \frac{\pi^2}{9}$$

$$I_{12} = I_{21} = \int_{-\infty}^{+\infty} x \left[\frac{g'(x)}{g(x)}\right]^2 g(x) dx = 0,$$

$$I_{22} = \int_{-\infty}^{+\infty} x^2 \left[\frac{g'(x)}{g(x)}\right]^2 g(x) dx - 1 = \frac{\pi^2 + 3}{9}.$$

$I_{12} = 0$  car  $g$  est symétrique, et une intégration par parties permet d'obtenir  $I_{11}$  et  $I_{22}$ .

**3. Choix des intervalles** sur lesquels on va comparer les fréquences observées et les fréquences théoriques :

Supposons que l'on ait choisi un vecteur  $\mathbf{p} = (p_1, p_2, \dots, p_k)^T$  de probabilités positives, par exemple :

$$p_1 = \dots = p_k = \frac{1}{k}, \quad y_i = G^{-1}\left(\frac{i}{k}\right) = -\frac{\sqrt{3}}{\pi} \ln\left(\frac{k}{i} - 1\right), \quad i = 1, \dots, k-1,$$

et notons  $\mathbf{v} = (v_1, \dots, v_k)^T$  le vecteur des effectifs que nous obtenons en regroupant les variables aléatoires  $X_1, \dots, X_n$  sur les intervalles

$$(-\infty, z_1], (z_1, z_2], \dots, (z_{k-1}, +\infty), \quad \text{où } z_i = \hat{\mu} + \hat{\sigma}y_i.$$

### 4. Test de $\chi^2$ . Posons

$$\mathbf{a} = (a_1, \dots, a_k)^T, \quad \mathbf{b} = (b_1, \dots, b_k)^T, \quad \mathbf{w} = -\frac{1}{\sigma} \|\mathbf{a}, \mathbf{b}\|, \quad \text{où}$$

$$a_i = g(y_i) - g(y_{i-1}) = \frac{\pi}{k^2\sqrt{3}}(k - 2i + 1),$$

$$b_i = y_i g(y_i) - y_{i-1} g(y_{i-1}) = \frac{1}{k^2} \left[ (i-1)(k-i+1) \ln \frac{k-i+1}{i-1} - i(k-i) \ln \frac{k-i}{i} \right],$$

$$\alpha(\mathbf{v}) = k \sum_{i=1}^k a_i v_i = \frac{\pi}{\sqrt{3}k} \left[ (k+1)n - 2 \sum_{i=1}^k i v_i \right],$$

$$\beta(\mathbf{v}) = k \sum_{i=1}^k b_i v_i = \frac{1}{k} \sum_{i=1}^{k-1} (v_{i+1} - v_i) i(k-i) \ln \frac{k-i}{i},$$

$$\lambda_1 = I_{11} - k \sum_{i=1}^k a_i^2 = \frac{\pi^2}{9k^2}, \quad \lambda_2 = I_{22} - k \sum_{i=1}^k b_i^2.$$

Comme  $g$  est symétrique on remarque que

$$\sum_{i=1}^k a_i = \sum_{i=1}^k b_i = 0.$$

Notons  $\mathbf{B} = \mathbf{D} - \mathbf{p}^T \mathbf{p} - \mathbf{W}^T \mathbf{I}^{-1} \mathbf{W}$ , où  $\mathbf{D}$  est la matrice diagonale avec les éléments  $1/k$  sur la diagonale principale (rang  $\mathbf{B} = k-1$ ). Notons les matrices précédentes dans lesquelles on supprime la dernière ligne pour  $\mathbf{W}$ ,  $\mathbf{p}$  et  $\mathbf{v}$  et les dernières ligne et colonne de  $\mathbf{D}$  et  $\mathbf{B}$ .

**Théorème 1.** *Sous l'hypothèse  $H_0$ , quand  $n \rightarrow \infty$ , le vecteur  $\tilde{\mathbf{v}}$  est asymptotiquement normalement distribué avec les paramètres*

$$\mathbf{E}\tilde{\mathbf{v}} = n\tilde{\mathbf{p}} + O(1) \text{ et } \mathbf{E}(\tilde{\mathbf{v}} - n\tilde{\mathbf{p}})^T (\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}) = n\tilde{\mathbf{B}} + O(1).$$

**Théorème 2.** *Sous l'hypothèse  $H_0$  la statistique*

$$Y_n^2 = \frac{1}{n} (\tilde{\mathbf{v}} - n\tilde{\mathbf{p}})^T \tilde{\mathbf{B}}^{-1} (\tilde{\mathbf{v}} - n\tilde{\mathbf{p}}) = X^2 + \frac{\lambda_1 \beta^2(\mathbf{v}) + \lambda_2 \alpha^2(\mathbf{v})}{n\lambda_1 \lambda_2},$$

converge en loi quand  $n \rightarrow \infty$  vers une distribution de  $\chi_{k-1}^2$ .

**Remarque.** Considérons l'hypothèse  $H_\eta$  selon laquelle  $X_i$  suit la loi  $G(\frac{x-\mu}{\sigma}, \eta)$ , où  $G(x, \eta)$  est continue,  $|x| < \infty$ ,  $\eta \in \mathbf{H}$  et  $G(x, 0) = G(x)$ ,  $\eta = 0$  est un point limite de  $\mathbf{H}$ . De plus, supposons qu'il existe

$$\frac{\partial}{\partial x} G(x, y) = g(x, y) \quad \text{et} \quad \frac{\partial}{\partial \eta} g(x, \eta) \Big|_{\eta=0} = \Psi(x),$$

où  $g(x, 0) = g(x) = G'(x)$ . Dans ce cas si  $\frac{\partial^2 g(x, \eta)}{\partial \eta^2}$  existe et est continue pour tout  $x$  au voisinage de  $\eta = 0$ , alors

$$\mathbf{P}\{y_{i-1} < X_i \leq y_i \mid H_\eta\} = p_i + \eta c_i + o(\eta),$$

$$\text{où} \quad c_i = \int_{y_{i-1}}^{y_i} \Psi(x) dx, \quad i = 1, \dots, k,$$

et donc

$$\lim_{n \rightarrow \infty} \mathbf{P}\{Y^2 \geq x \mid H_\eta\} = \mathbf{P}\{\chi_{k-1}^2(\lambda) \geq x\},$$

$$\lambda = \sum_{i=1}^k \frac{c_i^2}{p_i} + \frac{\lambda_2 \alpha^2(\mathbf{c}) + \lambda_1 \beta^2(\mathbf{c})}{\lambda_1 \lambda_2}, \quad \mathbf{c} = (c_1, c_2, \dots, c_k)^T.$$

Plus de détails on peut trouver dans Acuirre (1993), Acuirre et Nikulin (1994).

## 4.7 Test du chi-deux dans un problème d'homogénéité.

On a  $k$  groupes de souris soumises à des traitements par différents médicaments. Les souris d'un groupe, nommé "groupe de contrôle", ont reçu un médicament, dont les effets ont déjà été étudiés. Pour savoir si d'autres médicaments sont meilleurs ou moins bons, on compare les effets produit par ces médicaments à ceux du "groupe de contrôle". On vérifie l'hypothèse d'homogénéité : cette hypothèse est vraie s'il n'y a pas de changement d'effet. Autrement, l'hypothèse doit être rejetée. Dans ce cas, se pose le problème suivant : trouver les groupes pour lesquels on a des effets différents de ceux du "groupe de contrôle". Soient  $\mu_1, \dots, \mu_k$  des variables aléatoires indépendantes qui suivent la distribution binomiale de paramètres  $(n_1, p_1), \dots, (n_k, p_k)$  respectivement :

$$\mathbf{P}\{\mu_i = m\} = C_{n_i}^m p_i^m (1 - p_i)^{n_i - m}, \quad m \in \{0, 1, \dots, n_i\}, \quad i = 1, 2, \dots, k,$$

où les probabilités  $p_1, \dots, p_k$  sont inconnues ( $0 < p_i < 1; i = 1, \dots, k$ ). Supposons que la variable aléatoire  $\mu_k$  soit donnée pour "le contrôle"; notre but est alors de tester l'hypothèse que toutes les probabilités  $p_1, \dots, p_{k-1}$  ou quelques-unes d'entre elles sont égales à  $p_k$ . Ce problème peut être résolu si l'on suppose que  $\min(n_1, \dots, n_k) \rightarrow \infty$ .

Soit  $\xi_i = \frac{\mu_i}{n_i}, i = 1, \dots, k$ . Alors du théorème de de Moivre-Laplace on peut tirer que

$$\mathbf{P}\{\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_k \leq x_k\} \sim \prod_{i=1}^k \Phi \left[ (x_i - p_i) \sqrt{\frac{n_i}{p_i q_i}} \right],$$

si

$$\left( (x_i - p_i) \sqrt{\frac{n_i}{p_i q_i}} \right) = O(1),$$

où  $\Phi(\cdot)$  est la fonction de répartition de la loi normale  $N(0, 1)$  et  $q_i = 1 - p_i, i = 1, \dots, k$ . Soit  $\eta_i = \xi_i - \xi_k$ , et soit

$$\Delta_i = \mathbf{E}\eta_i = p_i - p_k \text{ et } \sigma_i^2 = \frac{p_i q_i}{n_i}, \quad i = 1, \dots, k.$$

Il est clair que le vecteur aléatoire  $\eta = (\eta_1, \dots, \eta_{k-1})^T$  a une distribution asymptotique normale de paramètres

$$\mathbf{E}\eta = \Delta = (\Delta_1, \dots, \Delta_{k-1})^T \text{ et } \mathbf{E}(\eta - \Delta)(\eta - \Delta)^T = \Sigma,$$

où

$$\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_{k-1}^2)^T + \sigma_k^2 \mathbf{E},$$

$\text{diag}(x_1, \dots, x_n)$  est la matrice diagonale ayant les éléments  $x_1, \dots, x_n$  sur la diagonale principale et  $\mathbf{E}$  est la matrice d'ordre  $(k-1) \times (k-1)$ , dont tous les éléments sont égaux à 1. Nous remarquons que la matrice  $\Sigma$  est non singulière et

$$\Sigma^{-1} = \text{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \dots, \sigma_{k-1}^{-2}) - \left( \sum_{i=1}^k \sigma_i^{-2} \right)^{-1} \|b_{ij}\|,$$

où

$$b_{ij} = \frac{1}{\sigma_i^2 \sigma_j^2}; i, j = 1, \dots, k-1.$$

Du fait que le vecteur  $\eta$  a une distribution asymptotique normale, il s'ensuit que la forme quadratique

$$\mathbf{Y}^2 = (\eta - \Delta)^T \Sigma^{-1} (\eta - \Delta)$$

a à la limite, lorsque  $\min(n_1, \dots, n_k) \rightarrow \infty$ , une distribution du chi-deux à  $k-1$  degrés de liberté.

Cette même forme quadratique peut être représentée sous une forme plus explicite :

$$Y^2 = \sum_{i=1}^{k-1} \left( \frac{\eta_i - \Delta_i}{\sigma_i} \right)^2 - \left( \sum_{i=1}^k \frac{1}{\sigma_i^2} \right)^{-1} \left[ \sum_{i=1}^{k-1} \frac{\eta_i - \Delta_i}{\sigma_i} \right]^2. \quad (1)$$

D'après la théorie générale des tests du chi-deux (voir, par exemple, Greenwood et Nikulin (1996), Nikulin (1991)), la distribution limite de la forme quadratique  $\mathbf{Y}^2$  sera la même si tous les paramètres inconnus  $\sigma_i^2$  sont remplacés par leurs meilleurs estimateurs sans biais

$$\hat{\sigma}_i^2 = \xi_i \frac{(1 - \xi_i)}{(n_i - 1)}, i = 1, \dots, k.$$

Soit  $P$  un coefficient de confiance donné,  $0.5 < P < 1$ , et soit  $x_p$  le quantile de niveau  $P$  de la distribution du chi-deux à  $k-1$  degrés de liberté. Dans ce cas, la probabilité  $\mathbf{P} \{ Y^2 \leq x_p \}$  est approximativement égale à  $P$  et toutes les valeurs du vecteur  $(\Delta_1, \dots, \Delta_{k-1})^T$ , satisfaisant l'inégalité  $Y^2 \leq x_p$ , donnent un intervalle de confiance dont le coefficient de confiance est proche de  $P$ . Ceci peut être utilisé dans la solution du problème proposé.

### Inférences statistiques.

On considère un ensemble d'hypothèses

$$H_r = H_r(i_1, \dots, i_r) : \Delta_{i_1} = \Delta_{i_2} = \dots = \Delta_{i_r} = 0, \\ r = 1, 2, \dots, k-1; 1 \leq i_1 < i_2 < \dots < i_r \leq k-1.$$

Nous dirons que l'hypothèse  $H_r(i_1, \dots, i_r)$  n'est pas contradictoire avec les données de l'expérience s'il existe  $\Delta_i^*$  ( $i \neq i_1, \dots, i_r$ ), pour lequel la valeur de la statistique  $Y^2$  est inférieure à  $x_p$ . Autrement dit, l'hypothèse  $H_r(i_1, \dots, i_r)$  doit être acceptée si, dans l'espace  $(\Delta_1, \dots, \Delta_{k-1})$  de dimension  $k-1$ , l'hyperplan défini par les équations

$$\Delta_{i_1} = \Delta_{i_2} = \dots = \Delta_{i_r} = 0$$

a une intersection non vide avec l'intérieur de l'ellipsoïde défini par l'inégalité  $Y^2 \leq x_p$ .

Le but final est de choisir un sous-ensemble d'hypothèses qui ne sont pas contradictoires avec les données de l'expérience ; puisque certaines hypothèses sont des conséquences des

autres nous ne nous intéresserons dans ce sous-ensemble qu'aux éléments dont l'indice  $r$  est maximal.

Considérons l'hypothèse  $H_r(1 \leq r \leq k-1)$ , et, sans restriction de généralité, supposons que  $i_1 = k-r, i_2 = k-r+1, \dots, i_r = k-1$ . Alors la statistique  $Y^2$  aura la forme

$$Y^2 = \sum_{i=1}^{k-r-1} \left( \frac{\eta_i - \Delta_i}{\sigma_i} \right)^2 + \sum_{i=k-r}^{k-1} \left( \frac{\eta_i}{\sigma_i} \right)^2 - \left[ \sum_{j=1}^{k-r-1} c_j(\eta_j - \Delta_j) + \sum_{j=k-r}^{k-1} c_j \eta_j \right], \quad (2)$$

où

$$c_j = \sigma_j^{-2} \left( \sum_{i=1}^k \sigma_i^{-2} \right)^{-1/2}, \quad j = 1, \dots, k-1.$$

Il est facile de voir que la plus petite valeur de la statistique  $Y^2$  est obtenue au point

$$\Delta_i = \Delta_i^* = \eta_i - c_i \hat{\sigma}_i^2 \left( 1 - \sum_{j=1}^{k-r-1} c_j^2 \sigma_j^2 \right)^{-1} \sum_{j=k-r}^{k-1} c_j \eta_j, \quad i = 1, \dots, k-1,$$

et sa valeur minimale est

$$Y_*^2 = \sum_{i=k-r}^{k-1} \left( \frac{\eta_i}{\hat{\sigma}_i} \right)^2 - \frac{\left( \sum_{i=k-r}^{k-1} c_i \eta_i \right)^2}{\left( 1 + \sum_{i=1}^{k-r-1} c_i^2 \hat{\sigma}_i^2 \right)} \quad (3)$$

(si  $r = k-1$ , alors le dénominateur de la fraction sera supposé à 1). Il est clair que l'hypothèse  $H_r(i_1, \dots, i_r)$  doit être rejetée si  $Y^2 \geq x_p$ .

**Exemple.** Soit

$$k = 4 \text{ et } n_1 = n_2 = n_3 = n_4 = 100, \text{ où } \mu_1 = 20, \mu_2 = 50, \mu_3 = 60 \text{ et } \mu_4 = 40.$$

Alors

$$\xi_1 = 0.2, \xi_2 = 0.5, \xi_3 = 0.6, \xi_4 = 0.4, \eta_1 = -0.2, \eta_2 = 0.1, \eta_3 = 0.2.$$

Si on utilise le meilleur estimateur sans biais  $\hat{\sigma}_i^2 = \xi_i(1 - \xi_i)/n_i$  pour estimer le paramètre inconnu  $\sigma_i^2, i = 1, \dots, 4$ , on obtient

$$\hat{\sigma}_1^2 = 0.0016, \hat{\sigma}_2^2 = 0.0025, \hat{\sigma}_3^2 = 0.0024 \text{ et } \hat{\sigma}_4^2 = 0.0024;$$

d'où

$$\hat{\sigma}_1^{-2} + \hat{\sigma}_2^{-2} + \hat{\sigma}_3^{-2} + \hat{\sigma}_4^{-2} = \frac{22.300}{12},$$

et

$$c_1 = 125 \sqrt{\frac{3}{223}}, c_2 = 80 \sqrt{\frac{3}{223}}, c_3 = \frac{250}{3} \sqrt{\frac{3}{223}}.$$

Puisque la statistique  $Y^2$  a ici approximativement une distribution du chi-deux à trois degrés de liberté, pour  $P = 0.95$  la valeur critique correspondante  $x_p$  est  $x_{0.95} = 7.815$ . Nous allons tester l'hypothèse  $H_r$ .

**A.** Test de l'hypothèse  $H_3(1, 2, 3)$ . En utilisant (3), nous avons

$$Y_*^2 = \frac{0.04}{0.0016} + \frac{0.01}{0.0025} + \frac{0.04}{0.0024} - \left( -125\sqrt{\frac{3}{223}} \times 0.2 + 80\sqrt{\frac{3}{223}} \times 0.1 + \frac{250}{3}\sqrt{\frac{3}{223}} \times 0.2 \right)^2 = 45.665,$$

et comme  $\mathbf{P}\{\chi_3^2 > 45.665\} < 10^{-7}$ , l'hypothèse  $H_3(1, 2, 3)$  doit être rejetée par tous les tests du chi-deux dont le niveau de signification n'est pas inférieur à  $10^{-7}$ .

**B<sub>1</sub>.** Test de l'hypothèse  $H_2(2, 3)$ . Dans ce cas

$$Y_*^2 = \frac{0.01}{0.0025} + \frac{0.04}{0.0024} - \left( 80\sqrt{\frac{3}{223}} \times 0.1 + \frac{250}{3}\sqrt{\frac{3}{223}} \times 0.2 \right)^2 \left( 1 + \frac{125 \times 125 \times 3}{223} \times \frac{16}{10.00} \right)^{-1} = 14.541.$$

Comme  $\mathbf{P}\{\chi_3^2 > 14.541\} = 0.00225$ , l'hypothèse  $H_2(2, 3)$  doit être rejetée par tous les tests du chi-deux dont le niveau de signification n'est pas inférieur à 0.00225.

**B<sub>2</sub>.** Test de l'hypothèse  $H_2(1, 3)$ . Comme

$$Y_*^2 = \frac{0.04}{0.0016} + \frac{0.04}{0.0024} - \left( -125\sqrt{\frac{3}{223}} \times 0.2 + \frac{250}{3}\sqrt{\frac{3}{223}} \times 0.2 \right)^2 \left( 1 + \frac{80 \times 80 \times 3}{223} \times \frac{25}{10.000} \right)^{-1} = 40.898,$$

l'hypothèse  $H_2(1, 3)$  doit être rejetée par tous les tests du chi-deux dont le niveau de signification n'est pas inférieur à  $\mathbf{P}\{\chi_3^2 > 40.898\} < 10^{-7}$ .

**B<sub>3</sub>.** Test de l'hypothèse  $H_2(1, 2)$ . Dans ce cas

$$Y_*^2 = \frac{0.04}{0.0016} + \frac{0.01}{0.0025} - \left( -125\sqrt{\frac{3}{223}} \times 0.2 + 80\sqrt{\frac{3}{223}} \times 0.1 \right)^2 \left( 1 + \frac{250 \times 250 \times 3}{3 \times 3 \times 223} \times \frac{24}{10.000} \right)^{-1} = 25.824.$$

Puisque la valeur minimale  $Y_*^2$  de la statistique  $Y^2$  dépasse la valeur critique  $x_{0.95} = 7.815$ , l'hypothèse  $H_2(1, 2)$  doit aussi être rejetée.

**C<sub>1</sub>.** Test de l'hypothèse  $H_1(1)$ . Comme

$$Y_*^2 = \frac{0.04}{0.0016} - \left( -125\sqrt{\frac{3}{223}} \times 0.2 \right)^2 \left( 1 + \frac{48}{223} + \frac{50}{223} \right)^{-1} = 19.159 > 7.815,$$

cette hypothèse doit être rejetée aussi.

C<sub>2</sub>. Test de l'hypothèse  $H_1(2)$ . Ici la plus petite valeur de la statistique  $Y^2$  est égale à

$$Y_*^2 = \frac{0.01}{0.0025} - \left(80\sqrt{\frac{3}{223}} \times 0.1\right)^2 \left(1 + \frac{75}{223} + \frac{50}{223}\right)^{-1} = 3.448,$$

ce qui est sensiblement plus petit que la valeurs critique choisie  $x_{0.95}$ , c'est pourquoi l'hypothèse  $H_1(2)$  n'est pas rejetée ; nous obtenons les estimateurs nouveaux

$$\begin{aligned} \Delta_1^* &= -0.2 - 80\sqrt{\frac{3}{223}} \times 0.1 \times 125\sqrt{\frac{3}{223}} \times 0.0016 \left(1 - \frac{75}{223} - \frac{50}{223}\right)^{-1} = \\ &= -0.249, \end{aligned}$$

et

$$\begin{aligned} \Delta_3^* &= -0.2 - 80\sqrt{\frac{3}{223}} \times 0.1 \times \frac{250}{3}\sqrt{\frac{3}{223}} \times 0.0024 \left(1 - \frac{75}{223} - \frac{50}{223}\right)^{-1} = \\ &= 0.151. \end{aligned}$$

C<sub>3</sub>. Test de l'hypothèse  $H_1(3)$ . Puisque

$$Y_*^2 = \frac{0.04}{0.0024} - \left(\frac{250}{3}\sqrt{\frac{3}{223}} \times 0.2\right)^2 \left(1 + \frac{75}{223} + \frac{48}{223}\right)^{-1} = 14.258,$$

alors la plus petite valeur dépasse la valeur critique et l'hypothèse  $H_1(3)$  est rejetée.

Conclusion : seule l'hypothèse  $H_1(2)$  peut-être acceptée d'après les résultats de l'expérience, d'où il s'ensuit que  $\Delta_2 = p_2 - p_4 = 0$ , i.e.  $p_2 = p_4$ . Si cette hypothèse est vraie , il est raisonnable de prendre comme estimateur de  $p_4$  la valeur de la statistique  $(\mu_2 + \mu_4)/(n_2 + n_4)$  ; dans l'exemple présent cette quantité est égale à  $(\xi_2 + \xi_4)/2 = 0.45$ . Puisque

$$p_1 - p_4 \approx \Delta_1^* = -0.249 \text{ et } p_3 - p_4 \approx \Delta_3^* = 0.151,$$

nous avons  $p_1 \approx 0.201$  et  $p_3 \approx 0.601$ .

**Remarque.** Pour utiliser cette approche, dans le cas général on doit tester

$$\sum_{r=1}^{k-1} C_{k-1}^r = 2^{k-1} - 1$$

hypothèses. Dans la pratique pourtant il suffit de tester  $k - 1$  hypothèses. Pour cela, il est nécessaire de calculer les relations

$$\frac{\eta_1^2}{\hat{\sigma}_1^2}, \dots, \frac{\eta_{k-1}^2}{\hat{\sigma}_{k-1}^2}$$

et de les ranger en une suite non décroissante

$$\left(\frac{\eta_1}{\hat{\sigma}_1}\right)^2 \geq \left(\frac{\eta_2}{\hat{\sigma}_2}\right)^2 \geq \dots \geq \left(\frac{\eta_{k-1}}{\hat{\sigma}_{k-1}}\right)^2$$

(les numéros peuvent être donnés après le rangement). Alors on teste successivement les hypothèses  $H_r = H_r(k-r, k-r+1, \dots, k-1)$  avec  $r = k-1, k-2, \dots$ . Si, en agissant de cette façon, on trouve que

$$\left(\frac{\eta_m}{\sigma_m}\right)^2 > \left(\frac{\eta_{m+1}}{\sigma_{m+1}}\right)^2 = \dots = \left(\frac{\eta_{m+t}}{\sigma_{m+t}}\right)^2 > \left(\frac{\eta_{m+t+1}}{\sigma_{m+t+1}}\right)^2$$

et qui l'hypothèse  $H_{k-m}$  est rejetée, alors il faut tester ensuite l'hypothèse  $H_{k-m-t}$  et non  $H_{k-m-1}$ .

On remarque enfin que Bolshev et Nikulin (1975) ont considéré la solution d'un problème de homogénéité plus général pour des distributions dépendant de paramètres de translation et d'échelle.

## 4.8 Test du $\chi^2$ d'homogénéité pour des lois multinomiales.

Observons  $I$  vecteurs aléatoires indépendants

$$\mu_1 = (\mu_{11}, \dots, \mu_{1r})^T, \mu_2 = (\mu_{21}, \dots, \mu_{2r})^T, \dots, \mu_I = (\mu_{I1}, \dots, \mu_{Ir})^T,$$

avec l'hypothèse  $H$  :

$$\mu_i \sim M_r(n_i, \mathbf{p}_i), \quad (1)$$

où  $n_1, n_2, \dots, n_I$  sont des entiers positifs,  $\mathbf{p}_i = (p_{i1}, \dots, p_{ir})^T \in \mathbf{R}^r$ ,

$$p_{i1} + p_{i2} + \dots + p_{ir} = 1, \quad i = 1, 2, \dots, I. \quad (2)$$

Puisque les vecteurs  $\mu_1, \dots, \mu_I$  sont indépendants, alors, sous l'hypothèse  $H$ , la fonction de vraisemblance  $L(\mathbf{p}_1, \dots, \mathbf{p}_I)$  est

$$L(\mathbf{p}_1, \dots, \mathbf{p}_I) = \frac{n_1! n_2! \dots n_I!}{\mu_{11}! \dots \mu_{1r}! \mu_{21}! \dots \mu_{2r}! \dots \mu_{I1}! \dots \mu_{Ir}!} p_{11}^{\mu_{11}} \dots p_{1r}^{\mu_{1r}} p_{21}^{\mu_{21}} \dots p_{2r}^{\mu_{2r}} \dots p_{I1}^{\mu_{I1}} \dots p_{Ir}^{\mu_{Ir}}. \quad (3)$$

Si nous supposons que toutes les probabilités  $p_{ij}$  sont connues, alors, d'après le théorème de Pearson, la statistique

$$X^2 = \sum_{i=1}^I \sum_{j=1}^r \frac{(\mu_{ij} - n_i p_{ij})^2}{n_i p_{ij}} \quad (4)$$

a pour distribution limite lorsque  $n \rightarrow \infty$  la distribution du  $\chi^2$  avec  $f = I(r-1)$  degrés de liberté :

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X^2 \leq x \mid H\} = \mathbf{P}\{\chi_{I(r-1)}^2 \leq x\}.$$

Supposons maintenant que tous les vecteurs  $\mathbf{p}_i$  sont inconnus. Dans ce cas, nous devons estimer  $I(r-1)$  paramètres  $p_{ij}$ . Sous l'hypothèse  $H$ , les estimateurs de vraisemblance des  $p_{ij}$  sont

$$\hat{p}_{ij} = \frac{\mu_{ij}}{N}, \quad j = 1, 2, \dots, r; \quad i = 1, 2, \dots, I, \quad (4)$$

où  $N = n_1 + n_2 + \dots + n_I$ . Supposons qu'on fasse l'hypothèse  $H_0$  :

$$\mathbf{p}_1 = \mathbf{p}_2 = \dots = \mathbf{p}_I = \mathbf{p}, \quad (5)$$

cela signifie que, sous cette hypothèse  $H_0$  toutes les distributions multinomiales (2) des vecteurs aléatoires  $\mu_i$  ont le même vecteur de probabilités  $\mathbf{p} = (p_1, \dots, p_r)^T$  qu'on a besoin d'estimer, si nous voulons faire de l'inférence statistique. Il est évident que sous l'hypothèse  $H_0$ , on a seulement besoin d'estimer  $r - 1$  paramètres  $p_1, p_2, \dots, p_{r-1}$ , puisque

$$p_1 + p_2 + \dots + p_r = 1.$$

Pour tester  $H_0$  on peut construire *le test d'homogénéité du  $\chi^2$*  bien connu, basé sur la variable aléatoire de Pearson (le paramètre  $\mathbf{p}$  est inconnu !), qui sous  $H_0$  peut s'écrire :

$$X^2 = \sum_{i=1}^I \sum_{j=1}^r \frac{(\mu_{ij} - n_i p_j)^2}{n_i p_j}. \quad (6)$$

Tout d'abord réécrivons la fonction de vraisemblance  $L(\mathbf{p})$  de nos données sous  $H_0$ . En utilisant (3) et en posant

$$\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_r)^T = \mu_1 + \dots + \mu_I, \quad (7)$$

où

$$\mathbf{v}_j = \sum_{i=1}^I \mu_{ij}, j = 1, 2, \dots, r \text{ and } \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_r = N, \quad (8)$$

on obtient d'après (3), (5) et (8) que

$$L(\mathbf{p}) = \frac{N!}{\mathbf{v}_1! \mathbf{v}_2! \dots \mathbf{v}_r!} p_1^{\mathbf{v}_1} p_2^{\mathbf{v}_2} \dots p_r^{\mathbf{v}_r}. \quad (9)$$

Pour trouver l'estimateur de maximum de vraisemblance  $\hat{\mathbf{p}}$  de  $\mathbf{p}$  sous  $H_0$ , on considère :

$$\ln L(\mathbf{p}) = \ln(\text{const}) + \sum_{i=1}^r \mathbf{v}_i \ln p_i, \quad (10)$$

d'où nous obtenons le système

$$\frac{\partial}{\partial p_j} L(\mathbf{p}) = \frac{\mathbf{v}_j}{p_j} - \frac{\mathbf{v}_r}{p_r} = 0, j = 1, 2, \dots, r-1, \quad (11)$$

pour lequel la solution est  $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_r)^T$ ,  $\hat{p}_r = 1 - \hat{p}_1 - \hat{p}_2 - \dots - \hat{p}_{r-1}$ , où

$$\hat{p}_j = \frac{\mathbf{v}_j}{N}, j = 1, 2, \dots, r. \quad (12)$$

Par suite, de (12) on obtient :

$$p_r \mathbf{v}_j = \mathbf{v}_r p_j, j = 1, 2, \dots, r, \quad (13)$$

ce qui implique

$$p_r \sum_{j=1}^r v_j = v_r \sum_{j=1}^r p_j, \quad (14)$$

d'où

$$\hat{p}_r = \frac{v_r}{N}. \quad (15)$$

En substituant (15) dans (13) on obtient (12).

Nous pouvons maintenant, pour tester  $H_0$ , utiliser la statistique de Pearson :

$$X^2(\hat{\mathbf{p}}) = \sum_{i=1}^I \sum_{j=1}^r \frac{(\mu_{ij} - n_i \hat{p}_j)^2}{n_i \hat{p}_j} = N \left( \sum_{i=1}^I \sum_{j=1}^r \frac{\mu_{ij}^2}{n_i v_j} - 1 \right). \quad (16)$$

D'après le théorème de Cramer,

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X^2(\hat{\mathbf{p}}) \leq x \mid H_0\} = \mathbf{P}\{\chi_{(I-1)(r-1)}^2 \leq x\}, \quad (17)$$

puisque le nombre de paramètres estimés est  $r - 1$ , d'où

$$f = I(r - 1) - (r - 1) = (I - 1)(r - 1), \quad (18)$$

et  $f$  est le nombre de degrés de liberté de la distribution limite du  $\chi^2$ .

**Exemple 1.** Supposons que deux groupes de 300 étudiants chacun passent le même examen. Dans le 1<sup>er</sup> groupe 144 étudiants obtiennent une très bonne note, 80 une bonne note, 43 une note passable, et 33 une mauvaise note. Pour le second groupe, la distribution est la suivante : 154 très bonnes notes, 72 bonnes, 35 moyennes et 39 mauvaises. Pouvons nous dire que les 2 groupes sont homogènes, ce qui signifie que nous avons observé les réalisations de 2 vecteurs aléatoires ayant la même distribution discrète ?

On peut présenter les données à l'aide du tableau suivant :

$i$	$\mu_{i1}$	$\mu_{i2}$	$\mu_{i3}$	$\mu_{i4}$
1	144	80	43	33
2	154	72	35	39

(19)

Soient  $\mu_i = (\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4})^T$  ( $i = 1, 2$ ) les 2 vecteurs aléatoires dont les réalisations sont présentées dans le tableau et soit  $H_0$  l'hypothèse nulle selon laquelle  $\mu_1$  and  $\mu_2$  ont la même distribution multinomiale  $M_4(300, \mathbf{p})$ , où  $\mathbf{p}$  est un vecteur inconnu de probabilités  $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$ , avec  $p_1 + p_2 + p_3 + p_4 = 1$ . Sous l'hypothèse  $H_0$ , l'estimateur de maximum de vraisemblance de  $\mathbf{p}$  est  $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)^T$ , où

$$\hat{p}_1 = \frac{298}{600}, \hat{p}_2 = \frac{152}{600}, \hat{p}_3 = \frac{78}{600}, \hat{p}_4 = \frac{72}{600}, \quad (20)$$

puisque dans l'exemple :

$$N = n_1 + n_2 = 300 + 300 = 600,$$

$$v_1 = \mu_{11} + \mu_{21} = 298, v_2 = \mu_{12} + \mu_{22} = 152,$$

$$v_3 = \mu_{13} + \mu_{23} = 78, v_4 = \mu_{14} + \mu_{24} = 72,$$

et  $\hat{p}_i = v_i/N$ . Pour tester  $H_0$  on peut construire un test du  $\chi^2$ , basé sur la statistique (16). D'après nos données nous avons :

$$\begin{aligned} X^2(\hat{\mathbf{p}}) &= 2 \left\{ \frac{(33 - 300 \frac{72}{600})^2}{36} + \frac{(43 - 300 \frac{78}{600})^2}{39} + \frac{(80 - 300 \frac{152}{600})^2}{76} + \frac{(144 - 300 \frac{298}{600})^2}{149} \right\} = \\ &= 2 \left( \frac{9}{36} + \frac{16}{39} + \frac{16}{76} + \frac{25}{149} \right) < 2 \left( \frac{1}{4} + \frac{1}{2} + \frac{1}{4} + \frac{1}{5} \right) = 2.4 < \chi_3^2(0.05) = 7.815. \end{aligned}$$

Puisque

$$X^2(\hat{\mathbf{p}}) < \chi_3^2(0.05) = 7.815, \quad (21)$$

on peut accepter  $H_0$ , si on prend  $\alpha = 0.05$ .

## 4.9 Test du $\chi^2$ pour l'indépendance dans une table de contingence.

Supposons que les données sont telles que chacune des  $n$  observations peut être classée dans une des  $K = I \cdot J$ , (nombre fini) de catégories possibles suivant deux attributs  $A_i, B_j$  ( $i = 1, 2, \dots, I; j = 1, 2, \dots, J$ ). Dans ce cas les données peuvent être présentées dans un tableau de contingence à  $I$  lignes et  $J$  colonnes. On notera  $p_{ij}$  la probabilité pour une observation d'être classée à la  $i$ -ème ligne et  $j$ -ème colonne du tableau, ce qui signifie que cette observation possède les attributs  $A_i$  et  $B_j$ . Notons  $v_{ij}$  le nombre des observations placées à la  $i$ -ème ligne et  $j$ -ème colonne. On a alors

$$\sum_{i=1}^I \sum_{j=1}^J v_{ij} = n \text{ and } \sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1. \quad (1)$$

Soit  $p_{i\cdot}$  la probabilité marginale que l'observation soit à la  $i$ -ème ligne et soit  $p_{\cdot j}$  la probabilité marginale que l'observation soit à la  $j$ -ème colonne du tableau. Il est clair que

$$p_{i\cdot} = \sum_{j=1}^J p_{ij} \text{ and } p_{\cdot j} = \sum_{i=1}^I p_{ij}. \quad (2)$$

Nous avons bien sûr :

$$\sum_{i=1}^I p_{i\cdot} = \sum_{j=1}^J p_{\cdot j} = 1. \quad (3)$$

On peut présenter le modèle avec les deux tableaux :

	$B_1$	$\cdots$	$B_j$	$\cdots$	$B_J$	
$A_1$	$p_{11}$	$\cdots$	$p_{1j}$	$\cdots$	$p_{1J}$	$p_{1\cdot}$
$A_i$	$p_{i1}$	$\cdots$	$p_{ij}$	$\cdots$	$p_{iJ}$	$p_{i\cdot}$
$A_I$	$p_{I1}$	$\cdots$	$p_{Ij}$	$\cdots$	$p_{IJ}$	$p_{I\cdot}$
	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	$p_{\cdot J}$	1

Tab. 1

	$B_1$	$\cdots$	$B_j$	$\cdots$	$B_J$	
$A_1$	$v_{11}$	$\cdots$	$v_{1j}$	$\cdots$	$v_{1J}$	$v_{1\cdot}$
$A_i$	$v_{i1}$	$\cdots$	$v_{ij}$	$\cdots$	$v_{iJ}$	$v_{i\cdot}$
$A_I$	$v_{I1}$	$\cdots$	$v_{Ij}$	$\cdots$	$v_{IJ}$	$v_{I\cdot}$
	$v_{\cdot 1}$	$\cdots$	$v_{\cdot j}$	$\cdots$	$v_{\cdot J}$	$n$

Tab. 2

Si on connaît les véritables probabilités  $p_{ij}$ , alors la statistique

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(v_{ij} - n_i p_{ij})^2}{n_i p_{ij}} \quad (4)$$

a pour distribution limite lorsque  $\min n_i \rightarrow \infty$  une distribution du  $\chi^2$  avec  $f$  degrés de liberté,

$$f = K - 1 = I \cdot J - 1, \quad (5)$$

où

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X^2 \leq x \mid H_0\} = \mathbf{P}\{\chi_{IJ-1}^2 \leq x\}. \quad (6)$$

Si les  $p_{ij}$  sont inconnus, nous devons les estimer. Supposons que nous nous intéressons à l'hypothèse  $H_0$  d'après laquelle les classements dans les lignes et les colonnes sont indépendants, i.e.,

$$\mathbf{P}\{A_i B_j\} = p_{ij} = \mathbf{P}\{A_i\} \mathbf{P}\{B_j\} = p_{i\cdot} p_{\cdot j}. \quad (7)$$

Dans notre modèle, la fonction de vraisemblance est :

$$L(\mathbf{p}) = \frac{n!}{v_{11}! \cdots v_{IJ}!} p_{11}^{v_{11}} \cdots p_{ij}^{v_{ij}} = \frac{n!}{v_{11}! \cdots v_{IJ}!} \prod_{i=1}^I \prod_{j=1}^J p_{ij}^{v_{ij}}. \quad (8)$$

Sous l'hypothèse  $H_0$  nous avons

$$\begin{aligned} L(\mathbf{p}) &= \frac{n!}{v_{11}! \cdots v_{IJ}!} \prod_{i=1}^I \prod_{j=1}^J p_{ij}^{v_{ij}} \\ &= \frac{n!}{v_{11}! \cdots v_{IJ}!} \left( \prod_{i=1}^I \prod_{j=1}^J p_{i\cdot}^{v_{ij}} \right) \left( \prod_{i=1}^I \prod_{j=1}^J p_{\cdot j}^{v_{ij}} \right) \end{aligned}$$

$$= \frac{n!}{v_{11}! \cdots v_{IJ}!} \left( \prod_{i=1}^I p_i^{v_{i\cdot}} \right) \left( \prod_{j=1}^J p_{\cdot j}^{v_{\cdot j}} \right), \quad (9)$$

où

$$v_{i\cdot} = \sum_{j=1}^J v_{ij} \text{ and } v_{\cdot j} = \sum_{i=1}^I v_{ij}, \quad (10)$$

et en prenant les logarithmes, on obtient

$$\ln L(\mathbf{p}) = \text{const} + \sum_{i=1}^I v_{i\cdot} \ln p_i + \sum_{j=1}^J v_{\cdot j} \ln p_{\cdot j}.$$

Pour trouver le vecteur informant  $\Lambda(\mathbf{p})$  nous dérivons  $\ln L(\mathbf{p})$  par rapport à  $p_i$  et  $p_{\cdot j}$  :

$$\Lambda(\mathbf{p}) = \frac{\partial}{\partial \mathbf{p}} \ln L(\mathbf{p}) = \left( \frac{\partial \ln L(\mathbf{p})}{\partial p_1}, \dots, \frac{\partial \ln L(\mathbf{p})}{\partial p_I}, \frac{\partial \ln L(\mathbf{p})}{\partial p_{\cdot 1}}, \dots, \frac{\partial \ln L(\mathbf{p})}{\partial p_{\cdot J}} \right)^T, \quad (12)$$

où

$$\frac{\partial \ln L(\mathbf{p})}{\partial p_i} = \frac{v_{i\cdot}}{p_i} - \frac{v_{I\cdot}}{p_I}, i = 1, 2, \dots, I-1; \quad (12)$$

et

$$\frac{\partial \ln L(\mathbf{p})}{\partial p_{\cdot j}} = \frac{v_{\cdot j}}{p_{\cdot j}} - \frac{v_{\cdot J}}{p_{\cdot J}}, j = 1, 2, \dots, J-1. \quad (13)$$

En utilisant

$$p_I = 1 - \sum_{i=1}^{I-1} p_i \text{ and } p_{\cdot J} = 1 - \sum_{j=1}^{J-1} p_{\cdot j}, \quad (14)$$

de (12)-(13) on tire les estimateurs de maximum de vraisemblance de  $p_i$  and  $p_{\cdot j}$  :

$$\hat{p}_{i\cdot} = \frac{v_{i\cdot}}{n} \text{ and } \hat{p}_{\cdot j} = \frac{v_{\cdot j}}{n}, \quad (15)$$

d'où les estimateurs de maximum de vraisemblance des probabilités  $p_{ij}$  sont

$$\hat{p}_{ij} = \hat{p}_{i\cdot} \cdot \hat{p}_{\cdot j} = \frac{v_{i\cdot} \cdot v_{\cdot j}}{n}. \quad (16)$$

Dans ce cas, d'après le théorème de Fisher, sous l'hypothèse  $H_0$  la statistique de Pearson

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(v_{ij} - n \hat{p}_{i\cdot} \hat{p}_{\cdot j})^2}{n \hat{p}_{i\cdot} \hat{p}_{\cdot j}} = n \left( \sum_{i=1}^I \sum_{j=1}^J \frac{v_{ij}^2}{v_{i\cdot} v_{\cdot j}} - 1 \right) \quad (17)$$

a pour distribution limite lorsque  $n \rightarrow \infty$ , la distribution du  $\chi^2$  à  $f$  degrés de liberté,

$$f = IJ - (I-1) - (J-1) - 1 = (I-1)(J-1)$$

et donc

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X^2 \leq x \mid H_0\} = \mathbf{P}\{\chi_{(I-1)(J-1)}^2 \leq x\}. \quad (18)$$

On peut utiliser ce résultat pour construire un test du  $\chi^2$  pour l'hypothèse  $H_0$  au seuil de signification  $\alpha$ . D'après ce test on doit rejeter  $H_0$  si

$$X^2 > \bar{\chi}_f^2(\alpha),$$

où  $\bar{\chi}_f^2(\alpha)$  est le  $\alpha$ -quantile supérieur ( $0 < \alpha < 0.5$ ) de la distribution du  $\chi^2$  à  $f = (I - 1)(J - 1)$  degrés de liberté.

Considérons le cas  $I = J = 2$ . Alors au lieu du tableau 2 nous avons le tableau connu comme le tableau  $2 \times 2$ .

$v_{11}$	$v_{12}$	$v_{1\cdot}$
$v_{21}$	$v_{22}$	$v_{2\cdot}$
$v_{\cdot 1}$	$v_{\cdot 2}$	$n$

Tab. 3

De façon évidente, sous l'hypothèse  $H_0$ , au lieu du tableau 1, nous aurons le tableau 4,

	$B_1$	$B_2$	
$A_1$	$pP$	$qP$	$P$
$A_2$	$pQ$	$qQ$	$Q$
	$p$	$q$	

Tab. 4

où

$$P = \mathbf{P}(A_1), Q = \mathbf{P}(A_2) = 1 - P, p = \mathbf{P}(B_1), q = \mathbf{P}(B_2) = 1 - p.$$

On peut vérifier qu'après quelques manipulations, la statistique de Pearson (17) peut s'écrire

$$X^2 = \frac{n(v_{11}v_{22} - v_{21}v_{12})^2}{v_{1\cdot}v_{2\cdot}v_{\cdot 1}v_{\cdot 2}}, \quad (19)$$

et d'après (18) il s'ensuit

$$\lim_{n \rightarrow \infty} \mathbf{P}\{X^2 \leq x \mid H_0\} = \mathbf{P}\{\chi_1^2 \leq x\}. \quad (20)$$

**Exemple 1.** Considérons un groupe de 300 étudiants qui ont passé un examen partiel en mathématiques. Parmi eux, 97 ont obtenu une très bonne note : A et les 203 autres une note inférieure : B. A la fin de l'année, ces étudiants passent l'examen final de mathématiques et cette fois-ci 48 d'entre eux obtiennent une très bonne note A et parmi eux 18 seulement ont obtenu une très bonne note au partiel. Cela signifie que 18 étudiants ont obtenu une très bonne note à la fois à l'examen partiel et à l'examen terminal.

En utilisant ces données nous pouvons construire un test du  $\chi^2$  au niveau de signification  $\alpha = 0.1$ , pour tester l'hypothèse  $H_0$  de l'indépendance d'obtention d'une très bonne note à chacun des 2 examens.

Tout d'abord, présentons les données dans le tableau  $2 \times 2$  suivant :

		<i>exam partiel</i>		<i>Total</i>
		<i>A</i>	<i>B</i>	
<i>exam final</i>	<i>A</i>	18	30	48
	<i>B</i>	79	173	252
<i>Total</i>		97	203	300

Tab. 5

		<i>exam partiel</i>		<i>Total</i>
		<i>A</i>	<i>B</i>	
<i>exam final</i>	<i>A</i>	$pP$	$qP$	$P$
	<i>B</i>	$pQ$	$qQ$	$Q$
<i>Total</i>		$p$	$q$	1

Tab. 6

Les estimateurs de maximum de vraisemblance de  $p$  et  $P$  sont

$$\hat{p} = \frac{v_{.1}}{n} = \frac{97}{300} \text{ et } \hat{P} = \frac{v_{1.}}{n} = \frac{48}{300}.$$

La valeur de la statistique de Pearson  $X^2$  donnée par (17), peut être évaluée en utilisant la formule (19) selon laquelle

$$\begin{aligned} X^2 &= \frac{\left(18 - \frac{48 \cdot 97}{300}\right)^2}{\frac{48 \cdot 97}{300}} + \frac{\left(30 - \frac{48 \cdot 203}{300}\right)^2}{\frac{48 \cdot 203}{300}} + \frac{\left(79 - \frac{252 \cdot 97}{300}\right)^2}{\frac{252 \cdot 97}{300}} + \frac{\left(173 - \frac{252 \cdot 203}{300}\right)^2}{\frac{252 \cdot 203}{300}} = \\ &= \frac{300(18 \cdot 173 - 30 \cdot 79)^2}{97 \cdot 203 \cdot 48 \cdot 252} = \frac{100(248)^2}{97 \cdot 203 \cdot 64 \cdot 7} \\ &= \frac{200}{203} \cdot \frac{93}{97} \cdot \frac{31}{42} < 1. \end{aligned}$$

Sous l'hypothèse  $H_0$  la statistique  $X^2$  de Pearson est distribuée approximativement comme  $\chi_1^2$ , et donc on accepte  $H_0$ , puisque la valeur observée de  $X^2$  est inférieure à  $\chi_1^2(0.1) = 2.706$ .

## 4.10 Test du Chauvenet pour la détection des observations aberrantes.

Le test de Chauvenet est une règle ancienne destinée à détecter au moins une valeur aberrante dans une série de mesures et à l'éliminer. Cette règle est basée sur une propriété simple de l'espérance mathématique. Ici nous allons suivre l'idée de L. Bolshev (1961) sur la présentation du test de Chauvenet (voir aussi Voinov et Nikulin (1996)).

Considérons  $n$  variables aléatoires indépendantes  $Y_1, \dots, Y_n$ ,  $n \geq 3$ , de même loi et soit  $y$  un nombre réel donné.

Soit

$$N = \sum_{j=1}^n \mathbf{1}_{[y, +\infty[}(Y_j)$$

La statistique  $N$  suit une loi binomiale de moyenne :

$$\mathbf{E}(N) = n\mathbf{P}(Y_1 \geq y) = np,$$

où  $p = \mathbf{P}\{Y_1 \geq y\}$ . Pour avoir l'égalité  $\mathbf{E}(N) = \alpha$ ,  $\alpha > 0$ , il faut choisir  $y = y(\alpha)$  comme la solution de l'équation

$$\mathbf{P}\{Y_1 > y\} = \frac{\alpha}{n}. \quad (1)$$

Dans ce cas il est facile de vérifier que

$$\begin{aligned} \beta &= \mathbf{P}\left\{\max_{1 \leq i \leq n} Y_i > y(\alpha)\right\} = 1 - \{1 - \mathbf{P}\{Y_1 \geq y(\alpha)\}\}^n = \\ &= 1 - \left(1 - \frac{\alpha}{n}\right)^n = 1 - e^{-\alpha} + o(1) \quad (n \rightarrow \infty), \end{aligned}$$

et donc si  $\alpha$  est suffisamment petit,

$$\mathbf{P}\left\{\max_{1 \leq i \leq n} Y_i > y(\alpha)\right\} \simeq \alpha.$$

Notons que Chauvenet lui-même a suggéré de choisir  $\alpha = 1/2n$ . Considérons l'hypothèse  $H_0$  selon laquelle

$$\mathbf{P}\{Y_i \leq y\} = F(y), \quad \forall i \in [1, n],$$

où  $F$  est une fonction de répartition donnée, et  $H_1$  est alternative d'après laquelle

$$\mathbf{P}\{Y_i \leq y\} = (1 - \varepsilon)F(y) + \varepsilon G(y), \quad i = 1, \dots, n, \quad (0 < \varepsilon < \frac{1}{2}),$$

où  $G$  est une fonction de répartition telle que  $G(y) < F(y)$  pour tout  $y$ .

Dans ce cas la région critique déterminée pour la règle de Chauvenet est :

$$\{N \geq 1\} \Leftrightarrow \left\{\max_{1 \leq i \leq n} Y_i > y(\alpha)\right\}$$

Le niveau de signification du test (pour  $n$  grand et  $\alpha$  petit) est approximativement  $\alpha$ . On peut même, en utilisant l'inégalité de Bonferroni, estimer l'erreur relative entre le seuil du test et  $\alpha$ , et ce pour tout  $\alpha$  et pour tout  $n$ .



# Chapitre 5

## REGRESSION

### 5.1 Régression linéaire

#### 5.1.1 Modèle de la régression linéaire

On considère le problème de la prédiction d'une ou plusieurs caractéristiques d'une variable aléatoire  $Y$  à l'aide de variables explicatives (covariables)  $x_1, \dots, x_m$ . Par exemple, on considère la prédiction de l'espérance du prix  $Y$  d'une voiture d'une certaine marque lorsqu'on connaît l'âge  $x_1$ , la puissance  $x_2$  et le kilométrage  $x_3$  de cette voiture.

Même si les valeurs de  $x_1, \dots, x_m$  sont fixées, la variable aléatoire  $Y$  peut prendre des valeurs différentes, parce qu'il y a souvent d'autres facteurs qui interviennent. Par exemple, les prix de voitures qui ont le même âge, la même puissance et le même kilométrage ne sont pas forcément les mêmes, à cause de facteurs tels que le nombre des pannes, la présence ou l'absence de garage spécifique, le régime de travail, les conditions climatiques, le lieu de vente, etc.

Notons

$$x = (x_0, x_1, \dots, x_m)^T, \quad x_0 = 1, \quad M(x) = \mathbf{E}(Y|x).$$

La fonction  $M(x)$  est appelée la *fonction de régression*. On suppose que  $M(x)$  est une combinaison linéaire des covariables  $x_i$  :

$$M(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m = \beta^T x, \quad (1)$$

où  $\beta = (\beta_0, \dots, \beta_m)^T$  est un paramètre inconnu.

Pour faire l'estimation on effectue  $n$  expériences. La  $i$ -ème expérience a lieu sous la covariable  $x^{(i)} = (x_{i0}, \dots, x_{im})$ ,  $x_{i0} = 1$ .

On observe des valeurs de la variable dépendante ( ou expliquée)  $Y_i$ . Donc on a un échantillon

$$(x^{(1)}, Y_1), \dots, (x^{(n)}, Y_n).$$

*Le modèle de la régression linéaire*

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} + e_i,$$

où  $e_1, \dots, e_n$  sont des variables aléatoires i.i.d.,

$$\mathbf{E}(e_i) = 0, \quad \mathbf{Var}(e_i) = \sigma^2, \quad i = 1, \dots, n.$$

Donc on a

$$Y_i = M(x^{(i)}) + e_i, \quad i = 1, \dots, n,$$

où  $M(x)$  est donné par la formule (1). Si  $m = 1$ , on a le modèle de *régression linéaire simple*, et si  $m > 1$ , on a le modèle de *régression linéaire multiple*.

Notons

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix}_{n \times (m+1)}, \quad e = (e_1, \dots, e_n)^T, \quad \mathbf{Y} = (Y_1, \dots, Y_n)^T.$$

Alors le modèle peut être écrit

$$\mathbf{Y} = X\beta + e, \quad \text{où } \mathbf{E}(e) = \mathbf{0}_n, \quad \mathbf{Var}(e) = \sigma^2 \mathbf{I}_n. \quad (2)$$

Dans ce modèle le vecteur  $e$  est interprété comme le *vecteur des erreurs*.

## 5.1.2 Codage des covariables

Si la  $j$ -ème variable explicative  $x_j$  dans (1) est dicrète et mesurée sur une échelle nominale, par exemple la couleur, la race, etc., et prend  $k_j$  valeurs différentes, on peut utiliser, au lieu de  $x_j$ , le vecteur  $z_j = (z_{j,1}, \dots, z_{j,k_j-1})$  des codes, qui prend  $k_j$  valeurs différentes :

$$z_j^{(0)} = (0, \dots, 0), \quad z_j^{(1)} = (1, 0, \dots, 0), \quad z_j^{(2)} = (0, 1, 0, \dots, 0), \dots, z_j^{(k_j-1)} = (0, \dots, 0, 1)$$

et le modèle (1) est modifié :

$$M(x) = \beta_0 + \beta_1 x_1 + \dots + \sum_{i=1}^{k_j-1} \beta_{ji} z_{ji} + \dots + \beta_m x_m. \quad (3)$$

On note que

$$\sum_{i=1}^{k_j-1} \beta_{ji} z_{ji} = \beta_j^T z_j$$

où  $\beta_j^T = (\beta_{j1}, \dots, \beta_{j,k_j-1})$ . Si, par exemple,  $x_j$  est la couleur qui prend 3 valeurs (noir, bleu, blanc), on considère le vecteur  $z_j = (z_{j1}, z_{j2})$  qui prend les valeurs

$$z_j^{(0)} = (0, 0) \text{ - (noir), } z_j^{(1)} = (1, 0) \text{ - (bleu), } z_j^{(2)} = (0, 1) \text{ - (blanc).}$$

Si  $x_j$  est le sexe (masculin, féminin), on considère la variable  $z_j$  qui prend les valeurs

$$z_j^{(0)} = 0 \text{ (masculin) et } z_j^{(1)} = 1 \text{ (féminin).}$$

Parfois le codage est différent :  $(-1, \dots, -1), (1, 0, \dots, 0), \dots, (0, 0, \dots, 1)$ , etc.

### 5.1.3 Interprétation des coefficients $\beta$ .

Notons que lorsqu'on prend deux valeurs  $x_j^{(1)}$  et  $x_j^{(2)}$  de  $x_j$  dans (1), alors

$$\beta_j(x_j^{(2)} - x_j^{(1)}) = M(x_1, \dots, x_j^{(2)}, \dots, x_m) - M(x_1, \dots, x_j^{(1)}, \dots, x_m).$$

Donc

$$\beta_j(x_j^{(2)} - x_j^{(1)})$$

(soit  $\beta_j$ , si  $x_j^{(2)} - x_j^{(1)} = 1$ ) représente *le changement de la valeur moyenne de la variable expliquée  $Y$  quand  $x_j$  passe de  $x_j^{(1)}$  à  $x_j^{(2)}$  tandis que toutes les autres covariables restent les mêmes.*

Il faut souligner que dans le modèle (1) le changement de la moyenne de  $Y$  est le même pour n'importe quelles valeurs fixées des autres covariables  $x_l$  ( $l \neq j$ ), c'est à dire qu' il n'y a pas d'interaction entre les covariables.

Si  $x_j$  est discrète et mesurée sur une échelle nominale, alors

$$\beta_{ji} = M(x_1, \dots, z_j^{(i)}, \dots, x_m) - M(x_1, \dots, z_j^{(0)}, \dots, x_m).$$

Donc  $\beta_{ji}$  représente le changement de la moyenne de la variable dépendante  $Y$  quand  $z_j$  passe de  $z_j^{(0)}$  à  $z_j^{(i)}$  tandis que toutes les autres covariables gardent les mêmes valeurs. Par exemple, si  $x_j$  est la couleur (noire, blanche ou bleue),  $\beta_{j2}$  représente le changement de la moyenne de  $Y$  qui correspond au changement de  $x_j$  de la couleur noire ( $z_j = z_j^{(0)}$ ) à la couleur blanche ( $z_j = z_j^{(2)}$ ).

### 5.1.4 Modèle avec interactions

Si l'effet du changement de la valeur de la covariable  $x_j$  est différent pour des valeurs différentes des autres covariables, c'est qu'on a une interaction entre  $x_j$  et ces covariables. Alors le modèle (1) peut être modifié pour mettre en lumière l'effet de cette interaction. Par exemple, dans le cas de deux covariables, on a le modèle

$$M(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \quad (4)$$

et dans le cas de trois covariables :

$$M(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3. \quad (5)$$

S'il y a une interaction dans le cas  $m = 2$  par exemple, alors

$$M(x_1^{(2)}, x_2) - M(x_1^{(1)}, x_2) = (\beta_1 + \beta_3 x_2)(x_1^{(2)} - x_1^{(1)}),$$

donc la moyenne de  $Y$  dépend non seulement de la différence  $x_1^{(2)} - x_1^{(1)}$  mais aussi de la valeur de la deuxième covariable  $x_2$ .

Si, par exemple,  $Y$  est le prix (en Frs.),  $x_1$  est l'âge (en années),  $x_2$  est la puissance (en  $cm^3$ ), d'une voiture d'une certaine marque et s'il y a une interaction entre l'âge et la puissance, il est évident que la valeur de la voiture diminue annuellement mais cette baisse du prix est différente pour des voitures de différentes puissances. Pour la voiture ayant la puissance  $x_2 cm^3$  la baisse du prix annuelle est de  $\beta_1 + \beta_3 x_2$  (Euros.). Voir aussi la section *Décomposition orthogonale de Fisher*.

## 5.1.5 Estimateurs des moindres carrés

On cherche l'estimateur  $\hat{\beta}$  qui minimise la somme des carrés

$$SS = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_m x_{im})^2 = (Y - X\beta)^T (Y - X\beta).$$

En dérivant  $SS$  par rapport à  $\beta_j$  on a

$$\frac{\partial SS}{\partial \beta_j} = -2 \sum_{i=1}^n x_{ij} (Y_i - \beta_0 - \dots - \beta_m x_{im}), \quad (j = 0, \dots, m),$$

d'où on obtient le système de  $(m+1)$  équations linéaires ( $j=0, \dots, m$ ) :

$$\beta_0 \sum_{i=1}^n x_{ij} x_{i0} + \beta_1 \sum_{i=1}^n x_{ij} x_{i1} + \dots + \beta_m \sum_{i=1}^n x_{ij} x_{im} = \sum_{i=1}^n \alpha_{ij} Y_i, \quad (6)$$

ou

$$X^T X \beta = X^T Y.$$

Si la matrice  $A_{(m+1) \times (m+1)} = X^T X$  n'est pas dégénérée, alors on a

*L'estimateur des moindres carrés de  $\beta$  :*

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (7)$$

Si  $x = (1, x_1, \dots, x_m)^T$  est un vecteur de covariables donné,

$$M(x) = \mathbf{E}(Y|x) = \beta^T x,$$

alors on obtient

*L'estimateur de l'espérance  $M(x) = \mathbf{E}(Y|x)$  est :*

$$\hat{M}(x) = \hat{\beta}^T x = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m.$$

Notons

$$\hat{Y}_i = \hat{M}(x^{(i)}) = \hat{\beta}^T x^{(i)}, \quad \hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)^T, \quad \hat{e} = (\hat{e}_1, \dots, \hat{e}_n)^T, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Les variables aléatoires  $\hat{Y}_i$  et  $Y_i$  sont appelées respectivement les valeurs *prédites* et *observées* des  $Y_i$ , et les  $\hat{e}_i = Y_i - \hat{Y}_i$  sont les *résidus estimés* @ù *des erreurs apparentes* . On a

$$\hat{Y} = X\hat{\beta}, \quad \hat{e} = Y - \hat{Y} = Y - X\hat{\beta} = e + X(\beta - \hat{\beta}). \quad (8)$$

## 5.1.6 Propriétés des estimateurs

Notons

$$B = B_{(m+1) \times n} = (X^T X)^{-1} X^T, \quad H = H_{n \times n} = \mathbf{I}_n - X(X^T X)^{-1} X^T.$$

Alors

$$\hat{\beta} = BY, \quad \hat{e} = HY. \quad (9)$$

Il faut remarquer que

$$HH = H, \quad H^T = H, \quad BB^T = B, \quad BH = \mathbf{0}_{(m+1) \times n}. \quad (10)$$

**Lemme 1.** Si  $\det(X^T X) \neq 0$ , alors

$$a) \quad X^T H = \mathbf{0}_{m \times n}, \quad X^T e = \mathbf{0}_{m+1}, \quad \hat{Y}^T e = 0, \quad (11)$$

$$b) \quad Y^T Y = \hat{Y}^T \hat{Y} + \hat{e}^T \hat{e}, \Leftrightarrow \|Y\|^2 = \|\hat{Y}\|^2 + \|\hat{e}\|^2, \quad (12)$$

$$c) \quad \sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2 + \sum_{j=1}^n (Y_j - \hat{Y}_j)^2, \quad (13)$$

$$d) \quad \sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i, \quad (14)$$

$$e) \quad e^T e = \hat{e}^T \hat{e} + (\hat{Y} - X\beta)^T (\hat{Y} - X\beta). \quad (15)$$

**Démonstration.**

a) On a

$$X^T H = X^T - X^T X (X^T X)^{-1} X^T = \mathbf{0}_{n \times n},$$

donc

$$X^T \hat{e} = X^T H Y = \mathbf{0}_{m+1}, \quad \hat{Y}^T \hat{e} = \hat{\beta}^T X^T \hat{e} = 0.$$

b) D'après (11)

$$Y^T Y = (\hat{Y} + \hat{e})^T (\hat{Y} + \hat{e}) = \hat{Y}^T \hat{Y} + \hat{Y}^T \hat{e} + \hat{e}^T Y + \hat{e}^T \hat{e} = \hat{Y}^T \hat{Y} + \hat{e}^T \hat{e}.$$

c) L'égalité (12) peut être écrite sous la forme

$$\sum Y_j^2 = \sum \hat{Y}_j^2 + \sum (Y_j - \hat{Y}_j)^2.$$

Alors

$$\sum_{j=1}^n Y_j^2 - n\bar{Y}^2 = \sum_{j=1}^n \hat{Y}_j^2 - n\bar{Y}^2 + \sum_{j=1}^n (Y_j - \bar{Y}_j)^2$$

et donc

$$\sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2 + \sum_{j=1}^n (Y_j - \hat{Y}_j)^2.$$

d) La première ligne de  $X^T$  est  $\mathbf{1}_n = (1, \dots, 1)^T$ , donc l'égalité  $X^T \hat{e} = \mathbf{0}_{m+1}$ , démontrée dans a), implique

$$\mathbf{1}_n^T \hat{e} = \sum_{i=1}^n \hat{e}_i = 0 \quad \text{et donc} \quad \sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i.$$

e) D'après (2) et (11) on a

$$\begin{aligned} e^T e &= (Y - X\beta)^T (Y - X\beta) = \\ &= (Y - \hat{Y} + \hat{Y} - X\beta)^T (Y - \hat{Y} + \hat{Y} - X\beta) = \\ &= (\hat{e} + \hat{Y} - X\beta)^T (\hat{e} + \hat{Y} - X\beta) = \\ &= \hat{e}^T \hat{e} + 2\hat{e}^T (\hat{Y} - X\beta) + (\hat{Y} - X\beta)^T (\hat{Y} - X\beta) = \\ &= \hat{e}^T \hat{e} + (\hat{Y} - X\beta)^T (\hat{Y} - X\beta). \end{aligned}$$

Le lemme est démontré.

**Théorème 1. Gauss-Markov.** Si  $\det(X^T X) \neq 0$ , alors

- $\mathbf{E}(\hat{\beta}) = \beta, \quad \mathbf{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1},$
- $\mathbf{E}(\hat{e}) = \mathbf{0}, \quad \mathbf{Var}(\hat{e}) = \sigma^2 H,$
- $\mathbf{Cov}(\hat{\beta}, \hat{e}) = \mathbf{0},$
- $\mathbf{E}(\hat{e}^T \hat{e}) = (n - m - 1)\sigma^2.$

**Démonstration.**

a) On a

$$\begin{aligned} \mathbf{E}(\hat{\beta}) &= (X^T X)^{-1} X^T \mathbf{E}(Y) = (X^T X)^{-1} X^T X \beta = \beta, \\ \mathbf{Var}(\hat{\beta}) &= (X^T X)^{-1} X^T \sigma^2 \mathbf{I}_n X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}. \end{aligned}$$

b)

$$\begin{aligned} \mathbf{E}(\hat{e}) &= \mathbf{E}(Y - X\hat{\beta}) = X\beta - X\beta = \mathbf{0}_n. \\ \mathbf{Var}(\hat{e}) &= \mathbf{Var}(HY) = H\sigma^2 \mathbf{I}_n H = \sigma^2 H. \end{aligned}$$

c)

$$\mathbf{Cov}(\hat{\beta}, \hat{e}) = \mathbf{Cov}(BY, HY) = B\sigma^2 \mathbf{I}_n H = \sigma^2 BH = \mathbf{0}_{(m+1) \times n}.$$

d) Notons

$$A = X^T X = (a_{ij}), \quad A^{-1} = (a^{ij}), \quad (i, j = 0, \dots, m).$$

Alors

$$\begin{aligned} \mathbf{E}((\hat{Y} - X\beta)^T (\hat{Y} - X\beta)) &= \mathbf{E}((\hat{\beta} - \beta)^T A (\hat{\beta} - \beta)) = \\ &= \sum_{i=0}^m \sum_{j=0}^m a_{ij} \mathbf{E}((\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)) = \sigma^2 \sum_{i=0}^m \sum_{j=0}^m a_{ij} a^{ij} = \\ &= \sigma^2 \text{Tr}(AA^{-1}) = \sigma^2 \text{Tr}(\mathbf{I}_{m+1}) = \sigma^2(m+1). \end{aligned}$$

On a

$$\mathbf{E}(e^T e) = \sum_{i=1}^n \mathbf{E}e_i^2 = \sum_{i=1}^n \mathbf{Var} e_i = n\sigma^2.$$

L'égalité (15) implique que

$$\mathbf{E}(\hat{e}^T \hat{e}) = (n - m - 1)\sigma^2.$$

Le théorème 1 est démontré.

**Corollaire.**

$$\hat{\beta} \quad \text{et} \quad \hat{\sigma}^2 = \frac{SS_R}{n - m - 1}$$

des estimateurs sans biais de  $\beta$  et de  $\sigma^2$  respectivement, et

$$\mathbf{Cov}(\hat{\beta}, \hat{\sigma}^2) = 0;$$

On a déjà vu que les paramètres qu'il est le plus important d'estimer et qui sont en même temps ceux dont l'interprétation est la plus évidente sont :

a) la moyenne  $M(x)$  de la variable expliquée  $Y$  sous n'importe quelle valeur de la covariable  $x$  ;

b) chacun des paramètres  $\beta_j$ , qui caractérise le changement de la moyenne de  $Y$  correspondant au changement de la covariable  $x_j$  (modèle sans interaction) ;

c) les combinaisons linéaires des paramètres  $\beta_j$ , qui caractérisent le changement de la moyenne de  $Y$  correspondant au changement d'une covariable sous des valeurs spécifiées des autres covariables (le modèle avec interactions). Par exemple, dans le modèle (4) la combinaison linéaire  $\beta_1 + \beta_3 x_2$  caractérise le changement de la moyenne de  $Y$  correspondant au changement de la covariable  $x_1$  sous des valeurs spécifiées de  $x_2$ .

Donc dans tous les cas l'estimation des combinaisons linéaires du type  $\mathbf{I}^T \beta$ , où  $\mathbf{I} = (l_0, \dots, l_m)^T$ , est importante.

Un estimateur de  $\mathbf{I}^T \beta$  est appelé *linéaire*, s'il a la forme

$$c^T Y, \quad c = (c_1, \dots, c_n)^T.$$

L'estimateur  $c^T Y$  de  $\mathbf{I}^T \beta$  est appelé *sans biais* si

$$\mathbf{E}(c^T Y) = \mathbf{I}^T \beta \quad \text{pour tout} \quad \beta \in R^{m+1},$$

i.e. pour le modèle de type (2) avec n'importe quel  $\beta \in R^{m+1}$ , l'espérance de  $c^T Y$  est égale à la vraie valeur de  $\mathbf{I}^T \beta$ .

Notons  $\mathcal{G}_l$  la classe des estimateurs linéaires sans biais de  $\mathbf{I}^T \beta$ .

**Théorème 2. (Gauss-Markov).** Si  $\det(X^T X) \neq 0$ , alors  $\mathbf{I}^T \hat{\beta}$  est l'unique estimateur de variance minimale dans la classe  $\mathcal{G}_l$ .

**Démonstration.** Si  $c^T Y \in \mathcal{G}_l$ , alors

$$\mathbf{I}^T \beta = \mathbf{E}(c^T Y) = \mathbf{E}(c^T Y - \mathbf{I}^T \hat{\beta} + \mathbf{I}^T \hat{\beta}) = (c^T X - \mathbf{I}^T) \beta + \mathbf{I}^T \beta,$$

donc

$$(c^T X - \mathbf{1}^T) \boldsymbol{\beta} = 0 \quad \text{pour tout } \boldsymbol{\beta} \in \mathbb{R}^{m+1}$$

et

$$c^T X - \mathbf{1}^T = \mathbf{0}_{m+1}^T. \quad (16).$$

On a

$$\begin{aligned} \mathbf{Var}(c^T Y) &= \mathbf{Var}(c^T Y - \mathbf{1}^T \hat{\boldsymbol{\beta}} + \mathbf{1}^T \hat{\boldsymbol{\beta}}) = \\ &= \mathbf{Var}(c^T Y - \mathbf{1}^T \hat{\boldsymbol{\beta}}) + \mathbf{Var}(\mathbf{1}^T \hat{\boldsymbol{\beta}}) + 2\mathbf{Cov}(c^T Y - \mathbf{1}^T \hat{\boldsymbol{\beta}}, \mathbf{1}^T \hat{\boldsymbol{\beta}}). \end{aligned}$$

L'égalité (16) et le Lemme 1 impliquent que

$$\begin{aligned} \mathbf{Cov}(c^T Y - \mathbf{1}^T \hat{\boldsymbol{\beta}}, \mathbf{1}^T \hat{\boldsymbol{\beta}}) &= \mathbf{Cov}((c^T - \mathbf{1}^T B)Y, \mathbf{1}^T B Y) = \\ &= (c^T - \mathbf{1}^T B) \sigma^2 \mathbf{I}_n B^T \mathbf{1} = \\ &= \sigma^2 (c^T X (X^T X)^{-1} - \mathbf{1}^T (X^T X)^{-1} X^T X (X^T X)^{-1}) \mathbf{1} = \\ &= \sigma^2 (c^T X - \mathbf{1}^T) (X^T X)^{-1} \mathbf{1} = 0, \end{aligned}$$

donc

$$\begin{aligned} \mathbf{Var}(c^T Y) &= \mathbf{Var}(\mathbf{1}^T \hat{\boldsymbol{\beta}}) + \mathbf{Var}(c^T - \mathbf{1}^T B)Y = \\ &= \mathbf{Var}(\mathbf{1}^T \hat{\boldsymbol{\beta}}) + \sigma^2 (c^T - \mathbf{1}^T B)(c^T - \mathbf{1}^T B)^T. \end{aligned}$$

On a  $\mathbf{Var}(c^T Y) \geq \mathbf{Var}(\mathbf{1}^T \hat{\boldsymbol{\beta}})$  et l'égalité est vérifiée si et seulement si  $c^T = \mathbf{1}^T B$ .

Le théorème est démontré.

**Corollaire.** Les estimateurs  $\hat{M}(x) = \hat{\boldsymbol{\beta}}^T x$  et  $\hat{\boldsymbol{\beta}}_j$  de la moyenne  $M(x)$  et du paramètre  $\boldsymbol{\beta}_j$ , respectivement, sont les estimateurs de variance minimale dans la classe des estimateurs linéaires sans biais de  $M(x)$  et  $\boldsymbol{\beta}_j$ .

Il s'ensuit par exemple que l'estimateur  $\hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\beta}}_3 x_2$  de  $\boldsymbol{\beta}_1 + \boldsymbol{\beta}_3 x_2$  est le meilleur estimateur dans la classe des estimateurs linéaires sans biais de  $\boldsymbol{\beta}_1 + \boldsymbol{\beta}_3 x_2$  (modèle (4)).

### 5.1.7 Décomposition des sommes de carrés

Le lemme 1 implique l'égalité

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2.$$

La somme

$$SS_R = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

caractérise la différence entre les valeurs prédites et observées et est appelée la *somme des carrés résiduelle*. La somme

$$SS_E = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

est appelée la *somme des carrés expliquée par régression*. La somme

$$SS_T = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

est appelée la *somme des carrés totale*. D'après le lemme 1

$$SS_T = SS_R + SS_E.$$

La somme  $SS_T$  mesure la variabilité des valeurs de  $Y$ , la somme  $SS_E$  mesure la partie de cette variabilité expliquée par la régression. Si le modèle de régression linéaire donne une bonne prédiction, c'est à dire si les  $\hat{Y}_i$  sont proches des  $Y_i$ , la somme  $SS_E$  est proche de  $SS_T$ . Donc  $SS_E$  explique une grande part de la variabilité des valeurs  $Y_i$  autour de  $\bar{Y}$ . Si la prédiction est mauvaise, la somme  $SS_E$  est petite par rapport à  $SS_T$  et  $SS_E$  n'explique pas beaucoup la variabilité des valeurs de  $Y_i$  autour de  $\bar{Y}$ . La somme des carrés résiduelle  $SS_R$  est ce qui reste de la variabilité totale après la soustraction de  $SS_E$ . D'où le nom de  $SS_R$ .

**Lemme 2.**

$$\mathbf{E}(SS_T) = (n-1)\sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [\beta^T(x^{(j)} - x^{(i)})]^2. \quad (17)$$

**Démonstration.** Notons  $M_j = \mathbf{E}(Y_j) = \beta^T x^{(j)}$ . Alors

$$\begin{aligned} \mathbf{E}(SS_T) &= \mathbf{E} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \mathbf{E} \left( \sum_{i=1}^n (Y_i - M_i + M_i - \bar{Y})^2 \right) = \\ &= \mathbf{E} \left( \sum_{i=1}^n (Y_i - M_i)^2 \right) - 2\mathbf{E} \left( \sum_{i=1}^n (Y_i - M_i)(\bar{Y} - M_i) \right) + \mathbf{E} \left( \sum_{i=1}^n (\bar{Y} - M_i)^2 \right) = \\ &= n\sigma^2 - \frac{2}{n} \mathbf{E} \left( \sum_{i=1}^n \sum_{j=1}^n (Y_i - M_i)(Y_j - M_j) \right) + \frac{1}{n^2} \sum_{i=1}^n \mathbf{E} \left( \sum_{j=1}^n (Y_j - M_j) \right)^2 = \\ &= (n-2)\sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{E} (Y_j - M_j)^2 = \\ &= (n-2)\sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{E} [(Y_j - M_j + M_j - M_i)]^2 = \\ &= (n-2)\sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{E} [\sigma^2 + (M_j - M_i)^2] = \\ &= (n-2)\sigma^2 + \sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (M_j - M_i)^2 = \\ &= (n-1)\sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [\beta^T(x^{(j)} - x^{(i)})]^2. \end{aligned}$$

Le lemme 2 est démontré.

D'après le théorème 1 et le lemme 2 on a

$$\mathbf{E}(SS_R) = (n-m-1)\sigma^2,$$

$$\mathbf{E}(SS_T) = (n-1)\sigma^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left[ \beta^T (x^{(j)} - x^{(i)}) \right]^2, \quad (18)$$

$$\mathbf{E}(SS_E) = \mathbf{E}(SS_T) - \mathbf{E}(SS_R).$$

**Corollaire.**

$$\mathbf{E}(SS_R) = \mathbf{E}(SS_T) \quad \text{et} \quad \mathbf{E}(SS_E) = 0,$$

si l'hypothèse  $H_0 : \beta_1 = \dots = \beta_m = 0$  est vérifiée, c'est-à-dire sous le modèle sans régression

$$Y_i = \beta_0 + e_i, \quad i = 1, \dots, n.$$

$$\mathbf{E}(SS_E) = \mathbf{E}(SS_T) \quad \text{et} \quad \mathbf{E}(SS_R) = 0,$$

si  $\sigma^2 = 0$ , c'est-à-dire le modèle de régression linéaire prédit sans erreur les valeurs de  $Y$ .

### 5.1.8 Le coefficient de détermination.

La variable aléatoire

$$R^2 = 1 - \frac{SS_R}{SS_T} = \frac{SS_E}{SS_T} \quad (19)$$

est appelée *le coefficient de détermination*.

$R^2$  prend ses valeurs dans le segment  $[0, 1]$ . Il représente la proportion de la variabilité des  $Y_i$  expliquée par la régression.

Si la prédiction est idéale, i.e.  $\hat{Y}_i = Y_i$ , alors  $SS_R = 0$  et  $R^2 = 1$ . S'il n'y a pas de régression, i.e. pour tous les  $x^{(i)}$  la prédiction de la moyenne  $M(x^{(i)})$  est la même :  $\hat{Y}_i = \bar{Y}$ , alors  $SS_R = SS_T$  et  $R^2 = 0$ . Donc  $R^2$  caractérise la qualité de la prédiction.

La variable aléatoire

$$R_{Y(12\dots m)} = \sqrt{R^2}$$

est appelée *le coefficient de corrélation empirique multiple*.

**Proposition.** *Le coefficient de corrélation empirique multiple est égal au coefficient de corrélation empirique simple entre les valeurs observées  $Y_i$  et les valeurs prédites  $\hat{Y}_i$  :*

$$R_{Y(12\dots m)} = r_{Y\hat{Y}} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

où  $\bar{\hat{Y}} = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i$ .

**Démonstration.** D'après le lemme 1 on a :  $\hat{Y}^T e = 0$ ,  $\bar{\hat{Y}} = \bar{Y}$  donc

$$\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})e_i = \sum_{i=1}^n \hat{Y}_i e_i = \hat{Y}^T e = 0,$$

$$\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y}) = \sum_{i=1}^n (Y_i - \bar{\hat{Y}})(e_i + \hat{Y}_i - \bar{\hat{Y}}) = \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2$$

et

$$r_{Y\hat{Y}} = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = R_{Y(12\dots m)}.$$

La proposition est démontrée.

### 5.1.9 Régression linéaire simple

Dans le cas d'une seule variable explicative ( $m = 1$ ), on a le modèle de *régression linéaire simple* :

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad (20)$$

où  $e_1, \dots, e_n$  sont les variables aléatoires i.i.d.,  $\mathbf{E}(e_i) = 0$ ,  $\mathbf{Var}(e_i) = \sigma^2$ .

On a un échantillon

$$(x_1, Y_1) \cdots (x_n, Y_n). \quad (21)$$

La réalisation de l'échantillon consiste en  $n$  paires de nombres réels qui peuvent être représentés dans le plan comme "un nuage" de points. Ces points sont dispersés autour de la *droite de régression*

$$y = \beta_0 + \beta_1 x \quad (22)$$

puisque  $\mathbf{E}(Y_i) = \beta_0 + \beta_1 x$ . Si la variance  $\sigma^2$  est petite, la plupart des points  $(x_i, Y_i)$  sont proches de cette droite. La droite (22) est inconnue parce que les paramètres  $\beta_0$  et  $\beta_1$  sont inconnus. La droite

$$y = \hat{\beta}_0 + \hat{\beta}_1 x \quad (23)$$

est la *droite de régression estimée*. Si on dispose de la réalisation de l'échantillon (21), la droite (23) peut être dessinée. Les points  $(x_i, Y_i)$  sont dispersés autour de cette droite.

Dans le cas de la régression linéaire simple, le système d'équations (6) devient

$$\beta_0 n + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n Y_i,$$

$$\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i Y_i,$$

donc

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

Si on note

$$r_{xY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

le coefficient empirique de corrélation de  $x$  et  $Y$  et

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

les variances empiriques de  $x$  et  $Y$ , alors

$$\hat{\beta}_1 = r_{xY} \frac{s_Y}{s_x}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

Les matrices  $X^T$  et  $X^T X$  sont

$$X^T = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix}, \quad X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

D'après le théorème 1,

$$\mathbf{E}(\hat{\beta}) = \beta,$$

$$\mathbf{Var}(\hat{\beta}) = \begin{pmatrix} \mathbf{Var}(\hat{\beta}_0) & \mathbf{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \mathbf{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathbf{Var}(\hat{\beta}_1) \end{pmatrix} =$$

$$\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix},$$

l'estimateur sans biais de  $\sigma^2$  est

$$\hat{\sigma}^2 = \frac{SS_R}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}$$

et

$$\mathbf{Cov}(\hat{\beta}, \hat{\sigma}^2) = 0.$$

D'après le théorème de Gauss-Markov les estimateurs

$$\hat{M}(x) = \hat{\beta}_0 + \hat{\beta}_1 x, \quad \hat{\beta}_0 \quad \text{et} \quad \hat{\beta}_1$$

sont de variance minimale dans la classes des estimateurs linéaires sans biais de  $M(x) = \mathbf{E}(Y|x)$ ,  $\beta_0$  et  $\beta_1$  respectivement,

$$\mathbf{E}(\hat{M}(x)) = M(x), \quad \mathbf{Var}(\hat{M}(x)) = \mathbf{Var}(\hat{\beta}_0) + 2x\mathbf{Cov}(\hat{\beta}_0, \hat{\beta}_1) + x^2\mathbf{Var}(\hat{\beta}_1).$$

Si  $x$  passe de  $x^{(1)}$  à  $x^{(2)}$ , alors le changement de la moyenne de  $Y$  est estimé par  $\hat{\beta}_1(x^{(2)} - x^{(1)})$ .

Notons que dans le cas de la régression linéaire simple

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad \bar{\hat{Y}} = \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

et donc

$$R^2 = \frac{[\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y})]^2}{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} =$$

$$\frac{[\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} = r_{xY}^2.$$

Le coefficient de détermination est égal au carré du coefficient de corrélation empirique des  $x_i$  et des  $Y_i$  :  $R^2 = r_{xY}^2$ . Le coefficient de corrélation empirique multiple est égal à la valeur absolue du coefficient de corrélation empirique simple :  $R_{Y(1)} = |r_{xY}|$ .

### 5.1.10 Régression normale

On a jusqu'à présent supposé seulement l'existence des deux premiers moments de  $Y_i$  dans le modèle (2). Si l'on veut obtenir des intervalles de confiance pour l'espérance  $m(x)$ , pour les paramètres  $\beta_i$ , pour des combinaisons linéaires  $\mathbf{I}^T \beta$ , ou si l'on veut vérifier des hypothèses sur les valeurs des paramètres inconnus, ou construire des tests d'ajustement, on doit faire des hypothèses supplémentaires, par exemple supposer que la répartition des  $Y_i$  appartient à une certaine classe de répartitions, la plus usuelle étant celle des lois normales.

On suppose par la suite dans ce chapitre que la loi des  $Y_i$  est normale, donc

$$Y = X\beta + e, \quad e \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n). \quad (24)$$

Certains cas où la loi de  $Y$  est différente de la loi normale sont considérés dans le chapitre sur la “régression log-linéaire”.

### 5.1.11 Estimateurs du maximum de vraisemblance

La fonction de vraisemblance sous le modèle (24) a la forme

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta^T X^{(i)})^2\right\}.$$

Pour n'importe quel  $\sigma^2 > 0$  la maximisation de  $L$  est équivalente à la minimization de

$$SS = \sum_{i=1}^n (Y_i - \beta^T x^{(i)})^2.$$

Donc les estimateurs du maximum de vraisemblance de  $\beta$  coïncident avec l'estimateur des moindres carrés  $\hat{\beta}$ . Notons que

$$\ln L(\hat{\beta}, \sigma^2) = -\frac{SS_R}{2\sigma^2} - \frac{n}{2}(\ln(2\pi) + \ln(\sigma^2)),$$

$$\frac{\partial}{\partial(\sigma^2)} \ln L(\hat{\beta}, \sigma^2) = \frac{SS_R}{2\sigma^4} - \frac{n}{2\sigma^2}$$

et donc l'estimateur du maximum de vraisemblance pour  $\sigma^2$  est :

$$\tilde{\sigma}^2 = \frac{1}{n} SS_R.$$

Cet estimateur est biaisé :

$$\mathbf{E}(\tilde{\sigma}^2) = \frac{n-m-1}{n} \sigma^2$$

est asymptotiquement ( $n \rightarrow \infty$ ) équivalent à l'estimateur

$$\hat{\sigma}^2 = SS_R / (n - m - 1),$$

considéré dans le corollaire du théorème 1.

### 5.1.12 Lois des estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ .

Considérons le théorème essentiel de la régression normale.

**Théorème 2.** Si  $\det(X^T X) \neq 0$ , alors

1. Les variables aléatoires  $\hat{\beta}$  et  $SS_R$  sont indépendantes ;

2. Les variables aléatoires  $SS_R$  et  $SS - SS_R$  sont indépendantes ;

$$3. \hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1}), \quad \frac{SS_R}{\sigma^2} \sim \chi_{n-m-1}^2, \quad \frac{SS-SS_R}{\sigma^2} \sim \chi_m^2.$$

**Démonstration.** D'après (11) on a  $X^T H = \mathbf{0}$  et

$$\begin{aligned} SS_R &= \hat{e}^T \hat{e} = Y^T H^T H Y = Y^T H Y = (Y^T - \beta^T X^T) H Y = \\ &= (Y^T - \beta^T X^T) H (Y - X\beta) = e^T H e. \end{aligned}$$

On a  $Y = X\beta + e$ , donc

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T Y - \beta = \beta + (X^T X)^{-1} X^T e - \beta =$$

$$(X^T X)^{-1} X^T e = B e,$$

$$SS_R / \sigma^2 = \tilde{e}^T H \tilde{e}, \quad (\hat{\beta} - \beta) / \sigma = B \tilde{e},$$

où  $\tilde{e} = e / \sigma \sim N(0, \mathbf{I}_n)$ . D'après (10)  $BH = \mathbf{0}$ . Donc le lemme 1 (annexe) implique que les variables aléatoires  $SS_R$  et  $\hat{\beta} - \beta$  sont indépendantes. D'après l'égalité (15) la différence

$$SS - SS_R = e^T e - \hat{e}^T \hat{e} = (\hat{Y} - X\beta)^T (\hat{Y} - X\beta) = (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)$$

est une fonction de  $\hat{\beta}$ . Donc les variables aléatoires  $SS_R$  et  $SS - SS_R$  sont aussi indépendantes.

Le vecteur  $\hat{\beta}$  est une fonction linéaire du vecteur normal  $Y$ . Donc

$$\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1}).$$

Le vecteur  $(\hat{\beta} - \beta) / \sigma \sim N(0, (X^T X)^{-1})$ . De plus,  $\text{rang}(X^T X) = m$  donc d'après le théorème 1 (annexe)

$$\frac{SS - SS_R}{\sigma^2} = \frac{1}{\sigma^2} (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)$$

suit la loi  $\chi_m^2$ .

On a obtenu

$$SS_R = \tilde{e}^T H \tilde{e}, \quad \tilde{e} \sim N(0, \mathbf{I}_n).$$

La matrice  $H$  est idempotente et

$$\text{Tr}(H) = \text{Tr} \mathbf{I}_n - \text{Tr}(X^T (X^T X)^{-1} X)$$

$$= n - \text{Tr}(X X^T (X^T X)^{-1}) = n - \text{Tr} \mathbf{I}_{m+1} = n - m - 1.$$

D'après le lemme 2 (annexe),  $SS_R \sim \chi_{n-m-1}^2$ . Le théorème est démontré.

### 5.1.13 Test de l'hypothèse $H_0 : \beta_{k+1} = \dots = \beta_m = 0$

Supposons qu'on ait le modèle de régression multiple

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} + e_i \quad (i = 1, \dots, n),$$

ou

$$Y = X\beta + e, \quad (25)$$

où

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix}, \quad \beta = (\beta_1, \dots, \beta_m)^T, \quad e = (e_1, \dots, e_n)^T.$$

Considérons le problème de la vérification de l'hypothèse

$$H_k : \beta_{k+1} = \dots = \beta_m = 0,$$

où  $k$  est un nombre fixé,  $k = 0, \dots, m-1$ . Sous  $H_k$  les covariables  $x_{k+1}, \dots, x_m$  n'améliorent pas la prédiction de la variable expliquée. Donc si  $H_k$  est vérifiée, on peut exclure ces covariables du modèle. Dans le cas  $k = 0$  on a l'hypothèse

$$H_0 : \beta_1 = \dots = \beta_m = 0.$$

On n'a pas de régression. La connaissance des valeurs des covariables ne dit rien sur les valeurs de  $Y$ .

Considérons le modèle réduit

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i \quad (i = 1, \dots, n)$$

ou

$$Y = X^{(k)}\beta^{(k)} + e, \quad (26)$$

où

$$X^{(k)} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \beta^{(k)} = (\beta_1, \dots, \beta_k)^T.$$

Notons

$$SS_R^{(k)} = \hat{e}^{(k)T} \hat{e}^{(k)} = (Y - X^{(k)}\hat{\beta}^{(k)})^T (Y - X^{(k)}\hat{\beta}^{(k)}),$$

$$SS_R^{(n)} = \hat{e}^T \hat{e} = (Y - X\hat{\beta})^T (Y - X\hat{\beta})$$

les sommes résiduelles des carrés pour le modèle (25) et (26).

**Théorème 1.** Si l'hypothèse  $H_k$  est vérifiée,  $\det(X^T X) \neq 0$ ,  $m+2 \leq n$ , alors

1.  $SS_R^{(m)}$  et  $SS_R^{(k)} - SS_R^{(m)}$  sont indépendantes.
2.  $SS_R^{(m)} \sim \sigma^2 \chi^2(n-m-1)$ ,  $SS_R^{(k)} - SS_R^{(m)} \sim \sigma^2 \chi^2(m-k)$ .

**Démonstration.** Notons que

$$SS_R^{(m)} = e^T H e, \quad SS_R^{(k)} = e^T H^{(1)} e,$$

où

$$H = \mathbf{I}_n - X(X^T X)^{-1} X^T, \quad H^{(1)} = \mathbf{I}_n - X^{(1)}(X^{(1)T} X^{(1)})^{-1} X^{(1)T}.$$

Notons  $X_0, \dots, X_m$  les colonnes de la matrice  $X$ . Considérons la suite des vecteurs orthonormaux d'ordre  $n$

$$V_0 = X\mathbf{l}_0, \dots, V_m = X\mathbf{l}_m$$

qui sont des combinaisons linéaires des  $X_0, \dots, X_m$  et sont obtenus par la méthode d'orthogonalisation de Gram-Schmidt ; ici

$$\mathbf{l}_0 = (l_{00}, 0, \dots, 0)^T, \quad \mathbf{l}_1 = (l_{10}, l_{11}, 0, \dots, 0)^T, \quad \dots, \quad \mathbf{l}_k = (l_{k0}, \dots, l_{kk}, 0, \dots, 0)^T,$$

$$\mathbf{l}_m = (l_{m0}, \dots, l_{mm})^T, \quad l_{ij} \in \mathbb{R}.$$

On a

$$V_i^T V_i = 1, \quad V_i^T V_j = 0 \quad (i \neq j).$$

Notons  $V_{m+1}, \dots, V_{n-1}$  les vecteurs orthonormaux d'ordre  $n$  qui sont orthogonaux à  $V_0, \dots, V_m$ .

Chaque vecteur  $V_i$  ( $i = 0, \dots, m$ ) est un vecteur propre de la matrice  $H$  correspondant à la valeur propre 0 : pour  $i = 0, \dots, m$  on a

$$H V_i = V_i - X(X^T X)^{-1} X^T V_i =$$

$$V_i - X(X^T X)^{-1} X^T X \mathbf{l}_i = V_i - X \mathbf{l}_i = 0.$$

Le vecteur  $V_i$  ( $i = m+1, \dots, n-1$ ) est un vecteur propre de  $H$  correspondant à la valeur propre 1 : pour  $i = m+1, \dots, n-1$  le vecteur  $V_i$  est orthogonal aux colonnes de la matrice  $X$ , donc  $X^T V_i = 0$  et

$$H V_i = V_i - X(X^T X)^{-1} X^T V_i = V_i.$$

La décomposition spectrale de  $H$  est

$$H = \sum_{i=m+1}^{n-1} V_i V_i^T,$$

donc

$$SS_R^{(m)} = e^T H e = \sum_{i=m+1}^{n-1} e^T V_i V_i^T e = \sum_{i=m+1}^{n-1} z_i^2,$$

où  $z_i = V_i^T e$ . Il faut remarquer que

$$\mathbf{E} z_i = 0, \quad \mathbf{Var} z_i = \sigma^2 V_i^T V_i = \sigma^2,$$

$$\mathbf{Cov}(z_i, z_j) = \mathbf{E} e^T V_i^T V_j e = 0 \quad (i \neq j).$$

Les variables aléatoires  $z_{m+1}, \dots, z_{n-1}$  sont indépendantes et  $z_i \sim N(0, \sigma^2)$ , ( $i = m+1, \dots, n-1$ ). Donc  $SS_R^{(m)} / \sigma^2 \sim \chi^2(n-m-1)$ .

Notons que pour  $i = 0, \dots, k$  on a  $V_i = X\mathbf{I}_i = X^{(k)}\mathbf{I}_i^*$ , où  $\mathbf{I}_i^* = (l_{i0}, \dots, l_{ii}, 0, \dots, 0)$  est le vecteur d'ordre  $k+1$ , et donc

$$H^{(k)}V_i = H^{(k)}X\mathbf{I}_i = H^{(k)}X^{(k)}\mathbf{I}_i^* = 0.$$

Pour  $i = k+1, \dots, n-1$  on a  $X^{(k)T}V_i = 0$ , donc

$$H^{(k)}V_i = V_i - X^{(k)}(X^{(k)T}X^{(k)})^{-1}X^{(k)T}V_i = V_i.$$

Par conséquent,  $V_0, \dots, V_k, V_{k+1}, \dots, V_{n-1}$  sont des vecteurs propres de  $H^{(k)}$  de valeurs propres respectives  $0, \dots, 0, 1, \dots, 1$ .

La décomposition spectrale de  $H^{(k)}$  est

$$H^{(k)} = \sum_{i=k+1}^{n-1} V_i V_i^T,$$

donc

$$SS_R^{(k)} = \sum_{i=k+1}^{n-1} e^T V_i V_i^T e_i \sim \sigma^2 \chi^2(n-k-1),$$

$$SS_R^{(k)} - SS_R = \sum_{i=k+1}^m e^T V_i V_i^T e_i \sim \sigma^2 \chi^2(m-k)$$

et les vecteurs  $SS_R$  et  $SS_R^{(k)} - SS_R$  sont indépendants.

**Corollaire.** Sous les hypothèses du théorème, la variable aléatoire

$$F = \frac{(SS_R^{(k)} - SS_R^{(m)})/(m-k)}{SS_R^{(m)}/(n-m-1)}$$

suit la loi de Fisher à  $m-k$  et  $n-k-1$  degrés de liberté.

Les sommes  $SS_R^{(m)}$  et  $SS_R^{(k)}$  caractérisent les différences entre les valeurs observées et les valeurs prédites. Sous l'hypothèse  $H_k$  la différence

$$SS_R^{(k)} - SS_R^{(m)}$$

ne doit pas être grande. Si  $H_k$  n'est pas vérifiée, alors les covariables  $x_{k+1}, \dots, x_m$  améliorent la prédiction et la différence  $SS_R^{(k)} - SS_R^{(m)}$  doit être plus grande. Donc on rejette  $H$  si

$$F > F_{1-\alpha}(m-k, n-k-1),$$

où  $F_{1-\alpha}$  est le  $(1-\alpha)$  quantile de la loi de Fisher.

L'hypothèse la plus intéressante de point de vue pratique est

$$H_{m-1} : \beta_m = 0.$$

Elle signifie que le modèle avec  $m-1$  covariables  $x_1, \dots, x_{m-1}$  donne la même prédiction que le modèle avec  $m$  covariables  $x_1, \dots, x_m$ , i.e. la covariable  $x_m$  peut être exclue du modèle. La statistique de test pour cette hypothèse est

$$F = \frac{SS_R^{(m-1)} - SS_R^{(m)}}{SS_R^{(m)}/(n-m-1)}.$$

L'hypothèse est rejetée avec le niveau de signification  $\alpha$ , si

$$F > F_{1-\alpha}(1, n-m).$$

Notons que dans le cas de l'hypothèse  $H_0$  on a

$$SS_R^{(0)} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = SS_T, \quad SS_R^{(0)} - SS_R^{(m)} = SS_E^{(m)},$$

où  $SS_T$  et  $SS_E^{(m)}$  sont la somme des carrés totale et la somme des carrés expliquée par la régression, respectivement, dans le modèle (25). La statistique de test pour  $H_0$  est

$$F = \frac{SS_E/m}{SS_R/(n-m-1)} \sim F_{m, n-m-1}.$$

Donc l'hypothèse  $H_0$  sur l'absence de la régression est rejetée avec le niveau de signification  $\alpha$ , si

$$F > F_{1-\alpha}(m, n-m-1).$$

Dans le cas du modèle linéaire simple cette hypothèse est équivalente à l'hypothèse

$$H_0 : \beta_1 = 0$$

et la statistique de test

$$F = \frac{SS_E}{SS_R/(n-2)} \sim F_{1, n-2}.$$

L'hypothèse est rejetée avec le niveau de signification  $\alpha$ , si

$$F > F_{1-\alpha}(1, n-2).$$

En utilisant la relation entre la loi de Fisher de 1 et  $(n-2)$  degrés de liberté et la loi de Student de  $(n-2)$  degrés de liberté, la région critique peut être écrite en forme équivalente :

$$t > t_{1-\alpha}(n-2),$$

où  $t = \sqrt{F}$  et  $t_{1-\alpha}(n-2)$  est la  $(1-\alpha)$  quantile de la loi de Student de  $(n-2)$  degrés de liberté.

### 5.1.14 Les coefficients empiriques de la corrélation partielles

Considérons la statistique

$$R_{Y(X_{k+1} \dots X_m)(1 \dots k)}^2 = \frac{SS_R^{(k)} - SS_R^{(m)}}{SS_R^{(k)}} = \frac{SS_E^{(m)} - SS_E^{(k)}}{SS_T - SS_E^{(k)}}.$$

La somme des carrés  $SS_E^{(m)}$  et  $SS_E^{(k)}$  mesurent les parties de variabilité des valeurs  $Y_i$  expliquées par la régression dans les modèles (25) et (26), respectivement, donc la statistique

$$SS_E^{(m)} - SS_E^{(k)}$$

mesure la partie de variabilité des valeurs de  $Y_i$ , expliqué par l'inclusion des covariables  $x_{k+1}, \dots, x_m$  complémentaires à  $x_1, \dots, x_k$ .

La statistique

$$SS_R^{(k)} = SS_T - SS_E^{(k)}$$

mesure la variabilité résiduelle des  $Y_i$ , i.e. la variabilité qui n'est pas expliquée par le modèle (26). Donc  $R^2$  est la *proportion de la variabilité résiduelle du modèle (26) expliquée par introduction des nouvelles covariables  $x_{k+1}, \dots, x_m$* .

Notons que

$$SS_R^{(m)} = SS_T(1 - R_{Y(1\dots m)}^2), \quad SS_R^{(k)} = SS_T(1 - R_{Y(1\dots k)}^2),$$

donc

$$R_{Y(X_{k+1}\dots X_m)(1\dots k)}^2 = \frac{R_{Y(1\dots m)}^2 - R_{Y(1\dots k)}^2}{1 - R_{Y(1\dots k)}^2}.$$

La statistique

$$R_{Y(X_{k+1}\dots X_m)(1\dots k)} = \sqrt{R_{Y(X_{k+1}\dots X_m)(1\dots k)}^2}$$

est appelée *le coefficient empirique de corrélation partiel de  $Y$  et  $X_{k+1}, \dots, X_m$* . Il mesure la corrélation entre  $Y$  et  $(X_{k+1}\dots X_m)$  après l'élimination de leur dépendance de  $X_1\dots X_k$ .

$R_{YX_m(1\dots k)}^2$  est la *proportion de la variabilité résiduelle du modèle avec  $(m-1)$  covariables  $x_1, \dots, x_{m-1}$  expliquée par introduction de la  $m$ -ème covariable  $x_m$* . On a

$$R_{YX_m(1\dots m-1)}^2 = \frac{R_{Y(1\dots m)}^2 - R_{Y(1\dots m-1)}^2}{1 - R_{Y(1\dots m-1)}^2}. \quad (27)$$

La statistique

$$R_{YX_m(1\dots m-1)} = \sqrt{R_{YX_m(1\dots m-1)}^2}$$

est appelée *le coefficient empirique de corrélation partielle de  $Y$  et  $X_m$* . Il mesure la corrélation entre  $Y$  et  $X_m$  après l'élimination de leur dépendance de  $X_1\dots X_{m-1}$ . L'égalité (27) implique

$$1 - R_{Y(1\dots m)}^2 = \left(1 - R_{YX_m(1\dots m-1)}^2\right) \left(1 - R_{Y(1\dots m-1)}^2\right).$$

### 5.1.15 Intervalles de confiance pour les coefficients $\beta$ et leur combinaisons linéaires

Considérons le modèle de régression multiple (25). Le théorème 1 implique que dans le cas normale

$$\hat{\beta} \sim N_{m+1}(\beta, \sigma^2(X^T X)^{-1}), \quad SS_R/\sigma^2 \sim \chi^2(n-m-1)$$

et les variables aléatoires  $\hat{\beta}$  et  $SS_R$  sont indépendantes. Notons  $s_{ii}$  les éléments diagonaux de la matrice  $(X^T X)^{-1} = (s_{ij})$ . Alors

$$\frac{\hat{\beta}_i - \beta_i}{\sigma s_{ii}} \sim N(0, 1), \quad \frac{SS_R}{\sigma^2} \sim \chi^2(n-m-1)$$

et donc

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\widehat{\mathbf{Var}}(\hat{\beta}_i)}} \sim St(n - m - 1),$$

où

$$\widehat{\mathbf{Var}}(\hat{\beta}_i) = s_{ii}\hat{\sigma}^2 = s_{ii}MS_R.$$

Le  $\gamma = 1 - \alpha$  intervalle de confiance pour  $\beta_i$  est

$$\hat{\beta}_i \pm s_{ii}\sqrt{MS_R}t_{1-\alpha/2}(n - m - 1),$$

où  $t_{1-\alpha/2}(n - m - 1)$  est le  $(1 - \alpha/2)$  quantile de la loi de Student de  $n - m - 1$  degrés de liberté.

Si on considère le modèle avec interactions, alors certaines combinaisons linéaires des paramètres  $\beta_0, \dots, \beta_m$  ont le sens pratique. Le paramètre

$$c = \sum_{i=0}^m l_i \beta_i = \mathbf{l}^T \boldsymbol{\beta}$$

est estimé par la statistique  $\hat{c} = \mathbf{l}^T \hat{\boldsymbol{\beta}}$ , donc

$$\mathbf{Var}(\hat{c}) = \mathbf{l}^T \mathbf{Var}(\hat{\boldsymbol{\beta}}) \mathbf{l} = \sigma^2 \mathbf{l}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{l}$$

et

$$t = \frac{\hat{c} - c}{\sqrt{\mathbf{l}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{l} MS_R}} \sim St(n - m - 1).$$

Le  $(1 - \alpha)$  intervalle de confiance pour  $c$  est

$$\hat{c} \pm \sqrt{\mathbf{l}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{l} MS_R} t_{1-\alpha/2}(n - m - 1). \quad (28)$$

### 5.1.16 Intervalles de confiance pour les valeurs de la fonction de régression $m(x)$

Fixons la valeur  $x_0$  du vecteur des covariables. Considérons la valeur

$$m(x_0) = \mathbf{E}(Y | x_0) = \beta_0 + \beta_1 x_{01} + \dots + \beta_m x_{0m} = x_0^T \hat{\boldsymbol{\beta}}$$

de la fonction de régression.

La formule (28) implique que  $(1 - \alpha)$  intervalle de confiance pour  $m(x_0)$  est

$$x_0^T \hat{\boldsymbol{\beta}} \pm \sqrt{x_0^T (\mathbf{X}^T \mathbf{X})^{-1} x_0 MS_R} t_{1-\alpha/2}(n - m - 1).$$

### 5.1.17 Prédiction de la nouvelle observation

Supposons que  $x_0 = (1, x_{01}, \dots, x_{0m})^T$  est un vecteur des covariables fixé et  $Y_{n+1}(x_0)$  est la  $(n+1)$  observation de la variable dépendante.

Intervalle aléatoire  $(U_1, U_2)$  tel que

$$\mathbf{P}\{U_1 < Y_{n+1}(x_0) < U_2\} = 1 - \alpha$$

est appelé la  $(1 - \alpha)$  *intervalle de prédiction pour  $Y_{n+1}(x_0)$* .

Les variables aléatoires  $Y_{n+1}(x_0)$  et  $x_0^T \hat{\beta}$  sont indépendantes et

$$Y_{n+1}(x_0) \sim N(x_0^T \beta, \sigma^2), \quad x_0^T \hat{\beta} \sim N(x_0^T \beta, \sigma^2 x_0^T (X^T X)^{-1} x_0),$$

donc

$$Y_{n+1}(x_0) - x_0^T \hat{\beta} \sim N(0, \sigma^2 (1 + x_0^T (X^T X)^{-1} x_0)).$$

La statistique

$$t = \frac{Y_{n+1}(x_0) - x_0^T \hat{\beta}}{\sqrt{MS_R (1 + x_0^T (X^T X)^{-1} x_0)}} \sim St(n - m - 1),$$

donc le  $(1 - \alpha)$  intervalle de prédiction pour  $Y_{n+1}(x_0)$  est

$$x_0^T \hat{\beta} \pm \sqrt{MS_R (1 + x_0^T (X^T X)^{-1} x_0)} F_{1-\alpha/2}(n - m - 1).$$

Il est plus large que l'intervalle de confiance pour la moyenne  $m(x_0) = x_0^T \beta$ .

Prédiction de la nouvelle observation  $Y_{n+1}(x_0)$  est plus incertaine que la prédiction de la moyenne de  $Y(x_0)$ .

Le chapitre n'est pas achevé. Les problèmes de diagnostique, step by step régression, liaison avec ANOVA, etc, sont à ajouter.

### 5.1.18 Analyse des résidus

Avant de faire inférences il est nécessaire de vérifier si le modèle est bien ajusté aux données réelles. Les suppositions principales du modèle de régression linéaire sont :

- l'égalité des variances des variables aléatoires  $e_i = Y_i - \beta^T x$  ;
- l'indépendance des  $e_i$  ;
- la linéarité de la fonction de régression  $M(x) = \mathbf{E}(Y(x))$  ;
- la normalité des variables aléatoires  $e_i$  (si l'on construit les intervalles de confiance ou vérifie des hypothèses).

Considérons des méthodes non formels de vérification des suppositions du modèle.

Dans le cas du modèle de régression linéaire simple des nuages des points  $(x_i, Y_i)$  peuvent être considérés. Si ces points sont dispersés autour d'une certaine courbe, qui n'est pas une droite, on peut supposer que le modèle n'est pas bien choisi.

Dans le cas  $m > 1$  des résidus  $\hat{e}_i$  peuvent être considérés. Notons que

$$\hat{e} = HY, \quad \mathbf{E}(\hat{e}) = \mathbf{0}, \quad \mathbf{Var}(\hat{e}) = \sigma^2 H,$$

où

$$H = I_n - X(X^T X)^{-1} X^T = (h_{ij}),$$

et donc

$$\mathbf{Var} \left( \frac{\hat{e}_i}{\sigma \sqrt{h_{ii}}} \right) = 1.$$

La variance est estimée par

$$\hat{\sigma}^2 = MS_R = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(n - m + 1)}.$$

Notons

$$\tilde{e}_i = \frac{\hat{e}_i}{\sqrt{MS_R h_{ii}}}.$$

On a  $\mathbf{E}(\tilde{e}_i) \approx 0$ ,  $\mathbf{Var}(\tilde{e}_i) \approx 1$ . Les variables aléatoires  $\tilde{e}_i$  sont appelées *les résidus standardisés*.

Si on considère le plan avec l'axe des abscisses  $Y$  et l'axe des ordonnées  $e$ , les points  $(\hat{Y}_i, \hat{e}_i)$ ,  $(i = 1, \dots, n)$ , sont dispersés autour de la droite horizontale avec l'axe de symétrie  $e = 0$ . Si les variances des  $e_i$  ne sont pas égales, on dit qu'on a *heterodescasité*. Si le modèle pour  $M(x)$  est bien choisi mais il y a heterodescasité, alors les points  $(\hat{Y}_i, \hat{e}_i)$ , sont dispersés aussi autour de la droite  $e = 0$ , mais la largeur de la bande n'est pas constante. Par exemple, si la variance augmente avec augmentation de  $\hat{Y}_i$ , alors la bande s'élargisse.

Si les points  $(\hat{Y}_i, \hat{e}_i)$  sont dispersés autour d'une autre courbe différente de  $e = 0$ , le modèle pour  $M(x)$  n'est pas bien choisi.

Au lieu des points  $(\hat{Y}_i, \hat{e}_i)$  on peut considérer les points  $(x_{ij}, \hat{e}_i)$ ,  $(i = 1, \dots, n)$  pour  $j$  fixé. Si le modèle est bien choisi, ces points doivent être dans la bande horizontale avec l'axe de symétrie  $e = 0$ . Sinon on peut supposer que la  $j$ -ème covariable n'influence pas  $M(x)$  linéairement ou il faut inclure plus de covariables dans le modèle.

Par exemple, si le vrai modèle est

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

mais on a choisi le modèle

$$Y'_i = \beta'_0 + \beta'_1 x_i + e'_i,$$

alors

$$\hat{e}_i = Y_i - \hat{Y}'_i = \beta_0 - \hat{\beta}'_0 + (\beta_1 - \hat{\beta}'_1) x_i + \beta_2 x_i^2 + e_i$$

et donc les points  $(x_i, \hat{e}_i)$  seront dispersés autour d'une parabole.

Si le vrai modèle est

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$$

mais on a choisi

$$Y'_i = \beta'_0 + \beta'_1 x_{i1} + e'_i,$$

alors

$$\hat{e}_i = \beta_0 - \hat{\beta}'_0 + (\beta_1 - \hat{\beta}'_1) x_{i1} + \beta_2 x_{i2} + e_i$$

et les points  $(x_{i1}, \hat{e}_i)$ ,  $(i = 1, \dots, n)$  (ainsi que les points  $(x_{i2}, \hat{e}_i)$ ,  $(i = 1, \dots, n)$ ) ne seront pas dispersés autour de la ligne horizontale  $e = 0$ .

Si les plots des résidus indiquent que le modèle est mal choisi, il suffit souvent de faire des transformations simples des  $x_i$  et  $Y_i$  pour obtenir le bon modèle.

Par exemple, si  $Y_i \sim LN(\beta_0 + \beta_1 \ln x_i, \sigma^2)$ , alors  $\ln Y_i = \beta_0 + \beta_1 \ln x_i + e_i$ , où  $e_i \sim N(0, \sigma^2)$ . Si on fait des transformations  $Y'_i = \ln Y_i$ ,  $x'_i = \ln x_i$ , alors on a le modèle linéaire simple  $Y'_i = \beta_0 + \beta_1 x'_i + e_i$ . Notons que dans ce cas les variances

$$\mathbf{Var}(Y_i) = x_i^{2\beta_1} e^{\sigma^2 + 2\beta_0} (e^{\sigma^2} - 1)$$

ne sont pas constants, mais les variances  $\mathbf{Var}(\ln Y_i) = \sigma^2$  sont constantes.

Considérons plusieurs exemples des transformations pour les modèles de régression à une covariable. Notons  $y = M(x)$ . On a

1. si  $y = \alpha x^\beta$ , alors  $y' = \ln y$ ,  $x' = \ln x$  et  $y' = \ln \alpha + \beta x'$  ;
2. si  $y = \alpha e^{\beta x}$ , alors  $y' = \ln y$  et  $y' = \ln \alpha + \beta x$  ;
3. si  $y = \frac{x}{\alpha x - \beta}$ , alors  $y' = 1/y$ ,  $x' = 1/x$  et  $y' = \alpha - \beta x'$  ;
4. si  $y = \alpha + \beta \ln x$ , alors  $x' = \ln x$  et  $y = \alpha + \beta x'$  ;
5. si  $y = e^{\alpha + \beta x} / (1 + e^{\alpha + \beta x})$ , alors  $y' = \ln \frac{y}{1-y}$  et  $y' = \alpha + \beta x$ .

Si  $Y$  compte le nombre de certains événements, la transformation  $Y' = \sqrt{Y}$  stabilise souvent la variance.

Considérons les méthodes non formels de vérification de la normalité des résidus. Si  $e_i \sim N(0, \sigma^2)$ , alors  $\hat{e} = HY \sim N(0, \sigma^2 H)$  et  $\tilde{e}_i \sim N(0, 1)$ . Souvent les corrélations entre  $\tilde{e}_i$  sont petites et on considère  $\tilde{e}_1, \dots, \tilde{e}_n$  comme i.i.d.  $N(0, 1)$ .

Pour tester la normalité grossièrement on peut faire l'hystogramme des  $\tilde{e}_i$ . On peut aussi faire des plots suivants :

soient  $\tilde{e}_{(1)} \leq \dots \leq \tilde{e}_{(n)}$  les statistiques d'ordre de  $\tilde{e}_1, \dots, \tilde{e}_n$ . Si  $Z_{(j)}$  est la  $j$ -ème statistique d'ordre de la loi  $N(0, 1)$ , ( $j = 1, \dots, n$ ), alors notons  $m_{(j)} = \mathbf{E}(Z_{(j)})$ . Les espérances  $m_{(j)}$  ne dépendent pas des paramètres inconnus. Si  $\tilde{e}_{(i)}$  sont des statistiques d'ordre de la loi  $N(0, 1)$ , alors les points  $(\tilde{e}_{(i)}, m_{(i)})$  doivent être dispersés autour de la droite  $e = m$  dans le plan  $(0em)$ .

On peut utiliser une autre méthode : mettre sur le plan  $(0eq)$  les points  $(\tilde{e}_{(i)}, q_{(i)})$ , où  $q_{(i)} = \Phi^{-1} \left( \frac{i-1/2}{n} \right)$  sont des  $\left( \frac{i-1/2}{n} \right)$ -quantiles de la loi  $N(0, 1)$ . Alors ces points doivent être dispersés autour de la droite  $e = q$ .

Indépendance des variables aléatoires  $e_i$  peut être vérifiée en utilisant le test de Durbin-Watson.

Considérons la statistique

$$r_1 = \frac{\sum_{i=2}^n (\hat{e}_{i-1} - \bar{\hat{e}})(\hat{e}_i - \bar{\hat{e}})}{\sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}})^2} = \frac{\sum_{i=2}^n \hat{e}_{i-1} \hat{e}_i}{\sum_{i=1}^n \hat{e}_i^2},$$

appelée la première autocorrelation des  $(\hat{e}_1, \hat{e}_2), (\hat{e}_2, \hat{e}_3), \dots, (\hat{e}_{n-1}, \hat{e}_n)$  ; ici  $\bar{\hat{e}} = \sum_{i=1}^n \hat{e}_i$ . Elle est très proche au coefficient de corrélation empirique linéaire de ces paires. Alors la statistique

$$d = \frac{\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n \hat{e}_i^2} \approx 2(1 - r_1)$$

est appelée la *statistique de Durbin-Watson*.  $r_1$  est proche à zéro, si les variables aléatoires  $\hat{e}_i$  sont indépendantes. Alors la statistique  $d$  est proche à 2 dans ce cas. La loi de  $d$  ne dépend pas des paramètres inconnus et les valeurs critiques  $d_i$  et  $d_s$  de  $d$  sont tabulées. On rejette l'hypothèse d'indépendance, si  $d < d_i$  ou  $d > d_s$ .

Même si le modèle est bien choisi, l'estimation peut être mauvaise, si parmi les points  $(x_i, Y_i)$  il y a des valeurs aberrantes, i.e. les points avec grands résidus  $\hat{e}_i = Y_i - \hat{Y}_i$ . La valeur aberrante est influente si son retrait change beaucoup la valeur de l'estimateur de  $\beta$ . L'influence de  $(x_i, Y_i)$  peut être mesurée à l'aide de la *distance de Cook*. Pour calculer cette distance, on calcule la *valeur prédite ajustée*  $\hat{Y}_{ia}$  qui est déterminée comme  $\hat{Y}_i$ , utilisant seulement les points

$$(x_1, Y_1), \dots, (x_{i-1}, Y_{i-1}), (x_{i+1}, Y_{i+1}), \dots, (x_n, Y_n).$$

La distance de Cook est déterminée par la formule suivante :

$$C_i^2 = \frac{\sum_{i=1}^n (\hat{Y}_{ia} - \hat{Y}_i)^2}{(m+1)MSR}.$$

La règle pratique : si  $C_i^2 > 1$ , le point est influent.

## 5.2 Annexe

Soit  $\mathbb{X} = (X_1, \dots, X_n)$  un échantillon,  $X_i \sim N(0, 1)$ . On considère la forme linéaire  $b^T X$ ,  $b = (b_1, \dots, b_n)^T$  et les formes quadratiques  $X^T A X$ ,  $X^T B X$ , où  $A$  et  $B$  sont des matrices symétriques.

### Lemme .

- a) Si  $b^T A = \mathbf{0}$ , alors  $X^T A X$  et  $b^T X$  sont indépendantes ;
- b) Si  $AB = \mathbf{0}$ , alors les formes quadratiques  $X^T A X$  et  $X^T B X$  sont indépendantes.

**Démonstration.** a). On suppose que  $\text{rang}(A) = r$ . Comme  $A$  est une matrice symétrique, on peut écrire sa décomposition spectrale :

$$A = \sum_{i=1}^r \lambda_i h_i h_i^T, \quad (28)$$

où  $\lambda_1, \dots, \lambda_r$  et  $h_1, \dots, h_r$  sont les valeurs propres positives et les vecteurs propres, respectivement, de la matrice  $A$ ,  $h_i^T h_j = 0$  ( $i \neq j$ ),  $h_i^T h_i = 1$ . On a

$$X^T A X = \sum_{i=1}^r \lambda_i (h_i^T X)^2 = (\sqrt{\lambda_1} h_1^T X, \dots, \sqrt{\lambda_r} h_r^T X) (\sqrt{\lambda_1} h_1^T X, \dots, \sqrt{\lambda_r} h_r^T X)^T.$$

L'égalité  $b^T X = 0$  implique

$$\text{Cov}(b^T X, h_i^T X) = b^T \text{Var}(X) h_i = b^T h_i = \lambda_i^{-1} b^T A h_i = 0,$$

$b^T X$  et  $h_i^T X$  sont des variables aléatoires normales et non-corrélées, donc indépendantes. Il s'ensuit que les variables aléatoires  $X^T A X$  et  $b^T X$  sont indépendantes.

b) On écrit la décomposition spectrale des matrices  $A$  et  $B$  :

$$A = \sum_{i=1}^r \lambda_i h_i h_i^T, \quad B = \sum_{j=1}^p \mu_j \mathbf{1}_j \mathbf{1}_j^T.$$

On a

$$X^T A X = \sum_{i=1}^r \lambda_i (h_i^T X)^2, \quad X^T B X = \sum_{j=1}^p \mu_j (b_j^T X)^2, \quad \mathbf{Cov}(h_i^T X, \mathbf{1}_j^T X) = h_i^T \mathbf{1}_j = 0,$$

donc  $h_i^T X$  et  $\mathbf{1}_j^T X$  et par conséquent  $X^T A X$  et  $X^T B X$  sont indépendantes. Le lemme est démontré.

**Lemme .** Soit  $A$  une matrice idempotente, i.e.  $A^2 = A$ , telle que  $\text{rang}(A) = r \leq n$ . Alors  $r = \text{Tr}A$  et  $X^T A X \sim \chi^2(r)$ .

**Démonstration.** On écrit la décomposition spectrale (28).  $A$  est idempotente, donc  $\lambda_1 = \dots = \lambda_r = 1$  et

$$X^T A X = \sum_{i=1}^r (h_i^T X)^2. \quad (29)$$

Les vecteurs propres  $h_i$  et  $h_j$  sont orthogonaux,

$$\mathbf{Cov}(h_i^T X, h_j^T X) = h_i^T h_j = 0$$

et donc les variables aléatoires  $h_i X$  et  $h_j X$  sont indépendantes. On a

$$\mathbf{E}(h_i^T X) = h_i^T \mathbf{E}(X) = \mathbf{0}, \quad \mathbf{Var}(h_i^T X) = h_i^T h_i = 1,$$

donc  $h_i^T X \sim N(0, 1)$  et d'après le lemme  $X^T A X \sim \chi_r^2$ . Le lemme est démontré.

**Théorème.** Si  $X \sim N(\mu, \Sigma)$ ,  $\det(\Sigma) \neq 0$ , alors

$$(X - \mu)^T \Sigma^{-1} (X - \mu) \sim \chi^2(n).$$

**Démonstration.** La matrice  $\Sigma$  est symétrique,  $\det(\Sigma) \neq 0$ . Donc il existe une matrice orthogonale  $H$  telle que  $H^T \Sigma H = D$  soit diagonale. De plus, les éléments de cette diagonale  $\lambda_1, \dots, \lambda_n$  sont positifs. Considérons le vecteur

$$Y = D^{-1/2} H^T (X - \mu).$$

On a

$$\mathbf{Var}(Y) = D^{-1/2} H^T \Sigma H D^{-1/2} = \mathbf{I}_n,$$

donc

$$(X - \mu)^T \Sigma^{-1} (X - \mu) = Y^T D^{1/2} H^T \Sigma^{-1} H D^{1/2} Y = Y^T Y \sim \chi_n^2.$$

Le théorème est démontré.

**Théorème.** Soit  $X = (X_1, \dots, X_n)$  un échantillon,  $X_i \sim N(\mu, \sigma^2)$ . Alors les moments empiriques

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

sont indépendants et

$$\sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1), \quad nS_X^2/\sigma^2 \sim \chi^2(n-1).$$

**Démonstration.** Notons  $Y_i = (X_i - \mu)/\sigma \sim N(0, 1)$ ,  $Y = (Y_1, \dots, Y_n)^T$ ,  $\bar{Y} = (\bar{X} - \mu)/\sigma$ ,  $S_Y^2 = S_X^2/\sigma^2$ .

Il suffit de démontrer que les variables aléatoires  $\bar{Y}$  et  $S_Y^2$  sont indépendantes.

Considérons le vecteur  $b = (1/n, \dots, 1/n)^T$  et la matrice  $B = (1/n)_{n \times n}$ . On a

$$\bar{Y} = b^T Y, \quad nS_Y^2 = (Y - BY)^T (Y - BY) = Y^T (\mathbf{I}_n - B)^2 Y.$$

La matrice  $\mathbf{I}_n - B$  est idempotente :

$$(\mathbf{I}_n - B)^2 = \mathbf{I}_n - 2B + B^2 = \mathbf{I}_n - B$$

et

$$b^T (\mathbf{I}_n - B) = b^T - b^T B = b^T - b^T = 0.$$

D'après le lemme, les variables aléatoires  $\bar{Y}$  et  $S_Y^2$  sont indépendantes. On a

$$\text{Tr}(\mathbf{I}_n - B) = \text{Tr}\mathbf{I}_n - \text{Tr}B = n - 1.$$

D'après le lemme  $nS_Y^2 \sim \chi^2(n-1)$ . Le théorème est démontré.

**Exemple 1.** Soient  $\mathbf{X} = (X_1, \dots, X_{n_1})^T$  et  $\mathbf{Y} = (Y_1, \dots, Y_{n_2})^T$  deux échantillons indépendants,  $X_i \sim N(\mu_1, \sigma_1^2)$ ,  $Y_i \sim N(\mu_2, \sigma_2^2)$ . Construire le test de rapport de vraisemblance pour tester l'hypothèse  $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ .

Solution. Notons que  $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \in \Theta = \mathbf{R} \times \mathbf{R} \times \mathbf{R}^+ \times \mathbf{R}^+$  et  $(\mu_1, \mu_2, \sigma^2) \in \Theta_0 = \mathbf{R} \times \mathbf{R} \times \mathbf{R}^+ \subset \Theta$ .

La fonction de vraisemblance pour  $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$  est

$$L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{1}{(2\pi\sigma_1^2)^{n_1/2}} \frac{1}{(2\pi\sigma_2^2)^{n_2/2}} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{n_1} (X_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right\}.$$

Le logarithme de la fonction de vraisemblance est

$$\begin{aligned} \ln L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= -\frac{n_1}{2} (\ln(2\pi) + \ln \sigma_1^2) - \frac{n_2}{2} (\ln(2\pi) + \ln \sigma_2^2) \\ &\quad - \frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} (X_i - \mu_1)^2 - \frac{1}{\sigma_2^2} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2. \end{aligned}$$

Les estimateurs de maximum de vraisemblance vérifient le système des équations

$$\begin{aligned} \frac{\partial L}{\partial \mu_1} &= \frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} (X_i - \mu_1) = 0, \\ \frac{\partial L}{\partial \mu_2} &= \frac{1}{\sigma_2^2} \sum_{i=1}^{n_2} (Y_i - \mu_2) = 0, \end{aligned}$$

$$\frac{\partial L}{\partial \sigma_1^2} = -\frac{n_1}{2\sigma_1^2} + \frac{1}{2\sigma_1^4} \sum_{i=1}^{n_1} (X_i - \mu_1)^2 = 0,$$

$$\frac{\partial L}{\partial \sigma_2^2} = -\frac{n_2}{2\sigma_2^2} + \frac{1}{2\sigma_2^4} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 = 0.$$

Donc  $\hat{\mu}_1 = \bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ ,  $\hat{\mu}_2 = \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$ ,  $\hat{\sigma}_1^2 = s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$ ,  $\hat{\sigma}_2^2 = s_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ .

Notons  $n = n_1 + n_2$ . Sous  $H_0$  la fonction de vraisemblance pour  $(\mu_1, \mu_2, \sigma^2)$  est

$$L_1(\mu_1, \mu_2, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^{n_1} (X_i - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right] \right\}.$$

Le logarithme de la fonction de vraisemblance est

$$\ln L_1(\mu_1, \mu_2, \sigma^2) = -\frac{n}{2} (\ln(2\pi) + \ln \sigma^2) - \frac{1}{\sigma^2} \left\{ \sum_{i=1}^{n_1} (X_i - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right\}.$$

Les estimateurs de maximum de vraisemblance vérifient le système des équations

$$\frac{\partial L_1}{\partial \mu_1} = \frac{1}{\sigma^2} \sum_{i=1}^{n_1} (X_i - \mu_1) = 0,$$

$$\frac{\partial L_1}{\partial \mu_2} = \frac{1}{\sigma^2} \sum_{i=1}^{n_2} (Y_i - \mu_2) = 0,$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \left\{ \sum_{i=1}^{n_1} (X_i - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right\} = 0.$$

Donc  $\hat{\mu}_1 = \bar{X}$ ,  $\hat{\mu}_2 = \bar{Y}$ ,  $\hat{\sigma}^2 = \frac{1}{n} (n_1 s_1^2 + n_2 s_2^2)$ . Les maximums des fonctions  $L$  et  $L_1$  sont

$$\hat{L} = L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2) = \frac{1}{(2\pi)^{n/2} s_1^{n_1} s_2^{n_2}} e^{-n/2}$$

et

$$\hat{L}_1 = L_1(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2) = \frac{1}{(2\pi)^{n/2} \left( \frac{n_1}{n} s_1^2 + \frac{n_2}{n} s_2^2 \right)^{n/2}} e^{-n/2}.$$

La région critique pour  $H_0$  est défini par l'inégalité

$$-2 \ln \left( \frac{\hat{L}_1}{\hat{L}} \right) > C.$$

On a

$$\begin{aligned} \frac{\hat{L}_1}{\hat{L}} &= \frac{s_1^{n_1/2} s_2^{n_2/2}}{\left( \frac{n_1}{n} s_1^2 + \frac{n_2}{n} s_2^2 \right)^{n/2}} \\ &= \left( \frac{n_1}{n} + \frac{n_2}{n} \frac{s_2^2}{s_1^2} \right)^{-n_1/2} \left( \frac{n_1}{n} \frac{s_1^2}{s_2^2} + \frac{n_2}{n} \right)^{-n_2/2}. \end{aligned}$$

Donc la région critique est défini par l'inégalité

$$n_1 \ln \left( \frac{n_1}{n} + \frac{n_2 s_2^2}{n s_1^2} \right) + n_2 \ln \left( \frac{n_1 s_1^2}{n s_2^2} + \frac{n_2}{n} \right) > C.$$

Posons  $x = s_2^2/s_1^2$  et étudions la fonction

$$g(x) = n_1 \ln \left( \frac{n_1}{n} + \frac{n_2}{n} x \right) - n_2 \ln \left( \frac{n_1}{n x} + \frac{n_2}{n} \right).$$

Sa dérivée est

$$g'(x) = n_1 n_2 \frac{n_2 x^2 + (n_1 - n_2)x - n_1}{(n_1 + n_2 x)(n_1 x + n_2 x^2)}.$$

Les racines de l'équation quadratique

$$n_2 x^2 + (n_1 - n_2)x - n_1 = 0$$

sont  $x_1 = 1$  et  $x_2 = -\frac{n_1}{n_2}$ . Donc  $g'(x) < 0$  si  $x \in ]0, 1[$ ,  $g'(x) > 0$  si  $x \in ]1, +\infty[$ . La fonction  $g$  est décroissante sur l'intervalle  $]0, 1[$ , croissante sur  $]1, +\infty[$  et le minimum est atteint dans le point 1. L'inégalité  $g(x) > C$  est vérifiée si et seulement si  $x < c_1$  ou  $x > c_2$ . Donc la région critique pour  $H_0$  est déterminée par les égalités

$$\frac{s_2^2}{s_1^2} < c_1 \quad \text{ou} \quad \frac{s_2^2}{s_1^2} > c_2.$$

Fixons le niveau de signification  $\alpha$ . Les constantes  $c_1$  et  $c_2$  sont trouvées de l'égalités

$$\mathbf{P}\left\{\frac{s_2^2}{s_1^2} < c_1 \mid H_0\right\} = \alpha/2, \quad \mathbf{P}\left\{\frac{s_2^2}{s_1^2} > c_2 \mid H_0\right\} = \alpha/2.$$

Donc  $c_1 = f_{\alpha/2}(n_2 - 1, n_1 - 1)$  et  $c_2 = f_{1-\alpha/2}(n_2 - 1, n_1 - 1)$  sont des quantiles de la répartition de Fisher de  $n_2 - 1$  et  $n_1 - 1$  degrés de liberté.

**Exemple 2.** Soit  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^T$  un échantillon de la loi normale bivarié, i.e.

$$\mathbf{Z}_i = (X_i, Y_i)^T \sim f_{X_i, Y_i}(x, y),$$

où la densité

$$f_{X_i, Y_i}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-a_1)^2}{\sigma_1^2} - \frac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} \right] \right\}$$

pour tout  $(x, y) \in \mathbf{R}^2$ .

Construire le test de rapport de vraisemblance pour tester l'hypothèse  $H_0 : \rho = 0$  qui est équivalente à l'hypothèse que les variables aléatoires  $X_i$  et  $Y_i$  sont indépendantes.

Notons que  $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \in \Theta = \mathbf{R} \times \mathbf{R} \times \mathbf{R}^+ \times \mathbf{R}^+ \times [-1, 1]$  et

$$(\mu_1, \mu_2, \sigma^2) \in \Theta_0 = \mathbf{R} \times \mathbf{R} \times \mathbf{R}^+ \subset \Theta$$

La fonction de vraisemblance pour  $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$  est

$$L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \\ \left( \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \right)^n \exp \left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ \frac{(X_i - a_1)^2}{\sigma_1^2} - \frac{2\rho(X_i - a_1)(Y_i - a_2)}{\sigma_1\sigma_2} + \frac{(Y_i - a_2)^2}{\sigma_2^2} \right] \right\}, \\ \ln L = -n \ln(2\pi) - \frac{n}{2} \ln \sigma_1^2 - \frac{n}{2} \ln \sigma_2^2 - \frac{n}{2} \ln(1-\rho^2) \\ - \frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ \frac{(X_i - a_1)^2}{\sigma_1^2} - \frac{2\rho(X_i - a_1)(Y_i - a_2)}{\sigma_1\sigma_2} + \frac{(Y_i - a_2)^2}{\sigma_2^2} \right],$$

donc

$$\frac{\partial L}{\partial a_1} = -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ -\frac{2(X_i - a_1)}{\sigma_1^2} + \frac{2\rho(Y_i - a_2)}{\sigma_1\sigma_2} \right] = 0, \\ \frac{\partial L}{\partial a_2} = -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ -\frac{2(Y_i - a_2)}{\sigma_2^2} + \frac{2\rho(X_i - a_1)}{\sigma_1\sigma_2} \right] = 0.$$

Ces équations impliquent

$$\sum_{i=1}^n (Y_i - a_2) = \sum_{i=1}^n (X_i - a_1) \frac{\sigma_2}{\sigma_1\rho}, \\ \sum_{i=1}^n (Y_i - a_2) = \sum_{i=1}^n (X_i - a_1) \frac{\rho\sigma_2}{\sigma_1},$$

donc

$$\sum_{i=1}^n (X_i - a_1) \left( \frac{\sigma_2}{\sigma_1\rho} - \frac{\rho\sigma_2}{\sigma_1} \right) = 0,$$

qui implique  $\hat{a}_1 = \bar{X}$ ,  $\hat{a}_2 = \bar{Y}$ .

En dérivant par rapport à  $\sigma_i^2$ , on a

$$\frac{\partial L}{\partial \sigma_1^2} = -\frac{n}{2\sigma_1^2} + \frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ \frac{(X_i - a_1)^2}{\sigma_1^4} - \frac{\rho(X_i - a_1)(Y_i - a_2)}{\sigma_2\sigma_1^3} \right] = 0, \\ \frac{\partial L}{\partial \sigma_2^2} = -\frac{n}{2\sigma_2^2} + \frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ \frac{(Y_i - a_2)^2}{\sigma_2^4} - \frac{\rho(X_i - a_1)(Y_i - a_2)}{\sigma_1\sigma_2^3} \right] = 0.$$

Notons

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{s_1 s_2}$$

le coefficient empirique de corrélation. Alors les dernières équations on s'écrivent (on remplace  $a_i$  par  $\hat{a}_i$ ) :

$$(1 - \rho^2) - \frac{s_1^2}{\sigma_1^2} + \rho r \frac{s_1 s_2}{\sigma_1 \sigma_2} = 0, \\ (1 - \rho^2) - \frac{s_2^2}{\sigma_2^2} + \rho r \frac{s_1 s_2}{\sigma_1 \sigma_2} = 0,$$

qui impliquent  $s_1/\sigma_1 = s_2/\sigma_2 = (1 - \rho^2)/(1 - \rho r)$ . La dernière dérivé est

$$\frac{\partial L}{\partial \rho} = \frac{n\rho}{1 - \rho^2} - \frac{\rho}{(1 - \rho^2)^2} \sum_{i=1}^n \left[ \frac{(X_i - a_1)^2}{\sigma_1^2} - \frac{2\rho(X_i - a_1)(Y_i - a_2)}{\sigma_1\sigma_2} + \frac{(Y_i - a_2)^2}{\sigma_2^2} \right] + \frac{1}{1 - \rho^2} \sum_{i=1}^n \frac{(X_i - a_1)(Y_i - a_2)}{\sigma_1\sigma_2} = 0,$$

donc

$$\rho - \frac{\rho}{1 - \rho^2} \left( \frac{s_1^2}{\sigma_1^2} - 2\rho r \frac{s_1}{\sigma_1} \frac{s_2}{\sigma_2} + \frac{s_2^2}{\sigma_2^2} \right) + r \frac{s_1}{\sigma_1} \frac{s_2}{\sigma_2} = 0.$$

Remplaçant  $s_i^2/\sigma_i^2$  par  $(1 - \rho^2)/(1 - \rho r)$  dans la dernière équation, on a  $\hat{\rho} = r$ . Donc  $s_i^2/\hat{\sigma}_i^2 = 1$ , qui donne  $\hat{\sigma}_i^2 = s_i^2$ .

Nous avons obtenu les estimateurs  $\hat{a}_1 = \bar{X}$ ,  $\hat{a}_2 = \bar{Y}$ ,  $\hat{\sigma}_i^2 = s_i^2$ ,  $\hat{\rho} = r$ .

Sous  $H_0$  les estimateurs sont obtenus dans le problème 1.

Les maximums des fonctions  $L$  et  $L_1$  sont

$$\hat{L} = L(\hat{a}_1, \hat{a}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}) = \frac{1}{(2\pi)^n s_1^n s_2^n (1 - r^2)^{n/2}} e^{-n},$$

$$\hat{L}_1 = L_1(\hat{a}_1, \hat{a}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2) = \frac{1}{(2\pi)^n s_1^n s_2^n} e^{-n}$$

La région critique est  $L_1/L < c$ , qui est équivalent à  $r^2 > C$ .

On peut remarquer, que dans la régression logistique simple :  $Y_i = \beta_0 + \beta_1 X_i$  les estimateurs de coefficients sont

$$\hat{\beta}_1 = r \frac{s_2}{s_1}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Sous l'hypothèse  $H : \beta_1 = 0$  la v.a.

$$F = \frac{SS_E}{SS_R/(n-2)}$$

suit la loi de Fisher de 1 et  $ne2$  degrés de liberté. Notons que

$$SS_E = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2,$$

où  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \bar{Y} + r \frac{s_2}{s_1} (X_i - \bar{X})$ , donc

$$SS_E = nr^2 s_2^2, \quad SS_R = SS_T - SS_E = \sum_{i=1}^n (Y_i - \bar{Y})^2 - nr^2 s_2^2 = ns_2^2 (1 - r^2),$$

et

$$F = \frac{r^2(n-2)}{1-r^2}.$$

On a  $r^2 > C$  si et seulement si  $F > C_1$ . On rejette  $H_0$  si  $F > F_{1-\alpha}(1, n-2)$ . Le niveau de signification est  $\alpha$ .

**Exercice 1.** Il est donné que entre  $x$  et  $y(x)$ , deux variables en étude, il y a une dépendance polynomiale

$$y(x) = a_0 + a_1 x + \dots + a_m x^m. \quad (30)$$

On suppose que les  $a_i$  et  $m$  sont inconnus et que pour tout  $x_i$  la quantité  $y_i = y(x_i)$  est mesurée avec une erreur  $e_i$ . On observe donc

$$Y_i = y_i + e_i, \quad i = 1, \dots, n. \quad (31)$$

On suppose aussi que le nombre de mesures  $n > m$ ,  $\mathbf{e} = (e_1, \dots, e_n)^T$  est un échantillon normale, c'est-à-dire  $e_i \sim N_1(0, \sigma^2)$  et donc  $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ .

Dans l'expérience on a reçu pour

$$x_i = h(i-1), \quad i = 1, \dots, 30, \quad \text{i.e.} \quad x_i - x_{i-1} = x_{i+1} - x_i = h = 0.1.$$

les 30 observations suivantes :

$$\begin{array}{lll} Y_1 = 1.911 & Y_{11} = 1.001 & Y_{21} = -1.756 \\ Y_2 = 1.970 & Y_{12} = 0.7129 & Y_{22} = -1.926 \\ Y_3 = 2.022 & Y_{13} = 0.4502 & Y_{23} = -2.001 \\ Y_4 = 1.990 & Y_{14} = 0.1543 & Y_{24} = -1.974 \\ Y_5 = 1.952 & Y_{15} = -0.1462 & Y_{25} = -1.875 \\ Y_6 = 1.881 & Y_{16} = -0.4793 & Y_{26} = -1.620 \\ Y_7 = 1.765 & Y_{17} = -0.7702 & Y_{27} = -1.256 \\ Y_8 = 1.636 & Y_{18} = -1.080 & Y_{28} = -0.7477 \\ Y_9 = 1.448 & Y_{19} = -1.342 & Y_{29} = -0.0425 \\ Y_{10} = 1.227 & Y_{20} = -1.578 & Y_{30} = 0.852 \end{array}$$

On suppose qu'il y a seulement 0 ou 1 observation aberrante et que  $m \leq 6$ .

1. Eliminer l'observation aberrante si elle existe parmi les  $Y_j$ .
2. Estimer la degré du polynôme dans le modèle et construire les estimateurs pour  $a_j$  et  $\sigma^2$  par la méthode des moindres carrés.
3. Construire la zone de confiance pour  $y(x)$ .

**Exercice 2.** Soit  $\mathbf{A} = \|a_{ij}\|_{n \times n} = \mathbf{A}_n$ ,  $\det \mathbf{A} \neq 0$ . Notons  $\mathbf{x} = (x_1, \dots, x_n)^T$  un vecteur arbitraire de  $\mathbf{R}^n$ . Montrer que pour tout  $k \in \mathbf{R}^1$

$$\frac{\det(\mathbf{A} + k\mathbf{x}\mathbf{x}^T)}{\det \mathbf{A}} = 1 + k\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x}.$$

**Exercice 3.** Soit  $\mathbf{A}$  une matrice nondégénérée,  $\mathbf{A} = \|a_{ij}\|_{n \times n}$ ,  $\mathbf{X} = \|x_{ij}\|_{n \times p}$ ,  $k \in \mathbf{R}^1$ , telles que

$$\mathbf{I}_p + k\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X}$$

est nondégénérée. Montrer que

$$(\mathbf{A} + k\mathbf{X}\mathbf{X}^T)^{-1} = \mathbf{A}^{-1} - k\mathbf{A}^{-1} \mathbf{X} (\mathbf{I}_p + k\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1}.$$

**Exercice 4.** Montrer que

$$\text{Tr}(\mathbf{A}_n + \mathbf{B}_n) = \text{Tr}(\mathbf{A}_n) + \text{Tr}(\mathbf{B}_n) \quad \text{et} \quad \text{Tr}(\mathbf{A}_n \cdot \mathbf{B}_n) = \text{Tr}(\mathbf{B}_n \cdot \mathbf{A}_n),$$

où  $\mathbf{B}_n = \|b_{ij}\|_{n \times n}$  et  $\text{Tr}(\mathbf{A}_n) = \sum_{i=1}^n a_{ii}$ .

**Exercice 5.** Montrer que

$$\det(\mathbf{A}_n \mathbf{B}_n) = \det \mathbf{A}_n \cdot \det \mathbf{B}_n.$$

**Exercice 6.** Soit  $\mathbf{A}$  une matrice symétrique,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  - les valeurs propres de  $\mathbf{A}_n$ .

Montrer que

$$\text{Tr}\mathbf{A}_n = \sum_{i=1}^n \lambda_i, \quad \det \mathbf{A}_n = \prod_{i=1}^n \lambda_i.$$

### 5.3 Régression logistique

On suppose que l'on observe un événement  $A$  et que la probabilité de cet événement dépend de la valeur d'un vecteur de variables explicatives (covariables)  $x_1, \dots, x_m$ . Notons  $x = (x_0, x_1, \dots, x_m)^T$ ,  $x_0 = 1$ ,

$$\pi(x) = \mathbf{P}\{A|x\}$$

et considérons la variable aléatoire  $Y$  qui prend deux valeurs : 0 et 1 et telle que

$$\mathbf{P}\{Y = 1|x\} = \pi(x).$$

On effectue  $n$  expériences indépendantes. La  $i$ -ème expérience est observée sous la covariable  $x^{(i)} = (x_{i0}, \dots, x_{im})^T$ ,  $x_{i0} = 1$ . On fixe les valeurs des variables aléatoires  $Y_i : Y_i = 1$ , si un événement  $A$  se produit pendant la  $i$ -ème expérience et  $Y_i = 0$  sinon. Donc on a un échantillon

$$(Y_1, x^{(1)}), \dots, (Y_n, x^{(n)}).$$

Les variables aléatoires  $Y_i$  suivent la loi de Bernoulli :  $Y_i|x^{(i)} \sim \text{Bi}(1, \pi(x^{(i)}))$ .

*Le but est d'estimer la probabilité  $\pi(x)$  pour tous les  $x \in \mathcal{E}$ , où  $\mathcal{E}$  est un ensemble de covariables. Si  $x \neq x^{(i)}$  et la forme de  $\pi(x)$  est complètement inconnue, l'estimation de  $\pi(x)$  sera impossible.*

On peut considérer le modèle linéaire

$$\pi(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m, \tag{1}$$

mais après estimation des paramètres  $\beta = (\beta_0, \dots, \beta_m)$  on peut obtenir un estimateur de  $\pi(x)$  qui n'appartient pas à l'intervalle  $[0, 1]$ . En règle générale la loi des estimateurs de maximum de vraisemblance des paramètres inconnus approche la loi normale quand  $n$  est grand, mais la vitesse de convergence vers la loi normale est plus grande quand la région des valeurs du paramètre est  $\mathbb{R}$ . C'est le deuxième argument défavorable à l'utilisation du modèle (1). Toutes ces restrictions peuvent être éliminées en considérant le modèle

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m = \beta^T x. \tag{2}$$

Alors

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}} = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}}.$$

Le domaine de variation de la *fonction-logit*  $g(x)$  est  $\mathbb{R}$  et pour n'importe quelle valeur de  $\beta$  la fonction  $\pi(x)$  prend ses valeurs dans  $]0, 1[$ . Donc on a

*Le modèle de régression logistique :*

$$Y_i \sim B(1, \pi(x^{(i)})), \quad \text{où} \quad \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m$$

et  $Y_1, \dots, Y_n$  sont des variables aléatoires indépendantes.

Si la  $j$ -ème variable explicative  $x_j$  est discrète avec une échelle nominale, par exemple, la couleur, l'ethnie, etc., et prend  $k_j$  valeurs différentes, on peut utiliser au lieu de  $x_j$  le vecteur  $(z_{j,1}, \dots, z_{j,k_j-1})$  des codes qui prend  $k_j$  valeurs différentes :  $(0, \dots, 0)$ ,  $(1, 0, \dots, 0)$ ,  $(0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$  et le modèle (2) est modifié :

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \sum_{i=1}^{k_j-1} \beta_{ji} z_{ji} + \dots + \beta_m x_m. \quad (3)$$

Si, par exemple,  $x_j$  est la couleur qui prend 3 valeurs (noir, bleu, blanc), alors on considère le vecteur  $(z_{j1}, z_{j2})$  qui prend les valeurs  $(0,0)$ - (noir),  $(1,0)$ - (bleu),  $(0,1)$ - (blanc). Si  $x_j$  est le sexe (masculin, féminin), alors on considère la variable codée  $z_{j1}$  qui prend les valeurs 0 (masculin) et 1 (féminin). Parfois le codage est différent :  $(-1, \dots, -1)$ ,  $(1, 0, \dots, 0), \dots, (0, 0, \dots, 1)$ , etc.

Notons que si on prend deux valeurs  $x_j^{(1)}$  et  $x_j^{(2)}$  de  $x_j$  dans (2), alors

$$g(x_1, \dots, x_j^{(2)}, \dots, x_m) - g(x_1, \dots, x_j^{(1)}, \dots, x_m) = \beta_j (x_j^{(2)} - x_j^{(1)})$$

et donc

$$\frac{\pi_j^{(2)} / (1 - \pi_j^{(2)})}{\pi_j^{(1)} / (1 - \pi_j^{(1)})} = \frac{\pi(x_1, \dots, x_j^{(2)}, \dots, x_m) / (1 - \pi(x_1, \dots, x_j^{(2)}, \dots, x_m))}{\pi(x_1, \dots, x_j^{(1)}, \dots, x_m) / (1 - \pi(x_1, \dots, x_j^{(1)}, \dots, x_m))} = e^{\beta_j (x_j^{(2)} - x_j^{(1)})}. \quad (4)$$

Le rapport des cotes est donc égal à  $e^{\beta_j (x_j^{(2)} - x_j^{(1)})}$  et si  $x_j^{(2)} - x_j^{(1)} = 1$ , alors il vaut  $e^{\beta_j}$ . La cote est le rapport des probabilités de succès et d'échec pour l'évènement  $A$ . Le rapport des cotes montre comment varie la cote quand  $x_j$  passe de  $x_j^{(1)}$  à  $x_j^{(2)}$ , toutes les autres covariables restant les mêmes. Si les probabilités  $\pi_j^{(i)}$  sont petites, alors le rapport des cotes est proche à  $\pi_j^{(2)} / \pi_j^{(1)}$ , i.e. au risque relatif. Dans ce cas  $e^{\beta_j (x_j^{(2)} - x_j^{(1)})}$  montre comment change la probabilité de succès quand  $x_j$  change sa valeur de  $x_j^{(1)}$  à  $x_j^{(2)}$  et toutes les autres covariables ne changent pas. Il faut souligner que dans le modèle (2) le rapport des cotes est le même pour n'importe quelles valeurs fixées des autres covariables  $x_l$  ( $l \neq j$ ), i.e. il n'y a pas d'interactions.

Si  $x_j$  est discrète avec une échelle nominale et  $(z_{j1}^{(i)}, \dots, z_{j,k_j-1}^{(i)}) = (0, \dots, 1, \dots, 0)$ , où 1 est dans la  $i$ -ème place,  $(z_{j1}^{(0)}, \dots, z_{j,k_j-1}^{(0)}) = (0, \dots, 0)$ , alors

$$g(x_1, \dots, z_{j1}^{(i)}, \dots, z_{j,k_j-1}^{(i)}, \dots, x_m) - g(x_1, \dots, z_{j1}^{(0)}, \dots, z_{j,k_j-1}^{(0)}, \dots, x_m) = \beta_{ji}$$

et alors

$$\frac{\pi_j^{(i)} / (1 - \pi_j^{(i)})}{\pi_j^{(0)} / (1 - \pi_j^{(0)})} = e^{\beta_{ji}}, \quad (5)$$

où  $\pi_j^{(l)} = \pi(x_1, \dots, z_{j1}^{(l)}, \dots, z_{j,k_j-1}^{(l)}, \dots, x_m)$  ( $l = 0, i$ ).

$e^{\beta_{ji}}$  est le rapport des cotes qui correspond au changement de valeur de la variable  $x_j$  de la première à la  $(i+1)$ -ème quand toutes les autres variables restent fixes. Par exemple, si  $x_j$  est la couleur (noire, blanche, bleue),  $e^{\beta_{j2}}$  exprime le rapport des cotes qui correspond au changement de  $x_j$  de la couleur noire à la couleur blanche ( $i = 1$ ).

Si l'effet de changement de la valeur de la covariable  $x_j$  est différent pour des valeurs différentes des autres covariables, on a une interaction entre  $x_j$  et ces covariables. Alors le modèle (2) peut être modifié pour tenir compte de l'effet d'interaction. Par exemple, dans le cas de deux covariables on a le modèle

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2,$$

dans le cas de trois covariables

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3. \quad (6)$$

S'il y a interaction, alors, par exemple ( $m = 2$ ),

$$g(x_1^{(2)}, x_2) - g(x_1^{(1)}, x_2) = (\beta_1 + \beta_3 x_2)(x_1^{(2)} - x_1^{(1)})$$

et

$$e^{(\beta_1 + \beta_3 x_2)(x_1^{(2)} - x_1^{(1)})} = \frac{\pi(x_1^{(2)}, x_2) / (1 - \pi(x_1^{(2)}, x_2))}{\pi(x_1^{(1)}, x_2) / (1 - \pi(x_1^{(1)}, x_2))}, \quad (7)$$

donc le rapport des cotes dépend non seulement de la différence  $x_1^{(2)} - x_1^{(1)}$  mais aussi de la valeur de la deuxième covariable  $x_2$ .

### 5.3.1 Estimation

On a un échantillon  $(Y_1, x^{(1)}), \dots, (Y_n, x^{(n)})$ , où  $x^{(i)} = (x_{i0}, \dots, x_{im})^T$ ,  $x_{i0} = 1$ . La variable aléatoire  $Y_i$  suit la loi de Bernoulli :

$$Y_i | x^{(i)} \sim B(1, \pi(x^{(i)})).$$

La fonction de vraisemblance

$$L(\beta) = \prod_{i=1}^n [\pi(x^{(i)})]^{Y_i} [1 - \pi(x^{(i)})]^{1-Y_i}$$

et

$$\begin{aligned} \ln L(\beta) &= \sum_{i=1}^n Y_i \ln \pi(x^{(i)}) + (1 - Y_i) \ln (1 - \pi(x^{(i)})) = \\ &= \sum_{i=1}^n Y_i \ln \frac{\pi(x^{(i)})}{1 - \pi(x^{(i)})} + \ln (1 - \pi(x^{(i)})) = \\ &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) - \ln (1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}}). \end{aligned}$$

Les fonctions score

$$U_l(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta_l} = \sum_{i=1}^n x_{il} [Y_i - \pi(x^{(i)})] \quad (l = 0, \dots, m).$$

Notons  $\hat{\beta}$  l'estimateur du maximum de vraisemblance. Il vérifie le système d'équations  $U_l(\beta) = 0$  ( $l = 0, \dots, m$ ). Alors la probabilité de l'événement  $A$  sous la covariable  $x = (1, x_1, \dots, x_m)^T$  est estimée par

$$\hat{\pi}(x) = \frac{e^{\hat{\beta}^T x}}{1 + e^{\hat{\beta}^T x}}.$$

Quelles sont les notions analogues aux notions de coefficient de détermination  $R^2$ , de sommes des carrés : totale  $SS_T$ , de régression  $SS_R$ , d'erreur  $SS_E$ , considérées dans le chapitre "régression linéaire" ?

Notons

$$\hat{Y}_i = \hat{\pi}(x^{(i)}) = \frac{e^{\hat{\beta} x^{(i)}}}{1 + e^{\hat{\beta}^T x^{(i)}}}$$

les valeurs prédites des  $\pi(x^{(i)})$ . Dans le cas de la régression linéaire  $\hat{Y}_i = \hat{\beta}^T x^{(i)}$ . La prédiction est bonne si les valeurs observées  $Y_i$  et les valeurs prédites  $\hat{Y}_i$  sont proches. Dans le cas de la régression linéaire la différence entre  $Y_i$  et  $\hat{Y}_i$  était déterminé par  $SS_R = \sum (Y_i - \hat{Y}_i)^2$ . Dans le cas normal  $SS_R/\sigma^2$  suit la loi du chi-deux à  $n - m - 1$  degrés de liberté..

Si la fonction  $\pi(x)$  est complètement inconnue et si

$$\mathbf{P}\{Y_i = 1 | x^{(i)}\} = \pi(x^{(i)}) = p_i,$$

on estime  $n$  paramètres inconnus  $p_1, \dots, p_n$ . On a le modèle saturé, parce que le nombre des paramètres à estimer est le même que la taille de l'échantillon.

La fonction de vraisemblance

$$L_0(p) = L_0(p_1, \dots, p_n) = \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1 - Y_i}$$

est maximisée au point  $\hat{p} = (\hat{p}_1, \dots, \hat{p}_n)$ , où  $\hat{p}_i = Y_i$ , donc

$$L_0(\hat{p}) = \prod_{i=1}^n Y_i^{Y_i} (1 - Y_i)^{1 - Y_i} = 1.$$

On suppose  $0^0 = 1$ . Si on considère le modèle (2) avec  $m + 1 < n$ , on a  $(m + 1)$  paramètres inconnus  $\beta_0, \dots, \beta_m$ . Le maximum de la fonction de vraisemblance

$$L(\beta) = \prod_{i=1}^n \pi(x^{(i)})^{Y_i} (1 - \pi(x^{(i)}))^{1 - Y_i}$$

est

$$L(\hat{\beta}) = \prod_{i=1}^n \hat{Y}_i^{Y_i} (1 - \hat{Y}_i)^{1 - Y_i} \leq L_0(\hat{p}).$$

Si  $\hat{Y}_i$  et  $Y_i$  sont proches, i.e. la prédiction est bonne, alors  $L_0(\hat{p})$  et  $L(\hat{\beta})$  sont proches, donc le rapport des vraisemblances  $L(\hat{\beta})/L_0(\hat{p})$  est proche de 1 et

$$D_R = -2 \ln \frac{L(\hat{\beta})}{L_0(\hat{p})} = -2 \ln L(\hat{\beta})$$

est proche de zero. Si  $n$  est grand et le modèle de régression logistique est vérifié, la loi de  $D_R$  est approchée par la loi du chi-deux à  $n - m - 1$  degrés de liberté. Donc un équivalent de la somme des carrés de régression  $SS_R$  dans la régression logistique est  $D_R$ .

Le nombre minimal de paramètres à estimer est égal à 1. On est dans ce cas, si

$$\beta_1 = \dots = \beta_m = 0 \quad \text{et} \quad \pi(x^{(i)}) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} = \pi = \text{const.}$$

Alors la fonction de vraisemblance

$$L_1(\pi) = \prod_{i=1}^n \pi^{Y_i} (1 - \pi)^{1 - Y_i}$$

est maximisée au point  $\hat{\pi} = \hat{Y} = \frac{1}{n} \sum Y_i$  et

$$L_1(\hat{\pi}) = \prod_{i=1}^n \bar{Y}^{Y_i} (1 - \bar{Y})^{1 - Y_i} \leq L(\hat{\beta}) \leq L_0(\hat{p}).$$

La loi de la variable aléatoire

$$D_T = -2 \ln \frac{L_1(\hat{\pi})}{L_0(\hat{p})} = -2 \ln L_1(\hat{\pi})$$

est proche de la loi du chi-deux à  $n - 1$  degrés de liberté. Donc un équivalent de la somme totale des carrés  $SS_T$  dans la régression logistique est  $D_T$ .

La loi de la variable aléatoire

$$D_E = -2 \ln \frac{L_1(\hat{\pi})}{L(\hat{\beta})} =$$

$$2 \left[ \sum_{i=1}^n Y_i \ln \hat{Y}_i + \sum_{i=1}^n (1 - Y_i) \ln (1 - \hat{Y}_i) - \sum_{i=1}^n Y_i \ln \bar{Y} - (n - \sum_{i=1}^n Y_i) \ln (1 - \bar{Y}) \right]$$

est proche de la loi du chi-deux à  $m$  degrés de liberté si  $\beta_1 = \dots = \beta_m = 0$  et  $n$  est grand. La variable aléatoire  $D_E$  est un équivalent de la somme des carrés d'erreur  $SS_E$ . On a

$$D_T = D_E + D_R.$$

L'équivalent du coefficient de détermination dans le cas de la régression logistique

$$R^2 = 1 - \frac{D_R}{D_T} = \frac{D_E}{D_T}.$$

Si  $\hat{Y}_i = Y_i$ , alors  $R^2 = 1$ . Si  $\bar{Y} \neq 0$  et  $\bar{Y} \neq 1$ ,  $\hat{Y}_i = \bar{Y}$ , alors  $R^2 = 0$ .

Considérons l'hypothèse

$$H_0 : \beta_1 = \dots = \beta_m = 0.$$

Cette hypothèse signifie qu'il n'y a pas de régression et la connaissance de la valeur de  $x$  n'améliore pas la prédiction de  $\pi(x)$ . L'hypothèse  $H_0$  peut être écrite comme  $H_0 : \pi(x) = \pi = \text{const}$ . Sous l'hypothèse  $H_0$  la loi de  $D_E$  est approchée par la loi du chi-deux à  $m$  degrés de liberté. L'hypothèse  $H_0$  est rejetée avec le niveau de signification  $\alpha$ , si  $D_E > \chi_{1-\alpha}^2(m)$ .

Considérons l'hypothèse

$$H_0 : \beta_{j_1} = \dots = \beta_{j_l} = 0, \quad (1 \leq j_1 < \dots < j_l \leq m, l < m).$$

Notons  $D_E^{(m)}$  et  $D_E^{(m-l)}$  la statistique  $D_E$  pour le modèle (2) avec tous  $\beta_0, \dots, \beta_m$  et sans  $\beta_{j_1}, \dots, \beta_{j_l}$ , respectivement. Sous l'hypothèse  $H_0$  la loi de la variable aléatoire  $D_E^{(m)} - D_E^{(m-l)}$  peut être approchée par la loi du chi-deux à  $k = m - (m - l)$  degrés de liberté.

On rejette l'hypothèse  $H_0$  avec le niveau de signification  $\alpha$ , si

$$D_E^{(m)} - D_E^{(m-l)} > \chi_{1-\alpha}^2(k).$$

En particulier ce test peut être appliqué pour tester l'hypothèse d'absence d'interactions entre des covariables. Par exemple, dans le modèle (6) cette hypothèse est équivalente à l'hypothèse

$$H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0.$$

La statistique de test  $D_E^{(7)} - D_E^{(3)}$  suit la loi de chi-deux de  $k = 4$  degrés de liberté.

L'hypothèse

$$H_0 : \beta_j = 0 \quad (j = 1, \dots, m)$$

peut aussi être testée en utilisant la matrice d'information de Fisher estimée.

On cherche la matrice d'information de Fisher  $\mathbf{I}(\beta) = (I_{ls}(\beta))$ . On a

$$I_{ls}(\beta) = -\mathbf{E} \frac{\partial^2 \ln L(\beta)}{\partial \beta_l \partial \beta_s} = \sum_{i=1}^n x_{il} x_{is} \pi(x^{(i)}) (1 - \pi(x^{(i)})) \quad (l, s = 0, \dots, m).$$

Donc  $\mathbf{I}(\hat{\beta}) = X^T \mathbf{V} X$ , où

$$X = \begin{pmatrix} x_{10} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n0} & \dots & x_{nm} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \hat{\pi}(x^{(1)})(1 - \hat{\pi}(x^{(1)})) & \dots & 0 \\ 0 & \dots & 0 \\ 0 & \dots & \hat{\pi}(x^{(m)})(1 - \hat{\pi}(x^{(m)})) \end{pmatrix}.$$

Si  $n$  est grand, la matrice de covariance de  $\hat{\beta}$  est approchée par  $\mathbf{I}^{-1}(\beta)$  ou  $\mathbf{I}^{-1}(\hat{\beta}) = (\hat{\sigma}_{ls})$ . Si  $x = (x_0, \dots, x_m)$ , alors la variance de  $\hat{\pi}(x)$  est approchée par

$$\begin{aligned} \hat{\sigma}^2(\hat{\pi}(x)) &= \left( \frac{\partial \hat{\pi}(x)}{\partial \beta_s} \right)_{1 \times (m+1)}^T \mathbf{I}^{-1}(\hat{\beta}) \left( \frac{\partial \hat{\pi}(x)}{\partial \beta_s} \right)_{(m+1) \times 1}^T = \\ &= \hat{\pi}^2(x) (1 - \hat{\pi}(x))^2 \sum_{l=0}^m \sum_{s=0}^m x_l x_s \hat{\sigma}_{ls}. \end{aligned} \quad (8)$$

Notons

$$\hat{\sigma}_{ll} = \hat{\sigma}^2(\hat{\beta}_l), \quad \hat{\sigma}_{ls} = \mathbf{Cov}(\hat{\beta}_l, \hat{\beta}_s).$$

Si  $n$  est grand alors la loi de  $\hat{\beta}$  est approchée par la loi normale  $N(\beta, \mathbf{I}^{-1}(\beta))$  et  $\mathbf{I}^{-1}(\beta)$  est estimée par  $\mathbf{I}^{-1}(\hat{\beta})$ .

La loi de la statistique

$$W_j = \frac{\hat{\beta}_j}{\hat{\sigma}(\hat{\beta}_j)}$$

est approchée par la loi  $N(0, 1)$ , quand  $n$  est grand. On rejette l'hypothèse  $H_0 : \beta_j = 0$  avec le niveau de signification  $\alpha$ , si  $|W_j| > w_{1-\alpha/2}$ .

La loi de  $(\hat{\pi}(x) - \pi(x))/\hat{\sigma}(\hat{\pi}(x))$  est approchée par la loi normale standard, donc l'intervalle de confiance de niveau de confiance  $\gamma = 1 - \alpha$  pour la probabilité  $\pi(x)$  est approché par

$$\hat{\pi}(x) \pm z_{1-\alpha/2} \hat{\sigma}(\hat{\pi}(x)),$$

où  $\hat{\sigma}(\hat{\pi}(x))$  est donnée par la formule (8).

La relation entre les coefficients  $\beta_j$  et les rapports des cotes donne la possibilité de construire des intervalles de confiance pour les rapports de cotes.

L'intervalle de confiance de niveau de confiance  $\gamma = 1 - \alpha$  pour le coefficient  $\beta_j$  est donné par la formule  $\hat{\beta}_j \pm z_{1-\alpha/2} \hat{\sigma}(\hat{\beta}_j)$  parce que la loi de  $(\hat{\beta}_j - \beta_j)/\hat{\sigma}(\hat{\beta}_j)$  est approchée par la loi normale standard réduite. Donc les intervalles de confiance pour les rapports des cotes (4), (5) et (7) sont

$$\exp\{(x_j^{(2)} - x_j^{(1)})(\hat{\beta}_j \pm z_{1-\alpha/2} \hat{\sigma}(\hat{\beta}_j))\},$$

$$\exp\{\hat{\beta}_{ji} \pm z_{1-\alpha/2} \hat{\sigma}(\hat{\beta}_{ji})\}$$

et

$$\exp\{(x_1^{(2)} - x_1^{(1)})(\hat{\beta}_1 + \hat{\beta}_3 x_2 \pm z_{1-\alpha/2} \sqrt{\hat{\sigma}^2(\hat{\beta}_1) + 2x_2 \mathbf{Cov}(\hat{\beta}_1, \hat{\beta}_3) + x_2^2 \hat{\sigma}^2(\hat{\beta}_3)})\},$$

respectivement.

# Chapitre 6

## ELEMENTS D'ANALYSE DES DONNEES CENSUREES ET TRONQUEES.

### 6.1 Distribution de survie.

Dans ce chapitre nous allons étudier les notions principales et les modèles de base de l'analyse de survie et de la fiabilité et seulement quelques nouveaux modèles proposés et étudiés par Bagdonavičius et Nikulin en 1994-2000. Les modèles plus généraux et plus récents et leurs analyses statistiques on peut trouver, par exemple, dans Bagdonavičius & Nikulin (1994, 1995, 1996, 1997, 1998, 1999, 2000), Dreesbeke & Fichet et Tassi (1989), Bagdonavičius, Gerville-Réache, Nikoulina & Nikulin (2000), Charlabidis, Koutras and Balakrishnan (2000), Meeker and Escobar (1998), Limnios and Nikulin (2000), Ionescu and Limnios (1999) etc.

Dans ce paragraphe, nous allons définir les fonctions permettant de décrire une distribution de survie et présenter quelques modèles paramétriques.

Admettons qu'à la date  $t = 0$  un élément (un sujet ou un système) commence à fonctionner (à vivre) et qu'à la date  $t$  il se produise une panne (la mort, le décès).

La variable *durée de vie*  $X$ , délai entre la date d'origine et la date du décès (panne) est une variable aléatoire non négative,  $X \in [0, \infty[$ .

Soit

$$F(t) = \mathbf{P}\{X \leq t\}, \quad t \in \mathbf{R}_+^1. \quad (1)$$

Nous ne considérons ici que le cas où  $X$  est continue, c'est-à-dire *que la probabilité de décès (de panne) à chaque instant est infiniment petite*.

Dans ce cas la fonction de répartition  $F(t)$  de la variable  $X$  est donnée par l'intégrale

$$F(t) = \int_0^t f(x) dx,$$

où  $f(t)$  est la densité de probabilité de  $X$

$$f(t) = F'(t) = \lim_{h \rightarrow 0} \frac{\mathbf{P}\{t \leq X \leq t+h\}}{h}, \quad h > 0. \quad (2)$$

Donc,  $F(t)$  est la probabilité de décéder entre 0 et  $t$ , ou la probabilité *de défaillance* (de panne) au cours de l'intervalle  $[0, t]$ .

**Définition 1.** La fonction

$$S(t) = \bar{F}(t) = 1 - F(t), \quad t \geq 0,$$

s'appelle *la fonction de survie* ou *la fonction de fiabilité* (*fonction de séjour*).

On remarque que  $S(t) = \bar{F}(t)$  est la probabilité de bon fonctionnement continu durant  $[0, t]$  :

$$S(t) = \bar{F}(t) = \mathbf{P}\{X > t\} = \mathbf{P}\{X \geq t\}, \quad t \in \mathbf{R}_+^1, \quad (3)$$

ou la probabilité du fonctionnement sans défaillance de l'élément au cours du temps  $t$ . La fonction  $S(t)$  est monotone décroissante :

$$S(0) = 1 \quad \text{et} \quad S(t) \rightarrow 0, \quad \text{quand} \quad t \rightarrow \infty.$$

La plus importante caractéristique numérique de la durée de survie  $X$  est le *temps moyen de survie*  $\mathbf{E}X$ . (On suppose que  $\mathbf{E}X$  existe). Dans ce cas

$$\begin{aligned} \mathbf{E}X &= \int_0^\infty t dF(t) = - \int_0^\infty t d[1 - F(t)] = \\ &= - \lim_{t \rightarrow \infty} t[1 - F(t)] + \int_0^\infty [1 - F(t)] dt = \int_0^\infty S(t) dt, \end{aligned}$$

i.e., si  $\mathbf{E}X$  existe, alors

$$\mathbf{E}X = \int_0^\infty S(t) dt. \quad (4)$$

De même, on peut montrer que, si  $\mathbf{Var}X$  existe, alors

$$\mathbf{Var}X = 2 \int_0^\infty t S(t) dt - (\mathbf{E}X)^2. \quad (5)$$

En pratique pour estimer  $\mathbf{E}X$  on utilise la formule :

$$\mathbf{E}X = \int_0^\infty S(t) dt \approx \sum_{i=1}^k \int_{(t_{i-1}+t_i)/2}^{(t_i+t_{i+1})/2} S(t) dt \approx \sum_{i=1}^k \hat{S}(t_i) \frac{t_{i+1} - t_{i-1}}{2}, \quad (6)$$

où  $0 = t_0 < t_1 < \dots < t_k$ , et  $\hat{S}(t_i)$  est un estimateur statistique de  $S(t_i)$ . Considérons un exemple dans lequel nous soumettons à l'essai  $n_i$  éléments identiques dans les mêmes conditions au cours du temps  $t_i$ . Si à l'instant  $t_i$ , où les essais se terminent,  $v_i$  éléments fonctionnent encore, alors la statistique  $v_i/n_i$  peut-être considérée comme un estimateur de  $S(t_i)$ , puisque d'après la loi de grands nombres

$$\mathbf{P}\left\{\frac{v_i}{n_i} \rightarrow S(t_i), \quad n_i \rightarrow \infty\right\} = 1.$$

Dans ce cas

$$S(t_i) \approx \frac{v_i}{n_i} = \hat{S}(t_i)$$

et donc

$$\mathbf{E}X \approx \frac{1}{2} \sum_{i=1}^k \frac{v_i}{n_i} (t_{i+1} - t_{i-1}). \quad (7)$$

Souvent, s'il n'y a pas de censure, pour estimer  $\mathbf{E}X$  on utilise aussi la *moyenne arithmétique*

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j = \frac{1}{n} \sum_{j=1}^n X_{(j)},$$

où  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  sont les *statistiques d'ordre*, associées avec les durées de survies  $X_1, X_2, \dots, X_n$ .

**Remarque 1.** Soit  $\mathbb{F}_n(t)$  la fonction empirique,

$$\mathbf{E}\mathbb{F}_n(t) = F(t).$$

Dans ce cas  $S_n(t) = 1 - \mathbb{F}_n(t)$  est l'estimateur empirique de la fonction de survie  $S(t)$ ,

$$\mathbf{E}S_n(t) = S(t), \quad \mathbf{Var} S_n(t) = \mathbf{Var} \mathbb{F}_n(t) = \frac{1}{n} F(t)S(t).$$

Puisque

$$\mathbf{Var} \{\ln S_n(t)\} \approx \frac{\mathbf{Var} S_n(t)}{S^2(t)} = \frac{F(t)}{nS(t)},$$

nous pouvons dire que l'estimateur  $S_n(t)$  n'est pas fiable quand  $S(t)$  est trop petite.

D'autres *caractéristiques empiriques* qui donnent des informations intéressantes sur la loi  $F$  sont :

*la fonction empirique*

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{]-\infty, x]}(X_j),$$

*la variance empirique*

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

*la moyenne harmonique*

$$\bar{X}_n^H = \frac{n}{\sum_{j=1}^n \frac{1}{X_j}},$$

*la moyenne géométrique*

$$\bar{X}_n^G = \sqrt[n]{X_1 X_2 \dots X_n},$$

*l'étendu*

$$R = X_{(n)} - X_{(1)},$$

*le coefficient de variation*

$$v = \frac{s_n}{\bar{X}_n}.$$

**Définition 2.** Soient  $X$  et  $Y$  deux durées de survie,

$$S(t) = \mathbf{P}\{X > t\}, \quad H(t) = \mathbf{P}\{Y > t\}, \quad t \in \mathbf{R}_+^1.$$

Nous disons que  $X$  est *stochastiquement plus grande* que  $Y$  et notons  $X \succeq Y$  si

$$S(t) \geq H(t) \quad \text{pour tout } t \in \mathbf{R}_+^1. \quad (8)$$

Le fait que  $X$  est stochastiquement plus grande que  $Y$  nous pouvons exprimer aussi en disant que  $Y$  est *stochastiquement plus petite* que  $X$  et en notant  $Y \preceq X$ . Il est clair que si  $Y \preceq X$ , alors

$$F_Y(t) = \mathbf{P}\{Y \leq t\} \geq F_X(t) = \mathbf{P}\{X \leq t\}, \quad t \in \mathbf{R}_+^1, \quad (9)$$

i.e.,

$$S(t) \leq H(t) \quad \text{pour tout } t \in \mathbf{R}_+^1.$$

**Théorème 1.** Si  $X \succeq Y$ , alors

$$\mathbf{E}X \geq \mathbf{E}Y.$$

En effet, puisque  $S(t) \geq H(t)$  on a

$$\mathbf{E}X = \int_0^\infty S(t)dt \geq \int_0^\infty H(t)dt = \mathbf{E}Y.$$

## 6.2 Risque de panne ou taux de défaillance.

Considérons tout d'abord le problème suivant : supposons que l'élément ait fonctionné sans défaillance jusqu'à l'instant  $u$ ,  $u > 0$ . Quelle est la probabilité pour qu'il ne tombe pas en panne dans l'intervalle  $]u, u+t]$ ,  $t > 0$  ? Donc, on s'intéresse à la probabilité

$$S_u(t) = \mathbf{P}\{X > u+t | X > u\}, \quad u > 0, t > 0.$$

La probabilité cherchée est alors la probabilité conditionnelle et on a

$$\mathbf{P}\{X > u+t | X > u\} = \frac{\mathbf{P}\{X > u+t\}}{\mathbf{P}\{X > u\}} = \frac{S(u+t)}{S(u)} = S_u(t). \quad (6.1)$$

De (1) on tire immédiatement que pour tout  $\Delta t > 0$

$$S(t + \Delta t) = \mathbf{P}\{X > t + \Delta t\} = S(t)_{\Delta t}p_t, \quad (6.2)$$

où

$$_{\Delta t}p_t = \mathbf{P}\{X > t + \Delta t | X > t\}.$$

C'est une notation utilisée en démographie. De (1) et (2) il suit que la probabilité de panne (de décès) au cours de  $(t, t + \Delta t]$ , sachant que  $X > t$  est

$$_{\Delta t}q_t = \mathbf{P}\{t < X \leq t + \Delta t | X > t\} = 1 - _{\Delta t}p_t = \frac{S(t) - S(t + \Delta t)}{S(t)}. \quad (6.3)$$

**Définition 1.** On appelle *risque instantané de décès* ou *taux de défaillance* ou *risque de panne* la fonction

$$\alpha(t) = \frac{f(t)}{\bar{F}(t)} = \frac{f(t)}{S(t)}, \quad t \geq 0. \quad (6.4)$$

De la définition 1 il suit que

$$\begin{aligned}\alpha(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}\{t < X \leq t + \Delta t\}}{\Delta t S(t)} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}\{t < X \leq t + \Delta t\}}{\Delta t \mathbf{P}\{X > t\}} = \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}\{t < X \leq t + \Delta t | X > t\}}{\Delta t} =\end{aligned}\quad (6.5)$$

$$= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t} = -\frac{S'(t)}{S(t)}.\quad (6.6)$$

**Remarque 1.** Des formules (2) et (6) on tire que

$$\begin{aligned}\alpha(t) &= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t} = \\ &= -\frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} = -\lim_{\Delta t \rightarrow 0} \frac{\Delta t p_t - 1}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \Delta t q_t.\end{aligned}\quad (6.7)$$

De cette remarque on tire aussi que

$$\Delta t q_t = \mathbf{P}\{t < X \leq t + \Delta t | X > t\} \approx \alpha(t) \Delta t, \quad \Delta t \rightarrow 0,\quad (6.8)$$

tandis que

$$\mathbf{P}\{t < X \leq t + \Delta t\} = f(t) \Delta t + o(\Delta t), \quad \Delta t \rightarrow 0.$$

Donc  $\alpha(t) \Delta t$  est approximativement égale (pour de petites valeurs de  $\Delta t$ ) à la probabilité de tomber en panne au cours de  $(t, t + \Delta t]$  à condition que l'élément ait fonctionné sans défaillance jusqu'à la date  $t$ . On voit que  $\alpha(t)$  est *une caractéristique locale* de fiabilité déterminant la fiabilité de l'élément à chaque instant de temps, d'où le nom de taux instantané de défaillance. Puisque

$$f(t) dt = S(t) \alpha(t) dt$$

il est clair que

$$\alpha(t) \approx f(t)$$

pour les petites valeurs de  $t$ .

**Remarque 2.** De (6) il suit que

$$\alpha(t) = -\frac{d \ln S(t)}{dt}, \quad S(0) = 1,$$

d'où on tire que

$$\ln S(t) = -\int_0^t \alpha(s) ds, \quad t > 0,$$

et donc

$$S(t) = \exp \left\{ -\int_0^t \alpha(s) ds \right\}.\quad (6.9)$$

On voit que le taux de défaillance détermine la distribution  $F(t) = 1 - \bar{F}(t) = 1 - S(t)$ .

**Définition 2.** On définit  $A(t)$ , fonction de risque cumulée de  $\alpha(s)$  entre 0 et  $t$  :

$$A(t) = \int_0^t \alpha(s) ds, \quad t \geq 0. \quad (6.10)$$

La fonction  $A(t)$  est aussi appelée *fonction du hasard* ou simplement *hasard*.

De (9) il suit que

$$A(t) = -\ln S(t), \quad t \geq 0, \quad (6.11)$$

et de (4) on tire que

$$f(t) = \alpha(t)S(t) = \alpha(t)\exp\{-A(t)\}, \quad (6.12)$$

puisque

$$S(t) = \exp\{-A(t)\}, \quad t \geq 0. \quad (6.13)$$

On peut définir la distribution de probabilité de la durée de survie  $X$  à partir de l'une quelconque des fonctions :  $f(t)$ ,  $\alpha(t)$ ,  $S(t)$ ,  $A(t)$ .

La fonction de risque fournit la description la plus concrète d'une distribution de survie.

**Remarque 3.** La fonction de survie conditionnelle  $S_u(t) = S(u+t)/S(u)$  s'exprime facilement en termes de la fonction de défaillance  $\alpha(t)$ . En effet, pour tout  $u > 0$  on a

$$\begin{aligned} S_u(t) &= \mathbf{P}\{X > u+t \mid X > u\} = \\ &= \frac{\exp\{-A(u+t)\}}{\exp\{-A(u)\}} = \exp\left\{-\int_u^{u+t} \alpha(x) dx\right\}, \quad t \in \mathbf{R}_+^1. \end{aligned} \quad (14)$$

En faisant le changement des variables  $v = x - u$ , on en tire que

$$S_u(t) = \exp\left\{-\int_0^t \alpha(v+u) dv\right\}, \quad (s, t) \in \mathbf{R}_+^1 \times \mathbf{R}_+^1. \quad (15)$$

De (14) il suit que

$$\mathbf{P}\{u < X \leq t+u \mid X > u\} = 1 - \exp\left\{-\int_u^{u+t} \alpha(x) dx\right\}.$$

**Remarque 4.** La fonction  $\alpha(t)$  peut-être déterminée d'après les résultats des essais. Si  $N = N(0)$  éléments sont soumis aux essais au moment  $t = 0$  et  $N(t)$  désigne le nombre d'éléments qui fonctionnent encore au moment  $t$ ,  $t > 0$ , alors ( $N \rightarrow \infty$ )

$$\alpha(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \approx \frac{N(t) - N(t + \Delta t)}{\Delta t N(t)} = \frac{\Delta N(t)}{\Delta t N(t)} = \hat{\alpha}_N(t), \quad (16)$$

$\hat{\alpha}_N(t)$  est le *taux de défaillance empirique*.

En pratique cela signifie que si on partage l'intervalle  $[0, t]$  en segments

$$[0, t_1[, [t_1, t_2[, \dots, [t_{k-1}, t_k[, \dots$$

de longueur  $h = \Delta t$ , et  $\mu_k$  désigne le nombre de pannes au cours de  $[t_{k-1}, t_k[$ , où

$$t_{k-1} = (k-1)h, \quad t_k = kh, \quad [(k-1)h, kh[ \subset [0, t],$$

i.e.,

$$\mu_k = N(t_{k-1}) - N(t_k) = N((k-1)h) - N(kh) = [N - N(kh)] - [N - N((k-1)h)],$$

dans ce cas le *taux de défaillance empirique*  $\hat{\alpha}_N(t)$  est donné par la formule :

$$\hat{\alpha}_N(t) = \frac{\mu_k}{(N - \mu_1 - \mu_2 - \dots - \mu_{k-1})h} = \frac{\mu_k}{hN(t_{k-1})}, \quad (k-1)h \leq t \leq kh. \quad (17)$$

**Remarque 5.** Soit  $X$  la durée de vie avec la fonction de répartition  $F(t)$  et la densité  $f_X(t) = F'(t)$ . Considérons la transformation  $Y = \sigma X + \mu$ ,  $\sigma > 0$ . Dans ce cas

$$\mathbf{P}\{Y \leq t\} = F\left(\frac{t-\mu}{\sigma}\right) \quad \text{et} \quad f_Y(t) = \frac{1}{\sigma} f_X\left(\frac{t-\mu}{\sigma}\right),$$

puisque

$$y = \sigma x + \mu \iff x = \frac{y-\mu}{\sigma}, \quad dx = \frac{dy}{\sigma},$$

d'où on tire que

$$\alpha_Y(t) = \frac{f_Y(t)}{S\left(\frac{t-\mu}{\sigma}\right)} = \frac{\frac{1}{\sigma} f_X\left(\frac{t-\mu}{\sigma}\right)}{S\left(\frac{t-\mu}{\sigma}\right)} = \frac{1}{\sigma} \alpha_X\left(\frac{t-\mu}{\sigma}\right).$$

**Remarque 6.** Dans le cas où  $X$  est une variable aléatoire discrète,

$$\mathbf{P}\{X = k\} = p_k, \quad k \in \mathbb{N} = \{1, 2, \dots\}$$

les fonctions de répartition  $F(k)$ , de survie  $S(k)$  et de risque de défaillance  $\alpha(k)$  de  $X$  sont données par les formules suivantes :

$$F(k) = \mathbf{P}\{X \leq k\} = \sum_{m \leq k} p_m, \quad (18)$$

$$S(k) = \mathbf{P}\{X > k\} = \mathbf{P}\{X \geq k+1\} = \sum_{m=k+1}^{\infty} p_m, \quad (19)$$

$$\begin{aligned} \alpha(k) &= \mathbf{P}\{X = k \mid X > k-1\} = \mathbf{P}\{X = k \mid X \geq k\} = \\ &= \frac{p_k}{\sum_{m=k}^{\infty} p_m} = \frac{p_k}{S(k-1)} \end{aligned} \quad (20)$$

pour tout  $k \in \mathbb{N}$  (on pose ici, que  $S(0) = 1$ ).

Comme

$$1 - \alpha(k) = \frac{S(k-1) - p_k}{S(k-1)} = \frac{S(k)}{S(k-1)}$$

on en tire que

$$S(k) = [1 - \alpha(k)]S(k-1) = \sum_{m=k+1}^{\infty} p_m = \prod_{m=1}^k [1 - \alpha(m)], \quad k \in \mathbb{N},$$

puisque

$$p_k = \alpha(k)S(k-1) = \alpha(k) \prod_{m=1}^{k-1} [1 - \alpha(m)], \quad k \in \mathbb{N},$$

en posant  $p_1 = \alpha(1)$ .

Enfin on remarque que

$$\begin{aligned} \mathbf{E}X &= \sum_{j=1}^{\infty} j p_j = \sum_{j=1}^{\infty} \sum_{k=1}^j p_k = \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} p_j = \\ &= \sum_{k=1}^{\infty} \mathbf{P}\{X \geq k\} = \sum_{k=1}^{\infty} \mathbf{P}\{X > k-1\} = \sum_{k=1}^{\infty} S(k-1). \end{aligned} \quad (21)$$

**Exemple 1.** Soit  $X$  est *uniforme* sur  $\mathcal{X} = \{1, \dots, N\}$ ,

$$p_k = \mathbf{P}\{X = k\} = \frac{1}{N}, \quad k \in \mathcal{X}.$$

Dans ce cas

$$\begin{aligned} F(k) &= \mathbf{P}\{X \leq k\} = \sum_{m=1}^k p_m = \frac{k}{N}, \quad k \in \mathcal{X}, \\ S(k) &= \mathbf{P}\{X > k\} = \mathbf{P}\{X \geq k+1\} = 1 - \frac{k}{n} = \frac{N-k}{N}, \\ \alpha(k) &= \frac{p_k}{S(k-1)} = \frac{1}{N-k+1}, \quad k \in \mathcal{X}. \end{aligned}$$

On remarque que

$$\frac{1}{N} = p_1 = \alpha(1) < \alpha(2) < \dots < \alpha(N) = 1.$$

**Exemple 2.** Soit  $X$  suit la loi *géométrique de paramètre*  $p$ ,  $p \in ]0, 1[$ . Dans ce cas

$$\begin{aligned} p_m &= \mathbf{P}\{X = m\} = p q^{m-1}, \quad m \in N, \\ F(k) &= \mathbf{P}\{X \leq k\} = p \sum_{m=1}^k q^{m-1} = 1 - q^k, \\ S(k) &= \mathbf{P}\{X > k\} = p \sum_{m=k+1}^{\infty} q^{m-1} = q^k \\ \alpha(k) &= \frac{p_k}{S(k-1)} = \frac{p q^{k-1}}{q^{k-1}} = p, \quad k \in N, \end{aligned}$$

d'où on tire que  $\alpha(k) = \text{const}$ .

Il est facile de démontrer que  $\alpha(k) = \text{const}$  si et seulement si la variable aléatoire discrète  $X$  suit une loi géométrique.

### 6.3 Modèles paramétriques de survie.

**Modèle exponentiel.** Le modèle de base est celui pour lequel la fonction de risque d'une variable aléatoire continue  $X$  est constante :

$$\alpha(t) = \lambda = \text{const}, \quad \lambda > 0, \quad t \in \mathbf{R}_+^1.$$

Dans ce modèle  $\lambda(t)$  est constante au cours du temps. On l'appelle modèle *exponentiel de paramètre*  $\lambda$  parce que la fonction de survie est exponentielle :

$$S(t) = S(t; \lambda) = \exp \left\{ - \int_0^t \alpha(s) ds \right\} = \exp \{-\lambda t\} = e^{-\lambda t}, \quad (1)$$

donc

$$F(t) = F(t; \lambda) = \mathbf{P}\{X \leq t\} = 1 - S(t) = 1 - e^{-\lambda t}, \quad t \geq 0. \quad (2)$$

Ce modèle ne dépend que du paramètre  $\lambda$  et on a

$$\mathbf{E}X = \frac{1}{\lambda} \quad \text{et} \quad \mathbf{Var}X = \frac{1}{\lambda^2}. \quad (3)$$

**Définition 1.** On dit que la variable de durée de survie  $X$  vérifie *la propriété d'indépendance temporelle (lack-of-memory)* si et seulement si

$$\alpha(t) = \lambda, \quad t > 0, \quad (4)$$

où  $\lambda = \text{const}$ ,  $\lambda > 0$ .

**Théorème 1.** Il y a indépendance temporelle si et seulement si la loi de la durée de survie  $X$  est exponentielle.

**Remarque 1.** La loi exponentielle est donc la seule loi continue à taux de défaillance constant.

**Théorème 2.** Il y a indépendance temporelle si et seulement si l'une des conditions suivantes est vérifiée :

1. les fonctions de survie conditionnelles  $\{S_u(t), u > 0, \}$  sont exponentielles de même paramètre  $\lambda$  ( $\lambda > 0$ ) :

$$S_u(t) = \frac{S(u+t)}{S(u)} = e^{-\lambda t}, \quad t \in \mathbf{R}_+^1$$

pour tout  $u \in \mathbf{R}_+^1$  ;

2.  $S(u+t) = S(t)S(u)$  pour tout  $t, u \in \mathbf{R}_+^1$ .

**Remarque 2.** La loi exponentielle est donc la seule loi continue possédant la propriété :

$$S_u(t) = \mathbf{P}\{X > t+u | X > u\} = \mathbf{P}\{X > t\}, \quad t \geq 0, \quad u > 0. \quad (5)$$

De cette relation il suit que pour tout  $u \in \mathbf{R}_+^1$

$$\mathbf{E}\{X | X > u\} = u + \int_0^\infty u e^{-\lambda u} du = u + \mathbf{E}X. \quad (6)$$

De l'autre côté on voit

$$\mathbf{E}\{X \mid X \leq t\} = \int_0^t \frac{\mathbf{P}\{t \geq X \geq u\}}{\mathbf{P}\{t \geq X\}} du = \int_0^t \frac{e^{-\lambda u} - e^{-\lambda t}}{1 - e^{-\lambda t}} du = \frac{1}{\lambda} - \frac{te^{-\lambda t}}{1 - e^{-\lambda t}}. \quad (7)$$

**Remarque 3.** Si  $h$  est petit ( $h \rightarrow 0$ ), alors

$$\mathbf{P}\{X \leq h\} = \lambda h + o(h).$$

L'interprétation de ce résultat est la suivant. Admettons qu'à la date  $t$ ,  $t > 0$ , l'élément fonctionne. Alors la probabilité de panne dans  $]t, t + h]$  vaut

$$\lambda h + o(h),$$

pour des petites valeurs de  $h$ ,  $h > 0$ .

**Modèle de Weibull.**

Soit

$$F(t) = F(t; \alpha, \lambda) = \mathbf{P}\{X \leq t\} = \left(1 - e^{-\lambda t^\alpha}\right) \mathbf{1}_{]0, \infty[}(t), \quad \lambda > 0, \alpha > 0, t \in \mathbf{R}^1, \quad (8)$$

i.e.,  $X$  suit une loi de Weibull  $W(\alpha, \lambda)$  de paramètres  $\alpha$  et  $\lambda$ . Dans ce modèle

$$S(t) = S(t; \alpha, \lambda) = e^{-\lambda t^\alpha} \mathbf{1}_{]0, \infty[}(t), \quad (9)$$

$$f(t) = f(t; \alpha, \lambda) = \alpha \lambda t^{\alpha-1} e^{-\lambda t^\alpha} \mathbf{1}_{]0, \infty[}(t). \quad (10)$$

On peut montrer que

$$\mathbf{E}X^k = \lambda^{-k/\alpha} \Gamma\left(\frac{k}{\alpha} + 1\right),$$

et par conséquent

$$\mathbf{E}X = \frac{1}{\lambda^{1/\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right), \quad \mathbf{E}X^2 = \frac{1}{\lambda^{2/\alpha}} \Gamma\left(\frac{2}{\alpha} + 1\right)$$

et donc

$$\mathbf{Var}X = \frac{1}{\lambda^{2/\alpha}} \Gamma\left(\frac{2}{\alpha} + 1\right) - \frac{1}{\lambda^{2/\alpha}} \Gamma^2\left(\frac{1}{\alpha} + 1\right).$$

On remarque que le coefficient de variation de  $X$  est

$$v = \frac{\sqrt{\mathbf{Var}X}}{\mathbf{E}X} = \sqrt{\frac{\Gamma\left(1 + \frac{2}{\alpha}\right)}{\Gamma^2\left(1 + \frac{1}{\alpha}\right)} - 1} = \frac{\pi}{\alpha\sqrt{6}} + O\left(\frac{1}{\alpha^2}\right), \quad \alpha \rightarrow \infty,$$

d'où on tire que la distribution de Weibull devient de plus en plus concentrée autour de  $\mathbf{E}X$ , quand  $\alpha$  devient de plus en plus grand. Il est évident que

$$\alpha(t) = \frac{f(t)}{S(t)} = \alpha \lambda t^{\alpha-1}. \quad (11)$$

Si  $\alpha > 1$ , le risque de panne  $\alpha(t)$  croit de façon monotone, à partir de 0, et on dit qu'il y a *usure*. Si  $\alpha < 1$ , le risque de panne  $\alpha(t)$  décroît de façon monotone et il n'est pas borné pour  $t = 0$ , dans ce cas on dit qu'il y a *rodage*. Si  $\alpha = 1$ , on obtient une loi exponentielle de

paramètre  $\lambda$ .

**Remarque 4.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon tel que

$$\mathbf{P}\{X_i \leq x\} = G(x; \alpha, \lambda) \mathbf{1}_{]0, \infty[}(x), \quad \alpha > 0, \quad \lambda > 0 \quad x \in \mathbb{R}^1, \quad (12)$$

où  $G(x; \alpha, \lambda)$  une fonction de répartition qui vérifie les conditions :

$$\lim_{x \downarrow 0} \frac{G(x; \alpha, \lambda)}{\lambda x^\alpha} = 1, \quad G(x; \alpha, \lambda) = 0, \quad x \leq 0,$$

pour tout  $\alpha$  et  $\lambda$  fixés.

Soit  $X_{(1)} = X_{(n1)} = \min(X_1, X_2, \dots, X_n)$ . Alors

$$n^{1/\alpha} X_{(n1)} \xrightarrow{\mathcal{L}} W(\alpha, \lambda), \quad \text{quand } n \rightarrow \infty.$$

En effet, pour tout  $x > 0$  on a

$$\mathbf{P}\{X_{(n1)} > x\} = [1 - G(x; \alpha, \lambda)]^n$$

et

$$\mathbf{P}\{n^{1/\alpha} X_{(n1)} > x\} = \left[ 1 - G\left(\frac{x}{n^{1/\alpha}}; \alpha, \lambda\right) \right]^n,$$

d'où on déduit que si  $n \rightarrow \infty$ , alors

$$\begin{aligned} \ln \mathbf{P}\{n^{1/\alpha} X_{(n1)} > x\} &= n \ln \left[ 1 - G\left(\frac{x}{n^{1/\alpha}}; \alpha, \lambda\right) \right] = \\ &= n \left[ -\lambda \left(\frac{x}{n^{1/\alpha}}\right)^\alpha + o\left(\frac{1}{n}\right) \right] = -\lambda x^\alpha + o(1), \end{aligned}$$

d'où on tire que pour tout  $x > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P}\{n^{1/\alpha} X_{(n1)} > x\} = e^{-\lambda x^\alpha} = S(x; \alpha, \lambda), \quad (13)$$

i.e. asymptotiquement ( $n \rightarrow \infty$ ) la statistique  $X_{(n1)}$  suit la loi de Weibull  $W(\alpha, \lambda)$  de paramètres  $\alpha$  et  $\lambda$ .

**Remarque 5.** Soit  $X \sim W(\alpha, \lambda)$ . Considérons la statistique  $Z = \ln X$ . On a

$$\begin{aligned} \mathbf{P}\{Z \leq z\} &= \mathbf{P}\{\ln X \leq z\} = \mathbf{P}\{X \leq e^z\} = 1 - \exp\{-(\lambda e^z)^\alpha\} \\ &= 1 - \exp\left\{-\exp\left[\alpha\left(z - \ln \frac{1}{\lambda}\right)\right]\right\} = 1 - \exp\left\{-\exp \frac{z - \mu}{\sigma}\right\}, \end{aligned} \quad (14)$$

où

$$\mu = \ln \frac{1}{\lambda} \quad \text{et} \quad \sigma = \frac{1}{\alpha} > 0.$$

**Modèle gamma.**

On suppose que la densité de  $X$  est

$$f(t) = f(t; \lambda, p) = \frac{\lambda^p}{\Gamma(p)} t^{p-1} e^{-\lambda t} \mathbf{1}_{]0, \infty[}(t), \quad \lambda > 0, \quad p > 0, \quad t \in \mathbb{R}^1. \quad (15)$$

Alors

$$\mathbf{E}X = \frac{p}{\lambda}, \quad \mathbf{Var}X = \frac{p}{\lambda^2},$$

et

$$\alpha(t) = \frac{f(t)}{S(t)} = \frac{t^{p-1}e^{-\lambda t}}{\int_t^\infty x^{p-1}e^{-\lambda x}dx}. \quad (16)$$

On peut montrer que si  $p > 1$ , alors  $\alpha(t)$  est croissante et

$$\lim_{t \rightarrow \infty} \alpha(t) = \lambda.$$

D'un autre côté, si  $0 < p < 1$ ,  $\alpha(t)$  est décroissante et

$$\lim_{t \rightarrow \infty} \alpha(t) = \lambda.$$

En effet,

$$\frac{1}{\alpha(t)} = \frac{\int_t^\infty x^{p-1}e^{-\lambda x}dx}{t^{p-1}e^{-\lambda t}} = \int_t^\infty \left(\frac{x}{t}\right)^{p-1} e^{-\lambda(x-t)}dx, \quad t > 0.$$

Après avoir fait le changement de variable dans l'intégrale :

$$u = x - t, \quad dx = du, \quad (u > 0),$$

on obtient que

$$\frac{1}{\alpha(t)} = \int_0^\infty \left(1 + \frac{u}{t}\right)^{p-1} e^{-\lambda u} du, \quad t > 0.$$

Posons

$$g_u(t) = \left(1 + \frac{u}{t}\right)^{p-1}, \quad t > 0,$$

pour tout  $u > 0$ .

Comme

$$\frac{dg_u(t)}{dt} = -(p-1)\frac{u}{t^2} \left(1 + \frac{u}{t}\right)^{p-2},$$

on en tire que pour tout  $u (u > 0)$

$$\frac{dg_u(t)}{dt} > 0, \quad \text{si } 0 < p < 1,$$

$$\frac{dg_u(t)}{dt} < 0, \quad \text{si } p > 1,$$

$$\frac{dg_u(t)}{dt} = 0, \quad \text{si } p = 1,$$

d'où on déduit que pour tout  $u > 0$

$g_u(t)$  est croissante, si  $0 < p < 1$ ,

$g_u(t) = 1$ , si  $p = 1$ ,

$g_u(t)$  est décroissante, si  $p > 1$ ,

et par conséquent on obtient que

$\alpha(t)$  est décroissante, si  $0 < p < 1$ ,  
 $\alpha(t) = \lambda$ , si  $p = 1$ , et donc on a la loi exponentielle de paramètre  $\lambda$ ,  
 $\alpha(t)$  est croissante, si  $p > 1$ .

Enfin, on remarque que pour tout  $p > 0$

$$\lim_{t \rightarrow \infty} \frac{1}{\alpha(t)} = \frac{1}{\lambda} \quad \text{et donc} \quad \lim_{t \rightarrow \infty} \alpha(t) = \lambda.$$

### Modèle de Rayleigh.

Dans ce modèle la fonction de survie d'une durée de survie  $X$  est

$$S(t) = S(t; c) = \exp\{-A(t)\} = \exp\left\{-\frac{ct^2}{2}\right\} \mathbf{1}_{[0, \infty[}(t), \quad c > 0, \quad (17)$$

$$f(t) = f(t; c) = -S'(t) = ct \exp\left\{-\frac{ct^2}{2}\right\} \mathbf{1}_{[0, \infty[}(t), \quad (18)$$

$$A(t) = \int_0^t \alpha(u) du = \int_0^t cu du = \frac{ct^2}{2}, \quad t \geq 0,$$

et par conséquent  $\alpha(t) = A'(t) = ct$ , et donc dans ce modèle  $\alpha(t)$  est une fonction linéaire. On peut montrer, que

$$\mathbf{E}X = \sqrt{\frac{\pi}{2c}}, \quad \mathbf{Var}X = \frac{4 - \pi}{2c}.$$

Il est clair que le modèle de Rayleigh représente le cas particulier du modèle de Weibull avec  $\alpha = 2$  et  $\lambda = c/2$ .

### Modèle de Pareto.

Dans ce modèle la fonction de survie est

$$S(t) = S(t; \alpha, \theta) = \left(\frac{\theta}{t}\right)^\alpha \mathbf{1}_{[\theta, +\infty[}(t), \quad t \in \mathbf{R}^1, \quad \theta > 0, \quad \alpha > 0, \quad (19)$$

$$f(t) = f(t; \alpha, \theta) = -S'(t) = \alpha \theta^\alpha \frac{1}{t^{\alpha+1}} \mathbf{1}_{[\theta, +\infty[}(t), \quad t \in \mathbf{R}^1, \quad (20)$$

par conséquent

$$\alpha(t) = \frac{\alpha}{t} \mathbf{1}_{[\theta, +\infty[}(t). \quad (21)$$

Il est évident que  $\alpha(t)$  est décroissante. On emploie ce modèle lorsqu'on est assuré que la survie dure au moins jusqu'à un instant  $\theta > 0$ .

### Modèles de Gompertz et de Makeham (taux de défaillance exponentiel).

Soit  $T$  est une durée de survie dont le taux de défaillance est

$$\alpha(t) = \alpha e^{\beta t} \mathbf{1}_{[0, \infty[}(t), \quad \alpha > 0, \quad \beta > 0. \quad (22)$$

Dans ce cas la densité de  $T$  est

$$f(t) = f(t; \alpha, \beta) = \alpha e^{\beta t} e^{-\alpha[e^{\beta t} - 1]/\beta} \quad (23)$$

et la fonction de survie est

$$S(t) = S(t; \alpha, \beta) = \exp\left\{\frac{\alpha}{\beta}(1 - e^{\beta t})\right\}. \quad (24)$$

Souvent on dit que  $T$  suit une *loi de Gompertz* ou *Makeham-Gompertz*. Parfois on considère

$$\alpha(t) = \gamma + \alpha e^{\beta t}, \quad (25)$$

où  $\gamma \geq 0$ , et dans ce cas on dit qu'il y a une loi de Makeham ou de Makeham-Gompertz. Récemment Gerville-Réache et Nikulin (2001) ont construit le test de type du chi-deux pour ce modèle.

### **Classe de Lehmann et le modèle de Cox à hasard proportionnel.**

Soit  $S(t)$ ,  $t \in \mathbf{R}_+^1$ , une fonction de survie, considérée comme la *fonction de survie de base*. A la base de  $S(t)$  on construit soit disant la *classe paramétrique de Lehmann*

$$H_\theta = \{S(t; \theta), \theta \in \Theta = ]0, \infty[ \}$$

de fonctions de survie  $S(t; \theta)$ , en posant

$$S(t; \theta) = S^\theta(t), \quad \theta \in \Theta = ]0, \infty[ = \mathbf{R}_+^1. \quad (26)$$

Soit  $T$  une durée de survie, dont la fonction de survie appartient à cette classe de Lehmann :

$$\mathbf{P}_\theta\{T > t\} = S(t; \theta), \quad t \in \mathbf{R}_+^1.$$

Il est évident que si  $\theta = 1$ , alors on obtient la fonction de survie de base  $S(t)$  :

$$\mathbf{P}_1\{T > t\} = S(t; 1) = S(t), \quad t \in \mathbf{R}_+^1.$$

De (26) il suit que la fonction de défaillance de  $T$  est

$$F(t; \theta) = \mathbf{P}_\theta\{T \leq t\} = 1 - S(t; \theta) = 1 - S^\theta(t), \quad (27)$$

d'où on tire que la densité de  $T$  est

$$f(t; \theta) = \theta[1 - F(t)]^{\theta-1} f(t), \quad (28)$$

où  $F(t) = 1 - S(t)$  et  $f(t) = F'(t)$ . De (26) et (28) on trouve que le taux de défaillance instantané de  $T$  est

$$\alpha(t; \theta) = \theta \frac{f(t)}{S(t)} = \theta \alpha(t), \quad (29)$$

où  $\alpha(t) = f(t)/S(t)$  est le *taux de défaillance de base*, correspondant à  $\theta = 1$ . Grâce à (29) ce modèle est connu sous le nom de *modèle à hasard proportionnel*.

**Remarque 6.** Il est clair que le modèle exponentiel entre dans une classe d'alternatives de Lehmann. En effet, en choisissant

$$S(t) = e^{-t} \mathbf{1}_{]0, \infty[}(t)$$

comme la fonction de survie de base, on obtient le modèle paramétrique exponentiel dont la fonction de survie est

$$S(t; \theta) = e^{-\theta t} \mathbf{1}_{[0, \infty[}(t), \quad \theta > 0.$$

Comme le taux de défaillance de base  $\alpha(t) = 1$ , on en tire que le taux de défaillance de modèle exponentiel est  $\alpha(t; \theta) \equiv \theta$ .

Soit  $Z = (Z_1, \dots, Z_p)^T \in \mathbf{R}^p$  un vecteur de  $p$  variables *exogènes* (*explicatives*), appelé *covariable* ou *stress*,  $Z \in \mathcal{E}$ , où  $\mathcal{E}$  est l'ensemble des tous les stresses admissibles (possibles). Soit

$$r(\cdot) : \mathcal{E} \rightarrow \mathbf{R}_+^1, \quad r(\mathbf{0}_p) = 1,$$

par exemple  $r(Z) = e^{\beta^T Z}$ , où et  $\beta = (\beta_1, \dots, \beta_p)^T \in \mathbf{R}^p$  est le vecteur représentant les *effets estimés* des variables exogènes  $Z_1, \dots, Z_p$  sur  $T$ , alors en introduisant les paramètres

$$\theta = r(Z) = e^{z^T \beta} \quad \text{et} \quad \alpha(t) = \alpha(t | Z = \mathbf{0}_p), \quad (30)$$

où  $z$  est une réalisation observée de  $Z$ , on obtient le modèle (29) dans la forme suivante :

$$\alpha(t; \mathbf{z}) = \alpha(t | Z = z) = \alpha(t) r(z) = \alpha(t) e^{z^T \beta}, \quad (31)$$

connu, au cas  $r(Z) = e^{\beta^T Z}$ , sous le nom de *modèle de régression de Cox avec des covariables constantes en temps*. Dans ce modèle pour tout  $t \in \mathbf{R}_+^1$  le logarithme du taux de hasard

$$\ln \alpha(t | Z = z) = \ln \alpha(t) + \ln r(Z) = \ln \alpha(t) + \sum_{j=1}^p z_j \beta_j$$

est donné par une *régression linéaire* sur des variables explicatives  $Z_1, \dots, Z_p$ .

Souvent  $Z$  ne dépend pas de temps, mais en principe le modèle de Cox générale l'admet.

Le modèle (31), comportant un paramètre  $\beta \in \mathbf{R}^p$  et un paramètre fonctionnel  $\alpha(t)$ , est appelé *semiparamétrique*.

On remarque que

$$\beta_k = \frac{\partial}{\partial z_k} \ln \alpha(t | Z = z), \quad k = 1, \dots, p, \quad (32)$$

et donc nous pouvons dire que le paramètre  $\beta_k$  fournit une bonne approximation de la modification du taux de hasard correspondant à une modification d'une unité de la variable explicative  $Z_k$ . En effet, si, par exemple, une seule variable  $Z_k$  est égale à 1, toutes les autres étant nulles, on obtient que

$$\alpha(t | Z_1 = \dots = Z_{k-1} = Z_{k+1} = \dots = Z_p = 0, Z_k = 1) = \alpha(t) e^{\beta_k},$$

i.e.

$$e^{\beta_k} = \frac{\alpha(t | Z_1 = \dots = z_{k-1} = Z_{k+1} = \dots = Z_p = 0, Z_k = 1)}{\alpha(t)}, \quad (33)$$

d'où on tire que

$$\beta_k = \ln \frac{\alpha(t | Z_1 = \dots = Z_{k-1} = Z_{k+1} = \dots = Z_p = 0, Z_k = 1)}{\alpha(t)}. \quad (34)$$

C'est intéressant de remarquer que les rapports dans (33) et (34) ne dépendent pas du taux de défaillance de base  $\alpha(t)$ . C'est pour cette raison qu'on utilise le vecteur  $\mathbf{b}$  pour évaluer

les effets des variables explicatives (de contrôle)  $Z_1, \dots, Z_p$  sur  $T$ .

**Modèle simple de la vie accélérées.**

Soit  $S(t)$ ,  $t \geq 0$ , une fonction de survie, considérée comme la fonction de survie de base. En utilisant  $S(t)$  nous pouvons construire une classe de Lehmann

$$\{S(t; \theta), \theta \in \Theta = ]0, \infty[ \}$$

de fonction de survie en posant

$$S(t; \theta) = S(\theta t). \quad (35)$$

Si  $T$  une durée de survie, dont la fonction de survie appartient à la classe

$$\{S(t; \theta) = S(\theta t), \theta > 0\}, \quad (36)$$

i.e., pour tout  $\theta > 0$

$$\mathbf{P}_\theta\{T > t\} = S(t; \theta) = S(\theta t), \quad (37)$$

nous disons que l'on a le *modèle simple de la vie accélérées*.

On remarque que si  $\theta = 1$ , alors

$$\mathbf{P}_1\{T > t\} = S(t; 1) = S(t), \quad (38)$$

i.e., la fonction de survie de base  $S(t)$  appartient à la classe (36).

Notons  $F(t) = 1 - S(t)$  la fonction de défaillance de base. Supposons qu'il existe la densité

$$f(t) = F'(t). \quad (39)$$

Dans ce cas pour tout  $\theta$  la fonction de défaillance

$$F(t; \theta) = F(\theta t), \quad t \geq 0, \quad (40)$$

a sa densité

$$f(t; \theta) = \theta f(\theta t), \quad t \in \mathbf{R}_+^1, \quad (41)$$

d'où on tire que le risque instantané de  $T$  est

$$\alpha(t; \theta) = \frac{f(t; \theta)}{S(t; \theta)} = \frac{\theta f(\theta t)}{S(\theta t)}. \quad (42)$$

Si  $T$  est une durée de survie qui suit la loi de base  $F(t)$ , alors il est clair que la durée de survie  $T/\theta$  suit la loi  $F(t; \theta) = F(\theta t)$ , puisque

$$\mathbf{P}_\theta\left\{\frac{T}{\theta} \leq t\right\} = \mathbf{P}_\theta\{T \leq \theta t\} = F(\theta t) = F(t; \theta), \quad (43)$$

d'où on voit clairement le rôle multiplicatif du paramètre  $\theta$  (de paramètre d'échelle) dans le modèle de la vie accélérée : si une durée de survie  $T$  suit une loi  $F(t)$ , considérée comme la loi de base, alors la loi de  $T/\theta$  est

$$F(t; \theta) = F(\theta t) \quad \text{pour chaque } \theta > 0.$$

On remarque que de (11) et (29) il suit que les deux modèles

$$\{S(\theta t), \theta \in \Theta = ]0, 1[ \} \text{ et } \{S^\theta(t), \theta \in \Theta = ]0, 1[ \}$$

coincident si et seulement si le risque instantané de base est

$$\alpha(t) = \gamma t^\beta, \quad \gamma > 0, \beta > 0,$$

i.e. si la fonction de survie de base  $S(t)$  appartient à une famille de Weibull  $W(\alpha, \lambda)$ , donnée par la formule (11) avec  $\alpha = 1 + \beta$  et  $\lambda = \gamma/(1 + \beta)$ .

Enfin on remarque qu'ici nous pouvons nous mettre dans la situation du modèle de Cox, en introduisant le paramètre  $\beta$  et le vecteur covariable  $Z$  telles que

$$\theta = r(Z) = e^{Z^T \beta}, \quad z \in \mathbf{R}^p, \beta \in \mathbf{R}^p,$$

$Z \in \mathcal{E}$ , où  $\mathcal{E}$  est l'ensemble des tous les stresses admissibles (possibles),

$$r(\cdot) : \mathcal{E} \rightarrow \mathbf{R}_+^1, \quad r(\mathbf{0}_p) = 1.$$

### Modèle log-logistique.

Soit  $X$  une variable aléatoire qui suit la *loi logistique standard*  $L(0, 1)$ , dont la densité est

$$g_X(x) = \frac{e^x}{[1 + e^x]^2} = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad x \in \mathbf{R}^1. \quad (44)$$

La fonction de répartition de  $X$  est

$$G(x) = \mathbf{P}\{X \leq x\} = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}.$$

En utilisant  $X$ , on construit une durée de survie  $T$  telle que

$$\ln T = -\ln \mu + \frac{1}{\lambda} X, \quad (45)$$

i.e.

$$T = \exp\left\{\frac{1}{\lambda} X - \ln \mu\right\} \quad (46)$$

pour tout  $\lambda > 0$  et  $\mu > 0$ . Par calcul direct on trouve que la densité de  $T$  est

$$f(t; \theta) = \frac{\lambda \mu (\mu t)^{\lambda-1}}{[1 + (\mu t)^\lambda]^2} \mathbf{1}_{[0, \infty[}(t), \quad (47)$$

où  $\theta = (\mu, \lambda)^T$ ,  $\mu > 0$ ,  $\lambda > 0$ . On dit que  $T$  suit une *loi log-logistique*  $LL(\mu, \lambda)$  de paramètre  $\theta = (\mu, \lambda)$ . De (47) on trouve

$$F(t; \theta) = \mathbf{P}_\theta\{T \leq t\} = \frac{(\mu t)^\lambda}{1 + (\mu t)^\lambda}, \quad t \in \mathbf{R}_+^1$$

et donc la fonction de survie de  $T$  est

$$\mathbf{P}_\theta\{T > t\} = S(t; \theta) = \frac{1}{1 + (\mu t)^\lambda}, \quad t \in \mathbf{R}_+^1. \quad (48)$$

De (47) et (48) on tire que le risque instantané de  $T$  est

$$\alpha(t; \theta) = \frac{\lambda \mu (\mu t)^{\lambda-1}}{1 + (\mu t)^\lambda} = \frac{\lambda}{t} [1 - S(t; \theta)] = \frac{\lambda}{t} F(t; \theta). \quad (49)$$

De (49) il suit que  $\alpha(t; \theta)$  est décroissante, si  $0 < \lambda < 1$ , i.e., dans ce cas  $T$  appartient à la classe DFR. Par contre, si  $\lambda > 1$ , alors  $\alpha(t; \theta)$  a un maximum

$$\lambda_{max} = \mu(\lambda - 1)^{(\lambda-1)/\lambda}$$

au point

$$t = \frac{1}{\mu}(\lambda - 1)^{1/\lambda}.$$

Enfin on remarque que si une durée de survie  $T$  suit une loi log-logistique  $LL(\mu, \lambda)$ , alors

$$\ln \frac{S(t; \theta)}{F(t; \theta)} = \ln \frac{\mathbf{P}_\theta\{T > t\}}{\mathbf{P}_\theta\{T \leq t\}} = -\lambda \ln t - \lambda \ln \mu, \quad (50)$$

i.e., le logarithme du rapport de probabilité de survie à la probabilité de défaillance est une fonction linéaire du logarithme du temps  $t$  (ici  $\theta = (\lambda, \mu)^T$  avec  $\lambda > 0, \mu > 0$ ).

**Remarque 7.** Il est évident que si une fonction de survie  $S(t; \theta)$  est considérée comme la fonction de survie de base, on peut construire la classe d'alternative de Lehmann, en introduisant

$$S(t; \theta, z) = [S(t; \theta)]^{e^{z^T \beta}}, \quad z \in \mathbf{R}^p, \beta \in \mathbf{R}^p.$$

Dans ce modèle le taux de défaillance  $\alpha(t; \theta, \beta)$  est

$$\alpha(t; \theta, z) = \alpha(t; \theta) e^{z^T \beta}.$$

Dans cette optique

$$S(t; \theta) = S(t; \theta, \mathbf{0}_p), \quad \text{et} \quad \alpha(t; \theta) = \alpha(t; \theta, \mathbf{0}_p),$$

où  $\mathbf{0}_p = (0, \dots, 0)^T \in \mathbf{R}^p$ .

## 6.4 Modèles nonparamétriques

**Définition 1 (La classe de Polya d'ordre 2 (PF<sub>2</sub>)).** On dit que  $T \sim PF_2$ , si pour tout  $s \in \mathbf{R}_+^1$  la fonction

$$g_s(t) = \frac{f(t)}{F(t+s) - F(t)}$$

est croissant en  $t$ .

**Théorème 1.**  $T \sim PF_2$  si et seulement si pour tout  $s \in \mathbf{R}_+^1$  la fonction

$$\frac{f(t-s)}{f(t)}$$

est croissant en  $t$ .

On remarque que  $T \sim PF_2$  si et seulement si

$$\frac{f(t+s)}{f(t)}$$

est décroissante en  $t$ , ce qui est équivalent au théorème 1.

**Définition 2 (IFR).** On dit qu'un élément, dont la durée de survie est  $T$ , est *vieillissant* si son taux de panne  $\alpha(t)$  est croissant (décroissant), i.e.

$$\alpha(s) \leq \alpha(t) \quad 0 < s < t, \quad (s, t) \in \mathbf{R}_+^1 \times \mathbf{R}_+^1.$$

On dit aussi que  $T$  a IFR (DFR) et on note  $T \sim IFR$  ( $T \sim DFR$ ).

IFR (DFR) vient de *Increasing (Decreasing) Failure Rate*

Souvent les différents modèles nonparamétriques sont classés suivant que le risque instantané  $\alpha(t)$  est croissant ou décroissant. La fonction  $F(t) = \mathbf{P}\{T \leq t\}$  est alors dite distribution IFR ou DFR respectivement et on note  $F \in IFR$ , ( $F \in DFR$ ).

De la définition 2 il suit que  $T \sim IFR$  si et seulement si la fonction  $\ln S(t)$  est concave.

**Théorème 2.** Soit  $T \sim IFR$ . Alors

$$S(t) > e^{-t/\mathbf{E}T}, \quad 0 < t < \mathbf{E}T.$$

**Démonstration.** Puisque  $T \sim IFR$ , on en tire que la fonction  $\alpha(t)$  est croissante. Comme  $A(t) = \int_0^t \alpha(s) ds$ , alors  $A''(t) = \alpha'(t) > 0$  et donc la fonction  $A(t)$  est convexe, d'où on tire que

$$A(t) \leq \frac{A(\mathbf{E}T)}{\mathbf{E}T} t, \quad 0 \leq t \leq \mathbf{E}T.$$

Puisque  $A(\mathbf{E}T) < 1$ , alors on a

$$S(t) = e^{-A(t)} > e^{-t/\mathbf{E}T}, \quad 0 < t \leq \mathbf{E}T.$$

**Théorème 3.** Soit  $T \sim DFR$ . Alors

$$S(t) \leq \begin{cases} e^{-t/\mathbf{E}T}, & \text{si } t \leq \mathbf{E}T, \\ \frac{1}{e^{t/\mathbf{E}T}}, & \text{si } t \geq \mathbf{E}T. \end{cases}$$

**Théorème 4.** Si  $T \sim IFR$ , alors

$$\mathbf{E}T^2 \leq 2(\mathbf{E}T)^2.$$

**Corollaire 1.** Si  $T \sim IFR$ , alors son coefficient de variation

$$v = \frac{\sqrt{\mathbf{Var}T}}{\mathbf{E}T} \leq 1.$$

**Théorème 5.** Si  $T \sim DFR$ , alors

$$\mathbf{E}T^2 \geq 2(\mathbf{E}T)^2.$$

**Corollaire 2.** Si  $T \sim DFR$ , alors son coefficient de variation

$$v = \frac{\sqrt{\mathbf{Var}T}}{\mathbf{E}T} \geq 1.$$

Souvent pour classer les modèles on emploie *le risque moyen*

$$RM(t) = \frac{1}{t} A(t) = \frac{1}{t} \int_0^t \alpha(s) ds.$$

**Définition 3 (IFRA).** Si  $RM(t)$  est croissant (décroissant), alors on dit que  $T$  a une distribution à taux de défaillance moyen croissant (décroissant) en temps et on note  $T \sim IFRA$  (DFRA).

IFRA (DFRA) vient de *Increasing (Decreasing) Failure Rate Average*.

**Théorème 7.** Soit  $F(t) = \mathbf{P}\{T \leq t\}$  la fonction de répartition de  $T$ ,  $S(t) = 1 - F(t)$  la fonction de survie de  $T$ . Alors  $T \sim IFRA$  si et seulement si pour tout  $\theta \in ]0, 1[$  on a

$$S(\theta t) \geq S^\theta(t), \quad t \in \mathbf{R}_+^1.$$

Autrement dit, si pour tout  $\theta \in ]0, 1[$  la durée de survie du modèle de la vie accélérée correspondante à ce  $\theta$  donné est stochastiquement plus grande que la durée de survie correspondante de la classe de Lehmann (au modèle de Cox).

**Théorème 8.** Si  $T \sim IFR$ , alors  $T \sim IFRA$ , i.e.  $IFR \subset IFRA$ .

**Démonstration.** En effet, comme  $T \sim IFR$ , on a

$$A(t) = \int_0^t \alpha(s) ds \leq t\alpha(t), \quad t \in \mathbf{R}_+^1,$$

puisque le risque instantané  $\alpha(t)$  est croissant, d'où on tire que

$$\left(\frac{A(t)}{t}\right)' = \frac{t\alpha(t) - A(t)}{t^2} \geq 0,$$

i.e.  $A(t)/t$  est croissante en  $t$ , donc  $T \sim IFRA$ .

**Définition 4 (NBU).** On dit que  $T \sim NBU$  (*New Better then Used* où *Neuf meilleur Usagé*) si pour tout  $u > 0$

$$S_u(t) \leq S(t), \quad t \in \mathbf{R}_+^1.$$

Ici  $S_u(t) = 1 - F_u(t) = \mathbf{P}\{T > u + t \mid T > u\}$ .

De cette définition on tire que  $T \sim NBU$  si et seulement si

$$S(u+t) \leq S(u)S(t) \quad \text{pour tout } u, t \in \mathbf{R}_+^1,$$

i.e., si et seulement si pour tout  $u, t \in \mathbf{R}_+^1$

$$A(u+t) \geq A(u) + A(t).$$

**Théorème 9.** Si  $T \sim IFRA$ , alors  $T \sim NBU$ , i.e.  $IFRA \subset NBU$ .

**Démonstration.** Soit  $T \sim IFRA$ . Dans ce cas

$$\frac{A(t)}{t} = \frac{1}{t} \int_0^t \alpha(s) ds, \quad t \in \mathbf{R}_+^1,$$

est croissante en  $t$ , d'où on tire que pour tout  $\theta \in ]0, 1[$  on a

$$A(\theta t) \leq \theta A(t) \quad \text{et} \quad A((1-\theta)t) \leq (1-\theta)A(t), \quad t \in \mathbf{R}_+^1.$$

De ces deux inégalités on déduit que pour tout  $\theta \in ]0, 1[$

$$A(\theta t) + A((1-\theta)t) \leq A(t), \quad t \in \mathbf{R}_+^1.$$

En posant  $\theta t = u$  et  $(1 - \theta)t = v$ , on obtient que

$$A(u) + A(v) \leq A(u + v), \quad u, v \in \mathbf{R}_+^1,$$

et donc  $T \sim NBU$ .

On remarque que  $IFR \neq IFRA$ , i.e. on peut construire une durée de survie  $T$  telle que  $\frac{A(t)}{t}$  est croissante en  $t$ ,  $t \in \mathbf{R}_+^1$ , mais  $\alpha(t)$  n'est pas croissante. En effet, soit

$$\alpha(t) = \begin{cases} t, & \text{si } 0 < t \leq 1, \\ 2 - t, & \text{si } 1 < t \leq \sqrt{2}, \\ 2 - \sqrt{2}, & \text{si } t > \sqrt{2}. \end{cases}$$

Dans ce cas

$$\frac{1}{t}A(t) = \frac{1}{t} \int_0^t \alpha(s) ds = \begin{cases} \frac{t}{2}, & \text{si } 0 < t \leq 1, \\ 2 - \frac{t}{2} - \frac{1}{t}, & \text{si } 1 < t \leq \sqrt{2}, \\ 2 - \sqrt{2}, & \text{si } t > \sqrt{2}. \end{cases}$$

On voit que

$$\left( \frac{1}{t}A(t) \right)' \geq 0, \quad t \in \mathbf{R}_+^1,$$

i.e.  $T \sim IFRA$ , mais la fonction  $\alpha(t)$  n'est pas croissante, et donc la distribution de  $T$  n'appartient pas à IFR.

### La durée de vie moyenne restante.

Soit  $T$  une durée de survie,

$$F(t) = \mathbf{P}\{T \leq t\}, \quad S(t) = 1 - F(t), \quad t \in \mathbf{R}_+^1.$$

Pour tout  $s \geq 0$  nous pouvons considérer la fonction de défaillance conditionnelle

$$\begin{aligned} F_{R_s}(t) = F_s(t) &= \mathbf{P}\{T \leq s + t \mid T > s\} = \mathbf{P}\{T - s \leq t \mid T > s\} = \\ &= \frac{\mathbf{P}\{s < T \leq s + t\}}{\mathbf{P}\{T > s\}} = \frac{F(s + t) - F(s)}{S(s)}, \quad t \in \mathbf{R}_+^1. \end{aligned} \quad (1)$$

On remarque que  $F_s(t) = F(t)$ , si  $s = 0$ . Comme  $F_s(t)$  a toutes les propriétés d'une fonction de répartition, du théorème de Kolmogorov il suit l'existence d'une variable aléatoire réelle  $R_s$  admettant  $F_s(t)$  en qualité de sa fonction de répartition :

$$F_s(t) = \mathbf{P}\{R_s \leq t\}, \quad t \in \mathbf{R}_+^1. \quad (2)$$

**Définition 5.** La variable aléatoire  $R_s$  est appelée *la durée de vie restante où résiduelle* (the residual life) de la durée de survie  $T$  qui a atteint l'âge  $s$ .

Donc, pour tout  $s$  fixé,  $s \in \mathbf{R}_+^1$ ,  $R_s$  est la durée de vie restante de  $T$ , sachant que  $T > s$ , et sa loi est une loi conditionnelle avec la fonction de survie

$$S_{R(s)}(t) = S_s(t) = 1 - F_s(t) = 1 - \frac{F(s + t) - F(s)}{S(s)} =$$

$$\frac{S(s) - [1 - S(s+t)] - [1 - S(s)]}{S(s)} = \frac{S(s+t)}{S(s)}, \quad t \in \mathbf{R}_+^1. \quad (3)$$

Donc  $S_s(t)$  est la probabilité de survie au cours de la période  $]s, s+t]$  sachant que le sujet a été vivant jusqu'à  $s$ ,  $s \in \mathbf{R}_+^1$ . Il faut remarquer ici que de la construction de  $S_s(t)$  il suit que la famille  $\{S_s(t), s \in \mathbf{R}_+^1\}$  de fonctions de survie conditionnelle caractérise la loi de  $T$ . Du théorème 2 il suit que  $F_s(t) = F(t)$  si et seulement si  $X$  suit une loi exponentielle (on suppose que  $F$  est continue).

Si  $T$  a la densité  $f(t) = F'(t)$ , alors la densité de  $R_s$  existe et s'exprime par la formule

$$f_{R_s}(t) = f_s(t) = \frac{f(s+t)}{S(s)}, \quad t \in \mathbf{R}_+^1. \quad (4)$$

Cette formule s'ensuit immédiatement de (1).

On remarque que si  $t = 0$ , alors

$$f_s(0) = \lim_{t \rightarrow 0} F_s(t) = \frac{f(s)}{S(s)} = \alpha(s), \quad (5)$$

où  $\alpha(t) = f(t)/S(t)$ ,  $t \in \mathbf{R}_+^1$ , est le risque instantané de  $T$ . Dans la remarque 3 de §3 on a montré que la fonction de survie  $S(t)$  de  $T$  s'exprime en terme du taux de défaillance instantané  $\alpha(t)$  :

$$S(t) = \exp\left\{-\int_0^t \alpha(u) du\right\}, \quad t \in \mathbf{R}_+^1,$$

d'où on tire que  $S_s(t)$  s'exprime aussi en terme de  $\alpha(t)$  :

$$S_s(t) = \frac{S(s+t)}{S(s)} = \exp\left\{-\int_s^{s+t} \alpha(x) dx\right\} = \exp\left\{-\int_0^t \alpha(u+s) du\right\}, \quad (6)$$

et donc le risque instantané  $\alpha_{R_s}(t)$  de  $R_s$  est

$$\alpha_{R_s}(t) = \alpha(s+t), \quad t \in \mathbf{R}_+^1, \quad (7)$$

où  $\alpha(t)$  est le risque instantané de  $T$ .

Soit  $T$  une durée de survie. Pour tout  $s \in \mathbf{R}_+^1$  on peut considérer sa vie restante  $R_s$ .

**Théorème 10.** *Le risque instantané de défaillance  $\alpha(t)$  de durée de survie  $T$  est croissant si et seulement si  $R_u$  est stochastiquement plus grande que  $R_v$  pour tous  $u < v$ , ( $u, v \in \mathbf{R}_+^1$ ).*

**Démonstration.** De (6) on a

$$S_s(t) = \exp\left\{-\int_s^{s+t} \alpha(x) dx\right\}, \quad t \in \mathbf{R}_+^1,$$

pour tout  $s \in \mathbf{R}_+^1$ , d'où on tire que

$$\frac{\partial}{\partial s} S_s(t) = [\alpha(s) - \alpha(s+t)] S_s(t). \quad (8)$$

Puisque  $S_s(t) > 0$ , on tire de (8) que  $S_s(t)$  est décroissante (croissante) en  $s$  si et seulement si le risque de défaillance  $\alpha(t)$  est croissant (décroissant). Mais si  $S_s(t)$  est décroissante en  $s$ , alors

$$S_u(t) \geq S_v(t) \quad \text{pour tout } u < v, \quad (9)$$

ce qui signifie que  $R_u$  est stochastiquement plus grande que  $R_v$ ,  $u < v$ . Le théorème est démontré.

**Définition 6.** L'espérance mathématique  $r(s) = \mathbf{E}R_s$ ,  $s \in \mathbf{R}_+^1$ , est appelée la *durée moyenne de la vie restante*  $R_s$ .

De cette définition on trouve que

$$r(s) = \mathbf{E}R_s = \mathbf{E}\{T - s \mid T > s\} = \mathbf{E}\{T \mid T > s\} - s, \quad s \in \mathbf{R}_+^1, \quad (10)$$

et en particulier  $r(0) = \mathbf{E}T$ .

**Théorème 11.** La durée moyenne  $r(s)$ ,  $s \in \mathbf{R}_+^1$ , de la vie restante  $R_s$  caractérise la loi de la durée de survie  $T$ .

Pour démontrer ce théorème il suffit de montrer que  $r(s)$  s'exprime en terme de  $S(s)$ , par exemple, ce qui n'est pas difficile, puisque de la définition de  $r(s)$  il suit que

$$r(s) = \frac{1}{S(s)} \int_s^\infty S(u) du, \quad s \in \mathbf{R}_+^1, \quad (11)$$

et réciproquement

$$S(t) = \frac{r(0)}{r(t)} \exp\left\{-\int_0^t \frac{1}{r(x)} dx\right\}, \quad t \in \mathbf{R}_+^1.$$

**Définition 7 (NBUE).** On dit que  $T \sim NBUE$  (*New is Better than Used in Expectation*) si pour tout  $s \in \mathbf{R}_+^1$

$$\mathbf{E}T \geq \mathbf{E}\{T - s \mid T > s\} = \mathbf{E}R_s.$$

**Théorème 12.** Si  $T \sim NBU$ , alors  $T \sim NBUE$ , i.e.  $NBU \subset NBUE$ .

**Démonstration.** En effet, pour tout  $s > 0$  on a

$$S_s(t) \leq S(t), \quad t \in \mathbf{R}_+^1,$$

d'où on tire que

$$\int_0^\infty S_s(t) dt \leq \int_0^\infty S(t) dt,$$

i.e.

$$r(s) = \mathbf{E}R_s \leq \mathbf{E}T = r(0), \quad s \in \mathbf{R}_+^1,$$

donc,  $T \sim NBUE$ .

**Définition 8 (DMRL).** On dit que  $T \sim DMRL$  (*Decreasing Mean Residual Life*), si pour tout  $0 \leq s < t < \infty$

$$\mathbf{E}\{T - s \mid T > s\} \geq \mathbf{E}\{T - t \mid T > t\}.$$

On dit aussi que  $T$  a la *durée de vie moyenne restante décroissante*.

**Théorème 13.** Si  $T \sim DMRL$ , alors  $T \sim NBUE$ , i.e.  $DMRL \subset NBUE$ .

**Définition 9 (HNBUE).** On dit que  $T \sim HNBUE$  (*Harmonic New Better than Used in Expectation*) si

$$\frac{1}{\mathbf{E}T} \int_s^\infty S(t) dt \leq e^{-s/\mathbf{E}T} \quad \text{pour tout } s \in \mathbf{R}_+^1.$$

**Exemple 2.** Soit  $\mathbf{P}\{T \geq t\} = S(t)$ , où

$$S(t) = \begin{cases} 1, & 0 \leq t < 1, \\ 0.25, & 1 \leq t < 3, \\ 0, & t \geq 3. \end{cases}$$

Il est facile de vérifier que  $T \sim HNBUE$ .

Enfin on introduit encore une classe qui est assez naturelle.

**Définition 10. (IDMRL)** On dit que  $T \sim IDMRL$  (*Increasing and Decreasing Mean Residual Life*), si la vie moyenne restante  $r(s)$  de  $T$  est unimodale i.e., s'il existe  $s_0 > 0$  tel que  $r(s)$  est croissante sur  $[0, s_0[$  et décroissante sur  $[s_0, \infty[$ .

De façon analogue à l'introduction de la classe  $DFR$  on peut introduire les classes suivantes :

DFRA - Decreasing Failure Rate on Average,

NWU - New Worse than Used,

NWUE - New Worse than Used in Expectation,

IMRL - Increasing Mean Residual Life,

HNWUE - Harmonic New Worse than Used in Expectation.

## 6.5 Types de censure.

### 1. Censure du type I : temps à censure fixé $C$ .

**Définition 1 (Censure à droite).** Étant donné un échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$  de durées de survie  $X_i$  et un nombre positif fixé  $C$ , on dit qu'il y a censure à droite de cet échantillon, si au lieu d'observer  $X_1, \dots, X_n$ , on observe  $n$  statistiques

$$(T_1, D_1), \dots, (T_n, D_n),$$

où

$$T_i = X_i \wedge C = \min(X_i, C), \quad D_i = \mathbf{1}_{\{T_i = X_i\}} = \begin{cases} 1, & \text{si } X_i \leq C, \\ 0, & \text{si } X_i > C. \end{cases} \quad (6.1)$$

Il est clair que

$$T_i = X_i \mathbf{1}_{\{X_i \leq C\}} + C \mathbf{1}_{\{X_i > C\}}.$$

Donc, en réalité on observe la défaillance (le décès) du sujet  $i$  si  $X_i \leq C$ , et la variable indicatrice  $D_i$  de l'état aux dernières nouvelles vaut 1 dans ce cas. Dans le cas contraire,  $X_i > C$  et donc l'observation est censurée et l'état aux dernières nouvelles  $D_i$  du sujet  $i$  vaut 0. Lorsqu'on ordonne les valeurs de  $T_i$  par ordre croissant, obtenant les statistiques d'ordre

$$T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)},$$

on ne perd aucune information.

C'est par exemple ce qui se passe lorsqu'on observe la durée de fonctionnement de  $n$  systèmes complexes au cours d'une expérience de durée  $C$ .

On remarque qu'en cas de censure non aléatoire à droite le nombre de décès (de pannes) et les durées exactes de survie des sujets décédés sont aléatoires. La période maximale de l'observation  $C$  est fixée. Soit  $f(x_i; \theta)$  la densité de  $X_i$ ,

$$X_i \sim f(x_i; \theta), \quad \theta \in \Theta, \quad x_i \geq 0,$$

et

$$S(x_i; \theta) = 1 - F(x_i; \theta) = \mathbf{P}_\theta\{X_i > x_i\}$$

sa fonction de survie,  $X_i$  est un élément de l'échantillon  $\mathbb{X}$ . Dans ce cas la densité de la statistique  $(T_i, D_i)$  est donnée par la formule

$$g(t_i, d_i; \theta) = [f(t_i; \theta)]^{d_i} [S(t_i; \theta)]^{1-d_i}, \quad t_i > 0; d_i \in \{0, 1\},$$

par rapport à la mesure  $d\lambda \times d\mu$ , où  $\lambda$  est la mesure de Lebesgues sur  $[0, \infty[$ , et  $\mu$  la mesure de comptage sur  $\{0, 1\}$ . Parce que la statistique  $D_i$ , représente la partie discrète de la statistique  $(T_i, D_i)$ , on a

$$\begin{aligned} \mathbf{P}_\theta\{T_i, D_i = 0\} &= \mathbf{P}_\theta\{C \leq, X_i > C\} = S(C; \theta), \\ &= \begin{cases} S(C; \theta) & \text{si } C \leq t_i, \\ 0 & \text{sinon,} \end{cases} = \int_0^{t_i} S(C; \theta) \mathbf{1}_{v > C} dv, \end{aligned}$$

et donc

$$g(t_i, 0; \theta) = S(C; \theta) \mathbf{1}_{t_i > C}.$$

De l'autre côté on a

$$\begin{aligned} \mathbf{P}_\theta\{T_i \leq t_i, D_i = 1\} &= \mathbf{P}_\theta\{X_i \leq t_i, X_i \leq C\} \\ &= \begin{cases} S(C; \theta) & \text{si } t_i \leq C, \\ 0 & \text{sinon,} \end{cases} = \int_0^{t_i} f(v; \theta) \mathbf{1}_{v \leq C} dv, \end{aligned}$$

et donc

$$g(t_i, 1; \theta) = f(t_i; \theta) \mathbf{1}_{t_i \leq C}.$$

Donc la fonction de vraisemblance, correspondant aux observations  $(T_1, D_1), \dots, (T_n, D_n)$ , est

$$L(\theta) = \prod_{i=1}^n [f(T_i; \theta)]^{D_i} [S(C; \theta)]^{1-D_i}. \quad (6.2)$$

On remarque que cette distribution est continue par rapport à  $T_i$ , et discrète par rapport à  $D_i$ .

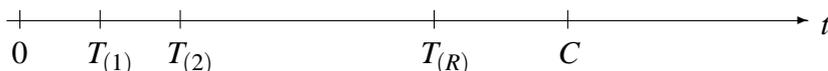
**Exemple 1.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim f(x_i; \theta) = \theta e^{-\theta x_i}, \quad x_i > 0, \quad \theta \in \Theta = ]0, +\infty[,$$

i.e., la durée de survie du sujet  $i$  suit une loi exponentielle de paramètre  $\theta$ . On remarque que dans ce modèle

$$\mathbf{E}X_i = \frac{1}{\theta} \quad \text{et} \quad \mathbf{Var}X_i = \frac{1}{\theta^2}.$$

Supposons que la durée  $C$  de l'étude est fixée à l'avance.



Soient  $\mathbf{T} = (T_1, \dots, T_n)^T$ , où  $T_i = \min(X_i, C)$ , et  $\mathbf{T}^{(c)} = (T_{(1)}, \dots, T_{(R)}, C, \dots, C)^T$  le vecteur des statistiques d'ordre associé à  $\mathbf{T}$ ,

$$0 < T_{(1)} < T_{(2)} < \dots < T_{(R)} < C.$$

La statistique

$$R = D_1 + D_2 + \cdots + D_n = D_{(1)} + D_{(2)} + \cdots + D_{(n)},$$

nous indique le nombre de décès observés,  $R \in \{0, 1, \dots, n\}$ . Ici  $D_{(i)}$  dénote la statistique  $D_i$  associée à  $T_{(i)}$ . De (2) il suit que la fonction de vraisemblance  $L(\theta)$ , correspondante à la statistique  $\mathbf{T}^{(\cdot)}$ , est donnée par la formule

$$L(\theta) = \prod_{i=1}^n \left( \theta e^{-\theta T_i} \right)^{D_i} \left( e^{-\theta T_i} \right)^{1-D_i} = \theta^R \exp \left\{ -\theta \sum_{i=1}^n T_i \right\}, \quad (6.3)$$

donc la statistique exhaustive est bidimensionnelle,

$$(R, T) = \left( \sum_{i=1}^n D_i, \sum_{i=1}^n T_i \right) = \left( \sum_{i=1}^n D_{(i)}, \sum_{i=1}^n T_{(i)} \right), \quad (6.4)$$

où

$$T = T_1 + T_2 + \cdots + T_n = T_{(1)} + T_{(2)} + \cdots + T_{(n)}.$$

On note encore une fois que la loi marginale de  $R$  est discrète, ici elle est binomiale  $B(n, p)$ ,

$$p = p(\theta) = S(C; \theta) = \mathbf{P}_\theta \{X_1 > C\} = e^{-C\theta}, \quad (6.5)$$

et la loi marginale de  $T$  est continue.

Puisque

$$T = \sum_{i=1}^n T_i = \sum_{i=1}^R T_{(i)} + (n-R)C, \quad (6.6)$$

on en tire que la statistique

$$\left( R, \sum_{i=1}^R T_{(i)} + (n-R)C \right) \quad (6.7)$$

est elle aussi exhaustive.

Pour estimer  $\theta$  nous pouvons utiliser la méthode du maximum de vraisemblance. De (3) et (5), de même que de (7), on déduit que

$$\ln L(\theta) = R \ln \theta - \theta \left[ \sum_{i=1}^R T_{(i)} + (n-R)C \right], \quad (6.8)$$

d'où

$$\Lambda(\theta) = \frac{\partial}{\partial \theta} \ln L(\theta) = \frac{R}{\theta} - \left[ \sum_{i=1}^R T_{(i)} + (n-R)C \right], \quad (6.9)$$

et l'estimateur de maximum de vraisemblance  $\hat{\theta}_n$  du paramètre  $\theta$ , qui vérifie l'équation  $\Lambda(\theta) = 0$ , est

$$\hat{\theta}_n = \frac{R}{\sum_{i=1}^R T_{(i)} + (n-R)C} = \frac{\sum_{i=1}^n D_i}{\sum_{i=1}^n T_i}. \quad (6.10)$$

On remarque que si  $R > 10$  et  $n$  assez grand pour que  $R/n < 0.1$ , alors on peut estimer en s'appuyant sur la loi des grands nombres que

$$\sum_{i=1}^R T_{(i)} = R \cdot \frac{1}{R} \sum_{i=1}^R T_{(i)} \approx R \cdot \frac{C}{2}.$$

On déduit alors de (10) que

$$\hat{\theta}_n \approx \frac{R}{\left[n - \frac{R}{2}\right] C}.$$

**Remarque 1.** On dit que la statistique

$$T = \sum_{i=1}^n T_i = \sum_{i=1}^R T_{(i)} + (n - R)C$$

est le temps global de survie (de fonctionnement) des sujets (des éléments) au cours des essais.

**Remarque 2.** Calculons  $M(\theta) = \mathbf{E}_\theta T_i$  et  $D(\theta) = \mathbf{Var}_\theta T_i$ . On a

$$\begin{aligned} M(\theta) &= \mathbf{E}_\theta T_i = \int_0^C t \theta e^{-\theta t} dt + \mathbf{CP}\{X_i > C\} = \\ &= \frac{1}{\theta} \left[ 1 - e^{-\theta C} - \theta C e^{-\theta C} \right] + C e^{-\theta C} = \frac{1}{\theta} \left( 1 - e^{-\theta C} \right). \\ D(\theta) &= \mathbf{Var}_\theta T_i = \mathbf{E} T_i^2 - (\mathbf{E} T_i)^2 = \\ &= \frac{2}{\theta^2} \left[ 1 - e^{-\theta C} - \theta C e^{-\theta C} \right] - \frac{1}{\theta^2} \left( 1 - 2e^{-\theta C} + e^{-2\theta C} \right) = \\ &= \frac{1}{\theta^2} \left[ 1 - 2\theta C e^{-\theta C} - e^{-2\theta C} \right]. \end{aligned}$$

Supposons  $\theta C \ll 1$ , c'est-à-dire que les éléments sont relativement sûrs. En décomposant l'exponentielle en série, on obtient

$$\begin{aligned} e^{-\theta C} &\approx 1 - \theta C + \frac{(\theta C)^2}{2} - \frac{(\theta C)^3}{6}, \\ e^{-2\theta C} &\approx 1 - 2\theta C + 2(\theta C)^2 - \frac{4}{3}(\theta C)^3, \end{aligned}$$

d'où, puisque  $\theta C \ll 1$ ,

$$\begin{aligned} M(\theta) &= \mathbf{E}_\theta T_i \approx C - \frac{\theta C^2}{2} + \frac{\theta^2 C^3}{6} = C \left[ 1 - \frac{\theta C}{2} + \frac{(\theta C)^2}{6} \right], \\ D(\theta) &= \mathbf{Var}_\theta T_i \approx \frac{1}{\theta^2} \left[ 2\theta C - 2(\theta C)^2 + \frac{4}{3}(\theta C)^3 - 2\theta C + 2(\theta C)^2 - (\theta C)^3 + \frac{(\theta C)^4}{3} \right] = \\ &= \frac{\theta C^3}{3} [1 - \theta C] \approx \frac{\theta C^3}{3}. \end{aligned}$$

Si on utilise le théorème limite central, on trouve que si  $n \rightarrow \infty$ , alors le temps global de fonctionnement

$$T = \sum_{i=1}^n T_i = \sum_{i=1}^R T_{(i)} + (n - R)C$$

est asymptotiquement normal de paramètres  $nM(\theta)$  et  $nD(\theta)$  :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{T - nM(\theta)}{\sqrt{nD(\theta)}} \leq x \right\} = \Phi(x),$$

d'où on tire que pour les grandes valeurs de  $n$  la variable aléatoire

$$\left\{ \frac{T - nC \left[ 1 - \frac{\theta C}{2} + \frac{(\theta C)^2}{6} \right]}{\sqrt{n \frac{\theta C^3}{3}}} \right\}^2$$

est distribuée approximativement comme  $\chi_1^2$ , autrement dit pour de grands  $n$  on peut admettre que

$$\left[ \frac{T - nM(\theta)}{\sqrt{nD(\theta)}} \right]^2 = \chi_1^2.$$

On peut utiliser ce résultat pour estimer  $\theta$  par intervalle.

## 2. Censure de type II : jusqu'au $r$ -ième "décès".

Si au lieu de décider à l'avance de la durée  $C$  de l'étude on décide d'attendre que parmi les  $n$  sujets initiaux ou les systèmes de l'étude,  $r$  soient morts ou en panne, on a affaire à une censure de type II. En pratique on applique ce type de censure quand la durée de vie moyenne avant la première panne du système est trop élevée par rapport à la durée de l'étude et on ne fixe pas la durée de l'expérience, mais le nombre  $r$  de pannes que l'on veut observer. Il est évident que dans cette situation le moment d'arrêt de l'expérience, le moment  $T$  du décès de  $r$ -ième sujet, c'est-à-dire la durée de l'expérience est aléatoire. On rappelle que dans le cas de la censure du type I la durée  $C$  de l'étude est fixée à l'avance, mais le nombre de décès observés  $R$  est aléatoire.

**Définition 2. (Censure du type II).** *Étant donné un échantillon*

$\mathbb{X} = (X_1, \dots, X_n)^T$  *de durées de survie*  $X_i$  *et un nombre entier positif*  $r$ , *on dit qu'il y a censure de type II, si au lieu d'observer*  $X_1, \dots, X_n$  *on observe*  $n$  *statistiques*

$$(T_1, D_1), \dots, (T_n, D_n),$$

où

$$T_i = X_i \wedge X_{(r)}, \quad D_i = \mathbf{1}_{\{T_i = X_i\}}, \quad (6.11)$$

$X_{(r)}$  *est la*  $r$ -*ième statistique d'ordre, i.e.*  $X_{(r)}$  *est la*  $r$ -*ième composante du vecteur des statistiques d'ordre*  $X^{(\cdot)} = (X_{(1)}, \dots, X_{(n)})^T$  *associé à l'échantillon*  $\mathbb{X}$ ,

$$0 < X_{(1)} < X_{(2)} < \dots < X_{(r)} < \dots < X_{(n)}. \quad (6.12)$$

C'est-à-dire que dans la situation considérée la date de censure est  $X_{(r)}$  et les observations sont :

$$\begin{aligned} T_{(i)} &= X_{(i)}, & i &= 1, 2, \dots, r, \\ T_{(i)} &= X_{(r)}, & i &= r, r+1, \dots, n. \end{aligned}$$

Si

$$X_i \sim f(x_i; \theta) \quad \text{et} \quad S(x_i; \theta) = \mathbf{P}_\theta\{X_i > x_i\}, \quad x_i > 0, \quad \theta \in \Theta,$$

alors la fonction de vraisemblance associée aux statistiques

$$(T_1, D_1), (T_2, D_2), \dots, (T_n, D_n)$$

est

$$\begin{aligned} L(\theta) &= \frac{n!}{(n-r)!} \prod_{i=1}^n f(T_{(i)}; \theta)^{D_{(i)}} S(T_{(i)}; \theta)^{1-D_{(i)}} = \\ &= \frac{n!}{(n-r)!} \prod_{i=1}^r f(X_{(i)}; \theta) S(X_{(r)}; \theta)^{n-r}, \end{aligned} \quad (6.13)$$

puisque  $\sum_{i=1}^n D_i = r$ , où  $r$  est donné.

**Exemple 2.** Soit

$$X_i \sim f(x_i; \theta) = \theta e^{-\theta x_i}, \quad x_i > 0, \quad \theta \in \Theta = ]0, +\infty[,$$

i.e.  $X_i$  suit une loi exponentielle de paramètre  $\theta$ ,  $\theta > 0$ . Dans ce cas la fonction de vraisemblance, associée aux données censurées (censure du type II) est

$$\begin{aligned} L(\theta) &= \frac{n!}{(n-r)!} \left( \prod_{i=1}^r \theta e^{-\theta X_{(i)}} \right) \left( e^{-\theta X_{(r)}} \right)^{n-r} = \\ &= \frac{n!}{(n-r)!} \theta^r \exp \left\{ -\theta \sum_{i=1}^r X_{(i)} \right\} \exp \left\{ -\theta X_{(r)} (n-r) \right\} = \\ &= \frac{n!}{(n-r)!} \theta^r \exp \left\{ -\theta \left[ \sum_{i=1}^r X_{(i)} + (n-r) X_{(r)} \right] \right\} = \\ &= \frac{n!}{(n-r)!} \theta^r \exp \left\{ -\theta \left[ \sum_{i=1}^r T_{(i)} + (n-r) T_{(r)} \right] \right\}. \end{aligned}$$

On voit que dans ce cas la statistique scalaire

$$T = \sum_{i=1}^r T_{(i)} + (n-r) T_{(r)} = \sum_{i=1}^n T_{(i)} = \sum_{i=1}^n T_i$$

est exhaustive. Elle représente le temps global de survie (de fonctionnement). Il est évident que l'estimateur de maximum de vraisemblance  $\hat{\theta}_n$  est

$$\hat{\theta}_n = \frac{\sum_{i=1}^n D_i}{\sum_{i=1}^n T_i} = \frac{r}{\sum_{i=1}^r T_{(i)} + (n-r) T_{(r)}}.$$

On peut démontrer que

$$\mathbf{P}\{T \leq x\} = \frac{n!}{(n-r)!} \theta^r \int \dots \int_{\substack{0 < t_1 < \dots < t_r \\ \sum_{i=1}^r t_i + (n-r)t_r \leq x}} \exp \left\{ -\theta \left[ \sum_{i=1}^r t_i + (n-r)t_r \right] \right\} dt_1 \dots dt_r =$$

$$= \mathbf{P}\{\chi_{2r}^2 \leq 2\theta x\},$$

i.e.

$$T = \frac{\chi_{2r}^2}{2\theta}.$$

En effet,

$$\begin{aligned} \mathbf{P}\{T \leq x\} &= \frac{n!}{(n-r)!} \theta^r \int \dots \int_{\substack{0 < t_1 < \dots < t_r \\ \sum_{i=1}^r t_i + (n-r)t_r \leq x}} \exp\left\{-\theta \left[\sum_{i=1}^r t_i + (n-r)t_r\right]\right\} dt_1 \dots dt_r = \\ &= \frac{n!}{(n-r)!} \theta^r \int \dots \int_{\substack{0 < t_1 < \dots < t_r \\ \sum_{i=1}^{r-1} t_i + (n-r+1)t_r \leq x}} \exp\left\{-\theta \left[\sum_{i=1}^{r-1} t_i + (n-r+1)t_r\right]\right\} dt_1 \dots dt_r. \end{aligned}$$

Après avoir fait le changement des variables :

$$t_1 = u_1, t_2 = u_1 + u_2, \dots, t_{r-1} = u_1 + \dots + u_{r-1}, \sum_{i=1}^{r-1} t_i + (n-r+1)t_r = u,$$

où  $u_i > 0$  et  $u \leq x$ , on a

$$\mathbf{P}\{T \leq x\} = \frac{n!}{(n-r)!} \theta^r \int \dots \int_{\substack{u_1 > 0, \dots, u_{r-1} > 0 \\ \sum_{i=1}^{r-1} (n-r+1)u_i < u \leq x}} \frac{1}{n-r+1} e^{-\theta u} du_1 \dots du_{r-1} du,$$

puisque

$$\det \left\| \frac{D(t_1, \dots, t_r)}{D(u_1, \dots, u_{r-1}, u)} \right\| = \frac{1}{n-r+1}.$$

En faisant un nouveau changement des variables :

$$(n-i+1)u_i = v_i, \quad i = 1, \dots, r-1, \quad u = v,$$

on trouve que

$$\begin{aligned} \mathbf{P}\{T \leq x\} &= \frac{n!}{(n-r+1)!} \theta^r \int \dots \int_{\substack{v_1 > 0, \dots, v_{r-1} > 0 \\ \sum_{i=1}^{r-1} v_i < v \leq x}} \prod_{i=1}^{r-1} \frac{1}{n-i+1} e^{-\theta v} dv_1 \dots dv_{r-1} dv = \\ &= \theta^r \int_0^x e^{-\theta v} dv \int \dots \int_{\substack{v_1 > 0, \dots, v_{r-1} > 0 \\ \sum_{i=1}^{r-1} v_i < v}} dv_1 \dots dv_{r-1} = \end{aligned}$$

$$= \theta^r \int_0^x v^{r-1} e^{-\theta v} dv = \frac{1}{\Gamma(r)} \int_0^{\theta x} y^{r-1} e^{-y} dy = \mathbf{P}\{\gamma_r \leq \theta x\} = \mathbf{P}\{\chi_{2r}^2 \leq 2\theta x\}.$$

De ce résultat il suit que

$$\mathbf{E}\hat{\theta}_n = \mathbf{E} \frac{2r\theta}{\chi_{2r}^2} = 2r\theta \int_0^\infty \frac{1}{x} \frac{1}{2^r \Gamma(\frac{2r}{2})} x^{\frac{2r}{2}-1} e^{-x/2} dx = \frac{r}{r-1} \theta,$$

et donc

$$\theta_n^* = \frac{r-1}{\sum_{i=1}^r T_{(i)} + (n-r)T_{(r)}}$$

est ici le meilleur estimateur sans biais (MVUE) pour  $\theta$ . On note que

$$\mathbf{Var}\theta_n^* = \frac{\theta^2}{r-2}, \quad r > 2.$$

Enfin, on remarque qu'en l'absence de censure, l'estimateur de maximum de vraisemblance  $\hat{\theta}_n$ , obtenu avec toutes les données  $X_1, \dots, X_n$ , est

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}.$$

**Exercice.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i$  suit une loi exponentielle de paramètre  $\theta$ . Notons

$$Z_i = (n-i+1)(X_{(i)} - X_{(i-1)}), \quad (i = 1, 2, \dots, n; X_{(0)} = 0),$$

$$\omega_i = (Z_1 + \dots + Z_i) / (Z_1 + \dots + Z_{i+1}), \quad (i = 1, 2, \dots, n-1),$$

$$\omega_n = Z_1 + \dots + Z_n, \quad V_i = \omega_i^i \quad (i = 1, 2, \dots, n-1).$$

Montrer que

- les variables aléatoires  $\omega_1, \dots, \omega_n$  sont indépendantes ;
- $V_i \sim \mathcal{U}(0, 1)$ .

Supposons qu'on ait une censure du type II avec  $r = 6$ , et que les 6 premiers moments de défaillance de  $n = 100$  téléviseurs sont :

$$60, 140, 240, 340, 400, 450 \quad (\text{jours}).$$

c) Vérifier l'hypothèse que la durée de survie des téléviseurs suit une loi exponentielle de paramètre  $\theta$ . Utiliser les résultats de a), b) et le critère de Kolmogorov.

### 3. Censure du type III : censure aléatoire.

**Définition 3.** Etant donné un échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$  de durées de survie  $X_i$ , on dit qu'il y a censure aléatoire de cet échantillon s'il existe un autre échantillon  $\mathbb{C}$

$$\mathbb{C} = (C_1, \dots, C_n)^T \in \mathbf{R}_+^n$$

indépendant de  $\mathbb{X}$ , tel que au lieu d'observer  $X_1, \dots, X_n$  on observe les statistiques

$$(T_1, D_1), (T_2, D_2), \dots, (T_n, D_n), \quad (6.14)$$

où

$$T_i = X_i \wedge C_i, \quad D_i = \mathbf{1}_{\{T_i = X_i\}}.$$

Donc en cas de censure aléatoire, on associe à chaque sujet  $i$  ( $i = 1, 2, \dots, n$ ) une statistique de dimension 2 :  $(X_i, C_i) \in \mathbf{R}_+^2$ , dont seulement la plus petite composante est observée :

$$\begin{cases} X_i & \text{est la survie,} \\ C_i & \text{est la censure.} \end{cases}$$

On sait de plus quelle est la nature de cette durée :

si  $D_i = 1$ , c'est une survie,

si  $D_i = 0$ , c'est une censure.

Nous avons supposé que le délai de censure  $C_i$  du sujet  $i$  est une variable aléatoire indépendante de la durée de survie  $X_i$ . Notons

$$H(t) = \mathbf{P}\{C_i \leq t\} \text{ et } Q(t) = \mathbf{P}\{C_i > t\}$$

la fonction de répartition et la fonction de survie de  $C_i$  et  $h(t) = H'(t)$ , densité de  $C_i$ ,  $i = 1, 2, \dots, n$ . Dans ce cas la densité  $g(t_i, d_i; \theta)$  de la statistique  $(T_i, D_i)$  est

$$f(t_i; \theta)Q(t_i), \quad \text{si } D_i = 1 \quad (X_i \text{ est la survie}),$$

$$h(t_i)S(t_i; \theta), \quad \text{si } D_i = 0 \quad (C_i \text{ est la censure}),$$

où  $S(x; \theta) = 1 - F(x; \theta)$ , i.e.,

$$(T_i, D_i) \sim g(t_i, d_i; \theta) = [f(t_i; \theta)Q(t_i)]^{d_i} [h(t_i)S(t_i; \theta)]^{1-d_i}.$$

On obtient donc la fonction de vraisemblance de l'échantillon (14)

$$L(\theta) = \prod_{i=1}^n [f(T_i; \theta)Q(T_i)]^{D_i} [h(C_i)S(C_i; \theta)]^{1-D_i}.$$

Comme  $Q(t)$  et  $h(t)$  ne dépendent pas de  $\theta$  on en tire que

$$L(\theta) = \text{const} \prod_{i=1}^n [f(T_i; \theta)]^{D_i} [S(C_i; \theta)]^{1-D_i}.$$

On remarque que ce résultat suit immédiatement du fait que  $T_1, T_2, \dots, T_n$  forment aussi un échantillon, où  $T_i$  suit la même loi  $H(t; \theta) = 1 - S(t; \theta)Q(t)$  :

$$H(t; \theta) = \mathbf{P}_\theta\{T_i \leq t\} = 1 - \mathbf{P}_\theta\{T_i > t\} = 1 - \mathbf{P}_\theta\{\min(X_i, C_i) > t\} =$$

$$1 - \mathbf{P}_\theta\{X_i > t, C_i > t\} = 1 - \mathbf{P}_\theta\{X_i > t\}\mathbf{P}\{C_i > t\} = 1 - S(t; \theta)Q(t).$$

## 6.6 Troncature.

**Définition 1.** On dit qu'il y a troncature gauche (respectivement droite) lorsque la variable d'intérêt  $T$  n'est pas observable quand elle est inférieure à un seuil  $c > 0$  fixé (respectivement supérieure à un seuil  $C > 0$  fixé).

On remarque que ce phénomène de troncature est très différent de celui de la censure, car dans le cas de la censure, on sait que la variable  $T$ , non observée, est supérieure (ou inférieure) à une valeur  $C$  qui, elle, a été observée. Donc, la troncature élimine de l'étude une partie des  $T$ , ce qui a pour conséquence que l'analyse pourra porter seulement sur la loi de  $T$  conditionnellement à l'événement ( $c < T \leq C$ ), en cas de troncature gauche et droite simultanées.

**Exemple 1.** Soit  $T$  une variable aléatoire, dont la fonction de répartition est

$$F(t) = \mathbf{P}\{T \leq t\}.$$

Supposons que  $T$  ait pour densité  $f(t) = F'(t)$ , et qu'il y ait troncature gauche et droite simultanées : pour cette raison  $T$  est observable seulement sur l'intervalle  $]c, C]$ . Donc, on a une distribution tronquée dont la fonction de répartition est

$$F(t|c < T \leq C) = \begin{cases} 0, & \text{si } t \leq c, \\ \frac{F(t)-F(c)}{F(C)-F(c)}, & \text{si } c < t \leq C, \\ 1, & \text{si } t > C. \end{cases}$$

En termes de fonction de survie de  $T$ ,

$$S(t) = \mathbf{P}\{T > t\} = 1 - F(t),$$

la fonction de survie de la loi tronquée est

$$S(t|c < T \leq C) = \begin{cases} 1, & \text{si } t \leq c, \\ \frac{S(t)-S(C)}{S(c)-S(C)}, & \text{si } c < t \leq C, \\ 0, & \text{si } t > C. \end{cases}$$

Si  $C = +\infty$  et  $c > 0$  on a une troncature à gauche,

si  $c = 0$  et  $C < \infty$  on a une troncature à droite.

Il est facile de vérifier que si  $f(t)$  existe alors la densité de la loi tronquée existe aussi et

$$f(t|c < T \leq C) = \begin{cases} \frac{f(t)}{F(C)-F(c)} = \frac{f(t)}{S(c)-S(C)}, & \text{si } c < t \leq C, \\ 0, & \text{sinon.} \end{cases}$$

Le risque de panne  $\alpha(t|c < T \leq C)$  de la loi tronquée est

$$\alpha(t|c < T \leq C) = \frac{f(t|c < T \leq C)}{S(t|c < T \leq C)} = \frac{f(t)}{S(t) - S(C)},$$

qui peut s'écrire aussi

$$\alpha(t|c < T \leq C) = \frac{f(t)}{S(t)} \frac{S(t)}{S(t) - S(C)} = \alpha(t) \frac{S(t)}{S(t) - S(C)}, \quad c < t \leq C.$$

On remarque que le risque de panne ne dépend que de  $C$ . Donc, s'il n'y a que la troncature à gauche ( $c > 0, C = \infty$ ), on a  $S(C) = 0$  et

$$\alpha(t|c < T) = \alpha(t),$$

i.e. la troncature à gauche ne change pas le risque de panne, tandis que la troncature à droite augmente ce risque.

Notons  $T_{c,C}$  la variable aléatoire, dont la fonction de répartition conditionnelle est

$$F(t|c < T \leq C) = F_{c,C}(t).$$

Il est évident que

$$\int_c^C f_{c,C}(t) dt = 1.$$

Nous pouvons calculer aussi son espérance mathématique

$$\mathbf{E}T_{c,C} = \mathbf{E}\{T|c < T \leq C\} = \int_c^C t f_{c,C}(t) dt.$$

Par exemple, si  $F(t)$  est la fonction de répartition de la loi uniforme sur  $[a, d]$ , i.e.

$$F(t) = \begin{cases} 0, & t \leq a, \\ \frac{t-a}{d-a}, & a < t \leq d, \\ 1, & t > d, \end{cases}$$

et

$$[c, C] \subset ]a, d],$$

alors

$$\begin{aligned} F(t|c < T \leq C) &= \begin{cases} 0, & \text{si } t \leq c, \\ \frac{F(t)-F(c)}{F(C)-F(c)}, & \text{si } c < t \leq C, \\ 1, & \text{si } t > C, \end{cases} \\ &= \begin{cases} 0, & \text{si } t \leq c, \\ \frac{t-c}{C-c}, & \text{si } c < t \leq C, \\ 1, & \text{si } t > C, \end{cases} \end{aligned}$$

et la distribution tronquée est de nouveau uniforme, mais sur l'intervalle  $]c, C]$ .

(Voir Woodroffe (1985), Huber (1989).)

**Exemple 2. Modèle de la loi normale tronquée.** Soit  $T$  une durée de survie dont la fonction de répartition est

$$F(t; \mu, \sigma^2) = \mathbf{P}_{\mu, \sigma^2}\{T \leq t\} = \frac{\Phi\left(\frac{t-\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)}{1 - \Phi\left(-\frac{\mu}{\sigma}\right)} \mathbf{1}_{[0, \infty[}(t), \quad t \in \mathbf{R}^1,$$

où  $\Phi(\cdot)$  est la fonction de répartition de la loi normale standard  $N(0, 1)$ ,  $|\mu| < \infty$ ,  $\sigma^2 > 0$ . On dit que la durée de survie  $T$  suit la loi normale, *tronquée au zéro*. La fonction de survie de  $T$  est

$$S(t; \mu, \sigma^2) = 1 - F(t; \mu, \sigma^2) = \frac{1 - \Phi\left(\frac{t-\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} \mathbf{1}_{[0, \infty[}(t), \quad t \in \mathbf{R}^1,$$

et la densité de  $T$  est

$$f(t; \mu, \sigma^2) = \frac{1}{\sigma \Phi\left(\frac{\mu}{\sigma}\right)} \varphi\left(\frac{t-\mu}{\sigma}\right) \mathbf{1}_{[0, \infty[}(t), \quad t \in \mathbf{R}^1,$$

où  $\varphi(\cdot) = \Phi'(\cdot)$ , d'où on tire que le risque de panne  $\alpha(t)$  est

$$\alpha(t) = \frac{\varphi\left(\frac{t-\mu}{\sigma}\right)}{\sigma \Phi\left(\frac{\mu-t}{\sigma}\right)} \mathbf{1}_{[0, \infty[}(t), \quad t \in \mathbf{R}^1,$$

puisque  $\Phi(x) + \Phi(-x) \equiv 1$ ,  $x \in \mathbf{R}^1$ .

La vie moyenne  $\mathbf{E}T$  de  $T$  est

$$\begin{aligned} \mathbf{E}T &= \int_0^\infty S(t; \mu, \sigma^2) dt = \frac{1}{\Phi\left(\frac{\mu}{\sigma}\right)} \int_0^\infty \Phi\left(\frac{\mu-t}{\sigma}\right) dt = \frac{\sigma}{\Phi\left(\frac{\mu}{\sigma}\right)} \int_{-\infty}^{\mu/\sigma} \Phi(u) du = \\ &= \frac{\sigma}{\Phi\left(\frac{\mu}{\sigma}\right)} \left[ \frac{\mu}{\sigma} \Phi\left(\frac{\mu}{\sigma}\right) - \int_{-\infty}^{\mu/\sigma} u \varphi(u) du \right] = \\ &= \mu + \frac{\sigma}{\Phi\left(\frac{\mu}{\sigma}\right)} \int_{-\infty}^{\mu/\sigma} \varphi'(u) du = \mu + \frac{\sigma \varphi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} > \mu. \end{aligned}$$

Pour étudier le comportement de  $\alpha(t)$  on remarque que

$$\varphi(x) \left(1 - \frac{4}{x^2}\right) < \varphi(x) < \varphi(x) \left(1 + \frac{1}{x^2}\right), \quad x > 0, \quad (1)$$

d'où on tire immédiatement que

$$\left(\frac{1}{x} - \frac{1}{x^2}\right) \varphi(x) < 1 - \Phi(x) < \frac{1}{x} \varphi(x), \quad x > 0, \quad (2)$$

puisque

$$\frac{1}{x} \varphi(x) = \int_0^\infty \varphi(u) \left(1 + \frac{1}{u^2}\right) du$$

et

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \varphi(x) = \int_x^\infty \varphi(u) \left(1 - \frac{4}{u^2}\right) du.$$

Comme

$$\alpha(t) = \frac{\varphi\left(\frac{t-\mu}{\sigma}\right)}{\sigma \Phi\left(\frac{\mu-t}{\sigma}\right)} \mathbf{1}_{[0, \infty[}(t), \quad t \in \mathbf{R}^1,$$

de (2) on tire que

$$\frac{1}{t} - \frac{1}{t^3} < \frac{1}{\alpha(t)} < \frac{1}{t},$$

d'où on obtient que

$$\lim_{t \rightarrow \infty} \frac{\alpha(t)}{t} = 1.$$

## 6.7 Estimateur de Kaplan-Meier.

Si l'on ne peut pas supposer a priori que la loi de la durée de survie  $X$  obéit à un modèle paramétrique, on peut estimer la fonction de survie  $S(t)$  grâce à plusieurs méthodes *non-paramétriques* dont la plus intéressante est celle de *Kaplan-Meier*, (1958).

Cet estimateur est aussi appelé *P-L estimateur* car il s'obtient comme un produit : la probabilité de survivre au delà de l'instant  $t_{(n)}$  est égale au produit suivant :

$$S(t_{(n)}) = \mathbf{P}\{X > t_{(n)}\} = \mathbf{P}(X > t_{(n)} | X > t_{(n-1)}) \cdot S(t_{(n-1)}) = \Delta_n p_{t_{(n-1)}} S(t_{(n-1)}),$$

où  $0 = t_{(0)} < t_{(1)} < \dots < t_{(n)}$ ,

$$\Delta_n p_{t_{(n-1)}} = S_{t_{(n-1)}}(\Delta_n), \quad \Delta_n = t_{(n)} - t_{(n-1)},$$

$t_{(n-1)}$  est une date antérieure à  $t_{(n)}$ .



Si on renouvelle l'opération en choisissant une date  $t_{(n-2)}$  antérieure à  $t_{(n-1)}$ , on aura de même

$$S(t_{(n-1)}) = \mathbf{P}\{X > t_{(n-1)}\} = \mathbf{P}(X > t_{(n-1)} | X > t_{(n-2)}) \cdot S(t_{(n-2)}),$$

et ainsi de suite, on obtient la formule :

$$S(t_{(n)}) = \prod_{i=1}^n \Delta_i p_{t_{(i-1)}} = \prod_{i=1}^n (1 - \Delta_i q_{t_{(i-1)}}),$$

sachant que  $S(0) = 1$ .

Cet estimateur est bien adopté aux cas de la présence de la censure. Si on choisit pour dates où l'on conditionne celles où s'est produit un événement, qu'il s'agisse d'une mort ou d'une censure,  $t_{(i)} = T_{(i)}$  on aura seulement à estimer des quantités de la forme :

$$p_i = \mathbf{P}\{X > T_{(i)} | X > T_{(i-1)}\} = \Delta_i p_{T_{(i-1)}},$$

qui est la probabilité de survivre pendant l'intervalle de temps  $\Delta_i = ]T_{(i-1)}; T_{(i)}]$  quand on était vivant au début de cet intervalle. On note que

$$0 = T_{(0)} \leq T_{(1)} \leq \dots \leq T_{(n)}.$$

Notons :

$R_i = \text{card } R(T_{(i)}^-)$  le nombre des sujets qui sont vivants juste avant l'instant  $T_{(i)}$ , en désignant par  $R(t^-)$  l'ensemble des sujets à risque à l'instant  $t^-$  ;

$M_i =$  le nombre de morts à l'instant  $T_{(i)}$  ;

$q_i = 1 - p_i$  la probabilité de mourir pendant l'intervalle  $\Delta_i$  sachant que l'on était vivant au début de cet intervalle.

Alors l'estimateur naturel de  $q_i$  est

$$\hat{q}_i = \frac{M_i}{R_i}.$$

Supposons d'abord qu'il n'y ait pas d'ex-aequo, i.e. on suppose que

$$0 = T_{(0)} < T_{(1)} < \dots < T_{(n)}.$$

Dans ce cas,

si  $D_{(i)} = 1$ , c'est qu'il y a eu un mort en  $T_{(i)}$  et donc  $M_i = 1$ ,

si  $D_{(i)} = 0$ , c'est qu'il y a eu une censure en  $T_{(i)}$  et donc  $M_i = 0$ .

Par suite,

$$\hat{p}_i = 1 - \frac{M_i}{R_i} = \left(1 - \frac{1}{R_i}\right)^{D_{(i)}} = \begin{cases} 1 - \frac{1}{R_i}, & \text{en cas de mort en } T_{(i)}, \\ 1, & \text{en cas de censure en } T_{(i)}, \end{cases}$$

donc  $\hat{p}_i$  n'est différent de 1 qu'aux instants de décès observés.

L'estimateur de Kaplan-Meier pour la fonction de survie  $S(t)$  est :

$$\begin{aligned} \hat{S}(t) &= \hat{S}_n(t) = \prod_{T_{(i)} \leq t} \hat{p}_i = \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{R_i}\right)^{D_{(i)}} = \\ &= \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n-i+1}\right)^{D_{(i)}}. \end{aligned}$$

Il est évident que en absence de la censure, i.e. si  $D_i = 1$  pour  $\forall i$ , alors

$$\hat{S}_n(t) = \begin{cases} 1, & t \leq T_{(1)}, \\ \frac{n-i}{n}, & T_{(i)} \leq t < T_{(i+1)}, \\ 0, & t \geq T_{(n)}. \end{cases}$$

On remarque que  $R_i = n - i + 1$  car, mort ou censuré le sujet disparaît de l'étude.

Il est évident que l'estimateur de Kaplan-Meier  $\hat{F}_n(t)$  de  $F(t) = 1 - S(t)$  est

$$\hat{F}_n(t) = 1 - \hat{S}_n(t) = \begin{cases} 0 & , \text{ si } t < T_{(1)}, \\ 1 - \prod_{T_{(i)} \leq t} \left(\frac{n-i}{n-i+1}\right)^{D_{(i)}} & , \text{ si } T_{(1)} \leq t < T_{(n)}, \\ 1, & \text{ si } t \geq T_n. \end{cases}$$

Pour estimer la variance de  $\hat{S}_n(t)$ , on utilise l'approximation de Greenwood, d'après laquelle

$$\mathbf{Var} [\hat{S}_n(t)] \approx [\hat{S}_n(t)]^2 \sum_{i: T_i \leq t} \frac{D_i}{(n-i)(n-i+1)}.$$

La moyenne  $\mathbf{E}X_i$  de survie  $X_i$  est estimée par  $\int_0^\infty \hat{S}_n(t) dt$ . Enfin on remarque que

$$\hat{A}_n(t) = -\ln \hat{S}_n(t)$$

peut-être considéré comme l'estimateur de Kaplan-Meier de la fonction de risque cumulée  $A(t)$ .

Quand  $n$  est assez grand pour évaluer  $\hat{A}_n(t)$  on peut utiliser l'approximation de Nelson :

$$\hat{A}_n(t) \approx \sum_{i: T_i \leq t} \frac{D_i}{n-i+1},$$

puisque

$$\log\left(1 - \frac{1}{n-j+1}\right) \approx -\frac{1}{n-j+1},$$

pour les grandes valeurs de  $n-j+1$ . La statistique

$$A_n^*(t) = \sum_{i:T_i \leq t} \frac{D_i}{n-i+1}$$

est connue comme l'estimateur de Nelson pour le taux de hasard cumulé  $A(t)$ .

**Théorème 1.** Si les lois  $F$  et  $H$  de la survie  $X_i$  et de la censure  $C_i$  n'ont aucune discontinuité commune, la suite d'estimateurs  $\{\hat{S}_n(t)\}$  de Kaplan-Meier de la fonction de survie  $S(t)$  est consistante.

**Théorème 2.** Si l'échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$  et l'échantillon de censure  $\mathbb{C} = (C_1, \dots, C_n)^T$  sont indépendants, alors dans les conditions du théorème 1

$$\sqrt{n}(\hat{S}_n(t) - S(t)) \xrightarrow{\mathcal{L}} W(t), \quad n \rightarrow \infty,$$

où  $W(t)$  est un processus gaussien centré,  $\mathbf{E}W(t) \equiv 0$ , dont la fonction de covariance est

$$k(s, t) = \mathbf{E}W(s)W(t) = S(s)S(t) \int_0^{s \wedge t} \frac{dF(u)}{[1-F(u)]^2[1-H(u)]}.$$

**Remarque 1.** Il est facile à voir que

$$\mathbf{E} \frac{\hat{S}_n(t)}{S(t \wedge T_{(n)})} = 1,$$

et donc

$$S(t) = \mathbf{E} \frac{S(t)}{S(t \wedge T_{(n)})} \hat{S}_n(t) > \mathbf{E} \hat{S}_n(t),$$

i.e. l'estimateur de Kaplan-Meier  $\hat{S}_n(t)$  n'est pas un estimateur sans biais pour  $S(t)$ .

**Remarque 2.** Si  $S(t)$  est continue, alors pour  $\forall t < H^{-1}(1)$

$$\hat{S}_n(t) = S(t) + \frac{1}{n} \sum_{i=1}^n \psi_i(t) + r_n(t),$$

où  $\psi_i(t)$  sont i.i.d.,  $\mathbf{E}\psi_i(t) = 0$ , uniformément bornées sur  $[0, T]$ , et

$$\sup_{t \in [0, T]} |r_n(t)| = O(n^{-1} \log n) \quad (\text{mod } \mathbf{P})$$

quand  $T < H^{-1}(1)$ ,  $H(t) = \mathbf{P}\{T_i \leq t\}$ .

**Théorème 3.** Dans les conditions du théorème 2 l'estimateur de Nelson  $A_n^*$  du taux de hasard cumulé  $A$  vérifie :

$$\sqrt{n}(\hat{A}_n^*(t) - A(t)) \xrightarrow{\mathcal{L}} W(t), \quad n \rightarrow \infty,$$

où  $W(t)$  est un processus gaussien centré,  $\mathbf{E}W(t) \equiv 0$ , dont la fonction de corrélation est

$$k(s, t) = \mathbf{E}W(s)W(t) = \int_0^{t_1 \wedge t_2} \frac{dG(t, 1)}{S^2(t)},$$

où  $G(t, 1) = \mathbf{P}\{T_i \geq t, D_i = 1\}$ .

**Exemple 1.** Sur 10 patients atteints de cancer des bronches on a observé les durées de survie suivantes, exprimées en mois :

$$1 \quad 3 \quad 4^+ \quad 5 \quad 7^+ \quad 8 \quad 9 \quad 10^+ \quad 11 \quad 13^+.$$

Les données suivies du signe  $+$  correspondent à des patients qui ont été perdues de vue à la date considérée, i.e. censurées.

L'estimateur de Kaplan-Meier  $\hat{S}(t) = \hat{S}_{10}(t)$  de la fonction de survie  $S(t)$  vaut :

$$\hat{S}(0) = 1 \quad \text{et} \quad \hat{S}(t) = 1 \quad \text{pour tout } t \text{ dans } [0; 1[$$

$$\hat{S}(t) = (1 - \frac{1}{10})\hat{S}(0) = 0.9, \quad 1 \leq t < 3,$$

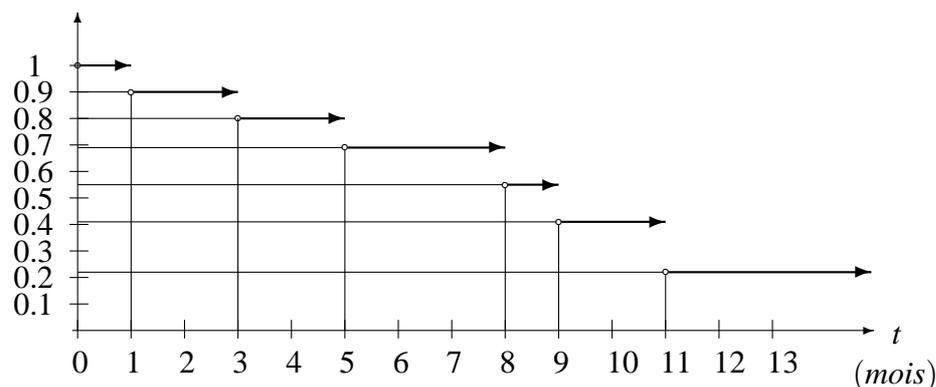
$$\hat{S}(t) = (1 - \frac{1}{9})\hat{S}(1) = 0.80, \quad 3 \leq t < 5,$$

$$\hat{S}(t) = (1 - \frac{1}{7})\hat{S}(3) = 0.694, \quad 5 \leq t < 8,$$

$$\hat{S}(t) = (1 - \frac{1}{5})\hat{S}(5) = 0.555, \quad 8 \leq t < 9,$$

$$\hat{S}(t) = (1 - \frac{1}{4})\hat{S}(8) = 0.416, \quad 9 \leq t < 11,$$

$$\hat{S}(t) = (1 - \frac{1}{2})\hat{S}(9) = 0.208.$$



Mais la plupart du temps il y a des ex-aequo, comme dans le premier exemple qui est celui des données de Freireich de l'exemple suivant.

**Exemple 2 (Données de Freireich).** Ces données, très souvent citées dans la littérature statistique médicale car les performances des diverses méthodes sont souvent testées sur elles, ont été obtenues par Freireich, en 1963, lors d'un essai thérapeutique ayant pour but de comparer les durées de rémission, exprimées en semaines, de sujets atteints de leucémie selon qu'ils ont reçu de la 6-mercaptopurine (notée 6-MP) ou un placebo. L'essai a été fait en double aveugle, c'est-à-dire que ni le médecin, ni le patient ne sont informés de l'attribution du traitement ou du placebo.

Le tableau ci-après donne, pour chacun des 42 sujets, la durée de rémission.

Traitement	Durée de rémission
6-MP	6, 6, 6, 6 <sup>+</sup> , 7, 9 <sup>+</sup> , 10, 10 <sup>+</sup> , 11 <sup>+</sup> , 13, 16, 17 <sup>+</sup> , 19 <sup>+</sup> , 20 <sup>+</sup> , 22, 23, 25 <sup>+</sup> , 32 <sup>+</sup> , 32 <sup>+</sup> , 34 <sup>+</sup> , 35 <sup>+</sup> .
Placebo	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.

Les chiffres suivis du signe  $+$  correspondent à des patients qui ont été perdus de vue à la date considérée. Ils sont donc exclus vivants de l'étude et on sait seulement d'eux que leur durée de vie est supérieure au nombre indiqué. Par exemple, le quatrième patient traité par 6-MP a eu une durée de rémission supérieure à 6 semaines. On dit que les perdus de vue ont été *censurés*, et ce problème de censure demande un traitement particulier. En effet, si l'on se contentait d'éliminer les observations incomplètes, c'est-à-dire les 12 patients censurés du groupe traité par le 6-MP, on perdrait beaucoup d'information : un test de Wilcoxon appliqué aux 9 patients restant dans le groupe 6-MP et aux 21 patients du groupe placebo sous-évaluerait très visiblement l'effet du traitement.

**Cas où il y a des ex-aequo :**  $0 = T_{(0)} \leq T_{(1)} \leq \dots \leq T_{(n)}$ .

1) Si ces ex-aequo sont tous de morts la seule différence tient à ce que  $M_i$  n'est plus égal à 1 mais au nombre des morts et l'estimateur de Kaplan-Meier devient :

$$\hat{S}(t) = \prod_{T_{(i)} \leq t} \left(1 - \frac{M_i}{R_i}\right).$$

2) Si ces ex-aequo sont des deux sortes, on considère que *les observations non censurées ont lieu juste avant les censurées*. Voyons ce que donne l'estimateur de Kaplan-Meier dans le cas des données de Freireich :

**Pour le 6-MP :**

$$\hat{S}(0) = 1 \quad \text{et} \quad \hat{S}(t) = 1 \quad \text{pour tout } t \text{ dans } [0; 6[,$$

$$\hat{S}(6) = \left(1 - \frac{3}{21}\right)\hat{S}(0) = 0.857,$$

$$\hat{S}(7) = \left(1 - \frac{1}{17}\right)\hat{S}(6) = 0.807,$$

$$\hat{S}(10) = \left(1 - \frac{1}{15}\right)\hat{S}(7) = 0.753,$$

$$\hat{S}(13) = \left(1 - \frac{1}{12}\right)\hat{S}(10) = 0.690,$$

$$\hat{S}(16) = \left(1 - \frac{1}{11}\right)\hat{S}(13) = 0.627,$$

$$\hat{S}(22) = \left(1 - \frac{1}{7}\right)\hat{S}(16) = 0.538,$$

$$\hat{S}(23) = \left(1 - \frac{1}{7}\right)\hat{S}(22) = 0.448.$$

**Pour le Placebo :**

$$\begin{aligned}
 \hat{S}(t) &= 1, \quad 0 \leq t < 1, \\
 \hat{S}(1) &= (1 - \frac{2}{21})\hat{S}(0) = 0.905, \\
 \hat{S}(2) &= (1 - \frac{2}{19})\hat{S}(1) = 0.895, \\
 \hat{S}(3) &= (1 - \frac{1}{17})\hat{S}(2) = 0.842, \\
 \hat{S}(4) &= (1 - \frac{2}{16})\hat{S}(3) = 0.737, \\
 \hat{S}(5) &= (1 - \frac{2}{14})\hat{S}(4) = 0.632, \\
 \hat{S}(8) &= (1 - \frac{4}{12})\hat{S}(5) = 0.421, \\
 \hat{S}(11) &= (1 - \frac{2}{8})\hat{S}(8) = 0.316, \\
 \hat{S}(12) &= (1 - \frac{2}{6})\hat{S}(11) = 0.210, \\
 \hat{S}(15) &= (1 - \frac{1}{4})\hat{S}(12) = 0.158, \\
 \hat{S}(17) &= (1 - \frac{1}{3})\hat{S}(15) = 0.105, \\
 \hat{S}(22) &= (1 - \frac{1}{2})\hat{S}(17) = 0.053, \\
 \hat{S}(23) &= (1 - \frac{1}{1})\hat{S}(22) = 0.
 \end{aligned}$$

Plus d'information sur le modèle de survie on peut voir dans Kaplan and Meier (1958), Turnbull (1974), (1976), Kalbfleisch and Prentice (1980), Lawless (1982), Drosbeke, Fichet & Tassi (1989), Bagdonavičius et Nikulin (1995, 1998, 1999).

## 6.8 Modèle de Cox.

Le modèle de Cox est employé lorsque on cherche à évaluer l'effet de certaines variables sur la durée de survie. D'après ce modèle on a les  $2n$  variables aléatoires indépendantes

$$X_1, X_2, \dots, X_n \quad \text{et} \quad C_1, C_2, \dots, C_n$$

que sont les durées de survie et les temps de censures des  $n$  individus considérés. En réalité, on observe la suite des  $n$  vecteurs  $(T_i, D_i)$ , où  $T_i$  date de départ du  $i$ -ème individu (en supposant qu'ils sont entrés à l'instant 0),  $D_i$  indicatrice de la cause de départ ( $D_i = 1$  si c'est la mort,  $D_i = 0$  sinon),

$$D_i = \mathbf{1}_{\{X_i \leq C_i\}}.$$

Mais on a aussi observé sur chacun des individus un vecteur  $Z_i = (Z_{i1}, \dots, Z_{ip})^T$  dont dépend la durée de survie  $X_i$ . Ce vecteur  $Z$  est généralement appelé *covariable*.

Le modèle des "*hasards proportionnels*", ou modèle de Cox suppose que

$$\alpha(t|Z = z) = \alpha_z(t) = \alpha_0(t) \exp \left\{ \beta^T z \right\},$$

$$\beta^T z = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p,$$

où  $\beta^T = (\beta_1, \dots, \beta_p)^T$  est le vecteur des coefficients de la régression,  $\alpha_0(t)$  est le *risque instantané de base*. En général, ils sont inconnus tous les deux. C'est pour cela on dit

souvent que le modèle de Cox est *semiparamétrique*.

**Remarque 1.** La famille des loi d'un tel modèle est du type suivant :

toutes les fonctions de survie sont égales à une même *fonction de survie*  $S_0(t)$  de base, élevée à des puissances variées :

$$S(t; \theta) = [S_0(t)]^\theta,$$

$$S_0(t) = \exp \left\{ - \int_0^t \alpha_0(u) du \right\},$$

$$\theta = \exp \left\{ \sum_{j=1}^p \beta_j z_j \right\} = \exp \left\{ \beta^T z \right\}.$$

**Exemple 1.** Prenons le cas le plus simple : 1 seule covariable ( $p = 1$ ),  $Z$  prenant seulement les valeurs 0 ou 1. Il peut s'agir par exemple d'un essai thérapeutique destiné à comparer l'effet d'un nouveau traitement ( $Z = 1$  pour les patient traités) à celui du traitement habituel ou d'un placebo ( $Z = 0$ ), sur la durée de survie.

On a alors deux populations :

$$\begin{aligned} \text{si } Z = 0, & \quad S(t) = S_0(t), \\ \text{si } Z = 1, & \quad S_1(t) = [S_0(t)]^\gamma, \end{aligned}$$

où  $\gamma = e^\beta$  mesure l'effet du traitement.

Ce modèle comporte donc un paramètre qui est une fonction  $\lambda_0$ , considérée en général comme nuisible et  $p$  paramètre réels  $\beta_1, \beta_2, \dots, \beta_p$  qui sont les quantités à estimer, où à tester, car elles représentent l'effet sur la durée de survie de chacune des covariables correspondantes.

Pour éliminer le "paramètre" nuisible totalement inconnu qu'est le risque instantané de base  $\alpha_0(t)$ , Cox (1972) considère la vraisemblance "partielle" suivante

$$V_C(\beta) = \prod_{\{i: D_{(i)}=1\}} \frac{\exp \left\{ \beta^T Z^{(i)} \right\}}{\sum_{k \in R_{(i)}} \exp \left\{ \beta^T Z^{(k)} \right\}},$$

où  $T_{(1)} < T_{(2)} < \dots < T_{(n)}$  désignent la suite des instant où a lieu un événement (mort ou censure), et à l'instant  $T_{(i)}$  sont observés :

$D_{(i)}$  la nature de l'événement ;  $D_{(i)} = 1$ , si c'est une mort,  $D_{(i)} = 0$ , si c'est une censure ;

$Z^{(i)}$  la covariable, de dimension  $p$ , de l'individu à qui est arrivé l'événement ;

$R_{(i)}$  l'ensemble des indices des individus encore à risque à l'instant  $T_{(i)}$  ainsi que la valeur de leur covariable,  $Z^{(k)}$ ,  $k \in R_{(i)}$ .

Cox traite cette vraisemblance partielle comme une vraisemblance ordinaire.

En temps continu, on fait l'hypothèse qu'il n'y a aucun ex-aequo, et dans ce cas

$$\mathcal{L} \left( \sqrt{n}(\hat{\beta}_n - \beta) \right) \rightarrow N \left( \mathbf{0}, I^{-1}(\beta) \right),$$

où  $\hat{\beta}_n$  est l'estimateur de maximum de vraisemblance partielle pour  $\beta$ ,

$$V_C(\hat{\beta}_n) = \max V_C(\beta).$$

## 6.9 Sur l'estimation semiparamétrique pour le modèle de Cox

On observe  $n$  individus. Notons  $X_i$  et  $C_i$  les durées de survie et les temps de censures. On suppose que la durée de survie du  $i$ -ème individu dépend du vecteur  $\mathbf{Z}_i(\cdot) = (Z_{i1}(\cdot), \dots, Z_{ip}(\cdot))^T$  des covariables. Posons

$$T_i = X_i \wedge C_i, \quad D_i = \mathbf{1}\{X_i \leq C_i\}.$$

Nous supposons que les statistiques  $(X_1, C_1), \dots, (X_n, C_n)$  soient indépendantes. On a un échantillon  $(T_i, D_i, Z_i(\cdot))$ ,  $(i = 1, \dots, n)$ .

Supposons que la *censure est indépendante* pour chaque individu, i.e.  $\alpha_{ci}(t) = \alpha_i(t)$  pour tout  $t : \mathbf{P}(T_i \geq t) > 0$ , où

$$\alpha_{ci}(t) = \lim_{h \downarrow 0} \frac{\mathbf{P}\{T_i \in [t, t+h[, D_i = 1 | T_i \geq t\}}{h},$$

$$\alpha_i(t) = \lim_{h \downarrow 0} \frac{\mathbf{P}\{X_i \in [t, t+h[ | X_i \geq t\}}{h}.$$

Supposons que les variables aléatoires  $X_1, \dots, X_n$  sont absolument continues. Soient  $N$  le nombre,  $X_{(1)} < \dots < X_{(N)}$  les moments des décès observés,  $(i)$  l'indice de l'individu décédé au moment  $X_{(i)}$ ,  $R_{(i)}$  l'ensemble des indices des individus à risque à l'instant  $X_{(i)}^-$ .

Supposons que le modèle de Cox ait vérifié :

$$\alpha_{Z_i(\cdot)}(t) = e^{\beta^T \mathbf{Z}_i(t)} \alpha_0(t),$$

où  $\beta = (\beta_1, \dots, \beta_p)^T$  est le vecteur des coefficients de regression inconnus,  $\alpha_0(t)$  le risque instantané de base inconnu. Alors

$$p_i(j|r, t) = \mathbf{P}\{(i) = j | R_{(i)} = r, X_{(i)} = t\} = \lim_{h \downarrow 0} \mathbf{P}\{(i) = j | R_{(i)} = r, X_{(i)} \in [t, t+h)\} =$$

$$\lim_{h \downarrow 0} \frac{\mathbf{P}\{(i) = j, R_{(i)} = r, X_{(i)} \in [t, t+h)\}}{\mathbf{P}\{R_{(i)} = r, X_{(i)} \in [t, t+h)\}} =$$

$$\lim_{h \downarrow 0} \frac{\mathbf{P}\{T_j \in [t, t+h), D_j = 1, T_l \geq t, l \in r \setminus \{j\}, T_l < t, l \notin r\}}{\sum_{s \in r} \mathbf{P}\{T_s \in [t, t+h), D_s = 1, T_l \geq t, l \in r \setminus \{s\}, T_l < t, l \notin r\}} =$$

$$\lim_{h \downarrow 0} \frac{\mathbf{P}\{T_j \in [t, t+h), D_j = 1\} \prod_{l \in r \setminus \{j\}} \mathbf{P}\{T_l \geq t\} \prod_{l \notin r} \mathbf{P}\{T_l < t\}}{\sum_{s \in r} \mathbf{P}\{T_s \in [t, t+h), D_s = 1\} \prod_{l \in r \setminus \{s\}} \mathbf{P}\{T_l \geq t\} \prod_{l \notin r} \mathbf{P}\{T_l < t\}} =$$

$$\lim_{h \downarrow 0} \frac{\mathbf{P}\{T_j \in [t, t+h), D_j = 1 | T_j \geq t\} \prod_{l \in r} \mathbf{P}\{T_l \geq t\}}{\sum_{s \in r} \mathbf{P}\{X_s \in [t, t+h), D_s = 1 | T_s \geq t\} \prod_{l \in r} \mathbf{P}\{T_l \geq t\}} =$$

$$\frac{\alpha_{cj}(t)}{\sum_{s \in r} \alpha_{cs}(t)} = \frac{\alpha_j(t)}{\sum_{s \in r} \alpha_s(t)} = \frac{e^{\beta^T \mathbf{Z}_j(t)}}{\sum_{s \in r} e^{\beta^T \mathbf{Z}_s(t)}}.$$

La fonction de vraisemblance partielle est déterminée comme le produit

$$L(\beta) = \prod_{i=1}^N p((i)|R(i); X(i)) = \prod_{i=1}^N \frac{e^{\beta^T \mathbf{z}_{(i)}(X(i))}}{\sum_{s \in R(i)} e^{\beta^T \mathbf{z}_s(X(i))}}.$$

Alors

$$\log L(\beta) = \sum_{i=1}^N \beta^T \mathbf{z}_{(i)}(X(i)) - \sum_{i=1}^N \log \sum_{s \in R(i)} e^{\beta^T \mathbf{z}_s(X(i))}$$

et

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^N \mathbf{z}_{(i)}(X(i)) - \sum_{i=1}^N \frac{\sum_{s \in R(i)} \mathbf{z}_s(X(i)) e^{\beta^T \mathbf{z}_s(X(i))}}{\sum_{s \in R(i)} e^{\beta^T \mathbf{z}_s(X(i))}}.$$

L'estimateur  $\hat{\beta}$  vérifie l'équation  $U(\hat{\beta}) = \mathbf{0}_p$ .

Alors on peut démontrer (voir la section suivante) que

$$\mathbf{E}N(t) = \mathbf{E} \int_0^t S^{(0)}(u, \beta) \alpha_0(u) du,$$

où

$$S^{(0)}(u, \beta) = \sum_{i=1}^n e^{\beta^T \mathbf{z}_i(u)} Y_i(u).$$

Cela implique l'estimateur  $\hat{A}_0(t)$  pour la fonction  $A_0(t) = \int_0^t \alpha_0(u) du$  :

$$N(t) = \int_0^t S^{(0)}(u, \hat{\beta}) d\hat{A}_0(u),$$

d'où

$$\hat{A}_0(t) = \int_0^t \frac{dN(u)}{S^{(0)}(u, \hat{\beta})}.$$

L'estimateur de la fonction

$$A_{\mathbf{z}(\cdot)}(t) = \int_0^t e^{\beta^T \mathbf{z}(u)} dA_0(u)$$

est

$$\hat{A}_{\mathbf{z}(\cdot)}(t) = \int_0^t e^{\hat{\beta}^T \mathbf{z}(u)} \frac{dN(u)}{S^{(0)}(u, \hat{\beta})},$$

et l'estimateur de la fonction de survie  $S_{\mathbf{z}(\cdot)}(t) = e^{-A_{\mathbf{z}(\cdot)}(t)}$  :

$$\hat{S}_{\mathbf{z}(\cdot)}(t) = e^{-\hat{A}_{\mathbf{z}(\cdot)}(t)}.$$

Les résultats obtenus nous permettent de construire des tests statistiques pour beaucoup de problèmes importants.

**Test d'homogénéité** Considérons l'hypothèse

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Sous cette hypothèse la loi de survie ne dépend pas des covariables. Elle peut être vérifiée en utilisant plusieurs tests.

a) Test du score

Sous  $H_0$  :

$$U(0) \approx N(0, \Sigma(0)),$$

où

$$U(0) = \sum_{i=1}^k \left\{ z(X_{(i)}) - \frac{\sum_{s \in R_i} z_s(X_{(i)})}{n_i} \right\},$$

$$\Sigma(0) = - \sum_{i=1}^k \left\{ \frac{\sum_{j \in R_i} z_{rj}(X_{(i)}) z_{sj}(X_{(i)})}{n_i} - \frac{\sum_{j \in R_i} z_{rj}(X_{(i)})}{n_i} \frac{\sum_{j \in R_i} z_{sj}(X_{(i)})}{n_i} \right\}$$

$n_i = Y(T_i^{(0)})$  est le nombre des sujets à risque juste avant  $T_i^{(0)}$ . Donc

$$U^T(0) \Sigma(0)^{-1} U(0) \approx \chi^2(p).$$

On rejette  $H_0$  au niveau de signification  $\alpha$ , si

$$U^T(0) \Sigma(0)^{-1} U(0) > \chi_{1-\alpha}^2(p).$$

b) Test de Wald

Sous  $H_0$

$$\hat{\beta} \approx N(0, \Sigma^{-1}(0)).$$

Donc

$$\hat{\beta}^T \Sigma(0) \hat{\beta} \approx \chi^2(p).$$

On rejette  $H_0$  au niveau de signification  $\alpha$ , si

$$\hat{\beta}^T \Sigma(0) \hat{\beta} > \chi_{1-\alpha}^2(p).$$

c) Test du rapport de vraisemblance

On peut montrer que

$$-2(\ln L(\beta) - \ln L(\hat{\beta})) \approx \chi^2(p).$$

Sous  $H_0$

$$-2(\ln L(0) - \ln L(\hat{\beta})) \approx \chi^2(p).$$

Notons que

$$\ln L(0) = - \sum_{i=1}^k \ln n_i,$$

$$\ln L(\hat{\beta}) = \sum_{i=1}^k \left\{ \hat{\beta}^T z_i(X_{(i)}) - \ln \sum_{s \in R_i} e^{\beta^T z_i(X_{(i)})} \right\}.$$

On rejette  $H_0$ , si

$$-2(\ln L(0) - \ln L(\hat{\beta})) > \chi_{1-\alpha}^2(p).$$

Si la seule caractéristique d'un individu est son appartenance à un groupe :

$$z(t) = \begin{cases} 1 & \text{pour les individus du 1 groupe} \\ 0 & \text{pour les individus du 2 groupe,} \end{cases}$$

le modèle de Cox a la forme

$$h(t | z) = \begin{cases} e^{\beta} h_0(t) & \text{pour les individus du 1 groupe} \\ h_0(t) & \text{pour les individus du 2 groupe.} \end{cases}$$

Dans ce cas l'hypothèse  $H_0 : \beta = 0$  signifie l'égalité des fonctions de risque de deux groupes qui est équivalent à l'égalité des fonctions de survie. Donc les tests du score, de Wald et du rapport de vraisemblance vérifient l'hypothèse de l'égalité des lois des deux groupes.

### Modèle stratifié

Supposons qu'on étudie l'effet des sous covariables  $z^{(s)} = (z_1, \dots, z_s)$  du vecteur des covariables  $z^{(p)} = (z_1, \dots, z_p)$  ( $p > s$ ) sur la survie, mais le modèle de Cox n'est pas vérifié par rapport à  $z^{(p)}$ . Parfois la modification suivante du modèle peut être utile.

Supposons que la région des valeurs de  $z_{s+1}, \dots, z_p$  est divisée en  $q$  strates et pour des sujets de  $j$ -ème strate le modèle de Cox est vérifié :

$$h_j(t | z^{(s)}) = e^{(\beta^{(s)})^T z^{(s)}(t)} h_{0j}(t) \quad (j = 1 \dots q).$$

Pour chaque strate la fonction de risque de base est différente mais l'effet des covariables  $z^{(s)}$  est le même pour toutes strates.

Pour estimer  $\beta$ , on commence par la vraisemblance partielle  $L_j$  à l'intérieur de chaque strate.

La vraisemblance partielle pour tous les sujets est le produit de toutes les vraisemblances :

$$L(\beta^{(s)}) = \prod_{j=1}^q L_j.$$

### Test graphique du modèle

Si des covariables sont constantes en temps, alors sous le modèle de Cox

$$H(t | z) = -\ln S(t | z) = e^{\beta^T z} H_0(t)$$

et donc

$$\ln H(t | z) = \beta^T z + \ln H_0(t).$$

Sous des valeurs différents de  $z$  les graphes des fonctions  $\ln H(t | z)$  sont parallèles. Donc, si  $z$  est discrète avec valeurs  $z^{(1)}, \dots, z^{(s)}$ , alors on considère les graphes des estimateurs

$$\ln \hat{H}(t | z^{(j)}) \quad (j = 1, \dots, s)$$

Sous le modèle de Cox ces graphes sont approximativement parallèles.

**Test de l'hypothèse  $H_l : \beta_{l+1} = \dots = \beta_p = 0$**

Considérons le problème de la vérification de l'hypothèse

$$H_l : \beta_{l+1} = \dots = \beta_p = 0,$$

où  $l = 1, \dots, p-1$ . Sous  $H_l$  les covariables  $z_{l+1}, \dots, z_p$  n'améliorent pas la prédiction. Donc si  $H_l$  est vérifié, on peut exclure ces covariables du modèle.

a) Test du rapport de vraisemblance

Soient

$$h(t | z^{(l)}) = e^{(\beta^{(l)})^T z^{(l)}(t)} h_0(t)$$

et

$$h(t | z^{(p)}) = e^{(\beta^{(p)})^T z^{(p)}(t)} h_0(t)$$

les modèles de Cox avec  $l$  et  $p$  covariables, respectivement. Alors

$$\begin{aligned} -2(\ln L_l(\beta^{(l)}) - \ln L_l(\hat{\beta}^{(l)})) &\approx \chi^2(l), \\ -2(\ln L_p(\beta^{(p)}) - \ln L_p(\hat{\beta}^{(p)})) &\approx \chi^2(p) \end{aligned}$$

Sous  $H_l$  :

$$L_l(\beta^{(l)}) = L_p((\beta^{(l)}, 0))$$

et la différence

$$\mathcal{L}_{l,p} = -2(\ln L_l(\beta^{(l)}) - \ln L_p(\hat{\beta}^{(p)})) \approx \chi^2(p-l),$$

Donc  $H_l$  est rejetée si

$$\mathcal{L}_{l,p} > \chi_{1-\alpha}^2(p-l).$$

L'hypothèse la plus intéressante de point de vue pratique est

$$H_{p-1} : \beta_p = 0.$$

Elle signifie que le modèle avec  $(p-1)$  covariables  $z_1, \dots, z_{p-1}$  donne la même prédiction que le modèle avec  $(p)$  covariables  $z_1, \dots, z_p$ , i.e. la covariable  $z_p$  peut être exclue du modèle.

L'hypothèse  $H_{p-1}$  est rejetée, si

$$\mathcal{L}_{p-1,p} > \chi_{1-\alpha}^2(1).$$

b) Test de Wald

On écrit l'inverse de la matrice d'information de Fisher sous la forme

$$\Sigma^{-1}(\beta) =$$

où  $A_{11}(\beta)$  et  $A_{22}(\beta)$  ont les dimensions  $l \times l$  et  $(p-l) \times (p-l)$ . Alors

$$(\hat{\beta}_{l+1}, \dots, \hat{\beta}_p) \approx N_{p-l}((\beta_{l+1}, \dots, \beta_p), A_{22}(\beta)).$$

Sous  $H_l$  :

$$W_{l,p} = (\hat{\beta}_{l+1}, \dots, \hat{\beta}_p)^T A_{22}^{-1}(\hat{\beta}_1, \dots, \hat{\beta}_l, 0, \dots, 0) (\hat{\beta}_{l+1}, \dots, \hat{\beta}_p) \approx \chi^2(p-l).$$

L'hypothèse  $H_l : \beta_{l+1} = \dots = \beta_p = 0$  est rejetée, si

$$W_{l,p} > \chi_{1-\alpha}^2(p-l)$$

Si  $l = p-1$ , alors

$$W_{p-1,p} = \hat{\beta}_p^2 / A_{22}(\hat{\beta}_1, \dots, \hat{\beta}_{p-1}, 0)$$

et l'hypothèse  $H_{p-1} : \beta_p = 0$  est rejetée, si

$$W_{p-1,p} > \chi_{1-\alpha}^2(1).$$

## 6.10 Processus de comptage et l'estimation non paramétrique

Soient  $X$  et  $C$  la durée de vie et le moment de censure, respectivement,

$$T = X \wedge C, \quad \delta = I(X \leq C), \quad N(t) = I(T \leq t, \delta = 1), \quad Y(t) = I(T \geq t).$$

$N(t)$  est le nombre des pannes observées dans l'intervalle  $[0, \tau]$ ,  $Y(t)$  est le nombre des unités à risque au moment  $t-$ .  $N(t)$  et  $Y(t)$  peuvent prendre des valeurs 0 et 1. On suppose que la variable aléatoire  $X$  est absolument continue et pour tout  $t$  tel que  $\mathbf{P}\{T \geq t\} > 0$  il existe la limite

$$\alpha_c(t) = \lim_{h \downarrow 0} \frac{\mathbf{P}\{T \in [t, t+h], \delta = 1 | T \geq t\}}{h}.$$

$\alpha_c(t)$  montre le risque de panne après ou au moment  $t$  sachant que une unité était à risque (pas censurée et pas en panne) juste avant le moment  $t$ .

On dit que la censure est indépendante, si

$$\alpha_c(t) = \alpha(t) = \lim_{h \downarrow 0} \frac{\mathbf{P}\{X \in [t, t+h] | X \geq t\}}{h}$$

pour tous  $t : \mathbf{P}\{T \geq t\} > 0$ .

Donc la censure ne influence pas le risque de panne d'une unité qui est "à risque".

Notons que

$$\begin{aligned} \alpha_c(t) &= \lim_{h \downarrow 0} \frac{\mathbf{P}\{t \leq X < t+h, X \leq C\}}{h \mathbf{P}\{X \geq t, C \geq t\}} = \\ \lim_{h \downarrow 0} \frac{\mathbf{P}\{X \leq C | t \leq X < t+h\} \mathbf{P}\{t \leq X < t+h\}}{h \mathbf{P}\{X \geq t, C \geq t\}} &= \\ \frac{\mathbf{P}\{C \geq X | X = t\} f_X(t)}{\mathbf{P}\{X \geq t, C \geq t\}} &= \frac{f_X(t)}{S_X(t)}. \end{aligned}$$

Donc l'égalité  $\alpha_c(t) = \alpha(t)$  est équivalente à l'égalité

$$\mathbf{P}\{C \geq t | X = t\} = \frac{\mathbf{P}\{X \geq t, C \geq t\}}{S_X(t)}.$$

Si  $X$  et  $C$  sont indépendantes, cette égalité est évidemment vérifiée. De l'autre côté on peut faire aussi une remarque intéressante :

$$\begin{aligned} \alpha_c(t) &= \lim_{h \downarrow 0} \frac{\mathbf{P}\{t \leq X < t+h, C \geq t\}}{h \cdot \mathbf{P}\{X \geq t, C \geq t\}} = \\ &= \frac{1}{\mathbf{P}\{X \geq t, C \geq t\}} \frac{\partial}{\partial s} [\mathbf{P}\{X \geq s, C \geq t\}] |_{s=t}. \end{aligned}$$

**Exemple.** Soit le vecteur  $(X, C)$  ait une loi exponentielle de trois paramètres  $\lambda > 0, \mu > 0, \theta > 0$  :

$$\mathbf{P}\{X \geq t, C \geq s\} = \exp(-\lambda t - \mu s - \theta t s), \quad t > 0, \quad s > 0,$$

d'où on tire que  $X$  suit une loi exponentielle de paramètre  $\lambda$ ,

$$\mathbf{P}\{X \geq t\} = \exp(-\lambda t),$$

et donc  $\alpha(t) = \lambda$ . De l'autre côté en utilisant la dernière remarque on trouve que

$$\alpha_c(t) = -\frac{1}{\mathbf{P}\{X \geq t, C \geq t\}} \frac{\partial}{\partial s} [\mathbf{P}\{X \geq s, C \geq t\}]|_{s=t} = \lambda + \theta t,$$

et donc on voit que dans cet exemple la censure n'est indépendante.

Notons

$$M(t) = N(t) - \int_0^t Y(u)\alpha(u)du.$$

**Proposition.** *Si la censure est indépendante, alors  $\mathbf{E}M(t) = 0$  pour tout  $t$  tel que  $\mathbf{P}\{T \geq t\} > 0$ .*

*Preuve.* L'égalité

$$\mathbf{P}\{C \geq t | X = t\} = \frac{\mathbf{P}\{X \geq t, C \geq t\}}{S_X(t)}.$$

implique

$$\begin{aligned} \mathbf{E}M(t) &= \mathbf{E}N(t) - \int_0^t \mathbf{E}Y(u)\alpha(u)du = \\ &= \mathbf{P}\{X \leq t, X \leq C\} - \int_0^t \mathbf{P}\{X \geq u, C \geq u\}\alpha(u)du = \\ &= \int_0^t \mathbf{P}\{C \geq u | X = u\}f_X(u)du - \int_0^t \mathbf{P}\{C \geq u | X = u\}S_X(u)\alpha(u)du = 0. \end{aligned}$$

La proposition est démontrée.

De plus on peut montrer le processus  $M(t)$  est une martingale par rapport à la filtration  $\mathcal{F}_t, t \geq 0$ , où  $\mathcal{F}_t$  est la  $\sigma$ -algèbre engendrée par les processus  $N(t)$  et  $Y(t)$  :

$$\mathcal{F}_t = \sigma\{N(s), Y(s) : 0 \leq s \leq t\}.$$

Dans ce cas on a :

$$\mathbf{E}\{M(t) | \mathcal{F}_s\} = M(s), \quad \text{pour } t \geq s,$$

ou

$$\mathbf{E}\{N(t) - N(s) | \mathcal{F}_s\} = \mathbf{E}\left\{ \int_s^t Y(u)\alpha(u)du | \mathcal{F}_s \right\},$$

d'où on tire que

$$\begin{aligned} \lim_{h \downarrow 0} \frac{1}{h} \mathbf{E}\{N(t) - N(s) | \mathcal{F}_s\} &= \\ \lim_{h \downarrow 0} \mathbf{E}\left\{ \int_s^{s+h} Y(u)\alpha(u)du | \mathcal{F}_s \right\} &= \mathbf{E}\{Y(s)\alpha(s) | \mathcal{F}_s\} = Y(s)\alpha(s). \end{aligned}$$

Cette relation montre que le processus

$$\lambda(t) = Y(t)\alpha(t)$$

est l'intensité du processus de comptage  $N(t)$ . Il représente le risque instantané observable au moment  $t$ . On dit aussi que l'intensité  $\lambda(t)$  est l'intensité multiplicative parce que dans

ce modèle elle est le produit d'un terme déterministe,  $\alpha(t)$ , et d'un processus  $Y(t)$ , qui est *prévisible*, c'est-à-dire sa valeur au moment  $t$  est connue si l'histoire dans l'intervalle  $[0, t[$  :

$$\mathcal{F}_{t-} = \sigma\{N(s), Y(s) : 0 \leq s < t\}$$

est connue.

Nous allons appliquer ces résultats dans la situation quand on observe  $n$  individus. Notons  $X_i$  et  $C_i$  les durées de survie et les temps de censures. Posons

$$T_i = X_i \wedge C_i, \quad D_i = \mathbf{1}\{X_i \leq C_i\}.$$

On a un échantillon  $(T_i, D_i)$ ,  $(i = 1, \dots, n)$ .

Supposons que la *censure est indépendante* pour chaque individu et que les variables aléatoires  $X_1, \dots, X_n$  sont absolument continues.

Notons

$$N_i(t) = I\{T_i \leq t, D_i = 1\}, \quad Y_i(t) = I\{T_i \geq t\},$$

$$N(t) = \sum_{i=1}^n N_i(t), \quad Y(t) = \sum_{i=1}^n Y_i(t).$$

$N(t)$  est un *processus de comptage* du nombre de défaillances observées sur  $[0, t]$  par sa valeur à l'instant  $t$ . Le processus  $N(t)$  est un processus *cadlag* : ses trajectoires sont des fonctions continues à droite et limitées à gauche. Enfin, le processus  $Y(t)$  représente le nombre des sujets à "*risque*" juste avant l'instant  $t$ , i.e.  $Y(t)$  montre le nombre de données restant encore en vie.

On introduit la *filtration*  $\mathcal{F}_t$  engendrée par tous les processus  $N_i(s)$  et  $Y_i(s)$  :

$$\mathcal{F}_t = \sigma\{N_i(s), Y_i(s) : 0 \leq s \leq t \quad (i = 1, \dots, n)\},$$

qui représente l'histoire des défaillances et des censures observées jusqu'à l'instant  $t$ . Notons

$$\Lambda_i(t) = \int_0^t \lambda_i(s) ds.$$

Parce que

$$M_i(t) = N_i(t) - \Lambda_i(t)$$

est une martingale avec  $\mathbf{E}\{M_i(t)\} = 0$ , on dit que  $\Lambda_i(t)$  est le *compensateur* du processus de comptage  $N_i(t)$ .

De même le processus

$$\Lambda(t) = \int_0^t \lambda(s) ds = \int_0^t Y(s) \alpha(s) ds = \int_0^t Y(s) dA(s)$$

est l'*intensité cumulée* du processus de comptage  $N(t)$ , où

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t),$$

d'où on obtient la *décomposition de Doob-Meyer* pour le processus  $N(t)$  :

$$N(t) = \Lambda(t) + M(t),$$

où  $M(t) = \sum_{i=1}^n M_i(t)$  est une  $\mathcal{F}_t$ -martingale,

$$\mathbf{E}\{M(t)|\mathcal{F}_s\} = M(s).$$

On dit que  $\Lambda(t)$  est le *le compensateur* du processus de comptage  $N(t)$ . Introduisons le processus

$$J(t) = I_{\{Y(t)>0\}}, \quad t > 0.$$

Pour estimer le taux de panne cumulé  $A(t)$  on utilise la méthode des moments. Parce que

$$\mathbf{E}\{N(t) - \int_0^t Y(s)dA(s)\} = 0,$$

on en tire que pour trouver l'estimateur  $\hat{A}_n(t)$  il nous faut résoudre l'équation suivante :

$$dN(t) - Y(t) \cdot dA(t) = 0,$$

d'où on obtient l'équation

$$dA(t) = J(t) \cdot \frac{dN(t)}{Y(t)},$$

ce qui nous donne le *fameux estimateur de Nelson-Aalen* :

$$\hat{A}_n(t) = \int_0^t J(u) \frac{dN(u)}{Y(u)} = \int_0^{t \wedge \tau} \frac{dN(u)}{Y(u)}$$

où  $\tau = \max X_i$ . Pour étudier les propriétés de l'estimateur Nelson-Aalen on utilise la relation suivante :

$$\begin{aligned} \hat{A}_n(t) - A(t) &= \int_0^t \left\{ J(u) \frac{dN(u)}{Y(u)} - J(u)dA(u) \right\} = \\ &= \int_0^t J(u) \frac{dN(u) - Y(u)dA(u)}{Y(u)} = \int_0^t J(u) \frac{dM(u)}{Y(u)}, \end{aligned}$$

où  $M(t)$  est la martingale définie plus haut, et donc  $\hat{A}_n(t) - A(t)$  est une  $\mathcal{F}_t$ -martingale, et donc pour nos études nous pouvons appliquer les résultats de R. Rebolledo (Central Limit Theorems for Local Martingales, 1984).

On va présenter l'estimateur de Nelson-Aalen en terme d'une somme.

Soit  $T_{(1)} < T_{(2)} < \dots < T_{(n)}$  la suite des instants où a lieu un événement (mort ou censure). A chaque instant  $T_{(i)}$  est observée  $D_{(i)}$  - *la nature de l'événement* :  $D_{(i)} = 1$ , si c'est une mort,  $D_{(i)} = 0$ , si c'est une censure. Il est évident que

$$Y(T_{(i)}) = n - i + 1,$$

d'où on tire l'estimateur de Nelson pour le taux de hazard cumulé  $A(t)$  :

$$\hat{A}_n(t) = \sum_{i: T_{(i)} \leq t} \frac{D_{(i)}}{n - i + 1} = \sum_{i: T_i \leq t} \frac{D_i}{n - i + 1}.$$

Ayant l'estimateur d'Aalen-Nelson pour le risque cumulé  $A(t)$  on peut facilement obtenir le *product-limite* (Kaplan-Meier) estimateur  $\hat{S}_n(t)$  pour la fonction de survie  $S(t) = \exp\{-A(t)\}$  :

$$\hat{S}_n(t) = \prod_{0 < s < t} \left( 1 - \frac{\Delta N(s)}{Y(s)} \right),$$

où  $\Delta N(t) = N(t) - N(t^-)$  est un processus  $\mathcal{F}_t$ -prévisible.

Pour obtenir cette formule on note d'abord que de l'équation

$$dS(t) = -S(t) dA(t), \quad S(0) = 1,$$

il suit que

$$S(t) = 1 - \int_0^t S(u) dA(u) = 1 - \int_0^t S(u-) dA(u).$$

Puisque

$$d\hat{A}_n(t) = \frac{dN(t)}{Y(t)}$$

on en tire que on a :

$$\hat{S}_n(t) = 1 - \int_0^t \frac{S(u-)}{Y(u)} dN(u)$$

et

$$d\hat{S}_n(t) = -\frac{\hat{S}_n(t-)}{Y(t)} dN(t).$$

Donc

$$\hat{S}_n(t-) - \hat{S}_n(t) = \int_0^t \frac{S_n(u-)}{Y(u)} dN(u) - \int_0^{t-} \frac{S_n(u-)}{Y(u)} dN(u) = \frac{\hat{S}_n(t-)}{Y(t)} \Delta N(t),$$

d'où on tire que

$$\hat{S}_n(t) = \hat{S}_n(t-) \left( 1 - \frac{\Delta N(t)}{Y(t)} \right), \quad \hat{S}_n(0) = 1,$$

et par la suite on obtient la formule de Kaplan-Meier :

$$\hat{S}_n(t) = \prod_{0 < s < t} \left( 1 - \frac{\Delta N(s)}{Y(s)} \right).$$

Le théorème suivant permet d'étudier les propriétés asymptotiques de l'estimateur de Kaplan-Meier.

**Théorème 4.** Si  $S(t) > 0$  alors

$$\frac{\hat{S}_n(t)}{S(t)} = 1 - \int_0^t \frac{\hat{S}_n(u-)}{S(u)Y(u)} dM(u).$$

*Démonstration.* On remarque d'abord que

$$\int_0^t u(s-) dv(s) = u(t)v(t) - u(0)v(0) - \int_0^t v(s) du(s).$$

En utilisant cette relation on trouve que

$$\int_0^t \hat{S}_n(u-) d\frac{1}{S(u)} = \frac{\hat{S}_n(t)}{S(t)} - \frac{\hat{S}_n(0)}{S(0)} - \int_0^t \frac{1}{\hat{S}_n(u)} d\hat{S}_n(u-).$$

Donc

$$\frac{\hat{S}_n(t)}{S(t)} = 1 - \int_0^t \frac{\hat{S}_n(u-)}{S^2(u)} dS(u) + \int_0^t \frac{1}{S(u)} d\hat{S}_n(u-).$$

Puisque on a

$$dS(t) = -S(t)dA(t), \quad d\hat{S}_n(t) = -\frac{\hat{S}_n(t-)}{Y(t)}dN(t),$$

et

$$dN(t) = dM(t) + Y(t)dA(t),$$

on trouve que

$$\begin{aligned} \frac{\hat{S}_n(t)}{S(t)} &= 1 + \int_0^t \frac{\hat{S}_n(u-)}{S(u)}dA(u) - \int_0^t \frac{\hat{S}_n(u-)}{S(u)Y(u)}dN(u) = \\ &= 1 - \int_0^t \frac{\hat{S}_n(u-)}{S(u)Y(u)}dM(u). \end{aligned}$$

Le théorème est démontré.

Ce théorème nous permet de calculer

$$\mathbf{Var} \hat{S}_n(t) = \mathbf{E} \left\{ S(t) \int_0^t \frac{\hat{S}_n(u-)}{S(u)Y(u)} I\{N(u) > 0\} dM(u) \right\}^2,$$

d'où on obtient son estimateur

$$\hat{\mathbf{Var}} \hat{S}_n(t) = \hat{S}_n^2(t) \int_0^t \frac{dN(u)}{(Y(u) - \Delta N(u))Y(u)},$$

connu comme la *formule de Greenwood*.

## 11. Comparaison des fonctions de survie

Supposons qu'on a deux groupes des individus (unités). Le  $i$ -ème groupe a  $n_i$  individus. Pour le premier groupe on a un échantillon

$$(X_{11}, \delta_{11}), \dots, (X_{1n_1}, \delta_{1n_1}),$$

où en forme équivalente

$$(N_{11}(t), Y_{11}(t), t \geq 0), \dots, (N_{1n_1}(t), Y_{1n_1}(t), t \geq 0).$$

Pour le deuxième groupe on observe

$$(X_{21}, \delta_{21}), \dots, (X_{2n_2}, \delta_{2n_2}),$$

où

$$(N_{21}(t), Y_{21}(t), t \geq 0), \dots, (N_{2n_2}(t), Y_{2n_2}(t), t \geq 0).$$

Soit  $S_i(t)$  la fonction de survie du  $i$ -ème groupe. On va tester l'hypothèse

$$\mathcal{H}_0 : S_1(t) = S_2(t) \quad \forall t \geq 0.$$

Notons  $H_i(t) = -\ln S_i(t)$  la fonction de risque cumulé pour le  $i$ -ème groupe.

L'estimateur de Nelson-Aalen pour  $H_i(t)$  est

$$\hat{H}_i(t) = \int_0^t \frac{dN_i(u)}{Y_i(u)}.$$

Si l'hypothèse  $\mathcal{H}_0$  est vérifiée, alors les estimateurs  $\hat{H}_1(t)$  et  $\hat{H}_2(t)$  doivent être proches. Donc le test est basé sur la statistique

$$V = \int_0^\infty K(u) d(\hat{H}_1(u) - \hat{H}_2(u)) = \int_0^\infty K(u) \frac{dN_1(u)}{Y_1(u)} - \int_0^\infty K(u) \frac{dN_2(u)}{Y_2(u)},$$

où  $K(u)$  est le poids,

$$N_i(u) = \sum_{j=1}^{n_i} N_{ij}(u), \quad Y_i(u) = \sum_{j=1}^{n_i} Y_{ij}(u).$$

Sous l'hypothèse  $\mathcal{H}_0$  les valeurs de la statistique  $V$  sont dispersées autour de zéro. En choisissant des poids différents, on obtient des statistiques différentes :

1. Test de logrank (Cox, Mantel - Haenkel) :

$$K_L(u) = a_n \frac{Y_1(u)/n_1 \cdot Y_2(u)/n_2}{Y(u)/n},$$

$$\text{où } Y = Y_1 + Y_2, \quad n = n_1 + n_2, \quad a_n = \sqrt{\frac{n_1 n_2}{n}}.$$

2. Test de Tarone-Ware :

$$K_{TW}(u) = a_n \frac{Y_1(u)/n_1 \cdot Y_2(u)/n_2}{\sqrt{Y(u)/n}}.$$

3. Test de Gehan (généralisation du test de Wilcoxon) :

$$K_G(u) = a_n \frac{Y_1(u)}{n_1} \frac{Y_2(u)}{n_2}.$$

4. Test de Prentice :

$$K_P(u) = a_n \tilde{S}(u-) \frac{Y(u)}{Y(u) + 1},$$

où

$$\tilde{S}(u) = \prod_{v \leq u} \left( 1 - \frac{\Delta N(v)}{Y(v) + 1} \right), \quad N = N_1 + N_2, \quad \Delta N(u) = N(u) - N(u-).$$

5. Test de Efron :

$$K_E(u) = a_n \hat{S}_1(u-) \hat{S}_2(u-) \mathbf{1}_{\{Y_1(u)Y_2(u) > 0\}},$$

où  $\hat{S}_i$  est l'estimateur de Kaplan-Meier de  $S_i$ .

Si  $n_1$  et  $n_2$  sont grands, la loi de  $V$  est approchée par la loi normale :

$$V \approx N(0, \sigma^2),$$

et la variance  $\sigma^2$  est estimée par :

$$\hat{\sigma}^2 = \int_0^\infty \frac{K^2(u)}{Y_1(u)Y_2(u)} \left(1 - \frac{\Delta N(u) - 1}{Y(u) - 1}\right) dN(u) \xrightarrow{\mathbf{P}} \sigma^2, \quad \mathbf{E}\hat{\sigma}^2 = \sigma^2.$$

Donc

$$\frac{V^2}{\hat{\sigma}^2} \approx \chi^2(1)$$

et  $\mathcal{H}_0$  est rejetée au niveau de signification  $\alpha$  si

$$\frac{V^2}{\hat{\sigma}^2} > \chi_{1-\alpha}^2(1).$$

Des intégrals peuvent être écrites en terme des sommes :

$$V = \sum_{j=1}^{m_1} K(T_{1j}^0) \frac{d_{1j}}{n_{1j}} - \sum_{j=1}^{m_2} K(T_{2j}^0) \frac{d_{2j}}{n_{2j}},$$

où

$T_{i1}^0 < \dots < T_{im_i}^0$  sont des moments distincts des décès observés du  $i$ -ème groupe,

$d_{ij}$  est le nombre des décès au moment  $T_{ij}^0$  pour le  $i$ -ème groupe,

$n_{ij}$  - le nombre des individus à risque juste avant le moment  $T_{ij}^0$  pour le  $i$  - ème groupe.

Par exemple, pour le test de Gehan

$$V_G = \int_0^\infty K_G(u) \left( \frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)} \right) = \sum_{j=1}^{m_1} K_G(T_{1j}^0) \frac{d_{1j}}{n_{1j}} - \sum_{j=1}^{m_2} K_G(T_{2j}^0) \frac{d_{2j}}{n_{2j}} =$$

$$\frac{a_n}{n_1 n_2} \left( \sum_{j=1}^{m_1} n_{2j} d_{1j} - \sum_{j=1}^{m_2} n_{1j} d_{2j} \right).$$

Considérons une autre expression pour ce test. Notons  $T_1^* < \dots < T_m^*$  les moments des décès observés de tous  $n = n_1 + n_2$  individus,

$D_{ij}, N_{ij}$  les nombres des décès au moment  $T_j^*$  et les nombres des individus à risque juste avant  $T_j^*$  pour les individus de  $i$ -ème groupe,

$$D_j = D_{1j} + D_{2j}, \quad N_j = N_{1j} + N_{2j};$$

Ici  $D_j > 0$  mais il est possible que  $D_{1j} = 0$  ou  $D_{2j} = 0$ . Alors

$$V_G = \int_0^\infty K_G \left\{ \frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)} \right\} =$$

$$\frac{a_n}{n_1 n_2} \left( \int_0^\infty Y_2(u) dN_1(u) - \int_0^\infty Y_1(u) dN_2(u) \right) =$$

$$\frac{a_n}{n_1 n_2} \left( \sum_{j=1}^m N_{2j} D_{1j} - \sum_{j=1}^m N_{1j} D_{2j} \right) =$$

$$\begin{aligned} & \frac{a_n}{n_1 n_2} \sum_{j=1}^m (N_{2j} D_{1j} + N_{1j} D_{1j} - N_{1j} D_{2j}) = \\ & \frac{a_n}{n_1 n_2} \sum_{j=1}^m (N_j D_{1j} - N_{1j} D_j) = \frac{a_n}{n_1 n_2} \sum_{j=1}^m N_j \left( D_{1j} - D_j \frac{N_{1j}}{N_j} \right). \end{aligned}$$

Dans la dernière formule  $D_{1j}$  représente le nombre des décès du premier groupe au moment  $T_j^*$ ,  $E_{1j} = D_j \frac{N_{1j}}{N_j}$  représente sous l'hypothèse  $\mathcal{H}_0$  le nombre attendu des décès du premier groupe sachant que le nombre des décès de tous les deux groupes est  $D_j$  et la proportion des individus à risque juste avant  $T_j^*$  est  $\frac{N_{1j}}{N_j}$ . Donc

$$V_G = \frac{a_n}{n_1 n_2} \sum_{j=1}^m N_j (D_{1j} - E_{1j}).$$

Si des autres statistiques sont considérées, les poids associés à  $(D_{1j} - E_{1j})$  sont différents :

$$\begin{aligned} V_L &= a_n \frac{n}{n_1 n_2} \sum_{j=1}^m (D_{1j} - E_{1j}); \\ V_{TW} &= a_n \frac{\sqrt{n}}{n_1 n_2} \sum_{j=1}^m \sqrt{N_j} (D_{1j} - E_{1j}); \\ V_P &= a_n \sum_{j=1}^m \tilde{S}(T_j^0 -) \frac{N_j^2}{(N_j + 1) N_{1j} N_{2j}} (D_{1j} - E_{1j}); \\ V_E &= a_n \sum_{j=1}^m \frac{\hat{S}_1(T_j^0 -) \hat{S}_2(T_j^0 -)}{N_{1j} N_{2j}} N_j \mathbf{1}_{\{N_{1j} N_{2j} > 0\}}. \end{aligned}$$

L'estimateur de la variance  $\hat{\sigma}^2$  de la statistique  $V$  peut être donnée en terme des sommes :

$$\hat{\sigma}^2 = \sum_{j=1}^m \frac{K^2(T_j^0)}{N_{1j} N_{2j}} \left( 1 - \frac{D_j - 1}{N_j - 1} \right) D_j.$$

## 6.11 Estimation dans des expériences accélérées

### 6.11.1 Modèles de vie accélérée

Supposons que des unités sont très fiables et il n'y a pas de possibilité d'obtenir des pannes pendant le temps  $t$  donné par expérience. Dans ce cas on effectue des expériences sous des stress qui sont supérieurs au stress usuel. On appelle ces expériences *expériences accélérées*. L'application des stress accélérés raccourci la durée de vie des unités et des pannes peuvent se produire pendant le temps  $t$ . Des exemples des stress : température, voltage, poids etc.

Dans le cas général des stress  $x$  peuvent varier en temps et peuvent être multidimensionnels :

$$x = x(\tau), \tau \geq 0, \quad \text{où } x : [0, \infty[ \rightarrow B \subset \mathbb{R}^m.$$

Supposons que la durée de vie  $T_{x(\cdot)}$  sous le stress  $x(\cdot)$  est la variable aléatoire non-négative absolument continue de fonction de survie

$$S_{x(\cdot)}(t) = \mathbf{P}\{T_{x(\cdot)} > t\}, \quad t \geq 0.$$

Considérons un ensemble des stress  $\mathcal{E}$ . Formellement, on dit qu'un stress  $x_1(\cdot)$  est supérieur à un stress  $x_0(\cdot)$ , si  $S_{x_0(\cdot)}(t) \geq S_{x_1(\cdot)}(t)$  pour tout  $t \geq 0$ .

Le but d'expériences accélérés est d'estimer la fiabilité des unités correspondante aux conditions usuelles  $x_0$  de fonctionnement en utilisant des données de ces expériences. La solution de ce problème exige construction des modèles qui déterminent de quelle façon la fonction de survie  $S_{x(\cdot)}$  ou une autre caractéristique (la densité, le taux de pannes, etc. ) change quand on change le stress  $x(\cdot)$ .

Soit  $f_{x(\cdot)}(t) = S_{x_0}^{-1} \circ S_{x(\cdot)}(t)$ , où  $x_0 \in \mathcal{E}$  est un stress usuel,  $S_{x_0}^{-1} = \inf\{s : S_{x_0}(s) \geq p\}$  est la fonction inverse de  $S_{x_0}$ . Alors pour tout  $x(\cdot) \in \mathcal{E}$

$$\mathbf{P}\{T_{x_0} \geq f_{x(\cdot)}(t)\} = \mathbf{P}\{T_{x(\cdot)} \geq t\}.$$

Pour tout  $x(\cdot) \in \mathcal{E}$  la probabilité de survivre jusqu'au moment  $t$  sous le stress  $x(\cdot)$  est la même que la probabilité de survivre jusqu'au moment  $f_{x(\cdot)}(t)$  sous le stress  $x_0(t)$ . Le nombre  $f_{x(\cdot)}(t)$  est appelé *la ressource utilisé sous le stress  $x(\cdot)$  jusqu'au moment  $t$* . Il est clair que  $f_{x(\cdot)}(0) = 0$  pour tout  $x(\cdot) \in \mathcal{E}$ . La variable aléatoire  $R = f_{x(\cdot)}(T_{x(\cdot)})$  est la ressource utilisé sous le stress  $x(\cdot)$  jusqu'au la panne. La fonction de survie de  $R$  est  $S_{x_0}$  et ne dépend pas de  $x(\cdot)$ .

Le modèle de vie accélérée (VA) est vérifié sur  $\mathcal{E}$  si'il existe une fonction  $r : \mathcal{E} \rightarrow \mathbb{R}^+$  telle que pour tout  $x(\cdot) \in \mathcal{E}$

$$\frac{d}{dt} f_{x(\cdot)}(t) = r[x(t)]. \quad (1)$$

Le modèle VA signifie que la vitesse d'utilisation de la ressource au moment  $t$  ne dépend que de la valeur du stress appliqué au moment  $t$ . La formule (1) implique que

$$S_{x(\cdot)}(t) = S_{x_0} \left( \int_0^t r[x(\tau)] d\tau \right). \quad (2)$$

Nous nous bornons au modèle (2). Pour nombreuses généralisations et applications voir Bagdonavičius & Nikulin (1995, 1997, 1998), voir aussi L.Gerville-Réache & V.Nikoulina (1998), V. Bagdonavičius, L.Gerville-Réache, V.Nikoulina & M.Nikulin (2000).

Dans le cas  $x(\tau) \equiv x = \text{const}$  le modèle (2) implique

$$S_x(t) = S_{x_0}(r(x)t), \quad (3)$$

donc le stress ne change que l'échelle. Notons que  $r(x_0) = 1$ .

Considérons deux plans d'expériences accélérées possibles.

Le premier plan : Soient  $x_1, \dots, x_k$  des stress accélérés :  $x_0 < x_1 < \dots < x_k$  et  $x_0$  le stress usuel.  $k$  groupes d'unités sont observés. On teste le  $i$ ème groupe sous le stress  $x_i$ . Donc le stress usuel  $x_0$  n'est pas utilisé.

Le deuxième plan peut être utilisé si le coefficient de variation de la durée de vie sous le stress usuel  $x_0$  n'est pas très grand et la plupart des pannes se produisent dans un certain intervalle  $[s_1, s_2]$ , où  $s_1$  est supérieur au temps  $t$  donné pour l'expérience. Alors on peut faire deux expériences : l'une sous un stress accéléré  $x_1$  et une autre sous le stress  $x_1$  jusqu'au moment  $t_1 < t$ , en remplaçant le stress  $x_1$  par le stress usuel  $x_0$  au moment  $t_1$ . Des unités utilisent beaucoup de ses "ressources" jusqu'au moment  $t_1$  sous le stress  $x_1$  donc même sous le stress usuel  $x_0$  on peut obtenir des pannes dans l'intervalle  $[t_1, t]$

Dans le cas du premier plan d'expériences on n'a pas d'expérience sous le stress usuel  $x_0$ . Si la fonction  $r(t)$  est complètement inconnue, la fonction  $S_{x_0}$  ne peut pas être estimée même si l'on connaît la famille des distributions à laquelle elle appartient.

Par exemple, si  $S_{x_0}(t) = e^{-(t/\theta)^\alpha}$ , alors

$$S_x(t) = \exp \left[ - \left( \frac{r(x)}{\theta} t \right)^\alpha \right].$$

Les paramètres  $\alpha, \frac{r(x_1)}{\theta}, \dots, \frac{r(x_k)}{\theta}$  et les fonctions  $S_{x_1}, \dots, S_{x_k}$  peuvent être estimés mais puisque  $r$  est complètement inconnu,  $r(x_0)$  et donc  $S_{x_0}(t)$  ne peuvent pas être estimés.

Donc la fonction  $r$  doit être choisie dans une certaine classe des fonctions.

Considérons choix possible de la fonction  $r(x)$ . Si le modèle (3) est vérifié sur un ensemble des stress  $\mathcal{E}$ , alors pour tous  $x_1, x_2 \in \mathcal{E}$

$$S_{x_2}(t) = S_{x_1}(\rho(x_1, x_2)t),$$

où  $\rho(x_1, x_2) = r(x_2)/r(x_1)$  montre comment l'échelle de distribution change quand le stress  $x_2$  est utilisé au lieu du stress  $x_1$ . Il est évident que  $\rho(x, x) = 1$ . Supposons que des stress  $x \in \mathcal{E}$  sont unidimensionnels :  $\mathcal{E} \subset \mathbb{R}$ . Le taux de changement d'échelle est déterminé par la dérivée

$$\delta(x) = \lim_{\Delta x \rightarrow 0} \frac{\rho(x, x + \Delta x) - \rho(x, x)}{\Delta x} = [\log r(x)]'.$$

Donc pour tout  $x \in \mathcal{E}$

$$r(x) = \exp \left\{ \int_{x_0}^x \delta(v) dv \right\}.$$

Supposons que  $\delta(x)$  est proportionnelle à une fonction connue  $u(x)$  de stress :

$$\delta(x) = \alpha u(x), \quad \alpha > 0. \quad (4)$$

Alors

$$r(x) = e^{\beta_0 + \beta_1 z(x)},$$

où  $z(x)$  est une fonction connue,  $\beta_0, \beta_1$  - des paramètres inconnus.

Des cas particuliers :

**a).**  $\delta(x) = \alpha$ , i.e. le taux de changement de l'échelle est constant. Alors

$$r(x) = e^{\beta_0 + \beta_1 x},$$

où  $\beta_1 > 0$ . C'est le modèle loglinéaire. Ce modèle est appliqué pour analyser des données de fatigue, testant divers composants électroniques.

b).  $\delta(x) = \alpha/x$ , alors

$$r(x) = e^{\beta_0 + \beta_1 \log x} = \alpha x^{\beta_1},$$

où  $\beta_1 > 0$ . C'est le modèle de la règle de puissance ("power rule model").

Ce modèle est appliqué quand le stress est le voltage, la charge mécanique.

c).  $\delta(x) = \alpha/x^2$ , alors

$$r(x) = e^{\beta_0 + \beta_1/x} = \alpha e^{\beta_1/x},$$

où  $\beta_1 < 0$ . C'est le modèle d'Arrhénius.

Ce modèle est largement appliqué quand le stress est la température.

S'il n'est pas clair laquelle de ces trois paramétrisations de  $r(x)$  à choisir, on peut considérer la plus large paramétrisation :

$$\delta(x) = \alpha x^\gamma,$$

qui est équivalente à

$$r(x) = \begin{cases} e^{\beta_0 + \beta_1(x^\varepsilon - 1)/\varepsilon}, & \text{si } \varepsilon \neq 0, \\ e^{\beta_0 + \beta_1 \log x}, & \text{si } \varepsilon = 0. \end{cases}$$

Dans le cas du deuxième plan la paramétrisation de  $r$  n'est pas nécessaire. Si le premier groupe est testé sous le stress accéléré  $x_1$  et le deuxième groupe sous le stress

$$x_2(t) = \begin{cases} x_1, & 0 \leq \tau \leq t_1, \\ x_0, & t_1 < \tau \leq t_2, \end{cases}$$

alors

$$S_{x_1}(u) = S_{x_0}(ru),$$

$$S_{x_2(\cdot)}(u) = \begin{cases} S_{x_0}(ru), & 0 \leq u \leq t_1, \\ S_{x_0}(r(u \wedge t_1) + (u - t_1) \vee 0), & t_1 < u \leq t_2, \end{cases}$$

où  $r = r(x_1)/r(x_0)$ . Les fonctions  $S_{x_1}$  et  $S_{x_2(\cdot)}$  peuvent être toujours estimées. On verra plus tard que des estimateurs de  $r$  et consécutivement de  $S_{x_0}$  peuvent être obtenues même dans le cas quand la fonction  $S_{x_0}$  est complètement inconnue.

Le modèle (4) peut être généralisé, en supposant que  $\delta(x)$  est la combinaison linéaire des fonctions connues du stress :

$$\delta(x) = \sum_{i=1}^k \alpha_i u_i(x).$$

Dans ce cas

$$r(x) = \exp \left\{ \beta_0 + \sum_{i=1}^k \beta_i z_i(x) \right\},$$

où  $r_i(x)$  sont des fonctions du stress connus,  $\beta_0, \dots, \beta_k$  des paramètres inconnus (peut être pas tous).

**Exemple.**

1.  $\delta(x) = 1/x + \alpha/x^2$ .

Alors  $r(x) = e^{\beta_0 + \beta_1 \log x + \beta_2/x} = \alpha_1 x e^{\beta_2/x}$ , où  $\beta_1 = 1$ ,  $\beta_2 < 0$ . C'est le modèle d'Eyring, on l'applique souvent quand le stress est une température.

2.  $\delta(x) = \sum_{i=1}^k \alpha_i/x^i$ . Alors

$$r(x) = \exp \left\{ \beta_0 + \beta_1 \log x + \sum_{i=1}^{k-1} \beta_i/x^i \right\}.$$

C'est le modèle d'Eyring généralisé.

Le stress peut être multidimensionnel :  $x = (x_1, \dots, x_m)^T$ . Alors on considère des caractéristiques infinitésimales  $\delta_i(x)$  données par des égalités :

$$\delta_i(x) = \lim_{\Delta x_i \rightarrow 0} \frac{\rho(x, x + \Delta x_i e_i) - \rho(x, x)}{\Delta x_i} = \frac{\partial \log r(x)}{\partial x_i},$$

où  $e_i = (0, \dots, 1, \dots, 0)$ . L'unité est dans la  $i$ -ème coordonné.

Généralisant le cas unidimensionnel,  $\delta_i(x)$  peut être paramétrisé de façon suivant

$$\delta_i(x) = \sum_{j=1}^{k_i} \alpha_{ij} u_{ij}(x),$$

où  $u_{ij}(x)$  sont des fonctions connues,  $\alpha_{ij}$  -des constantes inconnues. Dans ce cas

$$r(x) = \exp\left\{\beta_0 + \sum_{i=1}^m \sum_{j=1}^{k_i} \beta_{ij} z_{ij}(x)\right\},$$

où  $z_{ij}(x)$  sont des fonctions connues,  $\beta_{ij}$  sont des constantes inconnues.

#### Exemples.

1.  $\delta_1(x) = 1/x_1 + (\alpha_{11} + \alpha_{12}x_2)/x_1^2$ ,  $\delta_2(x) = \alpha_{21} + \alpha_{22}/x_1$ .

C'est le modèle d'Eyring généralisé. On l'applique pour certains matériels des semi-conducteurs, quand  $x_1$  est la température et  $x_2$  est le voltage.

2.  $\delta_i(x) = \alpha_i u_i(x_i)$ ,

où  $u_i$  sont connues. Alors

$$r(x) = \exp\left\{\sum_{i=1}^m \alpha_i \int_{x_i^0}^{x_i} u_i(v) dv\right\} = \exp\left\{\beta_0 + \sum_{i=1}^m \beta_i z_i(x_i)\right\},$$

où  $z_j$  sont des fonctions connues. C'est le modèle d'Arrhénius généralisé.

Donc dans tous les cas considérés les modèles (2) et (3) peuvent être écrits sous la forme

$$S_{x(\cdot)}(t) = S_{x_0} \left( \int_0^t e^{\beta^T z(\tau)} d\tau \right), \quad (5)$$

ou

$$S_x(t) = S_{x_0} \left( e^{\beta^T z_t} \right), \quad (6)$$

où  $\beta = (\beta_0, \dots, \beta_m)^T$  est un vecteur des paramètres,

$$z(t) = (z_0(t), \dots, z_m(t))^T = (z_0(x(t)), \dots, z_m(x(t)))^T, \quad z = (z_0(x), \dots, z_m(x))^T$$

sont des vecteurs des fonctions connues du stress, la première composante  $z_0$  est égale à 1.

Ces modèles peuvent être considérés comme paramétriques, si la fonction  $S_{x_0}$  appartient à une certaine classe des répartitions, ou comme semiparamétriques si  $S_{x_0}$  est complètement inconnue.

## 6.11.2 Estimation paramétrique

On suppose, que le modèle (6) est considéré et le premier plan d'expériences est utilisé :  $k$  groupes d'unités sont observés ; on fixe la durée maximale d'expérience  $t_i$  du  $i$ -ème groupe et on teste ce groupe sous le stress accéléré  $x_i$  ( $i = 1, \dots, k$ ). Notons

$$z_{il} = z_l(x_i), \quad z^{(i)} = (z_{i0}, \dots, z_{im})^T \quad (i = 1, \dots, k; l = 0, \dots, m).$$

On suppose que  $S_{x_0}$  appartienne à une classe des répartitions

$$S_{x_0}(t) = S_0((t/\theta)^\nu), \quad (\theta, \nu > 0). \quad (7)$$

Par exemple, si

$$S_0(t) = e^{-t}, \quad (1+t)^{-1}, \quad 1 - \Phi(\ln t),$$

alors on obtient des classes des répartitions de Weibull, loglogistique, lognormale respectivement. Ici  $\Phi$  est la fonction de répartition de la loi normale standard. Donc le modèle (6) peut être écrit sous la forme :

$$S_x(t) = S\left(\frac{\ln t - \gamma^T z}{\sigma}\right), \quad t > 0,$$

où

$$S(u) = S_0(e^u), \quad u \in \mathbb{R}, \quad \sigma = 1/\nu, \quad \gamma = (\gamma_0, \dots, \gamma_m), \quad \gamma_0 = \ln \theta - \beta_0, \\ \gamma_l = -\beta_l \quad (l = 1, \dots, m).$$

Dans les cas des lois de Weibull, loglogistique et lognormale

$$S(u) = e^{-e^u}, \quad (1 + e^u)^{-1}, \quad 1 - \Phi(u)$$

respectivement.

Notons  $T_{ij}$  la durée de vie (pas nécessairement observée) de  $j$ ème unité du  $i$ ème groupe,

$$X_{ij} = \ln(T_{ij} \wedge t_i), \quad \delta_{ij} = I\{T_{ij} \leq t_i\}, \quad f(u) = -S'(u), \quad \lambda(u) = \frac{f(u)}{S(u)}.$$

La fonction de survie et la densité de  $\ln T_{ij}$  sont

$$S_i(u; \gamma, \sigma) = S\left(\frac{u - \gamma^T z^{(i)}}{\sigma}\right), \quad f_i(u; \gamma, \sigma) = \frac{1}{\sigma} f\left(\frac{u - \gamma^T z^{(i)}}{\sigma}\right), \quad u \in \mathbb{R}.$$

Donc la fonction de vraisemblance

$$L(\gamma, \sigma) = \prod_{i=1}^k \prod_{j=1}^{n_i} \left[ \frac{1}{\sigma} \lambda\left(\frac{X_{ij} - \gamma^T z^{(i)}}{\sigma}\right) \right]^{\delta_{ij}} S\left(\frac{X_{ij} - \gamma^T z^{(i)}}{\sigma}\right).$$

En dérivant par rapport à  $\gamma_l$  et  $\sigma$  la fonction  $\ln L(\gamma, \sigma)$ , on obtient

$$U_l(\gamma; \sigma) = \frac{\partial \ln L(\gamma, \sigma)}{\partial \gamma_l} = \frac{1}{\sigma} \sum_{i=1}^k z_{il} \sum_{j=1}^{n_i} a_{ij}(\gamma, \sigma), \quad (l = 1, \dots, m),$$

$$U_{m+1}(\gamma; \sigma) = \frac{\partial \ln L(\gamma, \sigma)}{\partial \sigma} = \frac{1}{\sigma} \sum_{i=1}^k \sum_{j=1}^{n_i} \{v_{ij}(\gamma, \sigma) a_{ij}(\gamma, \sigma) - \delta_{ij}\},$$

où

$$v_{ij}(\gamma, \sigma) = \frac{X_{ij} - \gamma^T z^{(i)}}{\sigma}, \quad a_{ij}(\gamma, \sigma) = \lambda(v_{ij}(\gamma, \sigma)) - \delta_{ij} (\ln \lambda)'(v_{ij}(\gamma, \sigma)).$$

Des estimateurs de maximum de vraisemblance  $\hat{\sigma}$ ,  $\hat{\gamma}$  peuvent être obtenus en résolvant le système d'équations

$$U_l(\gamma, \sigma) = 0 \quad (l = 1, \dots, m+1).$$

Notons

$$\mathbf{I}(\gamma, \sigma) = (I_{lk}(\gamma, \sigma))_{(m+1) \times (m+1)}$$

la matrice avec des éléments suivants :

$$I_{ls}(\gamma, \sigma) = -\frac{\partial^2 \ln L(\gamma, \sigma)}{\partial \gamma_l \partial \gamma_s} = \frac{1}{\sigma^2} \sum_{i=1}^k z_{il} z_{is} \sum_{j=1}^{n_i} c_{ij}(\gamma, \sigma), \quad l, s = 0, \dots, m;$$

$$I_{l, m+1}(\gamma, \sigma) = -\frac{\partial^2 \ln L(\gamma, \sigma)}{\partial \gamma_l \partial \sigma} = \frac{1}{\sigma} U_l(\gamma, \sigma) + \frac{1}{\sigma^2} \sum_{i=1}^k z_{il} \sum_{j=1}^{n_i} v_{ij}(\gamma, \sigma) c_{ij}(\gamma, \sigma), \quad l = 0, \dots, m;$$

$$I_{m+1, m+1}(\gamma, \sigma) = -\frac{\partial^2 \ln L(\gamma, \sigma)}{\partial \sigma^2} = \frac{2}{\sigma} U_{m+1}(\gamma, \sigma) + \frac{1}{\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (v_{ij}^2(\gamma, \sigma) c_{ij}(\gamma, \sigma) + \delta_{ij}),$$

où

$$c_{ij}(\gamma, \sigma) = \lambda'(v_{ij}(\gamma, \sigma)) - \delta_{ij} (\ln \lambda)''(v_{ij}(\gamma, \sigma)).$$

Si  $T_{X_0}$  suit les lois de Weibull, loglogistique ou lognormale, alors

$$\lambda(t) = e^t; (1 + e^{-t})^{-1}; \varphi(t)/(1 - \Phi(t)).$$

respectivement, où

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Si les estimateurs de maximum de vraisemblance  $\hat{\gamma}$  et  $\hat{\sigma}$  sont obtenus, alors l'estimateurs de la fonction de survie  $S_{X_0}$  et de la  $p$ -quantile  $t_p(x_0)$  sont

$$\hat{S}_{X_0}(t) = S\left(\frac{\ln t - \hat{\gamma}^T z^{(0)}}{\hat{\sigma}}\right), \quad \hat{t}_p(x_0) = e^{\hat{\gamma}^T z^{(0)}} [S_0^{-1}(1-p)]^{\hat{\sigma}}.$$

La loi asymptotique de  $(\hat{\gamma}, \hat{\sigma})^T$  quand  $n_i$  sont grands est approchée par la loi normale  $N((\gamma, \sigma)^T, \mathbf{\Sigma}(\gamma, \sigma))$  et la matrice de covariance  $\mathbf{\Sigma}(\gamma, \sigma)$  peut être estimé par

$$\mathbf{I}^{-1}(\hat{\gamma}, \hat{\sigma}) = (I^{ls}(\hat{\gamma}, \hat{\sigma}))_{(m+2) \times (m+2)}.$$

L'estimateur  $\hat{t}_p(x_0)$  est la fonction régulière de  $\hat{\gamma}$  et  $\hat{\sigma}$ , donc la loi asymptotique de  $\hat{t}_p(x_0)$  est aussi normale. Mais  $t_p(x_0)$  prend des valeurs positives, donc la vitesse de convergence vers la loi normale est plus grande si on considère la loi limite de

$$\hat{K}_p(x_0) = \ln \hat{t}_p(x_0) = \hat{\gamma}^T z^{(0)} + \hat{\sigma} \ln [S_0^{-1}(1-p)].$$

La loi de  $\hat{K}_p(x_0)$  est approximée par la loi normale  $N(K_p(x_0), \sigma_{K_p}^2)$ , où la variance  $\sigma_{K_p}^2$  peut être estimée par

$$\begin{aligned} \hat{\sigma}_{K_p}^2 &= \left( \frac{\partial \hat{K}_p(x_0)}{\partial \hat{\gamma}_0}, \dots, \frac{\partial \hat{K}_p(x_0)}{\partial \hat{\gamma}_m}, \frac{\partial \hat{K}_p(x_0)}{\partial \hat{\sigma}} \right) \mathbf{I}^{-1}(\hat{\gamma}, \hat{\sigma}) \times \\ &\left( \frac{\partial \hat{K}_p(x_0)}{\partial \hat{\gamma}_0}, \dots, \frac{\partial \hat{K}_p(x_0)}{\partial \hat{\gamma}_m}, \frac{\partial \hat{K}_p(x_0)}{\partial \hat{\sigma}} \right)^T = \sum_{l=0}^m \sum_{s=0}^m z_{0l} z_{0s} I^{ls}(\hat{\gamma}, \hat{\sigma}) + \\ &2 \ln [S_0^{-1}(1-p)] \sum_{l=0}^m I^{l,m+1}(\hat{\gamma}, \hat{\sigma}) z_{0l} + \ln^2 [S_0^{-1}(1-p)] I^{m+1,m+1}(\hat{\gamma}, \hat{\sigma}). \end{aligned}$$

La loi de

$$\frac{\hat{K}_p(x_0) - K_p(x_0)}{\hat{\sigma}_{K_p}}$$

est approchée par la loi  $N(0, 1)$ . L'intervalle approximatif de confiance de niveau de confiance  $(1 - \alpha)$  pour  $K_p(x_0)$  est donné par la formule

$$\hat{K}_p(x_0) \pm \hat{\sigma}_{K_p} w_{1-\alpha/2},$$

où  $w_\alpha$  est la  $\alpha$ -quantile de la loi de  $N(0, 1)$ . L'intervalle approximatif pour  $t_p(x_0)$  est donné par la formule

$$\hat{t}_p(x_0) \exp\{\pm \hat{\sigma}_{K_p} w_{1-\alpha/2}\}.$$

L'estimateur  $\hat{S}_{x_0}(t)$  est aussi la fonction régulière de  $\gamma$  et  $\sigma$ . Notons

$$\hat{Q}_{x_0}(t) = \ln \frac{\hat{S}_{x_0}(t)}{1 - \hat{S}_{x_0}(t)} \quad \text{et} \quad Q_{x_0}(t) = \ln \frac{S_{x_0}(t)}{1 - S_{x_0}(t)}.$$

La fonction  $Q_{x_0}(t)$  prend ces valeurs dans  $\mathbb{R}$  donc la convergence de  $\hat{Q}_{x_0}(t)$  vers la loi limite est plus grande que la convergence de  $\hat{S}_{x_0}(t)$  vers sa loi limite. Comme dans le cas de  $t_p(x_0)$  on obtient que la loi de

$$(\hat{Q}_{x_0}(t) - Q_{x_0}(t)) / \hat{\sigma}_{Q_0}$$

est approximée par la loi normale  $N(0, 1)$ ; ici

$$\begin{aligned} \hat{\sigma}_{Q_0} &= \frac{S'(S^{-1}(\hat{S}_{x_0}(t)))}{\hat{\sigma}^2 \hat{S}_{x_0}(t)(1 - \hat{S}_{x_0}(t))} \times \\ &\sqrt{\hat{\sigma}^2 \sum_{l=0}^m \sum_{s=0}^m z_{0l} z_{0s} I^{ls}(\hat{\gamma}, \hat{\sigma}) - 2\hat{\gamma}^T z^{(0)} \sum_{l=0}^m z_{0l} \hat{\sigma} + (\hat{\gamma}^T z^{(0)})^2}. \end{aligned}$$

Donc les  $(1 - \alpha)$ -intervalles approximatifs de confiance pour  $Q_{x_0}(t)$  et  $S_{x_0}(t)$  sont  $\hat{Q}_{x_0}(t) \pm \hat{\sigma}_{Q_0} w_{1-\alpha/2}$  et

$$\left( 1 + \frac{1 - \hat{S}_{x_0}(t)}{\hat{S}_{x_0}(t)} \exp\{\mp \hat{\sigma}_{Q_0} w_{1-\alpha/2}\} \right)^{-1}.$$

**Exemple 1.** Si  $T_{x_0}$  suit la loi de Weibull, i.e.

$$S_{x_0}(t) = e^{-(t/\theta)^\nu}, \quad t \geq 0,$$

et la paramétrisation d'Arrhénius est choisie (le stress est la température, par exemple), i.e.

$$r(x) = e^{\beta_0 + \beta_1/x},$$

alors  $S(t) = \exp\{-\exp(t)\}$ ,  $z_{00} = 1$ ,  $z_{10} = 1/x_0$ , donc

$$\hat{S}_{x_0}(t) = \exp\left\{-\exp\left\{\frac{\ln t - \hat{\gamma}_0 - \hat{\gamma}_1/x_0}{\hat{\sigma}}\right\}\right\}, \quad \hat{t}_p(x_0) = e^{\hat{\gamma}_0 + \hat{\gamma}_1/x_0} (-\ln(1-p))^{\hat{\sigma}}.$$

**Exemple 2.** Si  $T_{x_0}$  suit la loi loglogistique, i.e.

$$S_{x_0}(t) = (1 + (t/\theta)^v)^{-1}, \quad t \geq 0,$$

et la paramétrisation de la règle de puissance est choisie (le stress est le voltage, par exemple), i.e.

$$r(x) = e^{\beta_0 + \beta_1 \ln x},$$

alors

$$S(t) = (1 + e^t)^{-1}, \quad z_{00} = 1, \quad z_{10} = \ln x_0,$$

donc

$$\hat{S}_{x_0}(t) = \left[1 + \exp\left(\frac{\ln t - \hat{\gamma}_0 - \hat{\gamma}_1 \ln x_0}{\hat{\sigma}}\right)\right]^{-1}, \quad \hat{t}_p(x_0) = e^{\hat{\gamma}_0 + \hat{\gamma}_1 \ln x_0} \left(\frac{p}{1-p}\right)^{\hat{\sigma}}.$$

**Exemple 3.** Si  $T_{x_0}$  suit la loi lognormale et la paramétrisation d'Eyring est choisie, i.e.

$$r(x) = e^{\beta_0 + \beta_1 \ln x + \beta_2/x},$$

alors

$$z_{00} = 1, \quad z_{10} = \ln x_0, \quad z_{20} = 1/x_0, \quad S(t) = 1 - \Phi(t)$$

et

$$\hat{S}_{x_0}(t) = 1 - \Phi\left(\frac{\ln t - \hat{\gamma}_0 - \hat{\gamma}_1 \ln x_0 - \hat{\gamma}_2/x_0}{\hat{\sigma}}\right), \quad \hat{t}_p(x_0) = e^{\hat{\gamma}_0 + \hat{\gamma}_1 \ln x_0 + \hat{\gamma}_2/x_0 + \hat{\sigma}\Phi^{-1}(p)}.$$

**Exemple 4.** Supposons que la durée de vie  $T_{x_0}$  suit la loi de Weibull et le stress  $x = (x_1, x_2)^T$  est bidimensionnel (le voltage et la température, par exemple) et le modèle d'Arrhénius généralisé avec  $\delta_1(x) = \alpha_1/x_1$ ,  $\delta_2/x_2^2$  est choisi. Alors

$$z_{00} = 1, \quad z_{10} = \ln x_{10}, \quad z_{20} = 1/x_{20}$$

et

$$\hat{S}_{x_0}(t) = \exp\left\{-\exp\left\{\frac{\ln t - \hat{\gamma}_0 - \hat{\gamma}_1 \ln x_{10} - \hat{\gamma}_2/x_{20}}{\hat{\sigma}}\right\}\right\},$$

$$\hat{t}_p(x_0) = e^{\hat{\gamma}_0 + \hat{\gamma}_1 \ln x_{10} + \hat{\gamma}_2/x_{20}} (-\ln(1-p))^{\hat{\sigma}}.$$

Les formules sont plus simples, si  $v = 1$  dans (7), par exemple dans le cas de la loi exponentielle :

$$S_{x_0}(t) = e^{-t/\theta}, \quad t \geq 0, \quad t_p(x_0) = -e^{-\gamma^T z^{(0)}} \ln(1-p).$$

Le modèle (6) peut être écrit

$$S_x(t) = \exp\{-\exp(\gamma^T z)t\},$$

où

$$\gamma = (\gamma_0, \dots, \gamma_m)^T, \quad \gamma_0 = \beta_0 - \ln \theta, \quad \gamma_i = \beta_i, \quad (i = 1, \dots, m).$$

La fonction de survie et le taux de pannes de  $T_{ij}$  sont

$$S_{x_i}(t) = \exp\{-\exp(\gamma^T z^{(i)})t\},$$

$$\lambda_{x_i}(t) = e^{-\gamma^T z^{(i)}}.$$

Notons  $X_{ij} = T_{ij} \wedge t_i$ ,  $\delta_{ij} = I(T_{ij} \leq t_i)$ . La fonction de vraisemblance

$$L(\gamma) = \prod_{i=1}^k \prod_{j=1}^{n_i} [\lambda_{x_i}(X_{ij})]^{\delta_{ij}} S_{x_i}(X_{ij}) = \exp\left\{-\sum_{i=1}^k \sum_{j=1}^{n_i} (\delta_{ij} \gamma^T z^{(i)} + e^{\gamma^T z^{(i)}} X_{ij})\right\}.$$

Les fonctions score

$$U_l(\gamma) = \frac{\partial \ln L(\gamma)}{\partial \gamma_l} = -\sum_{i=1}^k z_{il} (\delta_i + e^{\gamma^T z^{(i)}} X_i),$$

où

$$\delta_i = \sum_{j=1}^{n_i} \delta_{ij}, \quad X_i = \sum_{j=1}^{n_i} X_{ij}$$

et la matrice d'information de Fisher

$$\mathbf{I}(\gamma) = (I_{ls}(\gamma)), \quad (l, s = 0, \dots, m),$$

où

$$I_{ls}(\gamma) = -\mathbf{E} \left\{ \frac{\partial^2 \ln L(\gamma)}{\partial \gamma_l \partial \gamma_s} \right\} = \mathbf{E} \left\{ \sum_{i=1}^k z_{il} z_{is} e^{\gamma^T z^{(i)}} X_i \right\} = \sum_{i=1}^k n_i z_{il} z_{is} \left( 1 - e^{-e^{\gamma^T z^{(i)}} t_i} \right).$$

S'il n'y a pas de censures, i.e.  $t_i = \infty$ , alors

$$I_{ls}(\gamma) = \sum_{i=1}^k n_i z_{il} z_{is}$$

ne dépendent pas de  $\gamma$ .

Notons  $\hat{\mathbf{I}} = \mathbf{I}(\hat{\gamma})$  la matrice d'information de Fisher estimée.

La loi asymptotique de  $\hat{\gamma}$  quand  $n_i$  sont grands est approximée par la loi normale  $N(\gamma, \mathbf{I}^{-1}(\gamma))$  et donc la loi de

$$\hat{K}_p(x_0) = \ln \hat{t}_p(x_0) = -\hat{\gamma}^T z^{(0)} + \ln(-\ln(1-p))$$

est approximée par la loi normale  $N(K_p(x_0), \sigma_{K_p}^2)$ , où

$$\sigma_{K_p}^2 = \sum_{l=0}^m \sum_{s=0}^m I^{ls}(\gamma) z_{0l} z_{0s},$$

Notons

$$\hat{\sigma}_{K_p}^2 = \sum_{l=0}^m \sum_{s=0}^m I^{ls}(\hat{\gamma}) z_{0l} z_{0s}.$$

Donc

$$\frac{\hat{K}_p(x_0) - K_p(x_0)}{\hat{\sigma}_{K_p}}$$

est approximée par la loi  $N(0, 1)$ . L'intervalle approximatif de confiance du niveau  $1 - \alpha$  pour  $t_p(x_0)$  est donné par la formule

$$\hat{t}_p(x_0) \exp\{\pm \hat{\sigma}_{K_p} w_{1-\alpha/2}\}.$$

L'estimateur  $\hat{S}_{x_0}(t)$  est aussi la fonction régulière de  $\gamma$ . Notons

$$\hat{Q}_{x_0}(t) = \ln \frac{\hat{S}_{x_0}(t)}{1 - \hat{S}_{x_0}(t)} \quad \text{et} \quad Q_{x_0}(t) = \ln \frac{S_{x_0}(t)}{1 - S_{x_0}(t)}.$$

Comme dans le cas du quantile  $t_p(x_0)$  on obtient que la loi de

$$\frac{\hat{Q}_{x_0}(t) - Q_{x_0}(t)}{\hat{\sigma}_{x_0}}$$

peut être approximée par la loi  $N(0, 1)$ ; ici

$$\hat{\sigma}_{Q_0} = \ln \hat{S}_{x_0}(t) \sqrt{\sum_{l=0}^m \sum_{s=0}^m z_{0l} z_{0s} I^{ls}(\hat{\gamma})}.$$

Donc les  $(1 - \alpha)$ -intervalles de confiance approximatifs pour  $Q_{x_0}(t)$  et  $S_{x_0}(t)$  sont

$$\hat{Q}_{x_0}(t) \pm \hat{\sigma}_{Q_0} w_{1-\alpha/2}$$

et

$$\left(1 + \frac{1 - \hat{S}_{x_0}(t)}{\hat{S}_{x_0}(t)} \exp\{\mp \hat{\sigma}_{Q_0} w_{1-\alpha/2}\}\right)^{-1}.$$

Si  $\hat{\gamma}$  est l'estimateur de maximum de vraisemblance pour  $\gamma$ , alors

$$\hat{S}_{x_0}(t) = \exp\{-e^{\hat{\gamma}^T z^{(0)}} t\}, \quad \hat{t}_p(x_0) = -\exp\{-e^{\hat{\gamma}^T z^{(0)}} \ln(1 - p)\}.$$

Par exemple, dans le cas de modèles d'Arrhénius et de la règle de puissance il faut prendre  $z_{il} = 1/x_{il}$  et  $z_{il} = \ln x_{il}$  respectivement et on a

$$\hat{S}_{x_0}(t) = \exp\{-e^{\hat{\gamma}_0 + \hat{\gamma}_1/x_0} t\}, \quad \hat{S}_{x_0}(t) = \exp\{-e^{\hat{\gamma}_0 + \hat{\gamma}_1 \ln x_0} t\}$$

respectivement.

Le premier plan d'expérience a ses points faibles :

- 1) des strictes suppositions sur la forme de la fonction  $r(x)$  sont faites ;
- 2) comme dans le cas de tous les modèles de régression la prédiction de variable dépendante pour la valeur de stress  $x_0$  peut être mauvaise parce que cette valeur n'appartient pas à la région des stress utilisés pendant des expériences.

Donc supposons que le deuxième plan est utilisé : le premier groupe d'articles de taille  $n_1$  est testé sous le stress accéléré  $x_1$  et un échantillon complet  $T_{11} \leq \dots \leq T_{1n_1}$  est obtenu, le deuxième groupe de taille  $n_2$  est testé sous le stress

$$x_2(\tau) = \begin{cases} x_1, & \text{si } 0 \leq \tau \leq t_1, \\ x_0, & \text{si } t_1 \leq \tau \leq t_2 \end{cases}$$

et un échantillon censuré du premier type  $T_{21} \leq \dots \leq T_{2m_2}$  est obtenu ( $m_2 \leq n_2$ ).

Supposons que

$$S_{x_0}(t) = S_0((t/\theta)^\alpha),$$

donc le modèle (2) peut être écrit

$$S_{x(\cdot)}(t) = S_0\left(\left(\int_0^t r[x(\tau)]d\tau/\theta\right)^\alpha\right). \quad (8)$$

La formule (8) implique

$$S_{x_1}(t) = S_0\left(\left(\frac{rt}{\theta}\right)^\alpha\right),$$

$$S_{x_2}(t) = S_0\left(\left((r(t_1 \wedge t) + (t - t_1) \vee 0)/\theta\right)^\alpha\right),$$

où  $r = r(x_1)$ .

Notons

$$\rho = \ln r, \quad \psi = \ln \theta, \quad S(t) = S_0(e^t), \quad f(t) = -S'(t), \quad \lambda(t) = f(t)/S(t).$$

Alors

$$S_{x_1}(t) = S(\alpha(\ln t + \rho - \psi));$$

$$S_{x_2}(t) = \begin{cases} S(\alpha(\ln t + \rho - \psi)), & t \leq t_1, \\ S(\alpha(\ln(e^\rho t_1 + t - t_1) - \psi)), & t > t_1; \end{cases}$$

$$f_{x_1}(t) = f(\alpha(\ln t + \rho - \psi))\frac{\alpha}{t};$$

$$f_{x_2}(t) = \begin{cases} f(\alpha(\ln t + \rho - \psi))\frac{\alpha}{t}, & t \leq t_1, \\ f(\alpha(\ln(e^\rho t_1 + t - t_1) - \psi))\frac{\alpha}{e^\rho t_1 + t - t_1}, & t > t_1. \end{cases}$$

Notons  $r_2$  le nombre de pannes du deuxième groupe jusqu'au moment  $t_1$ . La fonction de vraisemblance

$$L = \prod_{j=1}^{n_1} f(\alpha(\ln T_{1j} + \rho - \psi))\frac{\alpha}{T_{1j}} \prod_{j=1}^{r_2} f(\alpha(\ln T_{2j} + \rho - \psi))\frac{\alpha}{T_{2j}} \times$$

$$\prod_{j=r_2+1}^{m_2} f(\alpha(\ln(e^\rho t_1 + T_{2j} - t_1) - \psi))\frac{\alpha}{e^\rho t_1 + T_{2j} - t_1} S^{n_2 - m_2}(\alpha(\ln(e^\rho t_1 + t_2 - t_1) - \psi)),$$

donc

$$U_1(\alpha, \rho, \psi) = \frac{\partial \ln L}{\partial \alpha} = \sum_{j=1}^{n_1} (\ln f)'(c(T_{1j}))\frac{c(T_{1j})}{\alpha} + \frac{n_1 + m_2}{\alpha} +$$

$$\sum_{j=1}^{r_2} (\ln f)'(c(T_{2j}))\frac{c(T_{2j})}{\alpha} + \sum_{j=r_2+1}^{m_2} (\ln f)'(d(T_{2j}))\frac{d(T_{2j})}{\alpha} - (n_2 - m_2)\lambda(d(t_2))\frac{d(t_2)}{\alpha},$$

$$U_2(\alpha, \rho, \psi) = \frac{\partial \ln L}{\partial \rho} = \sum_{j=1}^{n_1} (\ln f)'(c(T_{1j}))\alpha + \sum_{j=1}^{r_2} (\ln f)'(c(T_{2j}))\alpha +$$

$$\sum_{j=r_2+1}^{m_2} (\ln f)'(d(T_{2j})) \frac{\alpha e^{\rho t_1}}{e^{\rho t_1} + T_{2j} - t_1} - \sum_{j=r_2+1}^{m_2} \frac{e^{\rho t_1}}{e^{\rho t_1} + T_{2j} - t_1} - (n_2 - m_2)\lambda(d(t_2)) \frac{\alpha e^{\rho t_1}}{e^{\rho t_1} + t_2 - t_1},$$

$$U_3(\alpha, \rho, \psi) = \frac{\partial \ln L}{\partial \psi} = -\alpha \left[ \sum_{j=1}^{n_1} (\ln f)'(c(T_{1j})) + \sum_{j=1}^{r_2} (\ln f)'(c(T_{2j})) + \right.$$

$$\left. \sum_{j=r_2+1}^{m_2} (\ln f)'(d(T_{2j})) - (n_2 - m_2)\lambda(d(t_2)) \right],$$

où

$$c(u) = \alpha(\ln u + \rho - \psi), \quad d(u) = \alpha(\ln(e^{\rho t_1} + u - t_1) - \psi).$$

Dans les cas des lois de Weibull, loglogistique et lognormale

$$(\ln f)'(t) = e^t; \quad \frac{1 - e^t}{1 + e^t}; \quad -t,$$

respectivement, et

$$\lambda(t) = e^t; \quad (1 + e^{-t})^{-1}; \quad \frac{\Phi(t)}{1 - \Phi(t)},$$

respectivement.

Si les estimateurs de maximum de vraisemblance  $\hat{\alpha}$ ,  $\hat{\rho}$ ,  $\hat{\psi}$  sont obtenus, alors l'estimateurs de la fonction de survie  $S_{x_0}$  et de la  $p$ -quantile  $t_p(x_0)$  sont

$$\hat{S}_{x_0}(t) = S(\hat{\alpha}(\ln t - \hat{\psi})), \quad \hat{t}_p = \exp\{\hat{\psi} + \frac{1}{\hat{\alpha}} S^{-1}(1 - p)\}.$$

Dans les cas des lois de Weibull, loglogistique et lognormale

$$S^{-1}(p) = \ln(-\ln(1 - p)); \quad -\ln\left(\frac{1}{p} - 1\right); \quad \Phi^{-1}(1 - p).$$

Notons  $\mathbf{I}(\alpha, \rho, \psi) = (I_{ij}(\alpha, \rho, \psi))$  une  $(3 \times 4)$  matrice symétrique avec des éléments suivantes :

$$I_{11} = -\frac{\partial^2 \ln L}{\partial \alpha^2} = -\frac{1}{\alpha^2} \left\{ \sum_{j=1}^{n_1} (\ln f)''(c(T_{1j})) [c(T_{1j})]^2 - n_1 - m_2 + \right.$$

$$\sum_{j=1}^{r_2} (\ln f)''(c(T_{2j})) [c(T_{2j})]^2 + \sum_{j=r_2+1}^{m_2} (\ln f)''(d(T_{2j})) [d(T_{2j})]^2 -$$

$$\left. (n_2 - m_2)\lambda'(d(t_2)) [d(t_2)]^2 \right\},$$

$$I_{12} = I_{21} = -\frac{\partial^2 \ln L}{\partial \alpha \partial \rho} = -\sum_{j=1}^{n_1} (\ln f)''(c(T_{1j})) c(T_{1j}) -$$

$$\sum_{j=1}^{r_2} (\ln f)''(c(T_{2j})) - \sum_{j=r_2+1}^{m_2} (\ln f)''(d(T_{2j})) d(T_{2j}) + (n_2 - m_2)\lambda'(d(t_2)) \frac{e^{\rho t_1}}{e^{\rho t_1} + t_2 - t_1} -$$

$$\frac{1}{\alpha}U_2(\alpha, \rho, \psi) - \frac{1}{\alpha} \sum_{j=r_2+1}^{m_2} \frac{e^{\rho t_1}}{e^{\rho t_1} + T_{2j} - t_1},$$

$$I_{13} = I_{31} = -\frac{\partial^2 \ln L}{\partial \alpha \partial \psi} = \sum_{j=1}^{n_1} (\ln f)''(c(T_{1j}))c(T_{1j}) + \sum_{j=1}^{r_2} (\ln f)''(c(T_{2j}))c(T_{2j}) +$$

$$\sum_{j=r_2+1}^{m_2} (\ln f)''(d(T_{2j}))d(T_{2j}) - (n_2 - m_2)\lambda'(d(t_2))d(t_2) - \frac{1}{\alpha}U_3(\alpha, \rho, \psi),$$

$$I_{22} = -\frac{\partial^2 \ln L}{\partial \rho^2} = -\alpha^2 \sum_{j=1}^{n_1} (\ln f)''(c(T_{1j})) -$$

$$\alpha^2 \sum_{j=1}^{r_2} (\ln f)''(c(T_{2j})) - \alpha^2 \sum_{j=r_2+1}^{m_2} (\ln f)''(d(T_{2j})) \left( \frac{e^{\rho t_1}}{e^{\rho t_1} + T_{2j} - t_1} \right)^2 -$$

$$\sum_{j=r_2+1}^{m_2} [\alpha (\ln f)'(d(T_{2j})) - 1] \frac{e^{\rho t_1} (T_{2j} - t_1)}{(e^{\rho t_1} + T_{2j} - t_1)^2} +$$

$$(n_2 - m_2)\lambda'(d(t_2)) \left( \frac{\alpha e^{\rho t_1}}{e^{\rho t_1} + t_2 - t_1} \right)^2 + (n_2 - m_2)\lambda(d(t_2)) \frac{\alpha e^{\rho t_1} (t_2 - t_1)}{(e^{\rho t_1} + t_2 - t_1)^2},$$

$$I_{23} = I_{32} = -\frac{\partial^2 \ln L}{\partial \rho \partial \psi} = \alpha^2 \left\{ \sum_{j=1}^{n_1} (\ln f)''(c(T_{1j})) + \sum_{j=1}^{r_2} (\ln f)''(c(T_{2j})) +$$

$$\sum_{j=r_2+1}^{m_2} (\ln f)''(d(T_{2j})) \frac{e^{\rho t_1}}{e^{\rho t_1} + T_{2j} - t_1} - (n_2 - m_2)\lambda'(d(t_2)) \frac{e^{\rho t_1}}{(e^{\rho t_1} + t_2 - t_1)} \right\},$$

$$I_{33} = -\frac{\partial^2 \ln L}{\partial \psi^2} = -\alpha^2 \left[ \sum_{j=1}^{n_1} (\ln f)''(c(T_{1j})) + \sum_{j=1}^{r_2} (\ln f)''(c(T_{2j})) +$$

$$\sum_{j=r_2+1}^{m_2} (\ln f)''(d(T_{2j})) - (n_2 - m_2)\lambda'(d(t_2)) \right].$$

Dans les cas des lois de Weibull, loglogistique et lognormale

$$(\ln f)''(t) = e^t; \frac{-2e^t}{(1+e^t)^2}; -1,$$

respectivement, et

$$\lambda'(t) = e^t; \frac{e^t}{(1+e^t)^2}; -t \frac{\varphi(t)}{1-\Phi(t)} + \left( \frac{\varphi(t)}{1-\Phi(t)} \right)^2,$$

respectivement.

Si  $n$  est grand, la loi de  $(\hat{\alpha}, \hat{\rho}, \hat{\psi})$  peut être approchée par la loi normale de moyenne  $(\alpha, \rho, \psi)$  et la matrice de covariance estimée par

$$\mathbf{I}^{-1}(\hat{\alpha}, \hat{\rho}, \hat{\psi}) = (I^{ls}(\hat{\alpha}, \hat{\rho}, \hat{\psi}))_{3 \times 3}.$$

Notons

$$\hat{Q}_{x_0}(t) = \ln \frac{\hat{S}_{x_0}(t)}{1 - \hat{S}_{x_0}(t)}.$$

La loi de

$$(\hat{Q}_{x_0}(yt) - Q_{x_0}(t)) / \hat{\sigma}_{Q_0}$$

est approchée par la loi normale  $N(0, 1)$ , ici

$$\hat{\sigma}_{Q_0} = \frac{S'(S^{-1}(\hat{S}_{x_0}(t)))}{\hat{S}_{x_0}(t)(1 - \hat{S}_{x_0}(t))} \times$$

$$\sqrt{(\ln t - \hat{\psi})^2 I^{11}(\hat{\alpha}, \hat{\rho}, \hat{\psi}) - 2\hat{\alpha}(\ln t - \hat{\psi}) I^{13}(\hat{\alpha}, \hat{\rho}, \hat{\psi}) + \hat{\alpha}^2 I^{33}(\hat{\alpha}, \hat{\rho}, \hat{\psi})}.$$

Donc les  $(1 - \alpha)$  intervalles approximatifs de confiance pour  $Q_{x_0}(t)$  et  $S_{x_0}(t)$  sont

$$\hat{Q}_{x_0} \pm \hat{\sigma}_{Q_0} w_{1-\alpha/2} \quad \text{et} \quad \left( 1 + \frac{1 - \hat{S}_{x_0}(t)}{\hat{S}_{x_0}(t)} \exp\{\mp \hat{\sigma}_{Q_0} w_{1-\alpha/2}\} \right)^{-1},$$

respectivement.

Notons

$$\hat{K}_p(x_0) = \ln \hat{t}_p(x_0) = \hat{\psi} + \frac{1}{\hat{\alpha}} S^{-1}(1 - p).$$

La loi de

$$(\hat{K}_p(x_0) - K_p(x_0)) / \hat{\sigma}_{K_p}$$

est approchée par la loi normale  $N(0, 1)$ ; ici

$$\hat{\sigma}_{K_p}^2 = \left( \frac{S^{-1}(1 - p)}{\alpha^2} \right)^2 I^{11} - \frac{S^{-1}(1 - p)}{\alpha^2} I^{13} + I^{33}.$$

Donc les  $(1 - \alpha)$  intervalles approximatifs de confiance pour

$$K_p(x_0) = \ln t_p(x_0) \quad \text{et} \quad t_p(x_0)$$

sont

$$\hat{K}_p(x_0) + \pm w_{1-\alpha/2} \hat{\sigma}_{K_p} \quad \text{et} \quad \hat{t}_p(x_0) \exp\{\pm \hat{\sigma}_{K_p} w_{1-\alpha/2}\}$$

respectivement.

### 6.11.3 Estimation semiparamétrique

On suppose que le modèle (5) est considéré et la fonction  $S_{x_0}$  est inconnue. On considère le premier plan d'expériences. La fonction de survie sous le stress  $x_i$  est

$$S_{x_i}(t) = S_{x_0}(e^{\beta^T z_i t}).$$

Notons  $N_i(\tau)$  les nombres des pannes observées du  $i$ -ème groupe dans l'intervalle  $[0, \tau]$ ,  $Y_i(\tau)$  des nombres d'unités "à risque" (à l'état de fonctionnement et non-censurés) avant le

moment  $\tau$ ,  $T_{i1} \leq \dots \leq T_{im_i}$  les moments de pannes du  $i$ -ème groupe,  $m_i = N_i(t_i)$ . On suppose d'abord que  $\beta$  soit connu. Les variables aléatoires

$$e^{\beta^T z_i T_{ij}} \quad (i = 1, \dots, k; j = 1, \dots, m_i)$$

peuvent être considérées comme des pseudo-pannes "observées" dans une expérience où  $n = \sum_{i=1}^m n_i$  unités avec la fonction de survie  $S_{x_0}$  ont été testés et  $n_i$  parmi elles ont été censurées au moment  $e^{\beta^T z_i t_i}$  ( $i = 1, 2, \dots, k$ ). Alors

$$N^R(\tau, \beta) = \sum_{i=1}^k N_i(e^{-\beta z_i \tau})$$

est le nombre des pannes observées dans l'intervalle  $[0, \tau]$  et

$$Y^R(\tau, \beta) = \sum_{i=1}^k Y_i(e^{-\beta z_i \tau})$$

est le nombre des unités à risque avant le moment  $t$ .

La fonction de survie  $S_{x_0}$  peut être estimée par l'estimateur de Kaplan-Meier : pour tous  $s \leq \max_i \{e^{\beta z_i t_i}\}$

$$\tilde{S}_{x_0}(s, \beta) = \prod_{\tau \leq s} \left( 1 - \frac{\Delta N^R(\tau, \beta)}{Y^R(\tau, \beta)} \right) = \prod_{\tau \leq s} \left( 1 - \frac{\sum_{l=1}^m \Delta N_l(e^{-\beta z_l \tau})}{\sum_{l=1}^m Y_l(e^{-\beta z_l \tau})} \right),$$

où  $\Delta N^R(\tau, \beta) = N^R(\tau, \beta) - N^R(\tau-, \beta)$ . On écrit  $\tilde{S}_0$  de façon suivant :

$$\tilde{S}_{x_0}(s, \beta) = \prod_{(i,j): T_{ij} \leq \exp\{-\beta z_i\} s} \left( 1 - \frac{1}{\sum_{l=1}^m Y_l(e^{\beta(z_i - z_l) T_{ij}})} \right).$$

La fonction de vraisemblance

$$L(\beta) = \prod_{i=1}^k \prod_{j=1}^{m_i} [\tilde{S}_{x_0}(e^{\beta^T z_i T_{ij}-}, \beta) - \tilde{S}_{x_0}(e^{\beta^T z_i T_{ij}}, \beta)] \tilde{S}_{x_0}^{m_i - m_i}(e^{\beta^T z_i t_i}, \beta),$$

où

$$\tilde{S}_{x_0}(u-, \beta) = \lim_{\varepsilon \downarrow 0} \tilde{S}_{x_0}(u - \varepsilon, \beta).$$

Le facteur qui correspond à une panne est le saut de la fonction  $\tilde{S}_{x_0}$  parce que la densité  $f_{x_0} = -S'_{x_0}$  est inconnue et peut être approchée dans les points  $e^{\beta^T z_i T_{ij}}$  par le facteur proportionnel à

$$\tilde{S}_{x_0}(e^{\beta^T z_i T_{ij}-}, \beta) - \tilde{S}_{x_0}(e^{\beta^T z_i T_{ij}}, \beta).$$

Si on a des *ex aequo*, alors on note  $T_1^*(\beta) < \dots < T_q^*(\beta)$  les moments différents parmi  $\exp\{\beta^T z_i T_{ij}\}$ ,  $d_j$  - le nombre des pseudopannes au moment  $T_j^*(\beta)$ . Alors pour tout  $s \leq \max_i \{e^{\beta^T z_i t_i}\}$

$$\tilde{S}_{x_0}(s, \beta) = \prod_{j: T_j^*(\beta) \leq s} \left( 1 - \frac{d_j}{\sum_{l=1}^m Y_l(e^{-\beta^T z_l T_j^*(\beta)})} \right)$$

et

$$L(\beta) = \prod_{j=1}^q [\tilde{S}_{x_0}(T_{j-1}^*(\beta), \beta) - \tilde{S}_{x_0}(T_j^*(\beta), \beta)]^{d_j} \prod_{i=1}^m \tilde{S}_{x_0}^{n_i - m_i}(e^{\beta^T z_i t_i}, \beta).$$

Notons  $\hat{\beta} = \text{Argmax}_{\beta} L(\beta)$ . La fonction de survie sous le stress normale est estimée pour tous  $s \leq \max_i \{e^{\beta^T z_i t_i}\}$  par

$$\hat{S}_{x_0}(s) = \tilde{S}_{x_0}(s, \hat{\beta}).$$

Au lieu d'estimation par la méthode de maximum de vraisemblance on peut considérer la méthode des moments modifiée.

Si  $\beta$  est connu, le taux de pannes accumulé

$$A_{x_0}(t) = \exp\{-S_{x_0}(t)\}$$

peut être estimé par l'estimateur de Nelson-Aalen :

pour tout  $t \leq \max\{e^{\beta^T z_i t_i}\}$  on a

$$\tilde{A}_{x_0}(t, \beta) = \int_0^t \frac{dN^R(u)}{Y^R(u)} = \int_0^t \frac{d \sum_{i=1}^k N_i(e^{-\beta^T z_i u})}{\sum_{i=1}^k Y_i(e^{-\beta^T z_i u})}.$$

La proposition (annexe) implique que

$$\begin{aligned} \mathbf{E} \sum_{i=1}^k z_i \int_0^{\infty} dN_i(u) - Y_i(u) dA_{x_0}(e^{\beta^T z_i u}) = \\ \mathbf{E} \sum_{i=1}^k z_i \int_0^{\infty} dN_i(u) - Y_i(u) dA_i(u) = 0. \end{aligned}$$

Donc l'estimateur de  $\beta$  peut être trouvé en considérant la fonction

$$\tilde{U}(\beta) = \sum_{i=1}^k z_i \int_0^{\infty} dN_i(u) - Y_i(u) d\tilde{A}_{x_0}(e^{\beta z_i u}, \beta). \quad (9)$$

C'est la fonction en escaliers et ces valeurs sont dispersées autour zéro. L'estimateur de  $\beta$  peut être déterminé comme

$$\hat{\beta} = \sup_{\beta} \text{Arg min } \tilde{U}(\beta).$$

Le choix de poids  $z_i$  peut être justifié de façon suivante. Si  $A_{x_0}$  est connue, alors la fonction de vraisemblance pour  $\beta$

$$L(\beta) = \prod_{i=1}^k \prod_{j=1}^{n_i} \lambda_{x_i}^{\delta_{ij}}(X_{ij}) S_{x_i}(X_{ij})$$

et donc

$$\begin{aligned} U_l(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta_l} = \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} z_{il} [1 + e^{\beta z_i X_{ij}} \frac{\alpha'_{x_0}(e^{\beta z_i X_{ij}})}{\alpha_{x_0}(e^{\beta z_i X_{ij}})} - \alpha_{x_0}(e^{\beta z_i X_{ij}}) e^{\beta z_i X_{ij}}] = \\ \sum_{i=1}^k \int_0^{\infty} W_i(u) (dN_i(u) - Y_i(u) dA_{x_0}(e^{\beta z_i u})), \end{aligned}$$

où

$$W_{il} = z_{il} \left( 1 + e^{\beta z_i u} \frac{\alpha'_{x_0}(e^{\beta z_i u})}{\alpha_{x_0}(e^{\beta z_i u})} \right).$$

Les poids optimaux  $W_{il}$  dépendent de la loi de  $T_{x_0}$ . Si  $T_{x_0}$  suit la loi de Weibull, alors  $W_{il}(u) = z_{il}$ . Notons

$$U(\beta) = (U_0(\beta), \dots, U_m(\beta))^T, \quad W_i(u) = (W_{i0}, \dots, W_{im}(u)).$$

Alors

$$U(\beta) = \sum_i^k \int_0^\infty W_i(u) (dN_i(u) - Y_i(u) dA_{x_0}(e^{\beta^T z_i u})). \quad (10)$$

Remplaçant dans (10) la fonction inconnue  $A_{x_0}(v)$  par le pseudoestimateur  $\tilde{A}_{x_0}(v, \beta)$  et en prenant des poids  $W_{il}(u) = z_{il}$ , on obtient la fonction score modifiée  $\tilde{U}(\beta)$ , donnée par la formule (9).

Le choix de poids influence un peu l'effectivité mais pas la validité des procédures inférentielles. Les poids optimaux dépendent de la dérivée du taux de pannes et donc ne peuvent pas être bien estimés quand la loi est inconnue. Donc on utilise les poids les plus simples  $W_{il}(u) = z_{il}$  qui sont optimaux pour la loi de Weibull.

Après avoir trouvé l'estimateur  $\hat{\beta}$  par une des méthodes considérées on obtient un estimateur de la fonction de survie :

$$\hat{S}_{x_0}(t) = \tilde{S}_{x_0}(t, \hat{\beta})$$

ou, de façon alternative,

$$\hat{S}_{x_0}(t) = \exp\{-\tilde{A}_0(t, \hat{\beta})\}.$$

Considérons le deuxième plan d'expériences. Le premier groupe d'articles de taille  $n_1$  est testé sous le stress accéléré  $x_1$  et un échantillon complet  $T_{11} \leq \dots \leq T_{1n_1}$  est obtenu. Le deuxième groupe de taille  $n_2$  est testé sous le stress ( ) et un échantillon censuré  $T_{21} \leq \dots \leq T_{2m_2}$  est obtenu ( $m_2 \leq n_2$ ). Notons  $N_i(\tau)$  et  $Y_i(\tau)$  les nombres des pannes observées dans l'intervalle  $[0, \tau]$ ,  $Y_i(\tau)$  les nombres d'unités "à risque" avant le moment  $\tau$  du  $i$ ème groupe :

$$N_1(\tau) = \sum_{j=1}^{n_1} I(T_{1j} \leq \tau), \quad N_2(\tau) = \sum_{j=1}^{m_2} I(T_{2j} \leq \tau),$$

$$Y_1(\tau) = \sum_{j=1}^{n_1} I(T_{1j} \geq \tau), \quad Y_2(\tau) = \left[ \sum_{j=1}^{m_2} I(T_{2j} \geq \tau) + n_2 - m_2 \right] I(\tau \leq t).$$

Le modèle (8) implique que

$$S_{x_1}(t) = S_{x_0}(rt), \quad S_{x_2}(t) = S_{x_0}(r(t \wedge t_1) + (t - t_1) \vee 0),$$

où  $r = r(x_1)$ . Les moments

$$R_{ij} = rT_{ij} \quad \text{et} \quad R_{2j} = r(T_{2j} \wedge t_1) + (T_{2j} - t_1) \vee 0$$

peuvent être interprétés comme les moments de pannes obtenus dans une expérience pendant laquelle  $n = n_1 + n_2$  "unités" de fonction de survie  $S_{x_0}$  ont été observés et le temps de

censure pour les dernières  $n_2$  “unités” a été égale à  $(rt_1 + t - t_1)$ . Les nombres des pannes, “obsrvées” dans un intervalle  $[0, u]$  serait

$$N^R(u) = W_1(u/r) + N_2(u/r \wedge t_1 + (u - rt_1) \vee 0)$$

et le nombre des “unités à risque”

$$Y^R(u) = Y_1(u/r) + Y_2(u/r \wedge t_1) + (u - rt_1) \vee 0.$$

Donc le pseudoestimateur  $\tilde{A}_0(s, r)$ , dependant de  $r$ , de la fonction de pannes accumulées  $A_0(t) = -\ln S_0(t)$  est

$$\tilde{A}_0(s, r) = \int_0^s \frac{dN_1(u/r) + dN_2((u/r) \wedge t_1 + (u - rt_1) \vee 0)}{Y_1(u/r) + Y_2((u/r) \wedge t_1 + (u - rt_1) \vee 0)}$$

et le pseudoestimateur  $\tilde{S}_0(s, r)$  de la fonction de survie  $S_0$  est

$$\tilde{S}_0(s, r) = \prod_{(i,j) \in B(s)} \left( 1 - \frac{1}{Y_1(T_{1i}) + Y_2(t_1 \wedge T_{1i} + r((T_{1i} - t_1) \vee 0))} \right) \times \left( 1 - \frac{1}{Y_2(T_{2j}) + Y_1(t_1 \wedge T_{2j} + (\frac{T_{2j} - t_1}{r}) \vee 0)} \right),$$

où

$$B(s) = \{(i, j) \mid rT_{1i} \leq s \quad \text{et} \quad r(T_{2j} \wedge t_1) + (T_{2j} - t_1) \vee 0 \leq s\}.$$

Alors les pseudoestimateurs pour  $S_1 = S_{x_1}$  et  $S_2 = S_{x_2}$  sont

$$\tilde{S}_1(s, r) = \tilde{S}_0(rs, r), \quad \tilde{S}_2(s) = \tilde{S}_0(r(s \wedge t_1) + (s - t_1) \vee 0).$$

La fonction de vraisemblance observée

$$L(r) = \prod_{i=1}^{n_1} [\tilde{S}_0(rT_{1i}^-, r) - \tilde{S}_0(rT_{1i}, r)] \prod_{j=1}^{m_2} [\tilde{S}_0((r(T_{2j} \wedge t_1) + (T_{2j} - t_1) \vee 0)^-, r) - \tilde{S}_0(r(T_{2j} \wedge t_1) + (T_{2j} - t_1) \vee 0, r)] [\tilde{S}_0(rt_1 + t - t_1, r)]^{n_2 - m_2}.$$

Considérons la méthode des moments modifiée. De la même façon qu’au cas du premier plan, on a

$$\mathbf{E} \sum_{i=1}^2 \int_0^\infty x_i(\tau) (dN_i(\tau) - Y_i(\tau) dA_i(t)) = x_1 \mathbf{E} \int_0^\infty (dN_1(\tau) - Y_1(\tau) dA_0(r\tau)) + \mathbf{E} \int_0^\infty x_2(\tau) \{dN_2(\tau) - Y_2(\tau) dA_0(r(\tau \wedge t_1) + (\tau - t_1) \vee 0)\}$$

Notons  $n = n_1 + n_2$ . Considérons la fonction

$$\hat{U}(r) = \frac{1}{x_1 - x_0} \{x_1 \int_0^\infty (dN_1(\tau) - Y_1(\tau) d\tilde{A}_0(r\tau, r)) + \int_0^\infty x_2(\tau) \{dN_2(\tau) - Y_2(\tau) d\tilde{A}_0(r(t \wedge t_1) + (t - t_1) \vee 0, r)\}\}.$$

Il est facile à montrer que

$$\hat{U}(r) = \int_{t_1}^t \frac{Y_2(\tau) dN_1(t_1 + \frac{\tau-t_1}{r}) - Y_1(t_1 + \frac{\tau-t_1}{r}) dN_2(\tau)}{Y_1(t_1 + \frac{\tau-t_1}{r}) + Y_2(\tau)}$$

ou

$$\hat{U}(r) = \sum_{j: T_{1j} > t_1} \frac{Y_2(t_1 + r(T_{1j} - t_1))}{Y_1(T_{1j}) + Y_2(t_1 + r(T_{1j} - t_1))} - \sum_{j: T_{2j} > t_1} \frac{Y_1(t_1 + \frac{T_{2j}-t_1}{r})}{Y_1(t_1 + \frac{T_{2j}-t_1}{r}) + Y_2(T_{2j})}.$$

La fonction  $\hat{U}$  est décroissante et en escaliers,  $\hat{U}(0) > 0$ ,  $\hat{U}(\infty) < 0$  avec une probabilité 1.

L'estimateur du paramètre  $r$  :

$$\hat{r} = \hat{U}^{-1}(0) = \sup \{r : \hat{U}(r) \geq 0\}.$$

Alors

$$\hat{A}_0(s) = \int_0^s \frac{dN_1(\frac{u}{\hat{r}}) + dN_2[\frac{u}{\hat{r}} \wedge t_1 + (u - \hat{r}t_1) \vee 0]}{Y_1(\frac{u}{\hat{r}}) + Y_2[\frac{u}{\hat{r}} \wedge t_1 + (u - \hat{r}t_1) \vee 0]} = \sum_{j: T_{1j} \leq \frac{s}{\hat{r}}} \frac{1}{Y_1(T_{1j}) + Y_2[T_{1j} \wedge t_1 + \hat{r}(T_{1j} - t_1) \vee 0]} + \sum_{j: T_{2j} \leq \frac{s}{\hat{r}} \wedge t_1 + (s - \hat{r}t_1) \vee 0} \frac{1}{Y_1[T_{2j} \wedge t_1 + \frac{T_{2j}-t_1}{\hat{r}} \vee 0] + Y_2(T_{2j})}.$$

La fonction de survie  $S_{X_0}$  peut être estimée par l'estimateur

$$\hat{S}_{X_0}(t) = \tilde{S}_{X_0}(t, \hat{\beta})$$

ou de façon alternative

$$\hat{S}_{X_0}(t) = \exp\{-\hat{A}_0(t)\}.$$

Les propriétés asymptotiques des estimateurs et construction des intervalles de confiance approximatifs sont données dans Bagdonavičius & Nikulin (1997).



# Chapitre 7

## INFERENCE BAYESIENNE

### 7.1 La règle Bayésienne

Soit  $(X, \Theta)^T$  un vecteur aléatoire à valeurs dans l'espace  $\chi \times \Omega$  et soit  $p(x, \theta)$  sa densité. Alors

$$\pi(\theta) = \int_{\chi} p(x, \theta) dx \quad \text{et} \quad q(x) = \int_{\Omega} p(x, \theta) d\theta \quad (7.1)$$

sont les densités marginales de  $\Theta$  et  $X$ , respectivement. L'approche bayésienne suppose que pendant l'expérience on n'observe que des réalisations de  $X$ , i.e. on suppose que  $X$  est une variable (un vecteur) observable appelée un échantillon. Par contre la deuxième composante  $\Theta$  est inconnue et non observée et est considérée comme un paramètre. Supposons que la densité conditionnelle de  $X$  sachant la valeur de  $\Theta$  est connue. Notons

$$\pi(x|\theta) = \frac{p(x, \theta)}{\pi(\theta)} \quad (7.2)$$

la densité conditionnelle de  $X$  sachant que  $\Theta = \theta$ , et soit

$$q(\theta|x) = \frac{p(x, \theta)}{q(x)} \quad (7.3)$$

la densité conditionnelle de  $\Theta$  sachant que  $X = x$ . Puisque

$$p(x, \theta) = \pi(x|\theta)\pi(\theta) = q(\theta|x)q(x), \quad (7.4)$$

de (1)-(4) on tire les formules de Bayes :

$$q(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{q(x)} = \frac{\pi(x|\theta)\pi(\theta)}{\int_{\Omega} \pi(x|\theta)\pi(\theta) d\theta} \quad (7.5)$$

et

$$\pi(x|\theta) = \frac{q(\theta|x)q(x)}{\pi(\theta)} = \frac{q(\theta|x)q(x)}{\int_{\chi} q(\theta|x)q(x) dx}. \quad (7.6)$$

La densité marginale  $\pi(\theta)$  de  $\Theta$  est appelée la densité à *priori* et la densité conditionnelle  $q(\theta|x)$  de  $\Theta$  sachant  $X = x$  est appelée la densité à *posteriori*  $\square$

**Exemple 1.** Supposons que  $\Theta$  suit la loi normale  $N(\mu, \tau^2)$ , i.e. la densité à priori est

$$\pi(\theta) = \frac{1}{\tau} \varphi\left(\frac{\theta - \mu}{\tau}\right), \theta \in R^1, \quad (7.7)$$

où  $\varphi(u)$  est la densité de la loi normale standard  $N(0, 1)$ ,  $\mu$  et  $\tau$  sont connus,  $|\mu| < \infty, \tau^2 > 0$ . On suppose que la loi conditionnelle de  $X$  sachant que  $\Theta = \theta$  est normale  $N(\theta, \sigma^2)$  i.e., la densité conditionnelle de  $X$  sachant  $\Theta = \theta$  est

$$\pi(x|\theta) = \frac{1}{\sigma} \varphi\left(\frac{x - \theta}{\sigma}\right), \quad (7.8)$$

$\sigma^2 > 0, \sigma^2$  est connu. Calculons la densité  $q(x)$  de la loi marginale de  $X$ . D'après (1)-(3) on a

$$\begin{aligned} q(x) &= \int_{-\infty}^{\infty} \pi(x|\theta)\pi(\theta)d\theta = \int_{-\infty}^{\infty} \frac{1}{\sigma} \varphi\left(\frac{x - \theta}{\sigma}\right) \frac{1}{\tau} \varphi\left(\frac{\theta - \mu}{\tau}\right) d\theta \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\theta - \mu)^2\right\} d\theta \\ &= \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{x^2 - 2x\theta + \theta^2}{\sigma^2} + \frac{\theta^2 - 2\theta\mu + \mu^2}{\tau^2}\right]\right\} d\theta \\ &= \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2}\right)\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{\sigma^2} - 2\theta\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right) + \frac{\theta^2}{\tau^2}\right]\right\} d\theta \\ &= \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2}\right)\right\} \times \\ &\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right) + \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2 - \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2\right]\right\} d\theta \\ &= \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\frac{\tau^2 x^2 + \sigma^2 \mu^2}{\sigma^2 \tau^2}\right\} \times \\ &\times \int_{-\infty}^{\infty} \exp\left\{-\frac{\tau^2 + \sigma^2}{2\sigma^2 \tau^2} \left[\theta^2 - 2\theta \frac{\tau^2 x + \sigma^2 \mu}{\sigma^2 + \tau^2} + \left(\frac{\tau^2 x + \sigma^2 \mu}{\sigma^2 + \tau^2}\right)^2 - \left(\frac{\tau^2 x + \sigma^2 \mu}{\sigma^2 + \tau^2}\right)^2\right]\right\} d\theta \\ &= \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\frac{\tau^2 x^2 + \sigma^2 \mu^2}{\sigma^2 \tau^2}\right\} \times \\ &\int_{-\infty}^{\infty} \exp\left\{-\frac{\tau^2 + \sigma^2}{2\sigma^2 \tau^2} \left(\theta - \frac{\tau^2 x + \sigma^2 \mu}{\sigma^2 + \tau^2}\right)^2\right\} \exp\left\{\frac{(x\tau^2 + \mu\sigma^2)^2}{2\sigma^2 \tau^2 (\sigma^2 + \tau^2)}\right\} d\theta \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp\left\{-\frac{\tau^2 x^2 + \sigma^2 \mu^2}{\sigma^2 \tau^2} + \frac{(x\tau^2 + \mu\sigma^2)^2}{2\sigma^2 \tau^2 (\sigma^2 + \tau^2)}\right\} \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \left\{ -\frac{1}{2(\sigma^2 + \tau^2)} (x - \mu)^2 \right\}, \quad (7.9)$$

i.e. la loi marginale de  $X$  est normale  $N(\mu, \sigma^2 + \tau^2)$  de paramètres

$$\mathbf{E}X = \mu \quad \text{et} \quad \mathbf{Var}X = \sigma^2 + \tau^2. \quad (7.10)$$

D'après (2) la densité  $p(x, \theta)$  du vecteur  $(X, \Theta)^T$  est

$$p(x, \theta) = \pi(x|\theta)\pi(\theta) = \frac{1}{\sigma} \varphi \left( \frac{x - \theta}{\sigma} \right) \frac{1}{\tau} \varphi \left( \frac{\theta - \mu}{\tau} \right). = \quad (7.11)$$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu)^2}{\sigma^2 + \tau^2} - 2\rho \frac{(x-\mu)(\theta-\mu)}{\tau\sqrt{\sigma^2 + \tau^2}} + \frac{(\theta-\mu)^2}{\tau^2} \right] \right\},$$

où

$$\rho^2 = \frac{\tau^2}{\sigma^2 + \tau^2}, \quad \sigma_1^2 = \sigma^2 + \tau^2, \quad \sigma_2^2 = \tau^2,$$

i.e.  $(X, \Theta)^T$  suit la loi normale bidimensionnelle de paramètres  $a = (\mu, \mu)^T$  et  $\Sigma$ , où

$$\Sigma = \begin{vmatrix} \sigma^2 + \tau^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{vmatrix}$$

□

L'inférence statistique sur  $\Theta$  dans l'optique de l'approche bayésienne est donnée en utilisant la densité à posteriori  $q(\theta|x)$  basée sur l'échantillon  $X$ , puisque toute information probabiliste sur  $\Theta$  est exprimée en termes de  $q(\theta|X)$ . S'il est nécessaire d'estimer la valeur  $U(\theta)$ , où  $\theta$  est une réalisation non-observée du paramètre aléatoire  $\Theta$ , alors on utilise l'espérance conditionnelle  $\mathbf{E}\{U(\Theta)|X\}$  comme l'estimateur ponctuel pour  $U(\theta)$ . □

## 7.2 Estimation ponctuelle

Supposons que pendant une expérience une réalisation de  $X$  est observée et la réalisation correspondante de  $\Theta$  est inconnue. Il faut estimer la valeur  $\theta$  de la réalisation non observée de  $\Theta$ . Soit

$$\Theta^* = \Theta^*(X) \quad (7.1)$$

un estimateur ponctuel de  $\theta$ . L'erreur systématique de  $\Theta^*$  est

$$\mathbf{E}\{\Theta^* - \Theta|X\} = \mathbf{E}\{\Theta^*|X\} - \mathbf{E}\{\Theta|X\}, \quad (7.2)$$

où

$$\mathbf{E}\{\Theta|X = x\} = \int_{\Omega} \theta q(\theta|x) d\theta \quad \text{et} \quad \mathbf{E}\{\Theta^*|X = x\} = \Theta^*(x). \quad (7.3)$$

**Définition 1.** L'estimateur  $\hat{\Theta}(X)$  est sans biais si l'erreur systématique est égale à zéro, i.e., si

$$\hat{\Theta}(x) \equiv \mathbf{E}\{\Theta|X = x\}. \quad (7.4)$$

Il s'ensuit que l'estimateur sans biais est unique presque sûrement.  
Soit  $\tilde{\Theta} = \tilde{\Theta}(X)$  un autre estimateur de  $\theta$ . Puisque

$$\tilde{\Theta}(x) - \Theta = [\tilde{\Theta}(x) - \hat{\Theta}(x)] + [\hat{\Theta}(x) - \Theta],$$

on a

$$\begin{aligned} & \mathbf{E}\{(\tilde{\Theta} - \Theta)^2 | X = x\} \\ &= [\tilde{\Theta}(x) - \hat{\Theta}(x)]^2 + 2[\tilde{\Theta}(x) - \hat{\Theta}(x)]\mathbf{E}\{\hat{\Theta}(X) - \Theta | X = x\} + \mathbf{E}\{[\hat{\Theta}(X) - \Theta]^2 | X = x\} \\ &= [\tilde{\Theta}(x) - \hat{\Theta}(x)]^2 + \mathbf{E}\{[\hat{\Theta}(X) - \Theta]^2 | X = x\} \geq \mathbf{E}\{[\hat{\Theta}(X) - \Theta]^2 | X = x\}. \end{aligned} \quad (7.5)$$

Donc l'estimateur sans biais  $\hat{\Theta}$  minimize le risque quadratique à *posteriori*. L'inégalité implique qu'avec la probabilité 1

$$\mathbf{E}\{(\tilde{\Theta} - \Theta)^2 | X\} \geq \mathbf{E}\{(\hat{\Theta} - \Theta)^2 | X\}. \quad (7.6)$$

Prenant l'espérance de la gauche et de la droite, on a

$$\mathbf{E}\{(\tilde{\Theta} - \Theta)^2 | X\} \geq \mathbf{E}\{(\hat{\Theta} - \Theta)^2\} \quad (7.7)$$

L'inégalité (7) implique que l'estimateur sans biais

$$\hat{\Theta}(X) = \mathbf{E}\{\Theta | X\} = \int_{\Omega} \theta q(\theta | X) d\theta \quad (7.8)$$

est le meilleur dans le sens du minimum du risque quadratique.  $\square$

**Définition 2.** L'estimateur sans biais  $\hat{\Theta}(X)$  est appelé l'estimateur bayésien.

**Remarque 1.** L'estimateur bayésien  $\hat{\Theta}$  est l'espérance de la *répartition à posteriori*.  $\square$

**Remarque 2.** De (1.1) - (1.4) on a

$$\mathbf{E}q(\theta | X) = \int_{\mathcal{X}} q(\theta) | x q(x) dx = \int_{\mathcal{X}} p(x, \theta) dx = \pi(\theta),$$

i.e.,  $q(\theta | X)$  l'estimateur sans biais de la densité à priori  $\pi(\theta)$ .  $\square$

**Exemple 1.** Soit  $(X, \Theta)^T$  un vecteur aléatoire où  $\Theta$  est une variable aléatoire suivant la loi uniforme  $[0, 1]$ , i.e. la densité à priori est

$$\pi(\theta) = \begin{cases} 1, & \text{si } \theta \in \Omega = [0, 1], \\ 0, & \text{sinon.} \end{cases} \quad (7.9)$$

et la répartition conditionnelle de  $X$  sachant  $\Theta = \theta$  est la répartition de Bernoulli  $B(1, \theta)$ , i.e.,

$$\pi(x | \theta) = \begin{cases} \theta^x (1 - \theta)^{1-x}, & x \in \mathcal{X} = \{0, 1\}, \\ 0, & \text{sinon.} \end{cases} \quad (7.10)$$

On peut estimer  $\theta$  en utilisant l'estimateur du maximum de vraisemblance  $X$ , qui est le meilleur estimateur sans biais pour  $\theta$ , et le risque quadratique de  $X$  est  $\theta(1 - \theta)$ .

On va construire l'estimateur bayésien  $\hat{\Theta} = \hat{\Theta}(X)$ . De (1.5), (9) (10) on a

$$q(\theta | x) = \frac{\pi(x | \theta) \pi(\theta)}{\int_{\Omega} \pi(x | \theta) \pi(\theta) d\theta} = \begin{cases} \frac{\theta^x (1 - \theta)^{1-x}}{\int_0^1 \theta^x (1 - \theta)^{1-x} d\theta}, & \text{si, } x \in \mathcal{X}, \\ 0, & \text{sinon,} \end{cases}$$

$$= \begin{cases} 2(1-\theta), & \text{si } x=0, \theta \in \Omega, \\ 2\theta, & \text{si } x=1, \theta \in \Omega, \\ 0, & \text{sinon.} \end{cases}$$

Supposons que  $X = 0$ . Alors

$$\hat{\Theta}(0) = \int_0^1 \theta q(\theta|0) d\theta = 2 \int_0^1 \theta(1-\theta) d\theta = \frac{1}{3}.$$

Dans le cas  $X = 1$  on a

$$\hat{\Theta}(1) = \int_0^1 \theta q(\theta|1) d\theta = 2 \int_0^1 \theta^2 d\theta = \frac{2}{3}.$$

Le risque quadratique de l'estimateur bayésien  $\hat{\Theta} = \hat{\Theta}(X)$  est

$$\begin{aligned} \mathbf{E}\{(\hat{\Theta} - \Theta)^2 | \Theta = \theta\} &= \mathbf{E}\{(\hat{\Theta} - \theta)^2\} = \left(\frac{1}{3} - \theta\right)^2 \mathbf{P}\{X = 0\} + \left(\frac{2}{3} - \theta\right)^2 \mathbf{P}\{X = 1\} \\ &= \left(\frac{1}{3} - \theta\right)^2 (1 - \theta) + \left(\frac{2}{3} - \theta\right)^2 \theta = \frac{1}{3}(\theta^2 - \theta + \frac{1}{3}). \end{aligned}$$

Il peut être comparé avec le risque quadratique  $\theta(1-\theta)$  de l'estimateur de maximum de vraisemblance  $X$  de  $\theta$ .  $\square$

**Exemple 2.** Soit  $(X, \Theta)^T$  le modèle bayésien où

$$\Theta \sim N(\mu, \tau^2),$$

et la répartition conditionnelle de  $X$  sachant  $\Theta = \theta$  est normale  $N(\theta, \sigma^2)$ ,  $\mu, \tau^2, \sigma^2$  sont connus. Dans l'exemple 1.1 on a été montré que la répartition marginale  $q(x)$  de  $X$  est normale  $N(\mu, \sigma^2 + \tau^2)$ , i.e.,

$$q(x) = \frac{1}{\sqrt{2\pi\sqrt{\sigma^2 + \tau^2}}} \exp\left\{-\frac{1}{2(\sigma^2 + \tau^2)}(x - \mu)^2\right\} = \frac{1}{\sqrt{\sigma^2 + \tau^2}} \varphi\left(\frac{x - \mu}{\sqrt{\sigma^2 + \tau^2}}\right)$$

et la densité de  $(X, \Theta)^T$  est

$$p(x, \theta) = \pi(x|\theta)\pi(\theta) = \frac{1}{\sigma} \varphi\left(\frac{x - \theta}{\sigma}\right) \frac{1}{\tau} \varphi\left(\frac{\theta - \mu}{\tau}\right), \quad (7.11)$$

d'où la densité à posteriori  $q(\theta|x)$  peut être trouvé :

$$\begin{aligned} q(\theta|x) &= \frac{p(x, \theta)}{q(x)} = \frac{\frac{1}{\sigma} \varphi\left(\frac{x - \theta}{\sigma}\right) \frac{1}{\tau} \varphi\left(\frac{\theta - \mu}{\tau}\right)}{\frac{1}{\sqrt{\sigma^2 + \tau^2}} \varphi\left(\frac{x - \mu}{\sqrt{\sigma^2 + \tau^2}}\right)} \\ &= \frac{\sqrt{\sigma^2 + \tau^2}}{\sigma\tau} \varphi\left\{\frac{\sqrt{\sigma^2 + \tau^2}}{\sigma\tau} \left[\theta - \sigma^2 \rho^2 \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right]\right\}, \\ &= \frac{1}{\rho\sigma} \varphi\left\{\frac{1}{\rho\sigma} \left[\theta - \sigma^2 \rho^2 \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right]\right\}, \end{aligned} \quad (7.12)$$

où  $\rho^2 = \frac{\tau^2}{\sigma^2 + \tau^2}$ . i.e. la répartition à posteriori est normale de paramètres

$$\sigma^2 \rho^2 \left( \frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right) \quad \text{et} \quad \rho^2 \sigma^2 :$$

$$\mathbf{P}\{\Theta \leq \theta | X = x\} = \Phi \left\{ \frac{1}{\rho \sigma} \left[ \theta - \sigma^2 \rho^2 \left( \frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right) \right] \right\},$$

d'où l'estimateur bayésien est obtenu :

$$\hat{\Theta} = E\{\Theta | X\} = \int_{-\infty}^{\infty} \theta q(\theta | X) d\theta = \sigma^2 \rho^2 \left( \frac{X}{\sigma^2} + \frac{\mu}{\tau^2} \right).$$

La statistique  $X$  est l'estimateur de maximum de vraisemblance de  $\theta$  et est le meilleur estimateur sans biais. Notons que si  $\sigma^2$  est fixé et  $\tau^2 \rightarrow \infty$ , alors  $\rho^2 \rightarrow 1$ , d'où

$$\hat{\Theta}(x) \rightarrow x.$$

Ca signifie que quand  $\tau$  est grand alors

$$\hat{\Theta}(X) \approx X,$$

et le gain d'utilisation de l'estimateur bayésien est petit. De même, si  $\tau^2$  est fixé et  $\sigma^2 \rightarrow 0$ . D'autre part si  $\tau^2 \rightarrow 1$  et  $\sigma^2 \rightarrow 0$ , alors  $\rho^2 \rightarrow 1$  et  $\hat{\Theta}(x) \rightarrow \mu$ , i.e.  $\hat{\Theta} \approx \mu$ . Donc  $\tau^2$  et  $\sigma^2$  représentent les poids relatifs donnés à  $X$ , et à la moyenne à priori  $\mu$ .  $\square$

**Remarque 3.** Soit  $T = T(X)$  la statistique exhaustive pour  $\theta$  dans le modèle bayésien  $(X, \Theta)^T$  de paramètre  $\Theta, \Theta \in \Omega$ . Alors, pour toute répartition à priori  $\pi(\theta)$  on a

$$q(\theta | x) = q^*(\theta | t(x)),$$

où  $q(\theta | x)$  est la densité à posteriori de  $\Theta$  sachant  $X$  et  $q^*(\theta | t(x))$  est la densité à posteriori de  $\Theta$  sachant  $T$ . En effet, d'après la règle bayésienne on a

$$q(\theta | x) = \frac{\pi(x|\theta)\pi(\theta)}{q(x)}.$$

Puisque  $T$  est la statistique exhaustive pour  $\theta$  on a

$$p(x, \theta) = \pi(x|\theta)\pi(\theta) = \pi^*(T(x)|\theta)\pi(\theta)w(x)$$

où  $w(x)$  est une fonction nonnégative. Donc

$$\pi(x|\theta) = \pi^*(T(x)|\theta)w(x).$$

Il s'ensuit immédiatement que

$$q(\theta | x) = q^*(\theta | t(x)).$$

Notons que dans l'approche bayésienne la notion de l'exhaustivité joue le même rôle comme dans la statistique classique. De plus, la statistique  $T = T(X)$  est exhaustive si la répartition à posteriori de  $\Theta$ , sachant  $T$ , est la même que la répartition à posteriori de  $\Theta$ , sachant  $X$ . Donc,

$$\hat{\Theta} = \hat{\Theta}(T) = \frac{\int \theta \pi^*(T|\theta) w(X) \pi(\theta) d\theta}{\int \pi^*(T|\theta) w(X) \pi(\theta) d\theta} =$$

$$\frac{\int \theta \pi^*(T|\theta) \pi(\theta) d\theta}{\int \pi^*(T|\theta) \pi(\theta) d\theta} = \frac{\int \theta \pi^*(T|\theta) \pi(\theta) d\theta}{q^*(T)}. \square$$

**Exemple 3.** Supposons que sachant  $\Theta = \theta$  les composantes  $X_1, X_2, \dots, X_n$  du vecteur observé  $X = (X_1, X_2, \dots, X_n)^T$  sont des variables indépendantes Bernoulli  $B(1, \theta)$ ,  $X_i$  prend la valeur 1 avec la probabilité  $\theta$  et la valeur 0 avec la probabilité  $1 - \theta$ ,  $\theta \in \Omega = [0, 1]$ , i.e., pour tout  $i = 1, 2, \dots, n$  on a

$$\mathbf{P}\{X_i = x | \Theta = \theta\} = \theta^x (1 - \theta)^{1-x}, \quad x \in \mathcal{X} = \{0, 1\}.$$

Alors  $T_n = X_1 + X_2 + \dots + X_n$  est la statistique exhaustive et

$$\mathbf{P}\{T_n = k | \Theta = \theta\} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \quad k = 0, 1, \dots, n. \quad (7.13)$$

L'estimateur de maximum de vraisemblance  $\hat{\theta}$  de  $\theta$  est le meilleur estimateur de  $\theta$ ,

$$\hat{\theta} = \bar{X}_n = \frac{1}{n} T_n \quad (7.14)$$

$$\mathbf{E}\{\bar{X}_n | \Theta = \theta\} = \theta \text{ et } \mathbf{Var}\{\bar{X}_n | \Theta = \theta\} = \frac{\theta(1 - \theta)}{n}. \quad (7.15)$$

Construisons l'estimateur bayésien  $\hat{\Theta}$ . La densité de la répartition conditionnelle de la v.a.  $X_i$  sachant que  $\Theta = \theta$  est donné par la formule

$$\pi(x|\theta) = \theta^x (1 - \theta)^{1-x}, \quad x \in \mathcal{X} = \{0, 1\}.$$

Supposons que  $x_1, x_2, \dots, x_n$  sont les valeurs observées des variables aléatoires  $X_1, X_2, \dots, X_n$ . Dans ce cas la densité de la répartition à posteriori du paramètre  $\Theta$  sachant que

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$$

est

$$q(\theta|x_1, \dots, x_n) = q^*(\theta|t) = \frac{\pi(\theta) \theta^t (1 - \theta)^{n-t}}{\int_0^1 \pi(\theta) \theta^t (1 - \theta)^{n-t} d\theta} = \frac{\theta^t (1 - \theta)^{n-t}}{\int_0^1 \theta^t (1 - \theta)^{n-t} d\theta}, \quad (7.16)$$

où  $t = x_1 + x_2 + \dots + x_n$  est une réalisation de la statistique exhaustive

$$T_n = X_1 + X_2 + \dots + X_n, \quad (7.17)$$

et  $q^*(\theta|t)$  est la densité de la répartition à posteriori de  $\Theta$  sachant  $T_n$ . De plus, supposons que  $\Theta$  suit la loi uniforme sur  $\Omega = [0, 1]$ , i.e., la densité  $\pi(\theta)$  de la répartition à priori est

$$\pi(\theta) = \begin{cases} 1, & \text{si } \theta \in \Omega = [0, 1], \\ 0, & \text{sinon.} \end{cases} \quad (7.18)$$

Puisque

$$\mathbf{P}\{T_n \leq t | \Theta = \theta\} = \sum_{k=0}^t \binom{n}{k} \theta^k (1 - \theta)^{n-k} \equiv I_{1-\theta}(n-t, t+1) \equiv 1 - I_\theta(t+1, n-t) \quad (7.19)$$

et

$$\theta^t(1-\theta)^{n-t} \equiv \frac{\Gamma(t+1)\Gamma(n-t+1)}{\Gamma(n+2)} \frac{d}{d\theta} I_{\theta}(t+1, n-t+1), \quad (7.20)$$

on a

$$\int_0^1 \theta^t(1-\theta)^{n-t} d\theta = \frac{t!(n-t)!}{(n+1)!} = \frac{\Gamma(t+1)\Gamma(n-t+1)}{\Gamma(n+2)} \quad (7.21)$$

et donc

$$q^*(\theta|t) = \frac{\theta^t(1-\theta)^{n-t}}{\int_0^1 \theta^t(1-\theta)^{n-t} d\theta}$$

$$\frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)} \theta^t(1-\theta)^{n-t} = \frac{1}{B(t+1, n-t+1)} \theta^t(1-\theta)^{n-t}. \quad (7.22)$$

Pour tout  $t = 0, 1, \dots, n$ , la fonction  $I_{\theta}(t+1, n-t+1)$ , comme la fonction de  $\theta$  dans l'intervalle  $[0, 1]$  est la fonction de répartition, voir §2.3, avec la densité

$$f_{\beta}(\theta; t+1, n-t+1) = \frac{1}{B(t+1, n-t+1)} \theta^t(1-\theta)^{n-t}, \theta \in \Omega = [0, 1], \quad (7.23)$$

de la loi beta, i.e., pour tout  $\theta \in \Omega = [0, 1]$  on a

$$\mathbf{P}\{\Theta \leq \theta | T_n = t\} = I_{\theta}(t+1, n-t+1). \quad (7.24)$$

Soit  $Z$  une variable de la loi donnée par (23). Dans ce cas

$$\mathbf{E}Z^k = \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)} \frac{\Gamma(t+k+1)\Gamma(n-t+1)}{\Gamma(n+k+2)} = \frac{\Gamma(n+2)\Gamma(t+k+1)}{\Gamma(n+k+2)\Gamma(t+1)}.$$

Cette formule et (24) impliquent

$$\mathbf{E}\{\Theta | T_n = t\} = \frac{t+1}{n+2} \quad \text{et} \quad \mathbf{E}\{\Theta^2 | T_n = t\} = \frac{(t+1)(t+2)}{(n+2)(n+3)}, \quad (7.25)$$

i.e.,

$$\begin{aligned} \mathbf{Var}\{\Theta | T_n = t\} &= \mathbf{E}\{\Theta^2 | T_n = t\} - (\mathbf{E}\{\Theta | T_n = t\})^2 \\ &= \frac{(t+1)(n-t+1)}{(n+2)^2(n+3)} = \frac{1}{n+3} \frac{t+1}{n+2} \left(1 - \frac{t+1}{n+2}\right). \end{aligned} \quad (7.26)$$

Utilisant (16), (22), (23) et (25) on a

$$\hat{\Theta} = \mathbf{E}\{\Theta | X_1, \dots, X_n\} = \mathbf{E}\{\Theta | T_n\} = \frac{T_n + 1}{n + 2}. \quad (7.27)$$

(27) et (14) impliquent que pour grandes valeurs de  $n$  on a

$$\hat{\Theta} \sim \hat{\theta}_n \quad (7.28)$$

et (26) implique

$$\mathbf{Var}\{\hat{\Theta} | X_1 = x_1, \dots, X_n = x_n\} = \mathbf{Var}\{\hat{\Theta} | T_n = t\} \sim \frac{t}{n} \left(1 - \frac{t}{n}\right). \square \quad (7.29)$$

**Exemple 4.** (continuation de l'Exemple 2). Supposons que la densité à priori  $\pi(\theta)$  dans l'exemple 2 suit la loi beta de paramètres  $a$  et  $b, a > 0, b > 0$  :

$$\pi(\theta) = \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1}, \theta \in \Omega = [0, 1]. \quad (7.30)$$

Il est évident que si  $a = b = 1$  alors on a la densité  $\pi(\theta)$  de la loi uniforme sur  $[0, 1]$ , considérée en (18). La statistique

$$T_n = X_1 + X_2 + \dots + X_n$$

est suffisante et la loi conditionnelle de  $T_n$ , sachant  $\Theta = \theta$ , est donnée par (13). De (16) on a que la densité à posteriori  $q(\theta|t)$  sachant  $T_n = t$ ,

$$q(\theta|t) = \frac{\theta^{a+t-1} (1-\theta)^{b+n-1}}{\int_0^1 \theta^{a+t-1} (1-\theta)^{b+n-t-1} d\theta} = \frac{\theta^{a+t-1} (1-\theta)^{b+n-t-1}}{B(a+t, b+n-t)}, \quad (7.31)$$

i.e. c'est la densité beta de paramètres  $a+t$  et  $b+n-t$ , d'où l'estimateur bayésien  $\hat{\Theta} = \hat{\Theta}(T_n)$  est

$$\hat{\Theta} = \int_0^1 \theta q(\theta|T_n) d\theta = \frac{1}{B(a+T_n, b+n-T_n)} \int_0^1 \theta^{a+T_n} (1-\theta)^{b+n-T_n-1} d\theta = \frac{T_n + a}{a + b + n}. \quad (7.32)$$

On peut voir que pour grands valeurs de  $n$  on a

$$\hat{\Theta} \sim \hat{\theta}_n \quad (7.33)$$

pour tout  $a$  et  $b, a > 0, b > 0$ .  $\square$

**Exemple 5.** Supposons que, sachant  $\Theta$ , le vecteur aléatoire  $X = (X_1, \dots, X_n)^T$  est un échantillon de la loi exponentielle de la moyenne  $1/\Theta$ , i.e. , la densité conditionnelle  $\pi(x|\theta)$  de  $X_i$  sachant  $\Theta = \theta$  est

$$\pi(x|\theta) = \begin{cases} \theta e^{-\theta x}, & x > 0, \\ 0, & \text{sinon,} \end{cases} \quad (7.34)$$

$$\mathbf{E}\{X_i|\Theta = \theta\} = \frac{1}{\theta}, \quad \mathbf{Var}\{X_i|\Theta = \theta\} = \frac{1}{\theta^2}. \quad (7.35)$$

Trouvons l'estimateur bayésien pour

$$\mathbf{P}\{X_i > t|\Theta = \theta\} = e^{-\theta t}. \quad (7.36)$$

On suppose que la répartition à priori est gamma de  $p$  degrés de liberté et de paramètre d'échelle  $\lambda$  ( $p$  et  $\lambda$  sont connus), i.e., la densité  $\pi(\theta)$  de  $\Theta, \Theta \in \Omega = (0, +\infty)$ , est

$$\pi(\theta) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} \theta^{p-1} e^{-\lambda\theta}, & \theta > 0, \\ 0, & \text{sinon.} \end{cases} \quad (7.37)$$

On sait que

$$\mathbf{E}\Theta = \frac{\lambda}{p} \text{ et } \mathbf{Var}\Theta = \frac{\lambda}{p^2}. \quad (7.38)$$

Dans ce modèle la statistique

$$T_n = X_1 + \cdots + X_n \quad (7.39)$$

est exhaustive pour  $\theta$ , et

$$\mathbf{P}\{T_n \leq t | \Theta = \theta\} = \frac{\theta^n}{\Gamma(n)} \int_0^t x^{n-1} e^{-\theta x} dx, \quad t \geq 0, \quad (7.40)$$

i.e. sachant  $\Theta = \theta$  la statistique exhaustive  $T_n$  suit la loi gamma de  $n$  degrés de liberté et de paramètre  $\theta$ , i.e. la densité conditionnelle  $\pi(t|\theta)$  de la statistique exhaustive  $T_n$  sachant  $\Theta$  est

$$\pi(t|\theta) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t}, \quad t > 0, \quad \theta \in \Omega = (0, \infty), \quad (7.41)$$

et

$$\mathbf{E}\{T_n | \Theta = \theta\} = \frac{n}{\theta}, \quad \mathbf{Var}\{T_n | \Theta = \theta\} = \frac{n}{\theta^2}. \quad (7.42)$$

D'après (1.5) la densité  $q(\theta|t)$  de la loi à posteriori, i.e. la densité de  $\Theta$  sachant  $T_n = t$ , est

$$\begin{aligned} q(\theta|t) &= \frac{\pi(t|\theta)\pi(\theta)}{\int_0^\infty \pi(x|\theta)\pi(\theta)d\theta} = \frac{\theta^{n+p-1} e^{-\theta(t+\lambda)}}{\int_0^\infty \theta^{n+p-1} e^{-\theta(t+\lambda)} d\theta} = \\ &= \frac{(t+\lambda)^{n+p} \theta^{n+p-1} e^{-\theta(t+\lambda)}}{\int_0^\infty u^{n+p-1} e^{-u} du} = \frac{(t+\lambda)^{n+p}}{\Gamma(n+p)} \theta^{n+p-1} e^{-\theta(t+\lambda)}, \end{aligned} \quad (7.43)$$

i.e.,  $q(\theta|t)$  est la densité de la loi gamma de  $n+p$  degrés de liberté et de paramètre d'échelle  $t+\lambda$ . On peut trouver l'estimateur bayésien  $S(T_n)$  pour  $e^{-\theta t}$ . On a

$$\begin{aligned} S(T_n) &= \mathbf{E}\{e^{-\theta t} | T_n\} = \int_0^\infty e^{-\theta t} q(\theta|T_n) d\theta = \\ &= \int_0^\infty e^{-\theta t} \frac{(\lambda + T_n)^{n+p}}{\Gamma(n+p)} \theta^{n+p-1} \exp[-\theta(T_n + \lambda)] d\theta = \\ &= \frac{(\lambda + T_n)^{n+p}}{(\lambda + t + T_n)^{n+p}} \frac{1}{\Gamma(n+p)} \int_0^\infty u^{n+p-1} e^{-u} du = \left(1 + \frac{t}{\lambda + T_n}\right)^{-(n+p)}. \end{aligned} \quad (7.44)$$

On peut vérifier, voir par exemple Voinov & Nikulin (1993), que le meilleur estimateur sans biais de  $e^{-\theta t}$  est

$$U(T_n) = \begin{cases} \left(1 - \frac{t}{T_n}\right)^{n-1}, & \text{si } T_n > t, \\ 0, & \text{sinon. } \square \end{cases} \quad (7.45)$$

**Exemple 6.** Soit  $(X, \Theta)^T$  le modèle bayésien, où  $\Theta \sim N(0, 1)$ , i.e. la densité à priori  $\pi(\theta)$  de  $\Theta$  est la densité de la loi standard normale

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right) = \varphi(\theta), \theta \in \mathbb{R}^1. \quad (7.46)$$

De plus, supposons que sachant  $\Theta$ ,  $X = (X_1, \dots, X_n)^T$  est un échantillon de la loi normale  $N(\Theta, 1)$ . Dans ce cas, sachant  $\Theta = \theta$ , la statistique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

est suffisante pour  $\theta$ , et puisque la loi conditionnelle de  $\bar{X}_n$  est  $N(\theta, \frac{1}{n})$  i.e.

$$\mathbf{P}\{\bar{X}_n \leq x | \Theta = \theta\} = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{n}{2}(u - \theta)^2\right\} du = \Phi[\sqrt{n}(x - \theta)],$$

la densité conditionnelle de  $\bar{X}_n$ , sachant  $\Theta = \theta$ , est

$$\pi(x|\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}(x - \theta)^2\right\} = \sqrt{n}\phi(\sqrt{n}(x - \theta)), \quad x \in \mathbb{R}^1. \quad (7.47)$$

Utilisant (46) et (47) on peut trouver la densité  $q(x)$  de la loi marginale de  $\bar{X}_n$  :

$$\begin{aligned} q(x) &= \int_{-\infty}^{\infty} \pi(x|\theta)\pi(\theta)d\theta = \int_{-\infty}^{\infty} \sqrt{n}\phi(\sqrt{n}(x - \theta))\pi(\theta)d\theta = \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{n}{2}(x - \theta)^2\right\} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right) d\theta \\ &= \frac{\sqrt{n}}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{nx^2}{2} + nx\theta - \frac{n\theta^2}{2} - \frac{\theta^2}{2}\right\} d\theta \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+1}} \exp\left\{-\frac{nx^2}{2(n+1)}\right\} = \sqrt{\frac{n}{n+1}} \phi\left(\sqrt{\frac{n}{n+1}}x\right), \end{aligned} \quad (7.48)$$

i.e., la loi marginale de  $\bar{X}_n$  est normale  $N(0, \frac{n+1}{n})$  de paramètres 0 et  $(n+1)/n$ . De la formule de Bayes on peut obtenir la densité  $q(\theta|x)$  de la loi à posteriori, sachant  $\bar{X}_n = x$  :

$$\begin{aligned} q(\theta|x) &= \frac{\pi(x|\theta)\pi(\theta)}{q(x)} = \frac{\sqrt{n}\phi(\sqrt{n}(x - \theta))\pi(\theta)}{\sqrt{\frac{n}{n+1}}\phi\left(x\sqrt{\frac{n}{n+1}}\right)} \\ &= \frac{\sqrt{n+1}}{\sqrt{2\pi}} \exp\left\{-\frac{nx^2}{2} + \frac{nx^2}{2(n+1)} + nx\theta - \frac{\theta^2}{2}(n+1)\right\} \\ &= \frac{\sqrt{n+1}}{\sqrt{2\pi}} \exp\left\{-\frac{n+1}{2}\left(\theta - \frac{nx}{n+1}\right)^2\right\} = \sqrt{n+1}\phi\left[\sqrt{n+1}\left(\theta - \frac{nx}{n+1}\right)\right], \end{aligned} \quad (7.49)$$

i.e. la loi à posteriori de  $\Theta$ , sachant  $\bar{X}_n = x$ , est normale  $N\left(\frac{nx}{n+1}, \frac{1}{n+1}\right)$ ,

$$\mathbf{P}\{\Theta \leq \theta | \bar{X}_n = x\} = \Phi\left[\sqrt{n+1}\left(\theta - \frac{nx}{n+1}\right)\right],$$

d'où l'estimateur bayésien  $\hat{\Theta}$  de  $\theta$  est

$$\hat{\Theta} = \mathbf{E}\{\Theta | \bar{X}_n\} = \int_{-\infty}^{\infty} \theta q(\theta | \bar{X}_n) d\theta = \bar{X}_n \left(1 - \frac{1}{n+1}\right).$$

Nous savons déjà que l'estimateur de maximum de vraisemblance, qui est le meilleur estimateur sans biais pour  $\theta$ , est  $\hat{\theta}_n = \bar{X}_n$ .

On peut vérifier que si  $\Theta \sim N(\mu, \tau^2)$ , où  $\mu$  et  $\tau^2$  sont connus,  $\tau^2 > 0$ , et sachant  $\Theta = \theta$  on a  $X_i \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  est connu,  $\sigma^2 > 0$ . Alors

$$\mathbf{P}\{\Theta \leq \theta | \bar{X}_n\} = \Phi\left\{\frac{\sqrt{n\tau^2 + \sigma^2}}{\sigma\tau} \left(\frac{n\tau^2}{n\tau^2 + \sigma^2}\bar{X}_n + \frac{\sigma^2}{n\tau^2 + \sigma^2}\mu\right)\right\}, \quad (7.50)$$

i.e.,

$$\mathbf{E}\{\Theta|\bar{X}_n\} = \frac{n\tau^2}{n\tau^2 + \sigma^2}\bar{X}_n + \frac{\sigma^2}{n\tau^2 + \sigma^2}\mu \quad \text{et} \quad \mathbf{Var}\{\Theta|\bar{X}_n\} = \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}. \square$$

**Exemple 7.** Soit  $(X, \Theta)^T$  le modèle bayésien de paramètre  $\Theta, \Theta \in \Omega = (0, \infty)$ . Sachant  $\Theta$ , soit  $X$  la variable aléatoire de Poisson de paramètre  $\theta$  :

$$\mathbf{P}\{X = x|\Theta = \theta\} = \frac{\theta^x}{x!}e^{-\theta}, x = 0, 1, \dots \quad (7.51)$$

Supposons que la densité à priori  $\pi(\theta)$  est la densité de la loi gamma de  $m$  degrés de liberté et de paramètre d'échelle  $\alpha$ , i.e.

$$\pi(\theta) = \frac{\alpha^m}{\Gamma(m)}\theta^{m-1}e^{-\alpha\theta}, \quad (7.52)$$

$\alpha$  et  $m$  sont connus,  $\alpha > 0, m > 0$ . Dans ce cas, la densité marginale  $q(x)$  de  $X$  est

$$\begin{aligned} q(x) &= \int_0^\infty \pi(x|\theta)\pi(\theta)d\theta = \int_0^\infty \frac{\theta^x}{x!}e^{-\theta} \frac{\alpha^m}{\Gamma(m)}\theta^{m-1}e^{-\alpha\theta}d\theta \\ &= \frac{\alpha^m}{x!\Gamma(m)} \int_0^\infty \theta^{x+m-1}e^{-\theta(\alpha+1)}d\theta = \frac{\alpha^m}{\Gamma(x+1)\Gamma(m)(\alpha+1)^{x+m}} \int_0^\infty u^{x+m-1}e^{-u}du \\ &= \frac{\Gamma(x+m)\alpha^m}{\Gamma(x+1)\Gamma(m)(\alpha+1)^{x+m}} = \frac{\Gamma(x+m)}{\Gamma(x+1)\Gamma(m)} \left(\frac{1}{1+\alpha}\right)^m \left(\frac{\alpha}{1+\alpha}\right)^x, \end{aligned} \quad (7.53)$$

i.e. la loi marginale de  $X$  est la loi binomiale négative, donnée dans la section 0.3. L'estimateur bayésien  $\hat{\Theta} = \hat{\Theta}(X)$  pour  $\theta$  est

$$\begin{aligned} \hat{\Theta} &= \int_0^\infty \theta q(\theta|X)d\theta = \int_0^\infty \frac{\theta\pi(x|\theta)\pi(\theta)}{q(X)}d\theta = \frac{1}{q(X)} \int_0^\infty \theta \frac{\theta^x}{x!}e^{-\theta} \frac{\alpha^m}{\Gamma(m)}\theta^{m-1}e^{-\alpha\theta}d\theta \\ &= \frac{\alpha^m}{\Gamma(m)q(X)\Gamma(X+1)} \int_0^\infty \theta^{X+m}e^{-\theta(\alpha+1)}d\theta \\ &= \frac{\alpha^m\Gamma(X+m+1)}{\Gamma(m)q(X)\Gamma(X+1)(\alpha+1)^{X+m+1}} = \frac{X+m}{1+\alpha}. \square \end{aligned} \quad (7.54)$$

**Remarque 4.** Considérons le modèle bayésien  $(X, \Theta)^T$  de paramètre  $\Theta, \Theta \in \Omega \in \mathbb{R}^1$ , et soit  $q(\theta|x)$  la densité de la loi à posteriori de  $\Theta$  sachant  $X = x$ . Utilisant la densité à posteriori  $q(\theta|x)$  on peut construire  $(1 - \alpha)$ -intervalle de confiance  $(\underline{\Theta}(X), \bar{\Theta}(X))$  pour  $\Theta$ , tel que

$$\mathbf{P}\{\underline{\Theta} \leq \Theta \leq \bar{\Theta}|X = x\} = 1 - \alpha, \quad 0 < \alpha < 0.5. \quad (7.55)$$

En effet, soit  $\beta$  et  $\gamma$  deux nombres positifs tels que  $\beta + \gamma = \alpha$ . Définissons  $\bar{\Theta} = \bar{\Theta}(x, \gamma)$  comme le  $\gamma$ -quantile supérieur de la loi à posteriori, i.e.,  $\bar{\Theta}$  est la racine de l'équation

$$\mathbf{P}\{\Theta \leq \bar{\Theta}|X = x\} = \int_{-\infty}^{\bar{\Theta}(x, \gamma)} q(\theta|x)d\theta = 1 - \gamma. \quad (7.56)$$

De même, on peut trouver le  $\beta$ -quantile inférieur  $\underline{\Theta} = \underline{\Theta}(x, \beta)$  de la loi à posteriori comme la racine de l'équation

$$\mathbf{P}\{\underline{\Theta} \leq \Theta|X = x\} = \int_{-\infty}^{\underline{\Theta}(x, \beta)} q(\theta|x)d\theta = \beta. \quad (7.57)$$

Dans ce cas on obtient l'estimateur par intervalle  $(\underline{\Theta}(X), \overline{\Theta}(X))$  pour  $\Theta$  de coefficient de confiance  $P = 1 - \alpha$  :

$$\mathbf{P}\{\underline{\Theta} \leq \Theta \leq \overline{\Theta} | X = x\} = \int_{\underline{\Theta}}^{\overline{\Theta}} q(\theta|x) d\theta = 1 - \gamma - \beta = 1 - \alpha = P. \quad (7.58)$$

Il existe une autre approche qui permet de construire "le plus court" intervalle de confiance pour  $\Theta$ . Sachant  $X = x$  soit  $I(x, c)$  un ensemble dans  $\Omega$  tel que

$$I(x, c) = \{\theta : q(\theta|x) > c\}, \quad (7.59)$$

où  $c$  est la constante positive, et soit

$$\begin{aligned} P(x, c) &= \int_{I(x, c)} q(\theta|x) d\theta = \mathbf{P}\{\Theta \in I(x, c) | X = x\} \\ &= \mathbf{P}\{q(\Theta|X) > c | X = x\} = 1 - \mathbf{P}\{q(\Theta|X) \leq c | X = x\}. \end{aligned} \quad (7.60)$$

Choisissons  $c = c_\alpha$  tel que le coefficient de confiance  $P = 1 - \alpha$ , i.e., tel que

$$P(x, c) = P = 1 - \alpha. \quad (7.61)$$

Dans ce cas  $I(X, c_\alpha)$  est un estimateur par intervalle de  $\Theta$  de coefficient de confiance  $P = 1 - \alpha$ , on le tire de (58) et (60). Montrons que  $I(X, c_\alpha)$  est le plus court intervalle de confiance pour  $\Theta$  entre tous les intervalles avec le même coefficient de confiance  $P$ . En effet, soit  $J(X)$  un autre intervalle de confiance pour  $\Theta$ ,  $J(X) \subseteq \Omega$ , tel que

$$\mathbf{P}\{\Theta \in J(X) | X = x\} = P = 1 - \alpha. \quad (7.62)$$

Notons que

$$I = (I \cap J) \cup [I \setminus (I \cap J)] = (I \cap J) \cup \Delta_I \quad (7.63)$$

et

$$J = (I \cap J) \cup [J \setminus (I \cap J)] = (I \cap J) \cup \Delta_J. \quad (7.64)$$

De (58) et (60) et de la définition de  $\Delta_I$  et  $\Delta_J$  on a

$$\int_{\Delta_I} q(\theta|x) d\theta = \int_{\Delta_J} q(\theta|x) d\theta. \quad (7.65)$$

D'un autre côté on a

$$c_\alpha \text{mes} \Delta_J \geq \int_{\Delta_J} q(\theta|x) d\theta = \int_{\Delta_I} q(\theta|x) d\theta \geq c_\alpha \text{mes} \Delta_I, \quad (7.66)$$

d'où

$$\text{mes} \Delta_J \geq \text{mes} \Delta_I, \quad (7.67)$$

où

$$\text{mes} \Delta_J = \int_{\Delta_J} d\theta. \square$$

**Exemple 8.** Soit  $(X, \Theta)^T$  le modèle bayésien de paramètre  $\Theta$ ,  $\Theta \sim N(0, 1)$ . Sachant  $\Theta$ , les éléments  $X_1, X_2, \dots, X_n$  d'échantillon  $X$  sont les variables normales  $N(\Theta, 1)$  indépendantes. La loi à posteriori de  $\Theta$  est normale de paramètres

$$\mathbf{E}\{\Theta | \bar{X}_n\} = \bar{X}_n \left(1 - \frac{1}{n+1}\right) \quad \text{et} \quad \mathbf{Var}\{\Theta | \bar{X}_n\} = \frac{1}{n+1}, \quad (7.68)$$

où  $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$ . De la symétrie de la densité de la loi normale on tire que le plus court  $(1 - \alpha)$ -intervalle de confiance pour  $\Theta$  est

$$\left( \bar{X}_n \left( 1 - \frac{1}{n+1} \right) - \frac{x_{\alpha/2}}{\sqrt{n+1}}; \bar{X}_n \left( 1 - \frac{1}{n+1} \right) + \frac{x_{\alpha/2}}{\sqrt{n+1}} \right) \quad (7.69)$$

On peut remarquer que cet intervalle bayésien est plus court que  $(1 - \alpha)$ -intervalle de confiance classique

$$\left( \bar{X}_n - \frac{x_{\alpha/2}}{\sqrt{n}}; \bar{X}_n + \frac{x_{\alpha/2}}{\sqrt{n}} \right)$$

## 7.3 Approche bayésienne empirique

L'approche bayésienne empirique permet de faire des conclusions sur le paramètre non observé  $\Theta$  dans le modèle bayésien  $(X, \Theta)^T$  même si sa loi à priori  $\pi(\theta)$  est inconnue. Soit  $\pi(x|\theta)$  la densité de la répartition conditionnelle de  $X$  sachant  $\Theta$ . On suppose que  $\pi(x|\theta)$  est connu. L'estimateur bayésien pour  $\Theta$  est

$$\hat{\Theta} = \mathbf{E}\{\Theta|X\} = \int_{\Omega} \theta q(\theta|X) d\theta = \frac{1}{q(X)} \int_{\Omega} \theta \pi(x|\theta) \pi(\theta) d\theta, \quad (7.1)$$

où  $q(\theta|x)$  est la densité à posteriori de  $\Theta$  sachant  $X$ , et

$$q(x) = \int_{\Omega} \pi(x|\theta) \pi(\theta) d\theta \quad (7.2)$$

est la densité de la loi marginale de  $X$ . Si la densité à priori  $\pi(\theta)$  est inconnue, il est impossible de calculer les valeurs de  $\hat{\theta}$  et  $q(x)$ . Mais si la taille  $n$  de  $X = (X_1, X_2, \dots, X_n)^T$  est suffisamment grande, il est possible de construire un estimateur consistant  $\hat{q}(X)$  de  $q(x)$ . S.N. Bernstein (1941) a proposé d'estimer  $\Theta$  en remplaçant  $q(x)$  par  $\hat{q}(X)$  dans (2), et cherchant la solution  $\hat{\pi}(\theta)$  de cette équation intégrale. Après on peut estimer  $\Theta$ , en utilisant  $\hat{\pi}(\theta)$  et  $\hat{q}(X)$  au lieu de  $\pi(\theta)$  et  $q(x)$  dans (1). Cependant la méthode de Bernstein est difficile, puisque trouver la solution d'équation (2) est le problème difficile de la théorie des équations intégrales. Nous allons donner un exemple (Nikulin, 1978), où est démontré que la répartition à posteriori de la variable aléatoire  $X_n, X_n \sim B(n, \Theta)$  sachant  $\Theta$ , peut être approximée par la loi beta, si le paramètre  $n$  de la loi binomiale tend vers l'infini et la densité à priori  $\Theta$  est continue. Ici nous allons suivre l'article de Nikulin (1992).

## 7.4 Exemple

### 7.4.1 La loi beta et ses propriétés

Soit  $\beta$  la variable aléatoire suivant la loi beta de paramètres  $a$  et  $b$ . La densité de  $\beta$  est

$$p(y|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}, \quad 0 < y < 1, a > 0, b > 0, \quad (7.1)$$

la fonction de répartition de  $\beta$  est

$$\mathbf{P}\{\beta \leq y\} = I_y(a, b), \quad (7.2)$$

où  $I_y(a, b)$  vérifie l'identité

$$I_y(a, b) + I_{1-y}(b, a) \equiv 1, 0 \leq y \leq 1, 0, b > 0, \quad (7.3)$$

On suppose que

$$I_y(a, 0) \equiv 1 - I_{1-y}(0, a) \equiv 0, 0 \leq y \leq 1, a > 0. \quad (7.4)$$

Sous cette hypothèse pour tout  $x = 0, 1, 2, \dots, n$  ( $n$  est un entier positif) on a une identité

$$\sum_{k=0}^x \binom{n}{k} \theta^k (1 - \theta)^{n-k} \equiv I_{1-\theta}(n-x, x+1) \equiv 1 - I_\theta(x+1, n-x). \quad (7.5)$$

par rapport à  $\theta, \theta \in [0, 1]$ . Notons que

$$\mathbf{E}\beta = \frac{a}{a+b} \quad \text{et} \quad \mathbf{Var}\beta = \frac{ab}{(a+b)^2(a+b+1)} \leq \frac{1}{4(a+b+1)} \quad (7.6)$$

et donc de l'inégalité de Chebyshev's on obtient que

$$\mathbf{P}\left\{ \left| \beta - \frac{a}{a+b} \right| \geq \varepsilon \right\} \leq (a+b+1)^{-1/2}, \varepsilon = 1/2(a+b+1)^{-1/4}. \quad (7.7)$$

## 7.5 Résultats principaux.

Soit  $\{(X_n, \Theta)\}$  une suite des vecteurs aléatoires où  $\Theta$  est la variable aléatoire,  $\Theta \in [0, 1]$ , dont la densité  $p(\theta)$  est continue sur  $[0, 1]$ . On suppose que la loi conditionnelle de  $X_n$  sachant  $\Theta = \theta$  est binomiale  $B(n, \theta)$  :

$$\mathbf{P}\{X_n = x | \Theta = \theta\} = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, x = 0, 1, \dots, n. \quad (7.8)$$

Soit  $u(\theta)$  une fonction bornée sur  $[0, 1]$ ,  $|u(\theta)| \leq U$ , où  $U$  est une constante. On considère une fonction  $\mathcal{E}_n(u|x, p)$  qui représente l'espérance conditionnelle de la statistique  $u(\Theta)$  sachant  $X_n = x$ . D'après la formule de Bayes cette fonctionnelle peut être représentée comme le rapport

$$\mathcal{E}_n(u|x, p) = \mathbf{E}\{u(\Theta) | X_n = x\} = \frac{J_n(x; u, p)}{J_n(x; 1, p)}, \quad (7.9)$$

où, comme il s'ensuit de (1) et (8),

$$J_n(x; u, p) = \int_0^1 u(\theta) p(\theta | x+1, n-x+1) p(\theta) d\theta. \quad (7.10)$$

**Lemme.** Soit  $u(\theta)$  et  $v(\theta)$  deux fonctions continues sur  $[0, 1]$ . Alors lorsque  $n \rightarrow \infty$

$$R_n(x; u, p) = J_n(x; u, v) - v\left(\frac{x+1}{n+2}\right) \int_0^1 u(\theta) p(\theta | x+1, n-x+1) d\theta \rightarrow 0 \quad (7.11)$$

uniformément par rapport à  $x = 0, 1, 2, \dots, n$ .

*Démonstration.* On considère un ensemble

$$A_\varepsilon = \left\{ \theta : \left| \theta - \frac{x+1}{n+2} \right| < \varepsilon, 0 \leq \theta \leq 1 \right\},$$

où d'après (7)

$$2\varepsilon = (n+3)^{-1/4}.$$

Dans ce cas en utilisant (10) on obtient

$$R_n = R_n(x; u, v) = \int_0^1 \left[ v(\theta) - v\left(\frac{x+1}{n+2}\right) \right] u(\theta) p(\theta|x+1, n-x+1) d\theta = \int_{A_\varepsilon} + \int_{\bar{A}_\varepsilon}.$$

D'où d'après la définition de la variable aléatoire  $\beta$  (sous conditions  $a = x+1$  et  $b = n-x+1$ ) on a

$$\frac{|R_n|}{U} \leq \sup_{\theta \in A_\varepsilon} \left| v(\theta) - v\left(\frac{x+1}{n+2}\right) \right| + 2 \left[ \max_{0 \leq \theta \leq 1} v(\theta) \right] \mathbf{P} \left\{ \left| \beta - \frac{x+1}{n+2} \right| \geq \varepsilon \right\},$$

où  $U = \max_{0 \leq \theta \leq 1} u(\theta)$ . Le premier terme tend vers zéro uniformément par rapport à  $x$ , puisque  $v$  est une fonction continue sur  $[0, 1]$ . Le second terme tend vers zéro uniformément en  $x$  d'après (7), d'où (12) est démontré.  $\square$

Considérons un ensemble

$$V = \{v = v(\theta) : v \in C[0, 1], v(\theta) \geq 0\}$$

de toutes fonctions non négatives continues sur  $[0, 1]$ , telles que pour tout  $v \in V$  on a

$$\{\theta : v(\theta) = 0\} \subseteq \{\theta : p(\theta) = 0\},$$

d'où il s'ensuit que la densité à priori  $p(\theta)$  appartient à  $V$ .

**Corollaire 1.** Si  $v \in V$ , alors avec la probabilité 1

$$\mathcal{E}_n(u|X_n, v) - \int_0^1 u(\theta) p(\theta|X_n+1, n-X_n+1) d\theta \rightarrow 0 \quad (7.12)$$

lorsque  $n \implies \infty$ .

*Démonstration.* D'après (9), (10) et le lemme sachant  $X_n = x$  on a

$$\mathcal{E}(u|x, v) = \frac{v\left(\frac{x+1}{n+2}\right) \int_0^1 u(\theta) p(\theta|x+1, n-x+1) d\theta + R_n(x; u, v)}{v\left(\frac{x+1}{n+2}\right) + R_n(x; 1, v)},$$

où le reste  $R_n$  dans le numérateur et dans le dénominateur tend vers zéro uniformément par rapport à  $x = 0, 1, 2, \dots, n$ , lorsque  $n \longrightarrow \infty$  (bien sûr, la vitesse de convergence dépend du choix de la fonction  $v$ ). Pour la statistique  $(X_n + 1)/(n + 2)$  la loi de grands nombres est vérifiée et puisque la fonction  $v(\theta)$  est continue on a

$$\mathbf{P} \left\{ v\left(\frac{X_n+1}{n+2}\right) \longrightarrow v(\theta), \quad n \longrightarrow \infty | \Theta = \theta \right\} = 1. \quad (7.13)$$

De plus, puisque  $v \in V$  on a

$$\mathbf{P}\{v(\Theta) > 0\} = 1 - \mathbf{P}\{v(\Theta) = 0\} \geq 1 - \mathbf{P}\{p(\Theta) = 0\} = 1. \quad (7.14)$$

D'où, sachant  $\Theta$  la probabilité conditionnelle de la relation limite (12) égale à 1, et par conséquent la probabilité non conditionnelle est aussi égale à 1.

**Corollaire 2.** Si  $v \in V$ , alors avec la probabilité 1 on a lorsque  $n \rightarrow \infty$

$$\mathbf{P}\{\Theta \leq \theta | X_n = x\} - \sum_{k=x+1}^{n+1} \binom{n+1}{k} \theta^k (1-\theta)^{n-k+1} \rightarrow 0, \quad (7.15)$$

ou, qui est équivalent,

$$\mathbf{P}\{\Theta \leq \theta | X_n = x\} - \mathbf{P}\{X_{n+1} \geq x+1 | \Theta = \theta\} \rightarrow 0, \quad (7.16)$$

et

$$\mathbf{E}\{\Theta^k | X_n = x\} \cdot \frac{x!(n+k+1)!}{(x+k)!(n+1)!} \rightarrow 1 \quad (7.17)$$

uniformément par rapport à  $\theta, 0 \leq \theta \leq 1$ ; ( $k$  est un entier positif).

Pour démontrer (15) on peut remarquer que (15) suit immédiatement de (12), si on pose

$$u(t) \equiv \begin{cases} 1, & t \leq \theta, \\ 0, & t > \theta. \end{cases}$$

La formule (17) peut être obtenue de (12) avec  $u(t) = t^k$ .  $\square$

Particulièrement de (17) il s'ensuit que pour toutes les grandes valeurs de  $n$  l'estimateur bayésien

$$\hat{\Theta} = \mathbf{E}\{\Theta | X_n = x\},$$

qui est le meilleur estimateur ponctuel (dans le sens de minimum du risque quadratique) pour la valeur inconnue du paramètre  $\Theta$ , vérifie les relations

$$\hat{\Theta} = \mathbf{E}\{\Theta | X_n = x\} \approx \frac{x+1}{n+2} \quad \text{and} \quad \hat{\Theta}^2 = \mathbf{E}\{\Theta^2 | X_n = x\} \approx \frac{(x+2)(x+1)}{(n+3)(n+2)},$$

d'où

$$(\hat{\Theta}^2) - (\hat{\Theta})^2 \approx \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)} = \frac{1}{(n+3)} \left(1 - \frac{x+1}{n+2}\right) \frac{x+1}{n+2},$$

i.e., pour les grandes valeurs de  $n$  on a

$$\mathbf{E}\{\hat{\Theta} | X_n = x\} \approx \frac{x+1}{n+2} \approx \frac{x}{n},$$

$$\mathbf{Var}\{\hat{\Theta} | X_n = x\} \approx \frac{x+1}{(n+2)(n+3)} \left(1 - \frac{x+1}{n+2}\right) \approx \frac{x}{n} \left(\frac{x}{n}\right).$$

## 7.6 Aproximations

Le Corollaire 2 donne la possibilité de construire une approximation normale et de Poisson pour la loi à posteriori de  $\Theta$ .

**Approximation normale.** Si  $0 < \theta_0 \leq \theta \leq \theta_1 < 1$  et  $v \in V$ , alors avec la probabilité 1 on a lorsque  $n \rightarrow \infty$

$$\mathbf{P}\{\Theta > \theta | X_n = x\} - \Phi \left[ \frac{x - (n+1)\theta + 0.5}{\sqrt{(n+1)\theta(1-\theta)}} \right] \rightarrow 0. \quad (7.18)$$

Ce résultat ne diffère que par des détails non significatifs du théorème de S. Bernstein (1946), connu comme le théorème "inverse de Laplace".  $\square$

**Approximation de Poisson.** Si  $x \leq x_0$  ( $x_0$  est une constante positive) et  $v \in V$ , alors avec la probabilité 1 on a, lorsque  $n \rightarrow \infty$ ,

$$\mathbf{P}\{\Theta > \theta | X_n = x\} - \sum_{k=1}^x \frac{[\lambda(x, n, \theta)]^k}{k!} e^{-\lambda(x, n, \theta)} \rightarrow 0 \quad (7.19)$$

uniformément par rapport à  $\theta, \theta \in [0, 1]$ , où

$$\lambda(x, n, \theta) = (2n - x + 2)\theta / (2 - \theta). \square$$

**Remarque 1.** Supposons que la densité à priori  $p(\theta)$  est positive sur  $[0, 1]$ . Dans ce cas, dans les Corollaires (18) et (19) on peut omettre des mots "avec la probabilité 1" et après les relations (12)-(19) ajouter "uniformément par rapport à  $X_n = x = 0, 1, 2, \dots, n$ ",

**Remarque 2.** Les relations (15) - (17) dans certains sens approuvent le choix de M. De Groot de la famille des lois beta comme la famille *conjuguée des répartitions à priori* pour des échantillons de la loi Bernoulli.  $\square$

**Remarque 3.** (Nikulin (1978)). Considérons  $V = \{v = v(\theta) : v \in L^r[0, 1]\}$  tel que si  $x_0$  est un point de Lebesgue de  $v \in V$ , alors

$$\left| \frac{1}{2h} \int_{x_0-h}^{x_0+h} |v(x) - v(x_0)|^r dx \right|^{1/r} = o \left( \frac{1}{[\ln \ln \frac{1}{h}]^{1/2r}} \right).$$

Dans ce cas avec la probabilité 1 les relations (15) - (17) sont vérifiées pour toute densité à priori de  $V$ , lorsque  $n \rightarrow \infty$ .  $\square$ .

Plus de détails à ce problème on peut trouver dans Nikulin (1992), (1978), Voinov and Nikulin (1996), et C. Robert (1992).

# Chapitre 8

## EXERCICES.

1. Soit  $\mathbf{X} = (X_1, \dots, X_n)^T$  un vecteur aléatoire, dont la densité est  $\mathbf{f}(\mathbf{x}) = \mathbf{f}(x_1, x_2, \dots, x_n)$ ,  $\mathbf{x} \in \mathbf{R}^n$ .  
Notons

$$\mathbf{A} = \{\mathbf{x} \in \mathbf{R}^n : x_1 \leq x_2 \leq \dots \leq x_n\}.$$

Montrer que la densité

$$\mathbf{f}_{X_{(1)}, \dots, X_{(n)}}^*(x_1, x_2, \dots, x_n) = \mathbf{f}^*(x_1, x_2, \dots, x_n)$$

du vecteur des statistiques d'ordre

$$\mathbf{X}^{(n)} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})^T$$

est donnée par la formule

$$\mathbf{f}^*(x_1, x_2, \dots, x_n) = \begin{cases} \sum_{(r_1, \dots, r_n) \in \sigma_n} \mathbf{f}(x_{r_1}, x_{r_2}, \dots, x_{r_n}), & \text{si } \mathbf{x} \in \mathbf{A}, \\ 0, & \text{sinon,} \end{cases}$$

où  $\sigma_n$  est l'ensemble de toutes les permutations de  $(1, 2, \dots, n)$ .

2. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon tel que

$$\mathbf{P}\{X_i \leq x\} = F(x) \text{ et } f(x) = F'(x)$$

est la densité de  $X_i$ ,  $i = 1, \dots, n$ . Montrer que dans ce cas la densité de  $r$  premières statistiques d'ordre  $\mathbf{X}_n^{(r)} = (X_{(1)}, X_{(2)}, \dots, X_{(r)})^T$ , ( $1 \leq r \leq n$ ) est donnée par la formule suivante

$$\mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, x_2, \dots, x_r) = \frac{n!}{(n-r)!} [1 - F(x)]^{n-r} f(x_1) \cdot \dots \cdot f(x_r)$$

pour tout  $\mathbf{x} \in \mathbf{A}$ . Il est évident que si  $r = n$ , dans ce cas  $\mathbf{X}_n^{(n)} = \mathbf{X}^{(n)}$ .

3. Trouver la densité de

$$\mathbf{X}_n^{(r)} = (X_{(1)}, X_{(2)}, \dots, X_{(r)})^T, 1 \leq r \leq n,$$

quand

$$f(x; \theta) = \theta e^{-\theta x}, x \geq 0, \theta > 0.$$

4. (suite) On suppose que la durée de la vie de certains produits suit une loi exponentielle de paramètre  $\theta, \theta > 0$ . On considère un échantillon  $\mathbb{X} = (X_1, \dots, X_n)^T$  de cette distribution et on arrête l'expérience dès qu'on a obtenu la  $r$ -me ( $1 \leq r \leq n$ ) défaillance. Le résultat de l'expérience est donc une réalisation du vecteur

$$\mathbf{X}_n^{(r)} = (X_{(1)}, X_{(2)}, \dots, X_{(r)})^T.$$

- Trouver la statistique exhaustive minimale associée à ce problème et sa fonction de répartition.
- Estimer le paramètre  $\theta$  par la méthode du maximum de vraisemblance,
- Trouver le biais de cet estimateur. Construire le meilleur estimateur sans biais pour  $\mathbf{E}_\theta X_i$  sachant que  $\mathbf{E}_\theta X_i = 1/\theta$ .
- Trouver l'estimateur de maximum de vraisemblance et le meilleur estimateur sans biais  $S^*(t)$  pour la fonction de survie

$$S(t; \theta) = 1 - F(t; \theta) = \exp\{-\theta t\}, t > 0.$$

5. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon. Notons

$$W_n = X_{(n)} - X_{(1)}.$$

Cette statistique est appelée *l'étendue de l'échantillon*. On suppose que  $X_i$  suive une loi continue, dont la densité  $f$  ne dépend que des paramètres de *translation*  $\mu$  et d'*échelle*  $\sigma$ ,

$$X_i \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \quad |\mu| < \infty, \sigma > 0.$$

- Montrer qu'il existe une constante  $c_n$  telle que

$$\mathbf{E}W_n = c_n \sigma.$$

- Construire un estimateur sans biais pour  $\sigma$ .
- Trouver  $c_n$  quand  $X_i$  est uniforme sur  $[\mu, \mu + \sigma]$ .

6. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon et  $f(x)$  la densité de  $X_i$ . On désigne  $R_i$  le numéro de  $X_i$  dans la suite des statistiques d'ordre

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

On dit que  $R_i$  est le *rang* de  $X_i$ .

Montrer que la distribution conditionnelle de la statistique des rangs  $\mathbf{R} = (R_1, \dots, R_n)^T$ , à condition que

$$\mathbf{X}^{(n)} = \mathbf{x}, \quad \mathbf{x} = (x_1, \dots, x_n)^T \in \mathbf{A},$$

est donnée par la formule :

$$\mathbf{P}\{R_1 = r_1, \dots, R_n = r_n \mid X_{(1)} = x_1, \dots, X_{(n)} = x_n\} = \frac{\mathbf{f}(x_{r_1}, \dots, x_{r_n})}{\mathbf{f}^*(x_1, \dots, x_n)}$$

pour tout  $\mathbf{r} = (r_1, \dots, r_n)^T \in \sigma_n$ .

7. Soient  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon et  $f(x)$  la densité de  $X_i$ .  
Montrer que dans ce cas les statistiques  $\mathbf{R}$  et  $\mathbf{X}^{(n)}$  sont indépendantes et que

$$\mathbf{P}\{\mathbf{R} = \mathbf{r}\} = \mathbf{P}\{R_1 = r_1, \dots, R_n = r_n\} = \frac{1}{n!}, \quad \mathbf{r} = (r_1, \dots, r_n) \in \sigma_n,$$

$$\mathbf{P}\{R_{i_1} = r_{i_1}, \dots, R_{i_m} = r_{i_m}\} = \frac{(n-m)!}{n!}, \quad (i_1, \dots, i_m) \subseteq \{1, 2, \dots, n\},$$

$$\mathbf{E}R_i = \frac{n+1}{2}, \quad \mathbf{Var}R_i = \frac{n^2-1}{12}, \quad \mathbf{Cov}(R_i, R_j) = -\frac{n+1}{12}.$$

8. Soient  $\mathbb{X} = (X_1, \dots, X_n)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_n)^T$  deux échantillons peut être dépendants. On range  $(X_i, Y_i)$  de façon que les  $X_i$  forment une suite nondécroissante. On remplace les  $X_i$  et les  $Y_i$  par leur rangs. On a les statistiques de rangs :

$$\mathbb{R}^{(1)} = (R_{11}, R_{12}, \dots, R_{1n})^T \quad \text{et} \quad \mathbb{R}^{(2)} = (R_{21}, R_{22}, \dots, R_{2n})^T.$$

Le coefficient de corrélation linéaire empirique entre les vecteurs  $\mathbb{R}^{(1)}$  et  $\mathbb{R}^{(2)}$  :

$$r_s = \frac{\sum_{i=1}^n (R_{1i} - \bar{R}_1)(R_{2i} - \bar{R}_2)}{\sqrt{\sum_{i=1}^n (R_{1i} - \bar{R}_1)^2 \sum_{i=1}^n (R_{2i} - \bar{R}_2)^2}}$$

est appelé *le coefficient de corrélation de Spearman*.

Montrer que

a)  $r_s = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n (R_{1i} - R_{2i})^2$  ;

b)  $r_s = 1$ , si  $R_{1i} = R_{2i}$  et  $r_s = -1$ , si  $R_{2i} = n+1 - R_{1i}$  ;

c)  $\mathbf{E}r_s = 0$ ,  $\mathbf{Var}r_s = \frac{1}{n-1}$ , si  $X_i$  et  $Y_i$  sont indépendantes.

9. Soient  $\mathbb{X} = (X_1, \dots, X_n)^T$  et  $\mathbb{Y} = (Y_1, \dots, Y_m)^T$  deux échantillons indépendants,

$$\mathbf{P}\{X_i \leq x\} = F(x), \quad \mathbf{P}\{Y_j \leq y\} = G(y).$$

Notons  $\mathbf{X}^{(n)}$  et  $\mathbf{Y}^{(m)}$  les statistiques d'ordre correspondant à ces deux échantillons. Notons  $\mathbf{Z}^{(N)}$  le vecteur des statistiques d'ordre,  $N = n + m$ , correspondant à la statistique

$$\mathbf{Z} = (X_1, \dots, X_n, Y_1, \dots, Y_m)^T.$$

Soit  $R_i$  le rang de  $X_i$  dans  $\mathbf{Z}^{(N)}$ . On dit que

$$W = \sum_{i=1}^n R_i$$

est la *statistique de Wilcoxon*. Montrer que sous l'hypothèse  $H_0 : F(x) = G(x)$

$$\mathbf{E}W = \frac{n(N+1)}{2} \quad \text{et} \quad \mathbf{Var}W = \frac{nm(N+1)}{12}.$$

10. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \quad |\mu| < \infty, \sigma > 0,$$

où

$$f(x) = \exp(-x)\mathbf{1}_{]0, \infty[}(x).$$

- a) Estimer les paramètres  $\mu$  et  $\sigma$  en utilisant la méthode des moments ;  
 b) estimer les paramètres  $\mu$  et  $\sigma$  en utilisant la méthode de maximum de vraisemblance.

**11.** Supposons que, pour trouver une constante  $\mu$ , on ait fait  $n$  mesures indépendantes. Supposons de plus que les résultats de l'expérience sont libres d'erreur systématique et que les erreurs de mesure suivent une loi normale  $N(0, \sigma^2)$ . Pour estimer la variance  $\sigma^2$  de l'erreur de mesure on a proposé deux formules :

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \hat{\sigma}_2^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2.$$

Peut-on dire que  $\hat{\sigma}_1^2$  et  $\hat{\sigma}_2^2$  sont des valeurs de deux estimateurs sans biais pour la variance ? Quel est le meilleur de ces deux estimateurs ?

**12.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$H_0 : X_i \sim f_r(x, \theta) = \frac{1}{\theta^r \Gamma(r)} x^{r-1} e^{-x/\theta} \mathbf{1}_{]0, +\infty[}(x),$$

i.e.  $X_i$  suit une loi gamma avec deux paramètres  $r$  et  $\theta$ , qui sont inconnus,  $r \in \mathbb{N}$  et  $\theta \in \Theta = ]0, \infty[$ .

- a) Trouver par la méthode des moments les estimateurs  $r_n^*$  et  $\theta_n^*$  pour  $r$  et  $\theta$ .  
 b) Peut-on dire que les suites  $\{r_n^*\}$  et  $\{\theta_n^*\}$  sont consistantes ?  
 c) Supposons que  $n = 10$  et on a reçu :

$$X_1 = 0.117, X_2 = 0.438, X_3 = 0.054, X_4 = 0.732, X_5 = 0.601, \\ X_6 = 0.443, X_7 = 0.016, X_8 = 0.129, X_9 = 0.871, X_{10} = 0.104.$$

Calculer les réalisations des statistiques  $r_{10}^*$  et  $\theta_{10}^*$ .

**13.** Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon de taille  $n$ ,

$$H_0 : X_i \sim f(x; \theta) = \frac{\theta^x}{x!} e^{-\theta}, x \in \mathcal{X} = \{0, 1, \dots\}, \theta \in \Theta = ]0, \infty[,$$

i.e.  $X_i$  suit la loi de Poisson de paramètre  $\theta$ . Notons

$$T = X_1 + \dots + X_n$$

la statistique exhaustive pour  $\theta$ .

- a) Montrer que les statistiques

$$\theta_I = \frac{1}{2n} \chi_{1-\gamma_1}^2(2T) \text{ et } \theta_S = \frac{1}{2n} \chi_{\gamma_2}^2(2T + 2)$$

sont  $\gamma_1$ -limite inférieure de confiance et  $\gamma_2$ -limite supérieure de confiance pour  $\theta$ , où  $\chi_{\alpha}^2(n)$  désigne  $\alpha$ -quantile de la distribution du chi-deux de  $n$  degrés de liberté.

- b) Trouver  $\gamma$ -intervalle de confiance pour :

$$\theta = \mathbf{E}_{\theta} X, \quad b(\theta) = \mathbf{E}_{\theta} X^2, \quad c(\theta) = \frac{\ln(1 + \theta)}{1 + \theta}.$$

c) Le nombre de coups de téléphone venus *au* commutateur pendant une unité de temps est une réalisation d'une variable aléatoire qui suit la loi de Poisson de paramètre  $\theta$ . On a reçu  $X = 3$  coups de téléphone. Construire 0.95-intervalle de confiance pour  $\theta$  et 0.95-limites de confiance pour la probabilité

$$p_0(\theta) = \mathbf{P}_\theta\{X = 0\}.$$

14. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$H_0 : X_i \sim f(x; \theta) = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\} \mathbf{1}_{(x>0)},$$

i.e.  $X_i$  suit la loi exponentielle de paramètre d'échelle  $\theta, \theta > 0$ .

a) Construire  $\gamma$ -limites de confiance pour  $\theta$ .

b) Supposons que  $n = 5$  et que

$$X_1 = 0.71, X_2 = 1.02, X_3 = 0.28, X_4 = 2.49, X_5 = 0.62.$$

Construire 0.9-intervalle de confiance pour  $\theta$ .

c) Soit  $\mathbf{X}_n^{(r)} = (X_{(1)}, \dots, X_{(r)})^T$  un échantillon censuré, lié avec  $\mathbb{X}$  ( $r$  représente le nombre des défaillances observées de certains produits dans un expérience).

Trouver le  $\gamma$ -intervalle de confiance pour  $\theta$  et pour la fonction de survie

$$S(x; \theta) = \mathbf{P}_\theta\{X \geq x\}.$$

d) Soit  $n = 20$ ; le résultat d'expérience est donné par le vecteur

$$\mathbf{X}_{20}^{(8)} = (X_{(1)}, \dots, X_{(r)})^T = (10, 15, 41, 120, 159, 181, 222, 296)^T.$$

Trouver les 0.95-limites inférieures pour  $\theta$  et  $S(400; \theta)$ .

15. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$H_0 : X_i \sim f(x, \theta) = \theta^x (1 - \theta)^{1-x}, x \in \mathcal{X} = \{0, 1\}, \theta \in \Theta = ]0, 1[,$$

i.e.  $X_i$  suit la loi de Bernoulli de paramètre  $\theta$ .

a) Trouver les  $\gamma$ -limites de confiance pour  $\theta$ .

b) Soit  $n = 3$  et  $T_3 = X_1 + X_2 + X_3 = 2$ , i.e. on a eu 2 "succès". Trouver les 0.95-limites de confiance pour  $\theta$  et 0.95-intervalle de confiance pour  $\theta$ .

16. Soit  $X$  une variable aléatoire, dont la fonction de répartition

$$F(x; \theta), \theta \in \Theta = ]0, 1[,$$

est donnée par la formule :

$$F(x; \theta) = 1 - \theta^x, \text{ si } x > 0,$$

$$F(x; \theta) = 0, \text{ sinon.}$$

Supposons que dans l'expérience on a observé  $X = 1$ . Construire un intervalle de confiance de niveau  $P$  pour  $\theta$  dans deux cas :

a)  $X$  est continue ;

b)

$$X \text{ est discrète, } \mathbf{P}\{X \leq x\} = F([x]).$$

17. Soit  $X_1$  et  $X_2$  deux variables aléatoires indépendantes,

$$H_0 : X_i \sim e^{-(x-\theta)} \mathbf{1}_{[\theta, \infty[}(x), \theta \in \Theta = \mathbf{R}^1.$$

Trouver le plus petit  $\gamma$ -intervalle de confiance pour  $\theta$ .

18. Soit  $X_1$  et  $X_2$  deux variables aléatoires indépendantes,  $X_i$  suit la loi uniforme sur  $]\theta - 1, \theta + 1[$ .

Trouver le plus court 0.81-intervalle de confiance pour  $\theta$ .

19. Soit 20.76 et 20.98 deux mesures indépendantes d'un angle, qui ont la même précision, et soient

$$21.64, 21.54, 22.32, 20.56, 21.43 \text{ et } 21.07$$

6 autres mesures indépendantes du même angle, faites avec une précision 4 fois plus petite. On suppose que les erreurs aléatoires des mesures suivent une loi normale. Trouver les 0.95-limites de confiance pour la différence des erreurs systématiques des deux instruments utilisés pour obtenir les mesures données.

20. Quelles sont les valeurs de la moyenne et de la variance de la loi empirique construite d'après les valeurs successivement observées suivantes :

$$3.92, 4.04, 4.12, 4.35, 4.55?$$

Peut on, avec le niveau de signification  $\alpha = 0.05$ , retenir l'hypothèse  $H_0$  selon laquelle ces nombres sont les réalisations des variables aléatoires  $X_1, \dots, X_n$ , qui forment un échantillon ?

21. (suite de 9.) Montrer que la répartition de la statistique de Wilcoxon ne dépend pas des paramètres inconnus si l'hypothèse  $H_0 : F(x) = G(x)$  est vraie.

Comment définir la région critique pour l'hypothèse  $H_0$  contre l'alternative bilatérale  $H_1 : F(x) \neq G(x)$  et unilatérale  $H_2 : F(x) > G(x)$  ?

22. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,

$$X_i \sim f(x; \theta) = \theta e^{-\theta x} \mathbf{1}_{(]0, +\infty[)}, \quad \theta > 0.$$

Trouver le test uniformément le plus puissant (UPP) pour l'hypothèse simple  $H_0 : \theta = \theta_0$  contre l'alternative composée  $H_1 : \theta < \theta_0$ . Le niveau de signification est  $\alpha$ . Trouver la fonction de puissance et faire son graphe.

23. Quelle est le plus petit nombre des mesures indépendantes suivant la même loi normale avec l'espérance  $\mu$  et variance  $\sigma^2 = 1$  qui vérifie l'hypothèse  $\mu = 0$  contre l'alternative  $\mu = 1$  avec les probabilités d'erreurs de première et seconde espèce inférieures ou égales à 0.01 ?

24. Soit  $\mathbb{X} = (X_1, \dots, X_n)^T$  un échantillon,  $X_i \sim U(0, \theta)$ ,  $\theta > 0$ . Trouver

a) le test UPP pour l'hypothèse  $H_0 : \theta = \theta_0$  contre l'alternative  $H_1 : \theta > \theta_0$  ;

b) le test UPP pour l'hypothèse  $H_0 : \theta = \theta_0$  contre l'alternative  $H_2 : \theta < \theta_0$  ;

c) le test UPP pour l'hypothèse  $H_0 : \theta = \theta_0$  contre l'alternative  $H_3 : \theta \neq \theta_0$ .

25. Dans la suite des épreuves indépendantes de Bernoulli la probabilité de succès est égale à  $p$ . Construire un critère pour vérifier l'hypothèse  $p = 0$  contre l'alternative  $p = 0.01$  et déterminer la valeur minimale de taille d'échantillon, pour laquelle les probabilités d'erreurs de première et de seconde espèces sont inférieures ou égales à 0.01.

26. Cinq variables aléatoires indépendantes  $X_1, X_2, \dots, X_5$  qui suivent la même loi ont pris les valeurs : 47, 46, 49, 53, 50. Vérifier l'hypothèse  $H_0$ , avec le niveau de signification

$\alpha = 0.1$ , que  $X_i$  suit une loi de Poisson. Calculer, sous l'hypothèse  $H_0$ , la loi conditionnelle de  $X_i$  sachant  $\sum_{i=1}^5 X_i$ .

27. Après 8000 épreuves indépendantes les événements  $A, B, C$  se sont réalisés respectivement 2014, 5012 et 974 fois.

Tester l'hypothèse

$$H : \mathbf{P}(A) = 0.5 - 2a, \quad \mathbf{P}(B) = 0.5 + a, \quad \mathbf{P}(C) = a,$$

( $0 < a < 0.25$  ; niveau du test  $\alpha = 0.05$ ).

28. Au cours de la première heure de travail le compteur a enregistré 150 impulsions d'un processus poissonien, pendant les deux heures suivantes - 250 impulsions. Est-ce que l'intensité d'arrivée des impulsions à une unité de temps est la même ? (Prendre le niveau du test égal à 0.05).

29. Au cours du premier jour de travail on a enregistré 20026 impulsions d'un processus de Poisson, tandis que le jour suivant on n'a enregistré que 19580 impulsions. Y a-t-il des raisons d'affirmer que pendant le deuxième jour l'intensité d'arrivée des impulsions a diminué ? (Prendre le seuil  $\alpha = 0.05$ .)

30. Parmi 300 étudiants 97 ont obtenu d'excellentes notes à l'examen de fin d'études et 48 à l'examen d'entrée à l'université. 18 parmi eux ont eu d'excellentes notes aux deux à la fois. Vérifier l'hypothèse de l'indépendance des résultats des examens. Niveau de signification :  $\alpha = 0.1$ .

31. Le premier groupe de 300 étudiants a obtenu les notes suivantes à l'examen :

“excellent” : 144,

“bon” : 80 ;

“médiocre” : 43 ;

“mauvais” : 33.

Les résultats pour le deuxième groupe sont 154,72,35,39. Peut-on affirmer avec le niveau de signification  $\alpha = 0.05$  que les étudiants de ces groupes ont les mêmes connaissances ?

32. Soit  $\{X_t\}_{t \geq 0}$  un processus homogène de Poisson de paramètre  $\lambda$ ,  $\lambda > 0$ . ( $X_0 = 0$ ). Supposons que aux moments  $0 < t_1 < t_2 < \dots < t_n$  on observe les réalisations  $X_{t_1}, \dots, X_{t_n}$ .

Montrer que

$$\hat{\lambda}_n(X_{t_1}, \dots, X_{t_n}) = \frac{1}{t_n} X_{t_n}$$

est l'estimateur sans biais pour  $\lambda$  de variance minimale (MVUE).

33. Soit  $W(t)$ ,  $t \geq 0$ , un processus de Wiener,

$$\mathbf{E}W(t) = at, \quad \mathbf{Var} W(t) = \sigma^2 t$$

$$\mathbf{Cov}(W(s), W(t)) = \sigma^2 \min(s, t), \quad s \geq 0, t \geq 0,$$

$$|a| < \infty, \quad \sigma > 0.$$

Supposons que nous observons  $W(t_1), \dots, W(t_n)$  ( $n$  réalisations de  $W(t)$  dans les points  $0 < t_1 < \dots < t_n$ ). Notons

$$\Delta_k = t_k - t_{k-1}, \quad y_k = \frac{W(t_k) - W(t_{k-1})}{\Delta_k}, \quad t_0 = W(0) = 0.$$

Montrer que en statistique

$$\hat{a}_n = \frac{1}{t_n} \sum_{k=1}^n \Delta_k y_k \quad \text{et} \quad \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n \delta_k (y_k - \hat{a}_n)^2$$

sont les estimateurs sans biais pour  $a$  et  $\sigma^2$  de variances minimales (MVUE's).

**34.** Soit  $W(t)$ ,  $t \geq 0$ , un processus de Wiener

$$\mathbf{E}W(t) = t, \quad \mathbf{Var}W(t) = \sigma^2 t.$$

Supposons que l'on observe  $W(t)$  sur un intervalle  $[0, \varepsilon]$ ,  $\varepsilon > 0$ .

Soit  $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = \varepsilon$ ,  $t_i = \frac{i}{n}$ ,

$$S_n^2 = \frac{1}{\varepsilon} \sum_{i=0}^{n-1} [W(t_{i+1}) - W(t_i)]^2.$$

Montrer que

$$S_n^2 \xrightarrow{\mathbf{P}} \sigma^2, \quad n \rightarrow \infty.$$

# Chapitre 9

## SOLUTIONS.

1. Soit  $\mathbf{F}(\mathbf{x}) = \mathbf{F}(x_1, \dots, x_n)$  la fonction de répartition de  $\mathbf{X}^{(n)}$ ,  $\mathbf{x} \in \mathbf{R}^n$ . Dans ce cas pour tout

$$\mathbf{x} \in \mathbf{A} = \{\mathbf{x} \in \mathbf{R}^n : x_1 \leq x_2 \leq \dots \leq x_n\}$$

on a

$$\begin{aligned} \mathbf{F}(x_1, \dots, x_n) &= \mathbf{P}\{X_{(1)} \leq x_1, \dots, X_{(n)} \leq x_n\} = \\ &= \sum_{(r_1, \dots, r_n) \in \sigma_n} \mathbf{P}\{X_{r_1} \leq x_1, X_{r_2} \leq x_2, \dots, X_{r_n} \leq x_n\} = \\ &= \sum_{(r_1, \dots, r_n) \in \sigma_{n-\infty}} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \mathbf{f}_{X_{r_1}, \dots, X_{r_n}}(u_1, u_2, \dots, u_n) du_1 \dots du_n = \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \sum_{(r_1, \dots, r_n) \in \sigma_n} \mathbf{f}_{X_{r_1}, \dots, X_{r_n}}(u_1, u_2, \dots, u_n) du_1 \dots du_n, \end{aligned}$$

d'où on tire que pour tout  $\mathbf{x} \in \mathbf{A}$  on a

$$\begin{aligned} \mathbf{f}^*(x_1, x_2, \dots, x_n) &= \sum_{(r_1, \dots, r_n) \in \sigma_n} \mathbf{f}_{X_{r_1}, \dots, X_{r_n}}(x_1, x_2, \dots, x_n) = \\ &= \sum_{(r_1, \dots, r_n) \in \sigma_n} \mathbf{f}(x_{r_1}, x_{r_2}, \dots, x_{r_n}). \end{aligned}$$

On remarque que s'il existe au moins deux numéros  $i$  et  $j$  pour lesquels  $x_i > x_j$ , ( $i < j$ ), c'est à dire si  $\mathbf{x} \notin \mathbf{A}$ , dans ce cas

$$\mathbf{F}(x_1, \dots, x_n) = \mathbf{P}\{X_{(1)} \leq x_1, \dots, X_{(n)} \leq x_n\} = 0.$$

2.  $\mathbf{X} = \mathbb{X}$  est un échantillon, i.e.  $X_1, \dots, X_n$  sont indépendantes et suivent la même loi, dont la densité est  $f(x)$ . Dans ce cas, comme il suit du problème 1, pour tout  $\mathbf{x} \in \mathbf{A}$  la densité de

$$\mathbf{X}_n^{(r)} = (X_{(1)}, X_{(2)}, \dots, X_{(r)})^T$$

est donnée par la formule :

$$\mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, \dots, x_r) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{f}_{X_{(1)}, \dots, X_{(n)}}^*(\mathbf{x}) dx_{r+1} \dots dx_n.$$

Mais

$$\mathbf{f}_{X_{(1)}, \dots, X_{(n)}}^*(\mathbf{x}) = 0, \text{ si } \mathbf{x} \notin \mathbf{A},$$

et donc

$$\mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, \dots, x_r) = \int_{x_r}^{\infty} dx_{r+1} \int_{x_{r+1}}^{\infty} dx_{r+2} \dots \int_{x_{n-1}}^{\infty} \mathbf{f}_{X_{(1)}, \dots, X_{(n)}}^*(\mathbf{x}) dx_n.$$

Parce que  $\mathbf{X}$  est un échantillon, on en tire que pour tout  $\mathbf{x} \in \mathbf{A}$  :

$$\mathbf{f}^*(\mathbf{x}) = n! f(x_1) f(x_2) \dots f(x_n),$$

et donc

$$\begin{aligned} \mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, \dots, x_r) &= \\ &= n! f(x_1) f(x_2) \dots f(x_r) \int_{x_r}^{\infty} f(x_{r+1}) dx_{r+1} \dots \int_{x_{n-1}}^{\infty} f(x_n) dx_n. \end{aligned}$$

Notons que

$$\int_{x_{n-1}}^{\infty} f(x_n) dx_n = S(x_{n-1}),$$

où  $S(x) = 1 - F(x)$ , et donc

$$\int_{x_{n-2}}^{\infty} f(x_{n-1}) S(x_{n-1}) dx_{n-1} = - \int_{x_{n-2}}^{\infty} S(x_{n-1}) dS(x_{n-1}) = \frac{1}{2} S^2(x_{n-2}).$$

En procédant de la même façon on en tire que

$$\int_{x_r}^{\infty} f(x_{r+1}) \frac{1}{(n-r-1)!} S^{n-r+1}(x_{r+1}) dx_{r+1} = \frac{1}{(n-r)!} S^{n-r}(x_r),$$

et par conséquent on trouve que

$$\mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, \dots, x_r) = \frac{n!}{(n-r)!} S^{n-r}(x_r) f(x_1) \dots f(x_r).$$

**3.** En cas de la loi exponentielle on a  $X_i \sim f(x; \theta)$ ,  $\theta \in \Theta = ]0, \infty[$ , où pour tout  $\theta \in \Theta$

$$f(x; \theta) = \theta \exp\{-\theta x\}, x \geq 0,$$

et

$S(x; \theta) = 1 - F(x; \theta) = e^{-\theta x}, x \geq 0, F(x; \theta) = \mathbf{P}\{X_i \leq x; \theta\} = \mathbf{P}_\theta\{X_i \leq x\}$ , et donc avec la probabilité 1

$$\mathbf{X}^{(n)} \in \mathbf{A} = \{\mathbf{x} \in \mathbf{R}^n : 0 \leq x_1 \leq x_2 \leq \dots \leq x_n\},$$

d'où on trouve que la densité de  $\mathbf{X}_n^{(r)}$  est donnée par la formule :

$$\mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, \dots, x_r; \theta) = \frac{n!}{(n-r)!} \theta^r \exp\{-\theta t\}, \mathbf{x} \in \mathbf{A},$$

où

$$t = \sum_{i=1}^r x_i + (n-r)x_r.$$

**4. a)** La statistique

$$T = \sum_{i=1}^r X_{(i)} + (n-r)X_{(r)}$$

est exhaustive pour  $\theta$ , parce que la fonction de vraisemblance  $L(\mathbf{X}_n^{(r)}; \theta)$  de la statistique  $\mathbf{X}_n^{(r)}$  peut être présentée comme un produit

$$L(\mathbf{X}_n^{(r)}; \theta) = g(T; \theta)h(\mathbf{X}_n^{(r)}) = \frac{n!}{(n-r)!} \theta^r \exp\left\{-\theta \sum_{i=1}^r X_{(i)} - \theta(n-r)X_{(r)}\right\} \mathbf{1}_{\{X_{(1)} \geq 0\}},$$

et donc selon le critère de factorisation de Neyman-Fisher la statistique  $T$  est exhaustive. On remarque que la fonction de vraisemblance de la statistique  $\mathbf{X}^{(n)}$  est donnée par la formule :

$$L(\mathbf{X}^{(n)}; \theta) = n! f(X_{(1)}; \theta) f(X_{(2)}; \theta) \dots f(X_{(n)}; \theta).$$

Par la tradition on dit que  $T$  est *la survie sommaire* de tout  $n$  produits observés dans l'expérience.

**b)** Pour trouver l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  pour  $\theta$ , il nous faut maximiser  $L(\mathbf{X}_n^{(r)}; \theta)$  par rapport à  $\theta$ , ce qui est équivalent à la maximisation de  $\ln L(\mathbf{X}_n^{(r)}; \theta)$  par rapport à  $\theta$ . Donc pour trouver l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  il nous faut résoudre l'équation de maximum de vraisemblance

$$\frac{d}{d\theta} \ln L(\mathbf{X}_n^{(r)}; \theta) = 0.$$

Parce que

$$\ln L(\mathbf{X}_n^{(r)}; \theta) = \ln \frac{n!}{(n-r)!} + r \ln \theta - \theta T,$$

on trouve que

$$\hat{\theta}_n = \frac{r}{T}.$$

On remarque que  $\hat{\theta}_n$  ne dépend que de la statistique exhaustive  $T$ . Sachant  $\hat{\theta}_n$  nous pouvons construire tout de suite l'estimateur de maximum de vraisemblance  $\hat{S}(x)$  pour  $S(x; \theta)$  pour tout  $x$  fixé :

$$\hat{S}(x) = \exp \left\{ -\frac{rx}{T} \right\}.$$

Tout d'abord on remarque que dans ce problème on n'observe que la statistique

$$\mathbf{X}_n^{(r)} = (X_{(1)}, \dots, X_{(r)})^T, (1 \leq r \leq n)$$

et pas  $\mathbb{X}$  ou  $\mathbf{X}^{(n)}$ , et pour cette raison on dit que on a un *échantillon censuré*.

c). Pour apprendre des propriétés des estimateurs, basées sur la statistique exhaustive  $T$ , il nous faut savoir la distribution de  $T$ . On remarque que la statistique  $T$  peut être présentée dans la forme suivante :

$$T = nX_{(1)} + (n-1)(X_{(2)} - X_{(1)}) + \dots + (n-r-1)(X_{(r)} - X_{(r-1)}),$$

parce que  $n$  produits ont fonctionnés jusqu'à la première défaillance,  $(n-1)$  restants entre la première et la seconde défaillances, etc.

Soit

$$\mathbf{Z} = (Z_1, \dots, Z_r)^T = \mathbf{U}\mathbf{X}_n^{(r)},$$

une statistique dont les coordonnées  $Z_i$  sont déterminées par la transformation linéaire  $\mathbf{z} = \mathbf{U}\mathbf{x}$ ,  $\mathbf{x} \in \mathbf{A}$ , avec la matrice  $\mathbf{U}$ , dont les éléments  $u_{ij}$  sont

$$u_{ii} = n - i + 1, i = 1, \dots, r,$$

$$u_{ij} = -(n - j + 1), j = i - 1; i = 2, \dots, r,$$

$$u_{ij} = 0 \text{ dans tout les autres cas,}$$

et donc

$$z_i = (n - i + 1)(x_i - x_{i-1}), i = 1, \dots, r; x_0 = 0.$$

Dans ce cas

$$Z_i = (n - i + 1)(X_{(i)} - X_{(i-1)}), i = 1, \dots, r, X_{(0)} = 0,$$

d'où on tire que dans les terms de  $Z_i$  la statistique  $T$  est donnée par la formule suivante

$$T = Z_1 + \dots + Z_r.$$

Tout d'abord nous allons montrer que les statistiques  $Z_1, \dots, Z_r$  sont indépendantes et suivent la même loi. Pour prouver cela il nous faut trouver la densité  $\mathbf{f}_{\mathbf{Z}}(\mathbf{z}; \theta)$  de la statistique  $\mathbf{Z} = \mathbf{U}\mathbf{X}_n^{(r)}$ . Sachant que la densité de  $\mathbf{X}_n^{(r)}$  est

$$\mathbf{f}_{X_{(1)}, \dots, X_{(r)}}^*(x_1, \dots, x_r; \theta) = \frac{n!}{(n-r)!} \theta^r \exp \left\{ -\theta \left[ \sum_{i=1}^r x_i + (n-r)x_r \right] \right\},$$

pour trouver la densité  $\mathbf{f}_{\mathbf{Z}}(\mathbf{z}; \theta)$ , il nous faut calculer le Jacobian de la transformation  $\mathbf{U}^{-1}$ . Parce que

$$\det \mathbf{U} = n! / (n-r)!,$$

on trouve que

$$\mathbf{f}_{\mathbf{Z}}(\mathbf{z}; \theta) = \theta^r \exp\{-\theta(z_1 + \dots + z_r)\} = f(z_1; \theta) \dots f(z_r; \theta),$$

d'où on voit bien que les statistiques  $Z_1, \dots, Z_r$  sont indépendantes et suivent la loi exponentielle de paramètre  $\theta$ . Mais dans ce cas la statistique  $T$  suit la loi gamma avec  $r$  degrés de liberté, dont la densité est donnée par la formule :

$$f_T(t; \theta) = \frac{\theta^r}{\Gamma(r)} t^{r-1} e^{-\theta t}, \quad t > 0.$$

En utilisant ce résultat on trouve que

$$\mathbf{E}\hat{\theta}_n = \mathbf{E}\theta \frac{r}{T} = \int_0^{\infty} \frac{r}{t} f_T(t; \theta) dt = \frac{\theta r}{r-1},$$

d'où on trouve que le meilleur estimateur sans biais pour  $\theta$  est

$$\theta^* = \frac{r-1}{T}.$$

**d).** Pour trouver le meilleur estimateur sans biais  $S^*(x)$  pour  $S(x; \theta)$  nous pouvons appliquer l'approche de Rao-Blackwell-Kolmogorov, d'après laquelle tout d'abord il nous faut trouver n'importe quel estimateur sans biais, et après calculer son espérance conditionnelle par rapport à la statistique exhaustive  $T$ , qui est complète, parce que la famille  $\{f_T(t; \theta), \theta \in \Theta\}$  est complète. En qualité de l'estimateur primaire pour  $S(x; \theta)$  il est raisonnable de choisir la statistique

$$\tilde{S}(x) = \mathbf{1}_{\{Z_1 > x\}},$$

parce que

$$\mathbf{E}_\theta \tilde{S}(x) = \mathbf{P}\{Z_1 > x; \theta\} = e^{-\theta x} = S(x; \theta),$$

et donc le meilleur estimateur sans biais pour  $S(x; \theta)$  est

$$S^*(x) = \mathbf{E}_\theta\{\tilde{S}(x) \mid T\}.$$

On remarque que cette espérance conditionnelle *ne* dépend pas de  $\theta$ , parce que la statistique  $T$  est exhaustive. Pour trouver la densité conditionnelle de  $Z_1$  par rapport à  $T$ , il nous faut savoir la densité de la statistique  $(Z_1, T)^T$ . On remarque que la densité de

$$(Z_1, T - Z_1)^T = (Z_1, Z_2 + Z_3 + \dots + Z_r)^T,$$

est donnée par la formule

$$\mathbf{f}_{Z_1, T-Z_1}(z, v; \theta) = \theta e^{-\theta z} \frac{\theta^{r-1} v^{r-2}}{\Gamma(r-1)} e^{-\theta v}, \quad z \geq 0; v \geq 0,$$

sinon  $\mathbf{f}_{Z_1, T-Z_1}(z, v; \theta) = 0$ , d'où on trouve, par le changement de variables

$$z = z \text{ et } t = z + v,$$

la densité  $\mathbf{f}_{Z_1, T}(z, t; \theta)$  de la statistique  $(Z_1, T)^T$  :

$$\mathbf{f}_{Z_1, T}(z, t; \theta) = \theta e^{-\theta z} \frac{\theta^{r-1}}{\Gamma(r-1)} (t-z)^{r-2} e^{-\theta(t-z)}, t \geq z \geq 0,$$

parce que le Jacobien de la transformation est égal à 1. En utilisant ce résultat on trouve immédiatement la densité conditionnelle

$$f_{Z_1|T=t}(z) = \frac{\mathbf{f}_{Z_1, T}(z, t)}{f_T(t)} = \frac{r-1}{t^{r-1}} (t-z)^{r-2}, t \geq z \geq 0,$$

sinon  $f_{Z_1|T=t}(z) = 0$ . Donc si  $T \geq x$  on a

$$\mathbf{E}_\theta\{\tilde{S}(x) | T\} = \int_x^T 1 \cdot \frac{r-1}{T} (1 - \frac{z}{T})^{r-2} dz = (1 - \frac{x}{T})^{r-1},$$

sinon  $\mathbf{E}_\theta\{\tilde{S}(x) | T\} = 0$ . Donc

$$S^*(x) = \mathbf{E}_\theta\{\tilde{S}(x) | T\} = \begin{cases} (1 - \frac{x}{T})^{r-1}, & \text{si } T \geq x \\ 0, & \text{sinon.} \end{cases}$$

En fin on remarque que si  $T \gg x$ , alors

$$\begin{aligned} S^*(x) &= (1 - \frac{x}{T})^{r-1} = \exp\{(r-1)\ln(1 - \frac{x}{T})\} = \\ &= \exp\{-(r-1)[\frac{x}{T} + o(\frac{x}{T})]\} \cong \exp\{-r\frac{x}{T}\} = \hat{S}(x). \end{aligned}$$

6. Pour tout  $\mathbf{x} \in \mathbf{A}$  et  $\mathbf{r} = (r_1, \dots, r_n) \in \sigma_n$  on a :

$$\begin{aligned} \mathbf{P}\{\mathbf{R} = \mathbf{r} | \mathbf{X}^{(n)} = \mathbf{x}\} &= \lim_{h_1, \dots, h_n \downarrow 0} \mathbf{P}\{\mathbf{R} = \mathbf{r} | x_1 < X_{(1)} \leq x_1 + h_1, \dots, x_n < X_{(n)} \leq x_n + h_n\} = \\ &= \lim_{h_1, \dots, h_n \downarrow 0} \frac{\mathbf{P}\{R_1 = r_1, \dots, R_n = r_n, x_1 < X_{(1)} \leq x_1 + h_1, \dots, x_n < X_{(n)} \leq x_n + h_n\}}{\mathbf{P}\{x_1 < X_{(1)} \leq x_1 + h_1, \dots, x_n < X_{(n)} \leq x_n + h_n\}} = \\ &= \lim_{h_1, \dots, h_n \downarrow 0} \frac{\mathbf{P}\{x_{r_1} < X_1 \leq x_{r_1} + h_{r_1}, \dots, x_{r_n} < X_n \leq x_{r_n} + h_{r_n}\} h_1 h_2 \dots h_n}{\mathbf{P}\{x_1 < X_{(1)} \leq x_1 + h_1, \dots, x_n < X_{(n)} \leq x_n + h_n\} h_1 \dots h_n} \\ &= \frac{\mathbf{f}(x_{r_1}, \dots, x_{r_n})}{\mathbf{f}^*(\mathbf{x})}. \end{aligned}$$

13. a) La statistique exhaustive  $T = \sum_{i=1}^n X_i$  suit la loi de Poisson de paramètre  $n\lambda$  ( $\lambda > 0$ ) :  $T \sim \mathcal{P}(n\lambda)$ . La fonction de répartition de  $T$

$$G(k; \lambda) = \sum_{i=1}^k \frac{(n\lambda)^i}{i!} e^{-n\lambda} = \mathbf{P}\{\chi^2(2k+2) \geq 2n\lambda\} = \mathcal{P}(2n\lambda, 2k+2),$$

où

$$\mathcal{P}(x, n) = \mathbf{P}\{\chi^2(n) \geq x\}.$$

On a

$$G(k-0, \lambda) = \sum_{i=1}^{k-1} \frac{(n\lambda)^i}{i!} e^{-n\lambda} = \mathcal{P}(2n\lambda, 2k) \quad (k = 1, 2, \dots),$$

$$G(k-0, \lambda) = 0, \text{ si } k = 0.$$

Les fonctions  $I$  et  $S$  du théorème de Bolshev

$$I(\lambda; \mathbb{X}) = \mathcal{P}(2n\lambda, 2T), \text{ si } \mathbb{X} \neq 0,$$

$$I(\lambda; \mathbb{X}) = 0, \text{ si } \mathbb{X} = 0,$$

$$S(\lambda; \mathbb{X}) = \mathcal{P}(2n\lambda, 2T + 2).$$

La fonction  $S$  est strictement décroissante pour toutes valeurs de  $T$ , la fonction  $I$  est strictement décroissante pour  $T \neq 0$ .

On déduit du théorème de Bolshev que  $\gamma_1$ -limite inférieure de confiance  $\lambda_i$  et  $\gamma_2$ -limite supérieure de confiance  $\lambda_s$  pour  $\lambda$  peuvent être trouvées des équations

$$\mathcal{P}(2n\lambda_i, 2T) = \gamma_1,$$

$$\mathcal{P}(2n\lambda_s, 2T + 2) = 1 - \gamma_2$$

où

$$\lambda_i = \frac{1}{2n} \chi_{1-\gamma_1}^2(2T)$$

$$\lambda_s = \frac{1}{2n} \chi_{\gamma_2}^2(2T + 2). \quad (1)$$

Si  $T = 0$ ,  $I(\lambda; \mathbb{X}) = 0$ . Dans ce cas il n'existe pas  $\lambda$  tel que  $I(\lambda; \mathbb{X}) \geq \gamma_1 > 1/2$ . On déduit du théorème de Bolshev que

$$\lambda_i = \inf_{\lambda > 0} \lambda = 0.$$

**b)** Pour obtenir  $\gamma$ -intervalle de confiance  $]\lambda_i, \lambda_s[$  pour  $\lambda$  il faut prendre  $\gamma_1 + \gamma_2 = 1 + \gamma$  dans les formules (1). Dans le cas  $\gamma_1 = \gamma_2$  on a  $\gamma_1 = \gamma_2 = (1 + \gamma)/2$ .

**c)** Si  $n = 1$ ,  $T = X = 3$ , on a

$$\lambda_i = \frac{1}{2} \chi_{1-\gamma_1}^2(6), \quad \lambda_s = \frac{1}{2} \chi_{\gamma_2}^2(6).$$

Pour obtenir 0.95-intervalle de confiance il faut prendre

$$\gamma_1 = \gamma_2 = (1 + 0.95)/2 = 0.975.$$

On a

$$\lambda_i = \frac{1}{2} \chi_{0.025}^2(6) = \frac{1}{2} 1.237 = 0.6185,$$

$$\lambda_s = \frac{1}{2} \chi_{0.975}^2(8) = \frac{1}{2} 17.535 = 8.7675.$$

Si  $p_0(\lambda) = e^{-\lambda}$ , on a

$$\lambda_i < \lambda \Leftrightarrow e^{-\lambda_i} > e^{-\lambda}, \quad \lambda_s > \lambda \Leftrightarrow e^{-\lambda_s} < e^{-\lambda},$$

donc 0.95-intervalle de confiance pour  $p_0(\lambda)$  est  $]P_{i0}, P_{s0}[$  avec

$$P_{i0} = e^{-\frac{1}{2} \chi_{0.025}^2(6)} = e^{-\frac{1}{2} 1.237} \approx 0.000431.$$

$$P_{s0} = e^{-\frac{1}{2}\chi_{0.05}^2(6)} = e^{-\frac{1}{2}1.635} \approx 0.441.$$

**14. a)** Notons

$$T = X_1 + \dots + X_n.$$

La statistique  $T$  suit une loi gamma  $G(n; \frac{1}{\theta})$  de paramètres  $n$  and  $1/\theta$  :

$$\mathbf{P}\{T \leq t\} = \frac{1}{(n-1)!\theta^n} \int_0^t u^{n-1} e^{-u/\theta} du, t \geq 0,$$

et donc  $T/\theta$  suit la loi gamma  $G(n; 1)$ , et par conséquent

$$\frac{2T}{\theta} = \chi_{2n}^2.$$

Dans cet exemple les fonctions  $I$  et  $S$  peuvent être choisies de façons suivante

$$I(\theta; \mathbb{X}) = S(\theta; \mathbb{X}) = 1 - \mathcal{P}\left(\frac{2T}{\theta}, 2n\right).$$

Ces fonctions sont décroissantes en  $\theta$  et du théorème de Bolshev il suit que les limites inférieure  $\theta_i$  et supérieure  $\theta_s$  peuvent être trouvées des équations

$$1 - \mathcal{P}\left(\frac{2T}{\theta_i}, 2n\right) = \gamma \quad \text{et} \quad 1 - \mathcal{P}\left(\frac{2T}{\theta_s}, 2n\right) = 1 - \gamma,$$

c'est-à-dire

$$\frac{2T}{\theta_i} = \chi_{\gamma}^2(2n) \quad \text{and} \quad \frac{2T}{\theta_s} = \chi_{1-\gamma}^2(2n),$$

d'où on trouve que

$$\theta_i = \frac{2T}{\chi_{\gamma}^2(2n)} \quad \text{et} \quad \theta_s = \frac{2T}{\chi_{1-\gamma}^2(2n)}.$$

**c)** La statistique

$$T_r = \sum_{k=1}^r X_{(k)} + (n-r)X_{(r)}$$

suit une loi gamma  $G(r; \frac{1}{\theta})$ , et par conséquent  $\gamma$ -intervalle de confiance pour  $\theta$  est  $]\theta_i, \theta_s[$ , où

$$\theta_i = \frac{2T_r}{\chi_{\frac{1+\gamma}{2}}^2(2r)} \quad \text{et} \quad \theta_s = \frac{2T_r}{\chi_{\frac{1-\gamma}{2}}^2(2r)}.$$

Puisque la fonction de survie  $S(x; b) = e^{-x/\theta} \mathbf{1}_{[0, \infty[}(x)$  est croissante en  $\theta$ , nous avons  $\gamma$ -intervalle de confiance  $]S_i, S_s[$  pour  $S(x; \theta)$  avec

$$S_i = e^{-x/\theta_i} \quad \text{et} \quad S_s = e^{-x/\theta_s}.$$

**15. a)** Il est clair que la statistique

$$T = \sum_{i=1}^n X_i$$

suit une loi binomiale  $B(n, \theta)$  de paramètres  $n$  et  $\theta$ . La fonction de répartition de  $T$  est

$$G(k; \theta) = \mathbf{P}_{\theta}\{T \leq k\} = \sum_{i=0}^k \binom{n}{i} \theta^i (1-\theta)^{n-i} =$$

$$I_{1-\theta}(n-k, k+1) = 1 - I_{\theta}(k+1, n-k), \quad k = 0, 1, \dots, n-1,$$

$$G(k; \theta) = 1, \quad \text{si } k = n,$$

où  $I_x(a, b)$  est la fonction de répartition de la loi beta de paramètres  $a$  et  $b$ , et

$$G(k-0; \theta) = \sum_{i=0}^{k-1} \binom{n}{i} \theta^i (1-\theta)^{n-i} = 1 - I_{\theta}(k, n-k+1), \quad k = 1, 2, \dots, n,$$

$$G(k-0; \theta) = 0, \quad \text{si } k = 0.$$

Les fonctions  $I$  et  $S$  sont

$$I(\theta; \mathbb{X}) = \begin{cases} I_{1-\theta}(n-T+1, T), & \text{si } T \neq 0 \\ 0, & \text{sinon,} \end{cases}$$

$$S(\theta; \mathbb{X}) = \begin{cases} I_{1-\theta}(n-T, T+1), & \text{si } T \neq n \\ 1, & \text{si } T = n. \end{cases}$$

On remarque que  $S(\theta; \mathbb{X})$  est strictement décroissante en  $\theta$  pour  $T \neq n$ , et  $I(\theta; \mathbb{X})$  est strictement décroissante en  $\theta$  pour  $T \neq 0$ , et par conséquent du théorème de Bolshev il suit que

$$I_{1-\theta_i}(n-T+1, T) = \gamma_1 \quad \text{pour } T \neq 0,$$

et donc

$$\theta_i = 0, \quad \text{si } T = 0,$$

$$I_{1-\theta_s}(n-T, T+1) = 1 - \gamma_1 \quad \text{pour } T \neq n,$$

et donc

$$\theta_s = 1, \quad \text{si } T = n.$$

Donc,

$$\theta_i = \begin{cases} 1 - x(\gamma_1; n-T+1, T), & \text{si } T \neq 0 \\ 0, & \text{si } T = 0, \end{cases}$$

$$\theta_s = \begin{cases} 1 - x(1-\gamma_1; n-T, T+1), & \text{si } T \neq n \\ 1, & \text{si } T = n, \end{cases}$$

où  $x(\gamma_1; a, b)$  est le  $\gamma_1$ -quantile de la distribution beta de paramètres  $a$  et  $b$ .

**16. b)** Dans ce cas

$$I(X; \theta) = F(X-0; \theta) \quad \text{et} \quad S(X; \theta) = F(X; \theta).$$

Si  $X = 1$  alors

$$I(1; \theta) = F(1-0; \theta) = F(0; \theta) = 0.$$

Du théorème de Bolshev il suit que la limite inférieure de confiance  $\theta_i$  pour  $\theta$  du niveau de confiance supérieur ou égal à  $\gamma_1$  est

$$\theta_i = \inf \theta = \inf ]0, 1[ = 0.$$

Si  $\gamma_1 = 1$ , alors  $\mathbf{P}\{\theta_i \leq \theta\} = \gamma_1$ , et donc  $\theta_i = 0$  est la limite inférieure de 1-confiance pour  $\theta$ . De l'autre côté la fonction

$$S(1; \theta) = F(1; \theta) = 1 - \theta, \quad \theta \in ]0, 1[.$$

est décroissante en  $\theta$  et donc du théorème de Bolshev nous avons

$$S(1; \theta_s) = 1 - \gamma_2,$$

d'où il s'ensuit que  $\theta_s = \gamma_2$ . Donc  $\gamma_1 = 1$  et  $\gamma_2$  limites de confiance pour  $\theta$  sont 0 et  $\gamma_2$ , et  $\gamma$ -intervalle de confiance pour  $\theta$  est  $]0, \gamma[$ , puisque pour  $\gamma_1 = 1$  l'égalité  $\gamma = \gamma_1 + \gamma_2 - 1$  est juste quand  $\gamma_2 = \gamma$ .

**17.** La fonction de vraisemblance

$$L = \exp\{-(X_1 + X_2 - 2\theta)\} \mathbf{1}\{X_{(1)} \geq \theta\},$$

$$l = \ln L = (2\theta - X_1 - X_2) \mathbf{1}\{\theta \leq X_{(1)}\}.$$

$l = \max$ , si  $\hat{\theta} = X_{(1)}$ , parce que sur l'intervalle  $]-\infty, X_{(1)}[$  la fonction  $l$  est croissante. On a

$$\mathbf{P}\{X_{(1)} > x\} = \mathbf{P}\{X_1 > x, X_2 > x\} = \left( \int_x^\infty e^{-(x-\theta)} dx \right)^2 = e^{-2(x-\theta)}, \quad x \geq \theta.$$

La fonction de répartition de  $X_{(1)}$

$$G(x) = F_{X_{(1)}}(x) = 1 - e^{-2(x-\theta)}, \quad x \geq \theta.$$

Notons  $T = X_{(1)}$ . Les fonctions  $I$  et  $S$  du théorème de Bolshev

$$I(\theta; \mathbb{X}) = S(\theta; \mathbb{X}) = G(X_{(1)}) = 1 - e^{-2(X_{(1)} - \theta)}$$

sont décroissantes, d'où on déduit que

$$1 - e^{-2(X_{(1)} - \theta_i)} = \gamma_1,$$

$$1 - e^{-2(X_{(1)} - \theta_s)} = 1 - \gamma_2,$$

où

$$\theta_i = X_{(1)} + \frac{1}{2} \ln(1 - \gamma_1),$$

$$\theta_s = X_{(1)} + \frac{1}{2} \ln \gamma_2.$$

L'intervalle  $]\theta_i, \theta_s[$  est  $\gamma$ -intervalle de confiance pour  $\theta$  si  $\gamma = \gamma_1 + \gamma_2 - 1$ .

La longueur de cet intervalle

$$\theta_s - \theta_i = \frac{1}{2} (\ln \gamma_2 - \ln(1 - \gamma_1)).$$

On cherche  $\gamma_1$  et  $\gamma_2$  tels que

$$\gamma_1 + \gamma_2 = 1 + \gamma, \quad 0.5 < \gamma_i \leq 1 \quad (i = 1, 2)$$

et pour lesquels la longueur  $\theta_s - \theta_i$  est minimale. on considère  $\theta_s - \theta_i$  comme fonction de  $\gamma_2$ . la dérivée

$$(\theta_s - \theta_i)' = \frac{1}{2} (\ln \gamma_2 - \ln \gamma_2 - \gamma)' =$$

$$\frac{1}{2} \left( \frac{1}{\gamma_2} - \frac{1}{\gamma_2 - \gamma} \right) < 0.$$

cette fonction est décroissante, donc  $\theta_s - \theta_i = \min$  si  $\gamma_2 = 1$  et  $\gamma_1 = 1 + \gamma - \gamma_2 = \gamma$ , d'où on tire que

$$\begin{aligned} \theta_i &= x_{(1)} + \frac{1}{2} \ln(1 - \gamma); \\ \theta_s &= x_{(1)}. \end{aligned}$$

**18.** il est évident que  $y_i - \theta$  suit la loi uniforme sur  $[-1, 1]$ , d'où il suit que la répartition de la variable aléatoire

$$t = x_1 + x_2 - 2\theta = y_1 + y_2$$

ne dépend pas de  $\theta$ . il est facile à montrer que

$$g(y) = \mathbf{P}\{t \leq y\} = \begin{cases} 0, & y \leq -2, \\ \frac{1}{8}(y+2)^2, & -2 \leq y \leq 0, \\ 1 - \frac{(y-2)^2}{8}, & 0 \leq y \leq 2, \\ 1, & y \geq 2. \end{cases}$$

la fonction

$$g(t) = g(x_1 + x_2 - 2\theta), \theta \in \mathbf{r}^1,$$

est décroissant en  $\theta$ . du théorème de bolshev il s'ensuit que les limites de confiance, inférieure et supérieure, de niveau de confiance  $\gamma_1$  et  $\gamma_2$  respectivement ( $0.5 < \gamma_i \leq 1$ ) vérifient les équations

$$g(x_1 + x_2 - 2\theta_i) = \gamma_1 \quad \text{et} \quad g(x_1 + x_2 - 2\theta_s) = 1 - \gamma_2,$$

d'où nous trouverons

$$\theta_i = \frac{x_1 + x_2}{2} - 1 + \sqrt{2(1 - \gamma_1)} \quad \text{et} \quad \theta_s = \frac{x_1 + x_2}{2} + 1 - \sqrt{2(1 - \gamma_2)}.$$

il est facile à montrer que pour  $\gamma = \gamma_1 + \gamma_2 - 1$  donné la fonction

$$\theta_s - \theta_i = 2 - \sqrt{2(1 - \gamma_1)} - \sqrt{2(1 - \gamma_2)}$$

prend sa valeur minimale (considérée comme fonction de  $\gamma_1$ ,  $0.5 < \gamma_1 \leq 1$ ) quand

$$\gamma_1 = \frac{1 + \gamma}{2}.$$

dans ce cas  $\gamma_2 = \frac{1 - \gamma}{2}$ , et donc le  $\gamma$ -intervalle de confiance le plus court pour  $\theta$  est  $]\theta_i, \theta_s[$  avec

$$\theta_i = \frac{X_1 + X_2}{2} - 1 + \sqrt{1 - \gamma} \quad \text{et} \quad \theta_s = \frac{X_1 + X_2}{2} + 1 - \sqrt{1 - \gamma}.$$

**22.** La fonction de vraisemblance est :

$$L(\mathbb{X}; \theta) = \theta^n \exp \left\{ -\theta \sum_{i=1}^n X_i \right\} \mathbf{1}\{X_{(1)} > 0\}.$$

Le rapport de vraisemblance sera supérieur à  $c$  :

$$L(\mathbb{X}; \theta) / L(\mathbb{X}; \theta_0) = \left( \frac{\theta}{\theta_0} \right)^n \exp \left\{ -(\theta - \theta_0) \sum_{i=1}^n X_i \right\} > c$$

si et seulement si

$$\sum_{i=1}^n X_i > c_1$$

où  $c_1$  est une constante. On a utilisé le fait que  $\theta < \theta_0$ . On cherche  $c_1$  tel que :

$$\alpha = \mathbf{P}_{\theta_0} \left\{ \sum_{i=1}^n X_i > c_1 \right\} = \mathbf{P}_{\theta_0} \left\{ 2\theta_0 \sum_{i=1}^n X_i > 2\theta_0 c_1 \right\} = \mathbf{P} \{ \chi^2(2n) > 2\theta_0 c_1 \},$$

d'où

$$2\theta_0 c_1 = \chi_{1-\alpha}^2(2n)$$

et donc

$$c_1 = \frac{1}{2\theta_0} \chi_{1-\alpha}^2(2n).$$

Le test ne dépend pas de  $\theta$ , donc il est UPP pour l'alternative  $\theta < \theta_0$ . La fonction de puissance est :

$$\beta(\theta) = \mathbf{P}_{\theta} \left\{ \sum_{i=1}^n X_i > c_1 \right\} = \mathbf{P}_{\theta} \{ \chi^2(2n) > 2\theta c_1 \} = \mathcal{P}(2\theta c_1, 2n) = \mathcal{P} \left( \frac{\theta}{\theta_0} \chi_{1-\alpha}^2(2n), 2n \right),$$

où  $\mathcal{P}(x, n) = P \{ \chi^2(n) > x \}$ .  $\beta(\theta_0)$  est décroissante,

$$\lim_{\theta \rightarrow 0^+} \beta(\theta) = \mathcal{P}(0, n) = 1, \quad \beta(\theta_0) = \alpha.$$

### Figure 1.

Le test est biaisé pour l'alternatives  $\theta > \theta_0$ .

**23.** La fonction de vraisemblance est

$$L(\mathbb{X}; \theta) = \text{const} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2 \right\}.$$

Le rapport de vraisemblance sera supérieur à  $c$  :

$$L(\mathbb{X}; 1)/L(\mathbb{X}; 0) = \exp \left\{ -\frac{1}{2} \sum_{i=1}^n [(X_i - 1)^2 - X_i^2] \right\} = \exp \left\{ \sum_{i=1}^n (X_i - 1/2) \right\} > c$$

si et seulement si

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i > c.$$

Les risques de première et deuxième espèce sont :

$$\alpha = \mathbf{P}_0\{\bar{X} > c\} \leq 0.01,$$

$$\beta = \mathbf{P}_1\{\bar{X} \leq c\} \leq 0.01.$$

Si  $\theta = 0$ ,  $\bar{X} \sim N(0, \frac{1}{n})$ ,  $\sqrt{n}\bar{X} \sim N(0, 1)$ .

Si  $\theta = 1$ ,  $\bar{X} \sim N(1, \frac{1}{n})$ ,  $\sqrt{n}(\bar{X} - 1) \sim N(0, 1)$ .

Donc

$$1 - \Phi(\sqrt{nc}) \leq 0.01$$

$$\Phi(\sqrt{n}(c - 1)) \leq 0.01$$

où

$$\begin{aligned} \sqrt{nc} &\geq \Phi^{-1}(0.99) \\ \sqrt{n}(c - 1) &\leq 1 - \Phi^{-1}(0.99). \end{aligned} \quad (1)$$

Notons  $a = \Phi^{-1}(0.99) \approx 2.326$ ,  $m = \sqrt{n}$ . Il faut trouver le plus petit  $m$  vérifiant

$$mc \geq a,$$

$$m(c - 1) \leq -a,$$

où

$$c \geq \frac{a}{m}, \quad c \leq 1 - \frac{a}{m}. \quad (2)$$

### Figure 2

La fonction  $g(m) = \frac{a}{m}$  est décroissante, la fonction  $h(m) = 1 - \frac{a}{m}$  est croissante. On cherche le point d'intersection  $m^*$  :

$$\frac{a}{m} = 1 - \frac{a}{m},$$

donc  $m^* = 2a \approx 4.652$ .

$$\lim_{m \rightarrow \infty} g(m) = 0, \quad \lim_{m \rightarrow \infty} h(m) = 1, \quad \lim_{m \rightarrow 0+0} g(m) = +\infty,$$

$$\lim_{m \rightarrow 0+0} h(m) = -\infty; \quad h(m) = 0, \quad \text{si } m = a \approx 2.326.$$

Dans la région hachurée (figure 2) les inégalités (1) sont vérifiées.

Parce que

$$2 \cdot 2.325 < m^* < 2 \cdot 2.33$$

et

$$21.6 < (m^*)^2 < 21.8,$$

le plus petit nombre naturel pour lequel les inégalités (1) sont vérifiées est  $n = [(m^*)^2] + 1 = 22$ .

**24.** La fonction de vraisemblance est

$$L(\theta) = \frac{1}{\theta^n} \mathbf{1}\{0 \leq X_{(1)} \leq X_{(n)} \leq \theta\}.$$

a)  $H : \theta = \theta_0, \quad \bar{H} : \theta > \theta_0$ .

On cherche le test pur de Neyman-Pearson de niveau  $\alpha$  :

$$\varphi(\mathbb{X}) = \begin{cases} 1, & \text{si } L(\theta) > kL(\theta_0) \\ 0, & \text{sinon} \end{cases}$$

Si  $X_{(n)} \leq \theta_0$ , l'inégalité

$$L(\theta) > kL(\theta_0) \tag{1}$$

est vérifiée pour  $k > 0$ , si et seulement si

$$\left(\frac{\theta_0}{\theta}\right)^n > k.$$

Si  $X_{(n)} > \theta_0$ , l'inégalité (1) est toujours vérifiée. Prenons  $k < \left(\frac{\theta_0}{\theta}\right)^n$  :

$$\alpha = \mathbf{P}_{\theta_0}\{X_{(n)} \leq \theta_0\} + \mathbf{P}_{\theta_0}\{X_{(n)} > \theta_0\} = 1 + 0 = 1.$$

Il n'y a pas de test pur de niveau  $\alpha < 1$ .

Prenons  $k \leq \left(\frac{\theta_0}{\theta}\right)^n$  :

$$\alpha = \mathbf{P}_{\theta_0}\{X_{(n)} > \theta_0\} = 0.$$

Il n'y a pas non plus de test pur de niveau de signification  $\alpha$ . On cherche le test randomisé de Neyman-Pearson

$$\varphi(\mathbb{X}) = \begin{cases} 1, & \text{si } L(\theta) > kL(\theta_0), \\ \gamma, & \text{si } L(\theta) = kL(\theta_0), \\ 0, & \text{sinon.} \end{cases} \tag{2}$$

Si  $X_{(n)} \leq \theta_0$ , l'égalité

$$L(\theta) = kL(\theta_0) \tag{3}$$

est vérifiée pour  $k > 0$  si et seulement si

$$\left(\frac{\theta_0}{\theta}\right)^n = k.$$

Si  $X_{(n)} > \theta_0$ , l'égalité (3) n'est pas vérifiée. Prenons  $k = \left(\frac{\theta_0}{\theta}\right)^n$  :

$$\varphi(\mathbb{X}) = \begin{cases} 1, & X_{(n)} > \theta_0, \\ \gamma, & X_{(n)} \leq \theta_0, \end{cases}$$

car l'égalité (1) est vérifiée si  $X_{(n)} > \theta_0$ .

Le niveau de signification est :

$$\alpha = \mathbf{E}_{\theta_0} \varphi(\mathbb{X}) = \mathbf{P}_{\theta_0} \{X_{(n)} > \theta_0\} + \gamma \mathbf{P}_{\theta_0} \{X_{(n)} \leq \theta_0\} = \gamma.$$

Donc on a

$$\varphi(\mathbb{X}) = \begin{cases} 1, & X_{(n)} > \theta_0, \\ \alpha, & X_{(n)} \leq \theta_0. \end{cases}$$

D'après le lemme de Neyman-Pearson le test  $\varphi$  est UPP car il ne dépend pas de  $\theta > \theta_0$ .

b)  $H : \theta = \theta_0, \bar{H} : \theta < \theta_0$ . On cherche le test pur de Neyman-Pearson.

Si  $X_{(n)} \leq \theta$ , l'inégalité (1) est vérifiée pour  $k > 0$  si et seulement si

$$\left(\frac{\theta_0}{\theta}\right)^n > k.$$

Si  $X_{(n)} > \theta$ , l'inégalité (1) n'est pas vérifiée.

Prenons  $k < \left(\frac{\theta_0}{\theta}\right)^n$ . Dans ce cas

$$\varphi(\mathbb{X}) = \begin{cases} 1, & X_{(n)} \leq \theta, \\ 0, & \text{sinon,} \end{cases}$$

et

$$\alpha = \mathbf{P}_{\theta_0} \{X_{(n)} \leq \theta\} = \left(\frac{\theta}{\theta_0}\right)^n.$$

Le niveau de signification est  $\alpha$  pour l'alternative  $\theta_1 = \theta_0 \alpha^{1/n}$ . Sous cette alternative

$$\varphi(\mathbb{X}) = \begin{cases} 1, & X_{(n)} \leq \theta_0 \alpha^{1/n}, \\ 0, & \text{sinon.} \end{cases}$$

Dans le cas d'autres alternatives cherchons le test randomisé (2).

Si  $X_{(n)} \leq \theta$ , l'égalité (3) est vérifiée si et seulement si

$$\left(\frac{\theta_0}{\theta}\right)^n = k.$$

Pour  $X_{(n)} > \theta$ , l'égalité (3) n'est pas vérifiée.

Prenons  $k = \left(\frac{\theta_0}{\theta}\right)^n$ . Le test de Neyman-Pearson donne

$$\varphi_1(\mathbb{X}) = \begin{cases} \gamma, & X_{(n)} \leq \theta, \\ 0, & \text{sinon,} \end{cases}$$

$$\alpha = \mathbf{E}_{\theta_0} \varphi_1(\mathbb{X}) = \gamma \mathbf{P}\{X_{(n)} \leq \theta\} = \gamma \left(\frac{\theta}{\theta_0}\right)^n,$$

$$\gamma = \alpha \left(\frac{\theta_0}{\theta}\right)^n.$$

L'inégalité  $\gamma \leq 1$  est vérifiée si  $\theta \geq \theta_0 \alpha^{1/n}$ .

Le test de Neyman-Pearson n'existe pas quand  $\theta < \theta_0 \alpha^{1/n}$ .

Pour  $\theta \geq \theta_0 \alpha^{1/n}$

$$\varphi(\mathbb{X}) = \alpha \left(\frac{\theta_0}{\theta}\right)^n.$$

On cherche la puissance de  $\varphi$  et  $\varphi_1$  pour  $\theta \geq \theta_0 \alpha^{1/n}$  :

$$\mathbf{E}_{\theta} \varphi(\mathbb{X}) = \mathbf{P}_{\theta}\{X_{(n)} \leq \theta_0 \alpha^{1/n}\} = \left(\frac{\theta_0 \alpha^{1/n}}{\theta}\right)^n = \left(\frac{\theta_0}{\theta}\right)^n \alpha,$$

$$\mathbf{E}_{\theta} \varphi_1(\mathbb{X}) = \alpha \left(\frac{\theta_0}{\theta}\right)^n \mathbf{P}_{\theta}\{X_{(n)} \leq \theta\} = \left(\frac{\theta_0}{\theta}\right)^n \alpha.$$

La puissance de  $\varphi$  est la même que la puissance du test le plus puissant  $\varphi_1$  pour l'alternative  $\theta \geq \theta_0 \alpha^{1/n}$ . Si  $\theta < \theta_0 \alpha^{1/n}$

$$\mathbf{E}_{\theta} \varphi(\mathbb{X}) = \mathbf{P}_{\theta}\{X_{(n)} \leq \theta_0 \alpha^{1/n}\} = 1.$$

Donc, le test  $\varphi$  est le plus puissant pour toutes alternatives  $\theta > 0$ .

c) On a obtenu que le test

$$\varphi(\mathbb{X}) = \begin{cases} 1, & X_{(n)} > \theta_0 \\ \alpha, & X_{(n)} \leq \theta_0 \end{cases}$$

est le plus puissant pour l'alternative  $\theta > \theta_0$  et le test

$$\varphi_0(\mathbb{X}) = \begin{cases} 1, & X_{(n)} \leq \theta_0 \alpha^{1/n} \\ 0, & \text{sinon} \end{cases}$$

est le plus puissant pour l'alternative  $\theta < \theta_0$  et les puissances de ces tests

$$\mathbf{E}_{\theta} \varphi(\mathbb{X}) = \mathbf{P}_{\theta}\{X_{(n)} > \theta_0\} + \alpha \mathbf{P}_{\theta}\{X_{(n)} \leq \theta_0\},$$

$$\mathbf{E}_{\theta} \varphi_0(\mathbb{X}) = \mathbf{P}_{\theta}\{X_{(n)} \leq \theta_0 \alpha^{1/n}\}.$$

Définissons

$$\varphi_2(\mathbb{X}) = \begin{cases} 1, & \text{si } X_{(n)} > \theta_0 \text{ ou } X_{(n)} \leq \theta_0 \alpha^{1/n} \\ 0, & \text{sinon.} \end{cases}$$

Ce test a le niveau  $\alpha$  car

$$\mathbf{E}_{\theta_0} \varphi_2(\mathbb{X}) = \mathbf{P}_{\theta_0}\{X_{(n)} \leq \theta_0 \alpha^{1/n}\} = \alpha.$$

La puissance de  $\varphi_2$  :

$$\mathbf{E}_{\theta} \varphi_2(\mathbb{X}) = \mathbf{P}_{\theta}\{X_{(n)} > \theta_0\} + \mathbf{P}_{\theta}\{X_{(n)} \leq \theta_0 \alpha^{1/n}\}.$$

Si  $\theta < \theta_0$

$$\mathbf{E}_\theta \varphi_2(\mathbb{X}) = \mathbf{P}_\theta \{X_{(n)} \leq \theta_0 \alpha^{1/n}\} = \mathbf{E}_{\theta_0} \varphi_0(\mathbb{X}),$$

si  $\theta > \theta_0$

$$\mathbf{E}_\theta \varphi_2(\mathbb{X}) = \mathbf{P}_\theta \{X_{(n)} > \theta_0\} + \alpha \left(\frac{\theta_0}{\theta}\right)^n =$$

$$\mathbf{P}_\theta \{X_{(n)} > \theta_0\} + \alpha \mathbf{P}_\theta \{X_{(n)} \leq \theta_0\} = \mathbf{E}_\theta \varphi(\mathbb{X}),$$

$\varphi_2$  est le test UPP pour l'hypothèse  $H : \theta = \theta_0$  contre l'alternative bilatérale  $\bar{H} : \theta \neq \theta_0$ .

## BIBLIOGRAPHIE.

- Aguirre N.** (1993). *Test d'ajustement du chi-deux pour une loi logistique*. XXV Journée de Statistique, Vannes, Session **35** (191).
- Aguirre N. and Nikulin M.** (1994) *Chi squared goodness-of-fit test for the family of logistic distributions*. *Kybernetika*, **30** 3, p. 214-222.
- Aalen, O.** (1980). A model for nonparametric regression analysis of counting processes. In. *Mathematical Statistics and Probability Theory*, Lecture Notes in Statistics, **2**, (Eds. W. Klonecki, A. Kozek and J. Rosinski), New York : Springer Verlag, 1-25.
- Achtziger W., Bendsøe M.P. Taylor J.E.** (1998). Bounds on the effect of progressive structural degradation. *J. Mech. Phys. Solids*, **46**, 6, 1055-1087.
- Anderson T.W.** (1962). *On the distribution of the two-sample Cramer-von Mises criterion*. *Annals of the Mathematical Statistics*, **33**, p.1148- 1159.
- Anderson T.W. and Darling D.A.** (1952). *Asymptotic theory of certain "Goodness of fit" criteria based on stochastic processes*. *Annals of the Mathematical Statistics*, **23**, p.193-212.
- P.K.Andersen and R.D.Gill.** (1982). "Cox's regression model for counting processes : A large sample study", *Ann. Statist*, **10**, p. 1100-1120.
- P.K.Andersen, O.Borgan, R.D.Gill and N.Keiding**, (1993). *Statistical Models Based on Counting Processes*, New York : Springer-Verlag.
- Andersen, P.K.** (1991). Survival analysis 1981-1991 : The second decade of the proportional hazards regression model. *Statistics in Medicine*, **10**, # 12, 1931-1941.
- V.Bagdonavičius.** (1978.) "Testing the hypothesis of the additive accumulation of damages". *Probab. Theory and its Appl.*, **23**, pp. 403-408.
- V.Bagdonavičius and M.Nikulin.** (1994). " Stochastic models of accelerated life". In : *Advanced Topics in Stochastic Modelling*, (eds. J.Gutierrez, M.Valderrama), Singapore : World Scient.
- Bagdonavičius, V., Nikulin, M.** (1995). *Semiparametric models in accelerated life testing*. Queen's Papers in Pure and Applied Mathematics. Queen's University, Kingston, Ontario, Canada. **98**, 70p.
- V.Bagdonavičius and M.Nikulin.** (1996). "Analyses of generalized additive semiparametric models ", *Comptes Rendus, Academie des Sciences de Paris*, **323**, 9, Série I, 1079-1084.
- V.Bagdonavičius and M.Nikulin.** (1997a). "Transfer functionals and semiparametric regression models", *Biometrika*, vol. 84 pp. 365-378.
- V.Bagdonavičius and M.Nikulin.** (1997b). "Asymptotic analysis of semiparametric models in survival analysis and accelerated life testing", *Statistics*, vol. 29 pp. 261-283.
- V.Bagdonavičius and M.Nikulin.** (1997). "Semiparametric estimation in the generalized additive multiplicative model". In : *Probability and Statistics*, **2**, (Eds : I.A. Ibragimov, V.A. Sudakov), Proceeding of the Steklov Mathematical Institute, St. Petersburg, 7-27.
- V.Bagdonavičius and M.Nikulin.** (1997). "Statistical analysis of the generalized additive semiparametric survival model with random covariates", *Qüestiió, Qüestiió*, **21**, # 1-2, p. 273-291.
- V.Bagdonavičius and M.Nikulin.** (1997). "Sur l'application des stress en escalier dans les expériences accélérées ", *Comptes Rendus, Academie des Sciences de Paris*, **325**, Série I, p. 523-526.

- V.Bagdonavičius and M.Nikulin.** (1997). "Accelerated life testing when a process of production is unstable", *Statistics and Probability Letters*, **1997**, **35**, p. 269-279.
- V.Bagdonavičius and M.Nikulin.** (1997). "Transfer functionals and semiparametric regression models", *Biometrika*, **1997**, **84**, 2, p. 365-378.
- V.Bagdonavičius and M.Nikulin.** (1997). "Analysis of general semiparametric models with random covariates", *Revue Roumaine de mathématiques Pures et Appliquées*, **42**, # 5-6, p. 351-369.
- V.Bagdonavičius and M.Nikulin.** (1997). "Asymptotic analysis of semiparametric models in survival analysis and accelerated life testing", *Statistics*, **29**, p.261-283.
- V.Bagdonavičius and M.Nikulin.** (1997). "Some rank tests for multivariate censored data". In : *Advances in the Theory and Practice of Statistics : A volume in Honor of Samuel Kotz.* (eds. N.L.Johnson and N.Balakrishnan), New York : J. Wiley and Sons, 193-207.
- V.Bagdonavičius and M.Nikulin.** (1998a). *Additive and multiplicative semiparametric models in accelerated life testing and survival analysis.* Queen's Papers in Pure and Applied Mathematics, vol. 108, Kingston : Queen's University, Canada.
- V.Bagdonavičius and M.Nikulin.** (1998b). "Estimation in generalized proportional hazards model". *C.R.Acad.Sci.Paris, Serie I*, **326**, pp. 1415-1420.
- V.Bagdonavičius, S.Malov and M.Nikulin.** (1998). "Characterizations and semiparametric regression estimation in Archimedean copulas", *Journal of Applied Statistical Sciences*, **8**, 549-562.
- V.Bagdonavičius, V.Nikoulina and M.Nikulin.** (1998). "Bolshev's method of confidence interval construction", *Qüestiió*, **21**, # 3, 549-562.
- V.Bagdonavičius and M.Nikulin.** (1999). "Generalized proportional hazards model based on modified partial likelihood", *Life Data Analysis*, **5**, 329-350.
- Bagdonavičius, V. and Nikulin, M.** (2001). Estimation in Degradation Models with Explanatory variables, *Lifetime Data Analysis*, **7**, 85-103.
- V.Bagdonavičius and M.Nikulin.** (1999). "Model Buildings in Reliability", In : *Probabilistic and Statistical Models in Reliability*, (Eds. N. Limnios and D. Ionescu), Boston : Birkhauser, 51-74.
- V.Bagdonavičius and M.Nikulin.** (1999). "On Nonparametric Estimation From Accelerated Experiments", In : *Probabilistic and Statistical Models in Reliability*, (Eds. N. Limnios and D. Ionescu), Boston : Birkhauser, 75-90.
- V.Bagdonavičius and M.Nikulin.** (2000)"Modèle statistique de dégradation avec des covariables dépendant de temps", *Comptes Rendus, Academie des Sciences de Paris*, **2000**, **329**, Série I, p. 131-134.
- V.Bagdonavičius and M.Nikulin.** (2000). "On goodness-of-fit for the Linear Transformation and Frailty models", *Statistics and Probability Letters*, **47**, #2, 177-188.
- V.Bagdonavičius and M.Nikulin.** (2000). "On nonparametric estimation in accelerated experiments with step-stresses", *Statistics*, **33**, 349-365.
- V.Bagdonavičius, L.Gerville-Réache, V.Nikoulina, M.Nikulin.** (2000) "Expériences Accélérées : Analyse Statistique du Modèle Standard de Vie Accélérée", *Revue de Statistique Appliquée*, **XLVIII**, #3, 5-38.
- V.Bagdonavičius and M.Nikulin.** (2001). *Accelerated Life Models*, Chapman&Hall/CRC, 348p.
- V.Bagdonavičius, M.Nikulin.** (2003) Stochastic Modeling in survival analysis and its influence on duration analysis. In : " Advances in Survival Analysis. v.23 ". (by N.Balakrishnan

and C.R.Rao) North-Holland.

**V.Bagdonavičius, M.Nikulin.** (2003) " Semiparametric statistical analysis for aging and longevity ". In : "Advances in statistical inferential methods : theory and applications" (Ed. by V. Voinov), Gylym : Almaty, ISBN 9965-07-253-, p.17-30.

**Bagdonavičius, V., Bikelis, A., Kazakevičius, A. and Nikulin, M.** (2002). Non-parametric estimation from simultaneous degradation and failure data, *Comptes Rendus, Academie des Sciences de Paris*, v. **335**, 183-188.

**V.Bagdonavičius, A.Bikelis, V.Kazakevicius, M.Nikulin.** (2003) Estimation from simultaneous degradation and failure time data. In : *Mathematical and Statistical Methods in Reliability*,(B. Lindqvist and Kjell A Doksum, eds.), World Scientific Publishing, Series on Quality, Reliability and Engineering Statistics, **7**, p.301-318.

**Bagdonavičius, V., Hafdi, M., Himdi, K., Nikulin, M.** (2003). "Statistical analysis of the Generalised Linear Proportionnal Hazards model." Proceedings of the Steklov Mathematical Institute, St.Petersburg, : *Probability and Statistics*, 6., v.**294**, p.5-18, (ISSN 0373-2703).

**Bagdonavičius, V., Haghghi, F., Nikulin, M.** (2003). *Statistical Analysis of General Degradation Path Model and Failure time data with Multiple failure modes*, Preprint de l'IFR-99 Sané Publique, Université Victor Segalen Bordeaux 2.

**Bagdonavičius, V.** (1978). Testing the hypothesis of the additive accumulation of damages. *Probab. Theory and its Appl.*, **23**, No. 2, 403-408.

**Bagdonavičius V., M.Hafdi and Nikulin M.** (2002). The Generalized Proportional Hazards Model and its Application for Statistical Analysis of the Hsieh Model. In : Proceedings of *The Second Euro-Japanese Workshop on Stochastic Risk Modelling for Finance, Insurance, Production and Reliability*, September 18-20, Chamonix, France, (Eds. T.Dohi, N.Limnios, S.Osaki), p. 42-53.

**Bagdonavičius V., Hafdi, M., El Himdi, K. and Nikulin M.** (2002). *Analyse du modèle des hazards proportionnels généralisé. Application sur les données du cancer des poumons*. Preprint 0201, I.F.R. "Santé Publique".

**Bagdonavičius V., Hafdi, M., El Himdi, K. and Nikulin, M.** (2002). *Analysis of Survival Data with Cross-Effects of Survival Functions. Applications for Chemo and Radiotherapy Data*. Preprint 0202, I.F.R. "Santé Publique".

**Bagdonavičius, V. and Nikulin, M.** (2004). Semiparametric analysis of Degradation and Failure Time Models. In : *Semiparametric Models and Applications for Reliability, Survival Analysis and Quality of Life*, (Eds. : M.Nikulin, N.Balakrishnan, M.Mesbah, N.Limnios), Birkhauser : Boston.

**Balakrishnan N., Ed.** (1992) *Handbook of the logistic distribution*. New York : Marcel Dekker.

**Balakrishnan, E., Nelson, M. I., Wake, G. C.** (1999). Radiative ignition of combustible materials. I. Polymeric materials undergoing nonflaming thermal degradation :the critical storage problem.*Math. Comput. Modelling*,**30**, # 11-12, 177-195.

**Berger T., Zhang Z.** (1983). Minimum breakdown degradation in binary source encoding. *IEEE Trans. Inform. Theory*, **29**, # 6, 807-814.

**Boulanger, M., Escobar, L.A.** (1994). Experimental design for a class of accelerated degradation tests. *Technometrics*, **36**, 260-272.

**Burchard A.** (1994). Substrate degradation by a mutualistic association of two species in the chemostat. *J. Math. Biol.*, **32**, #5, 465-489.

**Busenberg S., Tang B.** (1994). Mathematical models of the early embryonic cell cycle :

the role of MPF activation and cyclin degradation. *J.Math.Biol.*, **32**, #6, 573-596.

**Birnbaum Z.W.** (1952). *Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size*. *JASA*, **v.47**, p.425.

**Bolshev L.N. and Nikulin M.S.** (1975) *One solution of the problem of homogeneity*. *Serdika, Bulgarsko Mathematischesko Spicanie*, **v.1**, p.104-109.

**Bolshev L.N. and Smirnov N.N.** (1968). *Tables of mathematical statistics*. Moscow : Nauka (in russian).

**S.C.Cheng, L.J.Wei and Z.Ying.** (1995). "Analysis of tranformation models with censored data", *Biometrika*, vol. 82 pp. 835-846.

**Chernoff H., Lehmann E.L.** (1954) *The use of maximum likelihood estimator in  $\chi^2$  tests for goodness of fit*. *Ann. Math. Stat.*, **25**, 579-586.

**Cantrell R.S., Cosner C., Fagan W. F.**(1998). Competitive reversals inside ecological reserves : the role of external habitat degradation. *J. Math. Biol.*, **37**, #6, 491-533.

**Carasso A.S., Sanderson J.G., Hyman J.M.** (1978). Digital removal of random media image degradations by solving the diffusion equation backwards in time. *SIAM J. Numer. Anal.* **15**, #2, 344-367.

**Carey, M.B., Koenig,R.N.** (1991). "Reliability assessment based on accelerated degradation : a case study. *IEEE Transactions on Reliability*" ,**40**, 499-506.

**Chiao, C.H., Hamada, M.** (1996). Using Degradation Data from an Experimet to Achive Robust Reliability for Light Emmitining Diodes, *Quality and Reliability Engineering International*, **12**, 89-94.

**Cinlar,E.** (1980). On a generalization of gamma processes, *J.Appl.Probab.*,**17**, 467-480.

**Cramer H.** (1946). *Mathematical methods of statistics*. Princeton University Press, Princeton, N.J.

**D.R.Cox.** (1972). "Regression models and life tables", *J.Roy.Statist. Soc.*, B, vol. 34 pp. 187-220.

**D.R.Cox and D.Oakes.** (1984). *Analysis of Survival Date*, London : Chapman and Hall.

**Cox, D.R.** (1975) Partial likelihood. *Biometrika*, **62**, 269-276.

**Cox, D.R.**(1999). Some Remarks on Failure-times, Surrogate Markers, Degradation, Wear, and the Quality of Life, *Lifetime Data Analysis*, **5**, 307-314, 1999.

**D.M.Dabrowska and K.A.Doksum.** (1988a). "Estimation and Testing in a Two-Sample Generalized Odds-Raparte Model", *JASA*, **83** pp. 744-749.

**D.M.Dabrowska and K.A.Doksum.** (1988b). "Partial likelihood in transformation model with censored data", *Scand. J. Statist.*, **15**, pp. 1-23.

**Darling D.A.** (1957) *The Kolmogorov-Smirnov, Cramer-fon-Mises tests*. *Ann. Math. Statist.*,**28**, p.1-7.

**Dowling, N.E.**(1993). *Mechanical Behavior of Materials*, Prentice Hall : Englewood Cliffs.

**Doksum, K.A., Hoyland, A.**(1992). Models for variable-stress accelerated life testing experiment based on Wiener processes and the inverse Gaussian distribution, *Technometrics*, **34**, 74-82.

**Doksum,K.A., Normand, S.-L.T.**(1995). "Gaussian Models for Degradation Processes - Part I : Methods for the Analysis of Biomarker Data", *Lifetime Data Analysis*,**1**, 131-144.

**Doksum K.A., Normand S.-L.T.** (1996). Models for degradation processes and event times based on Gaussian processes. *Lifetime data : models in reliability and survival analysis* (Cambridge, MA, 1994), 85-91.Dordrecht : Kluwer Acad. Publ.

- Droesbeke, J.-J., Fichet B. & Tassi P.**, (1989). *Analyse statistique des durées de vie*, Paris : Economica.
- Drost F.** (1988) *Asymptotics for generalized chi-square goodness-of-fit tests*, Amsterdam : Center for Mathematics and Computer Sciences, CWI Tracts, **48**.
- Dzhaparidze, K.O. and Nikulin M.S.** (1974). *On a modification of the standard statistics of Pearson*. *Theory of probability and its applications*, **19**, #4, p.851-852.
- Dzhaparidze, K.O. and Nikulin M.S.** (1982). *Probability distributions of the Kolmogorov and omega-square statistics for continuous distributions with shift and scale parameters*. *Journal of Soviet Mathematics*, **20**, p.2147-2163.
- Dzhaparidze, K.O., Nikulin, M.S.** (1995), On the computation of the chi-square type statistics, *Journal of Mathematical Sciences*, **75**, 5, 1910-1921.
- Fasano A., Primicerio M., Rosso F.** (1992). On quasi-steady axisymmetric flows of Bingham type with stress-induced degradation. *Computing*, **49**, # 3, 213-237.
- Friedrich J.** (1999). A dual reciprocity boundary element model for the degradation of strongly eroded archaeological signs. *Math. Comput. Simulation*, **48**, 3, 281-293.
- Gajewski, H., Sparing, H.-D.** (1992). On a model of a polycondensation process with thermal degradation. *Z. Angew. Math. Mech.*, **62**, #11, 615-626.
- Garrigoux, C., Meeker, W.Q.** (1994). A reliability model for planning in-service inspections for components subject to degradation failure. *Pakistan J. Statist.*, **10**, 1, 79-98.
- Gupta, R.** (1991). Analysis of a two-unit cold standby system with degradation and linearly increasing failure rates. *Internat. J. Systems Sci.*, **22**, #11, 2329-2338.
- Gerville-Réache L., Nikulin, M.** (2000). " Analyse statistique du modèle de Makeham " *Revue Roumaine Math. Pure et Appl.*, **45**, #6, 947-957.
- Gihman, I.I.** (1961) *On the empirical distribution function in the case of grouping data*. In : Selected Translation in Mathematical Statistics and Probability, **1**, p.77-81.
- Grizzle, J.E.** (1961) *A new method of testing hypotheses and estimating parameters for the logistic model*. *Biometrics*, **17**, p.372-385.
- Habib, M.G., Thomas, D.R.** (1986). Chi-square goodness-of-fit tests for randomly censored data. *Annals of Statistics*, **14**, 759-765.
- Haghighi, F., Nikulin, M** (2003). Chi-square type test for power generalized Weibull family. In : *Advances in statistical inferential methods : theory and applications*, (Ed. by V. Voinov), Gylym : Almaty, p.89-105.
- Hamada, M.**(1995). "Analysis of Experiments for Reliability Improvement and Robust Reliability", In : *Recent Advances in Life-Testing and Reliability*, (Ed. N. Balakrishnan), CRC Press : Boca Raton.
- Hald, A.** (1952) *Statistical Theory with Engineering Applications*. Wiley, New York.
- Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions, *Biometrika*, **73**, 3, 387-396.
- Hsieh, F.** (2001). On heteroscedastic hazards regression models : theory and application. *Journal of the Royal Statistical Society, Series B* **63**, 63-79.
- Huber-Carol C.** (1989). *Statistique au PCEM* . Masson, Paris .
- Huber C and Nikulin M.S.** (1993). *Classical random walks and some statistical problems*. In : Rings and modules. Limit theorems of probability theory.#3. St. Petersburg State University.
- Huber C.** (1991). *Modeles log-lineaires*. Preprinte de l'Université Paris 5, DEA STATISTIQUE ET SANTE, 50 p.

- Huber-Carol C.** (1991). *Statistique*. Preprint de l'Université Paris 5, Maitrise de Sciences Biologiques et Medicales d'Informatique, Statistique et Epidémiologie et DUPESB, 134 p.
- Huber C.** (1991). *Elements de statistique générale. Choix et réduction d'un modele statistique*. Preprint de l'Université Paris 5, 48 p.
- Huber C.** (2000). Censored and Truncated Lifetime Data. In : *Recent Advances in Reliability Theory*. (Eds. N. Limnios, M.Nikulin). Boston : Birkhauser, 291-306.
- C.Huber and M.Nikulin.** (1997). "Remarques sur le maximum de vraisemblance", *Qüestiió*, **21**, # 1-2, p. 37-58 (avec C. Huber).
- Igaki N., Sumita U., Kowada M.,** (1998). On a generalized M/G/1 queue with service degradation/enforcement. *J. Oper. Res. Soc. Japan*, **41**, 3, 415-429.
- Jayanti P, Chandra T. D., Toueg S.** (1999). The cost of graceful degradation for omission failures. *Inform. Process. Lett.*, **71**, # 3-4, 167-172.
- Kleinbaum, D.** (1996). *Survival Analysis : A Self-Learning text*. New York : Springer-Verlag.
- Klein, J.P. and Moeschberger, M.L.** (1997). *Survival Analysis*, New York : Springer.
- Kalbfleisch J.D., Prentice R.L.** (1980) *The Statistical Analysis of Failure Time Data*. New York : J. Wiley and Sons.
- Kaplan E.L. and Meier P.** (1958) *Nonparametric estimation from incomplete observations*. *J.Am.Stat.Assoc.*, **53**, p.457-481.
- Khalfina N.M.** (1983) *Some asymptotic results associated with the Chauvenet test for multidimensional random variables*. *Journal of Soviet Mathematics*, **23**, #1, p.99-106.
- Klinger D.J.(1992). "Failure time and rate constant of degradation : an argument for the inverse relationship". *Microelectronics and Reliability*,**32**, 987-994.
- Klimontovich, Yu. L.** (1997). *Chaoticity, degradation and self-organization in open systems. Self-organization of complex structures*, ( Berlin, 1995), 37-50. Amsterdam : Gordon and Breach.
- Koike T., Kameda H.** (1973). Reliability theory of structures with strength degradation in load history. *Mem. Fac. Engrg. Kyoto Univ.*, **35**, 331-360.
- Kolmogorov A.N.** (1933). *Sulla determinazione empirica di una legge di distribuzione*. *Giorn.Ist.Ital.Attuari*, **4**, p.83-91.
- Kolmogorov A.N.** (1951). *Une généralisation d'une formule de Poisson*. *Uspekhi Mat.Nauk.*, **6**,p. 133-134.
- Lawless J.F.** (1982) *Statistical Models and Methods for Lifetime Data*. New York : J. Wiley and Sons.
- Lawless,J.,Hu,J., and Cao, J.**(1995). Methods for the estimation of failure distributions and rates from automobile warranty data, *Lifetime Data Analysis*, **1**, 227-240.
- LeCam, L., Mahan,C., Singh, A.** (1983). An extension of a Theorem of H.Chernoff and E.L.Lehmann. In : *Recent advances in statistics*, Academic Press, Orlando, 303-332.
- Lehmann E.H.** (1973). *On two modification of the Cramer-von Mises statistic*. *Journal of the Royal Statist.Soc.,Ser*, **35**, p.523.
- Lin, D.Y., and Ying, Z.** (1996). " Semiparametric analysis of the general additive-multiplicative hazard models for counting processes", *Ann. Statist.*, **23**, p. 1712-1734.
- Lin, D.Y., Geyer, C.J.** (1992). Computational methods for semiparametric linear regression with censored data. *Journal Comput. and Graph. Statist.*, **1**, 77-90.
- Lu, C.J.** (1995). "Degradation processes and related reliability models", Ph.D. thesis, McGill University, Montreal, Canada.

- Lu, C.J., Meeker, W.Q.** (1993). "Using degradation Measures to Estimate a Time-to-Failure Distribution", *Technometrics*, **35**, 161-174.
- Lu C. J., Meeker W.Q., Escobar L.A.** (1996). A comparison of degradation and failure-time analysis methods for estimating a time-to-failure distribution. *Statist. Sinica*, **6**, 3, 531-546.
- Mann, N.R., Schafer, R.E. and Singpurwalla, N.** (1974) *Methods for Statistical Analysis of Reliability and Life Data*. New York : John Wiley and Sons.
- Mann H.B. and Whitney D.R.** (1947). *Annals of Mathematical Statistics*, **v.18**, p.50-60.
- Mardia K.V. and Zemroch P.J.** (1978). *Tables of the F- and related distributions with algorithms*. Academic Press.
- McKeague, I.W., Sasieni, P.D.** (1994). A partly parametric additive risk model. *Biometrika*, **81**, #3, 501-514.
- Meinhold R.J. and Singpurwalla N.D.** (1987) *A Kalman-Filter Smoothing Approach for Extrapolation in Certain Dose - Response. Damage Assessment. and Accelerated-Life-Testing Studies*. *The American Statistician*, **41**, p.101-106.
- Margolis S. B.** (1979). An analytical solution for the multidimensional degradation of a packed bed thermocline. *J. Franklin Inst.*, **307**, #1, 39-58.
- Meeker, W.Q., Escobar, L.A., Lu, C.J.** (1998). "Accelerated Degradation Tests : Modeling and Analysis", *Technometrics*, **40**, 89-99.
- Meeker, W.Q., Escobar, L.A.** (1998). *Statistical Methods for Reliability Data*, John Wiley and Sons : New York.
- Mine H., Kawai H.** (1976). Marginal checking of a Markovian degradation unit when checking interval is probabilistic. *J. Operations Res. Soc. Japan*, **19**, 2, 158-173.
- Mitsuo, F.** (1991). *Reliability and Degradation of Semiconductor Lasers and LEDs*", Artech House : Norwood.
- Meeker, W.Q. and Escobar, L.** (1998). *Statistical Analysis for Reliability Data*, John Wiley and Sons, New York.
- Miller L.** (1956). *Table of percentage points of Kolmogorov statistics*. *JASA*, **51**, p.111.
- Mises R. von** (1931). *Wahrscheinlichkeit, Statistik und Wahrheit*. Springer-Verlag.
- Molenaar W.** (1970). *Approximations to the Poisson, Binomial and Hypergeometric Distribution Functions* Amsterdam, Mathematical centre tracts, **31**.
- Moore D. and Spruill M.** (1975). Unified large-sample theory of general chi-squared statistics for tests of fit, *Ann. Statist.*, **3**, 599-616.
- S.A. Murphy.** (1995). "Asymptotic theory for the frailty model", *Annals of Statist.*, vol. 23 pp. 182-198.
- S.A. Murphy, A.J. Rossini and A.W. van der Vaart.** (1997). Maximum likelihood estimation in the proportional odds model, *JASA.*, **92**, p. 968-976.
- Nelson, W.** (1990). *Accelerated Testing : Statistical Models, Test Plans, and Data Analysis*, John Wiley and Sons : New York.
- Nikulin M.S.** (1973) Chi-square test for continuous distributions with shift and scale parameters. *Theory of probability and its applications*, **18**, p.559-568.
- Nikulin M.S.** (1973). On a chi-square test for continuous distributions. *Theory of probability and its applications*, **18**, p.638-639.
- Nikulin M.S.** (1979). Hypothesis testing for a parameter difference in binomial distributions. *Theory of probability and its applications*, **v.24**, #2, p.392-396.

**Nikulin M.S.** (1984). *F*-distributions and its relations with others distributions. In : **Mardia K.V. and Zemroch P.J.** *Tables of the F- and related distributions with algorithms*. Academic Press. Moscow, Nauka (in russian).

**Nikulin M.S.** (1991). *Some recent results on chi-squared tests*. Queen's papers in pure and applied mathematics, **86**, Queen's University, Kingston, Canada, 74 p.

**Nikulin M.S., Nacerra Seddik-Ameur** (1991). Analyse statistique des données binomiales. Seminaire 90-91, Université Paris 5, p.87-110.

**Nikulin M.S.** (1992). Gihman statistic and goodness-of-fit tests for grouped data. *C.R. Math. Rep. Acad. Sci. Canada*, **14**, #4, p.151-156.

**M.Nikulin and V.Solev.** (1999). Chi-squared goodness of fit test for doubly censored data, applied in Survival Analysis and Reliability, In : *Probabilistic and Statistical Models in Reliability*, (Eds. N. Limnios and D. Ionescu), Boston : Birkhauser, 101-112.

**M.Nikulin, M.Novak, D.Turetaev, V.Voinov.** (2000). Estimating Environmental Radioactive Contamination in Kazakhstan, *Central Asian Journal of Economics, Management and Social Research*, # 1, 59-71. (ISBN 9965-9047-3-1)

**Nikulin, M., Pya, N., Voinov, V.** (2003). *Chi-squared goodness-of-fit tests for the family of logistic distributions*. Preprinte "Statistique Mathématique et ses Applications, Université Victor Segalen Bordeaux 2, France.

**Oliver F.R.** (1964). Methods of estimating the logistic growth function. *Appl. Statist.*, **13**, p.57-66.

**Olson W.H.** (1977). Non-uniform breakage-mechanism branching processes and degradation of long-chain polymers. *J. Appl. Probability*, **14**, 1, 1-13.

**E.Parner.** (1998). Asymptotic theory for the correlated gamma-frailty model, *Ann. Statist.*, **26**, p. 183-214.

**Patnaik P.B.** (1949) . The non-central  $\chi^2$  and *F* distributins and their applications. *Biometrika*, **36**, p.202-232.

**Pearson E.S.** (1959). Note on an approximation to the distribution of non-central  $\chi^2$ . *Biometrika*, **46**, p.364.

**Pearson E.S. and Hartley H.O.** (1966). *Biometrika tables for statisticians*, **1**. Cambridge University Press.

**Pearson E.S. and Hartley H.O.** (1972). *Biometrika tables for statisticians*, **2**. Cambridge University Press.

**Pearson K.** (1934). *Tables of the incomplete  $\Gamma$ -fonction*. Cambridge University Press.

**Pearson K.** (1968). *Tables of the incomplete Beta-function*. Cambridge University Press.

**Pearl R., Reed L.J.** (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. of National Acad. Sci.*, **6**, p.275-288.

**Pearlman W.A.** (1976). A limit on optimum performance degradation in fixed-rate coding of the discrete Fourier transform. *IEEE Trans. Information Theory*, **IT-22**, 4, 485-488.

**Pinçon, C.** (2003) *Estimateurs non-paramétriques et semi-paramétriques efficaces dans l'analyse des données censurées multivariées*, Thèse de l'Université Paris XI, Faculté de MEDECINE PARIS-SUD.

**Prékopa A.** (1954). Statistical treatment of the degradation process of long chain polymers. *Magyar Tud. Akad. Alkalm. Mat. Int. Kozl.*, **2**, 103-123 .

**Pettit L. I., Young K. D. S.** (1999). Bayesian analysis for inverse Gaussian lifetime data with measures of degradation. *J. Statist. Comput. Simulation*, **63**, 3, 217-234.

**Redinbo G.R.** (1979). Optimum soft decision decoding with graceful degradation. *Inform. and Control*, **41**, #2, 165-185.

- Rao C.R.** (1965) *Linear Statistical Inference and its application*. New York : J.Wiley and Sons.
- Rao K.C., Robson D.S.** (1974). A chi-squared statistic for goodness-of-fit tests within the exponential distribution, *Commun. Statist.*, **3**, 1139-1153.
- Reed L.J., Berkson J.** (1929). The application of the logistic function to the experimental data. *Journal Physical Chemistry*, **33**, p.760-779.
- Sedyakin, N.M.** (1966). On one physical principle in reliability theory.(in russian). *Techn. Cybernetics*, **3**,80-87.
- Singpurwalla, N.D.**(1995). Survival in Dynamic Environments. *Statistical Science*,**1**,**10**, 86-103.
- Singpurwalla, N.D.**(1997). Gamma processes and their generalizations : an overview. In *Engineering Probabilistic Design and Maintenance for Flood Protection*, (R.Cook, M.Mendel and H.Vrijling, eds.) Kluwer Acad.Publishers, 67-73.
- Singpurwalla, N.D., Youngren, M.A.**(1998). Multivariate distributions induced by dynamic environments, *Scandinavian Journal of Statistics*, **20**, 251-261.
- Schiffer, M.** (1993). Quantum fog and the degradation of information by the gravitational field. *Gen. Relativity Gravitation*, **25**, # 7, 721-752.
- Srinivasan S. K., Mehata K. M.** (1972). A stochastic model for polymer degradation. *J. Appl. Probability*, **9**,43-53.
- Suzuki, K., Maki, K., Yokogawa, S.** (1993). An analysis of degradation data of a carbon film and properties of the estimators. In : *Statistical Sciences and Data Analysis*, (Eds. K.Matusita, M.Puri, T.Hayakawa), Utrecht, Netherlands :VSP.
- Smirnov N.V.** (1936). Sur la distribution de  $\omega^2$ . [*C.R.Acad.Sci. de Paris*, **202**, p.449-452.
- Smirnov N.V.** (1939). On estimating the discrepancy between empirical distribution functions in two independent samples. *The Bulletin of the Moscow's Gos.University, ser.A*, **2**, p.3-14.
- Smirnov N.V.** (1937). On the distribution of Mises  $\omega^2$ -test. *Math.Sbornik*, **2**, p.973-994.
- Smirnov N.V.** (1939). On deviation of the empirical distribution function. *Math. Sbornik*, **6**, p.3-26.
- Smirnov N.V.** (1944). Approximate distribution laws for random variables, constructed from empirical data. *Uspekhi Math.Nauk*, **10**, p.197-206.
- Stablein, D. M., Koutrouvelis, I. A.** (1985). A two sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics* **41**, 643-652.
- Thompson C.M.** (1963). Tables of percentage points of the  $\chi^2$  -distribution. *Biometrika*, **32**, p.187-191.
- Thompson C.M.** (1941). Tables of percentage of the incomplete Beta-function. *Biometrika*, **32**, p.151-181.
- Thompson W.R.** (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Annals of mathematical statistics*, **v.6**, p.214-219.
- Tseng, S.T., Hamada, M.S. and Chiao, C.H.**(1994). Using degradation data from a fractional experiment to improve fluorescent lamp reliability. *Research Report RR-94-05. The Institute for Improvement in Quality and Productivity, University of Waterloo, Waterloo, Ontario, Canada.*
- A.A.Tsiatis.** (1981). A large sample study of Cox's regression model, *Ann. Statist.*, **9**, p. 93-108.
- Tumanian S.Kh.** (1956). Asymptotic distribution of  $\chi^2$  criterion when the number of

observations and classes increase simultaneously. *Theory of Probability and its Applications*, **1**, #1, p.131-145.

**Turnbull B.W.** (1974). *Non parametric estimation of survivorship function with doubly censored data*. *JASA*, **69**, 169-173.

**Turnbull B.W.** (1976). *The empirical distribution function with arbitrarily grouped, censored, and truncated Data*. *Royal Statist. Soc. B* **38**, p.290-295.

**J.W.Vaupel, K.G.Manton and E.Stallard.** (1979). *The impact of heterogeneity in individual frailty on the dynamics of mortality*, *Demography*, **16**, p. 439-454.

**Van der Vaart, A. W.** (2000). *Asymptotic Statistics*. Cambridge : UK.

**Verdooren L.R.** (1963). *Extended tables of critical values for Wilcoxon's test statistic*. *Biometrika*, **v.50**, p.177-186.

**Voinov V.G. and Nikulin M.S.** (1993). *Unbiased estimators and their applications*, **v.1** Univariate case, Dordrecht : Kluwer Academic Publishers.

**Voinov V.G. and Nikulin M.S.** (1996). *Unbiased estimators and their applications*, **v.2** Multivariate case, Dordrecht : Kluwer Academic Publishers.

**Woodroffe M.** (1985). *Estimating a distribution function with truncated data*. *Ann. Statist.*, **13**, p.163-177.

**Wulfsohn, M. and Tsiatis, A.** (1997). *A Joint Model for Survival and Longitudinal Data Measured with Error*. *Biometrics*, **53**, 330-339.

**Whitmore,G.A.**(1995). *Estimating Degradation By a Wiener Diffusion Process Subject to Measurement Error*. *Lifetime Data Analysis*, **1**, 307-319.

**Whitmore, G.A., Schenkelberg,F.**(1997). *Modelling Accelerated Degradation data Using Wiener Diffusion With a Time Scale Transformation*, *Lifetime Data Analysis*, **3**, 27-45.

**Whitmore, G.A., Crowder,M.I. and Lawless, J.F.**(1998). *Failure inference from a marker process based on bivariate model*, *Lifetime Data Analysis*, **4**, 229-251.

**Wu S.-J., Shao J.** (1999). *Reliability analysis using the least squares method in nonlinear mixed-effect degradation models*. *Statist. Sinica*, **9**, # 3, 855–877.

**Yanagisava, T.** (1997). *Estimation of the degradation of amorphous silicon cells*, *Microelectronics and Reliability*, **37**, 549-554.

**Yu H.-F., Tseng S.-T.** (1999). *Designing a degradation experiment*. *Naval Res. Logist.*, **46**, #6, 689-706.

**Zeleny M.** (1995). *The ideal-degradation procedure : searching for vector equilibria*. *Advances in multicriteria analysis*, 117-127, *Nonconvex Optim. Appl.*, **5**, Kluwer Acad. Publ., Dordrecht.

**Zacks S.** (1971) *The theory of statistical inference*. New York : Wiley and Sons.

**Zerbet A.** (2001) *Statistical tests for normal family in the presence of outlying observations*. In : *Goodness-of-fit tests and Validity of Models* (Eds. C.Huber, N.Balakrishnan, M.Nikulin, M.Mesbah), Boston : Birkhauser.

**Zerbet, A., Nikulin, M.** (2003). *A new statistics for detecting outliers in exponential case*, *Communications in Statistics : Theory and Methods*,**32**, 573-584.

**Zhang B.** (1999) *A chi-squared goodness-of-fit test for logistic regression models based on case-control data*, *Biometrika*, **86**, #3, 531-539.

**Zdorova-Cheminade, O.** (2003) *Modélisation du processus d'évolution de l'incapacité chez les personnes âgées*, *Mémoire de DEA "Epidémiologie et Intervention en Santé Publique"*, Université Bordeaux 2, Juin 2003, 77 pages.