INTRODUCTION AUX PROBABILITES

Fabien Campillo
ESSI
1990–91

Table des matières

1 Probabilités, indépendance et variables aléatoires **1**
1.1 Probabilités . . . 1
1.2 Probabilités conditionnelles . . . 1
1.3 Evénements indépendants . . . 2
1.4 Variables aléatoires . . . 5
1.5 Variables aléatoires indépendantes . . . 6
1.6 Moments des variables aléatoires . . . 8

2 ESTIMATION PARAMETRIQUE **10**
2.1 CONCEPTS GENERAUX . . . 10
2.1.1 Modèles Statistiques . . . 10
2.1.2 Estimateur . . . 11
2.2 ESTIMATEURS PARAMETRIQUES . . . 11
2.2.1 Biais et Risque . . . 11
2.2.2 Estimateur du Maximum de Vraisemblance . . . 12
2.2.3 Estimateurs Bayésien et "MAP" . . . 15
2.3 THEORIE ASYMPTOTIQUE . . . 17
2.4 INTERVALLES DE CONFIANCE . . . 18
2.4.1 Cas des Grands Echantillons . . . 18
2.4.2 Cas Des Echantillons Moyens . . . 20
2.4.3 Cas Général . . . 20
2.5 LE MODELE LINEAIRE . . . 21
2.5.1 L'Estimateur des Moindres–Carrés . . . 21
2.5.2 Estimation de la Variance . . . 23
2.5.3 Le Modèle Linéaire Gaussien . . . 23
2.6 EXERCICES . . . 24

3 TEST D'HYPOTHESES **27**
3.1 CONCEPTS GENERAUX . . . 27
3.2 TEST DE 2 HYPOTHESES SIMPLES . . . 29
3.3 TEST DU RAPPORT DE VRAISEMBLANCE (TRV) . . . 32
3.4 APPLICATIONS DU TRV AU CAS GAUSSIEN . . . 33
3.4.1 Tests Portant sur L'Espérance . . . 33
3.4.2 Tests Portant sur la Variance . . . 35
3.4.3 Tests D'Indépendance de Deux Echantillons . . . 36
3.5 EXERCICES . . . 37

4 METHODES NON–PARAMETRIQUES **39**
4.1 ESTIMATION NON–PARAMETRIQUE . . . 39
4.2 TESTS NON–PARAMETRIQUES . . . 41
4.2.1 Test D'Adéquation . . . 41
4.2.2 Comparaison de Deux Echantillons . . . 43

4.2.3 Test D'Indépendance du χ^2 45

A Résumé de théorie de la mesure **47**
A.1 Espace mesurable 47
A.2 Mesure 47
A.3 Applications mesurables 48
A.4 Intégrale 49
A.4.1 Intégrale de fonctions mesurables positives 49
A.4.2 Intégrable de fonctios mesurables réelles 50
A.4.3 Espaces $L^p(E, \mathcal{A}, m)$ 50
A.5 Exemples 51
A.5.1 Mesures définies par des densités 51
A.5.2 Mesures images 51
A.5.3 Mesures produits 51

B EXERCICES CORRIGES ET PROBLEMES **52**
B.1 Correction des Exercices 52
B.2 Autres Exercices 56
B.3 Problèmes 63

Index **78**

NOTATIONS

- A^c complémentaire de A.
- fonction indicatrice: $\mathbf{1}_A(x) = 1$ si $x \in A$, 0 sinon.
- I_n matrice identité de dimension n.
- $\nabla_\theta f(x,\theta) = (\partial f(x,\theta)/\partial\theta_1 \cdots \partial f(x,\theta)/\partial\theta_n)^*$ gradient de $f(x,\theta)$ par rapport à θ.
- $p_X(x)$ densité de la loi de la variable aléatoire X.
- $p_{X|Y}(x|y)$ ou $p_{X|Y=y}(x)$ densité conditionnelle (de la loi) de X sachant que $Y = y$.
- "iid" indépendant(e)s et indentiquement distribué(e)s.

Chapitre 1

Probabilités, indépendance et variables aléatoires

1.1 Probabilités

Soit Ω une ensemble donné,

Définition 1.1.1 *Une famille $\mathcal{F}$ de sous ensembles de Ω est appelée* tribu, *ou σ–algèbre, si*

(i) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

(ii) $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$,

(iii) $\Omega \in \mathcal{F}$.

Les éléments de $\mathcal{F}$ sont appelés des évènements. *Le doublet $(\Omega, \mathcal{F})$ est appelé* espace probabilisable.

Soit $(\Omega, \mathcal{F})$ un espace probabilisable,

Définition 1.1.2 *On appelle* probabilité *(ou mesure de probabilité) une application $P : \mathcal{F} \to [0, 1]$ qui vérifie*

(i) $P(\Omega) = 1$,

(ii) Pour tout $\{A_n; n \geq 1\} \subset \mathcal{F}$, tel que $A_k \cap A_l = \emptyset$ si $k \neq l$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \ .$$

Un triplet $(\Omega, \mathcal{F}, P)$, où P est une mesure de probabilité, est appelé un espace de probabilité.

Dans toute la suite de ce chapitre, un tel espace de probabilité $(\Omega, \mathcal{F}, P)$ sera supposé donné.

Remarque 1.1.3 *Le parallèle avec la théorie de la mesure est le suivant : un espace probabilisable est un espace mesurable, une probabilité est une mesure positive μ telle que $\mu(\Omega) = 1$, c'est à dire de masse totale égale à 1.*

1.2 Probabilités conditionnelles

Définition 1.2.1 *Soient A, $B \in \mathcal{F}$, avec $P(A) > 0$. La probabilité de l'évènement B,* conditionné *par A (ou la probabilité conditionnelle de A sachant B) est donnée par*

$$P(B|A) \triangleq \frac{P(A \cap B)}{P(A)} \ . \tag{1.1}$$

A étant fixé tel que $P(A) > 0$, on vérifie aisément que Q définie par $Q(B) = P(B|A)$ est une mesure de probabilité.

L'égalité (1.1) peut s'écrire

$$P(A \cap B) = P(A)\, P(B|A) \ .$$

Cette dernière égalité est vraie même losrque $P(A) = 0$, à condition de convenir que $0 \times$ *terme indéterminé* $= 0$.

Proposition 1.2.2 *Soient $A_k \in \mathcal{F}$, $k = 1, \ldots, n$, alors*

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) \quad &= \quad P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2) \cdots \\ &\qquad \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_n) \ . \end{aligned} \tag{1.2}$$

Preuve Deux cas possibles

(*i*) $P(A_1 \cap A_2 \cap \cdots \cap A_n) = 0$,

(*ii*) $P(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$.

Cas 1 Si $P(A_1) = 0$, (1.2) est satisfaite. Si $P(A_1) > 0$, il existe $m \in \{2, \ldots, n\}$ tel que

$$0 = P(A_1 \cap \cdots \cap A_m) < P(A_1 \cap \cdots \cap A_{m-1}) \ .$$

D'où $P(A_m|A_1 \cap \cdots \cap A_{m-1}) = 0$, et (1.2) est satisfaite.

Cas 2 On démontre (1.2) par récurrence. Elle est vraie pour $m = 2$. Supposons que l'on ait

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_n) \ .$$

Alors en multipliant cette dernière égalité par

$$\frac{P(A_1 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})} = P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

on obtient le résultat. □

1.3 Evénements indépendants

Définition 1.3.1 *Deux événements A et $B \in \mathcal{F}$ sont dits* indépendants *si*

$$P(A \cap B) = P(A)\, P(B) \ .$$

Lorsque $P(A)\, P(B) \neq 0$, A et B sont indépendants si et seulement si $P(B) = P(B|A)$, ou $P(A) = P(A|B)$; c'est à dire A et B sont indépendants si et seulement si la probabilité de B n'est pas modifiée par l'information "A est réalisée".

Exemple 1.3.2 *Sur l'espace de probabilité produit*

$$(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \times P_2)$$

on considère les événements

$$A = A_1 \times \Omega_2 \ \textit{et} \ B = \Omega_1 \times B_2$$

où $A_1 \in \mathcal{F}_1$ et $B_2 \in \mathcal{F}_2$. A et B sont des événements indépendants, en effet $A \cap B = A_1 \times B_2$ donc

$$P(A \cap B) = P_1(A_1)\, P_2(B_2) = P(A)\, P(B) \ .$$

Exercice 1.3.3 *Soient A et B deux événements indépendants.*

(*i*) *Montrer que* A^c *et* B^c *[resp.* A^c *et* B*,* A *et* B^c*] sont indépendants.*

(*ii*) *Etablir la relation* $P(A \cup B) = P(A) + P(B) - P(A)\,P(B)$.

Définition 1.3.4 *Une suite* $\{A_1, \ldots, A_n\} \subset \mathcal{F}$ *est dite* indépendante *si*

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})\,P(A_{i_2}) \cdots P(A_{i_k})$$

pour toute suite d'indices deux à deux distincts $i_1, \ldots, i_k$ *compris entre 1 et* n.

Remarque 1.3.5 *Le fait que les événements* A_i *soient deux à deux indépendants n'entraîne pas que la suite* $\{A_1, \ldots, A_n\}$ *soit indépendante (cf. exemple suivant).*

Exemple 1.3.6 *On joue trois fois à "Pile ou Face" avec une pièce non truquée. On pose*

$$X_i = \begin{cases} 1 & \text{si on obtient Pile au } i\text{ème coup,} \\ 0 & \text{si on obtient Face au } i\text{ème coup.} \end{cases}$$

On suppose la suite $\{X_1 = 1\}$, $\{X_2 = 1\}$, $\{X_3 = 1\}$ *indépendante, chacun de ces événement étant de probabilité* $\frac{1}{2}$.

On pose $A_{ij} = \{X_i = X_j\}$. *Les événements* A_{12}, A_{23}, A_{31} *sont deux à deux indépendants, mais la suite ne l'est pas.*

Définition 1.3.7 *Le suite* $\mathcal{C}_1, \ldots, \mathcal{C}_n$ *de sous classes de* $\mathcal{F}$ *est dite* indépendante *si*

$$P(B_1 \cap B_2 \cap \cdots B_n) = P(B_1)\,P(B_2) \cdots P(B_n)$$

pour tout $B_i \in \mathcal{C}_i \cap \{\Omega\}$, $i = 1, cdotsn$.

Proposition 1.3.8 *Si la suite* $\{A_1, \ldots, A_n\}$ *est indépendante, les* 2^n *suites* $\{A'_1, \ldots, A'_n\}$ *obtenues en choisissant* $A'_i = A_i$ *ou* A_i^c *sont encore indépendantes.*

Preuve Il suffit d'établir que l'indépendance de la suite $\{A_1, \ldots, A_n\}$ entraine celle de

$$\{A_1, \ldots, A_{l-1}, A_l^c, A_{l+1}, \ldots, A_n\}$$

pour tout $l \in \{1, \ldots, n\}$ et d'itérer ce résultat. La notion d'indépendance d'une suite ne faisant pas intervenir l'ordre, il suffit d'établir le résultat pour $l = 1$. Pour établir l'indépendance de $\{A_1^c, A_2, \ldots, A_n\}$, il suffit d'etablir que pour tout $i_1, \ldots, i_k$ indices deux à deux distinct dans $\{2, \ldots, n\}$,

$$P(A_1^c \cap A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_1^c)\,P(A_{i_1}) \cdots P(A_{i_k})$$

mais le membre de gauche est égal à

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) - P(A_1 \cap A_{i_1} \cap \cdots \cap A_{i_k}) = (1 - P(A1))\,P(A_{i_1}) \cdots P(A_{i_k})\ .$$

□

Etant donnée une suite $\{A_1, \ldots, A_n\}$, définissons les tribus $\mathcal{F}_i = \{\emptyset, A_i, A_i^c, \Omega\}$. De la proposition (1.3.8) on déduit aisément le

Corollaire 1.3.9 *Pour que la suite* $\{A_1, \ldots, A_n\}$ *soit indépendante, il faut et il suffit que l'égalité*

$$P(B_1 \cap \cdots \cap B_n) = P(B_1) \cdots P(B_n)$$

soit satisfaite, pour tout $B_i \in \mathcal{F}_i$ $(i = 1 \ldots n)$.

Définition 1.3.10 *Une suite infinie d'événements* $\{A_n, n \in \mathbb{N}\}$ *est dite* indépendante *si toutes les sous–suites de* $\{A_n, n \in \mathbb{N}\}$ *sont indépendantes, c'est à dire, si pour tout* k *et* $i_1, \ldots, i_k$ *entiers deux à deux distincts,*

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})\ .$$

Lemme 1.3.11 (Borel–Cantelli)

(i) Pour toute suite infinie $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$,

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 0 \ .$$

(ii) Si la suite $\{A_n, n \in \mathbb{N}\}$ *est indépendante*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 1 \ .$$

Remarque 1.3.12

(*i*) *Dans le cas d'une suite indépendante, on a en particulier une loi 0–1*

$$P(\limsup_{n\to\infty} A_n) = 0 \ \textit{ou} \ 1 \ .$$

(*ii*) *(ii) est faux sans l'hypothèse d'indépendance, comme le montre l'exemple suivant :* $A_n \equiv A$, $0 < P(A) < 1$, $\limsup A_n = A$.

Preuve

(i)

$$\begin{aligned} P(\limsup_{n\to\infty} A_n) &= P\left(\lim_{m\to\infty} \downarrow \left(\bigcup_{n=m}^{\infty} A_n\right)\right) \\ &= \lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} A_n\right) \\ &\leq \lim_{m\to\infty} \sum_{n=m}^{\infty} P(A_n) = 0 \ , \end{aligned}$$

puisque par hypothèse la série $\sum P(A_n)$ converge.

(ii)

$$P\left(\bigcup_{n=m}^{\infty} A_n\right) = 1 - P\left(\bigcap_{n=m}^{\infty} A_n^c\right) \ .$$

Il suffit donc de montrer quer $P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0$, pour tout m. D'après l'indépendance,

$$P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = \prod_{n=m}^{\infty} (1 - P(A_n)) \ .$$

En posant $\log 0 = -\infty$, et en utilisant l'inégalité $\log(1+x) \leq x$, $\forall x \geq -1$, et l'hypothèse, on obtient

$$\log\left[\prod_{n=m}^{\infty} (1 - P(A_n))\right] = \sum_{n=m}^{\infty} \log(1 - P(A_n)) \leq -\sum_{n=m}^{\infty} P(A_n) = -\infty \ .$$

□

1.4 Variables aléatoires

Définition 1.4.1 *On appelle* variable aléatoire réelle *une application mesurable*

$$X \,:\, (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \ .$$

On appelle vecteur aléatoire *(de dimension d) une application mesurable*

$$X \,:\, (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}_d) \ .$$

On emploiera souvent "variable aléatoire" (en abrégé v.a.) pour désigner indifféremment une variable aléatoire réelle (en abrégé v.a.r.) ou un vecteur aléatoire.

Remarque 1.4.2 $X \,:\, (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ *est une application mesurable si*

$$\forall B \in \mathcal{B} \quad X^{-1}(B) \in \mathcal{F} \ .$$

Pour vérifier que X est mesurable il suffit n fait de démontrer que

$$\forall x \in \mathbb{R}, \quad \{\omega \in \Omega \,;\, X(\omega) < x\} \in \mathcal{F} \ .$$

Définition 1.4.3 *On appelle* tribu naturelle *du vecteur aléatoire X (de dimension d) la tribu*

$$\sigma(X) \triangleq X^{-1}(\mathcal{B}_d) = \left\{X^{-1}(B) \,;\, B \in \mathcal{B}_d\right\} \ .$$

Si $\mathbb{P}$ est un ensemble de parties de Ω, on note $\sigma(\mathbb{P})$ la plus petite tribu sur Ω qui contient $\mathbb{P}$, c'est la tribu *engendrée* par $\mathbb{P}$. Si $\mathcal{G}$ et $\mathcal{H}$ sont deux tribus de parties de Ω, on note

$$\mathcal{G} \vee \mathcal{H} = \sigma\left(\mathcal{G} \cup \mathcal{H}\right)$$

la plus petite tribu qui contient $\mathcal{G}$ et $\mathcal{H}$ ($\mathcal{G} \cup \mathcal{H}$ n'est pas nécessairement une tribu). De même

$$\mathcal{G}_1 \vee \mathcal{G}_2 \vee \cdots \vee \mathcal{G}_n \triangleq \sigma\left(\mathcal{G}_1 \cup \mathcal{G}_2 \cup \cdots \cup \mathcal{G}_n\right) \ .$$

Proposition 1.4.4 *Soit $\{X_i, i = 1 \cdots n\}$ des vecteurs aléatoires de dimensions respectives $d_1, \ldots, d_n$. On pose*

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \ .$$

Alors $\sigma(X) = \sigma(X_1) \vee \cdots \vee \sigma(X_n)$.

Preuve Dans le cas $n = 2$, $d_1 = d_2 = 1$.

$$\begin{aligned} \sigma(X_1) &= X_1^{-1}(\mathcal{B}_1) = X^{-1}\left(\{A_1 \times \mathbb{R}; A_1 \in \mathcal{B}_1\}\right) \ , \\ \sigma(X_2) &= X_2^{-1}(\mathcal{B}_1) = X^{-1}\left(\{\mathbb{R} \times A_2; A_2 \in \mathcal{B}_1\}\right) \ , \end{aligned}$$

donc $\sigma(X_1) \vee \sigma(X_2) = \sigma(X^{-1}(\mathcal{C}))$ où

$$\mathcal{C} = \{A_1 \times \mathbb{R}; A_1 \in \mathcal{B}_1\} \cup \{\mathbb{R} \times A_1; A_2 \in \mathcal{B}_1\} \ .$$

Or $\sigma(\mathcal{C}) = \mathcal{B}_2$, et $\sigma(X^{-1}(\mathcal{C})) = X^{-1}(\sigma(X)) = \sigma(X)$. □

Définition 1.4.5 *Etant donné un vecteur aléatoire X de dimension d, on appelle* loi de probabilité *de X (ou simplement loi de X), la mesure de probabilité notée P_X sur $(\mathbb{R}^d, \mathcal{B}_d)$ définie par*

$$P_X = P \circ X^{-1} \ .$$

Etant donnée une mesure de probabilité Q sur $(\mathbb{R}^d, \mathcal{B}_d)$, il existe un espace de probabilité et un vecteur aléatoire défini sur cet espace, de dimension d et de loi Q : $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}_d$, $P = Q$, $X(\omega) = \omega$.

Définition 1.4.6 *Soit X une variable aléatoire réelle [resp. un vecteur aléatoire de dimension d]. On appelle* fonction de répartition *de X la fonction*

$$F \,:\, \mathbb{R}\,[\textit{resp. } \mathbb{R}^d] \to [0,1]$$

définie par

$$F(x) = P(X < x) = P_X(]-\infty, x[)$$

[resp. $F(x_1, \ldots, x_d) = P(X_1 < x_1, \ldots, X_d < x_d) = P_X(]-\infty, x_1[\times \cdots \times]-\infty, x_d[)$].

Proposition 1.4.7 *Soit F une fonction de répartition réelle*

(i) F est croissante,

(ii) F est continue à gauche : $x_n \uparrow x \Rightarrow F(x_n) \to F(x)$,

(iii) $F(x) \to 0$ si $x \to -\infty$, $F(x) \to 1$ si $x \to +\infty$.

Soit $a, b \in \mathbb{R}^d$, avec $a < b$. Soit A le pavé $\prod_{i=1}^d [a_i, b_i[$. Ce pavé a 2^d sommets qui sont les $x = (x_1, \ldots, x_d)$, où chaque x_i est égal soit à a_i soit à b_i. On pose

$$\Delta_A F = \sum_{\text{sommet de } A} \text{signe}_A(x)\, F(x)$$

où $\text{signe}_A(x) = \pm 1$ suivant que le nombre de x_i égaux à a_i est pair ou impair.

Proposition 1.4.8 *Soit F une fonction de répartition sur $\mathbb{R}^d$*

(i) pour tout $A = \prod_{i=1}^d [a_i, b_i[$, $a < b$: $\Delta_A F \geq 0$,

(ii) si $x_i^{(k)} \uparrow x_i$ $(\forall i)$ alors $F(x^{(k)}) \to F(x)$,

(iii) $F(x) \to 0$ si $\exists i$ tel que $x_i \to -\infty$, $F(x) \to 1$ si $\forall i$, $x_i \to +\infty$.

Proposition 1.4.9 *Soit $F : \mathbb{R}^d \to \mathbb{R}$ qui vérifie (i), (ii) et (iii) de la proposition 1.4.8. Alors il existe une et une seule mesure de probabilité P sur $(\mathbb{R}^d, \mathcal{B}_d)$ dont F est la fonction de répartition.*

1.5 Variables aléatoires indépendantes

Définition 1.5.1 *Le suite des vecteurs aléatoires $X_1, \ldots, X_n$, de dimensions respectives $d_1, \ldots, d_n$, est dite* indépendante *si la suite de tribus $\sigma(X_1), \ldots, \sigma(X_n)$ est indépendante.*

Nous pouvons relier cette notion d'indépendance à celle du paragraphe 1.3.

Proposition 1.5.2 *Soit $(A_1, \ldots, A_n) \subset \mathcal{F}$. Pour tout $i \leq n$, on note*

$$\mathcal{F}_i = \{\emptyset, A_i, A_i^c, \Omega\}\ .$$

Les trois conditions suivantes sont équivalentes

(i) la suite d'événements $(A_1, \ldots, A_n)$ est indépendante (au sens de la définition 1.3.4),

(ii) la suite des tribus $(\mathcal{F}_1, \ldots, \mathcal{F}_n)$ est indépendante (au sens de la définition 1.3.7),

(iii) la suite de variables aléatoires réelle $(\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n})$ est indépendante (au sens de la définition 1.5.1).

Preuve L'équivalence $(i) \Leftrightarrow (ii)$ a été établie au corollaire 1.3.9. $(ii) \Leftrightarrow (iii)$ est immédiate car $\mathcal{F}_i = \sigma(\mathbf{1}_{A_i})$. □

Théorème 1.5.3 *Soient $\mathcal{C}_1, \ldots, \mathcal{C}_n$ des π–systèmes inclus dans $\mathcal{F}$. Si la suite $(\mathcal{C}_1, \ldots, \mathcal{C}_n)$ est indépendante, alors la suite de tribus $(\sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_n))$ est aussi indépendante.*

Preuve Soit $A_2, \ldots, A_n$ avec $A_i \in \mathcal{C}_i \cup \{\Omega\}$, et $P(A_2 \cap \cdots \cap A_n) > 0$. Soit $Q : \mathcal{F} \to [0,1]$ défini par

$$Q(A) = \frac{P(A \cap A_2 \cap \cdots \cap A_n)}{P(A_2 \cap \cdots \cap A_n)} \ .$$

Q est une mesure de probabilité, qui coïncide avec P sur $\mathcal{C}_1$, donc sur $\sigma(\mathcal{C}_1)$. L'égalité

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\, P(A_2) \cdots P(A_n)$$

est donc vraie pour tout $A_1 \in \sigma(\mathcal{C}_1)$, $A_2 \in \mathcal{C}_2 \cup \{\Omega\}, \ldots A_n \in \mathcal{C}_n \cup \{\Omega\}$. Par récurrence, on démontre que cette égalité est vraie pour tout $A_i \in \sigma(A_i)$, $i \leq n$. □

Théorème 1.5.4 *Pour que la suite de vecteurs aléatoires $X_1, \ldots, X_n$, de dimension respective $d_1, \ldots, d_n$, soit indépendante, il faut et il suffit que la loi du vecteur aléatoire $X = (X_1, \ldots, X_n)$ sur $\mathbb{R}^d$, $d = \sum_{i=1}^n d_i$, soit le produit des lois des X_i, c'est à dire*

$$P_X(dx_1, \ldots, dx_n) = P_{X_1}(dx_1) \cdots P_{X_d}(dx_n) \ .$$

Preuve Soit $B_1 \in \mathcal{B}_{d_1}, \ldots, B_n \in \mathcal{B}_{d_n}$, si la suite $X_1, \ldots, X_n$ est indépendante

$$\begin{aligned} P_X(B_1 \times \cdots \times B_n) &= P\left((X_1, \ldots, X_n) \in B_1 \times \cdots \times B_n\right) \\ &= P\left(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}\right) \\ &= P(X_1 \in B_1) \cdots P(X_n \in B_n) \\ &= P_{X_1}(B_1) \cdots P_{X_n}(B_n) \ . \end{aligned}$$

Réciproquement, si cette égalité est satisfaite, alors pour tout $B_i \in \mathcal{B}_i$

$$P\left(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}\right) = P(X_1 \in B_1) \cdots P(X_n \in B_n)$$

c'est à dire

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$$

pour tout $A_i \in \sigma(X_i)$. □

Corollaire 1.5.5 *Si la suite de vecteurs aléatoires $X_1, \ldots, X_n$, de dimensions respectives $d_1, \ldots, d_n$, est indépendante, et si chaque P_{X_i} admet une densité f_i par rapport à la mesure de Lebesgue sur $\mathbb{R}^{d_i}$, alors P_X admet la densité*

$$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n)$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, $d = \sum_{i=1}^n d_i$.

Proposition 1.5.6 *Soit $X_1, \ldots, X_n$ une suite de vecteurs aléatoires indépendants, $(X'_1, \ldots, X'_n)$ une permutation de $(X_1, \ldots, X_n)$. Alors les k vecteurs aléatoires*

$$(X'_1, \ldots, X'_{n_1}) \ , \ (X'_{n_1+1}, \ldots, X'_{n_2}) \ , \ldots , \ (X'_{n_2+1}, \ldots, X'_{n_k})$$

(ou $n_k \leq n$) sont indépendants.

Preuve Utiliser le théorème 1.5.4, et l'associativité du produit de mesures. □

Proposition 1.5.7 *Soit $X_1, \ldots, X_n$ une suite de vecteurs aléatoires indépendants de dimensions respectives $d_1, \ldots, d_n$, et $f_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d'}$, $i = 1, cdotsn$, des applications mesurables. Alors $f_1(X_1), \ldots, f_n(X_n)$ est une suite de vecteurs aléatoires indépendants.*

Preuve Il suffit de remarquer que

$$\sigma(f_i(X_i)) \subset \sigma(X_i) \ .$$

□

Proposition 1.5.8 *Une suite infinie de sous classes de $\mathcal{F}$ est indépendante si toutes les sous suites finies sont indépendantes.*

Une suite infinie de vecteurs aléatoires est dite indépendante si toutes les sous suites finies sont indépendantes.

1.6 Moments des variables aléatoires

Définition 1.6.1 *Soit X une variable aléatoire réelle définie sur l'espace de probabilité $(\Omega, \mathcal{F}, P)$. X est dite* quasi–intégrable *si*

$$\int_\Omega X^+ \, dP < \infty \ ou \ \int_\Omega X^- \, dP < \infty \ .$$

Elle est dite intégrable *si*

$$\int_\Omega |X| \, dP = \int_\Omega X^+ \, dP + \int_\Omega X^- \, dP < \infty \ .$$

Définition 1.6.2 *Soit X une variable aléatoire réelle intégrable (ou seulement quasi –intégrable). On définit alors* l'espérance mathématique *de X par*

$$E(X) \triangleq \int_\Omega X \, dP \ .$$

Soit X un vecteur aléatoire de dimension d, $X = (X_1, \ldots, X_n)^*$, tel que chaque X_i $(i = 1 \cdots d)$ soit une variable aléatoire réelle intégrable. On définit alors l'espérance de X, $E(X)$, comme le vecteur de coordonnées $E(X_i)$, $i = 1 \cdots d$.

Proposition 1.6.3 *Soit X un vecteur aléatoire de dimension d, $\varphi : \mathbb{R}^d \to \bar{\mathbb{R}}$ uen application mesurable. Alors $\varphi \circ X$ est P (quasi–)intégrable si et seulement si Φ est P_X (quasi–)intégrable, et dans ce cas*

$$\int_\Omega \varphi \circ X(\omega) \, dP(\omega) = \int_{\mathbb{R}^d} \varphi(x) \, dP_X(x) \ .$$

⚠ **Remarque 1.6.4** *Cette proposition est essentielle pour le calcul pratique des espérances.*

Etant donnée Q une probabilité sur $(\mathbb{R}, \mathcal{B})$, si $\int_\mathbb{R} |x| \, dQ(x) < \infty$, alors on peut définir *l'espérance de la probabilité Q* comme la valeur commune des espérances de variables aléatoires réelles de loi Q, grâce à la proposition 1.6.3.

Proposition 1.6.5 (Inégalité de Jensen) *Soit X un vecteur aléatoire de dimension d, et $\varphi : \mathbb{R}^d \to \mathbb{R}$ une fonction* convexe. *Si X est intégrable, alors $\varphi(X)$ est quasi–intégrable et*

$$\varphi\left[E(X)\right] \leq E\left[\varphi(X)\right] \ .$$

Preuve On va utiliser la propriété suivante des fonctions convexes : $\forall x_o \in \mathbb{R}^d$, $\exists a \in \mathbb{R}^d$ tel que

$$a^* \left(x - x_0\right) + \varphi(x_0) \leq \varphi(x), \ \forall x \in \mathbb{R}^d \ .$$

Donc avec $x_0 = E(X)$ on a : $\exists a \in \mathbb{R}^d$ tel que

$$a^* \left(X - E(X)\right) + \varphi(E(X)) \leq \varphi(x), \ \text{p.s.}$$

en prenant l'espérance à droite et à gauche de cette inégalité on en déduit le résultat. □

L'inégalité de Jensen ne s'étend pas au cas où l'on remplace l'espérance par une intégrale par rapport à une mesure quelconque.

Soit X une variable aléatoire réelle. Si X^2 est intégrable (on dit que X est de carré intégrable), alors X est intégrable (appliquer l'inégalité de Cauchy–Schwartz à $1 \times |X|$), et d'après l'inégalité de Jensen

$$[E(X)]^2 \leq E(X^2) \ .$$

Si X et Y sont de carré intégrable, alors $X+Y$ est de carré intégrable et XY est intégrable (utiliser les inégalités $(x+y)^2 \leq 2\,x^2 + 2\,y^2$, $|x\,y| \leq (x^2+y^2)/2$).

On peut alors poser les

Définition 1.6.6 *Soit X une variable al'eatoire de puissance p–ième intégrable ($p > 0$). On appelle* moment *d'ordre p de X la quantité $E(X^p)$.*

Si X est de carré intégrable, on appelle variance *de X la quantité*

$$Var(X) \triangleq E\left[(X - E(X))^2\right] = E(X^2) - [E(X)]^2 \ .$$

Si X et Y sont deux variables aléatoires réelles de carré intégrable, on appelle covariance *de X et Y la quantité*

$$Cov(X,Y) \triangleq E\,(X - E(X))\,(Y - E(Y))] = E(X\,Y) - E(X)\,E(Y) \ .$$

Définition 1.6.7 *Soit X un vecteur aléatoire de dimension d tel qye chaque X_i ($i = 1 \cdots d$) soit de carré intégrable. On appelle* matrice des moments *d'ordre 2 de X, la matrice $E(X\,X^*)$, et la* matrice de covariance *de X la natrice*

$$Cov(X) \triangleq E\left[(X - E(X))\,(X - E(X))^*\right] = E(X\,X^*) - E(X)\,E(X)^* \ .$$

Ces deux matrices sont auto–adjointes et définies non négatives.

Proposition 1.6.8 (Inégalité de Bienaymé–Tchebicheff) *Soit X une variable aléatoire et $p > 0$. Alors*

$$P(|X| \geq c) \leq \frac{E(|X|^p)}{c^p} \ , \qquad \forall c > 0 \ .$$

Preuve

$$E(|X|^p) \geq \int_{|X| \geq c} |X|^p \, dP \geq c^p \, P(|X| \geq c) \ .$$

□

Proposition 1.6.9 *Soient X et Y deux variables aléatoire réelle intégrables et indépendantes. Alors*

(i) *$X\,Y$ est intégrable,*

(ii) *$E(X\,Y) = E(X)\,E(Y)$.*

Preuve

(i) Supposons tout d'abord que $X \geq 0$ et $Y \geq 0$. On utilise la proposition 1.6.3, le théorème 1.5.4 et le lemme de Fubini

$$\begin{aligned} E(X\,Y) &= \int_{\mathbb{R}^2} x\,y\,dP_{(X,Y)}(x,y) \\ &= \int\int x\,y\,dP_X(x)\,dP_Y(y) \\ &= \int x\,dP_X(x) \times \int y\,dP_Y(y) \\ &= E(X)\,E(Y) \ . \end{aligned}$$

(ii) Cas général : $|X|$ et $|Y|$ sont indépendantes. Donc *(i)* résulte du cas précédent. *(ii)* se démontre alors comme ci–dessus.

□

Chapitre 2

ESTIMATION PARAMETRIQUE

2.1 CONCEPTS GENERAUX

2.1.1 Modèles Statistiques

On observe X le résultat d'un phénomène aléatoire. X est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ défini sur un espace de probabilité $(\Omega, \mathcal{T}, P)$. La loi de X n'est que partiellement connue (voire pas du tout). On suppose que cette loi appartient à une classe donnée $\{P_\theta\,;\,\theta \in \Theta\}$ de lois de probabilités, i.e. il existe $\theta_0 \in \Theta$ tel que $loi(X) = P_{\theta_0}$. Le problème d'estimation est typiquement la détermination de ce paramètre à partir de l'observation (d'une réalisation) de X.

Posons $\mathcal{X} = X(\Omega) \subset \mathbb{R}^n$, $\mathcal{X}$ est l'espace des observations. P_θ est une loi sur $\mathcal{X}$ munit de sa tribu borélienne.

Définition 2.1.1 *Un* modèle statistique *est un triplet* $(\mathcal{X}\,,\,\mathcal{F}\,,\,\{P_\theta; \theta \in \Theta\})$ *où*

$\mathcal{X}$	*est une partie de* $\mathbb{R}^n$ *représentant l'*espace des observations*;*
$\mathcal{F}$	*est la tribu borélienne sur* $\mathcal{X}$*;*
$\{P_\theta\,;\,\theta \in \Theta\}$	*est une famille de lois de probabilité définies sur* $(\mathcal{X}, \mathcal{F})$*.*

Losrqu'il n'existe aucune ambiguïté sur l'espace des observations, on appellera modèle statistique la famille $\{P_\theta\,;\,\theta \in \Theta\}$. Si on note $p(x;\theta)$ la densité de P_θ, la famille de densités $\{p(\,\cdot\,;\theta)\,;\,\theta \in \Theta\}$ sera aussi appelée modèle statistique. On notera E_θ (resp. E) l'espérance associée à P_θ (resp. P).

Remarque 2.1.2 *X est une variable aléatoire définie sur $(\Omega, \mathcal{F}, P)$, à valeurs dans $\mathbb{R}^n$. Donc si I est une partie de $\mathbb{R}^n$ on a*

$$P(X \in I) = P_\theta(I)\ .$$

Pour des raisons de commodité on utilisera souvent la notation suivante

$$P_\theta(X \in I)\ .$$

Cette notation, quoique formelle, présente l'intérêt de faire apparaître que l'on considère "la probabilité que l'observation X soit dans I, sous l'hypothèse que la loi de X est P_θ". De même si φ est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, on a

$$E\,\varphi(X) = E_\theta\,\varphi\ .$$

Dans la suite on utilisera aussi la notation $E_\theta\,\varphi(X)$.

On aura souvent à manipuler des suites $X_1, \ldots, X_n$ de variables indépendantes suivant toutes une même loi

Définition 2.1.3 *On appelle* n–échantillon *d'une loi F_θ, de densité $f(\,\cdot\,;\theta)$, sur un espace $\mathcal{A}$ toute suite $X_1, \ldots, X_n$ de variables aléatoires indépendantes de même loi F_θ.*

On considère un n–échantillon $X_1, \ldots, X_n$ d'une loi de densité $f(\,\cdot\,;\theta)$ sur $\mathcal{A} \in \mathbb{R}$ dépendant d'un paramètre inconnu $\theta \in \Theta$. Le modèle statistique associé à cette situation est le suivant

$$\begin{aligned} \mathcal{X} &= \mathcal{A}^n \ , \\ \mathcal{F} &= \text{boréliens de } \mathcal{X} \ , \\ P_\theta &= \text{loi de densité } p(x;\theta) = \textstyle\prod_{i=1}^n f(x_i;\theta) \ , \end{aligned}$$

où $x = (x_1, \ldots, x_n) \in \mathcal{X}$. $p(x;\theta)$ est appellée la densité conjointe de l'échantillon $X_1, \ldots, X_n$.

2.1.2 Estimateur

On se donne un modèle statistique $(\mathcal{X}\,,\,\mathcal{F}\,,\,\{P_\theta;\theta \in \Theta\})$. On cherche à estimer θ le paramètre inconnu.

Définition 2.1.4 *On appelle* estimateur *toute fonction mesurable*

$$T : \ \mathcal{X} \ \rightarrow \ \Theta \ ,$$

de l'espace des observations dans l'espace des paramètres. On parlera d'estimateur de θ.

Cette notion d'estimateur est très large. On verra dans les paragraphes suivants deux grandes classes d'estimateurs (ou plutôt de méthodes d'estimation), les estimateurs *paramétriques* et les estimateurs *non paramétriques.*

2.2 ESTIMATEURS PARAMETRIQUES

On se donne un modèle $(\mathcal{X}\,,\,\mathcal{F}\,,\,\{P_\theta;\theta \in \Theta\})$, et on suppose que Θ est une partie de $\mathbb{R}^d$. On note $p(x;\theta)$ la densité associée à la loi P_θ.

2.2.1 Biais et Risque

Définition 2.2.1 *Le* biais *de l'estimateur T est la fonction*

$$\theta \rightarrow b(\theta, T) \stackrel{\triangle}{=} E_\theta T - \theta \in \mathbb{R}^d \ .$$

Le risque *de l'estimateur T est la fonction*

$$\theta \rightarrow R(\theta, T) \stackrel{\triangle}{=} E_\theta |T - \theta|^2 \ .$$

Le biais définit donc l'erreur moyenne d'estimation. Le risque donne un critère de qualité de l'estimateur T.

Définition 2.2.2 *Un estimateur T est appelé* estimateur sans biais *(SB) si*

$$b(\theta, T) = 0 \ , \quad \forall \theta \in \Theta \ .$$

Etant donnés deux estimateurs T_1 et T_2 de θ, T_1 sera dit préférable *à T_2 si*

$$R(T_1, \theta) \leq R(T_2, \theta) \ , \quad \forall \theta \in \Theta \ .$$

Sauf dans certains cas triviaux, il n'existe pas d'estimateur préférable à tous les autres. Il est donc inutile d'en chercher un ! En revanche, il est possible de trouver un estimateur SB qui soit préférable à tout autre estimateur SB.

Définition 2.2.3 *Un estimateur T est dit* uniformément de variance minimum sans biais *(UVMB) s'il est sans biais et si, pour tout autre estimateur sans biais S, T est préférable à S, i.e.*

$$R(\theta, T) \leq R(\theta, S) \ , \quad \forall \theta \in \Theta \ .$$

Supposons que T soit un estimateur UVMB, il serait intéressant de pouvoir mesurer la qualité de cet estimateur. En d'autres termes, on aimerait pouvoir trouver une borne qui minimise le risque de tous les estimateurs sans biais. Cette borne est donnée dans le théorème 2.2.5, mais avant on introduit la notion d'information de Fisher

Définition 2.2.4 *On suppose que Θ est une partie ouverte de $\mathbb{R}^d$ et que $p(x;\theta)$ est régulière par rapport à θ. On définit la* matrice d'information de Fisher

$$[I(\theta)]_{i,j} \triangleq E_\theta\left(\frac{\partial \log p(X;\theta)}{\partial \theta_i}\frac{\partial \log p(X;\theta)}{\partial \theta_j}\right) .$$

$I(\theta) \in \mathbb{R}^d d$ est une matrice symétrique définie positive.

Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi de densité $f(\cdot;\theta)$ régulière par rapport à θ. On définit l'information de Fisher *(deuxième forme)*

$$[\mathcal{I}(\theta)]_{i,j} \triangleq E_\theta\left(\frac{\partial \log f(X_1;\theta)}{\partial \theta_i}\frac{\partial \log f(X_1;\theta)}{\partial \theta_j}\right) .$$

On peut vérifier (cf. exercice B.2.7) que

$$I(\theta) = n\,\mathcal{I}(\theta) .$$

On vérifie que la matrice d'information de Fisher s'écrit

$$[I(\theta)]_{i,j} = \begin{cases} \displaystyle\int_{x\in\mathcal{X}} \left(\frac{\partial \log p(x;\theta)}{\partial \theta_i}\frac{\partial \log p(x;\theta)}{\partial \theta_i}\right) p(x;\theta)\, dx , & \text{(dans le cas continu),} \\ \displaystyle\sum_{x\in\mathcal{X}} \left(\frac{\partial \log p(x;\theta)}{\partial \theta_i}\frac{\partial \log p(x;\theta)}{\partial \theta_i}\right) p(x;\theta) , & \text{(dans le cas discret).} \end{cases}$$

et

$$\mathcal{I}(\theta) = \begin{cases} \displaystyle\int_{\mathcal{A}} \left(\frac{\partial \log f(a;\theta)}{\partial \theta}\right)^2 f(a;\theta)\, da , & \text{(cas continu) ,} \\ \displaystyle\sum_{x\in\mathcal{A}} \left(\frac{\partial \log f(a;\theta)}{\partial \theta}\right)^2 f(a;\theta) , & \text{(cas discret) ,} \end{cases}$$

Théorème 2.2.5 (Inégalité de Cramer–Rao) *On suppose que Θ est une partie ouverte de $\mathbb{R}^d$ et que $p(x;\theta)$ est régulière par rapport à θ. Alors pour tout estimateur sans biais T,*

$$R(\theta, T) \geq \mathit{trace}\left(I(\theta)^{-1}\right) , \quad \forall \theta \in \Theta .$$

Un estimateur sans biais T qui réalise l'égalité dans l'inégalité précédente — pour tout $\theta \in \Theta$ — est dit efficace.

2.2.2 Estimateur du Maximum de Vraisemblance

Il n'existe pas de méthode d'estimation qui, dans tous les cas, minimise le risque (voir l'exercice 2.6.6). Toutefois, il existe une méthode qui donne habituellement de bons résultats. Considérons l'exemple suivant

Exemple 2.2.6 *Supposons que l'on jette une pièce de monnaie 10 fois de suite dans le but d'estimer p la probabilité d'obtenir face (F). Si on obtient 4 faces, il semble clair que le meilleur estimateur de p est 0.4. Pourquoi ? Il suffit de se dire que c'est la proportion de faces (=4/10). Mais en fait il existe une autre manière de raisonner: supposons que l'on observe la séquence*

$$x_0 = \; F\ P\ P\ P\ F\ P\ P\ F\ F\ P .$$

Si la probabilité d'obtenir F est p, la probabilité d'obtenir la séquence x_0 est

$$p^4(1-p)^6 .$$

Pour certaines valeurs de p, disons 0.9, la probabilité d'obtenir x_0 est trés petite. Donc si $p = 0.9$, on n'aurait vraisemblablement pas pû obtenir la séquence x_0. En fait, connaissant le résultat des 10 jets, il est raisonnable de penser que la vraie valeur du paramètre p est celle qui donne à la séquence observée la plus grande probabilité. C'est–à–dire, choisir comme estimation la valeur de p qui maximise $p^4(1-p)^6$. Cette fonction de p est nulle en 0 et 1, positive sur $]0,1[$. On peut trouver son maximum en dérivant

$$\begin{aligned} \frac{d}{dp}\left(p^4(1-p)^6\right) &= 0 \ , \\ \text{i.e.} \quad 4p^3(1-p)^6 - 6p^4(1-p)^5 &= 0 \ . \end{aligned}$$

Ainsi le maximum est atteint en $p = 0.4$. Dans ce cas, $p = 0.4$ est notre "estimateur du maximum de vraisemblance".

Comme on vient de le voir, si P_θ est une loi discrète, et si on observe x_0, l'estimateur du maximum de vraisemblance est la valeur de θ qui maximise la probabilité d'observer x_0, i.e.

$$\hat{\theta}(x_0) \in \operatorname{Arg}\max_{\theta\in\Theta} P_\theta(\{x_0\}) = \operatorname{Arg}\max_{\theta\in\Theta} p(x_0;\theta) \ .$$

Si P_θ est une loi continue $P(\{x_0\}) = 0$, l'argument précédent n'est donc plus valable. Mais maintenant si C_{x_0} est un carré de coté δ centré en x_0, on a

$$P_\theta\left(C_{x_0}\right) = \int_{C_{x_0}} p(x;\theta)\, dx \simeq p(x_0;\theta)\, \delta^n \ .$$

Si on veut rendre la probabilité de tomber dans le voisinage de x_0 la plus grande possible, on choisit (comme dans le cas discret)

$$\hat{\theta}(x_0) \in \operatorname{Arg}\max_{\theta\in\Theta} p(x_0;\theta) \ .$$

Exemple 2.2.7 *Avant de lancer sur le marché un nouveau type d'ampoule électrique, le producteur veut en déterminer la durée de vie moyenne. On prend donc un échantillon de 100 ampoules et on mesure $t = (t_1 \cdots t_{100})$, où t_i est la durée de vie de la ième ampoule. On suppose que la durée de vie de chacune de ces ampoules suit une loi exponentielle de paramètre λ et que la durée de vie de l'une est indépendante de celle de l'autre. L'observation admet donc une loi de densité*

$$p(t;\lambda) = \prod_{i=1}^{100} \frac{1}{\lambda} e^{-t_i/\lambda} = \frac{1}{\lambda^{100}}\, e^{-(t_1+\cdots+t_{100})/\lambda} \ , \quad t \in \{\mathbb{R}_+\}^{100} \ . \tag{2.1}$$

Supposons que les durées de vie observées soient $1900, 2534, \ldots, 1672$, on substitue $t_1 \cdots t_{100}$ par ces valeurs dans (2.1). Remarquons que seule la somme de ces durées de vie intervient dans (2.1). Si cette somme est égale à 175262, on obtient comme expression de la densite

$$\frac{1}{\lambda^{100}}\, e^{-175262/\lambda} \ . \tag{2.2}$$

En posant la dérivée du logarithme de cette dernière expression égale à 0, on obtient

$$\hat{\lambda}(t) = \frac{175262}{100} \ .$$

D'une manière générale, pour un échantillon de taille n, le logarithme de la densité s'écrit

$$l(\lambda) \triangleq \log p(t;\lambda) = -n \log \lambda - \frac{t_1 + \cdots + t_n}{\lambda} \ . \tag{2.3}$$

L'équation

$$\frac{d\, l(\lambda)}{d\lambda} = 0 \ ,$$

donne

$$\hat{\lambda}(t) = \frac{t_1 + \cdots + t_n}{n} \ .$$

Notez au passage qu'il n'est pas nécessaire de calculer une expression générale de la densité (cf. (2.3)) pour utiliser la méthode du maximum de vraisemblance. On peut aussi se placer dans le cas d'une réalisation particulière (cf. (2.2)). Ces deux démarches ont étés utilisées dans cet exemple.

La définition générale de l'estimateur du maximum de vraisemblance est la suivante

Définition 2.2.8 L'estimateur du maximum de vraisemblance *(EMV) de θ, noté $\hat{\theta}$, est défini par*

$$\hat{\theta}(x) \in \operatorname{Arg}\max_{\theta \in \Theta} p(x;\theta) \ , \quad \forall x \in \mathcal{X} \ .$$

La fonction $\theta \to L(\theta) \triangleq p(\,\cdot\,;\theta)$ est appelée *fonction de vraisemblance* ($L(\theta)$ est une variable aléatoire définie sur $(\mathcal{X}, \mathcal{F})$. Par définition, l'EMV est le maximum de cette fonction. Dans les deux exemples précédents, ce maximum a été calculé en posant la dérivée de $L(\theta)$ égale à 0, i.e.

$$\nabla_\theta \, p(x;\theta)|_{\theta=\hat{\theta}(x)} = 0 \ . \tag{2.4}$$

(2.4) est appelée *équation de la vraisemblance.* Pour écrire une telle équation, il est nécessaire que Θ soit une partie ouverte de $\mathbb{R}^d$ et que $L(\theta)$ soit régulier par rapport à θ. Mais cela n'est pas suffisant pour assurer l'existence et l'unicité d'une solution. En effet, il faut aussi que le point où s'annule la dérivée soit bien un maximum global (vieux problème !).

Dans l'exemple 2.2.7, au lieu d'écrire (2.4) on a annulé la dérivée du logarithme de la fonction de vraisemblance. On explique pourquoi dans l'exemple suivant

Exemple 2.2.9 *On dispose d'un n–échantillon d'une loi de densité $f(\,\cdot\,;\theta)$ ou θ appartient à une partie Θ de $\mathbb{R}^d$. La densité jointe de l'échantillon est donc*

$$p(x;\theta) \triangleq \prod_{i=1}^{n} f(x_i;\theta) \ , \quad \forall x \in \mathcal{X} \ .$$

Si on écrit l'équation de la vraisemblance (2.4), on obtient

$$\sum_{i=1}^{n} \left(\prod_{j \neq i} f(x_j;\theta) \right) \ \nabla_\theta \, f(x_i;\theta) = 0 \ . \tag{2.5}$$

Si on annule la dérivée du logarithme de $p(x;\theta)$, on obtient

$$\sum_{i=1}^{n} \frac{1}{f(x_i;\theta)} \ \nabla_\theta \, f(x_i;\theta) = 0 \ . \tag{2.6}$$

Une rapide comparaison de (2.5) et (2.6) explique l'intérêt de passer par le logarithme.

Comme on vient de le voir sur cet exemple plutôt général, il est plus intéressant de considérer l'équation de la vraisemblance sous la forme suivante (équivalente à (2.4))

$$\nabla_\theta \, \log p(x;\theta)|_{\theta=\hat{\theta}(x)} = 0 \ . \tag{2.7}$$

Il existe des cas irréguliers où on ne peut pas utiliser l'équation de la vraisemblance pour calculer l'EMV. L'exemple suivant en est une illustration

Exemple 2.2.10 *(EMV du paramètre d'une loi uniforme) On considère un n–échantillon d'une loi uniforme sur $[0,\theta]$, où θ est un paramètre positif inconnu. Cette loi admet une densité sur $\mathbb{R}$ définie par*

$$f(a;\theta) \triangleq \begin{cases} 0 \ , & a < 0 \ , \\ 1/\theta \ , & 0 \leq a \leq \theta \ , \\ 0 \ , & a > \theta \ , \end{cases} \qquad (a \in \mathbb{R})$$

la densité conjointe de l'échantillon est donc

$$p(x;\theta) \triangleq \begin{cases} 1/\theta^n \ , & \text{si } x \in \{x; 0 \le x_i \le \theta \, , \, i = 1 \cdots n\} \ , \\ 0 \ , & \text{sinon} \ , \end{cases} \qquad (x \in \mathbb{R}^n) \ . \tag{2.8}$$

Posons

$$l(\theta) \triangleq \log p(x;\theta) = -n \log \theta \ .$$

Maintenant on doit faire très attention, si on cherche à résoudre l'équation de la vraisemblance en prenant $p(x;\theta) = 1/\theta^n$, on obtient

$$-\frac{n}{\hat{\theta}} = 0 \Rightarrow \hat{\theta} = -\infty \ .$$

Mais en fait $1/\theta^n$ n'est pas la densité conjointe. L'EMV de θ se calcule en maximisant la densité (2.8) par rapport à θ, on obtient

$$\hat{\theta}(x) = \max\{x_1, \ldots, x_n\} \ ,$$

(à vérifier en exercice).

Proposition 2.2.11 *L'EMV possède les propriétés suivantes*

(*i*) *Principe d'Invariance: on se donne une fonction $h : \Theta \to \mathbb{R}^k$ alors $\widehat{h(\theta)} = h(\widehat{\theta})$ i.e. l'EMV de $h(\theta)$ est égale à h(EMV de θ).*

(*ii*) *Lien avec la borne de Cramer–Rao: s'il existe un estimateur sans biais qui atteint la borne de Cramer–Rao, alors il s'agit aussi de l'EMV.*

La propriété *(i)* signifie que l'EMV d'une certaine fonction de θ peut être calculée en appliquant cette fonction à l'EMV de θ. Ce résultat est très intéressant en pratique. La propriété *(ii)* est limitée par le fait qu'il est difficile de trouver un estimateur sans biais efficace. L'EMV possède d'autres propriétés intéressantes qu'on exposera dans le paragraphe 2.3 consacré à la théorie asymptotique.

2.2.3 Estimateurs Bayésien et "MAP"

On considère un modèle bayésien, i.e. un modèle statitistique $(\mathcal{X}, \mathcal{F}, \{P_\theta; \theta \in \Theta\})$ et une densité $p_{\underline{\theta}}(\,\cdot\,)$ de la loi d'une variable aléatoire $\underline{\theta}$ définie sur Θ. Etant donné T un estimateur de θ, on définit le *risque bayésien* associé à la loi a priori $p_{\underline{\theta}}$

$$\tilde{R}(T) \triangleq \begin{cases} \displaystyle\int_\Theta R(\theta, T)\, p_{\underline{\theta}}(\theta)\, d\theta \ , & (\text{si } p_{\underline{\theta}} \text{ est continue}), \\ \displaystyle\sum_{\theta \in \Theta} R(\theta, T)\, p_{\underline{\theta}}(\theta) \ , & (\text{si } p_{\underline{\theta}} \text{ est discrète}). \end{cases} \tag{2.9}$$

Définition 2.2.12 *Soit $\tilde{\theta}$ un estimateur du paramètre θ. Si pour tout estimateur T de θ on a*

$$\tilde{R}(\tilde{\theta}) \le \tilde{R}(T) \ ,$$

alors $\tilde{\theta}$ est appelé estimateur bayésien *du paramètre θ associé à la densité a priori $p_{\underline{\theta}}(\,\cdot\,)$.*

Proposition 2.2.13 *L'estimateur bayésien associé à la densité a priori $p_{\underline{\theta}}(\,\cdot\,)$, s'écrit explicitement en fonction de $p(\,\cdot\,;\theta)$, la densité de P_θ, de la manière suivante*

$$\tilde{\theta}(x) = \begin{cases} \dfrac{\displaystyle\int_\Theta \theta\, p(x;\theta)\, p_{\underline{\theta}}(\theta)\, d\theta}{\displaystyle\int_\Theta p(x;\theta)\, p_{\underline{\theta}}(\theta)\, d\theta} \ , & (\textit{si } p_{\underline{\theta}} \textit{ est continue}) \ , \\[2ex] \dfrac{\displaystyle\sum_{\theta \in \Theta} \theta\, p(x;\theta)\, p_{\underline{\theta}}(\theta)}{\displaystyle\sum_{\theta \in \Theta} p(x;\theta)\, p_{\underline{\theta}}(\theta)} \ , & (\textit{si } p_{\underline{\theta}} \textit{ est discrète}) \ , \end{cases} \tag{2.10}$$

Exemple 2.2.14 *Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi de Bernoulli de paramètre θ de densité conjointe*

$$p(x;\theta) = \theta^{s_n}\,(1-\theta)^{n-s_n} \quad , \quad (x \in \{0,1\}^n) \ ,$$

où $s_n = x_1 + \cdots + x_n$. On se donne une densité a priori sur le paramètre θ de la forme

$$p_{\underline{\theta}}(\theta) \triangleq \theta^{\alpha-1}\,(1-\theta)^{\beta-1} \ , (\theta \in [0,1]) \ ,$$

avec $\alpha > 0$ et $\beta > 0$. On rappelle les formules d'analyse suivantes. On définit la fonction

$$\beta(u,v) \triangleq \int_0^1 \theta^{u-1}\,(1-\theta)^{v-1}\,d\theta \ .$$

Elle vaut $\Gamma(u)\,\Gamma(v)/\Gamma(u+v)$ où $\Gamma(u)$ est la fonction Gamma définie par

$$\Gamma(u) \triangleq \int_0^\infty e^{-\theta}\,\theta^{u-1}\,d\theta \ .$$

Cette fonction vérifie $\Gamma(n+1) = n!$ pour $n \in \mathbb{N}$ et $\Gamma(u+1) = u\Gamma(u)$. On utilise ces résultats pour calculer l'estimateur bayésien de θ défini par (2.10) (cas continu)

$$\begin{aligned}
\tilde{\theta}(x) &= \frac{\int_0^1 \theta^{\alpha+s_n}\,(1-\theta)^{n-s_n+\beta-1}\,d\theta}{\int_0^1 \theta^{\alpha+s_n-1}\,(1-\theta)^{n-s_n+\beta-1}\,d\theta} \\
&= \frac{\Gamma(\alpha+s_n+1)}{\Gamma(\alpha+\beta+n+1)}\,\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s_n)} \\
&= \frac{\alpha+s_n}{n+\alpha+\beta} \ .
\end{aligned}$$

Cet estimateur est biaisé car

$$E_\theta\left(\tilde{\theta}-\theta\right) = \frac{\alpha-\theta\,(\alpha+\beta)}{n+\alpha+\beta} \ .$$

Le risque de cet estimateur est

$$R(\theta,\tilde{\theta}) = E_\theta\left(\tilde{\theta}-\theta\right)^2 = \frac{\theta^2\left((\alpha+\beta)^2-n\right)+\theta\,(n-2\alpha(\alpha+\beta))+\alpha^2}{(n+\alpha+\beta)^2} \ .$$

Prenons par exemple $\alpha = \beta = \sqrt{n}/2$. On obtient alors l'estimateur

$$\tilde{\theta}(x) = \frac{s_n+\sqrt{n}/2}{n+\sqrt{n}} \ ,$$

de risque

$$R(\theta,\tilde{\theta}) = \frac{1}{4\left(\sqrt{n}+1\right)^2} \ .$$

Ce risque est constant (ne dépend pas de θ).

Dans le cadre bayésien, il existe une seconde approche pour calculer un estimateur de θ.

Remarque 2.2.15 *Soit $p(x)$ la densité de la loi d'une variable aléatoire X à valeurs dans $\mathbb{R}$. Quel est le point x_0 de $\mathbb{R}$ qui peut le mieux représenter X ? Il existe deux approches. Premièrement, on peut dire que ce point est l'espérance de X, i.e. $x_0 = \int x\,p(x)\,dx$. Cette approche est justifiée par le fait que x_0 ainsi défini minimise la fonction $x \to E(X-x)^2$ (l'erreur quadratique). Deuxièment, on peut choisir un point tel que la probabilité que X tombe dans un voisinage de ce point soit la plus grande possible. Cela revient à choisir x_0 qui maximise la fonction $x \to P(x-\alpha \leq X \leq x+\alpha)$ (α petit). Donc x_0 est le point qui maximise $p(x)$, x_0 est appelé le* mode *de $p(x)$. Cette seconde approche suppose que $p(x)$ admette un maximum unique. Dans le cas d'une densité gaussienne, le mode et l'espérance coïncident.*

D'après cette remarque, on peut choisir comme estimateur celui qui maximise la densité a posteriori

$$p_{\underline{\theta}|X}(\theta|x) \triangleq \begin{cases} \dfrac{p(x;\theta)\, p_{\underline{\theta}}(\theta)}{\displaystyle\int_{\Theta} p(x;\theta)\, p_{\underline{\theta}}(\theta)\, d\theta} \ , & (\text{si } p_{\underline{\theta}} \text{ est continu}), \\ \dfrac{p(x;\theta)\, p_{\underline{\theta}}(\theta)}{\displaystyle\sum_{\theta\in\Theta} p(x;\theta)\, p_{\underline{\theta}}(\theta)} \ , & (\text{si } p_{\underline{\theta}} \text{ est discrète}), \end{cases}$$

Définition 2.2.16 L'estimateur "MAP" *(maximum a posteriori) de θ pour la densité a prioiri $p_{\underline{\theta}}(\theta)$, noté $\tilde{\theta}_{MAP}$, est défini par*

$$\tilde{\theta}_{MAP}(x) \in \operatorname{Arg}\max_{\theta\in\Theta} p_{\underline{\theta}|X}(\theta|X) = \operatorname{Arg}\max_{\theta\in\Theta} p(x;\theta)\, p_{\underline{\theta}}(\theta) \ , \quad \forall x \in \mathcal{X} \ .$$

Supposons la densité $p(x;\theta)$ unimodale, c'est–à–dire ne possédant "qu'une seule bosse". Prenons pour $p_{\underline{\theta}}(\theta)$ une densité uniforme sur un intervalle $[-K, K]$ (K grand). Il est alors facile de vérifier que l'estimateur MAP et l'EMV coïncident. En effet

$$\begin{aligned} \operatorname{Arg}\max_{\theta\in\Theta} p_{\underline{\theta}|X}(\theta|x) &= \operatorname{Arg}\max_{\theta\in\Theta} \frac{1}{c}\, p(x;\theta)\, p_{\underline{\theta}}(\theta) \\ &= \operatorname{Arg}\max_{-K\le\theta\le K} \frac{1}{2Kc}\, p(x;\theta) \\ &= \operatorname{Arg}\max_{-K\le\theta\le K} p(x;\theta) \\ &= \operatorname{Arg}\max_{\theta\in\Theta} p(x;\theta) \ . \end{aligned}$$

Cette dernière égalité est due au fait que, pour K suffisamment grand, le maximum en θ de $p(x;\theta)$ est atteint sur l'intervalle $[-K, K]$.

2.3 THEORIE ASYMPTOTIQUE

Soit $X = (X_1\, X_2 \cdots X_n \cdots)$ un échantillon de taille infinie d'une loi F_θ de densité $f(\,\cdot\,;\theta)$ sur $\mathcal{A} \subset \mathbb{R}^k$. On suppose que θ est unidimensionnel ($\Theta \subset \mathbb{R}$). Le modèle statistique correspondant est le suivant

$$\begin{array}{ll} \mathcal{X} & = \mathcal{A}^\infty \ , \\ \mathcal{F} & \text{borélien de } \mathcal{X} \ , \\ P_\theta & \text{loi de densité } p(x;\theta) = \prod_{i=1}^{\infty} f(x_i;\theta) \ , \quad x = (x_1 \cdots x_n \cdots) \in \mathcal{X} \ . \end{array}$$

On dispose d'une suite $\{T_n\}_{n\in\mathbb{N}}$ d'estimateurs de θ tels que, pour tout n,

$$T_n = T_n(X_1, \ldots, X_n) \ ,$$

(i.e. T_n ne dépend que de $X_1, \ldots, X_n$). Il s'agit d'étudier le comportement des estimateurs T_n lorsque $n \to \infty$.

Définition 2.3.1 *On dira que la suite d'estimateurs $\{T_n\}$ de θ est* consistante *si pour tout $\theta \in \Theta$*

$$T_n\,(X_1, \ldots, X_n) \underset{n\to+\infty}{\to} \theta \qquad P_\theta\text{–}p.s. \qquad (^1) \ .$$

Théorème 2.3.2 *Supposons la densité $f(\,\cdot\,;\theta)$ régulière par rapport à θ. On considère l'EMV $\hat{\theta}_n$ défini à partie de $X_1, \ldots, X_n$, i.e. $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ est solution de l'équation de la vraisemblance*

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i;\theta) = 0 \ .$$

La suite est consistante

$$\hat{\theta}_n \underset{n\to+\infty}{\to} \theta \ , \qquad P_\theta\text{–}p.s. \ , \ (\forall\theta \in \Theta).$$

1. $\xi_n \underset{n\to\infty}{\to} \xi$ P_θ–p.s. signifie que $P_\theta(\xi_n \underset{n\to\infty}{\to} \xi) = 1$.

Définition 2.3.3 *Soit $\{T_n\}$ une suite d'estimateurs de θ telle que*

$$loi\left(\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)}\right) \underset{n\to+\infty}{\to} N(0,1)$$

pour une suite $(\mu_n(\theta), \sigma_n(\theta)) \in \mathbb{R} \times \mathbb{R}$ donnée. Une telle suite d'estimateurs est dite asymptotiquement normale. *$\mu_n(\theta)$ est appelée la* moyenne asymptotique *et $\sigma_n(\theta)^2$ la* variance asymptotique. *La quantité suivante*

$$\frac{\mu_n(\theta) - \theta}{\sigma_n(\theta)}$$

définit le biais asymptotique. *La suite $\{T_n\}$ est dite* asymptotiquement sans biais *si le biais asymptotique tend vers 0 lorsque $n \to +\infty$.*

Théorème 2.3.4 *Supposons la densité $f(\,\cdot\,;\theta)$ régulière par rapport à θ. On considère l'EMV $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$. La suite $\hat{\theta}_n$ est asymptotiquement normale avec*

$$\mu_n(\theta) \triangleq \theta \ , \quad \sigma_n(\theta)^2 \triangleq (n \times \mathcal{I}(\theta))^{-1} \ .$$

C'est-à-dire

$$loi\left(\sqrt{n} \times \left(\hat{\theta}_n - \theta\right)\right) \underset{n\to+\infty}{\to} N\left(0, \mathcal{I}(\theta)^{-1}\right) \ , \quad (\forall\theta \in \Theta) \ ,$$

où $\mathcal{I}(\theta)$ est l'information de Fisher (définition 2.2.4).

2.4 INTERVALLES DE CONFIANCE

Dans toutes les méthodes d'estimation, une question naturelle vient à l'esprit: *qu'elle est la différence entre la vraie valeur d'un paramètre et son estimation?* Une première possibilité consiste à utiliser la fonction risque. Mais on peut vouloir une notion "plus parlante". Pour cela on utilise la notion d'intervalle de confiance.

On se donne un modèle statistique $\{P_\theta; \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, et un estimateur T de θ. Si par exemple, après estimation, on obtient $T = 7.0$, on se doute que θ sera proche de 7.0, mais rarement égale à 7.0. Ce que l'on aimerait obtenir, c'est un intervalle autour de la valeur estimée de la forme $[7.0 - 0.2, 7.0 + 0.2]$ (noté aussi 7.0 ± 0.2) tel qu'avec une grande probabilité, θ soit dans cet intervalle. La taille de l'intervalle nécessaire pour contenir θ donnera donc une bonne indication de la précision de l'estimation.

En général, on cherche deux fonctions $G(T)$ et $D(T)$ de l'estimateur T, avec $G(T) \leq T \leq D(T)$, telles que

$$\theta \in [G(T), G(T)] \ ,$$

avec une grande probabilité. On aimerait trouver une procédure générale, i.e. des fonctions $G(\,\cdot\,)$ et $D(\,\cdot\,)$ qui puissent être utilisées avec n'importe quel estimateur T de θ.

Définition 2.4.1 *Etant donné T une estimateur de θ, un* intervalle de confiance *(IC) de niveau $1 - \alpha$ est la donnée de deux fonctions $G(T)$ et $D(T)$ $(G(T) \leq T \leq D(T))$ de l'estimateur T telles que*

$$P_\theta\left(\theta \in [G(T), D(T)]\right) \geq 1 - \alpha \ , \quad (\forall\theta \in \Theta) \ .$$

2.4.1 Cas des Grands Echantillons

Dans beaucoup d'applications il existe une manière simple de calculer des IC fondée sur le fait que, pour les grands échantillons, l'estimateur à une loi simple. On se place dans le situation du paragraphe 2.3. On dispose donc d'une suite $\{T_n\}$ d'estimateurs de θ.

Supposons d'abord que

$$T_n \sim N\left(\theta\,,\, \sigma_n(\theta)^2\right) \ .$$

En utilisant une table de la loi de Gauss (cf. table p. **??**), on vérifie que

$$\begin{array}{rll} T_n \in & [\theta - 2\sigma_n(\theta)\,,\, \theta + 2\sigma_n(\theta)] \ , & \text{avec une probabilité de } 0.95 \ , \\ i.e. \quad \theta \in & [T_n - 2\sigma_n(\theta)\,,\, T_n + 2\sigma_n(\theta)] \ , & \text{avec une probabilité de } 0.95 \ . \end{array}$$

Si on suppose l'estimateur consistant, i.e. $T_n \simeq \theta$ (pour n grand), et que $\sigma_n(\theta)$ est une fonction régulière de θ, on peut remplacer dans l'expression précédente $\sigma_n(\theta)$ par $\sigma_n(T_n)$. On obtient alors

$$\theta \in [T_n - 2\sigma_n(T_n)\,,\, T_n + 2\sigma_n(T_n)]\ ,$$

avec une probabilité asymptotiquement ($n \to +\infty$) égale à 0.95. On a donc construit un IC de niveau (asymptotique) 0.95 .

De manière plus générale, on se fixe un niveau $1-\alpha$ et on calcule $z \geq 0$ tel que

$$\left\{ \begin{array}{l} P(-z \leq Z \leq z) = 1-\alpha\ , \\ Z \text{ v.a.r. de loi } N(0,1). \end{array} \right. \tag{2.11}$$

Alors

$$\begin{aligned} P_\theta\left(T_n \in [\theta - z\,\sigma_n(\theta)\,,\, \theta + z\,\sigma_n(\theta)]\right) &= 1-\alpha\ , \\ i.e. \qquad P_\theta\left(\theta \in [T_n - z\,\sigma_n(\theta)\,,\, T_n + z\,\sigma_n(\theta)\right) &= 1-\alpha\ . \end{aligned}$$

On remplace $\sigma_n(\theta)$ par $\sigma_n(T_n)$ pour conclure que

$$[T_n - z\,\sigma_n(T_n)\,,\, T_n + z\,\sigma_n(\mathbb{T}_n)]$$

est un IC de niveau (asymptotique) $1-\alpha$, où z est donné par (2.11). Donc dans le cas où T_n est une suite d'estimateurs non biaisés et de loi normale — $T_n \sim N(\theta\,,\,\sigma_n(\theta)^2)$ —, on a décrit une manière générale de construire des IC dans laquelle $G(T_n) = T_n - z\,\sigma_n(T_n)$ et $D(T_n) = T_n + z\,\sigma_n(T_n)$.

Cette approche se généralise au cas des suites $\{T_n\}$ d'estimateurs asymptotiquement normaux et sans biais, i.e. les suites telles qu'il existe une suite

$$(\mu_n(\theta)\,,\,\sigma_n(\theta))$$

telle que

$$loi\left(\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)}\right) \underset{n\to+\infty}{\to} N(0,1) \quad \text{et} \quad \frac{\mu_n(\theta)\theta}{\sigma_n(\theta)} \underset{n\to+\infty}{\to} 0\ .$$

En utilisant la même démarche que ci–dessus, on obtient le résultat suivant

Proposition 2.4.2 *Soit $\{T_n\}$ une suite d'estimateurs asymptotiquement normale et sans biais, de moyenne et de variance asymptotique $\mu_n(\theta)$ et $\sigma_n(\theta)^2$. Alors*

$$[T_n - z\,\sigma_n(\theta)\,,\, T_n + z\,\sigma_n(\theta)]$$

est un IC de niveau (asymptotique) $1-\alpha$, où z est défini en (2.11)

Dans la cas d'une suite d'EMV (cf. théorèmes 2.3.2 et 2.3.4), on obtient le résultat suivant

Proposition 2.4.3 *Soit $\hat\theta_n = \hat\theta_n(X_1 \cdot X_n)$ une suite d'EMV de θ, où $X_1, \ldots, X_n$ est un n–échantillon d'une loi de densité $f(\,\cdot\,;\theta)$ régulière par rapport à θ. Alors*

$$\left[\hat\theta_n - \frac{z}{\sqrt{n}}\frac{1}{\sqrt{\mathcal{I}(\hat\theta_n)}}\ ,\ \hat\theta_n + \frac{z}{\sqrt{n}}\frac{1}{\sqrt{\mathcal{I}(\hat\theta_n)}}\right]$$

est IC de niveau (asymptotique) $1-\alpha$, où z est défini en (2.11) et $\mathcal{I}$ est l'information de Fisher (définition 2.2.4).

Dans l'IC de la proposition 2.4.3, on voit apparaître un terme en $1/\sqrt{n}$. Pour diviser cet IC en deux il est donc nécessaire d'utiliser un échantillon 4 fois plus grand. Néanmoins, ces deux IC (propositions 2.4.2 et 2.4.3) donnent de bons résultats pour $n \geq 20$.

On peut retenir la *règle pratique* suivante

> *Etant donné une suite $\{T_n\}$ d'estimateurs de θ, la plupart du temps (dans les cas réguliers) on supposera cette suite asymptotiquement normale et sans biais. Il suffira donc de calculer sa variance (asymptotique) et d'utiliser l'IC décrit dans la proposition 2.4.2.*

2.4.2 Cas Des Echantillons Moyens

On se place ici aussi dans le cadre du chapitre 2.3. Lorsque l'échantillon est de taille moyenne, la méthode du paragraphe précédent n'est plus justifiée. On peut alors utiliser l'inégalité de Bienaymé–Chebychev

$$P(|Y| \geq \delta) \leq \frac{EY^2}{\delta^2} \ .$$

En effet, soit T_n une suite d'estimateurs de θ sans biais et de variance $r_n^2(\theta) = \mathrm{Var}_\theta T_n$. Dans l'inégalité précédente, remplaçons Y par $(T_n - \theta)/r_n(\theta)$. On obtient

$$\begin{aligned} P_\theta\left(\left|\frac{T_n-\theta}{r_n(\theta)}\right| \geq \delta\right) &\leq \frac{1}{\delta^2} \ , \\ i.e. \qquad P_\theta\left(\left|\frac{T_n-\theta}{r_n(\theta)}\right| \leq \delta\right) &\geq 1 - \frac{1}{\delta^2} \ , \\ i.e. \qquad P_\theta\left(\theta \in [T_n - \delta\, r_n(\theta), T_n + \delta\, r_n(\theta)]\right) &\geq 1 - \frac{1}{\delta^2} \ . \end{aligned} \tag{2.12}$$

Choisissons δ tel que

$$1 - \frac{1}{\delta^2} = 1 - \alpha \ , \quad i.e.\ \delta = \frac{1}{\sqrt{\alpha}} \ .$$

Si la suite $\{T_n\}$ est consistante et si $r_n(\theta)$ est une fonction régulière, on remplace $r_n(\theta)$ par $r_n(T_n)$ dans (2.12) pour obtenir le résultat suivant

Proposition 2.4.4 *Si $\{T_n\}$ est une suite consistante d'estimateurs de θ sans biais et de variance $r_n(\theta)^2$, alors*

$$\left[\ T_n - \frac{r_n(\theta)}{\sqrt{\alpha}} \ , \ T_n + \frac{r_n(\theta)}{\sqrt{\alpha}}\ \right]$$

est un IC de niveau (asymptotique) $1-\alpha$.

Cette méthode peut être utilisée pour des échantillons de taille moyenne ($n \geq 10$), mais elle présente l'inconvénient de donner des IC plutôt imprécis.

2.4.3 Cas Général

Plaçons–nous dans le cas où l'on dispose d'un estimateur T de θ. Supposons que l'on puisse trouver deux fonctions *monotones* $a(\theta)$ et $b(\theta)$ telles que

$$P_\theta\,(\,T \in [a(\theta), b(\theta)]\,) = 1 - \alpha \ , \qquad (\forall \theta \in \Theta) \ . \tag{2.13}$$

Puisque les fonctions $a(\,\cdot\,)$ et $b(\,\cdot\,)$ sont monotones, on peut trouver deux fonctions $G(\,\cdot\,)$ et $D(\,\cdot\,)$ telles que

$$a(\theta) \leq T \leq b(\theta) \qquad \Longleftrightarrow \qquad G(T) \leq \theta \leq D(T) \ .$$

Dans ce cas, $[G(T), D(T)]$ est un IC de niveau $1-\alpha$ car (2.13) est équivalent à

$$P_\theta\,(\theta \in [G(T), D(T)]) = 1 - \alpha \ , \qquad (\forall \theta \in \Theta) \ .$$

Proposition 2.4.5 *Soit T un estimateur de θ tel qu'il existe deux fonctions monotones $a(\theta)$ et $b(\theta)$ vérifiant (2.13) où $1-\alpha$ est un niveau fixé. Alors*

$$[G(T), D(T)]$$

définit un IC de niveau $1-\alpha$.

Ce résultat est intéressant à double titre. Premièrement, il est vrai quelque soit l'estimateur T. Deuxièmement, ce n'est pas un résultat asymptotique, il est donc utilisable quelque soit la taille de l'échantillon. En revanche, il est très difficile en pratique de trouver des fonctions $a(\theta)$ et $b(\theta)$ qui vérifient (2.13).

2.5 LE MODELE LINEAIRE

Exemple 2.5.1 *Un chaîne de grandes surfaces désire ouvrir un nouveau magasin et hésite entre plusieurs villes. Naturellement on souhaite que ce magasin soit le plus rentable possible. On retient l'hypothèse — quelque peu naïve — selon laquelle le chiffre d'affaires annuel (Y) d'une de ces grandes surfaces dépend linéairement des paramètres suivants:*

$$\begin{array}{lcl} a_1 & : & \textit{nombre d'habitants,} \\ a_2 & : & \textit{revenu moyen par habitant,} \\ a_3 & : & \textit{éloignement par rapport au centre ville.} \end{array}$$

Pour les dix magasins déjà existants les valeurs de Y, a_1, a_2, a_3 sont:

i	1	2	3	4	5	6	7	8	9	10
Y_i	237	514	1855	1685	929	4572	691	2942	1954	601
a_1^i	200	500	2000	1800	950	5000	700	3200	2100	600
a_2^i	70	80	77	84	89	95	88	83	87	89
a_3^i	0	6	16	5	8	25	18	10	2.8	6.5

(2.14)

L'hypothèse faite plus haut s'écrit donc

$$Y_i \simeq a_1^i\,\theta_1 + a_2^i\,\theta_2 + a_3^i\,\theta_3\ , \qquad (i = 1 \ldots 10)\ . \tag{2.15}$$

On veut estimer $\theta = (\theta_1\,\theta_2\,\theta_3)^*$, en essayant de faire correspondre "le mieux possible" les données réelles (2.14) avec le modèle théorique (2.15). Une approche consiste à minimiser l'erreur de modélisation que l'on peut représenter par la quantité suivante

$$S(\theta) \triangleq \sum_{i=1}^{10} (\,Y_i - a_1^i\,\theta_1 - a_2^i\,\theta_2 - a_3^i\,\theta_3\,)^2\ , \tag{2.16}$$

$S(\theta)$ mesure l'erreur que l'on fait lorsque l'on décide de choisir la valeur θ du paramètre. L'estimateur $\hat{\theta}$ qui minimise cette quantité est appelé estimateur des moindres carrés (EMC).

2.5.1 L'Estimateur des Moindres–Carrés

D'une manière générale, le problème s'introduit ainsi: on considère Y un vecteur aléatoire à valeurs dans $\mathbb{R}^n$, d'espérance $A\,\theta$ et de covariance $\sigma^2\,I_n$. $\theta \in \mathbb{R}^d$ est un paramètre inconnu, $A \in \mathbb{R}^{n\times d}$ est une matrice connue. On suppose, pour l'instant, σ^2 connu. Le but est d'estimer θ.

L'estimateur des moindres–carrés est celui qui minimise l'écart entre l'observation (d'une réalisation) de Y et l'espérance donnée par le modèle. Cet écart est mesuré par

$$S(\theta) = (Y - A\,\theta)^*\ (Y - A\,\theta)\ . \tag{2.17}$$

L'EMC est donc défini par

$$S(\hat{\theta}) = \min_{\theta\in\mathbb{R}^d} S(\theta)\ , \tag{2.18}$$

c'est–à–dire

$$\left.\frac{\partial S(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}} = 0\ , \quad i.e. \quad (A^*\,A)\,\hat{\theta} = A^*\,Y\ .$$

Donc si on suppose la matrice $(A^*\,A)$ inversible, l'EMC s'écrit

$$\hat{\theta} = (A^*\,A)^{-1}\,A^*\,Y\ . \tag{2.19}$$

Définition 2.5.2 L'estimateur des moindres–carrés $\hat{\theta}$ *de θ, est défini comme étant la solution de*

$$S(\hat{\theta}) = \min_{\theta\in\mathbb{R}^d} S(\theta)\ ,$$

si on suppose la matrice $(A^*\,A)$ inversible, alors cet estimateur s'écrit

$$\hat{\theta} = (A^*\,A)^{-1}\,A^*\,Y\ .$$

On fait les hypothèses

$$\begin{aligned} E_\theta\, Y &= A\,\theta\ , \\ E_\theta(Y - A\,\theta)\,(Y - A\,\theta)^* &= \sigma^2\, I_n\ , \end{aligned} \qquad (\forall\theta) \tag{2.20}$$

et

$$(A^*\, A) \text{ est inversible .} \tag{2.21}$$

Sous ces hypothèses, on a le résultat suivant

Proposition 2.5.3 *L'estimateur des moindres–carrés $\hat\theta$ vérifie*

(i) *$\hat\theta$ est un estimateur linéaire de θ,*

(ii) *$\hat\theta$ est un estimateur sans biais: $E_\theta\hat\theta = \theta \quad (\forall\theta)$,*

(iii) *$E_\theta(\hat\theta - \theta)\,(\hat\theta - \theta)^* = (A^*\, A)^{-1}\,\sigma^2$,*

(iv) *pour tout $\tilde\theta$ estimateur linéaire non biaisé:*

$$E_\theta(\tilde\theta - \theta)\,(\tilde\theta - \theta)^* \geq E_\theta(\hat\theta - \theta)\,(\hat\theta - \theta)^* \quad (\forall\theta)\ .$$

preuve
(i) $\hat\theta = LY$ avec $L = (A^*\, A)^{-1}\, A^*$, $\hat\theta$ est donc un estimateur linéaire.

(ii) $E_\theta\,\hat\theta = (A^*\, A)^{-1}\, A^*\ E_\theta\, Y = (A^*\, A)^{-1}\, A^*\, A\,\theta = \theta$.

(iii) $\hat\theta - \theta = (A^*\, A)^{-1}\, A^*\, Y - \theta = (A^*\, A)^{-1}\, A^*\,(Y - A\,\theta)$ donc

$$\begin{aligned} E_\theta(\hat\theta - \theta)\,(\hat\theta - \theta)^* &= (A^*\, A)^{-1}\, A^*\, E_\theta\,((Y - A\,\theta)\,(Y - A\,\theta)^*)\, A\,(A^*\, A)^{-1} \\ &= \sigma^2\,(A^*\, A)^{-1}\, A^*\, I_n\, A\,(A^*\, A)^{-1} \\ &= \sigma^2\,(A^*\, A)^{-1}\ . \end{aligned}$$

(iv) On se donne un estimateur linéaire sans biais

$$\tilde\theta = C\,Y\ , \qquad E_\theta\,\tilde\theta = \theta\,.$$

Donc $E_\theta(C\,Y) = \theta$ *i.e.* $C\,A\,\theta = \theta$ et ceci pour tout θ d'où

$$C\,A = I_d\ .$$

La covariance de $\tilde\theta$ est donnée par

$$\begin{aligned} E_\theta(\tilde\theta - \theta)\,(\tilde\theta - \theta)^* &= E_\theta(C\,Y - \theta)\,(C\,Y - \theta)^* \\ &= E_\theta(C\,Y - C\,A\,\theta + C\,A\,\theta - \theta) \\ &\qquad (C\,Y - C\,A\,\theta + C\,A\,\theta - \theta)^* \\ &= C\,E_\theta\,((Y - A\,\theta)\,(Y - A\,\theta)\,C^*) \qquad (\text{car } C\,A\,\theta - \theta = 0) \\ &= \sigma^2 C\,C^*\ . \end{aligned}$$

Posons $D = C - (A^*\, A)^{-1}\, A^*$

$$\begin{aligned} D\,D^* &= \left(C - (A^*\, A)^{-1}\, A^*\right)\left(C - (A^*\, A)^{-1}\, A^*\right)^* \\ &= C\,C^* - (A^*\, A)^{-1}\, A^*\, C^* - C\,A\,(A^*\, A)^{-1} + (A^*\, A)^{-1} \\ &= C\,C^* - (A^*\, A)^{-1}\ , \qquad (\text{car } C\,A = I_d) \\ &\geq 0\ . \end{aligned}$$

Donc $\sigma^2\, C\,C^* \geq \sigma^2 (A^*\, A)^{-1}$ ce qui termine la preuve de la proposition.

2.5.2 Estimation de la Variance

Dans la proposition 2.5.3, on a montré

$$E_\theta(\hat\theta-\theta)\,(\hat\theta-\theta)^* = (A^*\,A)^{-1}\,\sigma^2\ .$$

σ^2 est un paramètre de la loi de Y. En général, ce paramètre est inconnu et doit donc être estimé. Un estimateur non biaisé est décrit dans la proposition suivante

Proposition 2.5.4 *Un estimateur non biaisé de σ^2 est*

$$\hat\sigma^2 \triangleq \frac{S(\hat\theta)}{n-d}\quad , \tag{2.22}$$

où n est la dimension du vecteur d'observation, d la dimension du paramètre θ et $S(\hat\theta)$ est la somme des carrés des résidus

$$S(\hat\theta) = (Y - A\,\hat\theta)^*(Y - A\,\hat\theta)\ . \tag{2.23}$$

preuve (2.19) dans (2.23) donne

$$\begin{aligned}
S(\hat\theta) &= \left(Y - A\,(A^*\,A)^{-1}A^*\,Y\right)^*\ \left(Y - A\,(A^*\,A)^{-1}A^*\,Y\right)\\
&= Y^*\left(I_n - A\,(A^*\,A)^{-1}A^*\right)\ \left(I_n - A\,(A^*\,A)^{-1}A^*\right)Y\\
&= Y^*\left(I_n - A\,(A^*\,A)^{-1}A^*\right)Y\\
&= \mathrm{trace}\left(I_n - A\,(A^*\,A)^{-1}A^*\right)\ Y\,Y^*\ ,
\end{aligned}$$

$$\begin{aligned}
\text{donc}\quad E_\theta\,S(\hat\theta) &= \mathrm{trace}\left(I_n - A\,(A^*\,A)^{-1}A^*\right)\ (\sigma^2\,I_n + A\,\theta\,\theta^*\,A^*)\\
&= \sigma^2\,\mathrm{trace}\left(I_n - A\,(A^*\,A)^{-1}A^*\right)\\
&= \sigma^2\,\mathrm{trace}\,I_n - \sigma^2\,\mathrm{trace}\,A\,(A^*\,A)^{-1}\,A^*\\
&= \sigma^2\,\mathrm{trace}\,I_n - \sigma^2\,\mathrm{trace}\,(A^*\,A)^{-1}\,A^*A\\
&\qquad\qquad (\text{car trace } AB = \text{trace } BA)\\
&= \sigma^2\,\mathrm{trace}\,I_n - \sigma^2\,\mathrm{trace}\,I_d\\
&= \sigma^2(n-d)\ ,
\end{aligned}$$

i.e. $E_\theta\hat\sigma^2 = \sigma^2$.

2.5.3 Le Modèle Linéaire Gaussien

La proposition 2.5.3 établit que l'EMC est le meilleur des estimateurs *linéaires* (au sens de l'erreur quadratique). D'une manière générale, on ne sait pas démontrer plus, *sauf dans le cas gaussien.* En effet, dans ce cas on peut montrer que l'EMC est le meilleur des estimateurs sans biais et on peut aussi calculer explicitement sa loi. Tous ces résultats sont regroupés dans la proposition suivante

Proposition 2.5.5 *Supposons que $Y \sim N(A\,\theta, \sigma^2\,I_n)$, où $\theta \in \mathbb{R}^d$ et $\sigma \in \mathbb{R}$ sont inconnus, $A \in \mathbb{R}^{n\times d}$ connu. L'EMC $\hat\theta$ est sans biais, de variance minimum, i.e. pour tout estimateur $\tilde\theta$ sans biais:*

$$E_\theta(\tilde\theta-\theta)\,(\tilde\theta-\theta)^* \geq E_\theta(\hat\theta-\theta)\,(\hat\theta-\theta)^*\quad (\forall\theta)\ ,$$

c'est–à–dire, $\hat\theta$ est UVMB. Et $\hat\theta$ suit une loi de Gauss de moyenne θ et de covariance $\sigma^2\,(A^*\,A)^{-1}$. De plus,

$$\frac{\hat\theta_i - \theta_i}{\sqrt{C_{ii}\times\hat\sigma^2}} \sim \mathcal{T}_{n-d}\ ,$$

où $\mathcal{T}_{n-d}$ est la loi de Student à $n-d$ degrés de liberté et C_{ii} est le ième élément diagonal de $(A^*\,A)^{-1}$.

Exemple 2.5.6 *(suite de l'exemple 2.5.1) Dans cet exemple, on obtient*

$$\hat{\theta}_1 = 0.90 \ , \quad \hat{\theta}_2 = 0.77 \ , \quad \hat{\theta}_3 = -0.26 \ .$$

Ainsi pour une valeur $\bar{a} = (\bar{a}_1\, \bar{a}_2\, \bar{a}_3)$ *du paramètre a associé à une ville donnée, une nouvelle succursale installée dans cette ville "devrait" faire un chiffre d'affaires de*

$$\bar{Y} = 0.90\ \bar{a}_1 + 0.77\ \bar{a}_2 - 0.26\ \bar{a}_3 \ .$$

Par ailleurs, en utilisant (2.22), on obtient l'estimation suivante de la variance

$$\hat{\sigma}^2 = 4.42^2 \ ,$$

donc, si on suppose que l'on est dans le cas gaussien $Y \sim N(A\,\theta, \sigma^2\, I_n)$*, on peut affirmer que*

$$Y_i = 0.90\ a_1^i + 0.77\ a_2^i - 0.26\ a_3^i \quad (\pm 2 \times 4.42) \ , \qquad (i = 1 \ldots 10) \ ,$$

avec une probabilité de 0.95. La proposition 31 permet — dans le cas gaussien — de mieux préciser l'erreur d'estimation. En effet, dans cet exemple $n - d = 7$ *et à l'aide d'une table statistique de la loi de Student, on obtient*

$$P_\theta \left(\left| \frac{\hat{\theta}_i - \theta_i}{\sqrt{C_{ii} \times \hat{\sigma}^2}} \right| \geq 1.895 \right) = 0.1 \ ,$$

ou en d'autres termes

$$\hat{\theta}_i = \theta_i \pm 1.895 \sqrt{C_{ii} \times \hat{\sigma}^2} \quad \textit{avec une probabilité de 0.9 .}$$

2.6 EXERCICES

Exercice 2.6.1 *Soit* $X_1, \ldots, X_n$ *un* n*–échantillon d'une loi de Bernoulli* $B(\theta)$ *de paramètre inconnu* $\theta \in [0, 1]$.

- *Ecrivez le modèle statistique correspondant à cette situation, et calculez* T *la moyenne empirique.*
- *Calculez le biais et le risque de* T.

Exercice 2.6.2 *Soit* $X_1, \ldots, X_n$ *un* n*–échantillon d'une loi de Gauss sur* $\mathbb{R}$*, d'espérance* μ *et de variance* σ^2.

- *En supposant* σ^2 *connu (=1), calculez l'EMV de* μ.
- *En supposant* μ *connu (=0), calculez l'EMV de* σ^2.
- *Calculez l'EMV de* $\theta = (\mu, \sigma^2)$.

Dans chacun des cas étudiez le biais et le risque. Dans les deux premiers cas calculez la borne de Cramer–Rao et vérifiez si la borne est atteinte ou pas (cf. problème 2, annexe B).

Exercice 2.6.3 *(Exemple d'EMV biaisé) Soit* $X_1, \ldots, X_n$ *un* n*–échantillon d'une loi de densité*

$$f(x_i; \theta) \triangleq \theta^2\, x_i\, e^{-\theta x_i}$$

sur $\mathbb{R}_+$*, où* θ *est le paramètre inconnu.*

- *Calculez* $p(x; \theta)$ $(x = (x_1 \cdots x_n))$ *la densité conjointe.*
- *Posez l'équation de la vraisemblance et calculez* $\hat{\theta}$ *l'EMV de* θ.
- *Vérifiez que l'estimateur* $\hat{\theta}$ *est biaisé.*

Exercice 2.6.4 *Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi log–normale, i.e. la loi de $\log(X_i)$ est $N(\mu, \sigma^2)$ (X_i est à valeurs dans $\mathbb{R}_+$).*

– *On définit ξ et ν l'espérance et la variance de X_i, vérifiez que*

$$\xi = exp\left(\mu + \frac{\sigma^2}{2}\right) \ , \quad \nu = \xi^2 \ \left(e^{\sigma^2} - 1\right) \ .$$

– *Montrez que l'EMV de (ξ, ν) est*

$$\hat{\xi} = exp\left(\bar{X} + \frac{S^2}{2}\right) \ , \quad \hat{\nu} = \bar{X}^2 \ \left(e^{S^2} - 1\right) \ ,$$

où

$$\bar{X} \triangleq \frac{X_1 + \cdots + X_n}{n} \ , \quad S \triangleq \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n} \ .$$

Commentez (principe d'invariance).

Exercice 2.6.5 *(La méthode des moments) On considère un n–échantillon d'une loi — appelée exponentielle double — de densité sur $\mathbb{R}$ définie par*

$$f(a; \alpha, \beta) \triangleq \frac{1}{2\,\beta} \ e^{-|a-\alpha|/\beta} \ , \qquad (a \in \mathbb{R}) \ ,$$

où $\theta = (\alpha, \beta)$ est un paramètre inconnu.

– *Calculez les deux premiers moments $\mu_1(\alpha, \beta)$ et $\mu_2(\alpha, \beta)$ de X_i (réponse: α et $2\,\beta^2 + \alpha^2$).*

On considère les deux premiers moments empiriques de $X_1, \ldots, X_n$

$$\begin{aligned} \hat{\mu}_1 &\triangleq \frac{X_1 + \cdots + X_n}{n} \ , \\ \hat{\mu}_2 &\triangleq \frac{X_1^2 + \cdots + X_n^2}{n} \ . \end{aligned}$$

On pose les équations

$$\mu_1(\alpha, \beta) = \hat{\mu}_1 \quad , \quad \mu_2(\alpha, \beta) = \hat{\mu}_2 \ .$$

– *Déduisez de ces équations des estimateurs $\hat{\alpha}$ et $\hat{\beta}$ de α et β (réponse: $\hat{\alpha} = \hat{\mu}_1$ et $\hat{\beta} = \sqrt{(\hat{\mu}_2 - \hat{\mu}_1^2)/2}$).*

La méthode que l'on vient d'utiliser est une méthode générale (et ancienne) d'estimation appelée méthode des moments. *On considère un n–échantillon d'une loi de densité $f(\,\cdot\,; \theta)$, le paramètre inconnu θ est estimé en faisant correspondre les moments théoriques:*

$$\mu_k(\theta) \triangleq \begin{cases} \displaystyle\int_{\mathcal{A}} a^k \ f(a; \theta) \ da \ , & \textit{(cas continu)} \ , \\ \displaystyle\sum_{a \in \mathcal{A}} a^k \ f(a; \theta) \ , & \textit{(cas discret)}, \end{cases}$$

avec les moments empiriques:

$$\hat{\mu}_k \triangleq \frac{X_1^k + \cdots + X_n^k}{n} \ .$$

On obtient des équations de la forme

$$\mu_k(\theta) = \frac{X_1^k + \cdots + X_n^k}{n} \ .$$

Si θ est unidimensionnel on écrira une seule équation:

$$\mu_1(\theta) = \frac{X_1 + \cdots + X_n}{n} \ .$$

Si θ est bidimensionnel on écrira deux équations:

$$\begin{cases} \mu_1(\theta) &= \dfrac{X_1 + \cdots + X_n}{n} \ , \\ \mu_2(\theta) &= \dfrac{X_1^2 + \cdots + X_n^2}{n} \ . \end{cases}$$

Ainsi de suite.

- *Appliquez cette méthode au cas d'un n–échantillon d'une loi gaussienne*

$$g(\,\cdot\,;\mu,\sigma^2) \ , \quad \theta = (\mu,\sigma^2) \ .$$

La méthode des moments donne parfois des équations beaucoup plus simples à résoudre que les équations de la vraisemblance. Mais, en général, il est difficile de savoir si cette méthode va donner des estimations de qualité.

Exercice 2.6.6 *(Comparaison des risques de trois estimateurs) Soit $X_1 \cdots X_n$ un n–échantillon d'une loi de Bernoulli de paramètre $\theta \in [0,1]$. On considère les trois estimateurs de θ suivants*

$$T_1 \triangleq \frac{S_n}{n} \quad , \quad T_2 \triangleq \frac{S_n + \frac{\sqrt{n}}{2}}{n + \sqrt{n}} \quad , \quad T_3 \triangleq \theta_0 \ ,$$

où $S_n = (X_1 + \cdots + X_n)/n$. T_1 est l'EMV (ou la moyenne empirique). T_2 est l'estimateur bayésien associé à la densité a priori $p_{\underline{\theta}}(\theta) = \theta^{\sqrt{n}/2-1}\,(1-\theta)^{\sqrt{n}/2-1}$ (cf. exemple 2.2.14). T_3 est l'estimateur constant égale à un $\theta_0 \in [0,1]$donné.

- *Calculez $R_1(\theta)$, $R_2(\theta)$ et $R_3(\theta)$ les risques associés respectivement à T_1, T_2 et T_3.*
- *Tracez les trois courbes $\theta \to R_i(\theta)$ $(i = 1,2,3)$ et commentez.*

(Aucune des courbes n'est en–dessous des deux autres, i.e. aucun des estimateurs n'est préférable aux autres. Sur un intervalle centré en $1/2$, T_2 est meilleur que T_1, mais la longueur de cet intervalle tend vers 0 losrque $n \to \infty$).

Exercice 2.6.7 *Pendant N jours consécutifs on fait la même expérience: on irradie 1000 mouches et on compte le nombre de mutations. Le résultat de ces N expériences est noté $n_1, n_2, \ldots, n_n$. Le traitement est le même durant les N jours et le nombre de mutations est très faible comparé au nombre de mouches.*

Supposons que chaque mouche ait une probabilité p de muter et $q = 1-p$ de ne pas muter. Ainsi le nombre total n_i de mutations par jours est le résultat de 1000 réalisations indépendantes d'une loi de Bernoulli de paramètre p. Comme les n_i sont petits comparés à 1000 — le nombre total d'éssais —, on en déduit que p doit être relativement petit. Donc une bonne approximation pour le nombre total de mutations par jours est une loi de Poisson de paramètre λ. On retient donc le modèle statistique suivant: $(n_1 \cdots n_N)$ est un échantillon d'une loi de Poisson de paramètre λ.

- *Calculer $p(n_1 \cdots n_N;\lambda)$ la densité de la loi de l'échantillon.*
- *Poser l'équation de la vraisemblance et calculer $\hat{\lambda}$ l'EMV de λ.*

Chapitre 3

TEST D'HYPOTHESES

3.1 CONCEPTS GENERAUX

Dans un problème d'estimation on recherche la vraie valeur d'un paramètre θ, et la réponse est un point de l'espace des paramètres Θ. Dans un problème d'intervalle de confiance, on recherche un intervalle contenant la vraie valeur du paramètre et la réponse est un intervalle de Θ. Dans un problème de test, il n'existe que deux réponses possibles: *"oui"* ou *"non"*.

On se donne un modèle statistique $\{P_\theta\,;\,\theta \in \Theta\}$, le problème de test se formule ainsi: *étant donnée une partie Θ_0 de Θ, la vraie valeur de θ est–elle dans Θ_0 ou non?*

Exemple 3.1.1 *Un test d'intelligence non verbal a été mis au point sur un grand nombre d'enfants français de tel sorte que la répartition des résultats de ce test suit une loi normale de moyenne 100 et d'écart–type 10. Ce test est utilisé sur 50 enfants américains et le score moyen obtenu est 97.1 . Est–ce–que ce résultat est suffisamment plus petit que 100 pour indiquer une différence significative entre le score des enfants américains et celui des enfants français? En fait on essaie de savoir si la loi sous–jacente aux résultats obtenus par ces 50 enfants dévie de manière significative de la loi $N(100, 100)$.*

Ce problème peut se traiter à différents niveaux.

$\boxed{1}$ *Supposons que le résultat $X = (X_1 \cdots X_{50})$ est une suite de variables indépendantes, équidistribuées. Le modèle le plus simple consiste à choisir comme loi sous–jacente aux X_i une loi normale d'écart–type 10. Le modèle statistique correspondant est donc*

$$\{N(\mu, 100)\,;\,\mu \in \Theta\}\ , \qquad \textit{avec } \Theta \stackrel{\triangle}{=}]0, \infty[\ .$$

Comment choisir Θ_0 ? On peut prendre $\Theta_0 = \{100\}$: dans ce cas on veut savoir si la loi de X_i est $N(100, 100)$. On peut aussi prendre $\Theta_0 = \mathbb{R}_+ \setminus \{100\}$: dans ce cas on veut savoir si la moyenne de X_i n'est pas 100.

$\boxed{2}$ *Supposons que la variance empirique de l'échantillon $(X_1 \cdots X_{50})$ soit 59.5 . Dans ce cas il n'y a aucune raison de penser que l'écart–type des X_i soit 10. Le modèle statistique approprié est alors*

$$\left\{N(\mu, \sigma^2)\,;\,(\mu, \sigma) \in \Theta\right\}\ , \qquad \textit{avec } \Theta \stackrel{\triangle}{=}]0, \infty[\times]0, \infty[\ .$$

L'hypothèse selon laquelle la loi de X_i ne diffère pas sensiblement de $N(100, 100)$ s'écrit $(\mu, \sigma) = (100, 10)$. Dans ce cas $\Theta_0 = \Theta \setminus \{(100, 10)\}$. On peut maintenant ne pas être intéressé par l'écart–type mais seulement par la moyenne, i.e. on veut seulement tester si $\mu = 100$. Dans ce cas $\Theta_0 = \Theta \setminus \{(100, \sigma)\,;\,\sigma \in]0, \infty[\}$.

$\boxed{3}$ *A un autre niveau, on peut aussi remettre en cause l'hypothèse selon laquelle la loi des X_i est normale. Ce genre de situation sera étudié dans le chapitre consacré aux méthodes non-paramétriques.*

Définition 3.1.2 *Un* problème de test d'hypothèses *est la donnée d'un modèle statistique $\{P_\theta\,;\,\theta \in \Theta\}$ et d'une partie Θ_0 de Θ. L'hypothèse H_0: "$\theta \in \Theta_0$" est appelée hypothèse nulle. On pose $\Theta_1 = \Theta \setminus \Theta_0$, l'hypothèse H_1: "$\theta \in \Theta_1$" est appelée hypothèse alternative.*

Dans un problème de test, au vu d'une observation X, on veut décider si l'hypothèse H_0 est vraie ou fausse. On veut en fait une *"règle de décision"*

$$\mathcal{X} \ni x \longrightarrow \rho(x) = \begin{cases} 0 \text{ si on décide d'accepter } H_0 \text{ au vu de } x, \\ 1 \text{ si on décide de rejeter } H_0 \text{ au vu de } x. \end{cases}$$

Le plus souvent cette règle de décision sera de la forme suivante: $\rho(x) = \mathbb{I}_D(x)$, i.e. on accepte H_0 si $x \notin D$ et on rejette dans le cas contraire.

Définition 3.1.3 *Soit $\{P_\theta\,;\,\theta \in \Theta\}$ un modèle statistique et Θ_0 un ensemble de test. Soit D une partie de l'espace des observations $\mathcal{X}$. La* règle de décision ρ_D *— ou* test *d'hypothèse —, fondée sur l'ensemble D est*

$$\mathcal{X} \ni x \longrightarrow \rho_D(x) \triangleq \begin{cases} 0\ , & \textit{i.e. accepter } H_0 \quad \textit{si } x \notin D\ , \\ 1\ , & \textit{i.e. rejeter } H_0 \quad \textit{si } x \in D\ . \end{cases}$$

D est appelé région de rejet *de l'hypothèse H_0 (D sera parfois aussi appelé* test*).*

Pour un problème de test donné, il existe plusieurs tests — règles de décision — possibles (autant que de parties de $\mathcal{X}$). Il est alors important de pouvoir comparer tous ces tests, dans ce but on introduit les notations suivantes

Définition 3.1.4 *Soit ρ_D un test de région de rejet D. La* puissance *du test ρ_D est la fonction*

$$\Theta \ni \theta \longrightarrow \beta_D(\theta) \triangleq P_\theta(D)\ .$$

La taille *du test ρ_D est la quantité*

$$\alpha_D \triangleq \sup_{\theta \in \Theta_0} P_\theta(D)$$

Exemple 3.1.5 *Supposons que l'on vous donne une pièce de monnaie en vous demandant de vérifier si elle est biaisée ou non. Vous voilà dans l'embarras. Quel niveau de biais vous demande–t–on de détecter? Vous demande–t–on de mettre au point un test capable de détecter un biais de 1%, de 5%, ou alors est–il suffisant de pouvoir détecter un biais de 20%? On veut tester l'hypothèse*

$$H_0\ :\ \text{“}p = 0.5\text{”}\ .$$

On construit le test suivant: on jette la pièce n fois, soit $X = (X_1 \cdots X_n)$ le résultat obtenu ($X_i = 0$ si le ième jet est "face", 1 dans le cas contraire). On définit le test

$$\begin{cases} \textit{si } \left|X_1 + \cdots + X_n - \frac{n}{2}\right| \leq N & \textit{on accepte } H_0, \\ \textit{si } \left|X_1 + \cdots + X_n - \frac{n}{2}\right| > N & \textit{on rejette } H_0. \end{cases}$$

Où le seuil N est à déterminer. La région de rejet associée à ce test est

$$D \triangleq \left\{x \in \{0,1\}^n\,;\, \left|x_1 + \cdots + x_n - \frac{n}{2}\right| > N\right\}\ .$$

La puissance de ce test est

$$\beta_D(p) \triangleq P_p(D) = P_p\left(\left\{x\,;\, \left|x_1 + \cdots + x_n - \frac{n}{2}\right| > N\right\}\right)\ .$$

Pour n petit on peut calculer β_D en utilisant une table de la loi binomiale, pour n grand on utilise l'approximation normale

$$\textit{loi}\left(\frac{X_1 + \cdots + X_n - n\,p}{\sqrt{n\,p\,(1-p)}}\right) \underset{n\to\infty}{\to} N(0,1)\ ,$$

pour obtenir

$$\beta_D(p) \simeq 2 \times \int_a^\infty g(x;0,1)\,dx\ , \quad a = \frac{N + n/2 - n/p}{\sqrt{n\,p\,(1-p)}}\ ,$$

($g(\cdot\,;0,1)$ densité de la loi normale centrée réduite). En prenant $n = 1000$, $N = 16, 32, 48$ on obtient

N=16

p	0.46	0.48	0.50	0.52	0.54
$\beta_D(p)$	0.93	0.61	0.32	0.61	0.93

N=32

p	0.46	0.48	0.50	0.52	0.54
$\beta_D(p)$	0.69	0.23	0.05	0.23	0.69

N=48

p	0.46	0.48	0.50	0.52	0.54
$\beta_D(p)$	0.40	0.04	0.003	0.04	0.40

On considère maintenant la tailled de ce test $\alpha_D = P_{0.5}(D)$. Si $N = 16$, d'après les tableaux précédents, on obtient $\alpha_D = 0.32$. Cela signifie que, bien que H_0 soit vraie, 32 fois sur 100 l'observation X tombera dans D (i.e. rejet de H_0). En revanche, si on prend N plus grand ($N = 48$), alors $\alpha_D = 0.003$ et dans ce cas on ne rejettera (à tort) H_0 que 3 fois sur 1000.

Définition 3.1.6 *Soit $\alpha \in]0,1[$ fixé, un test ρ_D est dit de* niveau α *si $\alpha_D \leq \alpha$. Un test ρ_D est dit* uniformément le plus puissant *(UPP) au niveau α si*

$$\begin{cases} \rho_D \text{ est de niveau } \alpha, \\ \\ \text{pour tout test } \rho_C \text{ de niveau } \alpha: \quad \beta_D(\theta) \geq \beta_C(\theta) \ , \quad \forall \theta \in \Theta_1 \ . \end{cases}$$

Rechercher un test UPP de niveau α implique que les hypothèses H_0 et H_1 ne sont pas considérées de la même manière. Le tableau suivant donne les différents résultats possibles d'un test

VERITE	DECISION	RESULTAT
H_0	H_0	décision correcte
H_0	H_1	erreur de type I
H_1	H_0	erreur de type II
H_1	H_1	décision correcte

Donc rechercher un test UPP de niveau α revient à

$$\begin{array}{rl} \textit{choisir parmi tous les tests tels que} & P(\ \textit{erreur de type I}\ ) \leq \alpha \\ \textit{celui qui minimise} & P(\ \textit{erreur de type II}\ ) \end{array}$$

Cette dissymétrie s'introduit de manière naturelle dans beaucoup de problèmes concrets. Par exemple, si H_0 est "tel traitement médical est dangereux", en contrôlant l'erreur de type I on contrôle en fait le risque le plus grave: celui de juger inoffensif un traitement qui, en réalité, est dangereux. Juger ce traitement dangereux alors qu'il ne l'est pas est naturellement beaucoup moins grave.

3.2 TEST DE 2 HYPOTHESES SIMPLES

Dans la plupart des problèmes il n'existe pas de test UPP. Une exception est le cas où l'espace des paramètes est réduit à deux points

$$\Theta \triangleq \{\theta_0, \theta_1\} \quad \text{et} \quad \begin{cases} H_0 & : \ \text{“}\theta = \theta_0\text{”} \\ H_1 & : \ \text{“}\theta = \theta_1\text{”} \end{cases}$$

Dans ce cas nous allons voir qu'il existe un test UPP.

Exemple 3.2.1 *Lorsque un certain évènement se produit, un transmeteur radio à pour tâche d'envoyer un signal "+1" sur un court intervalle de temps. Comme le canal qui transmet ce signal est bruité, le signal est échantillonné 10 fois durant l'intervalle de temps en question. On obtient donc une observation de la forme*

$$X_i = S + b_i \ , \quad i = 1 \ldots 10 \ ,$$

où S — le signal — vaut 0 ou 1, b_i est un bruit d'observation additif que l'on suppose gaussien

$$\text{les } b_i \text{ sont indépendants et de même loi } N(0,1) \ .$$

Dans l'observation $X = (X_1 \cdots X_{10})$, on dit que le signal est présent si $S = 1$, qu'il est absent si $S = 0$. Au vu de l'observation X on veut savoir s'il le signal est présent, i.e. tester l'hypothèse

$$H_0 \ : \ \text{“ présence d'un signal (i.e. } S = 1\text{) ”} \ .$$

Le modèle statistique associé est $\{p(x;\theta); \theta \in \{0,1\}\}$ où

$$p(x;\theta) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\,\pi}} \exp\left(-\frac{(x_i-\theta)^2}{2}\right) \ , \quad x \in \mathbb{R}^{10} \ .$$

L'espace des paramètres Θ est constitué de deux valeurs $\theta_0 = 1$ et $\theta_1 = 0$. On se fixe $\alpha \in]0,1[$ et on cherche un test ρ_D tel que

$$P_{\theta_0}(D) \leq \alpha \ ,$$

et tel que $P_{\theta_1}(D)$ soit le plus grand possible. Une autre manière de construire D est la suivante: “D doit contenir tous les points x tels que la probabilité qu'ils soient effectivement observés est grande sous H_1 et petite sous H_0”. Ceci suggère la procédure suivante: étant donné un certain seuil c, on inclut x dans D si le rapport des densités vérifie

$$\frac{p(x;\theta_0)}{p(x;\theta_1)} \leq c \ ,$$

où c est calculé tel que $P_{\theta_0}(D) = \alpha$. Dans notre exemple

$$\frac{p(x;\theta_0)}{p(x;\theta_1)} = e^{-5} \ \exp(x_1 + \cdots + x_{10}) \ ,$$

et le test s'écrit

$$x_1 + \cdots + x_{10} \leq 5 + \log c = c' \ .$$

La règle de décision est la suivante: accepter H_0 si la somme des signaux reçus est inférieure à un certain seuil et rejeter H_0 dans le cas contraire. Le seuil c' est déterminé tel que

$$P_{\theta_0}\left(\{x; x_1 + \cdots + x_{10} \leq c'\}\right) = \alpha \ .$$

Notez que $Z = \sum_{i=1}^{10}((X_i - 1)/\sqrt{10})$ est une variable $N(0,1)$ sous l'hypothèse H_0. L'équation précédente peut s'écrire

$$P\left(Z \leq \frac{c'-10}{\sqrt{10}}\right) = \alpha \ .$$

*On détermine c' à l'aide d'une table (cf. Annexe **??** paragraphe **??**) de la loi normale. Si on veut un test de niveau 0.05 (i.e. $\alpha = 0.05$), à l'aide d'une table on vérifie que $P(Z \leq -1.65) = 0.05$ d'où $c' = 4.78$.*

L'approche utilisée dans cet exemple marche dans tous les cas où l'espace des paramètres Θ est réduit à deux points et aboutit au meilleur test:

Théorème 3.2.2 (lemme de Neyman–Pearson) *On considère un modèle statistique $\{P_\theta; \theta \in \Theta\}$ avec $\Theta = \{\theta_0, \theta_1\}$, P_θ de densité $p(x;\theta)$ ($x \in \mathcal{X}$). On veut tester l'hypothèse*

$$H_0 \ : \ \text{“}\theta = \theta_0\text{”} \ .$$

On définit la région de rejet

$$D \triangleq \left\{x \in \mathcal{X}; \frac{p(x;\theta_0)}{p(x;\theta_1)} \leq c\right\}$$

où c est choisi tel que $P_{\theta_0}(D) = \alpha$. Alors, le test ρ_D de région de rejet D, est un test UPP de niveau α.

preuve Soit D' une autre région de rejet de niveau α. Pour démontrer le théorème il suffit de vérifier que

$$P_{\theta_1}(D) \geq P_{\theta_1}(D') \ .$$

On pose $B = D \cap D'$, $A = D \setminus B$ et $A' = D' \setminus B$. On a

$$P_\theta(D) = P_\theta(A) + P_\theta(B) \ , \quad P_\theta(D') = P_\theta(A') + P_\theta(B) \ , \quad (\theta = \theta_0, \theta_1) \ .$$

Comme $P_{\theta_0}(D) = \alpha$ et D' est de niveau α

$$P_{\theta_0}(A') \leq P_{\theta_0}(A) \ .$$

Comme A' est en dehors de D, pour $x \in A'$ on a

$$p(x;\theta_0) \geq c \ p(x;\theta_1) \ ,$$

donc en intégrant (ou en sommant dans le cas discret) sur A' on a

$$P_{\theta_0}(A') \geq c \ P_{\theta_1}(A') \ .$$

Maintenant A est dans D, donc si $x \in A$ on a

$$p(x;\theta_0) \leq c \ p(x;\theta_1) \ ,$$

donc en intégrant (ou en sommant dans le cas discret) sur A on a

$$P_{\theta_0}(A) \leq c \ P_{\theta_1}(A) \ .$$

Des deux inégalités précédentes on déduit $P_{\theta_1}(D') \leq P_{\theta_1}(D)$.

Remarque 3.2.3 *Dans le cas où $p(x;\theta)$ est une densité continue, il est toujours possible de trouver c tel que $P_{\theta_0}(D) = \alpha$. Mais dans le cas discret cela est souvent impossible. Par exemple, s'il existe 50 valeurs possibles pour x, chacune ayant une probabilité de $1/50$ sous P_{θ_0}, on ne peut pas trouver c tel que $P_{\theta_0}(D) = 0.95$. En pratique, il suffira de choisir c de telle sorte que $P_{\theta_0}(D)$ soit le plus près possible de la valeur choisie.*

Dans le théorème 3.2.2, l'expression importante est le rapport

$$\lambda(x) \triangleq \frac{p(x;\theta_0)}{p(x;\theta_1)} \ ,$$

appelé *rapport de vraisemblance* de l'hypothèse " $\theta = \theta_0$ ". $\lambda(x)$ nous a donc permis de construire un test de région de rejet $\{\lambda \leq c\}$. Cette approche se généralise au cas où Θ ne se réduit plus à deux points, mais le test ainsi obtenu n'est plus UPP.

TEST DE 2 HYPOTHESES SIMPLES (Neyman–Pearson)		
H_0	H_1	On rejette H_0 au niveau α si
$\theta = \theta_0$	$\theta \neq \theta_0$	$\lambda(X) = \dfrac{p(X;\theta_0)}{p(X;\theta_1)} \leq c$

où c est choisi tel que $P_{\theta_0}(\{x; \lambda(x) \leq c\}) = \alpha$.

3.3 TEST DU RAPPORT DE VRAISEMBLANCE (TRV)

On considère le problème de test suivant

$$\{P_\theta\,;\,\theta\in\Theta\}\ ,\qquad H_0\ :\ \text{“}\theta\in\Theta_0\text{”}\ .$$

On note $p(x;\theta)$ la densité de la loi P_θ. On ne suppose plus que Θ se réduit à deux points. Le raisonnement du paragraphe précédent se généralise de la manière suivante

Définition 3.3.1 *Le* rapport de vraisemblance *de l'hypothèse H_0 est la fonction*

$$\lambda(x)\triangleq\frac{\sup_{\theta\in\Theta_0}p(x;\theta)}{\sup_{\theta\in\Theta}p(x;\theta)}\ .$$

Le test du rapport de vraisemblance *(TRV) de l'hypothèse H_0 est définit par*

$$\lambda(X)\ \begin{cases}>c & \textit{on accepte } H_0,\\[2ex] \le c & \textit{on rejette } H_0,\end{cases}$$

où c est un seuil fixé. Le TRV est le test de région de rejet $\{\lambda\le c\}$.

Remarque 3.3.2 *Le calcul du seuil c se fait de la manière suivante: soit α un niveau fixé. Lorsque c croit, la région de rejet $\{\lambda\le c\}$ grandit. On prend donc la plus grande valeur de c telle que $P_\theta(\lambda\le c)\le\alpha$ pour tout $\theta\in\Theta_0$*

$$c=Arg\max\{u;P_\theta(\lambda\le u)\le\alpha\,,\ \forall\theta\in\Theta_0\}\ .$$

Remarque 3.3.3 *Au lieu de considérer $\lambda(x)$, on peut considérer un autre rapport de vraisemblance*

$$\lambda'(x)\triangleq\frac{\sup_{\theta\in\Theta_0}p(x;\theta)}{\sup_{\theta\in\Theta_1}p(x;\theta)}\ .$$

C'est en fait λ' que l'on a utilisé dans le paragraphe précédent. L'intérêt de $\lambda(x)$ est le suivant: par définition de l'EMV de θ on a

$$\lambda(x)=\frac{\sup_{\theta\in\Theta_0}p(x;\theta)}{p(x;\hat\theta)}\ ,$$

donc si l'on dispose déjà de l'EMV de θ, le TRV de région de rejet $\{\lambda\le c\}$ est plus facile à construire que celui associé à $\{\lambda'\le c\}$.

Il est facile de vérifier que $\lambda(x)\le 1$, donc le seuil c vérifie nécéssairement $c<1$, mais si $\hat\theta(x)\in\Theta_0$ alors

$$\lambda(x)=\frac{\sup_{\theta\in\Theta_0}p(x;\theta)}{p(x;\hat\theta(x))}=\frac{p(x;\hat\theta(x))}{p(x;\hat\theta(x))}=1\ge c\ .$$

Cela signifie que $\{\hat\theta\in\Theta_0\}\subset\{\lambda\ge c\}$. On en déduit donc la règle générale

$$\textit{si } \hat\theta(x)\in\Theta_0 \textit{ alors accepter } H_0\ .$$

Naturellement il y a des cas où $\hat\theta(x)\notin\Theta_0$ et où le TRV accepte H_0.

Dans beaucoup d'exemple il est préférable d'utiliser la forme suivante du rapport de vraisemblance $l(x)\triangleq -2\log\lambda(x)$, qui est parfois plus simple à calculer (en particulier dans le cas gaussien).

Dans le paragraphe suivant nous allons appliquer le TRV à des échantillons gaussiens pour différents problèmes de test.

TEST DU RAPPORT DE VRAISEMBLANCE		
H_0	H_1	On rejette H_0 au niveau α si
$\theta\in\Theta_0$	$\theta\notin\Theta_0$	$\lambda(X)=\dfrac{\sup_{\theta\in\Theta_0}p(X;\theta)}{\sup_{\theta\in\Theta}p(X;\theta)}\le c$

où $c=\operatorname{Arg\,max}\{u;P_\theta(\lambda\le u)\le\alpha\,,\ \forall\theta\in\Theta_0\}$.

3.4 APPLICATIONS DU TRV AU CAS GAUSSIEN

Soit $X = (X_1, \cdots, X_n)$ un échantillon d'un loi $N(\mu, \sigma^2)$. Les EMV de μ et σ^2 sont

$$\hat{\mu}(x) \triangleq \frac{x_1 + \cdots + x_n}{n} , \tag{3.1}$$

$$\hat{\sigma}^2(x) \triangleq \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 . \tag{3.2}$$

Un autre estimateur de la variance est

$$\hat{s}^2(x) \triangleq \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 . \tag{3.3}$$

Les estimateurs $\hat{\mu}$ et $\hat{s}^2$ sont indépendants et

$$\frac{\sqrt{n}}{\sigma}(\hat{\mu}(X) - \mu) \sim N(0,1) \quad \text{loi normale centrée réduite,} \tag{3.4}$$

$$\frac{1}{\sigma^2}\sum(X_i - \mu)^2 \sim \chi^2_n \quad \text{loi du } \chi^2 \text{ à } n \text{ degrés de liberté,} \tag{3.5}$$

$$\frac{1}{\hat{s}^2(X)}\sum(X_i - \mu)^2 \sim \chi^2_{n-1} \quad \text{loi du } \chi^2 \text{ à } n-1 \text{ degrés de liberté,} \tag{3.6}$$

$$\frac{\sqrt{n}}{\hat{s}(X)}(\hat{\mu}(X) - \mu) \sim t_{n-1} \quad \text{loi de Student à } n-1 \text{ degrés de liberté,} \tag{3.7}$$

3.4.1 Tests Portant sur L'Espérance

On veut tester l'hypothèse

$$H_0 \ : \text{“ } \mu = \mu_0 \text{ ”} \quad \text{contre l'alternative} \quad H_1 \ : \text{“ } \mu \neq \mu_0 \text{ ”}$$

où $\mu_0 \in \mathbb{R}$ est une valeur donnée. On suppose que la variance σ^2 est inconnue (on présentera le cas où σ^2 est connue en fin de paragraphe). On veut calculer le rapport de vraisemblance

$$\lambda(x) \triangleq \frac{\sup_{\theta\in\Theta_0} p(x;\theta)}{\sup_{\theta\in\Theta} p(x;\theta)} .$$

Dans notre cas

$$\Theta = \{(\mu,\sigma)\,;\, \mu \in \mathbb{R},\ \sigma > 0\} \quad \text{et} \quad \Theta_0 = \{(\mu,\sigma)\,;\, \mu = \mu_0\,,\ \sigma > 0\} .$$

Le dénominateur du rapport de vraisemblance est égal à $p(x;\hat{\theta},\hat{\sigma})$, où $\hat{\theta}$ est l'EMV de θ.

$$\begin{aligned} p(x;\hat{\theta},\hat{\sigma}) &= \frac{1}{(2\pi\,\hat{\sigma})^{n/2}}\exp\left[-\frac{1}{2\hat{\sigma}^2}\sum(x_i - \hat{\mu})^2\right] \\ &= \frac{1}{(2\pi\,\hat{\sigma})^{n/2}}\exp\left[-\frac{n}{2}\right] . \end{aligned}$$

Le numérateur du rapport de vraisemblance est $p(x;\mu_0,\hat{\sigma}_0^2)$ où σ_0 est l'EMV de σ lorsque $\mu = \mu_0$, cet EMV est obtenu en résolvant l'équation de la vraisemblance

$$\left.\frac{\partial}{\partial\sigma}p(x;\mu_0,\sigma)\right|_{\sigma=\hat{\sigma}_0} = 0 ,$$

on obtient

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2 .$$

Le rapport de vraisemblance est donc

$$\lambda(x) = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} .$$

La région de rejet du TRV est donc de la forme

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \geq c^{-n/2} , \quad \text{i.e. } \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1 \geq c^{-n/2} - 1 = c' ,$$

On veut maintenant calculer le seuil c' en fonction d'un niveau α donné, pour cela on considère

$$\begin{aligned}
\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1 &= \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \\
&= \frac{1}{n\,\hat{\sigma}^2}\left(\sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \hat{\mu})^2\right) \\
&= \frac{(\hat{\mu} - \mu_0)^2}{\hat{\sigma}^2} \\
&= \frac{n}{n-1}\frac{(\hat{\mu} - \mu_0)^2}{\hat{s}^2} .
\end{aligned}$$

Il suffit de remarquer que (cf. (3.7))

$$\frac{\hat{\mu} - \mu_0}{\hat{s}/\sqrt{n}} \sim t_{n-1} \text{ (loi de Student à } n-1 \text{ degrés de liberté)},$$

quelque soit σ. Si T_{n-1} suit une loi de Student à $n-1$ degrés de liberté, c' est choisi tel que

$$\begin{aligned}
\alpha &= P_{(\mu_0,\sigma)}\left(\frac{n}{n-1}\frac{(\hat{\mu} - \mu_0)^2}{\hat{s}^2} \geq c'\right) \\
&= P_{(\mu_0,\sigma)}\left(\frac{(\hat{\mu} - \mu_0)^2}{\hat{s}^2/n} \geq (n-1)\,c'\right) \\
&= P\left(|T_{n-1}| \geq \sqrt{(n-1)\,c'}\right) \\
&= 2\,P\left(T_{n-1} \geq \sqrt{(n-1)\,c'}\right) ,
\end{aligned}$$

donc si t est tel que $P(T_{n-1} \geq t) = \alpha/2$ alors $c' = t^2/n - 1$.

Le tableau suivant rassemble, pour trois problèmes différents, les TRV lorsque σ est connu. Ces trois tests se construisent en utilisant la même démarche que ci–dessus.

Test sur μ (σ connu)		
H_0	H_1	On rejette H_0 au niveau α si
$\mu \leq \mu_0$	$\mu > \mu_0$	$\hat{\mu} \geq \mu_0 + \frac{\sigma}{\sqrt{n}}\,z(\alpha)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\hat{\mu} \leq \mu_0 + \frac{\sigma}{\sqrt{n}}\,z(1-\alpha)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert\hat{\mu} - \mu_0\rvert \geq \frac{\sigma}{\sqrt{n}}\,z(\alpha/2)$

où $P(Z \geq z(\alpha)) = \alpha$, $\quad Z \sim N(0,1)$.

Le tableau suivant rassemble, pour trois problèmes différents, les TRV lorsque σ est inconnu. Le dernier de ces tests est celui développé plus haut, les deux autres se construisent d'une manière similaire. Ces tests sont communément appelés tests de Student.

TEST DE STUDENT Test sur μ (σ inconnu)		
H_0	H_1	On rejette H_0 au niveau α si
$\mu \leq \mu_0$	$\mu > \mu_0$	$\hat{\mu} \geq \mu_0 + \frac{\hat{s}}{\sqrt{n}}\,\tau(n-1,\alpha)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\hat{\mu} \leq \mu_0 + \frac{\hat{s}}{\sqrt{n}}\,\tau(n-1,1-\alpha)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert\hat{\mu} - \mu_0\rvert \geq \frac{\hat{s}}{\sqrt{n}}\,\tau(n-1,\alpha/2)$

où $P(T_{n-1} \geq \tau(n-1,\alpha)) = \alpha$, T_{n-1} de loi de Student à $n-1$ degrés de liberté.

3.4.2 Tests Portant sur la Variance

On s'intéresse maintenant aux tests portant sur σ (μ connu et μ inconnu). Ces tests se construisent en utilisant la démarche du paragraphe précédent ainsi que (3.5) et (3.6).

Test sur σ (μ connu)		
H_0	H_1	On rejette H_0 au niveau α si
$\sigma \leq \sigma_0$	$\sigma > \sigma_0$	$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2 \geq \gamma(n,\alpha)$
$\sigma \geq \sigma_0$	$\sigma < \sigma_0$	$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2 \leq \gamma(n,1-\alpha)$
$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$\begin{cases} \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2 \leq \gamma(n,1-\alpha/2) \\ \text{ou} \\ \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2 \geq \gamma(n,1-\alpha/2) \end{cases}$

où $P(\chi_n^2 \geq \gamma(n,\alpha)) = \alpha$, χ_n^2 de loi du χ^2 à n degrés de liberté.

Test sur σ (μ inconnu)		
H_0	H_1	On rejette H_0 au niveau α si
$\sigma \leq \sigma_0$	$\sigma > \sigma_0$	$(n-1)\,\frac{\hat{s}^2}{\sigma_0^2} \geq \gamma(n-1,\alpha)$
$\sigma \geq \sigma_0$	$\sigma < \sigma_0$	$(n-1)\,\frac{\hat{s}^2}{\sigma_0^2} \leq \gamma(n-1,1-\alpha)$
$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$\begin{cases} (n-1)\,\frac{\hat{s}^2}{\sigma_0^2} \leq \gamma(n-1,1-\alpha/2) \\ \text{ou} \\ (n-1)\,\frac{\hat{s}^2}{\sigma_0^2} \geq \gamma(n-1,1-\alpha/2) \end{cases}$

où $P(\chi_{n-1}^2 \geq \gamma(n-1,\alpha)) = \alpha$, χ_{n-1}^2 de loi du χ^2 à $n-1$ degrés de liberté.

3.4.3 Tests D'Indépendance de Deux Echantillons

On considère un échantillon gaussien de la forme

$$(X,Y) = ((X_1,Y_1)\ (X_2,Y_2)\ \cdots\ (X_n,Y_n))\ ,$$

avec

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\ ,\ \begin{pmatrix} \sigma_1^2 & \sigma_1\,\sigma_2\,\rho \\ \sigma_1\,\sigma_2\,\rho & \sigma_2^2 \end{pmatrix}\right)\ ,\quad \begin{cases} \mu_1 & \text{espérance de } X_i, \\ \sigma_1 & \text{variance de } X_i, \\ \mu_2 & \text{espérance de } Y_i, \\ \sigma_2 & \text{variance de } Y_i, \\ \rho & \text{corrélation de } X_i \text{ et } Y_i. \end{cases}$$

L'espace des paramètres est

$$\Theta \triangleq \{\theta = (\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)\ ;\ \mu_1,\mu_2 \in \mathbb{R},\ \sigma_1,\sigma_2 > 0\ ,\ \rho \in [-1,1]\}\ .$$

La loi P_θ de (X,Y) admet la densité

$$p(x,y;\theta) = p(x,y;\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \exp\left(-\frac{1}{2}Q(x,y)\right)\sqrt{2\,\pi\,\sigma_1^2\,\sigma_2^2\,(1-\rho^2)^n}\ ,$$

$(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$, où

$$Q(x,y) = \frac{1}{1-\rho^2}\sum_{i=1}^{n}\left[\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2 - 2\,\rho\left(\frac{x_i-\mu_1}{\sigma_1}\right)\left(\frac{y_i-\mu_2}{\sigma_2}\right) + \left(\frac{y_i-\mu_2}{\sigma_2}\right)^2\right]\ .$$

On rappelle que par définition

$$\rho \triangleq \frac{\boldsymbol{E}(X_i-\mu_1)(Y_i-\mu_2)}{\left(\boldsymbol{E}(X_i-\mu_1)^2\ \boldsymbol{E}(Y_i-\mu_2)^2\right)^2} \in [-1,1]\ ,$$

et X_i et Y_i sont indépendants si et seulement si $\rho = 0$.

Il s'agit de tester l'indépendance des deux échantillons $(X_1 \cdots X_n)$ et $(Y_1 \cdots Y_n)$, i.e. tester

$$H_0\ :\ \text{“}\ \rho = 0\ \text{”}\ .$$

L'espace Θ_0 associé est

$$\Theta_0 \triangleq \{\theta = (\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)\ ;\ \mu_1,\mu_2 \in \mathbb{R},\ \sigma_1,\sigma_2 > 0\ ,\ \rho = 0\}\ .$$

On veut calculer le rapport de vraisemblance

$$\lambda(x,y) \triangleq \frac{\sup_{\theta\in\Theta_0}\ p(x,y;\theta)}{\sup_{\theta\in\Theta}\ p(x,y;\theta)}\ .$$

On a

$$\lambda(x,y) = \frac{p(x,y;\hat\theta_0(x,y))}{p(x,y;\hat\theta(x,y))}\ ,$$

où

$$\begin{cases} \hat\theta_0(x,y) \in \operatorname{Arg}\sup\limits_{\theta\in\Theta_0}\ p(x;\theta) & \text{est l'EMV de } \theta \text{ sous l'hypothèse } H_0, \\ \hat\theta(x,y) \in \operatorname{Arg}\sup\limits_{\theta\in\Theta}\ p(x;\theta) & \text{est l'EMV de } \theta. \end{cases}$$

Après calcul on obtient

$$\hat\theta_0 = (\hat\mu_1,\hat\mu_2,\hat\sigma_1,\hat\sigma_2,0) \quad \text{et} \quad \hat\theta = (\hat\mu_1,\hat\mu_2,\hat\sigma_1,\hat\sigma_2,\hat\rho)\ ,$$

où $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})$ est l'EMV de $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ donné par

$$\hat{\mu}_1 \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i \ , \tag{3.8}$$

$$\hat{\sigma}_1^2 \triangleq \frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{\mu}_1)^2 \ , \tag{3.9}$$

$$\hat{\mu}_2 \triangleq \frac{1}{n}\sum_{i=1}^{n} Y_i \ , \tag{3.10}$$

$$\hat{\sigma}_2^2 \triangleq \frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{\mu}_2)^2 \ , \tag{3.11}$$

$$\hat{\rho} \triangleq \frac{1}{n\,\hat{\sigma}_1\,\hat{\sigma}_2}\sum_{i=1}^{n} (X_i - \hat{\mu}_1)(Y_i - \hat{\mu}_2) \ . \tag{3.12}$$

Ainsi le rapport de vraisemblance est

$$\begin{aligned} \lambda(x,y) &= \frac{1}{(2\,\pi\,\hat{\sigma}_1\,\hat{\sigma}_2)^n}\ \exp(-n)\left(\frac{1}{(2\,\pi\,\hat{\sigma}_1\,\hat{\sigma}_2)^n}\ \frac{1}{(1-\hat{\rho}^2)^{n/2}}\ \exp(-n)\right)^{-1} \\ &= (1-\hat{\rho}^2)^{n/2} \ . \end{aligned}$$

Donc le TRV de H_0 a une région de rejet de la forme

$$D = \{(x,y)\ ;\ |\hat{\rho}(x,y)| \geq c\} \ .$$

Ce résultat est le plus intuitif possible: on accepte l'hypothèse d'indépendance H_0 si l'EMV de la corrélation ρ est suffisament petit. Enfin, pour calculer le seuil c on utilise le résultat suivant

$$\text{pour n grand et sous } P_\theta : Z = \sqrt{n}\ \frac{\hat{\rho} - \rho}{(1-\rho^2)} \sim N(0,1) \ .$$

Donc, sous P_θ avec $\theta \in \Theta_0$, $\sqrt{n}\,\hat{\rho}$ suit une loi $N(0,1)$ (pour n grand), on en déduit le test d'indépendance suivant

TEST D'INDEPENDANCE		
H_0	H_1	On rejette H_0 au niveau α si
$\rho = 0$	$\rho \neq 0$	$\lvert\hat{\rho}\rvert \geq \dfrac{z(\alpha)}{\sqrt{n}}$

où $z(\alpha)$ est tel que $P(|Z| \geq z(\alpha)) = \alpha$, $Z \sim N(0,1)$.

3.5 EXERCICES

Exercice 3.5.1 *D'après une théorie donnée, la loi d'une certaine variable est une loi exponentielle double de moyenne 0 et variance 1 (cf. exercice 2.6.5). Une autre théorie affirme que la loi de cette variable est* $N(0,1)$.

- *Sur la base de n–observations indépendantes, construire un test de 2 hypothèses simples de niveau α pour l'hypothèse que la première théorie est la bonne.*

Comme l'évaluation directe du seuil c est difficile, on supposera n grand et on utilisera une approximation normale pour une somme appropriée des observations.

Exercice 3.5.2 *Soit $X = (X_1 \cdots X_n)$ un n–échantillon d'une loi $N(\mu, \sigma^2)$, σ connu. Pour l'hypothèse H_0 : " $\mu = \mu_0$ " montrer que*

- *le rapport de vraisemblance est*

$$l(x) = \frac{n}{\sigma^2} (\hat{\mu}(x) - \mu_0)^2 \ , \quad x \in \mathbb{R}^n \ ,$$

  où $\hat{\mu}$ est l'EMV de μ.

- *Le TRV au niveau α rejette H_0 si*

$$|\hat{\mu} - \mu_0| \geq \frac{\sigma\, z}{\sqrt{n}} \ ,$$

  où $P(Z \geq z) = \alpha/2$, $(Z \sim N(0,1))$.

Exercice 3.5.3 *Un signal binaire (± 1) discret de longueur n noté*

$$S = (S_1 \cdots S_n)$$

est transmis par l'intermédiaire d'un canal bruité, i.e. on observe

$$X_i = S_i + b_i \ , \quad i = 1 \cdots n \ ,$$

où $S_i = \pm 1$ est la ième composante du signal et b_i est un bruit additif que nous supposerons gaussien: $b_1, \ldots, b_n$ indépendants de même loi $N(0,1)$

- *Etant donnés $X_1 \cdots X_n$ calculer $\hat{S}$ l'EMV de S (réponse: $\hat{S}_i = X_i/|X_i|$).*

- *Utiliser la réponse précédente pour déterminer la forme du TRV de l'hypothèse H_0 : " $S = S^0$ ", où $S^0 = (S_1^0 \cdots S_n^0)$ est un signal donné (réponse:*

$$\sum \{|X_i| \leq c; i \text{ tels que } \hat{S}_i \neq S_i^0\} \ . \qquad)$$

Chapitre 4

METHODES NON–PARAMETRIQUES

Les méthodes paramètriques exposées dans les chapitres précédents supposent toutes que l'on connaisse la forme de la loi de l'observation. Mais dans beaucoup de cas *celle–ci n'est pas connue* et on aimerait donc disposer de techniques statistiques ne dépendant pas de la forme de la loi de l'observation. Ces techniques sont dites *non–paramètriques*. Les notations utilisées dans ce chapitre sont indépendantes de celles utilisées dans les chapitres précédents.

4.1 ESTIMATION NON–PARAMETRIQUE

Supposons que l'on veuille des estimateurs utilisables pour un grand nombre de lois possibles. Le premier problème non trivial est de déterminer ce que l'on veut estimer. Dans un cas paramétrique ce problème est simple: la loi de l'observation appartient à une classe $\{P_\theta; \theta \in \Theta\}$ et on veut estimer le paramètre θ. Dans un cas non–paramètrique on peut vouloir estimer

(1) la loi de l'observation P_X,

(2) les moments de P_X,

(3) les quantiles[1] de P_X.

On se place sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ et on considère

$$X = X_1, \ldots, X_n$$

un n–échantillon d'une loi P sur $\mathbb{R}$. La loi de l'observation est donc $P_X = P^n$. On note F la fonction de répartition[2] associée.

Définition 4.1.1 *La* loi empirique *de l'échantillon $X = X_1, \ldots, X_n$ est la loi sur $\mathbb{R}$ définie par*[3]

$$\hat{P}_n(I) = \frac{1}{n} \# \{X_i \in I\}, \quad \forall I \text{ intervalle de } \mathbb{R} .$$

La fonction de répartition empirique *est définie par*

$$\hat{F}_n(x) = \frac{1}{n} \# \{X_i \leq x\} = \hat{P}_n(]-\infty, x]), \quad \forall x \in \mathbb{R} .$$

$\hat{F}_n$ est la fonction de répartition de la loi empirique $\hat{P}_n$.

1. Soit P une loi sur $\mathbb{R}$ et $\alpha \in [0,1]$, le *quantile* de cette loi est le nombre $x_\alpha \in [-\infty, +\infty]$ tel que $P(]-\infty, x_\alpha]) = \alpha$.
2. Par définition: $F(x) = P(]-\infty, x]) = \mathbb{P}(X_i \leq x), \quad \forall x \in \mathbb{R}$.
3. $\# \{X_i \in I\}$ est le nombre de i tels que $X_i \in I$.

Ainsi $\hat{P}_n$ peut aussi être définie comme la loi discrète qui associe la probabilité $1/n$ à chacun des points X_i (cette dernière définition se géneralise au cas multidimensionnel). On peut aussi remarquer que $\hat{P}_n$ dépend de l'observation $\{X_i\}$, donc $\hat{P}_n$ est aléatoire.

La loi empirique $\hat{P}_n$ est un estimateur de la loi P. En effet, soit I un intervalle de $\mathbb{R}$, si on veut estimer $P(I) = \mathbb{P}(X_i \in I)$ la probabilité pour que X_i appartienne à I, il suffit de calculer

$$\hat{P}_n(I) = \frac{1}{n} \#\{X_i \in I\} .$$

Soit $f : \mathbb{R} \to \mathbb{R}$, si on veut estimer $E\, f = \mathbb{E}\, f(X_i)$ ($\mathbb{E}$ désigne l'espérance associée à $\mathbb{P}$) on calcule

$$\hat{E}_n\, f \triangleq \frac{1}{n} \sum_{i=1}^{n} f(X_i) ,$$

$\hat{E}_n$ est l'espérance associée à $\hat{P}_n$. Pour estimer les moments $\mathbb{E}(X_i)^k$ on utilise les moments empiriques

Définition 4.1.2 *Les* moments empiriques *d'un n–échantillon $X = X_1, \ldots, X_n$ sont par définition les moments de la loi empirique*

$$\hat{\mu}_n^k \triangleq \frac{X_1^k + \ldots + X_n^k}{n} , \quad (k \in \mathbb{N}) .$$

La variance empirique *est la variance de la loi empirique*

$$\hat{\sigma}_n^2 \triangleq \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n^1)^2 .$$

Pour estimer les quantiles de la loi P on introduit les définitions suivantes: soit $X_{(1)}$ le plus petit élément parmi $\{X_1, \ldots, X_n\}$, i.e. $X_{(1)} = \min\{X_i\}$, $X_{(2)}$ le suivant, ainsi de suite

Définition 4.1.3 *On définit*

$$\begin{aligned} X_{(1)} &\triangleq \min_{i=1\cdots n} \{X_i\} , \\ X_{(j)} &\triangleq \min_{i=1\cdots n} \left\{X_i ; X_i \geq X_{(j-1)}\right\} , \end{aligned}$$

i.e. $X_{(j)}$ est le $j^{ième}$ plus petit élément parmi $\{X_1, \ldots, X_n\}$. $X_{(j)}$ est appelée la $j^{ième}$ statistique d'ordre *de l'échantillon $X = X_1, \ldots, X_n$.*

Pour estimer x_α le quantile d'ordre α de la loi P on utilise le quantile empirique défini par

Définition 4.1.4 *Soit $\alpha \in [0, 1]$, le* quantile empirique *d'ordre α de l'échantillon $X = X_1, \ldots, X_n$ est défini par*

$$\hat{x}_\alpha \triangleq X_{([n\alpha]+1)} ,$$

où $[n\,\alpha]$ désigne la partie entière de $n\,\alpha$.

Les propriétés de ces estimateurs sont rassemblées dans les deux résultats suivants

Théorème 4.1.5

(i) Pour tout I intervalle de $\mathbb{R}$

$$\mathbb{E}\, \hat{P}_n(I) = P(I) \quad , \quad \mathbb{E}\left(\hat{P}_n(I) - P(I)\right)^2 = \frac{1}{n} P(I)\,(1 - P(I)) .$$

(ii) Pour tout $x \in \mathbb{R}$

$$\mathbb{E}\,\hat{F}_n(x) = F(x) \quad , \quad \mathbb{E}\left(\hat{F}_n(x) - F(x)\right)^2 = \frac{1}{n}\,F(x)\,(1 - F(x)) \ .$$

(iii) Pour tout $f : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}\left(\hat{E}_n\, f\right) = E\, f \quad , \quad \mathbb{E}\left(\hat{E}_n\, f - E\, f\right)^2 = \frac{1}{n}\, Var(f(X_i)) \ .$$

(iv) Soit $p(x)$ la densité de la loi P, alors pour tout $\alpha \in [0,1]$ et pour n grand

$$\mathbb{E}\,\hat{x}_\alpha \simeq x_\alpha \quad , \quad \mathbb{E}\,(\hat{x}_\alpha - x_\alpha)^2 \simeq \frac{1}{n}\,\frac{\alpha\,(1-\alpha)}{p(x_\alpha)} \ .$$

A l'aide de ce dernier résultat on peut montrer le

Théorème 4.1.6 *Pour tout I intervalle de $\mathbb{R}$, $x \in \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$ et $\alpha \in [0,1]$*

$$\begin{array}{lll}
\hat{P}_n(I) & \underset{n\to\infty}{\to} P(I) & \mathbb{P}\text{–}p.s. \ , \\
\hat{F}_n(x) & \underset{n\to\infty}{\to} F(x) & \mathbb{P}\text{–}p.s. \ , \\
\hat{E}_n\, f & \underset{n\to\infty}{\to} E\, f & \mathbb{P}\text{–}p.s. \ , \\
\hat{x}_\alpha & \underset{n\to\infty}{\to} x_\alpha & \mathbb{P}\text{–}p.s. \ .
\end{array}$$

Un des résultats importants concernant la fonction de répartition empirique est le suivant

Théorème 4.1.7 (Théorème de Glivenko–Cantelli)

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| \underset{n\to\infty}{\to} 0 \ , \quad \mathbb{P}\text{–}p.s. \ .$$

Ces trois théorèmes sont vrais quelle que soit la loi de X_i. On peut donc utiliser ces estimateurs dans toutes les situations sans se soucier de la forme de la loi de l'observation.

4.2 TESTS NON–PARAMETRIQUES

Dans un cadre paramétrique $(\Omega, \mathcal{F}, \mathbb{P}, \{P_\theta; \theta \in \Theta\})$ les problèmes de test portent toujours sur le paramètre θ. Il existe pourtant des questions que l'on peut se poser sur un échantillon qui ne demandent pas de spécifier un modèle paramétrique, comme par exemple

(1) L'échantillon $X_1 \cdots X_n$ provient–il d'une loi P donnée?

(2) Les échantillons $X_1 \cdots X_n$ et $Y_1 \cdots Y_n$ proviennent–ils d'une même loi?

(3) Les échantillons $X_1 \cdots X_n$ et $Y_1 \cdots Y_n$ sont–ils indépendants?

Dans les trois paragraphes suivants nous allons présenter des outils statistiques permettant de répondre à ces questions sans avoir à préciser un modèle paramétrique.

4.2.1 Test D'Adéquation

On dispose d'un n–échantillon $X = X_1, \ldots, X_n$, i.e. les X_i sont des variables indépendantes et identiquement distribuées définies sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$. On se donne une loi de probabilité P sur $\mathbb{R}$ et on s'intéresse au problème de test de l'hypothèse

$$\mathcal{H} \quad : \quad \text{“ } X_1, \ldots, X_n \text{ est un échantillon de la loi } P \text{ ”}$$

Test d'adéquation de Kolmogorov–Smirnov

Soit $F : \mathbb{R} \to [0,1]$ la fonction de répartition de la loi P. On sait construire la fonction de répartition empirique de l'échantillon $X_1, \ldots, X_n$

$$\hat{F}_n(x) = \frac{1}{n} \# \{X_i \leq x\} \ .$$

Pour tester l'hypothèse $\mathcal{H}$, une méthode simple consiste à comparer F et $\hat{F}_n$.

Définition 4.2.1 *Le* test d'adéquation de Kolmogorov–Smirnov *consiste à accepter l'hypothèse $\mathcal{H}$ si*

$$\max_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| \leq c \ ,$$

pour un seuil c convenablement choisi.

Posons

$$d = \max_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \ .$$

Supposons que la loi P admette une densité continue sur $\mathbb{R}$. Sous l'hypothèse $\mathcal{H}$, on peut alors montrer que la loi de d ne dépend pas de la loi des X_i (qui est P sous $\mathcal{H}$). On peut alors calculer le seuil c en fonction d'un niveau $\alpha \in [0,1]$ fixé grâce au tableau

n	$\alpha = 0.90$	$\alpha = 0.95$
10	0.37	0.41
20	0.26	0.29
30	0.22	0.24

Par exemple, pour $n = 10$, $\mathbb{P}(d \leq 0.37) = 0.90$. L'inconvénient de ce test réside dans le fait que le calcul de d est difficile. Il existe une méthode plus simple.

Test d'adéquation du χ^2

On se donne $k+1$ points $-\infty \leq a_1 < a_2 < \cdots < a_{k+1} \leq \infty$ et on pose

$$\begin{aligned} I_j &\triangleq [a_j, a_{j+1}[\ , \quad j = 1 \cdots k \ , \\ n_j &\triangleq n \ P(I_j) \ , \\ \hat{n}_j &\triangleq \# \ \{X_i \in I_j\} \ . \end{aligned}$$

On définit alors le test suivant

Définition 4.2.2 *Le* test d'adéquation du χ^2 *consiste à accepter l'hypothèse $\mathcal{H}$ si*

$$\sum_{j=1}^{k} \frac{(\hat{n}_j - n_j)^2}{n_j} \leq c \ ,$$

pour un seuil c convenablement choisi.

Posons

$$d = \sum_{j=1}^{k} \frac{(\hat{n}_j - n_j)^2}{n_j} \ ,$$

et plaçons–nous sous l'hypothèse $\mathcal{H}$. Pour n grand, on peut montrer que d suit une loi du χ^2 à $k-1$ degrés de liberté. Soit $\alpha \in [0,1]$ donné, pour que le test de la définition 4.2.2 soit de niveau α, il suffit de choisir c comme étant le $(1-\alpha)$–quantile de la loi du χ^2 à $k-1$ degrés de liberté.

L'inconvénient majeur de ce dernier test réside dans le fait qu'il dépend de la partition de $\mathbb{R}$ que l'on se donne.

4.2.2 Comparaison de Deux Echantillons

Sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$ on considère deux échantillons indépendants de taille n

$$\begin{array}{rl} X_1, X_2, \ldots, X_n & n\text{–échantillon d'une loi } P \ , \\ Y_1, Y_2, \ldots, Y_n & n\text{–échantillon d'une loi } Q \ . \end{array}$$

On suppose que les lois admettent des densités continues. On veut savoir si ces deux échantillons proviennent de la même loi, en d'autres termes tester l'hypothèse

$$\mathcal{H} \quad : \quad \text{“ } P = Q \text{ ”}$$

Test du Signe

Posons

$$V_i = X_i - Y_i \ . \tag{4.1}$$

Proposition 4.2.3 *Soient X et Y deux variables aléatoires indépendantes. Si elles possèdent la même loi alors $V = X - Y$ admet une loi symétrique en 0.*

Ainsi sous l'hypothèse $\mathcal{H}$, les variables V_i définies par (4.1), admettent une loi dont la médiane[4] est 0, d'où

$$\mathbb{P}(V_1 > 0) = \cdots = \mathbb{P}(V_n > 0) = \frac{1}{2} \ . \tag{4.2}$$

Considérons maintenant le signe des variables V_i. Le test du signe est fondé sur le fait que, lorsque (4.2) est vrai, les signes "+" et les signes "–" se répartissent comme le nombre de "faces" parmi n jets d'une pièce non biaisée. Donc si S_n est le nombre de V_i positifs

$$S_n \triangleq \# \ \{V_i > 0\} \ , \tag{4.3}$$

alors la loi de S_n est celle du nombre de faces parmi n jets d'une pièce non biaisée.

Définition 4.2.4 *Le* test du signe *consiste à accepter l'hypothèse $\mathcal{H}$ lorsque*

$$\left| \frac{S_n}{n} - \frac{1}{2} \right| \leq c \ ,$$

où c est un seuil convenablement choisi.

Pour calculer le seuil c on peut soit utiliser une table de la loi binomiale ou, lorsque n est grand, utiliser l'approximation normale.

Test de Wilcoxon

Le test du signe est très simple à mettre en œuvre mais n'est pas trés précis (faible niveau de détection). Le test suivant donne de meilleurs résultats. On définit

$$U_i \triangleq |V_i| = |X_i - Y_i| \ , \quad i = 1 \cdots n \ . \tag{4.4}$$

On range les U_i en ordre croissant, i.e. on calcule

$$U_{(1)} < U_{(2)} < \cdots < U_{(n)} \ , \tag{4.5}$$

et

$$\left\{ \begin{array}{lcl} U^+ & \triangleq & \text{somme des rangs des } V_i \text{ positifs dans la suite (4.5)}, \\ U^- & \triangleq & \text{somme des rangs des } V_i \text{ négatifs dans la suite (4.5)}. \end{array} \right. \tag{4.6}$$

Exemple 4.2.5 *Supposons par exemple que l'on obtienne*

V_i	*–3.1*	*+2.5*	*+4.2*	*–0.8*	*+2.7*	*–5.3*	*–0.2*	*+2.3*	*+1.8*	*+6.5*

4. La médiane ν d'une loi P est le quantile d'ordre 1/2 de cette loi: $P(]-\infty, \nu]) = 1/2$.

On range par ordre croissant la valeur absolue de ces nombres

signe	$-$	$-$	$+$	$+$	$+$	$+$	$-$	$+$	$-$	$+$
$\lvert V_i \rvert$	*0.2*	*0.8*	*1.8*	*2.3*	*2.5*	*2.7*	*3.1*	*4.2*	*5.3*	*6.5*
rang	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*

Les rangs des V_i positifs sont 3,4,5,6,8,10; les rangs des V_i négatifs sont 1,2,7,9. Donc

$$U^+ = 36 \quad , \quad U^- = 19 \ .$$

Il est facile de vérifer que U^+ est à valeurs dans $\{0, 1, \ldots, N\}$ où $N = n(n+1)/2$.

Définition 4.2.6 *Le* test de Wilcoxon *consiste à accepter $\mathcal{H}$ si*

$$k < U^+ < N - k \ ,$$

où k est un seuil convenablement choisi.

On se place sous l'hypothèse $\mathcal{H}$, on se donne $\alpha \in [0, 1]$ et on veut calculer k tel que

$$\mathbb{P}\left(k < U^+ < N - k\right) = \alpha \ .$$

En pratique on prend k tel que $\mathbb{P}(T > k) = \alpha$ où T est une variable gaussienne d'espérance $n(n+1)/4$ et de variance $n(N = 1)(2n+1)/24$.

Test de Kolmogorov–Smirnov

On considère les fonctions de répartition empiriques associées aux deux échantillons

$$\hat{F_n}^X(x) = \frac{1}{n} \# \{X_i \leq x\} \quad , \quad \hat{F_n}^Y(x) = \frac{1}{n} \# \{Y_i \leq x\} \ .$$

Pour tester $\mathcal{H}$ on peut comparer $\hat{F_n}^X$ et $\hat{F_n}^Y$:

Définition 4.2.7 *Le* test de Kolmogorov–Smirnov *consiste à accepter l'hypothèse $\mathcal{H}$ si*

$$\max_{x \in \mathbb{R}} \left| \hat{F_n}^X(x) - \hat{F_n}^Y(x) \right| \leq c \ ,$$

où c est un seuil convenablement choisi.

Sous l'hypothèse $\mathcal{H}$ $(P = Q)$ on peut montrer que la loi de la variable aléatoire

$$d \triangleq \max_{x \in \mathbb{R}} \left| \hat{F_n}^X(x) - \hat{F_n}^Y(x) \right| \ ,$$

ne dépend pas de la loi sous–jacente P. On peut donc supposer que $P = Q$ =loi uniforme sur $[0, 1]$. Ceci permet de calculer le seuil c en fonction d'un niveau $\alpha \in [0, 1]$ donné, ainsi on acceptera l'hypothèse au niveau

$$\begin{array}{lll} \alpha = 0.10 & \text{si} & \leq \sqrt{2/n} \times 1.22 \ , \\ \alpha = 0.05 & \text{si} & \leq \sqrt{2/n} \times 1.36 \ , \\ \alpha = 0.01 & \text{si} & \leq \sqrt{2/n} \times 1.63 \ . \end{array}$$

Test du χ^2

Supposons dans un premier temps que les X_i et les Y_i soient à valeurs dans l'ensembel $\{1, 2, \ldots, J\}$. Posons

$$p_j^X \triangleq \mathbb{P}(X_i = j) \ , \quad p_j^Y \triangleq \mathbb{P}(Y_i = j) \ , \quad j = 1 \cdots J \ .$$

Les lois P et Q sont égales si et seulement si $p_j^X - p_j^Y = 0$, $j = 1 \cdots J$. Comme ces quantités sont inconnues, on utilise les estimateurs

$$\left\{ \begin{array}{l} \hat{p}_j^X \triangleq \frac{\hat{n}_j^X}{n} \ , \\ \hat{p}_k^Y \triangleq \frac{\hat{n}_k^Y}{n} \ , \end{array} \right. \quad \text{où} \quad \left\{ \begin{array}{l} \hat{n}_j^X \triangleq \# \ \{X_i = j\} \ , \\ \hat{n}_k^Y \triangleq \# \ \{Y_i = k\} \ . \end{array} \right.$$

Définition 4.2.8 *Le* test du χ^2 *consiste à accepter l'hypothèse $\mathcal{H}$ si*

$$\sum_{j=1}^{J} \frac{\left(\hat{n}_j^X - \hat{n}_j^Y\right)^2}{\hat{n}_j^X + \hat{n}_j^Y} \leq c \ ,$$

où c est un seuil convenablement choisi.

On peut montrer que sous l'hypothèse $\mathcal{H}$ $(P = Q)$, l'expression

$$d \triangleq \sum_{j=1}^{J} \frac{\left(\hat{n}_j^X - \hat{n}_j^Y\right)^2}{\hat{n}_j^X + \hat{n}_j^Y} \ ,$$

suit une loi du χ^2 à $J - 1$ degrés de liberté. Etant donné $\alpha \in [0, 1]$, pour que le test de la définition 4.2.8 soit de niveau α, il suffit de prendre c comme étant le $(1 - \alpha)$–quantile de la loi du χ^2 à $(J - 1)$ degrés de liberté.

4.2.3 Test D'Indépendance du χ^2

Sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$, on considère deux échantillons de taille n

$$X_1, X_2, \ldots, X_n \quad \text{et} \quad Y_1, Y_2, \ldots, Y_n \ . \tag{4.7}$$

Dans le paragraphe précédent on s'est intéressé aux tests permettant de vérifier si ces deux échantillons proviennent d'une même loi. Mais avant d'utiliser de telles méthodes il convient de vérifier si les deux échantillons sont indépendants. On s'intéresse donc au test de l'hypothèse

$$\mathcal{H} \ : \ \text{“} \ \{X_i \, ; \, i = 1 \cdots n\} \text{ indépendant de } \{Y_i \, ; \, i = 1 \cdots n\} \ \text{”}$$

Supposons que les X_i soient de variables aléatoires discrètes et prennent leur valeurs dans $\{1, 2, \ldots, J\}$ et les Y_i dans $\{1, 2, \ldots, K\}$. Posons

$$\begin{array}{rcl} p_{jk} & \triangleq & \mathbb{P}(X_i = j, Y_i = k) \ , \\ p_j^X & \triangleq & \mathbb{P}(X_i = j) \ , \\ p_k^Y & \triangleq & \mathbb{P}(Y_i = k) \ . \end{array} \tag{4.8}$$

Les X_i et Y_i sont indépendants si et seulement si $p_{jk} - p_j^X \, p_k^Y = 0$ pour tout $j = 1 \cdots J$ et $k = 1 \cdots K$. Donc pour tester $\mathcal{H}$ il suffit de considérer la différence $p_{jk} - p_j^X \, p_k^Y$, mais les termes qui la composent n'étant pas connus, nous sommes donc amenés à les estimer. Des estimateurs simples de p_{jk}, p_j^X, p_k^Y sont

$$\left\{ \begin{array}{l} \hat{p}_{jk} \triangleq \frac{\hat{n}_{jk}}{n} \ , \\ \hat{p}_j^X \triangleq \frac{\hat{n}_j^X}{n} \ , \\ \hat{p}_k^Y \triangleq \frac{\hat{n}_k^Y}{n} \ , \end{array} \right. \quad \text{où} \quad \left\{ \begin{array}{l} \hat{n}_{jk} \triangleq \# \ \{(X_i, Y_i) = (j, k)\} \ , \\ \hat{n}_j^X \triangleq \# \ \{X_i = j\} \ , \\ \hat{n}_k^Y \triangleq \# \ \{Y_i = k\} \ . \end{array} \right. \tag{4.9}$$

Définition 4.2.9 *(Cas discret) Soient $\{X_i\}$ et $\{Y_i\}$ deux échantillons de taille n, X_i à valeurs dans $\{1,\ldots,J\}$ et Y_i à valeurs dans $\{1,\ldots,K\}$. Le* test d'indépendance du χ^2 *consiste à accepter l'hypothèse $\mathcal{H}$ si*

$$n \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(\hat{n}_{jk} - \hat{n}_j^X \, \hat{n}_k^Y\right)^2}{\hat{n}_j^X \, \hat{n}_k^Y} \leq c \ , \tag{4.10}$$

pour un seuil c convenablement choisi, où les termes $\hat{p}_{jk}$, $\hat{p}_j^X$, $\hat{p}_k^Y$ sont définis par (4.9).

Supposons que les X_i et les Y_i soient des variables aléatoires à valeurs dans $\mathbb{R}$. On peut utiliser la même démarche que ci–dessus: on se donne deux partitions en intervalles de $\mathbb{R}$: $\{I_j^X; j = 1\ldots J\}$ et $\{I_k^Y; k = 1\ldots K\}$; on pose

$$\begin{aligned} p_{jk} &\triangleq \mathbb{P}\left(X_i \in I_j^X, Y_i \in I_k^Y\right) \ , \\ p_j^X &\triangleq \mathbb{P}\left(X_i \in I_j^X\right) \ , \\ p_k^Y &\triangleq \mathbb{P}\left(Y_i \in I_k^Y\right) \ . \end{aligned} \tag{4.11}$$

Si les X_i et Y_i sont indépendants, alors $p_{jk} - p_j^X \, p_k^Y = 0$ pour tout $j = 1 \cdots J$ et $k = 1 \cdots K$. Donc, comme ci–dessus, pour tester $\mathcal{H}$ on considére la différence $p_{jk} - p_j^X \, p_k^Y$, dont les termes qui la composent sont estimés par

$$\left\{ \begin{array}{l} \hat{p}_{jk} \triangleq \frac{\hat{n}_{jk}}{n} \ , \\ \hat{p}_j^X \triangleq \frac{\hat{n}_j^X}{n} \ , \\ \hat{p}_k^Y \triangleq \frac{\hat{n}_k^Y}{n} \ , \end{array} \right. \quad \text{où} \quad \left\{ \begin{array}{l} \hat{n}_{jk} \triangleq \# \ \left\{(X_i, Y_i) \in I_j^X \times I_k^Y\right\} \ , \\ \hat{n}_j^X \triangleq \# \ \left\{X_i \in I_j^X\right\} \ , \\ \hat{n}_k^Y \triangleq \# \ \left\{Y_i \in I_k^Y\right\} \ . \end{array} \right. \tag{4.12}$$

Définition 4.2.10 *(Cas continu) Soient $\{X_i\}$ et $\{Y_i\}$ deux échantillons de taille n à valeurs dans $\mathbb{R}$; $\{I_j^X; j = 1\ldots J\}$ et $\{I_k^Y; k = 1\ldots K\}$ deux partitions de $\mathbb{R}$ en intervalles. Le* test d'indépendance du χ^2 *consiste à accepter l'hypothèse $\mathcal{H}$ si*

$$n \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(\hat{n}_{jk} - \hat{n}_j^X \, \hat{n}_k^Y\right)^2}{\hat{n}_j^X \, \hat{n}_k^Y} \leq c \ , \tag{4.13}$$

pour un seuil c convenablement choisi où les termes $\hat{p}_{jk}$, $\hat{p}_j^X$, $\hat{p}_k^Y$ sont définis par (4.12).

Sous l'hypothèse $\mathcal{H}$ on peut montrer que l'expression

$$d \triangleq n \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(\hat{n}_{jk} - \hat{n}_j^X \, \hat{n}_k^Y\right)^2}{\hat{n}_j^X \, \hat{n}_k^Y} \ ,$$

suit une loi du χ^2 à $(J-1)\,(K-1)$ degrés de liberté. Soit $\alpha \in [0,1]$, pour que le test de la définition 4.2.10 soit de niveau α, il suffit de choisir c comme étant le $(1-\alpha)$–quantile de la loi du χ^2 à $(J-1)\,(K-1)$ degrés de liberté.

Annexe A

Résumé de théorie de la mesure

A.1 Espace mesurable

Définition A.1.1 *Soit E un ensemble et $\mathcal{A}$ un ensemble de parties de E, $\mathcal{A}$ est appelé* tribu *si*

(i) $E \in \mathcal{A}$,

(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,

(iii) $\{A_n; n \in \mathbb{N}\} \subset \mathcal{A} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Le doublet $(E, \mathcal{A})$ est appelé espace mesurable, *les éléments de $\mathcal{A}$ sont les ensembles mesurables.*

Exemple A.1.2 *Soit $\mathcal{C}$ un ensemble de parties de E, la tribu* engendrée *par $\mathcal{C}$, notée $\sigma(\mathcal{C})$, est la plus petite tribu contenant $\mathcal{C}$. Celle-ci est définie par*

$$\sigma(\mathcal{C}) \triangleq \bigcap \{\mathcal{A}; \mathcal{A} \text{ tribu sur } E \text{ contenant} \mathcal{C}\} \ ,$$

en effet, $\sigma(\mathcal{C})$ ainsi définie est une tribu. $\mathcal{C}$ est aussi appelée famille génératrice *de $\mathcal{A}$.*

Exemple A.1.3 *La tribu des* boréliens *de $\mathbb{R}^d$, notée $\mathcal{B}_d$, est la tribu engendrée par les ensembles ouverts de $\mathbb{R}^d$. C'est aussi la tribu engendrée par les pavés de la forme*

$$\prod_{i=1}^{d}]-\infty, a_i] \ .$$

A.2 Mesure

Soit $(E, \mathcal{A})$ un espace mesurable.

Définition A.2.1 *Un* mesure positive *définie sur $(E, \mathcal{A})$ est une application*

$$m \ : \ \mathcal{A} \to \bar{\mathbb{R}}_+$$

*qui possède la propriété de σ–*additivité *: pour toute famille dénombrable $\{A_n; n \in \mathbb{N}\}$ d'éléments deux à deux disjoints de $\mathcal{A}$*

$$m\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} m\left(A_n\right) \ .$$

Le triplet $(E, \mathcal{A}, m)$ s'appelle espace mesuré.

Une mesure m est dite finie *si $m(E) < \infty$, σ–*finie *si E est union d'une famille dénombrable d'ensembles mesurables de mesure finie.*

Exemple A.2.2

(1) La mesure de Dirac *au point a, notée δ_a, est définie par*

$$\delta_a(A) = \mathbf{1}_A(a) \ .$$

(2) Un mesure m est dite discrète *s'il existe des poids $\{\alpha_n; n \in \mathbb{N}\}$ $(\alpha_n > 0)$ et des points $\{a_n; n \in \mathbb{N}\}$ tels que*

$$m = \sum_{n \in \mathbb{N}} \alpha_n \, \delta_{a_n} \ .$$

(3) La mesure de Lebesgue *sur $\mathbb{R}^d$ est l'unique mesure σ–finie sur $\mathbb{R}^d$ telle que*

$$m\left(\prod_{i=1}^{d} [a_i, b_i]\right) = \prod_{i=1}^{d} (b_i - a_i) \ .$$

(4) Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction croissante continue à droite. Il existe une unique mesure σ–finie sur $\mathbb{R}$ telle que

$$m(]a, b]) = F(b) - F(a) \ .$$

Cette mesure est appelée mesure de Stieltjes.

Toute fonction F telle que $m(]a, b]) = F(b) - F(a)$ est appelée primitive *de m.*

A.3 Applications mesurables

Définition A.3.1 *Soient $(E, \mathcal{A})$ et $(E', \mathcal{A}')$ deux espaces mesurables. Une application*

$$f \, : \, (E, \mathcal{A}) \to (E', \mathcal{A}')$$

est dite mesurable *si*

$$f^{-1}(\mathcal{A}') \subset \mathcal{A}$$

(i.e. $\forall A' \in \mathcal{A}'$, $\exists A \in \mathcal{A}$ tel que $f^{-1}(A') = A$). Le triplet $(E, \mathcal{A}, m)$ est appelé espace mesuré.

Proposition A.3.2 *Soit $f : (E, \mathcal{A}) \to (E', \mathcal{A}')$ une application mesurable. $f^{-1}(\mathcal{A}')$ est une tribu de parties de E et si $\mathcal{C}'$ est une famille génératrice de $\mathcal{A}'$ alors $f^{-1}(\mathcal{C}')$ est une famille génératrice de $\mathcal{A}$. La famille $f^{-1}(\mathcal{C}')$ est dite engendrée par f.*

Exemple A.3.3 *Si $(E, \mathcal{A})$ et $(E', \mathcal{A}')$ sont des espaces boréliens, une application mesurable $f : (E, \mathcal{A}) \to (E', \mathcal{A}')$ est appelée application borélienne. Les applications continues sont boréliennes.*

Proposition A.3.4

(i) Soient f et g deux applications mesurables à valeurs dans $\mathbb{R}^d$ et α et β deux réels quelconques, alors $\alpha\, f + \beta\, g$ est mesurable.

(ii) Soient f et g deux applications mesurables à valeurs dans $\mathbb{R}$, $f \times g$ est mesurable.

(iii) Soit $\{f_n; n \in \mathbb{N}\}$ une famille dénombrable de fonctions mesurables à valeurs dans $\bar{\mathbb{R}}$, $\sup_{n \in \mathbb{N}} f_n$ et $\inf_{n \in \mathbb{N}} f_n$ sont mesurables.

(iv) Toute fonction mesurable positive (resp. mesurable positive bornée) peut s'écrire comme limite simple (resp. uniforme) d'une suite croissante $\{f_n; n \in \mathbb{N}\}$ de fonctions mesurables de la forme élémentaire *suivante*

$$f_n = \sum_{i \in I_n} \alpha_i \, \mathbf{1}_{A_n^i}$$

où I_n est fini, $A_n^i \in \mathcal{A}$ et $\alpha_i > 0$.

A.4 Intégrale

Soit $(E, \mathcal{A}, m)$ un espace mesuré avec une mesure m σ–finie.

A.4.1 Intégrale de fonctions mesurables positives

Sur l'ensemble des fonctions f mesurables à valeurs dans $\bar{\mathbb{R}}_+$ on associe un nombre

$$f \to \mathcal{I}(f) = \int f\, dm \in \bar{\mathbb{R}}_+ \ . \tag{A.1}$$

Théorème A.4.1 *Il existe une unique application (A.1) telle que*

(i) $\mathcal{I}$ *est seni–linéaire,*

(ii) $\forall A \in \ , \ \mathcal{I}(A) = m(A)$,

(iii) pour toute suite croissante $\{f_n; n \in \mathbb{N}\}$ *de fonctions mesurables positives*

$$\int \Big(\lim_{n\to\infty} \uparrow f_n \Big)\, dm = \lim_{n\to\infty} \uparrow \int f_n\, dm \ .$$

Exemple A.4.2 *Procédés de calcul :*

(1) Si m *est une mesure discrète* $\sum_{i=1}^n \alpha_i\, \delta_{a_i}$ *on a*

$$\int f\, dm = \sum_{i=1}^n \alpha_i\, f(a_i) \ .$$

(2) Si m *est une mesure quelconque, et si* f *est une fonction mesurable élémentaire*

$$f = \sum_{i=1}^n \alpha_i\, \mathbf{1}_{A_i}$$

alors

$$\int f\, dm = \sum_{i=1}^n \alpha_i\, m(A_i) \ .$$

Cet exemple et la proposition A.3.4 (iv) donne une méthode de calcul approchéde n'importe quelle intégrale.

(3) Si λ *est la mesure de Lebesgue sur* $(\mathbb{R}^d, \mathcal{B}_d)$, *l'intégrale* $\mathcal{I}$ *définie ci–dessus coïncide avec l'intégrale de Reimann.*

(4) Si m_F *est la mesure de Stieltjes associée à la fonction croissante* F, *l'intégrale* $\int f\, dm_F$ *est appelée intégrale de Stieltjes de* f *par rapport à* F *et est aussi notée* $\int f\, dF$.

Théorème A.4.3 (Fatou) *Pour toute suite* $\{f_n; n \in \mathbb{N}\}$ *de fonctions mesurables positives*

$$\int \Big(\liminf_{n\to\infty} f_n \Big)\, dm = \liminf_{n\to\infty} \int f_n\, dm \ .$$

A.4.2 Intégrable de fonctios mesurables réelles

Définition A.4.4 *Un fonction mesurable positive f est dite* intégrable *si*

$$\int f\, dm < \infty \ .$$

Une fonction mesurable f est dite intégrable *si*

$$\int |f|\, dm < \infty \ ,$$

et on définit son intégrale par

$$\int f\, dm = \int f^+\, dm - \int f^-\, dm \ .$$

On note $\mathcal{L}^1(E, \mathcal{A}, m)$ l'ensemble des fonctions réelles intégrables par rapport à m. C'est un espace vectoriel sur $\mathbb{R}$ sur lequel $f \to \int |f|\, dm$ définit une semi–norme. L'application $f \to \int f\, dm$ est une forme linéaire continue pour cette semi–norme.

A.4.3 Espaces $L^p(E, \mathcal{A}, m)$

Définition A.4.5 *Une partie B de E est dite* m–negligeable *si*

$$\exists A \in \mathcal{A},\ m(A) = 0 \ \textit{tel que}\ B \subset A \ .$$

Une fonction f définie sur E est dite m–negligeable *si elle est nulle en dehors d'un ensemble négligeable.*

Proposition A.4.6 *Une application mesurable f est m–négligeable si et seulement si*

$$\int |f|\, dm = 0 \ .$$

Une propriété sur les points de E est dite *vraie m–presque partout* (m–p.p.) si elle est vraie sur tous les points de E sauf sur un ensemble m–négligeable.

On note $\mathcal{L}^p(E, \mathcal{A}, m)$ l'espace vectoriel des fonctions réelles mesurables f telles que $\int |f|^p\, dm < \infty$. Dans $\mathcal{L}^p(E, \mathcal{A}, m)$ la relation d'équivalence

$$f \sim g \Longleftrightarrow f - g \text{ est } m\text{–négligeable}$$

est compatible avec la structure d'espace vectoriel et la passage à l'espace quotient transforme la semi–norme

$$f \to \int |f|\, dm$$

en norme et la forme bilinéaire positive

$$(f, g) \to \int f.g\, dm$$

en produit scalaire.

On note $L^p(E, \mathcal{A}, m)$ l'espace des calsses d'équivalence dans $\mathcal{L}^p(E, \mathcal{A}, m)$.

$L^p(E, \mathcal{A}, m)$ muni de la norme

$$\| f \|_p \triangleq \left(\int |f|^p\, dm \right)^{1/p}$$

est un espace de Banach.

$L^2(E, \mathcal{A}, m)$ muni du produit scalaire

$$< f, g >= \int f.g\, dm$$

est un espace de Hilbert.

Théorème A.4.7 (Théorème de convergence dominée) *Soit $\{f_n; n \in \mathbb{N}\}$ une suite de fonction réelles convergeant m–p.p. vers une fonction f. On suppose qu'il existe $g \in L^p(E, \mathcal{A}, m)$ telle que $|f| \leq g$ m–p.p., alors les fonctions f_n et f sont dans $L^p(E, \mathcal{A}, m)$ et*

$$\int \left(\lim_{n\to\infty} f_n\right) dm = \lim_{n\to\infty} \int f_n \, dm = \int f \, dm \ .$$

A.5 Exemples

A.5.1 Mesures définies par des densités

Soit f une application mesurable réelle. Pour tout $A \in \mathcal{A}$ on note

$$\int_A f \, dm \triangleq \int \mathbf{1}_A \, f \, dm \ .$$

Si f est positive

$$A \to \int_A f \, dm$$

définit une mesure positive, notée $f.m$, sur $(E, \mathcal{A})$, cette mesure est bornée si et seulement si $f \in \mathcal{L}^1(E, \mathcal{A}, m)$.

Soit ν la mesure positive $f.m$, une fonction mesurable g est ν–intégrable si et seulement si $|g| \, f$ est m–intégrable et dans ce cas

$$\int g \, d\nu = \int g \, f \, dm \ .$$

A.5.2 Mesures images

Soient $T : (E, \mathcal{A}) \to (E', \mathcal{A}')$ une application mesurable. L'application

$$A' \to m(T^{-1}(A'))$$

est une mesure sur $(E', \mathcal{A}')$ appelée *mesure image* de m par T.

Soit μ cette mesure image de m par T, une fonction $g : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ mesurable est μ–intégrable si et seulement si $g \circ T$ est m–intégrable et

$$\int g \, d\mu = \int g \circ T \, dm \ .$$

A.5.3 Mesures produits

Soient $(E_1, \mathcal{A}_1, m_1)$ et $(E_2, \mathcal{A}_2, m_2)$ deux espaces mesurés, m_1 et m_2 deux mesures finies.

On note $\mathcal{A}_1 \otimes \mathcal{A}_2$ la tribu engendrée sur $E_1 \times E_2$ par les rectangles mesurables $A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$.

Il existe une unique mesure positive σ–finie, notée $m_1 \times m_2$, sur $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ telle que pour tout rectangle mesurable $A_1 \times A_2$

$$m_1 \times m_2(A_1 \times A_2) = m_1(A_1) \, m_2(A_2) \ .$$

Cette mesure est appelée *mesure produit.*

Dans ce cas la régle de calcul provient du théorème de Fubini. Si f, définie sur $E_1 \times E_2$, est mesurable alors l'application $x_1 \to \int f(x_1, x_2) \, dm_2(x_2)$ est mesurable sur $(E_2, \mathcal{A}_2)$ et

$$\int f \, d(m_1 \times m_2) = \int \left(\int f(x_1, x_2) \, dm_2(x_2)\right) dm_1(x_1) \ .$$

Annexe B

EXERCICES CORRIGES ET PROBLEMES

B.1 Correction des Exercices

Exercice 2.6.3

La densité de l'échantillon est

$$p(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta^{2n}\, x_1 \times \cdots \times x_n\, e^{-\theta\,(x_1+\cdots+x_n)}$$

On a

$$\log\, p(x;\theta) = 2n\, \log(\theta) + (\log(x_1) + \cdots + \log(x_n)) - \theta\,(x_1 + \cdots + x_n)\ .$$

L'équation de la vraisemblance admet donc comme solution

$$\hat{\theta}(x) = 2\,n \times \left(\sum_{i=1}^{n} x_i\right)^{-1}\ .$$

Posons $b(\theta) = E_\theta(\hat{\theta}) - \theta$. On obtient

$$\begin{aligned}
\frac{\partial}{\partial\theta} b(\theta) &= \frac{\partial}{\partial\theta} \int_{(\mathbb{R}_+)^n} \hat{\theta}(x)\, \theta^{2n}\, x_1 \times \cdots \times x_n\, e^{-\theta\,(x_1+\cdots+x_n)}\, dx - 1 \\
&= 2n \int_{(\mathbb{R}_+)^n} \hat{\theta}(x)\, \theta^{2n-1}\, x_1 \times \cdots \times x_n\, e^{-\theta\,(x_1+\cdots+x_n)}\, dx \\
&\quad - \int_{(\mathbb{R}_+)^n} \hat{\theta}(x)\, (x_1 + \cdots + x_n)\, \theta^{2n-1}\, x_1 \times \cdots \times x_n\, e^{-\theta\,(x_1+\cdots+x_n)}\, dx - 1 \\
&= \frac{2\,n}{\theta}\, b(\theta) - 2n - 1\ ,
\end{aligned}$$

on en déduit aisément que $\hat{\theta}$ est biaisé.

Exercice 2.6.4

Soit X_i de loi log–normale, i.e. $Y_i = \log(X_i) \sim N(\mu, \sigma^2)$.

$$\begin{aligned}
E\, X_i &= E \exp(Y_i) \\
&= \int_{\mathbb{R}} \exp(y)\, \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, \exp\left(-(y-\mu)^2/(2\,\sigma^2)\right)\, dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{y^2 - 2\,\mu\,y + \mu^2 - 2\,\sigma^2\,y}{2\,\sigma^2}\right) dy \\
&= \exp(\mu + \sigma^2/2) \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{(y^2 - (\mu + \sigma^2))^2}{2\,\sigma^2}\right) dy \\
&= \exp(\mu + \sigma^2/2)\ ,
\end{aligned}
$$

La moyenne de X_i est donc $\xi = \exp(\mu + \sigma^2/2)$. Par ailleurs

$$
\begin{aligned}
E(X_i)^2 &= E\exp(2\,Y_i) \\
&= \int_{\mathbb{R}} \exp(2\,y) \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-(y-\mu)^2/(2\,\sigma^2)\right) dy \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{y^2 - 2\,\mu\,y + \mu^2 - 4\,\sigma^2\,y}{2\,\sigma^2}\right) dy \\
&= \exp(2\,\mu + 2\,\sigma^2) \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{(y^2 - (\mu + 2\,\sigma^2))^2}{2\,\sigma^2}\right) dy \\
&= \exp(2\,\mu + 2\,\sigma^2) \\
&= \xi^2\,\exp(\sigma^2)\ .
\end{aligned}
$$

Donc $E(X_i)^2 = \xi^2\,\exp(\sigma^2)$ et ainsi

$$
\nu \triangleq Var(X_i) = E(X_i)^2 - \xi^2 = \xi^2\,(\exp(\sigma^2) - 1)\ .
$$

On vérifie que la loi log–normale admet la densité

$$
f(a;\mu,\sigma^2) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \frac{1}{a} \exp\left(-\frac{1}{2\sigma^2}(\log(a) - \mu)^2\right) \quad , \quad (a \in]0,\infty[)\ .
$$

Soit $X_1 \cdots X_n$ un n–échantillon d'une loi log–mormale, sa densité est donc

$$
p(x;\mu,\sigma^2) = \frac{1}{x_1 \cdots x_n} \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (\log(x_i) - \mu)^2\right)\ .
$$

En résolvant l'équation de la vraisemblance on obtient

$$
\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^n \log(x_i) \quad , \quad \hat{\sigma}^2(x) = \frac{1}{n}\sum_{i=1}^n (\log(x_i) - \hat{\mu}(x))^2\ ,
$$

ce qui est naturel car $\log(X_1), \ldots, \log(X_n)$ est un n–échantillon de la loi $N(\mu,\sigma^2)$. Pour obtenir les EMV $\hat{\xi}$ et $\hat{\nu}$ de ξ et ν on utilise le principe d'invariance qui permet d'affirmer que

$$
\hat{\xi}(x) = \exp\left(\hat{\mu}(x) - \frac{\hat{\sigma}^2(x)}{2}\right) \quad , \quad \hat{\nu}(x) = \hat{\xi}(x)^2\left(e^{\hat{\sigma}^2(x)} - 1\right)\ .
$$

Une autre manière d'obtenir ce résultat est d'écrire la densité conjointe en fonction de ξ et ν et de résoudre l'équation de la vraisemblance associée (cette approche est beaucoup plus difficile).

Exercice 2.6.5

Soit X une variable aléatoire de loi exponentielle double de paramètre (α,β).

$$
\begin{aligned}
E\,X - \alpha &= \int_{-\infty}^{\infty} (x - \alpha)\,(2\,\beta)^{-1}\,e^{-|x-\alpha|/\beta}\,dx \\
&= \int_{-\infty}^{\infty} x\,(2\,\beta)^{-1}\,e^{-|x|/\beta}\,dx \\
&= 0\ , \quad \textit{(intégration d'une fonction impaire)}
\end{aligned}
$$

d'où $E\,X = \alpha$. Par ailleurs

$$\begin{aligned} E(X-\alpha)^2 &= \int_{-\infty}^{\infty} (x-\alpha)^2\,(2\,\beta)^{-1}\,e^{-|x-\alpha|/\beta}\,dx \\ &= \int_{-\infty}^{\infty} x^2\,(2\,\beta)^{-1}\,e^{-|x|/\beta}\,dx \\ &= 2\int_{0}^{\infty} x^2\,(2\,\beta)^{-1}\,e^{-|x|/\beta}\,dx \\ &= 2\,\beta^2\ , \quad \textit{(par intégration par parties)} \end{aligned}$$

d'où $E\,X^2 = 2\,\beta^2 + \alpha^2$.

La méthode des moments conduit dans ce cas aux équations suivantes

$$\hat{\alpha} = \hat{\mu}_1 \quad , \quad 2\,\hat{\beta}^2 + \hat{\alpha}^2 = \hat{\mu}_2\ ,$$

donc $\hat{\alpha} = \mu_1$ et $\hat{\beta} = \sqrt{\hat{\mu}_2 - \hat{\mu}_1^2}/2$.

Exercice 3.5.1

Soient F_0 la loi exponentielle double de paramètre $(0,1)$ de densité f_0 et F_1 la loi normale $N(0,1)$ de densité f_1. Les hypothèses sont

$$\begin{cases} H_0 : \ X_1 \cdots X_n \text{ est un } n\text{–échantillon de la loi } F_0\ , \\ H_1 : \ X_1 \cdots X_n \text{ est un } n\text{–échantillon de la loi } F_1\ . \end{cases}$$

Sous H_0 la densité de $X = (X_1, \ldots, X_n)$ est

$$p_0(x) = \prod_{i=1}^{n} f_0(x_i) \qquad \text{i.e.} \quad \log\,p_0(x) = -n\,\log 2 - \sum_{i=1}^{n} |x_i|\ .$$

Sous H_1 la densité de $X = (X_1, \ldots, X_n)$ est

$$p_1(x) = \prod_{i=1}^{n} f_1(x_i) \qquad \text{i.e.} \quad \log\,p_1(x) = -\frac{n}{2}\,\log(2\,\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i)^2\ .$$

Le log–rapport de vraisemblance s'écrit donc

$$l(x) \triangleq \log\frac{p_0(x)}{p_1(x)} = n\,k + \frac{1}{2}\sum_{i=1}^{n}\left(|x_i|-1\right)^2\ ,$$

où $k = \log(\sqrt{\pi/2}) - 1/2$. La région de rejet de H_0 est

$$D \triangleq \left\{ x\ ;\ n\,k + \frac{1}{2}\sum_{i=1}^{n}\left(|x_i|-1\right)^2 \le c \right\}\ .$$

Pour calculer la taille de ce test on utilise le théorème central limite. On pose $Y_i = (|X_i|-1)^2$, $Y_1 \cdots Y_n$ est une suite i.i.d. . Sous l'hypothèse H_0 — i.e. en supposant que X_i suit une loi exponentielle double — on vérifie aisément (merci Macsyma) que $E\,Y_i = 1$ et $\mathrm{Var}\,Y_i = 8$, d'où

$$\frac{Y_1 + \cdots + Y_n - n}{\sqrt{8\,n}} \underset{n\to\infty}{\overset{\text{loi}}{\to}} N(0,1)\ .$$

Ainsi

$$\begin{aligned}
\alpha_D &= P_0(D) \\
&= P\left(n\,k + \frac{1}{2}\sum_{i=1}^{n}(|X_i|-1)^2 \le c\right) \\
&\qquad (\{X_i\} \textit{ i.i.d. de loi exponentielle double}) \\
&= P\left(\sum_{i=1}^{n} Y_i \le 2\,(c - n\,k)\right) \\
&= P\left(\frac{Y_1 + \cdots + Y_n - n}{\sqrt{8\,n}} \le \frac{c - n\,\log\sqrt{\pi/2}}{\sqrt{2\,n}}\right) \\
&\simeq \int_{-\infty}^{a} \frac{1}{\sqrt{2\,\pi}}\, e^{-u^2/2}\, du\ . \\
&\qquad (\textit{ pour } n \textit{ grand avec } a = (c - n\,\log\sqrt{\pi/2})/\sqrt{2\,n})
\end{aligned}$$

Exercice 3.5.2

L'espace des paramètres est $\Theta = \{\mu; \mu \in \mathbb{R}\}$, l'espace des paramètres associé à l'hypothèse H_0 est $\Theta_0 = \{\mu_0\}$. Soit $\hat{\mu}(x) = \sum x_i/n$ l'EMV de μ. Si $p(x;\mu)$ désigne la densité d'un n–échantillon de la loi $N(\mu, \sigma^2)$ (σ^2 connu), le "$-2\times$log rapport de vraisemblance" s'écrit

$$\begin{aligned}
l(x) &= -2\,\log\,\lambda(x) \\
&= -2\,\log\,\frac{\sup_{\mu\in\Theta_0}\,p(x;\mu)}{\sup_{\mu\in\Theta}\,p(x;\mu)} \\
&= -2\,(\log\,p(x;\mu_0) - \log\,p(x;\hat{\mu}(x)))\ .
\end{aligned}$$

Après calcul on obtient

$$l(x) = \frac{n}{\sigma^2}\,(\hat{\mu}(x) - \mu_0)^2\ .$$

La région de rejet de H_0 est

$$D \triangleq \{x\ ;\ l(x) \ge c\} = \left\{x\ ;\ |\hat{\mu}(x) - \mu_0| \ge \frac{z\,\sigma}{\sqrt{n}}\right\}\ .$$

La taille de ce test est

$$\begin{aligned}
\alpha_D &= P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu_0\right| \ge \frac{z\,\sigma}{\sqrt{n}}\right) \qquad (\{X_i\} \textit{ i.i.d. } N(\mu_0, \sigma^2)) \\
&= P\left(\left|\frac{X_1 + \cdots + X_n - n\,\mu_0}{\sqrt{n\,\sigma^2}}\right| \ge z\right)\ ,
\end{aligned}$$

pour terminer il suffit de remarquer que $(X_1 + \cdots + X_n - n\,\mu_0)/\sqrt{n\,\sigma^2}$ est une variable aléatoire $N(0,1)$.

Exercice 3.5.3

Le modèle est le suivant: $X = (X_1, \ldots, X_n)$ est une suite de variables indépendantes et $X_i \sim N(S_i, 1)$ (attention X n'est pas un n–échantillon). La densité de X est

$$p(x;S) = \left(\frac{1}{\sqrt{2\,\pi}}\right)^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - S_i)^2\right)\ .$$

où $S \in \Theta = \{S \in \mathbb{R}^n ; S_i = \pm 1\}$. L'EMV de S_j est donné par

$$\hat{S}_j(x) = \text{Arg} \max_{S_j=\pm 1} \log p(x;S) = \text{Arg} \min_{S_j=\pm 1} \sum_{i=1}^{n} (x_i - S_i)^2 = \text{Arg} \min_{S_j=\pm 1} (x_j - S_j)^2 \ ,$$

on obtient donc $\hat{S}_j(x) = +1$ si $x_j > 0$ et $\hat{S}_j(x) = -1$ si $x_j < 0$ (si $x_j = 0$ on peut convenir que $\hat{S}_j(x) = +1$). Ainsi

$$\hat{S}(x) = \left(\frac{x_1}{|x_1|}, \ldots, \frac{x_n}{|x_n|} \right) \ .$$

On se donne $S^0 = (S_1^0 \cdots S_n^0) \in \Theta$, et on définit l'hypothèse

$$H_0 \ : \text{ “ } S = S^0 \text{ ” } .$$

L'espace des paramètres associé à H_0 est le singleton $\Theta_0 = \{S^0\}$ et le rapport de vraisemblance est (en utilisant le fait que $(\hat{S}_i(x))^2 = 1 = (S_i^0)^2$)

$$\begin{aligned} l(x) &\triangleq -2 \log \lambda(x) \\ &= -2 \log \frac{\sup_{S \in \Theta_0} p(x;S)}{\sup_{S \in \Theta} p(x;S)} \\ &= -2 \log \frac{p(x;S^0)}{p(x;\hat{S}(x))} \\ &= 2 \sum_{i=1}^{n} x_i \left(\hat{S}_i(x) - S_i^0 \right) \ . \end{aligned}$$

Maintenant il suffit de remarquer que

$$\sum_{i=1}^{n} x_i \left(\hat{S}_i(x) - S_i^0 \right) = \sum_{i ; \hat{S}_i \neq S_i^0} x_i \left(\hat{S}_i(x) - S_i^0 \right) = 2 \sum_{i ; \hat{S}_i \neq S_i^0} |x_i| \ .$$

Ainsi la forme de la région de rejet de H_0 est

$$D \triangleq \left\{ x \ ; \sum_{\{i ; \hat{S}_i(x) \neq S_i^0\}} |x_i| \geq c \right\} \ .$$

B.2 Autres Exercices

Exercice B.2.1 *Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi F de densité*

$$f(a;\theta) = \mathbb{I}_{[\theta - 1/2, \theta + 1/2]}(a) \ , \quad a \in \mathbb{R} \ ,$$

où $\theta \in \Theta = \mathbb{R}$. Calculez l'EMV de θ.

réponse La densité de l'observation est

$$p(x;\theta) \triangleq \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} 1 & \text{si } x_i \in [\theta - 1/2, \theta + 1/2] \text{ pour tout } i, \\ 0 & \text{sinon,} \end{cases}$$

$x = (x_1 \cdots x_n) \in \mathbb{R}^n$. Il s'agit d'un cas non régulier, on ne peut donc pas calculer $\hat{\theta}$ en dérivant. On doit maximiser

$$\hat{\theta}(x) = \text{Arg} \max_{\theta \in \mathbb{R}} p(x;\theta) \ .$$

Nécessairement $\hat{\theta}$ vérifie

$$\hat{\theta}(x) \in \left[x_i - \frac{1}{2}, x_i + \frac{1}{2}\right] \quad , \quad \forall i = 1, \ldots, n \ .$$

Donc

$$\hat{\theta}(x) \in \left[\max_i x_i - \frac{1}{2}, \min_i x_i + \frac{1}{2}\right] \ ,$$

et en fait tout $\hat{\theta}(x)$ appartenant à ce dernier intervalle est un EMV de θ. On peut prendre par exemple

$$\begin{aligned}
\hat{\theta}(x) &= \max_i x_i - \frac{1}{2} \ , \\
\hat{\theta}(x) &= \min_i x_i + \frac{1}{2} \ , \\
\hat{\theta}(x) &= \frac{1}{2}\left(\max_i x_i - \frac{1}{2} + \min_i x_i + \frac{1}{2}\right) = \frac{1}{2}\left(\max_i x_i + \min_i x_i\right) \ .
\end{aligned}$$

Exercice B.2.2 *Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi F de densité*

$$f(a; \mu, \sigma) = \frac{1}{2\sqrt{3}\,\sigma} \mathbb{I}_{[\mu - \sqrt{3}\,\sigma, \mu + \sqrt{3}\,\sigma]}(a) \ , \quad a \in \mathbb{R} \ ,$$

où $-\infty < \mu < \infty$ et $\sigma > 0$. Calculez l'EMV de (μ, σ).

réponse La densité de l'observation est

$$\begin{aligned}
p(x; \mu, \sigma) &\triangleq \prod_{i=1}^{n} f(x_i; \mu, \sigma) \\
&= \begin{cases} \left(\frac{1}{2\sqrt{3}\,\sigma}\right)^n & \text{si } x_i \in [\mu - \sqrt{3}\,\sigma, \mu + \sqrt{3}\,\sigma] \text{ pour tout } i, \\ 0 & \text{sinon,} \end{cases}
\end{aligned}$$

$x = (x_1 \cdots x_n) \in \mathbb{R}^n$. Il s'agit d'un cas non régulier, on ne peut donc pas calculer $(\hat{\mu}, \hat{\sigma})$ en dérivant. On doit maximiser

$$(\hat{\mu}, \hat{\sigma}) = \text{Arg} \max_{(\mu,\sigma) \in \mathbb{R} \times \mathbb{R}_+} p(x; \mu, \sigma) \ .$$

Si on pose $y_1 = \min_i x_i$ et $y_n = \max_i x_i$, on peut vérifier que

$$p(x; \mu, \sigma) = \left(\frac{1}{2\sqrt{3}\,\sigma}\right)^n \mathbb{I}_{[(\mu - y_1)/\sqrt{3}, \infty[}(\sigma) \ \mathbb{I}_{[(y_n - \mu)/\sqrt{3}, \infty[}(\sigma) \ ,$$

on doit donc trouver

$$(\hat{\mu}, \hat{\sigma}) = \text{Arg} \max_{(\mu,\sigma) \in D} \left(\frac{1}{2\sqrt{3}\,\sigma}\right)^n \ ,$$

où D est défini par

$$D \triangleq \left\{(\mu, \sigma) \ ; \ (\mu - y_1)/\sqrt{3} \leq \sigma \ , \ (y_n - \mu)/\sqrt{3} \leq \sigma)\right\}$$

En traçant le domaine D sur le plan (en (μ, σ)), on vérifie que le maximum de $(2\sqrt{3}\,\sigma)^{-n}$ sur le domaine D est atteint pour

$$\begin{aligned}
\hat{\mu} &= \frac{y_1 + y_n}{2} = \frac{\min_i x_i + \max_i x_i}{2} \ , \\
\hat{\sigma} &= \frac{y_n - y_n}{2\sqrt{3}} = \frac{\max_i x_i - \min_i x_i}{2\sqrt{3}} \ .
\end{aligned}$$

(Comparez avec l'exercice suivant).

Exercice B.2.3 *Dans l'exercice précédent utilisez la méthode des moments pour estimer μ et σ.*

réponse Les deux premiers moments de X_i sont

$$\begin{aligned} E\,X_i &= \frac{1}{2\sqrt{3}\,\sigma}\int_{\mu-\sqrt{3}\,\sigma}^{\mu+\sqrt{3}\,\sigma} a\;da = \mu\ , \\ E\,X_i^2 &= \frac{1}{2\sqrt{3}\,\sigma}\int_{\mu-\sqrt{3}\,\sigma}^{\mu+\sqrt{3}\,\sigma} a^2\;da = \mu^2+\sigma^2\ . \end{aligned}$$

Les deux premiers moments empiriques sont

$$\hat{\mu}^1 = \frac{x_1+\cdots+x_n}{n}\ , \quad \hat{\mu}^2 = \frac{x_1^2+\cdots+x_n^2}{n}\ .$$

Les équations de la méthode des moments s'écrivent

$$\left\{ \begin{array}{rcl} \mu &=& \hat{\mu}^1\ , \\ \mu^2+\sigma^2 &=& \hat{\mu}^2\ , \end{array} \right.$$

d'où les estimateurs

$$\frac{1}{n}\sum_{i=1}^n x_i \ \text{ pour } \mu\ , \quad \sqrt{\frac{1}{n}\sum_{i=1}^n\left(x_i-\frac{1}{n}\sum_{j=1}^n x_j\right)^2} \ \text{ pour } \sigma\ .$$

Exercice B.2.4 *Soit $X_1,\ldots,X_n$ un n–échantillon d'une loi F de densité*

$$f(a;\theta) = \theta\,e^{-\theta\,a}\ , \quad a\in]0,\infty[\ ,$$

où $\theta>0$. Calculez la borne de Cramer–Rao (cf. théorème 2.2.5).

réponse La densité de l'observation est

$$p(x;\theta) \triangleq \prod_{i=1}^n f(x_i;\theta) = \theta^n\,e^{-\theta\,(x_1+\ldots+x_n)}\ ,$$

$x=(x_1\cdots x_n)\in\mathbb{R}^n$. Donc

$$\frac{\partial}{\partial\theta}\log p(x;\theta) = \frac{\partial}{\partial\theta}\left(n\,\log\theta-\theta\,(x_1+\ldots+x_n)\right) = \frac{n}{\theta}-S(x)\ ,$$

où $S(x)=(x_1+\ldots+x_n)$.

$$\begin{aligned} E_\theta\left(\frac{n}{\theta}-S\right)^2 &= E\left(\frac{n}{\theta}-S(X)\right)^2 \\ &\qquad \text{(où } X=(X_1\cdots X_N) \text{ est une suite i.i.d. de loi } F) \\ &= E\left(\sum_{i=1}^n(X_i-\frac{1}{\theta})\right)^2 \\ &= E\left(\sum_{i=1}^n(X_i-E\,X_i)\right)^2 \quad (\text{car } E(X_i)=1/\theta) \\ &= \left(\sum_{i=1}^n E(X_i-E\,X_i)^2\right) \\ &= \frac{n}{\theta^2}\ . \quad (\text{car } \mathrm{Var}(X_i)=1/\theta^2) \end{aligned}$$

Si on note $K(\theta)$ la borne de Cramer–Rao, on obtient $K(\theta)=\theta^2/n$.

Exercice B.2.5 *Soit* $X_1, \ldots, X_n$ *un* n*–échantillon d'une loi* F *de densité*

$$f(a;\theta) = \theta\, e^{-\theta\, a} \ , \quad a \in]0,\infty[\ ,$$

où $\theta \in \Theta =]0,\infty[$. *On considère l'hypothèse*

$$H_0 \ : \ \text{“}\ \theta \leq \theta_0 \ \text{”} \ ,$$

où $\theta_0 \in \Theta$ *est donné. Ecrire la* forme *du test TRV de l'hypothèse* H_0.

réponse L'ensemble des paramètres correspondants à l'hypothèse H_0 est

$$\Theta_0 \triangleq \{\theta \, ; \, \theta \leq \theta_0\} \ .$$

La densité de l'échantillon est

$$p(x;\theta) \triangleq \prod_{i=1}^{n} f(x_i;\theta) = \theta^n \, e^{-\theta\, S(x)} \ ,$$

où $x = (x_1 \cdots x_n) \in \mathbb{R}^n$ et $S(x) = (x_1 + \ldots + x_n)$. Le rapport de vraisemblance s'écrit

$$\lambda(x) = \frac{\sup\limits_{\theta \in \Theta_0} p(x;\theta)}{\sup\limits_{\theta \in \Theta} p(x;\theta)} = \frac{p(x;\hat{\theta}_0(x))}{p(x;\hat{\theta}(x))} \ ,$$

où $\hat{\theta}(x)$ est l'EMV de θ, et $\hat{\theta}_0(x)$ est l'EMV de θ sous l'hypothèse H_0. On vérifie que

$$\hat{\theta}(x) = \frac{n}{S(x)} \ , \quad \hat{\theta}_0(x) = \begin{cases} \frac{n}{S(x)} & \text{si} \quad \frac{n}{S(x)} \leq \theta_0 \ , \\ \theta_0 & \text{si} \quad \frac{n}{S(x)} > \theta_0 \ . \end{cases}$$

Donc

$$\lambda(x) = \begin{cases} 1 & \text{si} \quad \frac{n}{S(x)} \leq \theta_0 \ , \\ \frac{\theta_0^n \, e^{-\theta_0\, S(x)}}{\left(\frac{n}{S(x)}\right)^n e^{-n}} & \text{si} \quad \frac{n}{S(x)} > \theta_0 \ . \end{cases}$$

Soit $c \in]0,1[$[1] Notons $\bar{x} = S(x)/n$, la région de rejet de H_0 est de la forme

$$\begin{aligned} D &= \{x \, ; \, \lambda(x) \leq c\} \\ &= \{x \, ; \, \lambda(x) \leq c \, , \, \bar{x}\, \theta_0 \geq 1\} \cup \{x \, ; \, \lambda(x) \leq c \, , \, \bar{x}\, \theta_0 < 1\} \\ &= \{x \, ; \, 1 \leq c \, , \, \bar{x}\, \theta_0 \geq 1\} \cup \left\{x \, ; \, (\theta_0\, \bar{x})^n \, e^{n - n\theta_0\, \bar{x}} \leq c \, , \, \bar{x}\, \theta_0 < 1\right\} \ . \end{aligned}$$

Le premier de ces deux ensembles est vide, donc on rejette l'hypothèse H_0 si et seulement si

$$(\theta_0\, \bar{x})^n \, e^{n - n\theta_0\, \bar{x}} \leq c \quad \text{et} \quad \theta_0\, \bar{x} < 1 \ , \tag{B.1}$$

c'est–à–dire ssi

$$\theta_0\, \bar{x} \leq k \ , \tag{B.2}$$

où $k \in]0,1[$. Le TRV de H_0 admet donc une région de rejet de la forme

$$D \triangleq \left\{ x \, ; \, \theta_0 \, \frac{1}{n} \sum_{i=1}^{n} x_i \leq k \right\} \ ,$$

1. il suffit de prendre c dans cet intervalle, cf. remarque 3.3.3, la valeur $c = 1$ n'est pas intéressante puisque dans ce cas $D = \mathbb{R}$.

avec $k \in]0,1[$. (Pour se convaincre de l'équivalence de (B.1) et (B.2) poser $\bar{y} = \theta_0\,\bar{x}$ et tracer le graphe de la fonction $]0,\infty[\ni \bar{y} \to \bar{y}^n\,e^{n-n\,\bar{y}}$. Vérifier alors que " $y^n\,e^{n-n\,\bar{y}} \leq c$ et $\bar{y} < 1$ " est équivalent à " $\bar{y} \leq k$ " où $k \in]0,1[$).

Exercice B.2.6 *Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi F de densité*

$$f(a;\theta) = \theta\,f_1(a) + (1-\theta)\,f_0(a), \quad a \in \mathbb{R}\ ,$$

où $\theta \in \Theta =]0,1[$, et f_0, f_1 sont deux densités sur $\mathbb{R}$ connues. Vérifiez que $f(a;\theta)$ est bien une densité sur $\mathbb{R}$. Estimez θ par la méthode des moments. Dans le cas d'un échantillon de taille 2 (i.e. $n = 2$), calculez l'EMV de θ.

réponse $f(\,\cdot\,)$ est positive, $\int f(a;\theta)\,da = 1$ et comme somme de deux fonctions mesurables, elle est aussi mesurable. Donc $f(\,\cdot\,;\theta)$ est une densité sur $\mathbb{R}$. Posons $\mu_j = \int a\,f_j(a)\,da$ $(j = 0,1)$, μ_i est la moyenne de f_i, on a

$$E\,X_i = \int_{\mathbb{R}} a\,f(a;\theta)\,da = \int_{\mathbb{R}} a\,(\theta\,f_1(a) + (1-\theta)\,f_0(a))\ da = \theta\,\mu_1 + (1-\theta)\,\mu_0\ .$$

Si $\hat{\mu}^1$ désigne la moyenne empirique de l'échantillon, la méthode des moments donne l'équation: $\theta\,\mu_1 + (1-\theta)\,\mu_0 = \hat{\mu}^1$, d'où on tire l'estimateur de θ suivant

$$\frac{1}{\mu_1 - \mu_0}\left(\hat{\mu}^1 - \mu_0\right) = \frac{1}{\mu_1 - \mu_0}\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)\ .$$

Dans le cas $n = 2$ la densité de l'échantillon s'écrit

$$p(x;\theta) = f(x_1;\theta)\ f(x_2;\theta) = (\theta\,f_1(x_1) + (1-\theta)\,f_0(x_1))\ \ (\theta\,f_1(x_2) + (1-\theta)\,f_0(x_2))\ .$$

On obtient donc

$$\frac{\partial}{\partial\theta}p(x;\theta) = 2\,\theta\,\alpha_1\,\alpha_2 + f_0(x_2)\,\alpha_1 + f_0(x_1)\,\alpha_2\ ,$$

où $\alpha_j = f_1(x_j) - f_0(x_j)$ $(j = 1,2)$. L'EMV de θ est donc

$$\hat{\theta}(x) = \frac{1}{2}\,\frac{f_0(x_2)\,\alpha_1 + f_0(x_1)\,\alpha_2}{\alpha_1\,\alpha_2}\ .$$

Exercice B.2.7 *Soit $X = X_1, \ldots, X_n$ un n–échantillon d'une loi F de densité $f(a;\theta)$ — régulière par rapport à θ — donnée. On suppose que θ appartient à Θ une partie ouverte de $\mathbb{R}$. On note $p(x;\theta)$ la densité de X. On définit*

$$\begin{aligned} I(\theta) &= E\left(\frac{\partial}{\partial\theta}\log\,p(X;\theta)\right)^2\ , \\ \mathcal{I}(\theta) &= E\left(\frac{\partial}{\partial\theta}\log\,f(X_i;\theta)\right)^2\ . \end{aligned}$$

(cf. définition 2.2.4). Montrez que

$$I(\theta) = n\,\mathcal{I}(\theta)\ . \tag{B.3}$$

Montrez que de plus

$$E\left(\frac{\partial}{\partial\theta}\log\,f(X_i;\theta)\right)^2 = -E\frac{\partial^2}{\partial\theta^2}\log\,f(X_i;\theta)\ . \tag{B.4}$$

Remarque: $I(\theta)$ est plus difficile à calculer que $\mathcal{I}(\theta)$, c'est donc ce dernier que l'on calculera dans le cas des échantillons pour obtenir l'information de Fisher (Les résultats (B.3) et (B.4) peuvent être admis comme des résultats du cours).

réponse Comme $\log p(x;\theta) = \sum_i \log f(x_i;\theta)$, on a

$$\begin{aligned} I(\theta) &= E\left(\frac{\partial}{\partial\theta}\log p(X;\theta)\right)^2 \\ &= \sum_i E\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2 + \sum_{i\neq j} E\frac{\partial}{\partial\theta}\log f(X_i;\theta)\; E\frac{\partial}{\partial\theta}\log f(X_j;\theta) \\ &= n\,\mathcal{I}(\theta) + n\,(n-1)\,\left(E\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2 . \end{aligned}$$

Pour démontrer (B.3) il suffit de remarquer que

$$\begin{aligned} E\frac{\partial}{\partial\theta}\log f(X_i;\theta) &= \int \frac{\frac{\partial}{\partial\theta}f(a;\theta)}{f(a;\theta)}\; f(a;\theta)\,da \\ &= \int \frac{\partial}{\partial\theta}f(a;\theta)\,da \\ &= \frac{d}{d\theta}\int f(a;\theta)\,da \\ &= 0 , \end{aligned}$$

car $\int f(a;\theta)\,da = 1$. Notons ∂ l'opérateur de dérivation par rapport à θ. Pour vérifier (B.4) on calcule

$$\begin{aligned} \int \left(\partial^2 \log f(a;\theta)\right)\; f(a;\theta)\,da &= \int \frac{\partial^2 f}{f}\, f\,da - \int \left(\frac{\partial f}{f}\right)^2 f\,da \\ &= \partial^2 \int f\,da - \int \left(\frac{\partial f}{f}\right)^2 f\,da \\ &= -\int \left(\frac{\partial f}{f}\right)^2 f\,da , \end{aligned}$$

d'où (B.4).

Exercice B.2.8 *Utilisez l'exercice précédent afin de calculer l'information de Fisher $I(\theta)$ dans l'exercice B.2.4.*

réponse On a $\log f(a;\theta) = \log(\theta) - \theta\, a$, donc

$$\frac{\partial^2}{\partial\theta^2}\log f(a;\theta) = -\frac{1}{\theta^2} .$$

D'après l'exercice B.2.7

$$\begin{aligned} I(\theta) &= n\,E\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2 \\ &= n\,E\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2 \\ &= n\,E\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2 \\ &= -n\,E\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta) \\ &= -\frac{n}{\theta^2} . \end{aligned}$$

Exercice B.2.9 *Soit $X_1,\ldots,X_n$ un n–échantillon d'une loi F de densité*

$$f(a;\theta) = \frac{e^{-\theta}\,\theta^a}{x!} , \quad a \in \mathbb{N} ,$$

où $\theta \in \Theta =]0,\infty[$. Calculez $K(\theta)$ la borne de Cramer–Rao associé à ce problème.

réponse D'après l'exercice B.2.7, l'information de Fisher est

$$I(\theta) = -n\, E \frac{\partial^2}{\partial \theta^2} \log f(X_i;\theta) = -n\, E\left(-\frac{X_i}{\theta^2}\right) = \frac{n}{\theta^2}\, E\, X_i \ .$$

Et

$$\begin{aligned}
E\, X_i &= \sum_{a=0}^{\infty} a\, f(a;\theta) \\
&= \sum_{a=0}^{\infty} a\, \frac{e^{-\theta}\, \theta^a}{a!} \\
&= \sum_{a=1}^{\infty} a\, \frac{e^{-\theta}\, \theta^a}{a!} \\
&= \sum_{a=1}^{\infty} \frac{e^{-\theta}\, \theta^a}{(a-1)!} \\
&= \theta \sum_{a=1}^{\infty} \frac{e^{-\theta}\, \theta^{a-1}}{(a-1)!} \\
&= \theta \sum_{a'=0}^{\infty} \frac{e^{-\theta}\, \theta^{a'}}{(a')!} \ , \quad (a' = a-1) \\
&= \theta \ . \quad (\text{car } \textstyle\sum_{a'=0}^{\infty} e^{-\theta}\, \theta^{a'}/(a')! = 1)
\end{aligned}$$

Ainsi la borne de Cramer–Rao est $K(\theta) = \theta/n$.

Exercice B.2.10 *Soit $X_1, \ldots, X_n$ un n–échantillon d'une loi de Bernoulli F de densité*

$$f(a;\theta) = \theta^a\, (1-\theta)^{1-a} \ , \quad a \in \{0,1\} \ ,$$

où $\theta \in \Theta = [0,1]$. On se donne une loi a priori sur le paramètre θ de densité

$$p_{\underline{\theta}}(\theta) = \mathbb{I}_{]0,1[}(\theta) \ ,$$

i.e. $\underline{\theta}$ est uniformément ditribué sur l'intervalle $]0,1[$. Calculez l'estimateur de Bayes de θ.[2]*. Vérifiez que cet estimateur est biaisé.*

réponse La densité de l'échantillon $X = X_1, \ldots, X_n$ est

$$p(x;\theta) = \theta^{S(x)}\, (1-\theta)^{n-S(x)} \ , \quad x \in \{0,1\}^n \ ,$$

où $S(x) = x_1 + \cdots + x_n$. La densité a posteriori de $\underline{\theta}$ est

$$p_{\underline{\theta}|X}(\theta|x) \triangleq \frac{p(x;\theta)\, p_{\underline{\theta}}(\theta)}{\int_\Theta p(x;\theta)\, p_{\underline{\theta}}(\theta)\, d\theta} \ .$$

L'estimateur bayésien de θ est par définition

$$\begin{aligned}
&E\,(\underline{\theta}\,|\, X_1 = x_1, \ldots, X_n = x_n) \\
&= \int_\Theta \theta\, p_{\underline{\theta}|X}(\theta|x)\, d\theta \\
&= \frac{\int_\Theta \theta\, p(x;\theta)\, p_{\underline{\theta}}(\theta)\, d\theta}{\int_\Theta p(x;\theta)\, p_{\underline{\theta}}(\theta)\, d\theta}
\end{aligned}$$

2. Utilisez la proposition 2.2.13, ainsi que les définitions des fonctions β et Γ de l'exemple 2.2.14

$$
\begin{aligned}
&= \frac{\int_\Theta \theta\, \theta^{S(x)}\, (1-\theta)^{n-S(x)}\, d\theta}{\int_\Theta \theta^{S(x)}\, (1-\theta)^{n-S(x)}\, d\theta} \\
&= \frac{\beta\,(S(x)+2\,,\, n-S(x)+1)}{\beta\,(S(x)+1\,,\, n-S(x)+1)} \\
&= \frac{\Gamma\,(S(x)+2)\,\Gamma\,(n-S(x)+1)}{\Gamma\,(n+3)}\;\frac{\Gamma\,(n+2)}{\Gamma\,(S(x)+1)\,\Gamma\,(n-S(x)+1)} \\
&= \frac{S(x)+1}{n+2}\;.
\end{aligned}
$$

L'estimateur bayésien de θ est donc

$$
\tilde{\theta}(x) = \frac{1+\sum_{i=1}^{n} X_i}{n+2}\;.
$$

On vérifie aisément que cet estimateur est biaisé

$$
\begin{aligned}
E_\theta \tilde{\theta} &= \frac{1}{n+2}\sum_{i=1}^{n} E\,X_i + \frac{1}{n+2} \quad (\text{où les } X_i \text{ sont i.i.d. de loi } B(\theta)) \\
&= \frac{n\,\theta+1}{n+2}\;. \quad (\text{car } E\,X_i = \theta)
\end{aligned}
$$

B.3 Problèmes

Problème 1 : Un exemple de contrôle de qualité

le problème

Une entreprise produit un certain type de circuit imprimé. On dispose d'un moyen pour contrôler le bon fonctionnement de ces circuits, ce contrôle est à deux issues: *le circuit fonctionne* ou *le circuit ne fonctionne pas*. Le dirigeants de cette entreprise se posent le problème du contrôle de la qualité des circuits qu'ils produisent. En fait ils veulent estimer le taux de pièces défectueuses en sortie de la chaîne de production. On pose

$$
\theta \triangleq \text{ taux de pièces défectueuses,}
$$

θ est un paramètre inconnu appartenant à l'intervalle $[0,1]$.

Le moyen le plus naturel pour estimer θ est le suivant: on prend n pièces dans la production du jour. On teste ces n pièces, on compte le nombre de pièces défectueuses et on pose

$$
\hat{\theta}_n \triangleq \frac{\text{nombre de pièces défectueuses}}{\text{nombre total de pièces } (=n)}
$$

C'est–à–dire, pour la ième pièce on pose

$$
x_i \triangleq \begin{cases} 1 & \text{si la pièce est défectueuse,} \\ 0 & \text{si la pièce est bonne,} \end{cases}
$$

alors

$$
\hat{\theta}_n \triangleq \frac{x_1+\cdots+x_n}{n}\;.
$$

On dit que $\hat{\theta}_n$ est une estimateur de θ. Cet estimateur — appelé *moyenne empirique* — est le plus intuitif possible. Supposons qu'on obtienne $\hat{\theta}_n = 0.005$, on peut alors dire que 0.5% de la production est défectueuse. Il existe d'autres questions que l'on peut se poser:

Question 1 Quelle est la qualité de cet estimateur. A–t–on $\theta = 0.005 \pm 0.0001$, $\theta = 0.005 \pm 0.01$ ou $\theta = 0.005 \pm 0.1$? On veut donc obtenir une information sur l'erreur commise.

Question 2 Supposons qu'après plusieurs jours de contrôle on puisse affirmer que 0.005 est une estimation suffisante du taux de pièces défectueuses en cours normal de production. On décide — par exemple tous les mois — de refaire un contrôle pour savoir si la qualité de la production se détèriore ou non (mauvais fonctionnement d'une machine, erreur dans la production). Si un jour on obtient $\hat{\theta}_n = 0.008$, cette valeur est–elle suffisamment plus grande que 0.005 pour signifier une réelle détérioration de la production?

Pour l'instant, on ne sait pas répondre à ces questions. C'est pourquoi on est amené à construire un modèle (étape de *modélisation*) censé décrire cette situation et qui permettra de répondre à ces questions (étape *d'identification*).

la modélisation

Cette étape se schématise ainsi

$$\left.\begin{array}{r}\text{connaissance du problème}\\ \text{hypothèses a priori}\end{array}\right\} \underset{\text{traduction}}{\rightarrow} \text{Modèle Mathématique.}$$

Considérons trois points

- Si θ est le taux de pièces défectueuses et si on considère une pièce donnée, on peut dire que cette pièce sera défectueuse avec une probabilité θ et bonne avec une probabilité $1-\theta$. Si la pièce est défectueuse on lui associe la valeur 1 et 0 si elle est bonne. En d'autres termes, le résultat du contrôle sur cette pièce suit une loi de Bernoulli $B(\theta)$ de paramètre θ. Cette loi admet la densité

$$f(a;\theta) \triangleq \begin{cases} \theta & \text{si} \quad a = 1 \ , \\ 1-\theta & \text{si} \quad a = 0 \ , \end{cases} \qquad \text{i.e. } f(a;\theta) = \theta^a (1-\theta)^{1-a} \ , \quad a \in \{0,1\} \ .$$

- On prend n pièces numérotées de 1 à n et on considère

$$X_i = \text{ le résultat du contrôle pour la pièce numéro } i.$$

D'après le point précédent, X_i est une variable aléatoire de loi de Bernoulli $B(\theta)$. On fait l'hypothèse suivante

$(H1)$ *le taux de pièces défectueuses est constant durant une journée.*

Si les n pièces en question ont été produites la même journée, alors les X_i $(i = 1 \cdots n)$ suivent la même loi $B(\theta)$. Donc

$$\text{loi}(X_i) = B(\theta) \ , \quad (i = 1 \cdots n) \ . \tag{B.5}$$

- Etant donnés deux pièces, on a remarqué que le fait qu'une des deux pièces soit défectueuse ou non est indépendant du fait que l'autre est défectueuse ou non (ou du moins on le suppose). On est donc amené à faire l'hypothèse suivante

$(H2)$ *les variables $X_1, \ldots, X_n$ sont indépendantes.*

Avec (B.5), on obtient donc

$$X_1, \ldots, X_n \quad \text{i.i.d. loi}\,(X_i) = B(\theta) \ . \tag{B.6}$$

Posons $X = (X_1, \ldots, X_n)$. X est un vecteur aléatoire à valeurs dans $\{0,1\}^n$. D'après (B.6), si on note $p(x;\theta)$ la densité de la loi de X (qui dépend de θ), on a

$$P(X_1 = x_1, \ldots, X_n = x_n) = p(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta) \ , \tag{B.7}$$

$x = (x_1 \cdots x_n) \in \{0,1\}^n$. X est appelée observation (ou données). On construit le modèle statistique

- Espace des observations: $\mathcal{X} = \{0,1\}^n$.

- P_θ loi de l'observation $P_\theta(\{x\}) = P(X = x) = p(x;\theta)$ $(x \in \mathcal{X})$, donnée par (B.7).
- Espace des paramètres $\Theta = [0,1]$.

Si $\mathcal{F}$ est l'ensemble des parties de $\mathcal{X}$ ($\mathcal{F}$ est une tribu sur $\mathcal{X}$), alors le modèle statistique s'écrit

$$(\mathcal{X}, \mathcal{F}, \{P_\theta; \theta \in \Theta\}) \ . \tag{B.8}$$

Sur ce modèle le problème se formule ainsi: on dispose d'une observation X dont la loi est inconnue, on sait seulement que loi(X) appartient à l'ensemble de lois $\{P_\theta; \theta \in \Theta\}$, cet ensemble de lois est connu (i.e. P_θ est Θ sont connus). On veut donc déterminer θ tel que loi$(X) = P_\theta$.

l'identification

identification totale: estimation de θ. Une fonction

$$\begin{array}{rcl} T \ : \ \mathcal{X} & \longrightarrow & \Theta \\ x & \longrightarrow & T(x) \end{array}$$

de l'espace des observations dans l'espace des paramètres est appelé *estimateur* de θ. Dans le chapitre **??** on a déjà construit un estimateur

$$\hat{\theta}_n(x) \triangleq \frac{x_1 + \cdots + x_n}{n} \ , \quad (x \in \mathcal{X}) \ . \tag{B.9}$$

(1) *calcul du biais de $\hat{\theta}_n$* On note E_θ l'espérance associée à P_θ,

$$\begin{array}{rcl} E_\theta \hat{\theta}_n & = & E\hat{\theta}(X) \qquad (\text{où loi}(X) = P_\theta)\,) \\ & = & E(X_1 + \cdots + X_n)/n \\ & = & (EX_1 + \cdots + EX_n)\,/n \ , \end{array}$$

X_i suit une loi $B(\theta)$ donc $EX_i = \theta$, d'où

$$E_\theta \hat{\theta}_n = \frac{n\,\theta}{n} = \theta \ ,$$

ce calcul est vraie pour toute valeur de θ, on en déduit

$$b(\theta, \hat{\theta}_n) = E_\theta \hat{\theta}_n - \theta = 0 \ , \quad \forall \theta \in \Theta \ ,$$

i.e. $\hat{\theta}_n$ est un estimateur non biaisé.

(2) *calcul du rique de $\hat{\theta}_n$*

$$\begin{array}{rcl} R(\theta, \hat{\theta}_n) & = & E_\theta |\hat{\theta}_n - \theta|^2 \\ & = & E|\hat{\theta}_n(X) - \theta|^2 \qquad (\text{où loi}(X) \sim P_\theta) \\ & = & \dfrac{1}{n^2}\, E(S_n - n\,\theta)^2 \ , \end{array}$$

où $S_n = X_1 + \ldots + X_n$. $X_1, \ldots, X_n$ est une suite de variables i.i.d. de loi $B(\theta)$, donc S_n suit une loi binomiale de paramètre (n, θ). Ainsi $E\,S_n = n\,\theta$ et $Var\,S_n = n\,\theta\,(1-\theta)$, d'où

$$R(\theta, \hat{\theta}_n) = \frac{\theta\,(1-\theta)}{n} \ , \quad \forall \theta \in \Theta \ . \tag{B.10}$$

(tracez la courbe $\theta \to R(\theta, \hat{\theta}_n)$, pour $\theta \in [0,1]$.)

Commentaires:

a. Pour $\theta = 0$ et $\theta = 1$ le risque est nul, pourquoi?

b. Lorsque $n \to \infty$, on s'apperçoit d'après la courbe, que $R(\theta, \hat{\theta}_n) \to 0$. C'est–à–dire, lorque $n \to \infty$ l'erreur d'estimation $E_\theta|\hat{\theta}_n - \theta|^2$ tend vers 0, on dit que $\hat{\theta}_n$ est un estimateur consistant au sens de l'erreur quadratique.

Problème 2 : Un exemple en traitement de signal

le problème

On transmet un signal $\mu \in \mathbb{R}$ par l'intermédiaire d'un canal bruité. Ce signal est échantilloné n fois, i.e. on obtient une observation de la forme

$$X_i = \mu + \sigma\, w_i \ , \quad i = 1 \cdots n \ , \tag{B.11}$$

où $\sigma\, w_i$ est un bruit additif. $\sigma \in]0, \infty[$ est l'intensité du bruit. On suppose que $\{w_i; i = 1 \cdots n\}$ est une suite i.i.d. de loi $N(0,1)$ ($\{w_i\}$ est appelé *bruit blanc gaussien*). μ représente le signal que l'on veut transmettre, μ est donc un paramètre inconnu; σ est l'intensité du bruit et le plus souvent il s'agit aussi d'un paramètre inconnu. Le problème est d'estimer les paramètres μ et σ^2. Nous allons d'abord construire le modèle statistique sous–jacent.

Il est clair que $\{X_i; i = 1 \cdots n\}$ donné par (B.11), forme une suite i.i.d. de variables de loi $N(\mu, \sigma^2)$. Donc, si $p(x; \mu, \sigma^2)$ est la densité de la loi de $X = (X_1 \cdots X_n)$ on obtient

$$p(x; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}} \right)^n \exp\left(- \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\,\sigma^2} \right) \ , \quad x = (x_1 \cdots x_n) \in \mathbb{R}^n \ , \tag{B.12}$$

et l'espace des paramètres est

$$\Theta \triangleq \{\theta = (\mu, \sigma^2)\ ;\ \mu \in]-\infty, \infty[\ ,\ \sigma^2 \in]0, \infty[\} \ .$$

On note P_{μ,σ^2} la loi de densité $p(x; \mu, \sigma^2)$ et E_{μ,σ^2} l'espérance associée. P_{μ,σ^2} est la loi $N(\mu, \sigma^2)$.

le problème d'estimation

Nous allons maintenant décrire les estimateurs du maximum de vraisemblance des différents paramètres. Les équations de la vraisemblance s'écrivent

$$\begin{cases} \dfrac{\partial}{\partial \mu} \log\, p(x; \mu, \sigma^2) = 0 \ , \\ \dfrac{\partial}{\partial \sigma^2} \log\, p(x; \mu, \sigma^2) = 0 \ , \end{cases}$$

c'est–à–dire

$$\frac{1}{2\,\sigma^2} \sum_{i=1}^n 2\,(x_i - \mu) = 0 \ , \tag{B.13}$$

$$-\frac{n}{2\,\sigma^2} + \frac{1}{2\,\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \ . \tag{B.14}$$

On considère trois problèmes d'estimation.

a. *Estimation de μ, σ connu (=1)* On résoud donc l'équation (B.13) avec $\sigma = 1$, ce qui donne

$$\hat{\mu}(x) = \frac{x_1 + \cdots + x_n}{n} \ . \tag{B.15}$$

Le biais de cet estimateur est

$$b(\mu; \hat{\mu}) = E_{\mu,1}\, \hat{\mu} - \mu = E\left(\frac{X_1 + \cdots + X_n}{n} \right) - \mu \ ,$$

où $X_1 \cdots X_n$ est une suite i.i.d. de loi $N(\mu, 1)$. Après calcul on obtient $b(\mu; \hat{\mu}) = 0$, donc $\hat{\mu}$ est un estimateur sans biais de μ. Le risque de cet estimateur est

$$R(\mu; \hat{\mu}) = E_{\mu,1} |\hat{\mu} - \mu|^2 = E\left(\frac{X_1 + \cdots + X_n}{n} - \mu \right)^2 \ ,$$

où $X_1 \cdots X_n$ est une suite i.i.d. de loi $N(\mu, 1)$. $\{X_i\}$ est un suite de variables indépendantes donc[3]

$$R(\mu; \hat{\mu}) = \frac{1}{n^2} E\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E(X_i - \mu)^2 \ ,$$

et comme X_i est de loi $N(\mu, 1)$, on obtient finalement $R(\mu; \hat{\mu}) = 1/n$. On va maintenant comparer ce risque à la borne de Cramer–Rao. L'information de Fisher s'écrit

$$I(\mu) = E_{\mu,1}\left(\frac{\partial}{\partial\mu} p(\,\cdot\,; \mu, 1)\right)^2 = E\left(\frac{\partial}{\partial\mu} p(X; \mu, 1)\right)^2 \ ,$$

où $X = (X_1 \cdots X_n)$ est une suite i.i.d. de loi $N(\mu, 1)$. On obtient donc

$$I(\mu) = E\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2 = \sum_{i=1}^{n} E(X_i - \mu)^2 = n \ .$$

Ainsi $R(\mu; \hat{\mu}) = 1/n = 1/I(\mu)$ donc $\hat{\mu}$ est un estimateur efficace de μ.
b. *Estimation de σ, μ connu (=0)* On résoud donc l'équation (B.14) avec $\mu = 0$, ce qui donné[4]

$$\hat{\sigma}^2(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2. \tag{B.16}$$

Le biais de cet estimateur est

$$b(\sigma^2; \hat{\sigma}^2) = E_{0,\sigma^2}\hat{\sigma}^2 - \sigma^2 = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \sigma^2 \ ,$$

où $X_1 \cdots X_n$ est une suite i.i.d. de loi $N(0, \sigma^2)$. Après calcul on obtient $b(\sigma^2; \hat{\sigma}^2) = 0$, donc $\hat{\sigma}^2$ est un estimateur sans biais de σ^2. Le risque de cet estimateur est

$$R(\sigma^2; \hat{\sigma}^2) = E_{0,\sigma^2}|\hat{\sigma}^2 - \sigma^2|^2 = E\left(\frac{X_1^2 + \cdots + X_n^2}{n} - \sigma^2\right)^2 \ ,$$

où $X_1 \cdots X_n$ est une suite i.i.d. de loi $N(0, \sigma^2)$. $\{X_i\}$ est un suite indépendante donc, en utilisant le même argument que pour le calcul de $\hat{\mu}$, on obtient

$$R(\sigma^2; \hat{\sigma}^2) = \frac{1}{n^2} E\left(\sum_{i=1}^{n}(X_i^2 - \sigma^2)\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E(X_i^2 - \sigma^2)^2 \ ,$$

et comme X_i est de loi $N(0, \sigma^2)$ ([5]) , on obtient finalement

$$R(\sigma^2; \hat{\sigma}^2) = \frac{1}{n^2}\sum_{i=1}^{n}\left(E\,X_i^4 - \sigma^4\right) = \frac{1}{n^2}\sum_{i=1}^{n}\left(3\,\sigma^4 - \sigma^4\right) = 2\,\sigma^4/n \ .$$

On va maintenant comparer ce risque à la borne de Cramer–Rao. L'information de Fisher s'écrit

$$I(\sigma^2) = E_{0,\sigma^2}\left(\frac{\partial}{\partial\sigma^2} p(\,\cdot\,; 0, \sigma^2)\right)^2 = E\left(\frac{\partial}{\partial\sigma^2} p(X; 0, \sigma^2)\right)^2 \ ,$$

3. $E\left(\sum_i (X_i - \mu)\right)^2 = \sum_i E(X_i - \mu)^2 + \sum_{i \neq j} E(X_i - \mu)\,(X_j - \mu) = \sum_i E(X_i - \mu)^2$ car si $i \neq j$, $E(X_i - \mu)(X_j - \mu) = E(X_i - \mu)\,E(X_j - \mu) = 0$

4. on utilise constamment le principe d'invariance: $\widehat{\sigma^2} = \hat{\sigma}^2$.

5. utiliser le fait que si $Y \sim N(\mu, \sigma^2)$, alors $E(Y - \mu)^4 = 3\,\sigma^4$

où $X = (X_1 \cdots X_n)$ est une suite i.i.d. de loi $N(0, \sigma^2)$. On obtient donc[6]

$$I(\sigma^2) = E\left(-\frac{n}{2\,\sigma^2} + \frac{1}{2\,\sigma^4}\sum_{i=1}^{n} X_i^2\right)^2 = \frac{n}{2\,\sigma^4}\ .$$

Dans cette dernière égalité on a utilisé le fait que

$$\begin{aligned} E\left(\sum_{i=1}^{n} X_i^2\right)^2 &= E\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^2\, X_j^2\right) \\ &= \sum_{i=1}^{n} E\, X_i^4 + \sum_{i\neq j} E(X_i^2)\, E(X_j^2) \\ &= n\, 3\,\sigma^4 + n\,(n-1)\,\sigma^4\ . \end{aligned}$$

Ainsi $R(\sigma^2; \hat{\sigma}^2) = 2\,\sigma^4/n = 1/I(\sigma^2)$ donc $\hat{\sigma}^2$ est un estimateur efficace de σ^2.
c. *Estimation de μ et σ* On résoud simultanément les équations (B.13) et (B.14), on obtient

$$\boxed{\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \hat{\sigma}^2(x) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu}(x))^2}$$

Pour $\hat{\mu}$ on a

$$\boxed{E_{\mu,\sigma^2}\hat{\mu}^2 = \mu \quad E_{\mu,\sigma^2}|\hat{\mu}^2 - \mu|^2 = \frac{\sigma^2}{n}}$$

Donc $\hat{\mu}$ est un estimateur sans biais. Calculons le biais de l'estimateur $\hat{\sigma}^2$.

$$E_{\mu,\sigma^2}\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\ ,$$

où $X = (X_1 \cdots X_n)$ est une suite i.i.d. de variable de loi $N(\mu, \sigma^2)$.

$$\begin{aligned} & E\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_i\right)^2 \\ &= E\left((X_i - \mu) - \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)\right)^2 \\ &= \sigma^2 - \frac{2}{n}E\left((X_i - \mu)\sum_{j=1}^{n}(X_j - \mu)\right) + \frac{1}{n^2}E\left(\sum_{j=1}^{n}(X_j - \mu)\right)^2\ . \end{aligned}$$

Si $i \neq j$ alors $E[(X_i - \mu)\,(X_j - \mu)] = E(X_i - \mu)\,E(X_j - \mu) = 0$, donc

$$E\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_i\right)^2 = \sigma^2 - \frac{2}{n}\sigma^2 + \frac{1}{n^2}\,n\,\sigma^2\ .$$

On obtient finalement

$$E_{\mu,\sigma^2}\hat{\sigma}^2 = \sum_{i=1}^{n} E\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_i\right)^2 = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2\ .$$

6. Nous utiliserons souvent le résultat suivant: $(\sum_i a_i)^2 = \sum_i a_i^2 + \sum_{i\neq j} a_i\, a_j$.

Donc $\hat{\sigma}^2$ est un estimateur biaisé de σ^2. Maintenant si on pose

$$\boxed{\hat{s}^2(x) \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}(x))^2 = \frac{n}{n-1}\,\hat{\sigma}^2(x)\ ,}$$

alors $\hat{s}^2$ est un estimateur non biaisé de σ^2.

$$\boxed{E_{\mu,\sigma^2}\hat{s}^2 = \sigma^2}$$

Le cacul du risque de $\hat{s}^2$ (ou de $\hat{\sigma}^2$) est en revanche plus difficile. On peut utiliser le fait que

$$\frac{(n-1)\,\hat{s}^2(X)}{\sigma^2} = \sum_{i=1}^{n}\frac{(X_i - \hat{\mu}(X))^2}{\sigma^2}$$

suit une loi du χ^2 à $n-1$ degrés de liberté, on sait alors que[7]

$$E_{\mu,\sigma^2}\left(\frac{(n-1)\,\hat{s}^2}{\sigma^2} - (n-1)\right)^2 = E\left(\frac{(n-1)\,\hat{s}^2(X)}{\sigma^2} - (n-1)\right)^2 = n-1\ ,$$

donc

$$\boxed{E_{\mu,\sigma^2}\left(\hat{s}^2 - \sigma^2\right)^2 = \frac{2\,\sigma^4}{n-1}}$$

Donc $E_{\mu,\sigma^2}\,\hat{s}^4 = (n+1)\,\sigma^4/(n-1)$. Vérifier ce dernier résultat sans utiliser la loi du χ^2 est plus long, en effet

$$\begin{aligned} E_{\mu,\sigma^2}\,\hat{s}^4 &= E\,\hat{s}^4(X)\ , \\ &= \frac{1}{(n-1)^2}\,E\left(\sum_i (X_i - \hat{\mu}(X))^2\right)^2\ , \\ &= \frac{1}{(n-1)^2}\,E\left(\sum_i (X_i - \mu)^2 - n\,(\mu(X) - \mu)^2\right)^2\ , \end{aligned} \tag{B.17}$$

où $X = (X_1 \cdots X_n)$ est une n–échantillon de la loi $N(\mu, \sigma^2)$. En développant le carré dans (B.17) on obtient

$$\begin{aligned} E_{\mu,\sigma^2}\,\hat{s}^4 = \frac{1}{(n-1)^2}\Bigg(& E\left(\sum_i (X_i - \mu)^2\right)^2 \\ & - 2\,n\,E\left(\sum_i (X_i - \mu)^2\,(\hat{\mu}(X) - \mu)^2\right) + n^2\,E(\mu(X) - \mu)^4\Bigg)\ . \end{aligned} \tag{B.18}$$

Posons

$$T_1 = E\left(\sum_i (X_i - \mu)^2\right)^2\ , \tag{B.19}$$

$$T_2 = E\left(\sum_i (X_i - \mu)^2\,(\hat{\mu}(X) - \mu)^2\right)\ , \tag{B.20}$$

$$T_3 = E(\hat{\mu}(X) - \mu)^4\ . \tag{B.21}$$

7. Si Y suit une loi du χ^2 à $n-1$ degrés de liberté, alors $E\,Y = n-1$ et $E(Y - E\,Y)^2 = n-1$ (cf. annexe **??** p.**??**).

FIG. B.1 – *observation*

Nous allons successivement calculer ces trois termes.

$$\begin{aligned} T1 &= E\left(\sum_i (X_i-\mu)^4 + \sum_{i\neq j}(X_i-\mu)^2\,(X_j-\mu)^2\right) \\ &= \sigma^4\,(3\,n + n(n-1)) \ . \end{aligned} \tag{B.22}$$

$$\begin{aligned} T_2 &= E\left(\sum_i (X_i-\mu)^2\left(\frac{1}{n}\sum_i (X_i-\mu)\right)^2\right) \\ &= \frac{1}{n^2}\,E\left(\sum_i (X_i-\mu)^2\left[\sum_i (X_i-\mu)^2 + \sum_{i\neq j}(X_i-\mu)\,(X_j-\mu)\right]\right) \\ &= \frac{1}{n^2}\,E\left(\sum_i (X_i-\mu)^2\right)^2 + \frac{1}{n^2}\sum_{\{i,j,k;j\neq k\}} E\left[(X_i-\mu)^2\,(X_j-\mu)\,(X_k-\mu)\right] \ , \end{aligned}$$

pour tout i, j, k tels que $j \neq k$ on a $E\left[(X_i-\mu)^2\,(X_j-\mu)\,(X_k-\mu)\right] = 0$, en utilisant (B.22) on obtient

$$T_2 = \frac{\sigma^4}{n^2}\,(3\,n + n(n-1)) \ . \tag{B.23}$$

Enfin, on vérifie aisément que $\hat{\mu}(X)$ suit une loi $N(\mu, \sigma^2/n)$, en effet $\hat{\mu}(X)$ est une somme de variables gaussiennes donc $\hat{\mu}(X)$ est gaussienne, son espérance et sa variance on été calculées plus haut. On en déduit

$$T_3 = E(\hat{\mu}(X)-\mu)^4 = 3\,\frac{\sigma^4}{n^2} \ , \tag{B.24}$$

En utilisant (B.22), (B.23) et (B.24) dans (B.18) et (B.19–B.21), on obtient

$$E_{\mu,\sigma^2}\,\hat{s}^4 = \frac{T_1 - 2\,n\,T_2 + n^2\,T_3}{(n-1)^2} = \frac{n+1}{n-1}\,\sigma^4 \ .$$

D'où le résultat cherché.

Problème 3 : Un exemple en traitement de signal

On veut transmettre un signal continu $\mu \in \mathbb{R}$ par l'intermédiaire d'un canal bruité. Le signal est segmenté N fois, en sortie du canal le signal obtenu ($n \to X_n$) est représenté sur la figure B.1. Ce signal de sortie se modélise ainsi: on observe μ plus un bruit additif d'intensité inconnue. Si on note $\{X_n\,;\,n = 1, \ldots, N\}$ le signal de sortie, celui–ci peut s'écrire

$$X_n = \mu + \sigma\,w_n \ , \quad n = 1 \cdots N \ ,$$

où $\sigma > 0$ et $\{w_n\}$ est un bruit blanc gaussien, i.e.

$$w_1, w_2, \ldots, w_N \quad \text{est une suite i.i.d. de variables } N(0,1) \ .$$

FIG. B.2 – *estimation*

Les estimateurs de μ et σ sont

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2 \ .$$

Ces deux estimateurs ce calcul de manière récursive

$$\begin{aligned} \hat{\mu}_{n+1} &= \hat{\mu}_n + \frac{X_{n+1} - \hat{\mu}_n}{n+1} \ , \\ \hat{\sigma}_{n+1}^2 &= \frac{n}{n+1}\left(\hat{\sigma}_n^2 + (\hat{\mu}_{n+1} - \hat{\mu}_n)^2\right) + \frac{1}{n+1}\left(X_{n+1} - \hat{\mu}_{n+1}\right)^2 \ , \end{aligned}$$

avec $\hat{\mu}_0 = 0$ et $\hat{\sigma}_0^2 = 0$. Avec les observations de la figure B.1, on obtient le résultat représenté sur la figure B.2, sur cette figure sont tracées les trois courbes suivantes

- $n \to \hat{\mu}_n + 2\sqrt{\hat{\sigma}_n^2/(n-1)}$.
- $n \to \hat{\mu}_n$.
- $n \to \hat{\mu}_n - 2\sqrt{\hat{\sigma}_n^2/(n-1)}$.

On sait que

$$\sqrt{\frac{n-1}{\hat{\sigma}_n^2}}\,(\hat{\mu}_n - \mu) \quad \text{suit une loi du } \chi^2 \text{ à } n-1 \text{ degrés de liberté} \ ,$$

en utilisant une table on peut alors vérifier que

$$P\left(\left|\sqrt{\frac{n-1}{\hat{\sigma}_n^2}}\,(\hat{\mu}_n - \mu)\right| \leq 2\right) \geq 0.95 \ ,$$

pour tout $n \geq 60$. C'est–à–dire

$$\mu \in \left[\hat{\mu}_n - 2\sqrt{\frac{\hat{\sigma}_n^2}{n-1}} \ , \ \hat{\mu}_n + 2\sqrt{\frac{\hat{\sigma}_n^2}{n-1}}\right] \quad \text{avec une probalilité} \geq 0.95,$$

pour tout $n \geq 60$.

Le Programme FORTRAN

Les courbes de simulation et d'estimations ont été obtenues avec $\mu = 5$ et $\sigma^2 = 2$.

```
      program simu
c ==========================================================
c     simulation d'une observation de la forme
c     X(i) = mu + sqrt(sigma2) w(i)  , i=1,...,n
c     w(i) bruit blanc gaussien  (i.i.d. N(0,1))
```

```
c     et estimation de mu et sigma2 par les moments
c     empiriques calcules de maniere recursive
c ==========================================================
c
      real      mu , muest
      dimension x(1000)
      data      pi /3.1415926535/
c
c-----lecture des donnees
c
      write(0,900)
      write(0,901)
c     nombre d'observations
      read(0,*)     n
      write(0,902)
c     moyenne
      read(0,*)     mu
      write(0,903)
c     variance
      read(0,*)     sigma2
      write(0,904)
c     germe du generateur aleatoire
      read(0,*)     seed
c
c-----simulation
c
      do 10 i=1,n,2
c
c         tir d'une loi uniforme
          call random_$uniform(seed,u1)
          call random_$uniform(seed,u2)
c         transformation de Box et Muller
          u3 = sqrt( -2*alog(u1) )
          v1 = u3 * sin(2*pi*u2)
          v2 = u3 * cos(2*pi*u2)
c         observations X(i) et X(i+1)
          x(i)   = mu + sqrt(sigma2)*v1
          x(i+1) = mu + sqrt(sigma2)*v2
c
10    continue
c
c-----estimation
c
      open(1,form='unformatted',file='simu.data')
c
c     initialisation
      muest = 0.
      siest = 0.
c
      do 20 i=1,n
c
c         formules recursives pour la moyenne et la variance
          dmu   = (x(i)-muest) / real(i)
          muest = muest + dmu
          siest = (i-1) *(siest+dmu*dmu)/real(i)
     &            + (x(i)-muest)*(x(i)-muest)/real(i)
c
          write(1) i,x(i),mu,sigma2,muest,siest
```

```
c
20    continue
c
      close (1)
      write(0,905)
c
900   format (/'Simulation d'une suite X(i) (i=1...n) i.i.d.'
     &        /'N(mu,sigma2)'
     &        /'et estimation de mu et sigma2'/)
901   format (' n                      : ',$)
902   format (' mu                     : ',$)
903   format (' sigma2                 : ',$)
904   format (' germe du tir aleatoire : ',$)
905   format (/'STOP (resultat -> simu.data)'/)
c
      stop
      end
```

Problème 4 : Sujet d'examen 86–87

l'énoncé

(1) Soit $X_1, X_2, \ldots, X_n$ un n–échantillon d'un loi gaussienne $N(\mu, 1)$, $\mu \in \Theta$.

a. Ecrire $\hat{\mu}(x)$ l'EMV (estimateur du maximum de vraisemblance) de μ dans le cas où $\Theta = \mathbb{R}$.

b. Calculer $\hat{\mu}_0(x)$ l'EMV de μ dans le cas où $\Theta = [-1, +1]$. [rép.:

$$\hat{\mu}_0(x) = \begin{cases} -1 & \text{si} \quad \bar{x} < -1 \ , \\ \bar{x} & \text{si} \quad -1 \leq \bar{x} \leq +1 \ , \\ +1 & \text{si} \quad +1 < \bar{x} \ , \end{cases}$$

où $\bar{x} = \sum_{i=1}^n x_i/n$].

(2) Soit $X_1, X_2, \ldots, X_n$ un n–échantillon d'une loi gaussienne $N(\mu, 1)$, $\mu \in \Theta = \mathbb{R}$. On veut tester l'hypothèse

$$H_0 \quad : \qquad \text{“}\mu \in [-1, +1]\text{''},$$

et on se propose pour cela d'écrire le TRV (test du rapport de vraisemblance).

a. Quel est l'espace Θ_0 des paramètres correspondant à l'hypothèse H_0?

b. Calculer $l(x) = \log \lambda(x)$ où $\lambda(x)$ est le rapport de vraisemblance de l'hypothèse H_0.

c. La forme de la région de rejet est donc $D = \{x \in \mathbb{R}^n \ ; \log \lambda(x) \leq c\}$ où $c < 0$. Expliciter la forme de D^c le complémentaire de la région de rejet D [rép.: $D^c = \{x \ ; -1 - s < \hat{\mu}(x) < 1 + s\}$ où $s > 0$].

d. (Difficile) On veut déterminer s tel que la probabilité d'erreur de type I soit inférieure à 0.05. Expliquer comment calculer s.

e. Expliquer pourquoi la forme de la région de rejet de l'hypothèse H_0 est naturelle.

correction

Le log de la densité de l'échantillon $X_1, \ldots, X_n$ est

$$\begin{aligned} \log p(x; \mu) &= \log \left[\left(\frac{1}{\sqrt{2\,\pi}} \right)^n \exp \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right) \right] \ , \\ &= -\frac{n}{2} \log(2\,\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \ . \end{aligned} \tag{B.25}$$

Si $\Theta = \mathbb{R}$, l'EMV de μ est

$$\hat{\mu}(x) = \operatorname{Arg} \max_{\mu \in \mathbb{R}} \log p(x; \mu) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \ . \tag{B.26}$$

Si $\Theta = [-1, +1]$, l'EMV de μ est

$$\begin{aligned}
\hat{\mu}_0(x) &= \operatorname{Arg} \max_{-1 \le \mu \le +1} \log p(x;\mu) \\
&= \operatorname{Arg} \max_{-1 \le \mu \le +1} -\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \\
&= \operatorname{Arg} \min_{-1 \le \mu \le +1} \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \\
&= \operatorname{Arg} \min_{-1 \le \mu \le +1} \left[\mu^2 - 2\mu \frac{1}{n} \sum_{i=1}^{n} x_i + \frac{1}{n} \sum_{i=1}^{n} x_i^2 \right] \\
&= \operatorname{Arg} \min_{-1 \le \mu \le +1} \left[\mu^2 - 2\mu \bar{x} + \frac{1}{n} \sum_{i=1}^{n} x_i^2 \right] .
\end{aligned}$$

Posons $F(\mu) = \mu^2 - 2\mu\bar{x} + \frac{1}{n}\sum x_i^2$. On veut donc minimiser $F(\mu)$ sous la contrainte $\mu \in [-1, +1]$. Trois cas

- Si $\bar{x} \in [-1, +1]$ alors le minimum de $F(\mu)$ sur $[-1, +1]$ est atteint en $\bar{x}$.
- Si $\bar{x} < -1$ alors le minimum de $F(\mu)$ sur $[-1, +1]$ est atteint en -1.
- Si $\bar{x} > +1$ alors le minimum de $F(\mu)$ sur $[-1, +1]$ est atteint en $+1$.

(Pour s'en convaincre il suffit de tracer le graphe de la fonction $\mathbb{R} \ni \mu \to F(\mu) \in [0, \infty[$). D'où le résultat cherché

$$\hat{\mu}_0(x) = \begin{cases} -1 & \text{si} \quad \bar{x} < -1 \ , \\ \bar{x} & \text{si} \quad -1 \le \bar{x} \le +1 \ , \\ +1 & \text{si} \quad +1 < \bar{x} \ . \end{cases} \tag{B.27}$$

On suppose que $\mu \in \Theta = \mathbb{R}$ et on veut tester l'hypothèse

$$H_0 \quad : \quad \text{“ } \mu \in [-1, +1] \text{ ”}.$$

L'espace des paramètres associé à cette hypothèse est $\Theta_0 = [-1, +1]$. La région de rejet de H_0 est donc de la forme

$$D = \{ x \in \mathbb{R}^n \ ; \log \lambda(x) \le c \} \ ,$$

où

$$\lambda(x) \triangleq \frac{\max_{\mu \in \Theta_0} p(x;\mu)}{\max_{\mu \in \Theta} p(x;\mu)} \ ,$$

avec $c < 0$. Mais d'après la première partie

$$\begin{aligned}
\log \lambda(x) &= \log \max_{\mu \in \Theta_0} p(x;\mu) - \log \max_{\mu \in \Theta} p(x;\mu) \\
&= \log p(x; \hat{\mu}_0(x)) - \log p(x; \hat{\mu}(x)) \\
&= -\frac{1}{2} \sum_{i=1}^{n} \left[(x_i - \hat{\mu}_0)^2 - (x_i - \hat{\mu})^2 \right] \\
&= -\frac{1}{2} \sum_{i=1}^{n} (x_i - \hat{\mu}_0 + x_i - \hat{\mu}) \ (x_i - \hat{\mu}_0 - x_i + \hat{\mu}) \\
&= -\frac{n}{2} (\hat{\mu} - \hat{\mu}_0) \frac{1}{n} \sum_{i=1}^{n} (2\, x_i - \hat{\mu}_0 - \hat{\mu}) \\
&= -\frac{n}{2} (\hat{\mu} - \hat{\mu}_0)(2\, \hat{\mu} - \hat{\mu}_0 - \hat{\mu}) \\
&= -\frac{n}{2} (\hat{\mu} - \hat{\mu}_0)^2 \ .
\end{aligned}$$

Donc D est de la forme $\{ x \, ; \, (\hat{\mu} - \hat{\mu}_0)^2 \ge c' \}$ avec $c' = -2c/n > 0$. Donc, avec (B.27), on obtient

$$\begin{aligned}
D^c &= \left\{ x \, ; \, (\hat{\mu}(x) - \hat{\mu}_0(x))^2 < c' \right\} \\
&= \left\{ x \, ; \, -\sqrt{c'} < \hat{\mu}(x) - \hat{\mu}_0(x) < \sqrt{c'} \right\}
\end{aligned}$$

$$
\begin{aligned}
&= \left\{ x\,;\, -\sqrt{c'} < \hat{\mu}(x) - \hat{\mu}_0(x) < \sqrt{c'}\ ,\ \hat{\mu}(x) < -1 \right\} \\
&\quad \cup \left\{ x\,;\, -\sqrt{c'} < \hat{\mu}(x) - \hat{\mu}_0(x) < \sqrt{c'}\ ,\ -1 \le \hat{\mu}(x) \le +1 \right\} \\
&\quad \cup \left\{ x\,;\, -\sqrt{c'} < \hat{\mu}(x) - \hat{\mu}_0(x) < \sqrt{c'}\ ,\ +1 < \hat{\mu}(x) \right\} \\
&= \left\{ x\,;\, -\sqrt{c'} - 1 < \hat{\mu}(x) < \sqrt{c'} - 1\ ,\ \hat{\mu}(x) < -1 \right\} \\
&\quad \cup \left\{ x\,;\, -\sqrt{c'} < 0 < \sqrt{c'}\ ,\ -1 \le \hat{\mu}(x) \le +1 \right\} \\
&\quad \cup \left\{ x\,;\, -\sqrt{c'} + 1 < \hat{\mu}(x) < \sqrt{c'} + 1\ ,\ +1 < \hat{\mu}(x) \right\} \\
&= \left\{ x\,;\, -\sqrt{c'} - 1 < \hat{\mu}(x) < -1 \right\} \\
&\quad \cup \{ x\,;\, -1 \le \hat{\mu}(x) \le +1 \} \\
&\quad \cup \left\{ x\,;\, +1 < \hat{\mu}(x) < \sqrt{c'} + 1 \right\} \\
&= \left\{ x\,;\, -\sqrt{c'} - 1 < \hat{\mu}(x) < \sqrt{c'} + 1 \right\}\ ,
\end{aligned}
$$

on pose $s = \sqrt{c'} > 0$, le complémentaire de la région de rejet a donc pour forme

$$D^c = \{ x \in \mathbb{R}^n\,;\, -1 - s < \hat{\mu}(x) < 1 + s \}\ .$$

On veut calculer le seuil s tel que

$$P(\text{ rejeter } H_0 \mid H_0 \text{ est vraie }) \le 0.05\ ,$$

c'est–à–dire

$$
\begin{aligned}
\max_{-1 \le \mu \le +1} P_\mu(D) &\le 0.05\ , \\
\max_{-1 \le \mu \le +1} (1 - P_\mu(D^c)) &\le 0.05\ , \\
\max_{-1 \le \mu \le +1} (1 - P\,(-1 - s < \hat{\mu} < 1 + s)) &\le 0.05\ ,
\end{aligned}
$$

où $\hat{\mu} \sim N(\mu, 1/n)$. On obtient donc???

$$
\begin{aligned}
\min_{-1 \le \mu \le +1} P(-1 - s < \hat{\mu} < 1 + s) &> 0.95\ , \\
\text{i.e.} \quad \min_{-1 \le \mu \le +1} P(\sqrt{n}(-1 - s - \mu) < N(0,1) < \sqrt{n}(1 + s - \mu)) &> 0.95\ .
\end{aligned}
$$

Le minimum dans l'expression précédente est atteint en $\mu = -1$ ou $\mu = +1$, d'où

$$P\left(-\sqrt{n}s < N(0,1) < \sqrt{n}(2 + s)\right) > 0.95\ ,$$

c'est–à–dire, s est choisi tel que

$$\int_{-\sqrt{n}s}^{\sqrt{n}(2+s)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx > 0.95\ .$$

Problème 5 : Sujet d'examen 87–88

l'énoncé

On considère trois échantillons indépendants

- $X_1, \ldots, X_n$ un n–échantillon d'une loi $N(\mu_1, \sigma^2)$,
- $Y_1, \ldots, Y_m$ un m–échantillon d'une loi $N(\mu_2, \sigma^2)$,
- $Z_1, \ldots, Z_l$ un l–échantillon d'une loi $N(\mu_3, \sigma^2)$.

Le paramètre $\theta = (\mu_1, \mu_2, \mu_3, \sigma^2)$ est inconnu. Le but du problème est de construire le test du maximum de vraisemblance de l'hypothèse

$$H_0\ :\ \text{“}\mu_1 = \mu_2 = \mu_3\text{”}\ .$$

(1) Ecrire Θ l'espace des paramètres associé à ce problème et Θ_0 l'espace des paramètres associé à l'hypothèse H_0.

(2) Calculer $p(x, y, z; \theta)$ ($x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_m)$, $z = (z_1, \ldots, z_l)$) la densité de la loi du vecteur aléatoire $(X_1, \ldots, X_n, Y_1, \ldots, Y_m, Z_1, \ldots, Z_l)$.
[rép.:

$$p(x, y, z; \theta) = \prod_{i=1}^{n} f(x_i; \mu_1, \sigma^2) \prod_{i=1}^{m} f(y_i; \mu_2, \sigma^2) \prod_{i=1}^{l} f(z_i; \mu_3, \sigma^2) \ ,$$

où $f(a; \mu, \sigma^2)$ est la densité d'une loi $N(\mu, \sigma^2)$ réelle.]

(3) Calculer $\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\sigma}^2)$ l'estimateur du maximum de vraisemblance de θ (poser et résoudre l'équation de la vraisemblance).

(4) Calculer $\hat{\theta}_0$ l'estimateur du maximum de vraisemblance de θ sous l'hypothèse H_0 (poser et résoudre l'équation de la vraisemblance).
[rép: $\hat{\theta}_0 = (\hat{\mu}_0, \hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0^2)$ où

$$\begin{aligned}
\hat{\mu}_0 &\triangleq \frac{1}{n+m+l} \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} y_i + \sum_{i=1}^{l} z_i \right) \ , \\
\hat{\sigma}_0^2 &\triangleq \frac{1}{n+m+l} \left(\sum_{i=1}^{n} (x_i - \hat{\mu}_0)^2 + \sum_{i=1}^{m} (y_i - \hat{\mu}_0)^2 + \sum_{i=1}^{l} (z_i - \hat{\mu}_0)^2 \right) \ .]
\end{aligned}$$

(5) Ecrire la région de rejet de l'hypothèse H_0.
[rép: on rejette l'hypothèse H_0 si

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \leq c \ .]$$

correction

(1) Les espaces des paramètres sont

$$\begin{aligned}
\Theta &\triangleq \left\{ \theta = (\mu_1, \mu_2, \mu_3, \sigma^2) \ ; \ \mu_i \in \mathbb{R}, \sigma^2 > 0 \right\} \ , \\
\Theta_0 &\triangleq \left\{ \theta = (\mu, \mu, \mu, \sigma^2) \ ; \ \mu \in \mathbb{R}, \sigma^2 > 0 \right\} \ .
\end{aligned}$$

(2) $\mathcal{X} = (X_1, \ldots, X_n, Y_1, \ldots, Y_m, Z_1, \ldots, Z_l)$ est un vecteur aléatoire de $N = n + m + l$ composantes mutuellement indépendantes (car $\mathcal{X}$ est composé de trois échantillons indépendants). De plus chaque composante est gaussienne, $X_i \sim N(\mu_1, \sigma^2)$, $Y_i \sim N(\mu_2, \sigma^2)$, $Z_i \sim N(\mu_3, \sigma^2)$, donc la densité de la loi de $\mathcal{X}$ est

$$\begin{aligned}
&\log p(x, y, z; \theta) = \\
&= \log p(x, y, z; \mu_1, \mu_2, \mu_3, \sigma^2) \\
&= \log \left(\prod_{i=1}^{n} f(x_i; \mu_1, \sigma^2) \prod_{i=1}^{m} f(y_i; \mu_2, \sigma^2) \prod_{i=1}^{l} f(z_i; \mu_3, \sigma^2) \right) \\
&= -\frac{n+m+l}{2} \log(2\,\pi) - \frac{n+m+l}{2} \log(\sigma^2) \\
&\quad - \frac{1}{2\,\sigma^2} \left(\sum_{i=1}^{n} (x_i - \mu_1)^2 + \sum_{i=1}^{m} (y_i - \mu_2)^2 + \sum_{i=1}^{l} (z_i - \mu_3)^2 \right) \ .
\end{aligned}$$

(3) L'équation de la vraisemblance s'écrit

$$\left\{ \begin{array}{lcl}
\frac{\partial}{\partial \mu_1} \log p(x, y, z; \mu_1, \mu_2, \mu_3, \sigma^2) &=& 0 \ , \\
\frac{\partial}{\partial \mu_2} \log p(x, y, z; \mu_1, \mu_2, \mu_3, \sigma^2) &=& 0 \ , \\
\frac{\partial}{\partial \mu_3} \log p(x, y, z; \mu_1, \mu_2, \mu_3, \sigma^2) &=& 0 \ , \\
\frac{\partial}{\partial \sigma^2} \log p(x, y, z; \mu_1, \mu_2, \mu_3, \sigma^2) &=& 0 \ .
\end{array} \right.$$

On obtient

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i \ ,$$

$$
\begin{aligned}
\hat{\mu}_2 &= \frac{1}{m}\sum_{i=1}^{m} y_i \ , \\
\hat{\mu}_3 &= \frac{1}{l}\sum_{i=1}^{l} z_i \\
\hat{\sigma}^2 &= \frac{1}{n+m+l}\left(\sum_{i=1}^{n}(x_i-\hat{\mu}_1)^2+\sum_{i=1}^{m}(y_i-\hat{\mu}_2)^2+\sum_{i=1}^{l}(z_i-\hat{\mu}_3)^2\right) \ .
\end{aligned}
$$

(4) L'équation de la vraisemblance s'écrit

$$
\left\{
\begin{array}{lcl}
\frac{\partial}{\partial\mu}\ \log p(x,y,z;\mu,\mu,\mu,\sigma^2) &=& 0 \ , \\
\frac{\partial}{\partial\sigma^2}\ \log p(x,y,z;\mu,\mu,\mu,\sigma^2) &=& 0 \ .
\end{array}
\right.
$$

On obtient

$$
\begin{aligned}
\hat{\mu}_0 &= \frac{1}{n+m+l}\left(\sum_{i=1}^{n} x_i+\sum_{i=1}^{m} y_i+\sum_{i=1}^{l} z_i\right) \ , \\
\hat{\sigma}_0^2 &= \frac{1}{n+m+l}\left(\sum_{i=1}^{n}(x_i-\hat{\mu}_0)^2+\sum_{i=1}^{m}(y_i-\hat{\mu}_0)^2+\sum_{i=1}^{l}(z_i-\hat{\mu}_0)^2\right) \ .
\end{aligned}
$$

(5) La région de rejet de l'hypothèse H_0 est

$$
D \triangleq \left\{(x,y,z)\ ;\ \frac{\max_{\theta\in\Theta_0} p(x,y,z;\theta)}{\max_{\theta\in\Theta} p(x,y,z;\theta)} \leq s\right\} = \left\{(x,y,z)\ ;\ \frac{p(x,y,z;\hat{\theta}_0)}{p(x,y,z;\hat{\theta})} \leq s\right\} \ ,
$$

et

$$
\begin{aligned}
p(x,y,z;\hat{\theta}) &= \frac{1}{(2\pi\,\hat{\sigma}^2)^{(n+m+l)/2}} e^{(n+m+l)/2} \ , \\
p(x,y,z;\hat{\theta}_0) &= \frac{1}{(2\pi\,\hat{\sigma}_0^2)^{(n+m+l)/2}} e^{(n+m+l)/2} \ ,
\end{aligned}
$$

donc

$$
\frac{p(x,y,z;\hat{\theta})}{p(x,y,z;\hat{\theta}_0)} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{(n+m+l)/2} \ .
$$

Ainsi la région de rejet est de la forme $\{\hat{\sigma}^2/\hat{\sigma}_0^2 \leq c\}$.

Index

échantillon, 10
évènement, 1

biais asymptotique, 18
biais d'un estimateur, 11

covariance, 9
Cramer–Rao (inégalite de), 12

densité conjointe d'un échantillon, 11
densité mode d'une, 16

espérance mathématique, 8
espace de probabilité, 1
espace des observations, 10
espace probabilisable, 1
estimateur, 11
estimateur asymptotiquement normal, 18
estimateur bayésien, 15
estimateur consistant, 17
estimateur du maximum de vraisemblance, 12, 14
estimateur efficace, 12
estimateur MAP, 17
estimateur préférable, 11
estimateur sans biais, 11
estimateur uniformément de variance minimum sans biais, 11
estimation paramétrique, 10

Fisher (information de), 12
fonction de répartition, 6

gamma (fonction), 16

indépendance d'une suite finie de v.a., 3
indépendance d'une suite inifinie de v.a., 3
indépendance de deux événements, 2
indépendance de variables aléatoires, 6
intervalle de confiance, 18

loi de probabilité, 5

matrice de covriance, 9
modèle statistique, 10
moment, 9
moyenne asymptotique, 18

principe d'invariance de l'EMV, 15
probabilité, 1
probabilité conditionnelle, 1

risque bayésien, 15
risque d'un estimateur, 11

tribu, 1
tribu naturelle, 5

variable aléatoire, 5
variable aléatoire intégrable, 8
variable aléatoire quasi–intégrable, 8
variance, 9
variance asymptotique, 18
vecteur aléatoire, 5
vraisemblance équation de la, 14
vraisemblance fonction de, 14