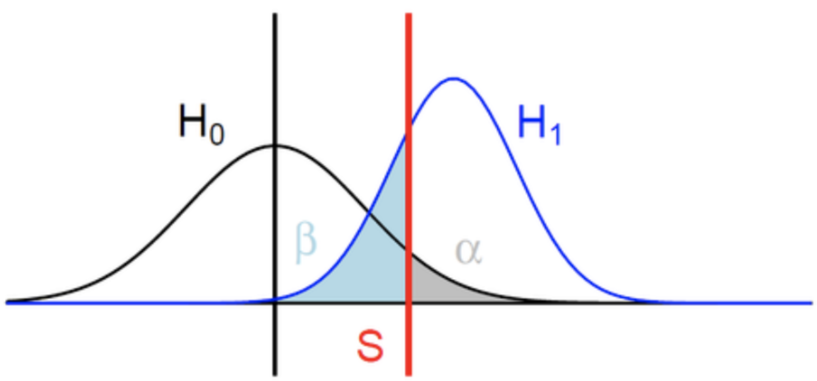
Année 2023-2024

Statistique Mathématique

Vincent Rivoirard



Licence 3 - Mathématiques appliquées
Département MIDO
Université Paris Dauphine - PSL

Table des matières

Avant-propos

Ces notes de cours présentent une introduction à la théorie statistique classique. Elles prolongent le cours du premier semestre en insistant sur le formalisme mathématique de la statistique. On se place donc volontairement dans un cadre un peu abstrait. L'objectif de ce cours est de présenter des méthodes quantitatives qui s'appuient sur des principes relativement généraux et qui permettent de retrouver les paramètres d'un modèle et de prendre des décisions à partir d'observations issues de ce modèle. L'approche décrite dans ce polycopié s'appuie sur le triptyque classique « estimer – encadrer – tester ». On accorde une attention particulière à la question des tests d'hypothèses qui illustrent les difficultés inhérentes à la modélisation mathématique de questions simples et naturelles que l'on se pose dans des situations très concrètes.

Chapitre 1

Outils de probabilité

L'objectif de ce chapitre est de décrire les outils de probabilité qui seront indispensables pour l'étude mathématique des problématiques statistiques envisagées dans les chapitres ultérieurs.

1.1 Loi d'une variable aléatoire réelle

On désigne par $(\Xi, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Les points $w \in \Xi$ s'interprètent comme les résultats d'une expérience aléatoire. On s'intéresse aux événements, c'est-à-dire aux éléments de la tribu $\mathcal{A}$. On rappelle dans la suite un certain nombre de notions et de propriétés associées. On note $\mathcal{B}$ la tribu des boréliens sur $\mathbb{R}$.

Définition 1.1. *Une **variable aléatoire réelle** (souvent abrégée en v.a.r.) est une application mesurable*

$$X : (\Xi, \mathcal{A}) \longmapsto (\mathbb{R}, \mathcal{B}).$$

Définition 1.2. *On appelle **loi ou distribution** d'une v.a.r. X la mesure image de $\mathbb{P}$ par X sur $(\mathbb{R}, \mathcal{B})$, notée $\mathbb{P}^X$ et définie, pour tout $B \in \mathcal{B}$, par*

$$\mathbb{P}^X(B) := \mathbb{P}(X \in B).$$

A chaque variable aléatoire réelle, on peut associer sa fonction de répartition.

Définition 1.3. *La **fonction de répartition** de la variable aléatoire réelle X est l'application $F : \mathbb{R} \longmapsto [0; 1]$ définie, pour $x \in \mathbb{R}$, par*

$$F(x) := \mathbb{P}(X \leq x) = \mathbb{P}\Big(\big\{w \in \Xi : \ X(w) \leq x\big\}\Big).$$

Proposition 1.1. *Si F est la fonction de répartition d'une v.a.r. X, elle vérifie :*

1. *F est croissante*

2. *F est continue à droite en tout point de $\mathbb{R}$*
3. *F admet une limite à gauche en tout point de $\mathbb{R}$*
4. *$\lim_{x\to+\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$*

Remarque 1.1. *En notant pour tout $x \in \mathbb{R}$,*

$$F(x^-) := \lim_{u \nearrow x} F(u) = \mathbb{P}(X < x),$$

on a

$$\mathbb{P}(X = x) = F(x) - F(x^-).$$

La fonction de répartition caractérise la loi d'une variable aléatoire via la formule

$$\mathbb{P}^X(]a,b]) = F(b) - F(a), \quad \forall\, (a,b) \in \mathbb{R}^2.$$

Remarque 1.2. *Puisque la fonction de répartition F caractérise $\mathbb{P}^X$, par abus de langage, on dit parfois que F est la loi de X.*

1.1.1 Variables discrètes

Une v.a.r. X est **discrète** si elle prend un ensemble de valeurs au plus dénombrable $\{x_i,\ i \in \mathcal{I}\} \subset \mathbb{R}$, avec $\mathcal{I} \subset \mathbb{Q}$. Dans ce cas, les valeurs $(\mathbb{P}(X = x_i))_{i\in\mathcal{I}}$ caractérisent la loi de X. A noter que $\mathbb{P}(X \notin \{x_i,\ i \in \mathcal{I}\}) = 0$. Les exemples suivants définissent respectivement **les lois de Bernoulli, binomiale, de Poisson et géométrique.**

Exemple 1.1. *Loi de Bernoulli. Pour $p \in]0;1[$, on note $X \sim Ber(p)$ si on a $X \in \{0,1\}$ et*

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

Exemple 1.2. *Loi binomiale. Pour $n \in \mathbb{N}^*$ et $p \in]0;1[$, on note $X \sim Bin(n,p)$ si $X \in \{0,1,\dots,n\}$ et*

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

Exemple 1.3. *Loi de Poisson. Pour $\lambda > 0$, on note $X \sim \mathcal{P}(\lambda)$ si $X \in \mathbb{N}$ et*

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \forall\, k \in \mathbb{N}.$$

Exemple 1.4. *Loi géométrique. Pour $p > 0$, on note $X \sim \mathcal{G}(p)$ si $X \in \mathbb{N}^*$ et*

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad \forall\, k \in \mathbb{N}^*.$$

1.1.2 Variables de loi absolument continue

Une v.a.r. X est **de loi absolument continue (ou à densité)** si sa fonction de répartition s'écrit

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad \forall\, x \in \mathbb{R}.$$

La fonction f, définie à un ensemble négligeable près, est une densité de probabilité, i.e.

$$f \geq 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} f(t)dt = 1.$$

Dans ce cas, la fonction de répartition F est dérivable presque partout et on a, pour presque tout x, $F'(x) = f(x)$.

Exercice 1.1. *Redémontrer que si f est continue sur $\mathbb{R}$, alors $\forall\, x \in \mathbb{R}$, $F'(x) = f(x)$.*

Si elle existe, la densité d'une v.a.r. détermine entièrement sa fonction de répartition et donc caractérise sa loi.

Proposition 1.2. *La loi d'une variable absolument continue est diffuse : pour tout $x \in \mathbb{R}$, on a $\mathbb{P}(X = x) = 0$.*

Démonstration. Cela découle du fait que pour tout $x \in \mathbb{R}$, $\int_x^x f(t)dt = 0$. □

Les exemples suivants définissent respectivement **les lois normale, uniforme, exponentielle et de Cauchy.**

Exemple 1.5. *Loi normale (ou gaussienne). Pour $\mu \in \mathbb{R}$ et $\sigma > 0$, on note $X \sim \mathcal{N}(\mu, \sigma^2)$, si X admet pour densité*

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad \forall\, t \in \mathbb{R}.$$

Exemple 1.6. *Loi uniforme. Pour $-\infty < a < b < +\infty$, on note $X \sim U[a; b]$ si X admet pour densité la fonction*

$$f(t) = \frac{1}{b-a} 1_{[a;b]}(t), \quad \forall\, t \in \mathbb{R}.$$

Exemple 1.7. *Loi exponentielle. Pour $\lambda > 0$, on note $X \sim \mathcal{E}(\lambda)$ si X admet pour densité*

$$f(t) = \lambda e^{-\lambda t} 1_{[0;+\infty[}(t), \quad \forall\, t \in \mathbb{R}.$$

Exemple 1.8. *Loi de Cauchy. On note $X \sim \mathcal{C}$ si X admet pour densité*

$$f(t) = \frac{1}{\pi(1+t^2)}, \quad \forall\, t \in \mathbb{R}.$$

Remarque 1.3. *Une v.a.r. peut être ni discrète ni absolument continue. Par exemple, on considère $Y \sim \mathcal{N}(0, 1)$ et on pose $X = Y 1_{\{Y \geq 0\}}$ qui n'est ni discrète ni absolument continue. En effet, elle est à valeurs dans $[0; +\infty[$ mais n'est pas diffuse car $\mathbb{P}(X = 0) =$ 0.5.*

1.1.3 Formules d'intégration

Si X est une v.a.r. de fonction de répartition F et de loi $\mathbb{P}^X$, on a alors pour toute fonction mesurable ϕ bornée, positive ou intégrable (i.e. telle que $\mathbb{E}[|\phi(X)|] < \infty$)

$$\mathbb{E}[\phi(X)] = \int_\Xi \phi(X(w))\mathbb{P}(dw) = \int_\mathbb{R} \phi(x)\mathbb{P}^X(dx) = \text{"} \int_\mathbb{R} \phi(x)dF(x)\text{"}.$$

Exemple 1.9. *Cas discret :* $X \in \{x_i,\ i \in \mathcal{I}\}$

$$\mathbb{E}[\phi(X)] = \sum_{i\in\mathcal{I}} \phi(x_i)\mathbb{P}(X = x_i).$$

Exemple 1.10. *Cas absolument continu :* X *de densité* f

$$\mathbb{E}[\phi(X)] = \int \phi(x)f(x)dx.$$

Exemple 1.11. $X = Y1_{\{Y\geq 0\}}$ *et* $Y \sim \mathcal{N}(0, 1)$.

$$\mathbb{E}[\phi(X)] = \frac{1}{2}\phi(0) + \int_0^{+\infty} \phi(t)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$$

1.2 Paramètre de positions

Etant donnée une v.a.r., on peut chercher à décrire sa loi à l'aide d'indicateurs déterministes les plus simples possibles. On utilise en première approximation 2 indicateurs (s'ils existent) basés sur les 2 premiers moments, ou plus précisément sur la moyenne et la variance. On pourrait également considérer le coefficient d'asymétrie (ou skewness) et le coefficient d'aplatissement (ou kurtosis). Ils sont basés sur les moments d'ordres 3 et 4. Un autre type d'approximation se base sur les quantiles de la loi considérée, qui mesurent, dans un certain sens, la dispersion de la loi. Plus difficiles à manipuler, ils présentent l'avantage d'être toujours définis.

1.2.1 Espérance - variance

Une variable aléatoire réelle X admet un moment d'ordre $p \in \mathbb{N}^*$ si

$$\mathbb{E}[|X|^p] = \int_\Xi |X(w)|^p\, \mathbb{P}(dw) = \int_\mathbb{R} |x|^p\, \mathbb{P}^X(dx) < \infty.$$

Dans ce cas, le moment d'ordre p est

$$\mathbb{E}[X^p] = \int_\Xi X^p(w)\, \mathbb{P}(dw) = \int_\mathbb{R} x^p\, \mathbb{P}^X(dx).$$

Remarque 1.4. *L'intégrale $\int_{\mathbb{R}} x^p \, \mathbb{P}^X(dx)$ s'écrit sous la forme d'une somme dans le cas discret.*

Définition 1.4. *La **moyenne ou espérance** de la v.a.r. X, notée μ_X, si elle existe, est le moment d'ordre 1 de X :*

$$\mu_X := \mathbb{E}[X].$$

*La **variance** de la v.a.r. X, notée σ_X^2, si elle existe, est le moment d'ordre 2 de la v.a.r. X recentrée :*

$$\sigma_X^2 := \mathbb{E}[(X-\mu_X)^2] = \mathbb{E}[X^2] - \mu_X^2.$$

Proposition 1.3. *Si X admet un moment d'ordre 2, alors*

$$\mathbb{E}[(X-\mu_X)^2] = \min_{c\in\mathbb{R}} \mathbb{E}[(X-c)^2].$$

Démonstration. Pour tout $c \in \mathbb{R}$ déterministe, on a la décomposition biais-variance

$$\begin{aligned}\mathbb{E}[(X-c)^2] &= \mathbb{E}\Big[(X-\mu_X)^2 + (\mu_X - c)^2 + 2(X-\mu_X)(\mu_X - c)\Big]\\ &= \sigma_X^2 + (\mu_X - c)^2 + 0.\end{aligned}$$

□

Le couple espérance/variance fournit un indicateur très simple pour contrôler les fluctuations de X autour de μ_X, via **l'inégalité de Bienaymé-Tchebychev** :

$$\mathbb{P}\Big(|X-\mu_X| \geq t\Big) \leq \frac{\sigma_X^2}{t^2}, \quad \forall\, t > 0. \tag{1.1}$$

1.2.2 Quantiles

Définition 1.5. *A toute fonction de répartition F, on associe son **inverse généralisé** $F^{(-1)}$ défini, pour tout $q \in]0;1[$, comme suit :*

$$F^{(-1)}(q) := \inf\big\{x \in \mathbb{R} : \; F(x) \geq q\big\}.$$

*La fonction $F^{(-1)}$ est aussi appelée la **fonction quantile**.*

Proposition 1.4. *On a les résultats suivants :*

1. *Lorsque F est bijective, $F^{(-1)} = F^{-1}$.*
2. *$F^{(-1)}$ est croissante.*
3. *Pour tout $x \in \mathbb{R}$, pour tout $q \in]0;1[$, on a :*

$$F(x) \geq q \iff x \geq F^{(-1)}(q).$$

Démonstration. On pose pour tout $q \in]0;1[$,

$$A_q := \{x \in \mathbb{R} : F(x) \geq q\}$$

et $F^{(-1)}(q) := \inf A_q$. On montre :

Lemme 1.1. *Pour tout $q \in]0;1[$,*

$$F(F^{(-1)}(q)) \geq q.$$

Démonstration. On considère une suite $(u_n)_n$ telle que pour tout n, $u_n \in A_q$ et $(u_n)_n$ tend en décroissant vers $F^{(-1)}(q)$. Par exemple, on peut poser pour tout n, $u_n = F^{(-1)}(q)+n^{-1}$. On a donc que pour tout n, $F(u_n) \geq q$. Et par continuité de F à droite, $F(F^{(-1)}(q)) \geq q$. □

On démontre à présent les 3 points de la proposition.

1. Lorsque F est bijective de $\mathbb{R}$ dans $]0;1[$ alors $\forall\, q \in]0;1[$, $\exists\,!\, x \in \mathbb{R}$ tel que $F(x) = q$. Alors, pour tout $y > x$, $F(y) > F(x) = q$ et pour tout $z < x$, $F(z) < F(x) = q$. Donc
$$x = F^{-1}(q) = F^{(-1)}(q).$$
2. Pour tous $0 < q_1 \leq q_2 < 1$, on a $A_{q_2} \subset A_{q_1}$ et $\inf A_{q_1} \leq \inf A_{q_2}$. Cela implique que $F^{(-1)}(q_1) \leq F^{(-1)}(q_2)$.
3. On suppose d'abord $F(x) \geq q$. Donc $x \in A_q$ et $x \geq \inf A_q = F^{(-1)}(q)$.
On suppose $x \geq F^{(-1)}(q)$ et donc $F(x) \geq F(F^{(-1)}(q))$. On conclut en utilisant le Lemme 1.1.

□

Exercice 1.2. *Soit $q \in]0;1[$. On note*

$$I_q = \{x \in \mathbb{R} : \ F(x) = q\}.$$

Déterminer (par exemple à l'aide d'un graphique) $F^{(-1)}(q)$ dans les 3 cas suivants : $I_q = \emptyset$, I_q est réduit à un singleton et I_q est un intervalle.

Pour la construction de tests et de régions de confiance, on s'appuie sur la notion de quantile d'une loi.

Définition 1.6. *Soit F une fonction de répartition. Pour $\beta \in]0;1[$, on appelle **quantile d'ordre β de (la loi associée à) F** la quantité*

$$q_\beta := F^{(-1)}(\beta) = \inf \{x \in \mathbb{R} : \ F(x) \geq \beta\}.$$

*La **médiane de F** est $q_{\frac{1}{2}}$, le quantile d'ordre $\frac{1}{2}$.*
*Les **quartiles de F** désignent $q_{\frac{1}{4}}$, $q_{\frac{1}{2}}$ et $q_{\frac{3}{4}}$.*

Exemple 1.12. *Pour le calcul de la médiane, on a par exemple les résultats suivants.*
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$*, alors* $q_{\frac{1}{2}} = \mu$.
- Si $X \sim Ber(p)$*, alors* $q_{\frac{1}{2}} = 0$ *si* $p \leq \frac{1}{2}$ *et* $q_{\frac{1}{2}} = 1$ *si* $p > \frac{1}{2}$.

Proposition 1.5. *Pour* $\beta \in]0;1[$ *et* X *de fonction de répartition* F*,*

$$\mathbb{P}(X \geq q_\beta) \geq 1 - \beta \quad et \quad \mathbb{P}(X \leq q_\beta) \geq \beta.$$

Démonstration. Soit $(u_n)_n$ une suite strictement croissante convergeant vers q_β (on peut prendre par exemple $u_n = q_\beta - n^{-1}$). On note

$$C_{u_n} := \{w : \ X(w) \leq u_n\}, \quad C^-_{q_\beta} := \{w : \ X(w) < q_\beta\}.$$

Pour tout n, $u_n < q_\beta$ donc $C_{u_n} \subset C^-_{q_\beta}$ et

$$\bigcup_n C_{u_n} \subset C^-_{q_\beta}.$$

Soit $w \in C^-_{q_\beta}$ donc $X(w) < q_\beta$. Par définition de $(u_n)_n$, il existe n tel que $X(w) \leq u_n < q_\beta$ et $w \in C_{u_n}$. On a donc

$$\bigcup_n C_{u_n} = C^-_{q_\beta}.$$

Cela implique

$$\lim_{n \to +\infty} F(u_n) = \lim_{n \to +\infty} \mathbb{P}(w : \ X(w) \leq u_n) = \lim_{n \to +\infty} \mathbb{P}(C_{u_n}) = \mathbb{P}(C^-_{q_\beta}) = F(q^-_\beta).$$

A présent, comme $u_n < q_\beta$, $F(u_n) < \beta$. Donc, par passage à la limite, $F(q^-_\beta) \leq \beta$. Cette dernière inégalité est équivalente à $\mathbb{P}(X \geq q_\beta) \geq 1 - \beta$.
Pour le dernier point, en utilisant le Lemme 1.1, on obtient :

$$\mathbb{P}(X \leq q_\beta) = F(q_\beta) = F(F^{(-1)}(\beta)) \geq \beta.$$

□

Remarque 1.5. *En prenant* $\beta = \frac{1}{2}$*, on obtient pour la médiane :*

$$\mathbb{P}(X \geq q_{\frac{1}{2}}) \geq \frac{1}{2} \quad et \quad \mathbb{P}(X \leq q_{\frac{1}{2}}) \geq \frac{1}{2}.$$

Proposition 1.6. *Si* X *admet un moment d'ordre 1 alors pour tout* $a \in \mathbb{R}$ *vérifiant*

$$\mathbb{P}(X \geq a) \geq \frac{1}{2} \quad et \quad \mathbb{P}(X \leq a) \geq \frac{1}{2}$$

on a

$$\mathbb{E}[|X - a|] = \min_{c \in \mathbb{R}} \mathbb{E}[|X - c|].$$

En particulier,

$$\mathbb{E}[|X - q_{\frac{1}{2}}|] = \min_{c \in \mathbb{R}} \mathbb{E}[|X - c|].$$

Démonstration. Soit $c \in \mathbb{R}$. On suppose d'abord $c > a$ et on montre que

$$\mathbb{E}[|X - c|] \geq \mathbb{E}[|X - a|]. \tag{1.2}$$

On a $(a + c)/2 \in]a; c[$ et

$$\begin{array}{lll} |X - c| = & |X - a| + (c - a) & \text{sur } \{X \leq a\} \\ |X - c| \geq & |X - a| & \text{sur } \{a < X < (a + c)/2\} \\ |X - c| \geq & |X - a| - (c - a) & \text{sur } \{X \geq (a + c)/2\}, \end{array}$$

où la dernière inégalité est une simple conséquence de l'inégalité triangulaire. Donc

$$|X - c| \geq |X - a| + (c - a)1_{\{X \leq a\}} - (c - a)1_{\{X \geq (a+c)/2\}}$$

Donc,

$$\mathbb{E}[|X - c|] \geq \mathbb{E}[|X - a|] + (c - a)\big(\mathbb{P}(X \leq a) - \mathbb{P}(X \geq (a + c)/2)\big).$$

Mais

$$\mathbb{P}(X \geq (a + c)/2) \leq \mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a),$$

ce qui implique que

$$\mathbb{E}[|X - c|] \geq \mathbb{E}[|X - a|] + (c - a)\big(2\mathbb{P}(X \leq a) - 1\big) \geq \mathbb{E}[|X - a|]$$

et donc que (1.2) est vraie.
Le cas $c < a$ se traite de la même manière mais en utilisant que $\mathbb{P}(X \geq a) \geq \frac{1}{2}$. □

Chapitre 2

Concepts fondamentaux de la statistique

Dans ce chapitre, on décrit le cadre mathématique pour entreprendre la démarche statistique.

2.1 Exemples et problématiques

Exemple 1 : Un projet de loi est soumis à référendum. A cette occasion, on interroge n individus dans la population française. On note $X_i = 1$ lorsque le i–ème individu répond ≪ *oui* ≫ et $X_i = 0$ lorsqu'il répond ≪ *non* ≫. On suppose que les sondés sont choisis au *hasard* (par exemple, par tirage au sort de numéros téléphoniques), de sorte que l'on peut modéliser les variables $X_1, \ldots, X_n$ comme étant indépendantes et identiquement distribuées, de loi commune la loi de Bernoulli de paramètre $\theta \in]0;1[$ inconnu. Ce paramètre θ est la proportion de Français qui voteraient ≪ *oui* ≫ si l'élection se déroulait le jour où le sondage est réalisé. Au vu des réalisations des variables aléatoires X_i, on cherche à déterminer la valeur de θ et on désire savoir si le projet de loi va être adopté. C'est-à-dire que l'on souhaite d'une part *estimer* θ et d'autre part *tester* si θ est supérieur à $1/2$ ou non.

Exemple 2 : Une entreprise de téléphonie pense que le nombre journalier d'appels téléphoniques que passe chacun de ses clients suit une loi de Poisson. Cette hypothèse (fondée par exemple sur l'approximation de la loi binomiale par la loi de Poisson) lui permettant de fixer ses tarifs, elle souhaite la *tester* à l'aide du décompte (sur une journée) du nombre d'appels téléphoniques passés par n clients choisis au hasard dans un grand fichier de clients. Si elle accepte cette hypothèse, elle souhaitera alors également *estimer* le paramètre θ de la loi. On suppose donc disposer de l'observation de n variables aléatoires $X_1, \ldots, X_n$ indépendantes et identiquement distribuées de loi inconnue et on se demande si cette loi est une loi de Poisson.

L'objet de la *statistique inférentielle* est de répondre aux problèmes décrits par ces exemples. Comme dans la théorie des probabilités, le *hasard* intervient fortement. Mais dans la théorie des probabilités, on suppose la loi connue précisément et on cherche à caractériser le comportement d'une variable aléatoire qui suit cette loi. L'objectif de la statistique est le contraire : à partir de la connaissance de la variable, que peut-on dire de la loi de cette variable ?

Par ailleurs, pour chacun de ces exemples, nous avons introduit *un modèle statistique* (cette notion sera définie précisément dans le paragraphe suivant). Bien entendu, un modèle est toujours faux. Ce n'est qu'une approximation simple de la réalité, que l'on espère pas trop mauvaise. Le choix d'un modèle repose sur une connaissance partielle préalable du phénomène étudié, de la façon dont une expérience a été menée et/ou sur des représentations graphiques des données recueillies.

2.2 Modèle statistique

La notion de modèle statistique fournit le cadre mathématique nécessaire pour la présentation rigoureuse des problèmes statistiques décrits dans le paragraphe précédent et pour leur résolution.

Définition 2.1. *Un* ***modèle statistique*** *est la donnée d'un espace Ω mesuré par une tribu $\mathcal{B}$ et une famille $(\mathbb{P}_\theta)_{\theta\in\Theta}$ de lois de probabilité. On note $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in \Theta)$ le modèle associé. Quand il existe $p \in \mathbb{N}^*$ tel que $\Theta \subset \mathbb{R}^p$, le modèle est dit* ***paramétrique****. Sinon, il est dit* ***non-paramétrique****.*

Remarque 2.1. *Le paramètre θ est inconnu mais l'ensemble Θ des paramètres possibles est connu.*

Remarque 2.2. *Très souvent, on prend pour Ω l'espace $\mathbb{R}^n$ pour un certain $n \in \mathbb{N}^*$. Dans ce cas, $\mathcal{B}$ est $\mathcal{B}(\mathbb{R}^n)$, la tribu des boréliens sur $\mathbb{R}^n$.*

La démarche statistique consiste à donner de l'information sur le paramètre θ (on parle d'inférence et donc de **statistique inférentielle**) en s'appuyant sur la notion d'observation.

Définition 2.2. *Une* ***observation*** *X est une variable aléatoire à valeurs dans Ω et dont la loi appartient à la famille $(\mathbb{P}_\theta)_{\theta\in\Theta}$.*

Remarque 2.3. *L'espace sur lequel X est défini n'a pas d'intérêt statistique et on peut le confondre avec l'espace d'arrivée Ω (on en a le droit, quitte à considérer l'identification canonique, ce qui revient à prendre pour X l'identité). Si la loi de X est $\mathbb{P}_\theta$, on écrira $\mathbb{P}_\theta(X \in B)$ pour tout $B \in \mathcal{B}$.*

Dans la suite de ce cours, par abus de langage, le terme « *observation* » se rapportera soit à la variable aléatoire, soit à sa réalisation. Pour être rigoureux, on précise que par **réalisation de** X, on entend la valeur $X(\omega)$ que prend la variable X en un certain ω de l'ensemble sur lequel X est défini.

La plupart du temps, nos observations auront la structure d'un n–échantillon.

Définition 2.3. *Pour $n \in \mathbb{N}^*$, un n–**échantillon de loi** ν est la donnée de n variables $X_1, \dots, X_n$ indépendantes et identiquement distribuées selon la loi ν.*

Remarque 2.4. *Lorsque notre observation X a la structure d'un n–échantillon de loi ν_θ, alors $\mathbb{P}_\theta = (\nu_\theta)^{\otimes n}$.*

La réalisation de l'observation contient la totalité de l'information disponible sur la loi $\mathbb{P}_\theta$. C'est donc uniquement à partir de la valeur de cette réalisation que le statisticien peut accéder au paramètre θ inconnu.

Pour **l'Exemple 1**, les réalisations sont dans $\Omega = \{0,1\}^n$, la tribu $\mathcal{B}$ est l'ensemble des parties de Ω, l'ensemble des paramètres est donné par $\Theta =]0;1[$ et la loi de l'observation est celle d'un n–échantillon de loi de Bernoulli de paramètre θ : $\mathbb{P}_\theta = Ber(\theta)^{\otimes n}$.

Pour **l'Exemple 2**, les réalisations sont dans $\Omega = \mathbb{N}^n$, la tribu $\mathcal{B}$ associée est l'ensemble des parties de Ω, l'ensemble des paramètres vaut

$$\Theta = \left\{ \theta = (p_0, p_1, \dots) : \quad p_i \geq 0, \ \sum_{i=0}^{+\infty} p_i = 1 \right\}$$

et la loi de l'observation est $\mathbb{P}_\theta = (\nu_\theta)^{\otimes n}$, où ν_θ est la loi sur $\mathbb{N}$ donnée par la suite θ.

Si le modèle est paramétré par θ, on peut inférer θ à condition que $\theta \mapsto \mathbb{P}_\theta$ soit injective. On dit alors que le modèle est **identifiable**.

Dans toute la suite de ce cours, on suppose donné un modèle statistique identifiable, que l'on note $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in \Theta)$.

Chapitre 3

Estimation ponctuelle

On étudie à présent la problématique de l'estimation. Pour cela, on considère fixé un modèle statistique et donnée une observation X issue de ce modèle statistique et à valeurs dans $(\Omega, \mathcal{B})$. Notre objectif est d'estimer une quantité dépendant d'un paramètre inconnu θ, où θ appartient à un ensemble connu Θ et où $\mathbb{P}_\theta$ est la loi de l'observation X. Quand X aura la structure d'un n–échantillon, $X = (X_1, \ldots, X_n)$, on notera plutôt $\mathbb{P}_\theta$ la loi commune des variables X_i. On notera $\mathbb{E}_\theta$ et var_θ respectivement l'espérance et la variance associées à la loi $\mathbb{P}_\theta$. La quantité à estimer est notée $g(\theta)$, où g est une fonction de Θ dans $\mathbb{R}^p$ (pour un certain p dans $\mathbb{N}^*$).

3.1 Estimateurs

Définition 3.1. *Un **estimateur** $\widehat{g}$ de $g(\theta)$ est toute application mesurable en l'observation X et indépendante de θ. Tout estimateur est donc de la forme $\widehat{g} = h(X)$ où h est une application mesurable $\Omega \to \mathbb{R}^p$.*

Concrètement, c'est la **réalisation** de la variable aléatoire $\widehat{g}$ qui fournit une estimation de $g(\theta)$; on l'appelle une **estimée** de $g(\theta)$.

La définition d'un estimateur est très générale. La condition de mesurabilité est indispensable pour donner un sens mathématique à la variable aléatoire $\widehat{g}$. Par ailleurs, comme le paramètre θ est inconnu, h ne doit pas en dépendre. Hormis ces deux conditions, rien n'est imposé. Ainsi, on peut considérer des estimateurs très naïfs pour estimer $g(\theta)$, comme par exemple l'estimateur constant $\widehat{g} = 0_{\mathbb{R}^p}$. Ne s'appuyant pas sur les observations, cet estimateur s'avèrera très mauvais la plupart du temps.

Le but de ce chapitre est de décrire deux grandes classes de méthodes pour construire des estimateurs : la **méthode des moments** et la **méthode du maximum de vraisemblance.** Avant cela, on rappelle la notion de consistance qui est la première propriété qu'un estimateur est susceptible de posséder.

3.2 Consistance

La consistance traduit la proximité de l'estimateur (en un certain sens) à la quantité à estimer lorsque la taille de l'observation est très grande.

Définition 3.2. *On fixe une observation $X = (X_1, \ldots, X_n)$ et $\widehat{g}_n = h_n(X_1, \ldots, X_n)$ un estimateur de $g(\theta)$. Cet estimateur est dit*
- ***consistant** si pour tout $\theta \in \Theta$, $\widehat{g}_n$ converge en $\mathbb{P}_\theta$–probabilité vers $g(\theta)$ lorsque n tend vers $+\infty$,*
- ***fortement consistant** si pour tout $\theta \in \Theta$, $\widehat{g}_n$ converge $\mathbb{P}_\theta$–presque sûrement vers $g(\theta)$ lorsque n tend vers $+\infty$.*

Remarque 3.1. *Un estimateur fortement consistant est donc en particulier consistant.*

Remarque 3.2. *Par abus de langage, on parle de consistance d'un estimateur là où l'on devrait parler de la consistance d'une suite d'estimateurs.*

Remarque 3.3. *La consistance découle souvent de l'application de la loi des grands nombres.*

Dans l'Exemple 1 du Chapitre 2, notre but est d'estimer la proportion θ de la population en faveur du projet de loi, à l'aide de l'observation $X = (X_1, \ldots, X_n)$. Il est naturel de considérer l'estimateur de la moyenne empirique, qui est fortement consistant. Ce dernier sera dorénavant noté $\overline{X}_n$ et il est défini par

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

La notion de consistance conduit à une première méthode de construction d'estimateurs, la méthode des moments.

3.3 Méthode des moments

On suppose donné un n–échantillon $(X_1, \ldots, X_n)$ de loi $\mathbb{P}_\theta$ et une fonction ϕ mesurable à valeurs vectorielles telle que $\mathbb{E}_\theta[\|\phi(X_1)\|] < \infty$ pour tout $\theta \in \Theta$. Pour estimer $g(\theta)$ que l'on suppose écrite sous la forme

$$g(\theta) = \mathbb{E}_\theta\big[\phi(X_1)\big],$$

on s'appuie sur la convergence, pour tout $\theta \in \Theta$,

$$\widehat{g}_n = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i) \overset{n\to+\infty}{\longrightarrow} g(\theta) \qquad \mathbb{P}_\theta\text{–p.s.}$$

donnée par la loi forte des grands nombres. Elle indique que la suite des estimateurs $\widehat{g}_n$ est fortement consistante.

L'idée de la méthode des moments repose sur le fait que les paramètres d'intérêt s'expriment très fréquemment comme des fonctions des moments de la loi des observations X_i. Dans ce cas, la méthode dite des moments propose automatiquement des estimateurs fortement consistants pour l'estimation de ces paramètres. On estime par exemple l'espérance

$$\mu_1 = \mathbb{E}_\theta[X_1] \qquad \text{par} \qquad \widehat{\mu}_{1,n} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

et, si les variables X_i sont à valeurs réelles et lorsque la loi admet un moment d'ordre deux,

$$\mu_2 = \mathbb{E}_\theta\Big[X_1^2\Big] \qquad \text{par} \qquad \widehat{\mu}_{2,n} = \frac{1}{n}\sum_{i=1}^n X_i^2.$$

On peut alors estimer la variance $\sigma^2(\theta) = \mathrm{var}_\theta(X_1)$ par

$$\widehat{s}_n^2 = \widehat{\mu}_{2,n} - \big(\widehat{\mu}_{1,n}\big)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n \Big(X_i - \overline{X}_n\Big)^2. \tag{3.1}$$

La continuité de l'application $(x, y) \in \mathbb{R}^2 \mapsto y - x^2$ assure la forte consistance de $\widehat{s}_n^2$.

Un autre exemple souvent utile concerne l'estimation de la probabilité d'occurrence d'un événement $A \in \mathcal{B}$. La probabilité $g(\theta) = \mathbb{P}_\theta(X_1 \in A)$ est estimée selon la méthode des moments par

$$\widehat{\mu}_{A,n} = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \in A\}}.$$

Plus généralement, supposons données q fonctions mesurables à valeurs vectorielles $\phi_1, \ldots, \phi_q$ ayant chacune un moment $\mathbb{E}_\theta\big[\|\phi_j(X_1)\|\big]$ fini (pour tout $j \in \{1, \ldots, q\}$ et tout $\theta \in \Theta$) où on a noté $\|\cdot\|$ la norme euclidienne. La méthode des moments consiste à estimer

$$g(\theta) = \psi\big(g_1(\theta), \ldots, g_q(\theta)\big)$$

où $g_j(\theta) = \mathbb{E}_\theta\big[\phi_j(X_1)\big]$, pour $1 \leq j \leq q$, par

$$\widehat{g}_n = \psi\left(\frac{1}{n}\sum_{i=1}^n \phi_1(X_i),\ \ldots,\ \frac{1}{n}\sum_{i=1}^n \phi_q(X_i)\right).$$

Si par exemple ψ est continue, q applications de la loi forte des grands nombres assurent que les $\widehat{g}_n$ sont fortement consistants.

Exemple 3.1. *On se place dans le cadre de l'Exemple 2 du Chapitre 2 concernant les appels téléphoniques. On suppose que la distribution du nombre d'appels suit effectivement une loi de Poisson de paramètre $\theta \in \Theta := \mathbb{R}_+^*$ inconnu. On souhaite estimer θ à l'aide d'un n–échantillon $X_1, \ldots, X_n$. Puisque l'on a $\mathbb{E}_\theta[X_1] = \theta$, une première application de la méthode des moments suggère l'estimateur de la moyenne empirique $\overline{X}_n$. Mais comme $var_\theta(X_1) = \theta$, un autre estimateur de θ, toujours donné par la méthode des moments, est $\widehat{s}_n^2$ défini en* (3.1).

L'Exemple 3.1 montre que la méthode des moments permet parfois de déterminer différents estimateurs pour un même paramètre. Le choix de l'estimateur idoine parmi eux se fait alors au cas par cas. C'est la force et la faiblesse de la méthode : elle permet d'exhiber facilement des estimateurs simples, tellement facilement que souvent, plusieurs sont proposés. Par ailleurs, les estimateurs ainsi obtenus peuvent avoir de mauvaises performances.

Exemple 3.2. *On suppose que les variables X_i sont i.i.d. de loi $U]0;\theta]$, $\theta \in \Theta := \mathbb{R}_+^*$. On souhaite estimer θ. On a $\mathbb{E}_\theta[|X_1|] < \infty$ et $\mathbb{E}_\theta[X_1] = \frac{\theta}{2}$. En particulier $\theta = 2\mathbb{E}_\theta[X_1]$ et la méthode des moments suggèrent d'estimer θ par*

$$\widehat{\theta}_n := \frac{2}{n}\sum_{i=1}^{n} X_i \overset{n\to+\infty}{\longrightarrow} \theta \qquad \mathbb{P}_\theta\text{–}p.s.$$

De plus

$$\sqrt{n}(\widehat{\theta}_n - \theta) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}(0, 4var_\theta(X_1)) \qquad \textit{en loi sous } \mathbb{P}_\theta.$$

Mais en considérant

$$\tilde{\theta}_n := \max\{X_1, \ldots, X_n\},$$

on obtient

$$n(\theta - \tilde{\theta}_n) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{E}\big(\theta^{-1}\big) \qquad \textit{en loi sous } \mathbb{P}_\theta.$$

Ce second estimateur est donc plus rapide.

3.4 Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance ne s'applique que dans le cas où toutes les lois du modèle sont dominées par une mesure commune μ. Cette dernière est typiquement la mesure de Lebesgue sur $\mathbb{R}^m$, pour un certain $m \in \mathbb{N}^*$, ou la mesure de comptage sur un ensemble dénombrable. Dans les deux cas, pour tout θ on note f_θ la densité de $\mathbb{P}_\theta$ par rapport à μ. Dans le cas où μ est une mesure de comptage, on rappelle qu'on a alors $f_\theta(x) = \mathbb{P}_\theta(X = x)$ pour toutes les valeurs $x \in \Omega$.

La méthode du maximum de vraisemblance consiste à estimer $g(\theta) = \theta$ par le paramètre de Θ maximisant la **vraisemblance**

$$\begin{array}{rccl} V_X : & \Theta & \longrightarrow & \mathbb{R} \\ & \theta & \longmapsto & V_X(\theta) = f_\theta(X). \end{array}$$

On appelle ce point $\widehat{\theta}$ **l'estimateur du maximum de vraisemblance**. Dit autrement, la vraisemblance est la densité appliquée à l'observation.

Heuristiquement, la méthode consiste à dire qu'elle privilégie la valeur de θ la « plus probable » au vu de l'observation. L'estimateur associé possède en général de bonnes propriétés. Sous certaines conditions, le maximum de vraisemblance possède même des propriétés d'optimalité. L'inconvénient est que d'une part ce maximum peut ne pas exister ou ne pas être unique et que d'autre part, il peut être difficile à exhiber.

Remarque 3.4. *Quand on dispose d'un n–échantillon $(X_1, \ldots, X_n)$ de loi commune $\mathbb{P}_\theta$, la vraisemblance prend alors une forme produit,*

$$V_{X_1,\ldots,X_n}(\theta) = \prod_{i=1}^n f_\theta(X_i).$$

Cela découle de l'indépendance des X_i. Plutôt que de maximiser la vraisemblance directement, on préfère alors maximiser son logarithme (la log–vraisemblance). Cela conduit à travailler avec une somme plutôt qu'avec un produit, ce qui donne en général des calculs plus simples.

Exemple 3.3. *On poursuit l'étude de l'Exemple 3.1. La log–vraisemblance vaut*

$$\log V_{X_1,\ldots,X_n}(\theta) = -n\theta + \sum_{i=1}^n X_i \log\theta + \sum_{i=1}^n \log\frac{1}{X_i!}.$$

Elle admet un unique maximum, en $\widehat{\theta} = \overline{X}_n$, comme le montre un calcul simple (par étude de fonction, par exemple : recherche du point où la dérivée s'annule, calcul du signe de la dérivée seconde).

Proposition 3.1. *L'estimateur du maximum de vraisemblance ne dépend pas du choix de la mesure dominante μ dans le calcul de la vraisemblance.*

Démonstration. Soit ν une autre mesure dominante. Les mesures μ et ν sont elles-mêmes dominées par $\mu + \nu$. Donc, pour toute fonction test ϕ, on a :

$$\begin{aligned} \int \phi(x)\mathbb{P}_\theta(dx) &= \int \phi(x)\frac{d\mathbb{P}_\theta}{d(\mu+\nu)}(x)(\mu+\nu)(dx) \\ &= \int \phi(x)\frac{d\mathbb{P}_\theta}{d\mu}(x)\frac{d\mu}{d(\mu+\nu)}(x)(\mu+\nu)(dx) \\ &= \int \phi(x)\frac{d\mathbb{P}_\theta}{d\nu}(x)\frac{d\nu}{d(\mu+\nu)}(x)(\mu+\nu)(dx). \end{aligned}$$

Les densités $\frac{d\mathbb{P}_\theta}{d\mu}(x)$ et $\frac{d\mathbb{P}_\theta}{d\nu}(x)$ ne diffèrent que d'un facteur multiplicatif qui ne dépend pas de θ (sauf éventuellement sur un ensemble $(\mu+\nu)$–négligeable). Donc, presque sûrement,

$$\prod_{i=1}^n \frac{d\mathbb{P}_\theta}{d\mu}(X_i) \quad \text{et} \quad \prod_{i=1}^n \frac{d\mathbb{P}_\theta}{d\nu}(X_i)$$

ne diffèrent que d'une fonction de $X_1, \ldots, X_n$ qui ne dépend pas de θ. On ne modifie pas l'estimateur du maximum de vraisemblance selon que l'on maximise la vraisemblance sur l'une ou l'autre des mesures dominantes. □

Exercice 3.1. *On dispose d'un n–échantillon $(X_1, \ldots, X_n)$ de loi commune $\mathbb{P}_\theta$. Calculer le maximum de vraisemblance dans les 3 cas suivants :*

1. *La loi des X_i est uniforme sur l'intervalle $]0;\theta]$ avec $\theta \in \Theta := \mathbb{R}_+^*$.*
2. *La loi des X_i est la loi de Bernoulli de paramètre $\theta \in \Theta :=]0;1[$.*
3. *Les X_i admettent pour densité (par rapport à la mesure de Lebesgue sur $\mathbb{R}$) la fonction $x \longmapsto f_\theta(x) := f(x-\theta)$ avec $\theta \in \Theta := \mathbb{R}$ et $\forall x \in \mathbb{R}$,*

$$f(x) = \frac{e^{-\frac{|x|}{2}}}{2\sqrt{2\pi|x|}}.$$

3.5 Normalité asymptotique

La normalité asymptotique est une propriété plus forte que la consistance. En effet, on ne s'intéresse plus seulement à la convergence d'une suite d'estimateurs $\widehat{g}_n$ vers une quantité $g(\theta)$, mais à la vitesse de la convergence de $\widehat{g}_n - g(\theta)$ vers 0. Le théorème de la limite centrale joue ici le même rôle-clé que la loi des grands nombres pour établir la consistance d'estimateurs. Sa version vectorielle part d'un n–échantillon $(X_1, \ldots, X_n)$, de loi commune admettant un moment d'ordre deux, d'espérance μ et de matrice de variance–covariance Σ, et assure la convergence en loi sous $\mathbb{P}_\theta$:

$$\sqrt{n}\Big(\overline{X}_n - \mu\Big) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}(0, \Sigma).$$

Exemple 3.4. *Dans l'Exemple 3.1, on avait préconisé l'utilisation de l'estimateur de la moyenne empirique $\overline{X}_n$, dont on sait déjà qu'il est (fortement) consistant. Le théorème de la limite centrale conduit à la convergence en loi sous $\mathbb{P}_\theta$:*

$$\sqrt{n}\Big(\overline{X}_n - \theta\Big) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}(0, \theta).$$

En général, on observe que

1. la vitesse est $\sqrt{n}$,

2. la convergence a lieu en loi,
3. la loi limite est normale.

Lorsque ces trois points sont satisfaits, on dit que l'estimateur étudié est **asymptotiquement normal (et de vitesse $\sqrt{n}$)**. Nous redonnons l'énoncé du **lemme de Slutzky** qui montre qu'un estimateur asymptotiquement normal est consistant.

Lemme 3.1. *On considère $(U_n)_n$ et $(V_n)_n$ deux suites de vecteurs aléatoires. Si $(U_n)_n$ converge en loi vers U, et si $(V_n)_n$ converge en probabilité vers une constante c, alors le couple $(U_n, V_n)_n$ converge en loi vers le couple (U, c).*

Remarque 3.5. *Un contre-exemple classique aux points 1 et 3 et donné par l'estimateur du maximum de vraisemblance pour l'Exemple 3.2.*

Pour la consistance, on utilisait le fait que les convergences en probabilité ou presque sûre sont stables par passage aux fonctions continues. Ici, pour tirer d'autres formules de normalité asymptotique à partir de briques élémentaires données par des applications du théorème de la limite centrale, on a recours à des fonctions de classe C^1 et à la **méthode delta**. Dans ce qui suit, on note indistinctement $\|\cdot\|$ la norme (euclidienne) dans $\mathbb{R}^m$ ou $\mathbb{R}^p$, pour m et p fixés dans $\mathbb{N}^*$.

Théorème 3.1 (Méthode delta). *On se donne une suite $(U_n)_{n\in\mathbb{N}}$ de vecteurs aléatoires de $\mathbb{R}^m$, une suite déterministe de réels $(a_n)_{n\in\mathbb{N}}$ et une application $\ell : \mathbb{R}^m \to \mathbb{R}^p$ telles que*

- *$a_n \to +\infty$,*
- *il existe $U \in \mathbb{R}^m$, un vecteur déterministe, et un vecteur aléatoire V tels que l'on a la convergence en loi suivante :*

$$a_n(U_n - U) \overset{n\to+\infty}{\rightsquigarrow} V,$$

- *ℓ est une fonction différentiable en U, de différentielle notée $D\ell(U) \in M_{p,m}(\mathbb{R})$.*

Alors, on a la convergence en loi suivante :

$$a_n\big(\ell(U_n) - \ell(U)\big) \overset{n\to+\infty}{\rightsquigarrow} D\ell(U) \times V.$$

Démonstration. La preuve utilise le lemme suivant.

Lemme 3.2 (Lemme de Portmanteau). *Soient X une variable aléatoire réelle et $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles. Les cinq assertions suivantes sont équivalentes :*

1. *$(X_n)_{n\geq 1}$ converge en loi vers X lorsque n tend vers $+\infty$*
2. *Pour toute fonction φ bornée et uniformément continue sur $\mathbb{R}$,*

$$\lim_{n\to+\infty} \mathbb{E}\left[\varphi(X_n)\right] = \mathbb{E}\left[\varphi(X)\right]$$

3. Pour tout fermé F de $\mathbb{R}$,

$$\limsup_{n\to+\infty} \mathbb{P}\left(X_n \in F\right) \;\leq\; \mathbb{P}\left(X \in F\right)$$

4. Pour tout ouvert O de $\mathbb{R}$,

$$\liminf_{n\to+\infty} \mathbb{P}\left(X_n \in O\right) \;\geq\; \mathbb{P}\left(X \in O\right)$$

5. Pour tout borélien A de $\mathbb{R}$ tel que $\mathbb{P}\left(X \in \partial A\right) = 0$

$$\lim_{n\to+\infty} \mathbb{P}\left(X_n \in A\right) \;=\; \mathbb{P}\left(X \in A\right)$$

Par le lemme de Slutzky, il suffit de montrer que

$$a_n\big(\ell(U_n) - \ell(U)\big) - D\ell(U) \times a_n(U_n - U) \overset{n\to+\infty}{\longrightarrow} 0 \qquad \text{en proba.}$$

Or, fixant $\epsilon > 0$ et $M > 0$, et choisissant convenablement $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\bigg(\Big\| a_n\big(\ell(U_n) - \ell(U)\big) - D\ell(U) \times a_n(U_n - U)\Big\| \geq \epsilon\bigg) \\
&\leq \quad \mathbb{P}\Big(a_n\|\ell(U_n) - \ell(U) - D\ell(U) \times (U_n - U)\| \geq \epsilon, \quad \|U_n - U\| \leq \delta\Big) \\
&\quad + \mathbb{P}\big(\|U_n - U\| > \delta\big) \\
&\leq \;\mathbb{P}\big(a_n\|U_n - U\| \geq M\big) + \mathbb{P}\big(a_n\|U_n - U\| > a_n\delta\big) \\
&\leq \; 2\,\mathbb{P}\Big(a_n\|U_n - U\| \geq \min\big\{M,\, a_n\delta\big\}\Big)
\end{aligned}$$

où l'on a utilisé que par définition de la différentiabilité, il existe $\delta > 0$ tel que pour tout $U' \in \mathbb{R}^m$,

$$\|U' - U\| \leq \delta \quad\Longrightarrow\quad \|\ell(U') - \ell(U) - D\ell(U) \times (U' - U)\| \leq \frac{\varepsilon}{M}\|U' - U\|.$$

En particulier, puisque $a_n \to \infty$ et en utilisant le lemme de Portmanteau,

$$\begin{aligned}
&\limsup_{n\to\infty}\; \mathbb{P}\bigg(\Big\| a_n\big(\ell(U_n) - \ell(U)\big) - D\ell(U) \times a_n(U_n - U)\Big\| \geq \epsilon\bigg) \\
&\qquad\qquad \leq \limsup_{n\to\infty}\; 2\,\mathbb{P}\big(a_n\|U_n - U\| \geq M\big) \leq 2\;\mathbb{P}\big(\|V\| \geq M\big).
\end{aligned}$$

Ceci valant pour tout $M > 0$, on en déduit finalement que

$$\mathbb{P}\bigg(\Big\| a_n\big(\ell(U_n) - \ell(U)\big) - D\ell(U) \times a_n(U_n - U)\Big\| \geq \epsilon\bigg) \longrightarrow 0,$$

ce qu'il suffisait de montrer. $\square$

Corollaire 3.1. *On se donne* $(X_1, \ldots, X_n)$ *un* n*–échantillon de vecteurs de* $\mathbb{R}^m$*, de loi commune admettant un moment d'ordre deux, d'espérance notée* μ *et de matrice de variance–covariance* Σ*. Si* ℓ *est une fonction différentiable en* μ*, de différentielle notée* $D\ell(\mu)$*, alors*

$$\sqrt{n}\Big(\ell\big(\overline{X}_n\big) - \ell(\mu)\Big) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}\Big(0,\, D\ell(\mu)\,\Sigma\,D\ell(\mu)^T\Big) \qquad \textit{en loi sous } \mathbb{P}_\theta.$$

Exercice 3.2. *Pour l'Exemple 3.1, montrer que*

$$\sqrt{n}\Big(\widehat{s}_n^2 - \theta\Big) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}\big(0,\, \theta + 2\theta^2\big) \qquad \textit{en loi sous } \mathbb{P}_\theta,$$

en utilisant

$$var(X_1) = \theta, \quad cov(X_1, X_1^2) = 2\theta^2 + \theta, \quad var(X_1^2) = 4\theta^3 + 6\theta^2 + \theta.$$

On avait écrit à l'Exemple 3.4 la formule de normalité asymptotique pour le premier estimateur, $\overline{X}_n$*. De la comparaison des variances asymptotiques, on tire que l'estimateur* $\overline{X}_n$ *converge plus rapidement vers* θ *que* $\widehat{s}_n^2$*.*

3.6 Estimation sans biais

Nous avons introduit deux critères asymptotiques pour mesurer la performance d'un estimateur : la consistance et la normalité asymptotique. L'objet de ce paragraphe est de proposer un critère non-asymptotique. Par exemple, on peut exiger d'un estimateur qu'en moyenne, il soit proche de la quantité estimée.

Définition 3.3. *Soient* X *une observation et* $\widehat{g} = h(X)$ *un estimateur de* $g(\theta)$*, vérifiant* $\mathbb{E}_\theta\big[\|\widehat{g}\|\big] < \infty$ *pour tout* θ*. On appelle* ***biais*** *la fonction* $b : \Theta \to \mathbb{R}^p$ *définie pour tout* $\theta \in \Theta$ *par*

$$b(\theta) = \mathbb{E}_\theta\big[\widehat{g}\big] - g(\theta).$$

On dit alors que $\widehat{g}$ *est* ***sans biais*** *si la fonction* b *est identiquement nulle sur* Θ*, ce qui revient à dire que pour tout* $\theta \in \Theta$*, on a* $\mathbb{E}_\theta\big[\widehat{g}\big] = g(\theta)$*.*

Le caractère non-biaisé d'un estimateur se vérifie donc en calculant $b(\theta)$ pour *tout* $\theta \in \Theta$. Lorsque l'on travaille avec un n–échantillon, cette propriété est d'autant plus importante que n est petit (puisqu'alors les deux critères asymptotiques précédents ne peuvent pas nous guider dans le choix d'un estimateur).

Pour être exhaustifs, on mentionne la définition suivante.

Définition 3.4. *Soient* $X = (X_1, \ldots, X_n)$ *une observation et* $\widehat{g}_n = h_n(X_1, \ldots, X_n)$ *un estimateur de* $g(\theta)$*, vérifiant* $\mathbb{E}_\theta\big[\|\widehat{g}_n\|\big] < \infty$ *pour tout* $\theta \in \Theta$*. Il est dit* asymptotiquement sans biais *si pour tout* $\theta \in \Theta$*,*

$$\lim_{n\to+\infty} \mathbb{E}_\theta\big[\widehat{g}_n\big] = g(\theta).$$

Exemple 3.5. *Dans l'Exemple 3.1, le premier estimateur* $\overline{X}_n$ *est bien sans biais. Ce n'est cependant pas le cas du second estimateur, comme l'indique l'exemple suivant.*

Exemple 3.6. *L'estimateur* $\widehat{s}_n^2$ *de la variance introduit au paragraphe 3.3 est biaisé (mais asymptotiquement sans biais). On corrige cela en le multipliant par le facteur* $n/(n-1)$*. Avec les notations introduites pour (3.1), on considère donc plutôt en pratique, surtout lorsque* n *est petit,*

$$\widehat{\sigma}_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\Big(X_i - \overline{X}_n\Big)^2.$$

Ce chapitre se prolongerait naturellement par l'introduction de critères d'optimalité. Ils ont été largement abordés lors du premier semestre.

Chapitre 4

Intervalles et régions de confiance

Dans le chapitre précédent, nous avons proposé différentes méthodes pour construire un estimateur ponctuel (proposant une valeur) de la quantité $g(\theta)$ à l'aide de l'observation X. Nous avons montré comment mesurer sa proximité à $g(\theta)$. Ici, plutôt qu'une valeur, nous construisons un ensemble de valeurs possibles qui contient $g(\theta)$ avec une probabilité élevée et connue. C'est ce qu'on appelle une région de confiance.

4.1 Définitions et premières constructions

On reprend les mêmes notations que celles introduites dans les chapitres précédents. On estime $g(\theta)$ à l'aide de l'observation X.

Définition 4.1. *Pour tout $\alpha \in [0;1]$, une **région de confiance de $g(\theta)$ de niveau (de confiance)** $1-\alpha$ est un ensemble $\widehat{C}$ construit mesurablement par rapport à X et tel que pour tout $\theta \in \Theta$,*

$$\mathbb{P}_\theta\Big(\Big\{g(\theta) \in \widehat{C}\Big\}\Big) \geq 1-\alpha.$$

*Lorsque cette inégalité devient une égalité, on dit que **le niveau de confiance est exactement égal à** $1-\alpha$.*

Remarque 4.1. *Comme un estimateur, une région de confiance ne peut pas dépendre de θ, inconnu. Mais elle peut dépendre de α. A noter que la plupart du temps, plus α est petit, plus grande est la région de confiance.*

Dans la définition suivante, X est supposée de la forme $X = (X_1, \ldots, X_n)$.

Définition 4.2. *Pour tout $\alpha \in [0;1]$, une **région de confiance asymptotique de $g(\theta)$ de niveau (de confiance)** $1-\alpha$ est une suite d'ensembles $\widehat{C}_n$, chacun construit mesurablement sur $(X_1, \ldots, X_n)$ et tels que pour tout $\theta \in \Theta$,*

$$\liminf_{n\to+\infty} \mathbb{P}_\theta\Big(\Big\{g(\theta) \in \widehat{C}_n\Big\}\Big) \geq 1-\alpha.$$

Les valeurs usuelles (et quelque peu arbitraires) de α sont $1\,\%$, $5\,\%$ ou $10\,\%$. A niveau de confiance fixé, une région de confiance est d'autant meilleure que son aire est petite. Dans le cas unidimensionnel, ces régions s'écrivent la plupart du temps sous la forme d'un intervalle (dit **bilatère** ou **unilatère** selon que les deux bornes de l'intervalle sont finies ou non). On parle alors **d'intervalles de confiance**.

Si on considère un intervalle de confiance de niveau $95\,\%$, alors la probabilité que pour une réalisation de l'expérience considérée il contienne la vraie valeur inconnue $g(\theta)$ est d'au moins $95\,\%$. Ici, de la même manière que l'on a distingué estimateur et estimée, il faudrait distinguer l'estimation par intervalle ou région de confiance de l'estimée correspondante.

La construction de régions de confiance s'appuie sur les quantiles d'une loi tels qu'introduits dans la Définition 1.6. Rappelons que pour $\beta \in]0;1[$, le quantile d'ordre β de (la loi associée à) F est la quantité

$$q_\beta := F^{(-1)}(\beta) = \inf\{x \in \mathbb{R} : \ F(x) \geq \beta\}.$$

Par ailleurs, si F est bijective sur $\mathbb{R}$, $F^{(-1)} = F^{-1}$. Par exemple, le quantile d'ordre $97,5\,\%$ pour la loi $\mathcal{N}(0,1)$ (respectivement, $95\,\%$) vaut $1,960$ (respectivement, $1,645$). Ces deux valeurs sont à retenir, car elles sont fréquemment utilisées. On donne un premier exemple de construction.

Exemple 4.1. *On observe un n–échantillon $X = (X_1, \ldots, X_n)$ de loi $\mathcal{N}(\theta, 1)$. L'espérance $\theta \in \mathbb{R}$ de cet échantillon est inconnue mais la variance est connue égale à 1. L'estimateur (du maximum de vraisemblance) $\overline{X}_n$ vérifie*

$$\overline{X}_n \sim \mathcal{N}\left(\theta, \frac{1}{n}\right) \qquad \text{soit} \qquad \sqrt{n}\left(\overline{X}_n - \theta\right) \sim \mathcal{N}(0,1).$$

Pour l'estimation de θ, on propose alors typiquement un des trois intervalles de confiance suivants, tous de niveau exactement égal à $95\,\%$:

$$\left[\overline{X}_n - \frac{1,960}{\sqrt{n}},\ \overline{X}_n + \frac{1,960}{\sqrt{n}}\right], \qquad \left[\overline{X}_n - \frac{1,645}{\sqrt{n}},\ +\infty\right[, \qquad \left]-\infty,\ \overline{X}_n + \frac{1,645}{\sqrt{n}}\right];$$

le premier est bilatère, les deux suivants sont unilatères.

A travers cet exemple, on a illustré la **méthode du pivot** qui peut être résumée ainsi :

1. on choisit un estimateur,
2. on calcule sa loi en fonction de θ,
3. on transforme cet estimateur pour obtenir une variable aléatoire dont la loi (ou la loi asymptotique) ne dépend plus de θ,
4. on détermine les quantiles de la loi qui sont nécessaires pour construire une région de niveau de confiance adéquat.

Les deux prochains paragraphes proposent des méthodes de construction d'intervalles de confiance. Nous nous plaçons donc dans un cadre unidimensionnel.

4.2 Utilisation d'inégalités de probabilité pour l'obtention d'intervalles de confiance exacts

Ce paragraphe propose deux méthodes de construction d'intervalles de confiance à l'aide de deux inégalités de probabilité classiques, l'inégalité de Bienaymé–Tchebychev et l'inégalité de Hoeffding. Elles nécessitent respectivement une hypothèse de majoration uniforme de la variance des lois et une inclusion des supports des lois du modèle dans un compact donné.

4.2.1 Cas de variances uniformément bornées

On rappelle l'inégalité de Bienaymé-Tchebychev.

Proposition 4.1. *Soit une variable aléatoire réelle Y admettant un moment d'ordre 2 : pour tout $\varepsilon > 0$,*

$$\mathbb{P}\Big(\big|Y - \mathbb{E}[Y]\big| \geq \varepsilon\Big) \leq \frac{var(Y)}{\varepsilon^2}.$$

Lorsque les variances $\mathrm{var}_\theta(X)$ des observations X sont bornées indépendamment de θ, alors on peut en déduire un intervalle de confiance sur $g(\theta) = \mathbb{E}_\theta[X_1]$.

Exemple 4.2. *On observe $X = (X_1, \ldots, X_n)$ où les X_i sont i.i.d. de loi $Ber(\theta)$, pour $\theta \in]0;1[$. On cherche à estimer l'espérance θ de la loi de ce n–échantillon. La variance d'une loi de Bernoulli $Ber(\theta)$ est majorée uniformément par $\theta(1-\theta) \leq 1/4$. L'inégalité de Bienaymé–Tchebychev donne alors, pour la moyenne empirique $\overline{X}_n$ et pour tout $\varepsilon > 0$,*

$$\mathbb{P}_\theta\Big(\big|\,\overline{X}_n - \theta\big| \geq \varepsilon\Big) \leq \frac{1}{4n\varepsilon^2}.$$

Le choix de $\varepsilon = 1/\big(2\sqrt{n\alpha}\big)$ conduit à l'intervalle de confiance de niveau $1-\alpha$ suivant :

$$\left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}},\ \overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right].$$

Pour $\alpha = 5\,\%$ et $n = 100$, la précision de l'intervalle (sa demi-longueur) est environ de 0.23. On notera que la majoration obtenue ici n'est pas très précise lorsque la vraie valeur du paramètre θ est loin de $1/2$, auquel cas $\theta(1-\theta)$ est bien plus petit que $1/4$.

4.2.2 Cas de lois à supports tous inclus dans un compact donné

Ce paragraphe s'appuie sur le résultat suivant, l'inégalité de Hoeffding.

Lemme 4.1. *Soient* $(Y_1, \ldots, Y_n)$ *des variables aléatoires réelles indépendantes. On suppose l'existence de* $(a_1, \ldots, a_n)$ *et* $(b_1, \ldots, b_n)$ *deux* n*–uplets de réels tels que pour tout* i*, on a* $\mathbb{E}[Y_i] = 0$ *et* $a_i \leq Y_i \leq b_i$ *presque sûrement. Alors, pour tout* $t > 0$*,*

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

En particulier, en utilisant la symétrie de l'inégalité (i.e., en prenant les $-Y_i$ au lieu des Y_i), on obtient le corollaire suivant dont la démonstration est immédiate (la faire en guise d'exercice).

Corollaire 4.1. *Soit un* n*–échantillon* $(X_1, \ldots, X_n)$*. Sous* $\mathbb{P}_\theta$*, on suppose que la loi commune de chacun des* X_i *a son support inclus dans* $[a, b]$*. Alors, en notant* $\mu(\theta) = \mathbb{E}_\theta[X_1]$*, on a pour tout* $\varepsilon > 0$ *:*

$$\mathbb{P}_\theta\left(\left|\overline{X}_n - \mu(\theta)\right| \geq \varepsilon\right) \leq 2\, \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right).$$

Exemple 4.3. *Sur l'Exemple 4.2 on pose* $Y_i = X_i - \theta$*, d'où* $a_i = -\theta$ *et* $b_i = 1 - \theta$*, de sorte que le choix de*

$$\varepsilon = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}$$

conduit à l'intervalle de confiance de niveau $1 - \alpha$

$$\left[\overline{X}_n - \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)},\ \overline{X}_n + \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right].$$

Pour $\alpha = 5\,\%$ *et* $n = 100$*, la précision de l'intervalle est d'environ* 0.14*.*

De manière générale, lorsqu'une variable aléatoire Y est telle que $a \leq Y \leq b$ presque sûrement, alors $\mathrm{var}(Y) \leq (b-a)^2/4$. Cela procède de la caractérisation de la variance comme meilleure approximation au sens des moindres carrés (voir Proposition 1.3),

$$\mathrm{var}(Y) = \inf_{\mu \in \mathbb{R}}\ \mathbb{E}\left[(Y - \mu)^2\right],$$

et du choix (sous-optimal) $\mu = (b+a)/2$. Dans ce cas, comparer la précision des deux méthodes que nous avons introduites revient à comparer

$$\frac{1}{\sqrt{2\alpha}} \qquad \text{et} \qquad \sqrt{\log\left(\frac{2}{\alpha}\right)}.$$

Pour les petites valeurs de α (qui sont celles qui nous intéressent), c'est l'intervalle fourni par l'inégalité de Hoeffding qui fournit le meilleur résultat. Son application requiert cependant des hypothèses plus fortes que celles de l'inégalité de Bienaymé–Tchebychev.

Le paragraphe suivant montre que l'on peut encore améliorer la précision des intervalles au prix d'une plus grande incertitude sur leur niveau (en ne fixant ce dernier qu'à la limite en n).

4.3 Intervalles de confiance asymptotiques

Lorsque nous disposons d'un nombre important d'observations et qu'il s'agit d'estimer l'espérance $\mu(\theta)$ de la loi inconnue $\mathbb{P}_\theta$, il est naturel d'utiliser les résultats de normalité asymptotique fournis par l'application du théorème de la limite centrale. Les garanties sur l'intervalle obtenu ne sont alors qu'asymptotiques et elles ne valent en théorie que lorsque n est très grand. En pratique, on considère cependant leur utilisation acceptable dès que le nombre de données est supérieur à 30.

Partant d'un n-échantillon $(X_1, \ldots, X_n)$, le théorème de la limite centrale donne, pour l'estimation de l'espérance $\mu(\theta)$ et en faisant apparaître la variance $\sigma^2(\theta)$ de chacune des variables X_i, la convergence en loi suivante sous $\mathbb{P}_\theta$:

$$\sqrt{n}\Big(\overline{X}_n - \mu(\theta)\Big) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}\big(0,\, \sigma^2(\theta)\big).$$

De cette convergence on ne peut pas tirer directement d'intervalle de confiance. En effet, la variance asymptotique dépend de θ, de sorte que l'intervalle suivant n'est pas un intervalle de confiance asymptotique puisqu'il ne peut pas être calculé uniquement à partir des données :

$$\left[\overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\sigma^2(\theta)}{n}},\ \overline{X}_n + z_{1-\alpha/2}\sqrt{\frac{\sigma^2(\theta)}{n}}\right],$$

où l'on a noté $z_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. Nous allons voir essentiellement trois moyens de remédier à ce problème.

4.3.1 Cas de variances uniformément bornées

Lorsque les variances sont uniformément bornées, il suffit de réinjecter la majoration uniforme de ces variances dans l'intervalle proposé ci-dessus.

Exemple 4.4. *Toujours pour l'Exemple 4.2 et en utilisant que la variance est majorée par* $1/4$*, on obtient :*

$$\left[\overline{X}_n - \frac{z_{1-\alpha/2}}{2\sqrt{n}},\ \overline{X}_n + \frac{z_{1-\alpha/2}}{2\sqrt{n}}\right].$$

Pour $\alpha = 5\,\%$ *et* $n = 100$*, la précision de cet intervalle est environ de* 0.10.

4.3.2 Estimation consistante de la variance

Une alternative est fournie par le lemme de Slutzky (Lemme 3.1), en considérant une suite d'estimateurs $\left(\widehat{\sigma}_n^2\right)$ consistants de la variance $\sigma^2(\theta)$. On obtient pour tout θ la convergence en loi suivante sous $\mathbb{P}_\theta$:

$$\sqrt{\frac{n}{\widehat{\sigma}_n^2}}\left(\overline{X}_n - \mu(\theta)\right) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}(0,1).$$

On en tire alors l'intervalle de confiance asymptotique de niveau exactement $1-\alpha$

$$\left[\overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\widehat{\sigma}_n^2}{n}},\ \overline{X}_n + z_{1-\alpha/2}\sqrt{\frac{\widehat{\sigma}_n^2}{n}}\right].$$

C'est une illustration de la méthode du pivot.

Exemple 4.5. *Toujours, dans le cadre de l'Exemple 4.2, un estimateur consistant de la variance $\theta(1-\theta)$ est, en vertu de la loi des grands nombres,*

$$\widehat{\sigma}_n^2 = \overline{X}_n\left(1-\overline{X}_n\right).$$

4.3.3 Stabilisation de la variance par méthode delta

Enfin, une dernière méthode (qui illustre également la méthode du pivot) consiste à identifier une fonction φ continûment dérivable, admettant un inverse et telle que, pour tout θ,

$$\sigma^2(\theta)\times\left[\varphi'\left(\mu(\theta)\right)\right]^2 = 1.$$

La méthode delta implique la convergence en loi sous $\mathbb{P}_\theta$:

$$\sqrt{n}\left(\varphi\left(\overline{X}_n\right) - \varphi\left(\mu(\theta)\right)\right) \overset{n\to+\infty}{\rightsquigarrow} \mathcal{N}(0,1).$$

Si φ est par exemple strictement croissante, on tire l'intervalle de confiance asymptotique

$$\left[\varphi^{-1}\left(\varphi\left(\overline{X}_n\right) - \frac{z_{1-\alpha/2}}{\sqrt{n}}\right),\ \varphi^{-1}\left(\varphi\left(\overline{X}_n\right) + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)\right].$$

Exemple 4.6. *Dans le cas de l'Exemple 4.2, il suffit de considérer pour tout $x\in]0;1[$*

$$\varphi(x) = 2\arcsin\sqrt{x}.$$

Son inverse est définie sur $]0;\pi[$ par $\varphi^{-1}(u) = \sin^2(u/2)$ pour tout $u\in]0;\pi[$. On a également pour tout $x\in]0;1[$, $\varphi'(x) = \frac{1}{\sqrt{x(1-x)}}$. L'intervalle asymptotique proposé est alors

$$\left[\sin^2\left(\arcsin\sqrt{\overline{X}_n} - \frac{z_{1-\alpha/2}}{2\sqrt{n}}\right),\ \sin^2\left(\arcsin\sqrt{\overline{X}_n} + \frac{z_{1-\alpha/2}}{2\sqrt{n}}\right)\right].$$

Chapitre 5

Tests d'hypothèses - Généralités

Les chapitres précédents s'intéressaient à l'estimation, ponctuelle ou par intervalle, du paramètre θ ou d'une fonction $g(\theta)$ à l'aide de l'observation X. La problématique des tests est différente, au sens où ils se contentent de vérifier la compatibilité, ou plutôt, l'absence d'incompatibilité de X avec une ou des valeurs de référence θ_0 du paramètre. Nous verrons néanmoins que les outils de l'estimation jouent un rôle majeur pour la construction de tests et qu'il existe une dualité entre tests et régions de confiance.

5.1 Formalisme et démarche expérimentale

Dans le cadre de notre modèle statistique $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in \Theta)$, on se donne deux sous-ensembles Θ_0 et Θ_1 disjoints, inclus dans Θ (on n'impose pas que leur union soit Θ). Au vu d'une observation X de loi $\mathbb{P}_\theta$ inconnue du statisticien, on veut décider si $\theta \in \Theta_0$ ou pas ; dans ce dernier cas, on considère alors que $\theta \in \Theta_1$. On définit respectivement

- $H_0 : \theta \in \Theta_0$ comme **l'hypothèse nulle**,
- $H_1 : \theta \in \Theta_1$ comme **l'hypothèse alternative**.

Comme nous le verrons en détails par la suite, H_0 et H_1 ne jouent pas un rôle symétrique. En termes de vocabulaire, pour $j \in \{0, 1\}$, on dit que H_j est **simple** si Θ_j est réduit à un singleton, **composite** sinon.

Définition 5.1. *On appelle **test de l'hypothèse** H_0 **contre l'hypothèse** H_1 toute fonction $\phi(X)$ à valeurs dans $\{0, 1\}$, où ϕ est mesurable et peut dépendre de Θ_0 et Θ_1 :*

- *lorsque $\phi(X) = 0$, on conserve H_0,*
- *lorsque $\phi(X) = 1$, on rejette H_0 et on accepte H_1.*

Un test peut donc toujours s'écrire sous la forme $\phi(X) = 1_{\{X \in R\}}$ où $R \in \mathcal{B}$. Parfois, il est plus naturel de l'écrire sous la forme $\phi(X) = 1_{\{h(X) \in R'\}}$, où h est mesurable et $R' \in \mathcal{B}$. Ainsi, les ensembles $\{X \in R\} = \{h(X) \in R'\}$ et R (et par extension R' également même si cela constitue un abus de dénomination) sont appelés **région de rejet** du test. Leur complémentaire est parfois défini comme la **région d'acceptation**. Lorsqu'il existe une écriture naturelle de $\phi(X)$ en termes de $h(X)$ et R', on appelle $h(X)$ la **statistique de test**.

5.1.1 Mesure de la qualité d'un test

Lorsque l'on effectue un test, il y a deux manières de se tromper :
- rejeter H_0 alors que $\theta \in \Theta_0$,
- accepter H_0 alors que $\theta \in \Theta_1$.

On introduit donc naturellement les quantités suivantes.

Définition 5.2. *Soit $\phi(X) = 1_{\{X \in R\}} = 1_{\{h(X) \in R'\}}$ un test de l'hypothèse $H_0 : \theta \in \Theta_0$ contre l'hypothèse $H_1 : \theta \in \Theta_1$.*
*- **L'erreur de première espèce** correspond à la probabilité maximale de rejeter H_0 alors qu'elle est vraie. On la définit donc par :*

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\Phi(X)] = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\Phi(X) = 1) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(X \in R) \quad \left(= \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(h(X) \in R') \right).$$

*- **L'erreur de seconde espèce** correspond à la probabilité maximale d'accepter H_0 alors qu'elle est fausse. On la définit donc par :*

$$\sup_{\theta \in \Theta_1} \mathbb{E}_\theta[1 - \Phi(X)] = \sup_{\theta \in \Theta_1} \mathbb{P}_\theta(\Phi(X) = 0) = \sup_{\theta \in \Theta_1} \mathbb{P}_\theta(X \notin R) \quad \left(= \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(h(X) \notin R') \right).$$

L'erreur de première espèce mesure la probabilité (maximale) de rejeter H_0 à tort, et l'erreur de seconde espèce d'accepter H_0 à tort.

Définition 5.3. *La **fonction puissance du test** $\phi(X)$ est l'application*

$$\begin{array}{rccl} \Pi_\phi : & \Theta_1 & \longrightarrow & [0;1] \\ & \theta & \longmapsto & \mathbb{E}_\theta\big[\phi(X)\big]. \end{array}$$

On observe que l'erreur de seconde espèce correspond à $\sup_{\theta \in \Theta_1}\{1 - \Pi_\phi(\theta)\}$. Un test sera d'autant meilleur que sa fonction puissance prendra de grandes valeurs. La fonction puissance est utile notamment lorsque l'espace Θ_1 est grand.

Exemple 5.1. *On suppose données des observations $X_1, \ldots, X_n$ indépendantes et identiquement distribuées selon $\mathcal{N}(\theta, 1)$, pour un paramètre inconnu $\theta \in \Theta = \mathbb{R}$. On veut tester $H_0 : \theta \geq 1$ contre $H_1 : \theta < 1$, ce qui correspond à $\Theta_0 = [1, +\infty[$ et $\Theta_1 =]-\infty, 1[$. Un statisticien inexpérimenté pourrait penser à utiliser le test*

$$\phi(X) = 1_{\left\{\overline{X}_n < 1\right\}}, \qquad \text{où} \qquad \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,$$

fondé donc sur la statistique de test $\overline{X}_n$ et la région de rejet $]-\infty, 1[$. Il est aisé d'exprimer les erreurs α et β de première et seconde espèce en fonction de la fonction de répartition F de la loi normale standard $\mathcal{N}(0,1)$:

$$\alpha := \sup_{\theta \geq 1} \mathbb{E}_\theta[\phi(X)] = \sup_{\theta \geq 1} F\big(\sqrt{n}(1-\theta)\big) = F(0) = \frac{1}{2}.$$

$$\beta := \sup_{\theta < 1} \mathbb{E}_\theta[1-\phi(X)] = \sup_{\theta < 1} \big\{1 - F\big(\sqrt{n}(1-\theta)\big)\big\} = 1 - F(0) = \frac{1}{2}.$$

5.1.2 Dissymétrie des rôles des hypothèses H_0 et H_1

L'Exemple 5.1 n'est pas représentatif de la pratique statistique : dans la théorie des tests, H_0 et H_1 ne jouent absolument pas le même rôle et on n'attache pas autant d'attention au rejet à tort de H_1 qu'à celui de H_0.

Un test veut mesurer l'adéquation de l'hypothèse H_0 avec les observations et vérifie si le modèle associé à H_0 n'est pas en contradiction avec les données. Pour cela, le statisticien détermine les valeurs typiques de l'observation X sous H_0. S'il y a un désaccord grave, c'est-à-dire que la valeur réalisée de X n'est pas l'une de ces valeurs typiques, il rejette H_0 ; sinon, et peut-être faute de mieux, il conserve H_0.

Par ailleurs, les deux risques ne jouent pas le même rôle : celui de première espèce est privilégié et contrôlé.

Définition 5.4. *Soit $\alpha \in [0;1]$. Un* ***test est dit de niveau*** *α si*

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi(X)] \leq \alpha.$$

Un ***test est dit de taille*** *α si on a l'égalité dans le contrôle précédent :*

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi(X)] = \alpha.$$

Le niveau α peut être vu comme le risque maximal que l'on accepte de prendre en rejetant H_0 à tort. On fixe d'abord α, puis on cherche des tests de niveau α les plus puissants possibles.

Ces considérations nous aident à déterminer H_0 et H_1. On prendra ainsi pour H_0 :
- une hypothèse communément établie ou un parti pris subjectif,
- une hypothèse de prudence (motivée par un critère de coût ou de sécurité),
- la seule hypothèse facile à formuler.

L'hypothèse H_1 est souvent la négation de l'hypothèse H_0, mais pas nécessairement, un *a priori* peut conduire à un ensemble d'alternatives plus réduit.

En pratique, deux groupes avec des visées et intérêts différents auront souvent des couples (H_0, H_1) inversés. Ainsi, les industriels et les associations de consommateurs ne partent pas en général de la même hypothèse H_0.

5.1.3 Démarche de construction et mise en œuvre d'un test

On déduit des considérations précédentes la démarche de construction d'un test. Il sera question plus loin de la variante en termes de p–valeur.

Construction préliminaire

(1) Choix des hypothèses à tester H_0 et H_1

(2) Détermination de la statistique de test $h(X)$, dont on doit connaître le comportement sous H_0

(3) Allure de la zone de rejet en fonction de la forme de H_1 (c'est-à-dire du comportement de $h(X)$ sous H_1)

Suite classique

(4) Calcul de la zone de rejet R_α en fonction d'un niveau α déterminé par le statisticien (typiquement, de 5 %)

(5) Observation de la réalisation $h(x)$ de $h(X)$ et détermination de son appartenance, ou non, à la zone de rejet

Variante en termes de p–valeur

(4') Observation de la réalisation $h(x)$ de $h(X)$

(5') Calcul de la p–valeur associée $\widehat{\alpha}(x)$ et comparaison à un seuil α déterminé par un responsable

Conclusion

(6) Conservation ou rejet de H_0.

L'étude, même sommaire, de la puissance est nécessaire pour écarter les mauvais tests comme $\phi_\alpha \equiv 0$, qui contrôlent l'erreur de première espèce sans réellement prendre en compte H_1 et l'erreur de seconde espèce.

On reprend maintenant l'Exemple 5.1 dans l'optique de construire un test plus avisé.

Exemple 5.2. *La statistique de test naturelle est toujours $\overline{X}_n$ et, au vu de H_1, on a tendance à rejeter H_0 lorsque $\overline{X}_n$ est petit : la région de rejet est de la forme $]-\infty, \cdots[$. On fixe $\alpha \in]0;1[$ et on choisit donc le test $\phi_\alpha(X)$ de la forme*

$$\phi_\alpha(X) = 1_{\left\{\overline{X}_n < k_\alpha\right\}},$$

avec k_α, le seuil plus grand possible, de sorte que

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi_\alpha(X)] \leq \alpha.$$

Un raisonnement similaire à celui de l'Exemple 5.1 montre que ce seuil est tel que

$$\alpha = \sup_{\theta \geq 1} F\big(\sqrt{n}(k_\alpha - \theta)\big) = F\big(\sqrt{n}(k_\alpha - 1)\big), \tag{5.1}$$

soit $k_\alpha = 1 + z_\alpha/\sqrt{n}$*, où* z_α *désigne le quantile d'ordre* α *de la loi normale standard* $\mathcal{N}(0,1)$*. On peut déterminer l'erreur de seconde espace :*

$$\beta = \sup_{\theta<1} \mathbb{P}_\theta(\phi_\alpha(X) = 0) = \sup_{\theta<1} \mathbb{P}_\theta\Big(\sqrt{n}(\overline{X}_n - \theta) \geq \sqrt{n}(k_\alpha - \theta)\Big) = 1 - F\big(\sqrt{n}(k_\alpha - 1)\big) = 1 - \alpha.$$

La fonction puissance Π_{ϕ_α} *est définie sur* $]-\infty, 1[$ *et vaut pour tout* $\theta < 1$*,*

$$\Pi_{\phi_\alpha}(\theta) = \mathbb{E}_\theta[\phi_\alpha(X)] = F\big(\sqrt{n}(k_\alpha - \theta)\big) = F\big(z_\alpha + \sqrt{n}(1-\theta)\big).$$

On suppose que $n = 4$ *et que l'on observe les réalisations* $x_1 = 0,15$*,* $x_2 = 0,45$*,* $x_3 = 0,90$ *et* $x_4 = 0,50$*. Si l'on choisit un niveau* $\alpha = 10\,\%$*, alors* $z_\alpha = -1.28$ *et* $k_{10\,\%} \approx 0,36$ *tandis que* $\overline{x}_4 = 0,50$*, ce qui conduit à la conclusion (théorique) de conserver* H_0*. On le fait évidemment faute de mieux et à cause du niveau relativement élevé. La vraie conclusion (pratique) d'un statisticien est sans doute qu'on manque de données et qu'il faut en collecter davantage.*

Remarque 5.1. *Face à des échantillons de faible taille, il arrive ainsi couramment que* H_0 *soit acceptée face à* H_1 *et qu'il en soit de même pour* $H_0' = H_1$ *face à* $H_1' = H_0$*. Cela illustre encore une fois la tendance des tests à conserver l'hypothèse nulle.*

Exercice 5.1. *Reprendre l'Exemple 5.2 en inversant les rôles de* H_0 *et* H_1*.*

5.1.4 Première propriété d'un test

Comme un estimateur, un test peut posséder diverses propriétés. La première d'entre elles est d'être sans biais. Pour montrer l'utilité de cette dernière, on utilise les notations précédentes et on considère le test $\Phi(X) = 1_{\{Y=1\}}$ ou Y est indépendante de X et suit une loi de Bernoulli de paramètre α. Ce test a bien un niveau α, mais son erreur de seconde espèce est égale à $1-\alpha$ (et sa fonction puissance est constante, égale à α). Pour éliminer ce genre de tests, on pose :

Définition 5.5. *Un test* $\phi(X)$ *est dit* ***sans biais*** *lorsque sa fonction puissance vérifie*

$$\Pi_\phi(\theta) > \sup_{\theta' \in \Theta_0} \mathbb{E}_{\theta'}[\phi(X)], \qquad \forall\, \theta \in \Theta_1.$$

Remarque 5.2. *Comme pour les intervalles de confiance, la notion de test s'étend également au cas asymptotique : étant donnée une suite* $(\phi_n)_n$ *de tests, et notant* α_n^* *la taille de* ϕ_n *à tout rang fini* $n \geq 1$*, on dit que la suite de tests est de niveau asymptotique* α*, pour* $\alpha \in]0;1[$*, si*

$$\limsup_{n\to\infty} \alpha_n^* \leq \alpha.$$

On parle également de test asymptotique de niveau α*.*

5.2 p–valeur

L'approche précédente avait l'avantage de la simplicité mais conduisait à une information pauvre, parce que binaire : conservation ou rejet de H_0. Un point majeur en statistique mathématique est la quantification des assertions. A cet égard, la p–valeur permet de quantifier précisément la crédibilité de l'hypothèse H_0 au vu de l'observation X.

Définition 5.6. *Pour tout $\alpha \in]0;1[$, supposons avoir construit un test de niveau α, noté $\phi_\alpha(X)$ pour tester H_0 contre H_1. La p–**valeur** est le réel défini par*

$$\widehat{\alpha}(X) = \inf \big\{ \alpha \in]0;1[: \ \phi_\alpha(X) = 1 \big\}.$$

En termes plus informels, $\widehat{\alpha}(X)$ est le plus petit niveau pour lequel on rejette H_0. La p–valeur mesure l'adéquation entre l'hypothèse H_0 et les observations : plus $\widehat{\alpha}(X)$ est faible, plus nos observations contredisent H_0. **En pratique, on rejette H_0 lorsque la p–valeur est plus petite qu'un seuil α par exemple $5\,\%$.** L'intérêt de cette notion est qu'elle éclaire la décision sans la prendre : le statisticien, de manière objective, indique une p–valeur à un responsable politique ou économique qui, lui, doit prendre la décision de conserver ou rejeter H_0.

Remarque 5.3. *On notera les faits suivants.*

- *La p–valeur est une variable aléatoire.*
- *Une p–valeur importante n'implique pas forcément que H_0 soit vraie. Il se peut que H_0 soit fausse mais que le test ne soit pas puissant. Au risque de nous répéter, nous rappelons qu'un test dit simplement « peut-être » ou « non » à H_0, mais jamais « oui ».*
- *Une p–valeur petite est donc une bonne nouvelle, puisqu'elle permet un progrès négatif par le rejet de H_0.*

Ce calcul est particulièrement simple à mener dans le cas des régions de rejet formées d'un intervalle, comme le montre l'exemple suivant.

Exemple 5.3. *Reprenons l'Exemple 5.2 : la statistique de test est $\overline{X}_n$ et les zones de rejet sont de la forme $]-\infty, k_\alpha[$. On note $\overline{x}_n$ la valeur observée de $\overline{X}_n$. La p–valeur sur l'observation x de X est caractérisée par le cas limite*

$$k_{\hat{\alpha}(x)} = \overline{x}_n.$$

En effet $\alpha \longmapsto k_\alpha$ est une fonction croissante et $\phi_\alpha(x) = 1$ si et seulement si $\overline{x}_n < k_\alpha$, soit en utilisant l'égalité (5.1),

$$\widehat{\alpha}(x) = F\Big(\sqrt{n}\big(k_{\hat{\alpha}(x)} - 1\big)\Big) = F\Big(\sqrt{n}\big(\overline{x}_n - 1\big)\Big).$$

On observait sur les données x une valeur réalisée $\overline{x}_4 = 0,50$ pour la statistique de test, qui conduit à une p–valeur de $\widehat{\alpha}(x) = F(-1) = 15,9\,\%$. Soit α un seuil donné. Si $\alpha > 15,9\,\%$, *on rejette H_0 ; si $\alpha < 15,9\,\%$, on conserve H_0. En particulier, les calculs précédents donnent*

$$\widehat{\alpha}(X) = F\Big(\sqrt{n}\big(\overline{X}_n - 1\big)\Big).$$

Dans la plupart des cas, le calcul de la p–valeur est aisé, en utilisant la propriété suivante.

Proposition 5.1. *Si l'on dispose pour tout $\alpha \in]0;1[$ d'un test ϕ_α de taille α de la forme*

$$\phi_\alpha(X) = 1_{\left\{h(X)\geq k_\alpha\right\}},$$

avec $k_\alpha \in \mathbb{R}$, alors on a pour tout x :

$$\widehat{\alpha}(x) = \sup_{\theta\in\Theta_0} \mathbb{P}_\theta\big\{h(X) \geq h(x)\big\}\ .$$

Remarque 5.4. *Bien entendu, quitte à changer $h(X)$ en $-h(X)$ ou $|h(X)|$, une formule similaire s'obtient dans tous les cas où la zone de rejet est de la forme $]-\infty, k_\alpha[$ ou $]-\infty, -k_\alpha[\, \cup\,]k_\alpha, +\infty[$. Nous insistons sur le fait que ce calcul n'est valable que dans le cas de tests de taille α ; il ne suffit pas que leur niveau soit α.*

Exemple 5.4. *Reprenons l'Exemple 5.3 : L'Equation* (5.1) *montre que le test de de l'Exemple 5.2 est de taille α. A l'aide de la remarque précédente, on en déduit que*

$$\widehat{\alpha}(x) = \sup_{\theta\geq 1} \mathbb{P}_\theta\Big(\overline{X}_n < \overline{x}_n\Big) = \sup_{\theta\geq 1} \mathbb{P}_\theta\Big(\sqrt{n}(\overline{X}_n - \theta) < \sqrt{n}(\overline{x}_n - \theta)\Big) = F\Big(\sqrt{n}\big(\overline{x}_n - 1\big)\Big).$$

On retrouve bien l'expression déjà obtenue.

5.3 Botanique des tests

Comme il est impossible de faire l'inventaire de tous les tests existants, nous les classifions en quatre catégories principales.

Tests de conformité (ou d'ajustement à un paramètre). Ils partent d'un paramètre de référence θ_0 et testent, pour une certaine fonction g, l'hypothèse $H_0 : g(\theta) = g(\theta_0)$ contre l'alternative bilatère $H_1 : g(\theta) \neq g(\theta_0)$ ou une des deux alternatives unilatères $H_1 : g(\theta) > g(\theta_0)$ ou $H_1 : g(\theta) < g(\theta_0)$.

Exemple 5.5. *On suppose disposer d'un n–échantillon $X = (X_1, \ldots, X_n)$ de loi admettant un moment d'ordre deux et on veut tester que sa moyenne μ vaut $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$. On peut voir par le théorème de la limite centrale que le test fondé sur la statistique*

$$h_n(X) = \sqrt{n}\left(\frac{\overline{X}_n - \mu_0}{\sqrt{\widehat{\sigma}_n^2}}\right)$$

et la zone de rejet $R_\alpha =]-\infty, -z_{1-\alpha/2}[\, \cup\,]z_{1-\alpha/2}, +\infty[$ est un test asymptotique de niveau α.

Tests d'ajustement à une loi ou à une famille de lois. Le test porte ici sur la loi de X, dont on veut vérifier l'appartenance ou non à une certaine famille de lois.

Exemple 5.6. *On suppose disposer d'un n–échantillon $X = (X_1, \ldots, X_n)$ de loi ν. On peut vouloir tester $H_0 : \nu = \mathcal{N}(0,1)$ contre $H_1 : \nu \neq \mathcal{N}(0,1)$. On appliquera alors le test de Kolmogorov, dont il sera question au Chapitre 7. Ce test s'étend au cas du test dit de normalité, i.e., de H_0 : il existe μ et σ^2 tels que $\nu = \mathcal{N}(\mu, \sigma^2)$ contre H_1* ***fantôme*** *(la simple négation de H_0)*

Tests d'indépendance. Ils testent si deux observations X et Y sont de lois indépendantes (H_0) ou non (H_1). Le test d'indépendance du χ^2 sera vu au Chapitre 7. Il ne requiert qu'un éventuel regroupement préalable des observations.

Remarque 5.5. *Un test d'indépendance peut bien sûr être reformulé comme un test d'ajustement à l'ensemble des lois-produits.*

Tests d'homogénéité (forts et faibles). Ils testent si deux observations X et Y ont le même comportement (H_0) ou non (H_1). Lorsqu'il ne s'agit que d'un comportement moyen, c'est-à-dire que l'on teste si leurs espérances μ_X et μ_Y sont égales ($H_0 : \mu_X = \mu_Y$), on parle de test d'homogénéité faible. Sinon, si on veut réellement comparer leurs lois ν_X et ν_Y, de test d'homogénéité fort ($H_0 : \nu_X = \nu_Y$). On distingue tout d'abord le cas où les observations sont appariées, c'est-à-dire dépendantes et de même nombre parce que réalisées sur les mêmes objets. Dans ce cas, de $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, Y_n)$ on construit soit un test d'homogénéité faible en menant un test de conformité de l'espérance $\mu_X - \mu_Y$ de $Z = X - Y$ à $\mu_0 = 0$, soit un test d'homogénéité fort. Le second cadre est constitué d'observations indépendantes $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, Y_m)$, non nécessairement de même taille. Il survient lors de la réalisation d'une expérience sur deux groupes d'individus distincts et sans interaction.

Remarque 5.6. *Le test d'homogénéité peut être vu comme un test d'indépendance. En effet, si $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, Y_m)$ sont deux échantillons, pour tester l'homogénéité de X et Y il suffit de tester l'indépendance entre les deux composantes des couples $\big((X_1, 0), \ldots, (X_n, 0), (Y_1, 1), \ldots, (Y_m, 1)\big)$, une fois ceux-ci permutés dans un ordre aléatoire (voir Exercice 2 de l'Examen 2019-2020).*

5.4 Dualité entre tests et régions de confiance

On parle de dualité entre tests et régions de confiance parce qu'à partir de toute région de confiance, on peut construire un test d'ajustement à un paramètre de référence et réciproquement.

Si on dispose d'une région de confiance $\widehat{C}$ de niveau $1-\alpha$ pour l'estimation de $\theta \in \Theta$, alors, pour tout $\theta_0 \in \Theta$ le test

$$\phi(X) = 1_{\left\{\theta_0 \notin \widehat{C}\right\}}$$

est un test de niveau α pour le test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. En effet, avec $\Theta_0 = \{\theta_0\}$, on a :

$$\begin{aligned}\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\Phi(X)] = \mathbb{E}_{\theta_0}[\Phi(X)] &= 1 - \mathbb{P}_{\theta_0}(\theta_0 \in \widehat{C}) \\ &\leq 1 - (1-\alpha) = \alpha.\end{aligned}$$

Réciproquement, si pour tout $\theta_0 \in \Theta$, on dispose d'un test $\phi_{\alpha,\theta_0}(X)$ de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ de niveau α, alors la région d'acceptation

$$\widehat{C} = \left\{\theta \in \Theta : \ \phi_{\alpha,\theta}(X) = 0\right\}$$

est une région de confiance de niveau $1-\alpha$ pour l'estimation de θ. En effet, on a pour tout $\theta \in \Theta$,

$$\begin{aligned}\mathbb{P}_\theta(\theta \in \widehat{C}) &= 1 - \mathbb{P}_\theta\big(\left\{\phi_{\alpha,\theta}(X) = 1\right\}\big) \\ &= 1 - \mathbb{E}_\theta\left[\phi_{\alpha,\theta}(X)\right] \\ &\geq 1 - \alpha.\end{aligned}$$

Ce qui précède vaut évidemment pour des régions de confiance et des tests asymptotiques. Remarquons que le niveau de confiance $1-\alpha$ d'une région de confiance conduit à un test de niveau d'erreur α.

Dans le cas où l'estimation d'un paramètre n'est pas nécessaire pour construire un test de conformité, alors que cela l'est pour construire un intervalle de confiance, la dualité ci-dessus fonctionne mais ne conduit pas toujours au test ou à l'intervalle de confiance le plus naturel.

Exemple 5.7. *Ainsi pour α fixé, si l'observation $X = (X_1, \ldots, X_n)$ est un n–échantillon de loi de Bernoulli p, le test naturel (asymptotique, de niveau α) de $H_0 : p = p_0$ contre $H_1 : p \neq p_0$ est donné par la statistique de test*

$$h_n(X) = \sqrt{n}\left(\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)}}\right)$$

et la zone de rejet $R_\alpha =]-\infty, -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}, +\infty[$ (avec $z_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$). En revanche l'intervalle de confiance naturel (asymptotique, de niveau $1-\alpha$) est

$$\widehat{C}_n = \left[\overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n\big(1-\overline{X}_n\big)}, \ \ \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\overline{X}_n\big(1-\overline{X}_n\big)}\right] \ .$$

Ils ne s'induisent pas (strictement) l'un l'autre.

Exercice 5.2. *Démontrer les assertions de l'exemple précédent.*

Chapitre 6

Tests fondés sur la vraisemblance

Ce chapitre décrit les tests fondés sur la vraisemblance du modèle que l'on considère. Ils ont la propriété importante d'être non-asymptotiques. On les privilégie donc lorsque le nombre d'observations est peu important. On considère le même cadre et le même formalisme que le chapitre précédent. On note $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in \Theta)$ le modèle statistique et on se donne deux sous-ensembles Θ_0 et Θ_1 disjoints, inclus dans Θ. Au vu d'une observation X de loi $\mathbb{P}_\theta$ inconnue du statisticien, on veut décider si $\theta \in \Theta_0$ ou pas ; dans ce dernier cas, on considère alors que $\theta \in \Theta_1$. On définit respectivement

- $H_0 : \theta \in \Theta_0$ comme l'hypothèse nulle,
- $H_1 : \theta \in \Theta_1$ comme l'hypothèse alternative.

On notera μ une mesure dominante de la famille $(\mathbb{P}_\theta)_{\theta\in\Theta}$.

6.1 Tests UPP

La définition suivante constitue un critère de comparaison de deux tests.

Définition 6.1. *Pour $\phi(X)$ et $\phi'(X)$ deux tests de niveau $\alpha \in]0;1[$, on dit que $\phi(X)$ est* ***uniformément plus puissant (UPP)*** *que $\phi'(X)$ si leurs fonctions puissance respectives Π_ϕ et $\Pi_{\phi'}$ vérifient $\forall\, \theta \in \Theta_1$,*

$$\Pi_\phi(\theta) \geq \Pi_{\phi'}(\theta),$$

c'est-à-dire que $\forall\, \theta \in \Theta_1$,

$$\mathbb{P}_\theta(\Phi(X) = 1) \geq \mathbb{P}_\theta(\Phi'(X) = 1).$$

Par ailleurs, $\Phi(X)$ est dit ***UPP(α)*** *s'il est uniformément plus puissant que tout test de niveau α.*

Il n'existe pas toujours des tests UPP(α), mais nous allons étudier dans la suite les situations qui nous permettent d'en bâtir. Une méthode générale et intuitive pour construire des tests UPP(α) est de considérer des rapports de vraisemblance. A cet effet, on reprend le cadre et les notations du paragraphe 3.4. La statistique du test consiste en le rapport

$$h(X) = \frac{\sup_{\theta \in \Theta_1} V_X(\theta)}{\sup_{\theta \in \Theta_0} V_X(\theta)},$$

où l'on rappelle que $V_X(\theta) = f_\theta(X)$ est appelée la vraisemblance du modèle au point θ. Puisque sous H_1, le numérateur de $h(X)$ aura tendance à être grand et son dénominateur petit (et donc que $h(X)$ aura elle-même tendance à prendre de grandes valeurs), la zone de rejet est de la forme $]k_\alpha, +\infty[$ (où k_α est une constante à fixer dépendant du niveau α). Le test s'écrit donc comme

$$\phi(X) = 1_{\{h(X) > k_\alpha\}}.$$

Il est appelé **test de rapport de vraisemblance**.

Exemple 6.1. *Reprenons l'Exemple 5.1. Soit l'observation de $X = (X_1, \ldots, X_n)$ où les X_i sont i.i.d. selon $\mathcal{N}(\theta, 1)$, pour un paramètre inconnu $\theta \in \Theta = \mathbb{R}$. On veut tester $H_0 : \theta \geq 1$ contre $H_1 : \theta < 1$. Pour ce modèle, la vraisemblance au point $\theta \in \Theta$ s'écrit*

$$V_X(\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^n (X_i - \theta)^2\right).$$

Les variations de V_X sont dictées par la fonction dans l'exponentielle : V_X est strictement croissante sur l'intervalle $\left]-\infty, \overline{X}_n\right]$ et strictement décroissante sur l'intervalle $\left[\overline{X}_n, +\infty\right[$, de sorte que

$$h(X) = \begin{cases} V_X\left(\overline{X}_n\right) / V_X(1) & \text{si } \overline{X}_n < 1, \\ V_X(1) / V_X\left(\overline{X}_n\right) & \text{si } \overline{X}_n \geq 1. \end{cases}$$

Un calcul immédiat montrant que pour tous points θ_0 et θ_1 de Θ,

$$\frac{V_X(\theta_1)}{V_X(\theta_0)} = \exp\left(\frac{n}{2}\left(\theta_0^2 - \theta_1^2 + 2\overline{X}_n(\theta_1 - \theta_0)\right)\right), \tag{6.1}$$

la statistique de test s'exprime finalement comme

$$h(X) = \exp\left(\frac{n}{2}\left(\overline{X}_n - 1\right)^2 \left(2\,1_{\left\{\overline{X}_n < 1\right\}} - 1\right)\right).$$

C'est donc une fonction strictement décroissante de $\overline{X}_n$, de sorte que le test de rapport de vraisemblance peut se réécrire sous la forme

$$\phi(X) = 1_{\left\{\overline{X}_n < k_\alpha\right\}},$$

où k_α est une constante à fixer une fois le niveau α choisi. On retrouve ainsi le test introduit dans l'Exemple 5.2.

6.2 Tests d'hypothèses simples - Lemme de Neyman-Pearson

Le résultat suivant montre l'optimalité des tests de rapport de vraisemblance introduits précédemment dans le cadre du test de deux hypothèses simples.

Théorème 6.1 (lemme de Neyman–Pearson)**.** *On fixe θ_0 et θ_1 deux points distincts de Θ et on s'intéresse au test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. Pour $\alpha \in \,]0;1[$ donné, lorsqu'il existe un seuil k_α tel que le test de rapport de vraisemblance*

$$\phi(X) = 1_{\{h(X,\theta_0,\theta_1)>k_\alpha\}} \qquad \text{où} \qquad h(X,\theta_0,\theta_1) = \frac{V_X(\theta_1)}{V_X(\theta_0)} = \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}$$

vérifie

$$\mathbb{P}_{\theta_0}\big(\phi(X) = 1\big) = \alpha,$$

alors ce test est UPP(α).

Démonstration. On considère un autre test $\psi(X)$ de niveau α, i.e.,

$$\mathbb{E}_{\theta_0}\big[\psi(X)\big] = \mathbb{P}_{\theta_0}\big(\psi(X) = 1\big) \leq \alpha, \qquad \text{soit} \qquad \mathbb{E}_{\theta_0}\big[\phi(X) - \psi(X)\big] \geq 0.$$

Or, il s'agit de montrer que

$$\mathbb{E}_{\theta_1}\big[\phi(X) - \psi(X)\big] = \mathbb{P}_{\theta_1}\big(\phi(X) = 1\big) - \mathbb{P}_{\theta_1}\big(\psi(X) = 1\big) \geq 0.$$

Pour tout x, si $f_{\theta_0}(x) > 0$ on a $f_{\theta_1}(x) = \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)$ et donc

$$\begin{aligned}
&\mathbb{E}_{\theta_1}\big[\phi(X) - \psi(X)\big] - k_\alpha \, \mathbb{E}_{\theta_0}\big[\phi(X) - \psi(X)\big] \\
&= \int (\phi(x) - \psi(x))\Big(\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) 1_{\left\{f_{\theta_0}(x)>0\right\}} + f_{\theta_1}(x) 1_{\left\{f_{\theta_0}(x)=0\right\}}\Big) d\mu(x) \\
&\quad - k_\alpha \, \mathbb{E}_{\theta_0}\big[\phi(X) - \psi(X)\big] \\
&= \mathbb{E}_{\theta_0}\left[\big(\phi(X) - \psi(X)\big)\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right] + \mathbb{E}_{\theta_1}\left[\big(\phi(X) - \psi(X)\big) 1_{\left\{f_{\theta_0}(X)=0\right\}}\right] \\
&\quad - k_\alpha \, \mathbb{E}_{\theta_0}\big[\phi(X) - \psi(X)\big] \\
&= \mathbb{E}_{\theta_0}\left[\big(\phi(X) - \psi(X)\big)\left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} - k_\alpha\right)\right] + \mathbb{E}_{\theta_1}\left[\big(\phi(X) - \psi(X)\big) 1_{\left\{f_{\theta_0}(X)=0\right\}}\right].
\end{aligned}$$

Par construction de ϕ,

$$\phi(X) = 1 \iff \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} > k_\alpha.$$

La variable aléatoire dont on prend l'espérance pour former le premier terme est donc positive. Pour le second terme, on observe que

$$\mathbb{E}_{\theta_1}\Big[\big(\phi(X)-\psi(X)\big)1_{\big\{f_{\theta_0}(X)=0\big\}}\Big] = \mathbb{E}_{\theta_1}\Big[\big(\phi(X)-\psi(X)\big)1_{\big\{f_{\theta_0}(X)=0\big\}}1_{\big\{f_{\theta_1}(X)>0\big\}}\Big]$$

et que puisque $\alpha > 0$, alors k_α est fini, de sorte que $\phi(X) = 1$ si $f_{\theta_1}(X) > 0$ et $f_{\theta_0}(X) = 0$: il est donc également positif. Finalement, on a prouvé

$$\mathbb{E}_{\theta_1}\big[\phi(X)-\psi(X)\big] \geq k_\alpha\ \mathbb{E}_{\theta_0}\big[\phi(X)-\psi(X)\big] \geq 0,$$

ce qui montre que ϕ est plus puissant que ψ. □

Remarque 6.1. *L'hypothèse $\alpha < 1$ peut être omise. L'hypothèse $\alpha > 0$ est nécessaire seulement pour avoir k_α fini. On peut s'en passer en considérant le test*

$$\phi(X) = 1_{\big\{V_X(\theta_1)>k_\alpha V_X(\theta_0)\big\}},$$

où k_α peut alors être infini si on adopte la convention "$0 \times +\infty = 0$". A noter que les cas $\alpha = 0$ et $\alpha = 1$ n'ont pas d'intérêt pratique.

Remarque 6.2. *Pour l'application du Théorème 6.1, il est important d'avoir un test de taille α. Dans certaines situations, ce n'est pas possible. On contourne cette difficulté en considérant des tests randomisés (hors programme).*

Exemple 6.2. *On suppose donnée une observation X de loi de Poisson de paramètre $\theta \in \Theta := \mathbb{R}_+^*$. Pour $\theta_0 < \theta_1$, on veut tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. La densité de X par rapport à la mesure de comptage vaut*

$$f_\theta(x) = \exp(-\theta)\frac{\theta^x}{x!}, \quad x \in \mathbb{N}.$$

Le test de Neyman-Pearson s'écrit :

$$\phi(X) = 1_{\Big\{\exp(-(\theta_1-\theta_0))\big(\frac{\theta_1}{\theta_0}\big)^X > k_\alpha\Big\}} = 1_{\Big\{X > \frac{\log(k_\alpha)+\theta_1-\theta_0}{\log\theta_1-\log\theta_0}\Big\}}.$$

Le seuil k_α s'obtient en résolvant

$$\mathbb{P}_{\theta_0}\left(X > \frac{\log(k_\alpha)+\theta_1-\theta_0}{\log\theta_1-\log\theta_0}\right) = \alpha.$$

Comme la fonction de répartition de la loi de Poisson de paramètre θ_0 n'est pas continue, ce n'est pas forcément possible. Par exemple, avec $\theta_0 = 5$ et $\alpha = 5\%$, on a

$$\mathbb{P}_{\theta_0}(X > 9) = 0,032 \quad \mathbb{P}_{\theta_0}(X > 8) = 0,068.$$

En pratique, on considère $\phi(X) = 1_{\{X>9\}}$. Ce test est de niveau 5% mais on ne peut pas garantir qu'il est UPP(5%).

Exemple 6.3. *On suppose données des observations $X_1, \dots, X_n$ indépendantes et identiquement distribuées selon $\mathcal{N}(\theta, 1)$, pour un paramètre inconnu $\theta \in \Theta = \mathbb{R}$. Pour $\theta_0 < \theta_1$, on veut tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. D'après l'Equation (6.1), on a*

$$\frac{V_X(\theta_1)}{V_X(\theta_0)} = \exp\left(\frac{n}{2}\Big(\theta_0^2 - \theta_1^2 + 2\overline{X}_n(\theta_1 - \theta_0)\Big)\right).$$

On obtient que pour $k_\alpha \in \mathbb{R}_+^*$,

$$\frac{V_X(\theta_1)}{V_X(\theta_0)} > k_\alpha \iff \overline{X}_n > \frac{\theta_0 + \theta_1}{2} + \frac{\log(k_\alpha)}{n(\theta_1 - \theta_0)}$$

Avec $z_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi $\mathcal{N}(0,1)$, on obtient alors qu'avec

$$k_\alpha = \exp\Big(\sqrt{n}(\theta_1 - \theta_0)z_{1-\alpha} - \frac{n}{2}(\theta_1 - \theta_0)^2\Big),$$

le test de rapport de vraisemblance

$$\phi(X) = 1_{\{h(X,\theta_0,\theta_1) > k_\alpha\}} \qquad \text{où} \qquad h(X, \theta_0, \theta_1) = \frac{V_X(\theta_1)}{V_X(\theta_0)} = \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}$$

vérifie

$$\mathbb{P}_{\theta_0}\big(\phi(X) = 1\big) = \alpha.$$

Par le lemme de Neyman-Pearson, ce test est UPP(α).
Notons que la statistique du rapport de vraisemblance est une fonction croissante de $\overline{X}_n$ et $\phi(X)$ peut donc s'écrire

$$\phi(X) = 1_{\{\overline{X}_n > \tilde{k}_\alpha\}},$$

avec $\tilde{k}_\alpha$ de sorte que

$$\mathbb{P}_{\theta_0}\big(\overline{X}_n > \tilde{k}_\alpha\big) = \alpha.$$

On obtient $\tilde{k}_\alpha = \theta_0 + z_{1-\alpha}/\sqrt{n}$.

Proposition 6.1. *On reprend le cadre du Théorème 6.1. Si le test de rapport de vraisemblance du lemme de Neyman-Pearson est de taille α, il vérifie*

$$\mathbb{E}_{\theta_1}[\phi(X)] \geq \alpha.$$

Démonstration. Soit U une variable de loi uniforme sur $]0;1[$. On considère le test $\tilde{\phi}$ défini par

$$\tilde{\phi}(X) = 1_{\{U \leq \alpha\}}.$$

On a $\mathbb{E}_{\theta_0}[\tilde{\phi}(X)] = \alpha$ et donc $\tilde{\phi}(X)$ est de niveau α. Comme ϕ est UPP(α), on a donc

$$\mathbb{E}_{\theta_1}[\phi(X)] \geq \mathbb{E}_{\theta_1}[\tilde{\phi}(X)] = \alpha.$$

□

6.3 Tests d'hypothèses composites

On suppose dans ce paragraphe que Θ est un intervalle ouvert de $\mathbb{R}$. On suppose toujours que la famille $(\mathbb{P}_\theta)_{\theta\in\Theta}$ est dominée et on note μ une mesure dominante. On pose

$$f_\theta(x) := \frac{d\mathbb{P}_\theta}{d\mu}(x), \quad x \in \Omega.$$

On souhaite tester pour $\tilde{\theta} \in \Theta$, $H_0 : \theta \leq \tilde{\theta}$ contre $H_1 : \theta > \tilde{\theta}$. Pour pouvoir appliquer le lemme de Neyman-Pearson, il faudrait d'une certaine manière pouvoir traiter tous les tests de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ simultanément pour tous les $\theta_0 \leq \tilde{\theta}$ et $\theta_1 > \tilde{\theta}$.

Définition 6.2. *On suppose que $\forall\theta \in \Theta$,*

$$f_\theta(x) > 0 \ \mu(dx) - p.s.$$

*La famille $(f_\theta)_{\theta\in\Theta}$ est dite à **rapport de vraisemblance monotone** s'il existe une application $T : \Omega \longmapsto \mathbb{R}$ mesurable de sorte que pour tous $\theta_1 < \theta_2$, $\frac{f_{\theta_2}(X)}{f_{\theta_1}(X)}$ est une fonction monotone de $T(X)$, c'est à dire qu'il existe $\psi : \mathbb{R} \longmapsto \mathbb{R}_+$ une fonction monotone telle que*

$$\frac{V_X(\theta_2)}{V_X(\theta_1)} = \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} = \psi(T(X)).$$

Remarque 6.3. *Dans la définition précédente, quitte à changer T en $-T$, on peut toujours supposer que la fonction ψ est croissante.*

Pour l'Exemple 6.3, la famille de densités gaussiennes est à rapport de vraisemblance monotone avec $T(X) = \overline{X}_n$ et $\psi(x) = \exp\Big(\frac{n}{2}\Big(\theta_0^2 - \theta_1^2 + 2x(\theta_1 - \theta_0)\Big)\Big)$ qui est croissante puisque $\theta_1 > \theta_0$.

On a le résultat suivant que l'on admet. Néanmoins, la preuve dans le cas particulier de la loi exponentielle est proposée dans l'exercice 1 du sujet d'examen 2022-2023 (Appendice C.8). Elle se généralise aisément à toute distribution.

Théorème 6.2. *Soit $\alpha \in]0;1[$. On suppose que $\forall\theta \in \Theta$,*

$$f_\theta(x) > 0 \ \mu(dx) - p.s.$$

et que la famille $(f_\theta)_{\theta\in\Theta}$ est à rapport de vraisemblance monotone où la fonction ψ associée est strictement croissante. Si pour $\tilde{\theta} \in \Theta$, il existe $\rho(\tilde{\theta}, \alpha) \in \mathbb{R}$ tel que $\mathbb{P}_{\tilde{\theta}}(T(X) > \rho(\tilde{\theta}, \alpha)) = \alpha$ alors le test

$$\phi(X) = 1_{\{T(X) > \rho(\tilde{\theta}, \alpha)\}}$$

est de taille α pour tester $H_0 : \theta \leq \tilde{\theta}$ contre $H_1 : \theta > \tilde{\theta}$ et UPP(α).

Hormis quelques cas particuliers, comme les familles à rapport de vraisemblance monotone, on ne sait pas, en général, exhiber des tests optimaux pour hypothèses composites. Une manière de contourner cette difficulté est de considérer le cadre asymptotique.

Chapitre 7

Tests asymptotiques - Tests du χ^2

On a vu dans le chapitre précédent que, mis à part quelques cas particuliers, la construction de tests optimaux est délicate. L'utilisation du cadre asymptotique permet de contourner partiellement cette difficulté. On se placera donc dans la suite dans un régime asymptotique où la taille des observations tend vers l'infini. Cependant, on ne pourra pas obtenir l'optimalité d'une suite de tests de niveau asymptotique donné aussi facilement qu'au chapitre précédent. On se contentera d'une notion plus faible : la consistance. On étudiera alors les tests classiques de Wald et ceux fondés sur la loi du χ^2.

7.1 Consistance d'une suite de tests

Dans tout le chapitre, on note $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in \Theta)$ le modèle statistique. La plupart du temps, on omettra la dépendance des tests en l'observation X de loi $\mathbb{P}_\theta$ inconnue du statisticien. On rappelle la définition suivante déjà introduite au Chapitre 5.

Définition 7.1. *Soit $\alpha \in]0;1[$. Etant donnée une suite $(\phi_n)_n$ de tests, et notant α_n^* la taille de ϕ_n à tout rang fini $n \geq 1$, on dit que la* ***suite de tests est de niveau asymptotique*** *α si*

$$\limsup_{n \to \infty} \alpha_n^* \leq \alpha.$$

Remarque 7.1. *La terminologie des tests asymptotiques est relativement floue et en lieu et place de « suite de tests de niveau asymptotique α » on parle souvent de « test de niveau asymptotique α » ou de « test asymptotique de niveau α ». Lorsque la limite de la suite des tailles des tests $(\phi_n)_n$ vaut α, on parle aussi de « test asymptotiquement de taille α. » Le contexte est en général dénué d'ambiguïté.*

La consistance se définit alors de la manière suivante.

Définition 7.2. *Soit $\alpha \in]0;1[$. On dit que $(\phi_n)_n$, une suite de tests de niveau asymptotique α, est* ***consistante*** *si pour tout $\theta \in \Theta_1$, on a*

$$\lim_{n \to +\infty} \beta_n(\theta) = 0$$

où $\beta_n(\theta)$ est définie par

$$\beta_n(\theta) := \mathbb{E}_\theta[1 - \phi_n].$$

7.2 Test de Wald

Le test de Wald permet de traiter le problème du test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ pour tout $\theta_0 \in \Theta$. Pour cela, on suppose que pour tout $\theta \in \Theta$ on dispose d'un estimateur $\hat{\theta}_n$ de θ tel que sous $\mathbb{P}_\theta$, on a la convergence en loi suivante :

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, v(\theta)),$$

avec $v(\theta) > 0$ et d'un estimateur $\hat{v}_n$ de $v(\theta)$ tel que sous $\mathbb{P}_\theta$, $\hat{v}_n$ converge en $\mathbb{P}_\theta$-probabilité vers $v(\theta)$. Pour tout n, on pose alors

$$T_n := \frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sqrt{\hat{v}_n}}$$

et on définit le **test de Wald** associé à cette statistique de test :

$$\phi_n := 1_{\left\{|T_n| \geq z_{1-\frac{\alpha}{2}}\right\}}$$

où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$. On obtient le résultat suivant :

Proposition 7.1. *Soient $\alpha \in]0;1[$ et $\theta_0 \in \Theta$. Pour le problème du test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$, la suite de tests $(\phi_n)_n$ est de niveau asymptotique α et consistante.*

Démonstration. Le lemme de Slutzky implique que, sous $\mathbb{P}_{\theta_0}$, T_n converge en loi vers la loi $\mathcal{N}(0,1)$. On a donc que

$$\mathbb{E}_{\theta_0}[\phi_n] = \mathbb{P}_{\theta_0}\left(|T_n| \geq z_{1-\frac{\alpha}{2}}\right) \to \alpha$$

et $(\phi_n)_n$ est de niveau asymptotique α. On considère $\theta_1 \in \Theta$ avec $\theta_1 \neq \theta_0$. On décompose $T_n = U_n + V_n$ avec

$$U_n := \frac{\sqrt{n}(\hat{\theta}_n - \theta_1)}{\sqrt{\hat{v}_n}}$$

et

$$V_n := \frac{\sqrt{n}(\theta_1 - \theta_0)}{\sqrt{\hat{v}_n}}.$$

On a

$$T_n^{-1} = \frac{1}{U_n + V_n} = \frac{V_n^{-1}}{V_n^{-1}U_n + 1}.$$

Sous $\mathbb{P}_{\theta_1}$, U_n converge en loi vers $Z \sim \mathcal{N}(0,1)$ et V_n^{-1} converge en probabilité vers 0. Le lemme de Slutzky implique que, sous $\mathbb{P}_{\theta_1}$, le vecteur (U_n, V_n^{-1}) converge en loi vers le vecteur $(Z, 0)$. La fonction $\Psi : (x, y) \longmapsto y/(xy+1)$ étant continue, on a que $T_n^{-1} = \psi(U_n, V_n^{-1})$ tend en loi vers $\psi(Z, 0) = 0$. Donc T_n^{-1} tend vers 0 en $\mathbb{P}_{\theta_1}$-probabilité. On a finalement que

$$\begin{aligned}\mathbb{E}_{\theta_1}[\phi_n] &= \mathbb{P}_{\theta_1}\big(|T_n| \geq z_{1-\frac{\alpha}{2}}\big)\\ &= \mathbb{P}_{\theta_1}\big(|T_n|^{-1} \leq z_{1-\frac{\alpha}{2}}^{-1}\big)\\ &= 1 - \mathbb{P}_{\theta_1}\big(|T_n|^{-1} > z_{1-\frac{\alpha}{2}}^{-1}\big) \to 1,\end{aligned}$$

quand $n \to +\infty$. Cela établit la consistance de $(\phi_n)_n$. □

7.3 Test du χ^2

Le test de Wald vu dans la section précédente, aurait pu s'écrire sous la forme

$$\phi_n := 1_{\left\{T_n^2 \geq q_{1-\alpha}\right\}}$$

où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(1)$. En effet, sous $\mathbb{P}_{\theta_0}$, T_n^2 converge en loi vers la loi $\chi^2(1)$. On peut donc réécrire le test de Wald sous la forme d'un test asymptotique du χ^2. Les tests du χ^2 forment un des outils statistiques les plus populaires. Il en existe plusieurs variantes, selon que l'on teste un ajustement simple, un ajustement à une famille de lois ou une propriété d'indépendance. Ils ne s'appliquent qu'à des observations discrètes ou rendues discrètes grâce à un partitionnement de l'ensemble des valeurs possibles. Ce sont des tests asymptotiques qui reposent essentiellement sur le TCL vectoriel, ainsi que sur le théorème de Cochran que l'on rappelle ci-dessous. Dans la suite, on note $\|\cdot\|$ la norme euclidenne.

Théorème 7.1 (Cochran)**.** *On considère un vecteur gaussien $X \sim \mathcal{N}(0, I_d)$ et une décomposition $E_1 \oplus \cdots \oplus E_r$ de $\mathbb{R}^d$ en sous-espaces orthogonaux de dimensions respectives $d_1, \ldots, d_r$. Les projections orthogonales $\Pi_{E_1}X$, ..., $\Pi_{E_r}X$ forment des vecteurs gaussiens indépendants. Par ailleurs, pour $j \in \{1, \ldots, r\}$,*

$$\|\Pi_{E_j}X\|^2 \sim \chi^2(d_j).$$

7.3.1 Test d'ajustement à une loi donnée

Dans ce paragraphe, l'observation est formée par un n-échantillon $X = (X_1, \ldots, X_n)$ donc chacune des variables X_i prend ses valeurs dans l'ensemble discret $\Omega = \{x_1, \ldots, x_d\}$ pour $d \in \mathbb{N}^*$. On note $p = (p_1, \ldots, p_d)$ la loi commune des X_i, i.e. $\forall\, i \in \{1, \ldots, n\}$, $\forall\, j \in \{1, \ldots, d\}$,

$$\mathbb{P}_p(X_i = x_j) = p_j.$$

Le vecteur p joue également le rôle du paramètre θ inconnu. On se donne une loi de référence $p^0 = (p_1^0, \ldots, p_d^0)$ de **support plein**, i.e. $\forall\, j \in \{1, \ldots, d\}$, $p_j^0 > 0$ et $\sum_{j=1}^d p_j^0 = 1$. On désire tester si les X_i ont pour loi commune p^0, i.e. faire le test d'hypothèses de $H_0 : p = p^0$ contre $H_1 : p \neq p^0$, ce qui est équivalent à tester

$$H_0 : \forall j \ p_j = p_j^0 \qquad \textit{contre} \qquad H_1 : \exists j \ p_j \neq p_j^0.$$

On utilise la méthode des moments pour estimer p. Pour tout $j \in \{1, \ldots, d\}$, on pose

$$\widehat{p}_{j,n} := \frac{N_{j,n}}{n}, \quad N_{j,n} := \sum_{i=1}^n 1_{\{X_i = x_j\}}.$$

Ces estimateurs sont des estimateurs fortement consistants et sans biais des p_j.

Remarque 7.2. *La statistique $N_n := (N_{1,n}, \ldots, N_{d,n})$ suit une* ***loi multinomiale de paramètres*** n ***et*** p *: Pour tout d-uplet d'entiers $k = (k_1, \ldots, k_d)$ tels que $k_1 + k_2 + \cdots + k_d = n$,*

$$\mathbb{P}_p(N_n = k) = \mathbb{P}_p((N_{1,n}, \ldots, N_{d,n}) = (k_1, \ldots, k_d)) = \frac{n!}{k_1! \cdots k_d!} p_1^{k_1} \cdots p_d^{k_d}.$$

Sous H_1, il y a lieu de penser que $\widehat{p}_n := (\widehat{p}_{1,n}, \ldots, \widehat{p}_{d,n})$ est loin de p^0 lorsque n est grand. La statistique de test consiste en une mesure de l'écart entre $\widehat{p}_n$ et p^0. On ne considère pas la distance euclidienne (voir la preuve du théorème ci-dessous) mais plutôt la **pseudo-distance du** χ^2 ou **statistique de Pearson** :

$$D_n^2(\widehat{p}_n, p^0) := n \sum_{j=1}^d \frac{(\widehat{p}_{j,n} - p_j^0)^2}{p_j^0} = \sum_{j=1}^d \frac{(N_{j,n} - np_j^0)^2}{np_j^0}.$$

Sous H_0, lorsque $n \to +\infty$, la loi limite des $D_n^2(\widehat{p}_n, p^0)$ est non-seulement connue mais également indépendante de p^0. Le théorème ci-dessous précise cela et indique le comportement de la statistique sous H_1.

Théorème 7.2. *La statistique de Pearson $D_n^2(\widehat{p}_n, p^0)$ admet les comportements limites suivants : Lorsque $n \to +\infty$,*
- sous H_0 ($p = p^0$), $D_n^2(\widehat{p}_n, p^0)$ converge en loi vers la loi $\chi^2(d-1)$,
- sous H_1 ($p \neq p^0$), $D_n^2(\widehat{p}_n, p^0)$ tend vers $+\infty$ presque sûrement.

Démonstration. On commence par le comportement sous H_0. Pour tout $i \in \{1, \ldots, n\}$, on définit les vecteurs aléatoires

$$Z_i = \left(\frac{1}{\sqrt{p_1^0}} \Big(1_{\{X_i = x_1\}} - p_1^0 \Big), \ \ldots, \ \frac{1}{\sqrt{p_d^0}} \Big(1_{\{X_i = x_d\}} - p_d^0 \Big) \right)^T.$$

Les vecteurs $Z_1, \dots, Z_n$ sont indépendants et identiquement distribués selon une loi admettant un moment d'ordre deux, d'espérance le vecteur nul et de matrice de variance–covariance Γ définie pour tous $(j, j') \in \{1, \dots, d\}^2$, $j \neq j'$, par

$$\begin{aligned}\Gamma_{jj} &= \operatorname{var}\left(\frac{1}{\sqrt{p_j^0}}\left(1_{\{X_1 = x_j\}} - p_j^0\right)\right) = \frac{p_j^0\left(1 - p_j^0\right)}{p_j^0} = 1 - p_j^0 \qquad \text{et} \\ \Gamma_{jj'} &= \operatorname{cov}\left(\frac{1}{\sqrt{p_j^0}}\left(1_{\{X_1 = x_j\}} - p_j^0\right),\ \frac{1}{\sqrt{p_{j'}^0}}\left(1_{\{X_1 = x_{j'}\}} - p_{j'}^0\right)\right) = -\sqrt{p_j^0}\sqrt{p_{j'}^0}.\end{aligned}$$

On note pour la suite $\sqrt{p^0} = \left(\sqrt{p_1^0}, \dots, \sqrt{p_d^0}\right)^T$. Par le théorème de la limite centrale multi-dimensionnel, on a ainsi la convergence en loi, lorsque n tend vers $+\infty$:

$$\sqrt{n}\left(\frac{1}{\sqrt{p_1^0}}\left(\widehat{p}_{1,n} - p_1^0\right), \dots, \frac{1}{\sqrt{p_d^0}}\left(\widehat{p}_{d,n} - p_d^0\right)\right)^T = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i \rightsquigarrow V$$

où $V \sim \mathcal{N}\left(0, I_d - \sqrt{p^0}\sqrt{p^0}^T\right)$. Par image continue, en considérant l'application $\|\cdot\|^2$: $(x_1, \dots, x_d) \mapsto x_1^2 + \dots + x_d^2$, il vient la convergence en loi

$$D_n^2\left(\widehat{p}_n, p^0\right) \rightsquigarrow \|V\|^2.$$

La matrice $\sqrt{p^0}\sqrt{p^0}^T$ est la matrice de projection orthogonale sur $\operatorname{vect}\left(\sqrt{p^0}\right)$, l'espace vectoriel engendré par le vecteur $\sqrt{p^0}$. On en déduit que $I_d - \sqrt{p^0}\sqrt{p^0}^T$ est la matrice de projection orthogonale sur le sous-espace vectoriel orthogonal à $\operatorname{vect}\left(\sqrt{p^0}\right)$, qui est donc de dimension $d-1$. Par conséquent, V a même loi que la projection d'un vecteur normal standard sur ce sous-espace. Par le théorème de Cochran, on a ainsi que $\|V\|_2^2 \sim \chi^2(d-1)$, ce qui conclut la preuve du premier point.

Pour prouver le second résultat, on remarque qu'il existe, sous H_1, un entier j tel que $p_j \neq p_j^0$. Or, on a

$$D_n^2\left(\widehat{p}_n, p^0\right) \geq n\,\frac{\left(\widehat{p}_{j,n} - p_j^0\right)^2}{p_j^0} \sim n\,\frac{\left(p_j - p_j^0\right)^2}{p_j^0} \longrightarrow +\infty \qquad \text{p.s.,}$$

où l'équivalence procède de la loi forte des grands nombres. □

On obtient alors le corollaire suivant.

Corollaire 7.1. *Soit $\alpha \in]0;1[$. Le test défini par*

$$\phi(X_1, \dots, X_n) = 1_{\left\{D_n^2(\widehat{p}_n, p^0) > c_{d-1,1-\alpha}\right\}}$$

où $c_{d-1,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(d-1)$ est un test asymptotiquement de taille α. En outre, il est consistant.

Démonstration. On calcule tout d'abord l'erreur de première espèce :

$$\begin{aligned}\mathbb{E}_{p_0}[\phi(X_1,\ldots,X_n)] &= \mathbb{P}_{p_0}\left(D_n^2(\widehat{p}_n,p^0) > c_{d-1,1-\alpha}\right)\\ &= 1-\mathbb{P}_{p_0}\left(D_n^2(\widehat{p}_n,p^0) \leq c_{d-1,1-\alpha}\right),\end{aligned}$$

qui tend vers α lorsque n tend vers $+\infty$, par le premier point du théorème précédent. Dans un second temps, on se place sous H_1. Soit $p \neq p_0$. Par le second point du théorème précédent, on a que $D_n^{-2}(\widehat{p}_n,p^0)$ tend vers 0 presque sûrement donc en probabilité. On a alors

$$\begin{aligned}\mathbb{E}_{p}[1-\phi(X_1,\ldots,X_n)] &= \mathbb{P}_{p}\left(D_n^2(\widehat{p}_n,p^0) \leq c_{d-1,1-\alpha}\right)\\ &= \mathbb{P}_{p}\left(D_n^{-2}(\widehat{p}_n,p^0) \geq c_{d-1,1-\alpha}^{-1}\right),\end{aligned}$$

qui tend vers 0 lorsque n tend vers $+\infty$. □

Remarque 7.3. *Les garanties données par le test du χ^2 reposent sur le Théorème 7.2 et sont donc asymptotiques. En pratique, on considère cette asymptotique respectée lorsque la taille d'échantillon vérifie $n \geq 30$ et que les probabilités p_j^0 des modalités sont telles que $np_j^0 \geq 5$ pour tout $j \in \{1,\ldots,d\}$. Si ce n'est pas le cas, il faut effectuer un regroupement d'une ou plusieurs modalités en une seule classe.*

Exemple 7.1. *Avec $d=3$ et si $p_1^0 = p_2^0 = \frac{1}{10}$ et $p_3^0 = \frac{4}{5}$ avec $n=30$, on a $np_1^0 = np_2^0 = 3$ et $np_3^0 = 24$. On regroupe les classes x_1 et x_2 en une classe x_{12} de probabilité (sous H_0) $p_{12}^0 := \mathbb{P}_{p_0}(X_1 \in \{x_1,x_2\}) = p_1^0 + p_2^0 = \frac{1}{5}$ et $np_{12}^0 = 6$, ce qui convient.*

Nous illustrons ce qui précède sur un exemple.

Exemple 7.2. *On suppose que la couleur d'une variété de tulipe est déterminée par deux allèles R et B d'un gène. Une tulipe est rouge, rose ou blanche selon que les deux allèles de ce gène sont RR, RB ou BB. On croise des tulipes roses entre elles : la théorie mendélienne de l'hérédité (brassage des allèles au hasard) postule que l'on devrait obtenir à la nouvelle génération des tulipes rouges, roses et blanches dans les proportions respectivement égales à 25 %, 50 % et 25 %. C'est ce fait que l'on souhaite tester. Un protocole correspondant à l'expérience précédente est mis en œuvre et 52 croisements sont effectués indépendamment et dans les mêmes conditions. On observe que l'on obtient 31 tulipes roses, 12 blanches et 9 rouges. Ces observations sont la réalisation d'un 52–échantillon dont la loi commune p a un support inclus dans {RR, RB, BB}.*

On pose comme loi de référence $p^0 = (1/4,\ 1/2,\ 1/4)$. Il s'agit de tester $H_0 : p = p^0$ contre l'hypothèse fantôme $H_1 : p \neq p^0$. Les conditions asymptotiques de la Remarque 7.3 sont bien vérifiées, il n'y a pas lieu de procéder à un regroupement de classes. La réalisation de $\widehat{p}_{52}$ est

$$\widehat{p}_{52} = (9/52,\ 31/52,\ 12/52),$$

de sorte que la réalisation de la statistique de test $D^2_{52}\big(\widehat{p}_{52}, p^0\big)$ vaut 2.27. On en déduit que la p–valeur (asymptotique) associée est égale à

$$\widehat{\alpha} := \mathbb{P}\{Z > 2.27\} = 0.32,$$

où Z est une variable qui suit une loi du χ^2 à 2 degrés de liberté. Si l'on choisit un seuil d'erreur de première espèce supérieur à 32 %, on rejette H_0, sinon on la conserve. Ici, il faut noter que l'on a effectué le calcul de la p–valeur en se plaçant dans le cadre asymptotique et en utilisant la loi limite. Les valeurs que l'on donne ne sont donc qu'approximatives.

Le test du χ^2 est bien adapté aux lois discrètes mais peut en fait être utilisé pour toutes les lois, même si, dans le cas d'une loi continue, on préfèrera utiliser le test de Kolmogorov. Dans le cas d'une loi continue, on construit une partition finie de l'ensemble des observations en d classes $K_1, \ldots, K_d$. A partir des observations $X_1, \ldots, X_n$, on construit alors les modalités observées X'_i selon $X'_i = j$ si et seulement si $X_i \in K_j$, puis on applique le test présenté précédemment aux variables X'_i. Bien entendu, le choix de la partition $K_1, \ldots, K_d$ est un problème délicat et crucial et des partitions différentes peuvent conduire à des conclusions différentes (conservation ou rejet de H_0). Une règle classique est d'avoir des sous-ensembles de probabilités à peu près égales sous H_0.

7.3.2 Test d'ajustement à une famille paramétrée de lois

Dans ce paragraphe, on se donne une famille de lois de référence sur Ω,

$$\mathcal{F} = \big\{p(\theta),\ \theta \in \Theta\big\},$$

où Θ est un ouvert de $\mathbb{R}^k$. On suppose que la fonction $p : \mathbb{R}^k \longmapsto \mathbb{R}^d$ est injective. Cela implique en particulier que $k \leq d$. En utilisant les même principes que ceux étudiés en Section 7.3.1, il s'agit ici de tester l'ajustement à $\mathcal{F}$:

$$H_0 : p \in \mathcal{F} \qquad \text{contre} \qquad H_1 : p \notin \mathcal{F}.$$

Exemple 7.3. *Tester si la loi étudiée est une loi binomiale.*

Le principe est le même que dans la Section 7.3.1, à ceci près qu'il faut déterminer au préalable une loi de référence particulière dans $\mathcal{F}$. A cet effet, on estime le paramètre θ par $\widehat{\theta}_n$, obtenu par maximum de vraisemblance, et on considère alors

$$p\Big(\widehat{\theta}_n\Big) = \Big(p_1\Big(\widehat{\theta}_n\Big), \ldots, p_d\Big(\widehat{\theta}_n\Big)\Big).$$

Notons qu'il peut arriver, pour n petit mais pas lorsque $n \to +\infty$, que $\widehat{\theta}_n$ ne soit pas dans Θ mais seulement dans sa fermeture. Cela conduit à la statistique de test

$$D^2_n\big(\widehat{p}_n, \mathcal{F}\big) = n \sum_{j=1}^{d} \frac{\Big(\widehat{p}_{j,n} - p_j\Big(\widehat{\theta}_n\Big)\Big)^2}{p_j\Big(\widehat{\theta}_n\Big)} = \sum_{j=1}^{d} \frac{\Big(N_{j,n} - n\, p_j\Big(\widehat{\theta}_n\Big)\Big)^2}{n\, p_j\Big(\widehat{\theta}_n\Big)}.$$

Théorème 7.3. *Soient d et k deux entiers tels que $k < d-1$ et Θ un ouvert de $\mathbb{R}^k$. On considère l'application*

$$p: \ \theta \in \Theta \longmapsto p(\theta) = \big(p_1(\theta), \ \ldots, \ p_d(\theta)\big),$$

d'image notée $\mathcal{F}$ et contenue dans l'ensemble des probabilités sur d éléments. On suppose que p est injective, de classe C^2 et qu'aucune de ses composantes p_j ne s'annule sur Θ lorsque $j \in \{1, \ldots, d\}$. Nous supposons également que les k dérivées partielles de la fonction p, $\frac{\partial p(\theta)}{\partial \theta_1}, \ldots, \frac{\partial p(\theta)}{\partial \theta_k}$, sont linéairement indépendantes en tout point de Θ. Alors, si $(X_1, \ldots, X_n)$ est un n–échantillon de loi $p(\theta)$ et que les estimateurs du maximum de vraisemblance $\widehat{\theta}_n$ définissent une suite d'estimateurs fortement consistants de θ, lorsque $n \to +\infty$, on a la convergence en loi

$$D_n^2\big(\widehat{p}_n, \mathcal{F}\big) \ \rightsquigarrow \ \chi^2(d-1-k).$$

Ce théorème décrit le comportement sous H_0 de la statistique de test $D_n^2\big(\widehat{p}_n, \mathcal{F}\big)$: elle converge vers la loi du χ^2 à $d-1-k$ degrés de liberté. Sous H_1, on voit par loi forte des grands nombres (comme dans la preuve du Théorème 7.2) que si la distance de p (le vrai paramètre) à $\mathcal{F}$ est strictement positive, la statistique de test diverge presque sûrement vers $+\infty$.

Le test d'ajustement à $\mathcal{F}$ est ainsi

$$\phi(X_1, \ldots, X_n) = 1_{\left\{D_n^2(\hat{p}_n, \mathcal{F}) > c_{d-1-k, 1-\alpha}\right\}},$$

où $c_{d-1-k,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(d-1-k)$. Sous les hypothèses du théorème, il est asymptotiquement de taille α. Par ailleurs, il est consistant, au remplacement près de l'hypothèse alternative par $H_1 : p \notin \overline{\mathcal{F}}$, où $\overline{\mathcal{F}}$ est la fermeture de $\mathcal{F}$, qui assure que sous H_1 la distance de p à $\mathcal{F}$ est bien strictement positive. Enfin, on dispose du même jeu de conditions pratiques d'application qu'à la Remarque 7.3, à ceci près qu'on y remplace la loi de référence par la loi dans $\mathcal{F}$ correspondant au paramètre proposé par l'estimation par maximum de vraisemblance.

Remarque 7.4. *On note une diminution du nombre de degrés de liberté dans la loi limite de la statistique de Pearson sous H_0, ce qui conduit à une région de rejet plus grande et donc, à une augmentation de la puissance du test. Cela peut sembler paradoxal, puisqu'il y a estimation de paramètres et donc plus grande incertitude en apparence. Cependant, ce problème d'estimation est évacué à la limite, par consistance. Il ne reste donc plus que le fait que la famille de probabilités considérées est de dimension plus petite, grâce aux contraintes données par la paramétrisation, ce qui facilite le test.*

7.3.3 Application : test du χ^2 d'indépendance

On dispose ici de n couples d'observations $(X_1, Y_1), \ \ldots, \ (X_n, Y_n)$ prenant des valeurs discrètes, indépendants et identiquement distribués selon une loi notée ν. On veut tester H_0 : « *les observations X_i sont indépendantes des observations Y_i.* » Tester cette indépendance revient à tester l'ajustement de la loi jointe ν à la famille des lois-produits. Le test décrit ici est donc une conséquence du test général décrit au paragraphe 7.3.2. Nous allons voir comment appliquer concrètement le Théorème 7.3.

On note $\mathcal{X} = \{x_1, \ldots, x_r\}$ l'ensemble des valeurs possibles pour les X_i et

$$\mathcal{P}(\mathcal{X}) = \left\{ p = (p_1, \ldots, p_r) : \quad \sum_{j=1}^{r} p_j = 1, \ \inf_{j \in \{1, \ldots, r\}} p_j > 0 \right\}.$$

On procède de même pour les Y_i en notant $\mathcal{Y} = \{y_1, \ldots, y_s\}$ et $\mathcal{P}(\mathcal{Y})$ les ensembles analogues. Il s'agit de tester ici l'appartenance de ν à la famille

$$\mathcal{F} = \big\{ p \otimes q, \quad p \in \mathcal{P}(\mathcal{X}) \text{ et } q \in \mathcal{P}(\mathcal{Y}) \big\}.$$

L'hypothèse nulle se réécrit $H_0 : \nu \in \mathcal{F}$. Ici, la paramétrisation injective peut être réalisée en termes des $p_1, \ldots, p_{r-1}, q_1, \ldots, q_{s-1}$, soit avec $k = (r-1) + (s-1) = r + s - 2$ paramètres. Ce nombre k est à comparer à la dimension $d = rs$ dans laquelle vit le paramètre d'intérêt ν.

On introduit des notations pour les décomptes d'occurrences : pour tous $j \in \{1, \ldots, r\}$ et $\ell \in \{1, \ldots, s\}$,

$$N_{(j,\cdot),n} = \sum_{i=1}^{n} 1_{\{X_i = x_j\}} , \qquad N_{(\cdot,\ell),n} = \sum_{i=1}^{n} 1_{\{Y_i = y_\ell\}}$$

et

$$N_{(j,\ell),n} = \sum_{i=1}^{n} 1_{\{(X_i, Y_i) = (x_j, y_\ell)\}}.$$

A partir de ces décomptes, on définit les fréquences empiriques

$$\widehat{p}_{j,n} = \frac{N_{(j,\cdot),n}}{n}, \qquad \widehat{q}_{\ell,n} = \frac{N_{(\cdot,\ell),n}}{n} \qquad \text{et} \qquad \widehat{\nu}_{(j,\ell),n} = \frac{N_{(j,\ell),n}}{n},$$

puis les deux marginales empiriques $\widehat{p}_n = \big(\widehat{p}_{1,n}, \ldots, \widehat{p}_{r,n}\big)$ et $\widehat{q}_n = \big(\widehat{q}_{1,n}, \ldots, \widehat{q}_{s,n}\big)$.

L'estimateur du maximum de vraisemblance de ν sous H_0 est donné par le produit des marginales empiriques $\widehat{p}_n \otimes \widehat{q}_n$. Il suffit pour le voir d'étudier la fonction de vraisemblance

$$\begin{aligned}(p_1, \ldots, p_{r-1}, q_1, \ldots, q_{s-1}) \longmapsto & \left(\big(1 - (p_1 + \cdots + p_{r-1})\big)^{N_{(r,\cdot),n}} \prod_{j=1}^{r-1} p_j^{N_{(j,\cdot),n}} \right) \\ & \times \left(\big(1 - (q_1 + \cdots + q_{s-1})\big)^{N_{(\cdot,s),n}} \prod_{\ell=1}^{s-1} q_\ell^{N_{(\cdot,\ell),n}} \right).\end{aligned}$$

Sous H_0, cet estimateur est fortement consistant d'après la loi forte des grands nombres. La statistique de test $D_n^2\big(\widehat{p}_n, \mathcal{F}\big)$ prend alors la forme suivante :

$$\begin{aligned} D_n^2\big(\widehat{p}_n,\, \widehat{q}_n,\, \widehat{\nu}_n\big) &= n \sum_{j=1}^{r} \sum_{\ell=1}^{s} \frac{\big(\widehat{\nu}_{(j,\ell),n} - \widehat{p}_{j,n}\widehat{q}_{\ell,n}\big)^2}{\widehat{p}_{j,n}\widehat{q}_{\ell,n}} \\ &= \sum_{j=1}^{r} \sum_{\ell=1}^{s} \frac{\big(N_{(j,\ell),n} - N_{(j,\,\cdot\,),n}\, N_{(\,\cdot\,,\ell),n}/n\big)^2}{N_{(j,\,\cdot\,),n}\, N_{(\,\cdot\,,\ell),n}/n}. \end{aligned}$$

Par ailleurs, on peut vérifier que les $k = r + s - 2$ dérivées partielles de l'application définie sur l'ensemble ouvert idoine

$$(p_1, \ldots, p_{r-1}, q_1, \ldots, q_{s-1}) \longmapsto p \otimes q \in \mathbb{R}^{rs}$$

sont linéairement indépendantes en tout point. Toutes les hypothèses du Théorème 7.3 étant vérifiées, on en déduit que sous H_0,

$$D_n^2\big(\widehat{p}_n,\, \widehat{q}_n,\, \widehat{\nu}_n\big) \;\rightsquigarrow\; \chi^2(d-1-k) = \chi^2\big(rs - 1 - (r+s-2)\big) = \chi^2\big((r-1)(s-1)\big).$$

Sous H_1, la statistique de test tend presque sûrement vers $+\infty$.

Le test d'ajustement à la famille $\mathcal{F}$ des lois-produits est

$$\phi(X_1, \ldots, X_n) = 1_{\left\{D_n^2(\hat{p}_n,\, \hat{q}_n,\, \hat{\nu}_n) > c_{(r-1)(s-1),1-\alpha}\right\}},$$

où $c_{(r-1)(s-1),1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2((r-1)(s-1))$. Il est consistant et asymptotiquement de taille α. On effectue les mêmes remarques que précédemment quant aux conditions de son application pratique.

Exemple 7.4. *On teste l'indépendance entre le nombre d'enfants d'un ménage et son revenu sur une population de n = 25 263 ménages en Suède au milieu du siècle passé. Les ménages sont classés en 4 catégories selon leur revenu : de R1 pour les revenus les plus faibles à R4 pour les revenus les plus importants. On obtient les résultats suivants.*

	R1	R2	R3	R4	TOTAL
0 enfant	2 161	3 577	2 184	1 636	9 558
1 enfant	2 755	5 081	2 222	1 052	11 110
2 enfants	936	1 753	640	306	3 635
3 enfants	225	419	96	38	778
$\geq$ 4 enfants	39	98	31	14	182
TOTAL	6 116	10 928	5 173	3 046	25 263

Table 7.1 – Effectifs des ménages suédois en fonction du nombre d'enfants et selon le revenu.

On calcule

$$D_n^2\big(\widehat{p}_n,\,\widehat{q}_n,\,\widehat{\nu}_n\big)=568,5$$

ce qui donne une p-valeur

$$\widehat{\alpha}:=\mathbb{P}\left\{Z>568,5\right\}=0$$

où Z est une variable qui suit une loi du χ^2 à $(5-1)\times(4-1)=12$ degrés de liberté. On vérifie au préalable que la proportion estimée de chaque case multipliée par n est supérieure à 5. Notons qu'il suffit de le faire pour la proportion estimée de ménages suédois avec au moins 4 enfants de revenu R4 qui vaut

$$\widehat{\nu}_{(4,5),n}=\frac{182}{25263}\times\frac{3046}{25263}\approx 0,0009$$

et

$$\widehat{\nu}_{(4,5),n}\times 25263\approx 21,9\geq 5.$$

Annexe A

Définitions et propriétés des lois classiques

A.1 Lois discrètes classiques

Dans le tableau ci-dessous, $n \in \mathbb{N}^*$, $p \in]0;1[$, $\lambda \in \mathbb{R}_+^*$. La notation δ désigne la masse de Dirac.

Loi de la variable X	$\mathbb{P}^X$	Espérance	variance	Fonction caractéristique
Bernoulli $Ber(p)$	$(1-p)\delta_0 + p\delta_1$	p	$p(1-p)$	$\varphi(t) = 1 - p + pe^{it}$
Binomiale $Bin(n,p)$	$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \delta_k$	np	$np(1-p)$	$\varphi(t) = (1 - p + pe^{it})^n$
Poisson $\mathcal{P}(\lambda)$	$\sum_{k=0}^{+\infty} \frac{e^{-\lambda}\lambda^k}{k!} \delta_k$	λ	λ	$\varphi(t) = e^{\lambda(e^{it}-1)}$
Géométrique $\mathcal{G}(p)$	$\sum_{k=1}^{+\infty} p(1-p)^{k-1} \delta_k$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\varphi(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$

A.2 Lois classiques à densité

Dans le tableau ci-dessous, $m \in \mathbb{R}$, $\sigma \in \mathbb{R}_+^*$, $\lambda \in \mathbb{R}_+^*$, $d \in \mathbb{N}^*$

Loi de la variable X	densité de $\mathbb{P}^X$	Espérance	variance	Fonction caractéristique
Normale $\mathcal{N}(m, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$\varphi(t) = e^{imt - \frac{\sigma^2 t^2}{2}}$
Uniforme $U(]0;1[)$	$f(x) = 1_{]0;1[}(x)$	$\frac{1}{2}$	$\frac{1}{12}$	$\varphi(t) = \frac{e^{it}-1}{it} 1_{\{t \neq 0\}} + 1_{\{t=0\}}$
Exponentielle $\mathcal{E}(\lambda)$	$f(x) = \lambda e^{-\lambda x} 1_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\varphi(t) = \frac{\lambda}{\lambda - it}$
Cauchy $\mathcal{C}$	$f(x) = \frac{1}{\pi(1+x^2)}$	$\nexists$	$\nexists$	$\varphi(t) = e^{-\lvert t \rvert}$
χ^2 centrée $\chi^2(d)$	$f(x) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} 1_{\mathbb{R}_+}(x)$	d	$2d$	$\varphi(t) = (1 - 2\,i\,t)^{-d/2}$

La notation $\nexists$ signifie « n'existe pas ». La fonction Γ est définie par pour $t > 0$ par $\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx$.

Annexe B

Vecteurs gaussiens

B.1 Définitions

Rappelons la définition des variables aléatoires gaussiennes réelles.

Définition B.1. *On définit :*

— *Une variable aléatoire réelle Z est dite* **gaussienne centrée réduite** *si elle admet pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ la fonction :*

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

On note $Z \sim \mathcal{N}(0,1)$.

— *Une variable aléatoire réelle X est dit* **gaussienne** *s'il existe $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$ et $Z \sim \mathcal{N}(0,1)$ tels que $X = \mu + \sigma Z$. La densité de X est*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

On note $X \sim \mathcal{N}(\mu, \sigma^2)$. Quand $\sigma = 0$, on dit que X est une variable gaussienne dégénérée.

Une variable gaussienne est caractérisée par sa fonction caractéristique donnée par la proposition suivante.

Théorème B.1. *La fonction caractéristique de $X \sim \mathcal{N}(\mu, \sigma^2)$ est donnée par*

$$\forall\, t \in \mathbb{R}, \quad \phi_X(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

Preuve : ϕ_X se calcule à l'aide de ϕ_Z où $Z \sim \mathcal{N}(0,1)$ et on montre que

$$\forall\, t \in \mathbb{R}, \quad \phi_Z(t) = -t\phi_Z'(t).$$

□

Introduisons à présent les vecteurs gaussiens.

Définition B.2. *Un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est dit* **gaussien** *si toute combinaison linéaire de ses composantes est une variable aléatoire gaussienne.*
Si $X = (X_1, \ldots, X_d)^T$ est un vecteur gaussien, on définit son **vecteur moyenne** $\mathbb{E}(X)$ *par*

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d))^T$$

et sa **matrice de variance-covariance** *$var(X)$ par*

$$var(X) = \mathbb{E}((X - \mathbb{E}(X)) \times (X - \mathbb{E}(X))^T).$$

Notons que $\mathrm{var}(X)$ est symétrique et

$$\forall\ (i,j) \in \{1, \ldots, d\}^2, \quad \mathrm{var}(X)_{ij} = \mathrm{cov}(X_i, X_j).$$

Remarque B.1. *Si $(X_1, \ldots, X_n)$ est un n-échantillon de loi gausienne alors on a évidemment que $X = (X_1, \ldots, X_n)^T$ est un vecteur gaussien dont la matrice de variance-covariance est proportionnelle à I.*

B.2 Propriétés des vecteurs gaussiens

Donnons la fonction caractéristique d'un vecteur gaussien et les conséquences importantes qui en découlent.

Théorème B.2. *Soit $X = (X_1, \ldots, X_d)^T$ un vecteur gaussien. On note $m = \mathbb{E}(X)$ et $\Sigma = var(X)$. On a que X admet pour fonction caractéristique la fonction*

$$\forall\ t \in \mathbb{R}^d, \quad \phi_X(t) = \mathbb{E}[\exp(it^T X)] = \exp\left(it^T m - t^T \Sigma t/2\right).$$

La loi de X est donc entièrement déterminée par m et Σ. On note $X \sim \mathcal{N}(m, \Sigma)$.

Preuve : On note que

$$\forall\ t \in \mathbb{R}^d, \quad t^T X \sim \mathcal{N}(t^T m, t^T \Sigma t).$$

□

Proposition B.1. (Propriété de linéarité)
Soit $X = (X_1, \ldots, X_d)^T$ un vecteur gaussien. On note $m = \mathbb{E}(X)$ et $\Sigma = var(X)$. On a pour toute matrice A possédant d colonnes et pour tout vecteur $b \in \mathbb{R}^d$,

$$AX + b \sim \mathcal{N}(Am + b, A\Sigma A^T).$$

Preuve : Elle découle du Théorème B.2 $\square$

Proposition B.2. (Propriété pour l'indépendance)
Soit $X = (X_1, \dots, X_d)^T$ un vecteur gaussien. On note $m = \mathbb{E}(X)$ et $\Sigma = var(X)$. Pour tout $(i, j) \in \{1, \dots, d\}^2$ tel que $i \neq j$, X_i et X_j sont indépendantes si et seulement si $cov(X_i, X_j) = 0$.

Preuve : Elle découle du Théorème B.2 $\square$

Remarque B.2. *Les composantes d'un vecteur gaussien sont des variables aléatoires gaussiennes mais la réciproque est fausse. En effet, on considère $X \sim \mathcal{N}(0, 1)$ et $\varepsilon \sim Ber(0.5)$ indépendante de X. Alors $X_1 = X$ et $X_2 = (2\varepsilon - 1)X$ sont des variables gaussiennes mais $(X_1, X_2)^T$ n'est pas un vecteur gaussien. Notons que $cov(X_1, X_2) = 0$ mais que X_1 et X_2 ne sont pas indépendantes.*

Proposition B.3. (Propriété pour l'espérance conditionnelle)
Soit $(Y, X_1, \dots, X_d)$ un vecteur gaussien alors $\mathbb{E}(Y|X_1, \dots, X_d)$ est une fonction affine de $(X_1, \dots, X_d)$.

Preuve : Soit $p_{1,X_1,\dots,X_d}(Y)$ la projection de Y sur $\mathrm{vect}(1, X_1, \dots, X_d)$ pour le produit scalaire associé à l'espérance. Donc $\mathbb{E}[(Y - p_{1,X_1,\dots,X_d}(Y))Z] = 0$ pour toute variable $Z \in \mathrm{vect}(1, X_1, \dots, X_d)$. Avec $Z = 1$, on déduit que $\mathbb{E}[Y - p_{1,X_1,\dots,X_d}(Y)] = 0$. Puis, pour toute variable $Z \in \{X_1, \dots, X_d\}$, $(Y - p_{1,X_1,\dots,X_d}(Y), X_1, \dots, X_d)$ étant un vecteur gaussien, $0 = \mathbb{E}[(Y - p_{1,X_1,\dots,X_d}(Y))Z] = \mathrm{cov}(Y - p_{X_1,\dots,X_d}(Y), Z)$ montre que $Y - p_{1,X_1,\dots,X_d}(Y)$ et Z sont indépendantes. Donc $Y - p_{1,X_1,\dots,X_d}(Y)$ est indépendante de toutes fonction de $(X_1, \dots, X_d)$ et $p_{1,X_1,\dots,X_d}(Y) = \mathbb{E}(Y|X_1, \dots, X_d)$. $\square$

A l'aide de la fonction caractéristique, on démontre le TCL vectoriel.

Théorème B.3. *Soient $X_1, \dots, X_n$ des vecteurs aléatoires de $\mathbb{R}^d$ i.i.d. admettant un moment d'ordre 2. On note m leur espérance et Γ leur matrice de variance-covariance. Alors,*

$$\sqrt{n}(\overline{X}_n - m) \overset{n \to +\infty}{\longrightarrow} \mathcal{N}(0, \Gamma) \text{ en loi.}$$

Preuve : On calcule pour tout n la fonction caractéristique de $Z_n = \sqrt{n}(\overline{X}_n - m)$:

$$\forall\, t \in \mathbb{R}^d, \quad \phi_{Z_n}(t) = \mathbb{E}[\exp(it^T Z_n)].$$

On a par le TCL,

$$t^T Z_n \overset{n \to +\infty}{\longrightarrow} \mathcal{N}(0, t^T \Gamma t) \text{ en loi.}$$

Donc

$$\forall\, t \in \mathbb{R}^d, \quad \phi_{Z_n}(t) \overset{n\to+\infty}{\longrightarrow} \exp\left(-\frac{1}{2}t^T\Gamma t\right).$$

□

Théorème B.4. *Soit $X = (X_1, \ldots, X_d)^T$ un vecteur gaussien. On note $m = \mathbb{E}(X)$ et $\Sigma = var(X)$. X admet une densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement $det(\Sigma) \neq 0$.*

— *Si $det(\Sigma) = 0$, la loi de $X - m$ est presque sûrement portée par un espace vectoriel engendré par les vecteurs propres associés aux valeurs propres non nulles de Σ.*
— *Si $det(\Sigma) \neq 0$,*

$$\forall\, x \in \mathbb{R}^d, \quad f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{det(\Sigma)}} \exp\left(-\frac{(x-m)^T\Sigma^{-1}(x-m)}{2}\right).$$

Preuve : La matrice Σ est symétrique. Donc il existe U une matrice orthogonale (composée des vecteurs propres de Σ notés $u_1, u_2, \ldots, u_d$) et il existe $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ ($r = \text{rang}(\Sigma) \leq d$) tels que

$$\Sigma = U\Gamma U^T,$$

avec

$$\Gamma = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

Si $\det(\Sigma) = 0$, on a $r < d$. Pour $i \in \{r+1, \ldots, d\}$, $\mathbb{E}[(u_i^T(X-m))^2] = u_i^T\Sigma u_i = 0$. Donc $u_i^T(X-m) = 0$ p.s. et $X - m$ prend ses valeurs dans $\text{vect}(u_1, \ldots, u_r)$ qui est de mesure de Lebesgue nulle dans $\mathbb{R}^d$.
Si $\det(\Sigma) \neq 0$,, $U\sqrt{\Gamma}$ est inversible. On pose $Y \sim \mathcal{N}(0, I_d)$. Alors $U\sqrt{\Gamma}Y + m \sim X$. Pour toute fonction g continue bornée,

$$\begin{aligned}
\mathbb{E}(g(X)) &= \mathbb{E}(g(U\sqrt{\Gamma}Y + m)) \\
&= \int_{\mathbb{R}^d} g(U\sqrt{\Gamma}y + m)\left(\frac{1}{\sqrt{2\pi}}\right)^d \exp\left(-\frac{\|y\|^2}{2}\right) dy \\
&= \int_{\mathbb{R}^d} g(x)\left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{(x-m)^T\Sigma^{-1}(x-m)}{2}\right) dx.
\end{aligned}$$

B.3 Lois du χ^2, de Student et de Fisher - Théorème de Cochran

Dans cette section, nous nous placerons dans $\mathbb{R}^d$ muni du produit scalaire euclidien et on notera $\|.\|$ la norme euclidienne dans $\mathbb{R}^d$.

Définition B.3. *Soit X un vecteur gaussien de $\mathbb{R}^d$ tel que $\mathbb{E}(X) = m$ et $var(X) = I_d$. La loi de $\|X\|^2$ ne dépend que de d et $\|m\|$. On note*

$$\|X\|^2 \sim \chi^2(d, \|m\|^2)$$

et on dit que $\|X\|^2$ suit une **loi du χ^2 (décentrée si $\|m\| \neq 0$)**. *d est le* **nombre de degrés de liberté**, *$\|m\|^2$ est le* **paramètre de décentrage**.
Lorsque $\|m\| = 0$, on note plus simplement,

$$\|X\|^2 \sim \chi^2(d).$$

Enfin, pour k et ℓ deux entiers, la **loi de Fisher à k et ℓ degrés de liberté**, *notée $\mathcal{F}(k, \ell)$, est la loi de la variable aléatoire*

$$Z = \frac{U/k}{V/\ell} \qquad \text{lorsque} \qquad U \sim \chi^2(k) \quad \text{et} \quad V \sim \chi^2(\ell)$$

sont indépendantes.

Preuve : Soit $Y \in \mathbb{R}^d$ tel que $Y \sim \mathcal{N}(m', I_d)$ avec $\|m\| = \|m'\|$. Il existe U matrice orthogonale telle que $m = Um'$. Donc $UY \sim \mathcal{N}(m, I_d) \sim X$ et

$$\|Y\|^2 = \|UY\|^2 \sim \|X\|^2.$$

□

On a la proposition suivante.

Proposition B.4. *Si $Z_d \sim \chi^2(d)$, on montre que la densité de Z_d est la fonction f telle que*

$$\forall\, x \in \mathbb{R}, \quad f(x) = \frac{\exp(-x/2)x^{d/2-1}}{2^{d/2}\Gamma(d/2)} \mathbb{I}_{\mathbb{R}_+}(x),$$

avec

$$\forall\, a > 0, \quad \Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx.$$

On a :

$$\mathbb{E}(Z_d) = d, \quad var(Z_d) = 2d.$$

Preuve : s'obtient par le calcul. $\square$

Enonçons le résultat principal de cette section.

Théorème B.5. *(Cochran - DDC) Soit $E_1 \oplus \cdots \oplus E_r$ une décomposition de $\mathbb{R}^d$ en sous-espaces deux à deux orthogonaux de dimension respective $d_1, \ldots, d_r$. Si $X \sim \mathcal{N}(m, I_d)$, les vecteurs aléatoires $X_{E_1}, \ldots, X_{E_r}$, projections orthogonales de X sur $E_1, \ldots, E_r$ sont indépendants. Les variables aléatoires $\|X_{E_1}\|^2, \ldots, \|X_{E_r}\|^2$ sont indépendantes et*

$$(\|X_{E_1}\|^2, \ldots, \|X_{E_r}\|^2)^T \sim (\chi^2(d_1, \|m_{E_1}\|^2), \ldots, \chi^2(d_r, \|m_{E_d}\|^2))^T$$

où $m_{E_1}, \ldots, m_{E_r}$, sont les projections de m sur $E_1, \ldots, E_r$.

Preuve : Soit $(e_{j1}, \ldots, e_{jd_j})$ une base orthonormée de E_j. On a

$$\forall\, j \in \{1, \ldots, d\}, \quad X_{E_j} = \sum_{k=1}^{d_j} e_{jk} e_{jk}^T X.$$

Les variables $e_{jk}^T X$ sont indépendantes de loi $\mathcal{N}(e_{jk}^T m, 1)$ donc les vecteurs aléatoires $X_{E_1}, \ldots, X_{E_r}$ sont indépendants. Pour achever la preuve, on remarque que

$$\forall\, j \in \{1, \ldots, d\}, \quad \|X_{E_j}\|^2 = \sum_{k=1}^{d_j} (e_{jk}^T X)^2.$$

$\square$

Une application importante du théorème de Cochran est la suivante.

Proposition B.5. *Soit $X = (X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$. Reprenons les estimateurs obtenus par la méthode des moments ou par le principe du maximum de vraisemblance pour l'estimation de μ et σ^2 :*

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

Alors, on a :

- *$\overline{X}_n$ et S_n^2 sont des variables aléatoires indépendantes.*
- *Les lois des ces variables sont explicites :*

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1).$$

Preuve : On pose pour tout $i \in \{1,\ldots,n\}$, $Y_i = \sigma^{-1}(X_i - \mu)$. On a alors que $(Y_1,\ldots,Y_n)$ est un n-échantillon de loi $\mathcal{N}(0,1)$. On pose ensuite $e = (1,\ldots,1)^T$ et $E = \mathrm{vect}(e)$. On a alors

$$\mathbb{R}^n = E \oplus E^\perp.$$

Les projections de $Y = (Y_1,\ldots,Y_n)^T$ sur E et $E^\perp$, Y_E et $Y_{E^\perp}$ sont indépendantes et valent

$$Y_E = \frac{1}{n}\sum_{i=1}^n Y_i \times e, \quad Y_{E^\perp} = \begin{pmatrix} Y_1 - \frac{1}{n}\sum_{i=1}^n Y_i \\ \vdots \\ Y_n - \frac{1}{n}\sum_{i=1}^n Y_i \end{pmatrix}.$$

On a

$$\frac{1}{\sigma}(\overline{X}_n - \mu) \times e = Y_E, \quad \frac{nS_n}{\sigma^2} = \|Y_{E^\perp}\|^2.$$

□

Ce résultat nous permet de construire des intervalles de confiance pour l'estimation de μ et σ^2 à l'aide de la définition suivante.

Définition B.4. *Si X et Y sont deux variables aléatoires indépendantes telles que*
- $X \sim \mathcal{N}(\mu, 1)$
- $Y \sim \chi^2(d)$,

alors la loi de la variable

$$Z = \frac{X}{\sqrt{\frac{Y}{d}}}$$

est appelée **loi de Student (décentrée si $\mu \neq 0$) à d degrés de liberté**. *On note*

$$Z \sim t(d, \mu).$$

Si **le paramètre de décentrage** $\mu = 0$, *on note plus simplement*

$$Z \sim t(d).$$

Proposition B.6. *Si $Z_d \sim t(d)$, on montre que la densité de Z_d est la fonction f telle que*

$$\forall\, x \in \mathbb{R}, \quad f(x) = \frac{\Gamma((d+1)/2)}{\sqrt{d\pi}\Gamma(d/2)}\left(1 + \frac{x^2}{d}\right)^{-(d+1)/2},$$

avec

$$\forall\, a > 0, \quad \Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx.$$

Pour $d > 1$, on a :

$$\mathbb{E}(Z_d) = 0.$$

Pour $d > 2$, on a :

$$var(Z_d) = \frac{d}{d-2}.$$

On a

$$\lim_{d \to +\infty} Z_d = Z \text{ en loi,}$$

où $Z \sim \mathcal{N}(0,1)$.

Preuve : les premiers points s'obtiennent par le calcul. Pour le dernier point, on peut utiliser le lemme de Slutsky, on peut aussi utiliser la formule de Stirling :

$$\Gamma(x+1) \overset{x \to +\infty}{\sim} \sqrt{2\pi x}\left(\frac{x}{e}\right)^x$$

et le **lemme de Scheffé**.

Lemme B.1. *Soit $(f_n)_n$ est une suite de densités de probabilité par rapport à la mesure de Lebesgue qui converge simplement vers une densité f. Alors si pour tout n, Z_n est une variable aléatoire de densité f_n et si Z est une variable aléatoire de densité f, on a*

$$\lim_{n \to +\infty} Z_n = Z \text{ en loi.}$$

Preuve du lemme : On pose pour tout n, $g_n = f - f_n$. On a

$$\forall\, x \in \mathbb{R}, \quad 0 \le g_n(x)\mathbb{I}_{g_n(x) \ge 0} \le f$$

et par le théorème de convergence dominée

$$\lim_{n \to +\infty} \int g_n(x)\mathbb{I}_{g_n(x) \ge 0} dx = 0.$$

Comme $\int g_n(x)dx = 0$,

$$\begin{aligned}
\int |g_n(x)|dx &= \int g_n(x)\mathbb{I}_{g_n(x) \ge 0}dx - \int g_n(x)\mathbb{I}_{g_n(x) < 0}dx \\
&= 2\int g_n(x)\mathbb{I}_{g_n(x) \ge 0}dx \\
&\longrightarrow 0
\end{aligned}$$

Donc pour tout $t \in \mathbb{R}$,

$$\lim_{n \to +\infty} \left| \int (f_n(x) - f(x))\mathbb{I}_{x \le t} dx \right| \le \lim_{n \to +\infty} \int |g_n(x)|dx = 0.$$

$\square$

Comme la loi normale, la loi de Student est symétrique mais ses queues sont plus épaisses que celles de la loi normale. On déduit de la définition précédente que

$$\frac{\sigma^{-1}\sqrt{n}(\overline{X}_n-\mu)}{\sigma^{-1}\sqrt{\frac{nS_n}{n-1}}}=\frac{(\overline{X}_n-\mu)}{\sqrt{\frac{S_n}{n-1}}}\sim t(n-1).$$

En notant $t_{n-1,1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ pour la loi $t(n-1)$ et $c_{n-1,1-\alpha}$ le quantile d'ordre $1-\alpha$ pour la loi $\chi^2(n-1)$, un intervalle de confiance de niveau de confiance exactement égal à $1-\alpha$ pour μ est :

$$I_{n,\alpha}=\left[\overline{X}_n-t_{n-1,1-\alpha/2}\sqrt{\frac{S_n}{n-1}},\overline{X}_n+t_{n-1,1-\alpha/2}\sqrt{\frac{S_n}{n-1}}\right]$$

et un intervalle de confiance de niveau de confiance exactement égal à $1-\alpha$ pour σ^2 est :

$$J_{n,\alpha}=\left[\frac{nS_n}{c_{n-1,1-\alpha}},+\infty\right[.$$

On déduit de ces intervalles de confiance les tests de taille α de $\mu=\mu_0$ contre $\mu\neq\mu_0$ et de $\sigma^2=\sigma_0^2$ contre $\sigma^2<\sigma_0^2$. Notons que l'on obtient une région de confiance de niveau de confiance $1-2\alpha$ pour l'estimation de $\theta=(\mu,\sigma^2)$ en considérant $I_{n,\alpha}\times J_{n,\alpha}$.

Annexe C

Annales

C.1 Partiel 2019-2020

Statistique mathématique

PARTIEL – 12 mars 2020 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Questions de cours : On observe un n-échantillon $X = (X_1, \ldots, X_n)$ de loi $\mathcal{N}(\theta, 1)$, d'espérance $\theta \in \mathbb{R}$ inconnue mais de variance connue égale à 1. Un estimateur de θ est donné par la moyenne empirique $\bar{X}_n$.

1. Proposer un intervalle de confiance bilatère de niveau exactement égal à 95%.
2. Proposer un intervalle de confiance unilatère de niveau exactement égal à 95%.

Exercice 1 : On observe un n-échantillon $X = (X_1, \ldots, X_n)$ de loi uniforme sur $[\theta - \frac{1}{2}; \theta + \frac{1}{2}]$ où $\theta \in \mathbb{R}$ un inconnu

1. Ecrire la vraisemblance du modèle.
2. L'estimateur du maximum de vraisemblance de θ est-il bien défini ? Justifiez votre réponse.
3. Proposer alors un estimateur consistant de θ.

Exercice 2 : Pour $\theta \in]-1; 1[$, on considère la fonction f_θ définie pour tout $x \in \mathbb{R}$ par

$$f_\theta(x) = (1-\theta)1_{\{-\frac{1}{2} \leq x \leq 0\}} + (1+\theta)1_{\{0 < x \leq \frac{1}{2}\}}.$$

1. Vérifier que f_θ est bien une densité de probabilité.

   On observe alors un n-échantillon $X = (X_1, \ldots, X_n)$ de densité f_θ et on cherche à estimer θ à l'aide de $X = (X_1, \ldots, X_n)$. Dans la suite, on note $M = \sum_{i=1}^n 1_{\{-\frac{1}{2} \leq x \leq 0\}}$.
2. Calculer $\mathbb{E}_\theta[X_1]$, l'espérance de X_1.
3. Montrer que la vraisemblance s'écrit

   $$V_X(\theta) = (1-\theta)^M (1+\theta)^{n-M}.$$

   En déduire que l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est de la forme

   $$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n Y_i,$$

   avec $Y_1, \ldots, Y_n$ des variables indépendantes identiquement distribuées de variance finie que vous préciserez. Vous justifierez soigneusement toutes les étapes.

4. L'estimateur $\hat{\theta}_n$ est-il biaisé ?
5. L'estimateur $\hat{\theta}_n$ est-il consistant ?
6. Donner la loi limite de $\sqrt{n}(\hat{\theta}_n - \theta)$.
7. En déduire un intervalle de confiance asymptotique de niveau égal à 95%.

Exercice 3 : Soit $X = (X_1, \dots, X_n)$ un n-échantillon de loi exponentielle de paramètre $\theta \in \mathbb{R}_+^*$. La densité des variables X_i est donc définie pour $x \in \mathbb{R}_+$ par

$$f_\theta(x) = \theta e^{-\theta x}.$$

Les variables X_i ne sont pas observées, on observe uniquement les variables aléatoires $Y_1, \dots, Y_n$ définies pour $i \in \{1, \dots, n\}$ par

$$Y_i = 1_{\{X_i > 2\}}.$$

1. Donner la loi de Y_1.
2. Calculer $\mathbb{E}_\theta[Y_1]$, l'espérance de Y_1.
3. Ecrire la vraisemblance des observations $(Y_1, \dots, Y_n)$.
4. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ en justifiant soigneusement les étapes.
5. Montrer que $\hat{\theta}_n$ est un estimateur consistant de θ.
6. Donner la loi limite de $\sqrt{n}(\hat{\theta}_n - \theta)$.
7. En déduire un intervalle de confiance asymptotique de niveau égal à 95%.

C.2 Examen 2019-2020

Statistique mathématique

EXAMEN – Mai 2020 – Durée : Environ 3h00

L'utilisation d'internet sera utile pour retrouver la densité et le calcul des valeurs des fonctions de répartition des lois usuelles. La qualité de la rédaction sera essentielle dans la notation : soyez lisibles, clairs et précis (on évitera les longues tirades).

Exercice 1 : On observe un n-échantillon $X = (X_1, \dots, X_n)$ de densité

$$f_\theta(x) = \theta^2 \frac{\log x}{x^{\theta+1}} 1_{]1;+\infty[}(x)$$

où $\theta \in]1, +\infty[$ est inconnu.

1. Vérifier que f_θ est bien une densité de probabilité et calculer l'espérance des variables X_i en fonction de θ.
2. Démontrer que
$$2\theta \sum_{i=1}^{n} \log(X_i) \sim \chi^2(4n).$$
3. Pour $\theta_0 > 1$, on souhaite tester
$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$
Soit $\alpha \in]0, 1[$. Construire un test de H_0 contre H_1 uniformément le plus puissant parmi les tests de niveau α.
4. Exprimer le test précédent en fonction de l'estimateur du maximum de vraisemblance $\hat{\theta}$ pour l'estimation de θ.
5. Un vendeur affirme que la durée moyenne des batteries des téléphones qu'il commercialise est au moins égale à 4 ans. On modélise la durée de vie (en années) des batteries des téléphones qu'il vend à l'aide de la loi associée à la densité f_θ.
   (a) Montrer que tester l'affirmation du vendeur revient à réaliser le test précédent pour la valeur $\theta_0 = 2$.
   (b) On observe $n = 15$ et $\sum_{i=1}^{n} \log(X_i) = 11$. Déterminer la p-valeur associée à ces observations. On justifiera soigneusement son calcul.
   (c) Quelle est la conclusion du test pour $\alpha = 5\%$?

Exercice 2 : Sur l'espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$, on observe 2 échantillons indépendants $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, Y_m)$ de tailles respectives n et m. On suppose que les variables X_i et Y_ℓ sont à valeurs dans l'ensemble discret $\mathcal{X} = \{x_1, \ldots, x_d\}$. On note pour tout $j \in \mathcal{X}$,

$$p_j = \mathbb{P}(X_1 = x_j), \quad q_j = \mathbb{P}(Y_1 = x_j),$$

et on suppose que pour tout $j \in \mathcal{X}$, $p_j > 0$ et $q_j > 0$. On veut tester que les variables X_i ont même distribution que les variables Y_ℓ :

$$H_0 : p = q \quad \text{vs} \quad H_1 : p \neq q,$$

où on a noté $p = (p_1, \ldots, p_d)$ et $q = (q_1, \ldots, q_d)$.

1. On considère les $n + m$ couples d'observations $(U_1, V_1), \ldots, (U_{n+m}, V_{n+m})$ définis pour $k \in \{1, \ldots, n + m\}$ par

$$(U_k, V_k) = \begin{cases} (X_k, 0) & \text{si} \quad 1 \leq k \leq n \\ (Y_{k-n}, 1) & \text{si} \quad n + 1 \leq k \leq n + m \end{cases}$$

   On considère une permutation tirée aléatoirement des vecteurs $(U_k, V_k)_{1 \leq k \leq n+m}$. On note $(T_k, W_k)_{1 \leq k \leq n+m}$ les $n + m$ vecteurs obtenus. On admet que ceux-ci sont i.i.d. Montrer que notre test d'hypothèses est équivalent à tester l'indépendance entre les T_k et les W_k.

2. A l'aide du test d"indépendance vu en cours, démontrer que le test mettra en jeu la statistique

$$D_n^2(X, Y) := \sum_{j=1}^{d} \left(\frac{(N_{j,n} - n\hat{p}_{j,n,m})^2}{n\hat{p}_{j,n,m}} + \frac{(M_{j,m} - m\hat{p}_{j,n,m})^2}{m\hat{p}_{j,n,m}} \right),$$

   où

$$N_{j,n} := \sum_{i=1}^{n} 1_{\{X_i = x_j\}}, \quad M_{j,m} := \sum_{\ell=1}^{m} 1_{\{Y_\ell = x_j\}}, \quad \hat{p}_{j,n,m} := \frac{N_{j,n} + M_{j,m}}{n + m}.$$

3. En déduire que lorsque $n + m$ tend vers $+\infty$,
   - sous H_0, $D_n^2(X, Y)$ converge en loi vers la loi $\chi^2(d - 1)$
   - sous H_1, $D_n^2(X, Y)$ tend vers $+\infty$ ps.

4. Pour $\alpha \in]0, 1[$, on considère le test

$$\phi(X, Y) = 1_{\{D_n^2(X,Y) > c_{d-1,1-\alpha}\}},$$

   où $c_{d-1,1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi $\chi^2(d - 1)$. Quelles sont les propriétés de ce test ?

5. On considère, pour une journée donnée, le nombre d'appels téléphoniques passés par deux groupes d'étudiants. Le tableau suivant donne le nombre de personnes pour chaque groupe (A et B) qui ont passé un nombre fixé d'appels téléphoniques.

nombre d'appels	0	1	2	3	4	5	6	7	≥ 8
Groupe A	85	138	104	49	15	8	0	1	0
Groupe B	112	109	92	55	8	4	0	0	0

Tester si la distribution des appels téléphoniques est la même pour les deux groupes d'étudiants.

6. Tester si la distribution des appels téléphoniques passés par les étudiants du groupe A suit une loi de Poisson.

C.3 Partiel 2020-2021

Statistique mathématique

PARTIEL – 11 mars 2021 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Exercice 1 : A l'aide d'un thermomètre, on réalise n mesures indépendantes de la température d'un liquide maintenu à température constante égale à μ. On considère ces mesures comme des réalisations de variables aléatoires $(X_1, \dots, X_n)$, de même loi normale $\mathcal{N}(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$ sont des paramètres inconnus. Construire un intervalle de confiance bilatère pour μ au niveau de confiance $1 - \alpha$ pour $\alpha \in]0; 1[$. Les différentes étapes de la construction sont à justifier rigoureusement.

Exercice 2 : On observe $(X_1, \dots, X_n)$ un n–échantillon de loi exponentielle de paramètre $\theta \in \mathbb{R}_+^*$. On rappelle que la densité de chacune des variables X_i est donnée par la fonction

$$f_\theta(x) = \theta \exp(-\theta x) 1_{]0;+\infty[}(x), \quad x \in \mathbb{R}.$$

1. Montrer que $\mathbb{E}_\theta[X_1] = \theta^{-1}$ et $\mathrm{var}_\theta[X_1] = \theta^{-2}$.
2. A l'aide de la méthode des moments, donner un estimateur sans biais de $g_1(\theta) = \theta^{-1}$.
3. A l'aide de la méthode des moments, donner deux estimateurs fortement consistants de $g_2(\theta) = \theta^{-2}$.
4. Montrer que la fonction génératrice des moments de X_1 satisfait pour tout $t \in]-\infty; \theta[$,
$$\mathbb{E}_\theta\left[e^{tX_1}\right] = \frac{\theta}{\theta - t}.$$
5. On pose $S_n = \sum_{i=1}^n X_i$. Montrer que $\theta \times S_n$ suit la loi Gamma de paramètres $(n, 1)$. On rappelle que si Y suit la loi Gamma de paramètres (k, λ) alors sa fonction génératrice des moments satisfait pour tout $t \in]-\infty; \lambda^{-1}[$,
$$\mathbb{E}\left[e^{tY}\right] = (1 - \lambda t)^{-k}.$$
6. On revient à présent à l'estimation de $g_1(\theta) = \theta^{-1}$. On s'appuie sur la question précédente. On suppose que $n = 25$ et on donne les quantiles suivants de la loi Gamma de paramètres $(25, 1)$:

Ordre	1%	2,5%	5%	10%	50%	90%	95%	97,5%	99%
Quantile	14.9	16.2	17.4	18.9	24.7	31.6	33.8	35.7	38.1

(a) On souhaite déterminer une fonction des observations qui soit, avec probabilité égale à 95%, une borne inférieure pour $g_1(\theta)$. Construire la région de confiance unilatère associée à ce problème. Réaliser l'application numérique lorsque $S_{25} = 50$.

(b) Déterminer un intervalle de confiance bilatère pour l'estimation de $g_1(\theta)$ de niveau de confiance exactement égal à 95%. Réaliser l'application numérique lorsque $S_{25} = 50$.

Exercice 3 : Un joueur s'amuse à lancer un dé à 6 faces. La probabilité de chaque chiffre est non-nulle. On note θ la probabilité inconnue d'obtenir un 6 et X le nombre de lancers nécessaires pour l'obtention d'un 6. On note $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in]0;1[)$ le modèle statistique associé à l'observation X.

1. Démontrer que pour tout $k \in \mathbb{N}^*$,
$$\mathbb{P}_\theta(X = k) = \theta(1-\theta)^{k-1}.$$

2. On note F_X la fonction de répartition associée à X.
   (a) Soit $k \in \mathbb{N}^*$. Montrer que la fonction de répartition de X en k vaut
   $$F_X(k) = 1 - (1-\theta)^k.$$
   (b) En déduire la valeur de $F_X(t)$ pour tout $t \in \mathbb{R}$.
   (c) Soit $\beta \in]0;1[$. Calculer alors le quantile de F_X d'ordre β uniquement en fonction de β et θ.

3. On admettra que
$$\mathbb{E}_\theta[X] = \frac{1}{\theta} \quad \text{et} \quad \text{var}_\theta[X] = \frac{1-\theta}{\theta^2}.$$
On observe $X_1, \ldots, X_n$ n variables aléatoires indépendantes de même loi que X. On souhaite estimer θ à l'aide des variables $X_1, \ldots, X_n$.
   (a) Déterminer un estimateur de θ à l'aide de la méthode des moments.
   (b) Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance.
   (c) Etablir que $\widehat{\theta}_n$ est asymptotiquement normal.
   (d) Montrer que lorsque $n \to +\infty$, on a la convergence en loi suivante :
   $$\frac{\sqrt{n}(\overline{X}_n - \theta^{-1})}{\sqrt{\overline{X}_n(\overline{X}_n - 1)}} \rightsquigarrow \mathcal{N}(0,1), \quad \text{où on a posé} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

4. On suppose à présent que le dé est déformé et que le chiffre 6 est le plus probable de tous les chiffres. La probabilité de chaque chiffre reste non-nulle. On souhaite garantir, avec grande probabilité, une borne inférieure et supérieure pour θ dépendant des observations. Pour cela, on fixe $\alpha \in]0;1[$. Déterminer une région de confiance non-asymptotique de niveau de confiance $1-\alpha$ qui réponde au problème.

   Indication : appliquer l'inégalité de Bienaymé-Tchebychev à $\overline{X}_n$.

Exercice 4 : On observe un n-échantillon $X = (X_1, \dots, X_n)$ de loi de Poisson de paramètre $\theta \in \mathbb{R}_+^*$ inconnu. On a donc pour tout $i \in \{1, \dots, n\}$ et pour tout $k \in \mathbb{N}$,

$$\mathbb{P}_\theta(X_i = k) = \frac{e^{-\theta}\theta^k}{k!}.$$

1. Démontrer que $\mathbb{E}_\theta[X_1] = \theta$.
2. Calculer $\mathrm{cov}_\theta(X_1, X_1 - 1)$, la covariance de X_1 et $X_1 - 1$, en fonction de θ et proposer un estimateur fortement consistant de cette quantité.
3. Calculer la valeur de la variance de X_1 à l'aide de la question précédente.
4. On pose pour $\theta \in \mathbb{R}_+^*$,
$$g(\theta) = \sqrt{\theta}.$$
   Exhiber un estimateur issu de la méthode des moments asymptotiquement normal pour l'estimation de $g(\theta)$. Déterminer précisément la variance de la loi limite.

   Indication : On utilisera la méthode delta.

C.4 Examen 2020-2021

Statistique mathématique

EXAMEN – 20 mai 2021 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Exercice 1 : Un enseignant souhaite que ses élèves se familiarisent à la notion de hasard et leur demande, en guise de devoirs, de lancer 200 fois une pièce bien équilibrée pour le lendemain et de noter les résultats. Il relève donc, pour chaque élève, une suite $x = (x_1, \dots, x_{200})$ de 0 (pile) et de 1 (face). Il souhaite identifier les élèves qui ne lancent pas honnêtement leur pièce, et écrivent des 0 ou 1 selon un certain schéma qui n'est pas du hasard pur. Pour cela, il s'appuie sur le résultat suivant : si $X = (X_1, \dots, X_{200})$ est une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi de Bernoulli de paramètre 1/2, la probabilité que X contienne une sous-suite formée de six 0 ou six 1 à la suite est de 97 %.

Proposer un test $\phi(X)$ de H_0 : « *L'élève a honnêtement effectué ses lancers* », contre l'hypothèse H_1 : « *L'élève a inventé les résultats.* » Spécifier la taille du test $\phi(X)$ mais pas l'erreur de seconde espèce.

Exercice 2 : Une usine fabrique des machines. On suppose que le temps aléatoire X au bout duquel une machine tombe en panne suit une loi exponentielle $\varepsilon(\lambda)$ de paramètre d'intensité $\lambda \in \mathbb{R}_+^*$ inconnu, *i.e.*, la densité de X est définie pour $x \in \mathbb{R}$ par

$$f_\lambda(x) = \lambda \exp(-\lambda x) 1_{x \in \mathbb{R}_+^*}.$$

Les machines sont vendues à une grande surface avec laquelle le fabricant passe un contrat sur la qualité des machines fournies. Sur ce contrat on veut que la probabilité pour qu'une machine tombe en panne avant un instant donnée T soit plus petite qu'une quantité donnée. On dit que la convention est respectée si $\lambda \leq \lambda_0$ pour $\lambda_0 \in \mathbb{R}_+^*$. Pour s'en assurer, le fournisseur lance régulièrement des études dans lesquelles il observe au hasard n machines. On note $X_1, \dots, X_n$ le n-échantillon suivant la loi $\varepsilon(\lambda)$ correspondant aux temps de panne observés.

Partie I. On considère le test d'hypothèses suivant

$$H_0 : \lambda \leq \lambda_0 \quad \text{contre} \quad H_1 : \lambda > \lambda_0.$$

1. Justifier que ce choix d'hypothèses est cohérent avec le point de vue du fabricant.

2. Construire un test uniformément plus puissant de taille $\alpha \in]0;1[$.
   Indication : On pourra utiliser sans démonstration que si $X_1, \dots, X_n$ est un n-échantillon suivant la loi $\varepsilon(\lambda)$ alors $2\lambda \sum_{i=1}^n X_i$ suit une loi du $\chi^2(2n)$.
3. Déterminer l'erreur de seconde espèce et la fonction puissance de ce test de taille α.
4. Donner l'expression explicite de $\widehat{\alpha}(X)$, la p-valeur du test, en fonction de λ_0 et $\sum_{i=1}^n X_i$.
5. Pour une p-valeur de 0.86, donner la conclusion du test ainsi que l'erreur de seconde espèce pour les tailles de test suivantes : 1%, 5%, 10%.

Partie II. On considère à présent le test d'hypothèses suivant

$$H_0 \,:\, \lambda = \lambda_0 \quad \text{contre} \quad H_1 \,:\, \lambda \neq \lambda_0.$$

1. Déterminer l'expression de l'estimateur du maximum de vraisemblance $\widehat{\lambda}_n$ de λ.
2. Proposer le test de Wald de niveau asymptotique $\alpha \in]0;1[$ pour tester H_0 contre H_1.

Exercice 3 : On observe un échantillon de n personnes atteintes d'une maladie particulière. Pour chacune de ces personnes, on relève son groupe sanguin. On introduit donc les variables $X_1, \dots, X_n$ i.i.d. à valeurs dans l'ensemble $\Omega := \{A, B, AB, O\}$ et le vecteur de probabilités

$$p := \big(p_A, p_B, p_{AB}, p_O\big)$$

avec

$$p_A = \mathbb{P}(X_1 = A), \quad p_B := \mathbb{P}(X_1 = B), \quad p_{AB} := \mathbb{P}(X_1 = AB), \quad p_O = \mathbb{P}(X_1 = O).$$

On suppose dans la suite que le vecteur p inconnu a toutes ses composantes strictement positives.

1. On souhaite estimer le vecteur p. Pour cela, on introduit les variables aléatoires

$$N_A := \sum_{i=1}^n 1_{\{X_i = A\}},\ N_B := \sum_{i=1}^n 1_{\{X_i = B\}},\ N_{AB} := \sum_{i=1}^n 1_{\{X_i = AB\}},\ N_O := \sum_{i=1}^n 1_{\{X_i = O\}}$$

   et le vecteur N défini par

$$N = \big(N_A, N_B, N_{AB}, N_O\big).$$

   (a) On suppose $n \geq 7$. Calculer la probabilité

$$\mathbb{P}\Big((N_A, N_B, N_{AB}, N_O) = (7, 0, 0, n-7)\Big)$$

   en fonction de n et des composantes de p.

(b) Donner pour tout 4-uplet $k = (k_A, k_B, k_{AB}, k_O) \in \mathbb{N}^4$, la probabilité

$$\mathbb{P}\Big((N_A, N_B, N_{AB}, N_O) = (k_A, k_B, k_{AB}, k_O)\Big)$$

en fonction de n, des composantes de p et de celles de k.

(c) Proposer un estimateur fortement consistant et sans biais de p en vous appuyant sur la méthode des moments. Vous justifierez soigneusement votre réponse.

(d) En justifiant soigneusement vos calculs, démontrer que l'estimateur du maximum de vraisemblance pour l'estimation de p est donné par le vecteur

$$\widehat{p}_n := \Big(\frac{N_A}{n}, \frac{N_B}{n}, \frac{N_{AB}}{n}, \frac{N_O}{n}\Big).$$

Indication : On considèrera tout d'abord le cas où toutes les composantes de N sont non-nulles puis on traitera le cas où au moins une composante de N est nulle.

(e) Prouver que le vecteur $\widehat{p}_n$ est asymptotiquement normal et préciser la matrice de variance-covariance de la loi limite.

2. On considère à présent la construction de régions de confiance pour le vecteur p.

(a) Pour $j \in \Omega$, et $\alpha \in]0;1[$, construire un intervalle de confiance asymptotique de niveau exactement $1 - \alpha/4$ pour l'estimation de p_j. En déduire une région de confiance asymptotique de niveau $1 - \alpha$ pour l'estimation du vecteur p.

(b) En utilisant directement les résultats du cours, montrer que la variable

$$D_n^2(\widehat{p}_n, p) := n \sum_{j\in\Omega} \frac{\Big(\frac{N_j}{n} - p_j\Big)^2}{p_j}$$

converge en loi lorsque n tend vers $+\infty$ et donner la loi limite.

(c) Proposer un vecteur aléatoire observable $\widehat{w} = (\widehat{w}_A, \widehat{w}_B, \widehat{w}_{AB}, \widehat{w}_O)$ de sorte que la variable

$$\widetilde{D}_n^2(\widehat{p}_n, p) := n \sum_{j\in\Omega} \widehat{w}_j \Big(\frac{N_j}{n} - p_j\Big)^2$$

converge en loi lorsque n tend vers $+\infty$ et donner la loi limite.

(d) Déduire de la question précédente une région de confiance asymptotique de niveau exactement $1 - \alpha$ pour l'estimation du vecteur p.

3. On considère à présent la problématique du test. Dans la population, la proportion théorique de chaque groupe sanguin est la suivante :

Type	A	B	AB	O
Proportion théorique	45%	9%	4%	42%

On note
$$p^0 := \Big(0.45,\ 0.09,\ 0.04,\ 0.42\Big).$$
A l'aide de ces données, on souhaite tester si l'échantillon de personnes malades est conforme à la population.

(a) Formuler le problème de test et écrire le test du χ^2 qui répond à la question.

(b) On observe $n = 200$ et on obtient
$$D_n^2(\widehat{p}_n, p^0) = 7.6.$$
A l'aide des tables du χ^2 données ci-dessous, donner une valeur approchée de la p-valeur du test. Au risque $\alpha = 5\%$ quelle décision prend-on ? On justifiera soigneusement les réponses à ces questions.

ddl	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

TABLE C.1 – Quantiles de la loi du χ^2 en fonction du nombre de degré de libertés (ddl)

ddl	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

TABLE C.2 – Quantiles de la loi du χ^2 en fonction du nombre de degré de libertés (ddl)

C.5 Partiel 2021-2022

Statistique mathématique

PARTIEL – 10 mars 2022 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Exercice 1 : Soit $(X_1, \ldots, X_n)$ un n-échantillon de la loi de paramètre inconnu $(\lambda, b) \in \mathbb{R}_+^* \times \mathbb{R}$ et de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$f(x; \lambda, b) = \lambda \exp\left[\lambda(x - b)\right] 1_{x \leq b}.$$

1. Déterminer, s'il existe, l'estimateur du maximum de vraisemblance de (λ, b).
2. Obtient-on un estimateur sans biais de λ ?

Exercice 2 : Soit $(X_1, \ldots, X_n)$ un n-échantillon de la loi de paramètre $p \in]0, 1[$ et de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ donnée par

$$f(x; p) = \frac{p}{a} 1_{x \in [0,a]} + \frac{1-p}{b} 1_{x \in [0,b]}, \quad \text{avec} \quad 0 < a < b \text{ connus.}$$

On note A_n le nombre d'observations dans l'intervalle $[0, a]$ et on admet que A_n suit une loi binomiale de paramètre (n, λ).

1. Déterminer un estimateur de p par la méthode des moments, que l'on note $\widehat{p}_n$.
2. Déterminer l'expression de λ en fonction de p, a et b.
3. Montrer que l'estimateur du maximum de vraisemblance de p est
$$\widehat{p}_{\text{EMV}} = \frac{bA_n - na}{n(b-a)}.$$
4. (a) Montrer que
$$\frac{1}{\sqrt{ng\left(\widehat{p}_n\right)\left[1 - g\left(\widehat{p}_n\right)\right]}} \left[A_n - ng\left(p\right)\right] \xrightarrow[n \to +\infty]{d} \mathcal{N}(0, 1), \quad \text{avec} \quad g : x \mapsto x + \frac{(1-x)a}{b}.$$
   (b) Obtiendrait-on un résultat différent en remplaçant $\widehat{p}_n$ par $\widehat{p}_{\text{EMV}}$? Justifier brièvement votre réponse.

(c) En déduire un intervalle de confiance asymptotique bilatère pour p de niveau $1-\alpha$, $\alpha \in]0,1[$.

Exercice 3 : Soit $(X_1,\dots,X_n)$ un n-échantillon de la loi de Cauchy de paramètre $\beta \in \mathbb{R}$, notée $\mathcal{C}(\beta)$, dont la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est donnée par

$$f_\beta(x) = \frac{1}{\pi[1+(x-\beta)^2]}.$$

On note F la fonction de répartition de la loi $\mathcal{C}(\beta)$ et $\widehat{F}_n$ la fonction de répartition empirique associée à $(X_1,\dots,X_n)$: pour tout $x \in \mathbb{R}$

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{k=1}^n 1_{X_k \leq x}.$$

On rappelle que pour $\alpha \in]0,1[$,

$$\widehat{F}_n^{(-1)}(\alpha) = \inf\left\{x \in \mathbb{R} \mid \widehat{F}_n(x) \geq \alpha\right\} \quad \text{et} \quad F^{(-1)}(\alpha) = \inf\left\{x \in \mathbb{R} \mid F(x) \geq \alpha\right\}.$$

1. On rappelle que la fonction caractéristique de X_1 est donnée pour tout $t \in \mathbb{R}$ par

   $$\Phi_{X_1}(t) = \exp\left(i\beta t - |t|\right).$$

   Donner la fonction caractéristique de $(X_1,\dots,X_n)$.
2. Déterminer la limite en loi de $\frac{1}{n}\sum_{k=1}^n X_k$.
3. Montrer que la médiane de la loi $\mathcal{C}(\beta)$ est β.
4. Montrer que pour tout $x \in \mathbb{R}$, $\widehat{F}_n(x)$ est un estimateur fortement consistant de $F(x)$.
5. (a) Soit $\delta > 0$. Montrer qu'il existe un rang N tel que pour tout $n \geq N$, on a presque-sûrement

   $$\widehat{F}_n\left[F^{(-1)}(\alpha)-\delta\right] < \alpha \quad \text{et} \quad \widehat{F}_n\left[F^{(-1)}(\alpha)+\delta\right] \geq \alpha.$$

   (b) En déduire que pour tout $n \geq N$, on a presque sûrement

   $$F^{(-1)}(\alpha)-\delta < \widehat{F}_n^{(-1)}(\alpha) \leq F^{(-1)}(\alpha)+\delta.$$

   (c) Montrer que par conséquent, on a presque sûrement

   $$\liminf_{n\to+\infty} \widehat{F}_n^{(-1)}(\alpha) \geq F^{(-1)}(\alpha) \quad \text{et} \quad \limsup_{n\to+\infty} \widehat{F}_n^{(-1)}(\alpha) \leq F^{(-1)}(\alpha).$$

   (d) En utilisant les questions précédentes, montrer que pour tout $\alpha \in]0,1[$, $\widehat{F}_n^{(-1)}(\alpha)$ est un estimateur fortement consistant de $F^{(-1)}(\alpha)$.

   (e) Déduire de ce résultat un estimateur fortement consistant de β.

C.6 Examen 2021-2022

Statistique mathématique

EXAMEN – 19 mai 2022 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Exercice 1 : On dispose du cumul des précipitations du mois de mars 2022 pour $n = 1342$ stations météorologiques situées en France métropolitaine et dans les territoires d'Outre-Mer. Les mesures sont résumées de la façon suivante :

Cumul de précipitations (en cm)	C_1	C_2	C_3	C_4	C_5	C_6	C_7
	$[0,4[$	$[4,6[$	$[6,8[$	$[8,10[$	$[10,12[$	$[12,14[$	$[14,+\infty[$
Effectifs	966	192	93	38	23	15	15

On note $X = (X_1, \dots, X_n)$ le n-échantillon correspondant au cumul des pécipitations et on définit les variables catégorielles $Y_1, \dots, Y_n$ indépendantes et identiquement distribuées telles que pour $i \in \{1, \dots, n\}$ et $k \in \{1, \dots, 7\}$,

$$\mathbb{P}[Y_i = k] = \mathbb{P}[X_i \in C_k] =: p_k.$$

On souhaite tester si la quantité de précipitation est distribuée suivant une loi exponentielle de paramètre $\lambda = 0.3$. On rappelle que la loi exponentielle de paramètre λ admet pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ la fonction donnée par

$$f_\lambda(x) = \lambda \exp(-\lambda x) 1_{\mathbb{R}_+^*}(x). \tag{C.1}$$

Cela revient à construire un test qui vérifie si nos observations ont pour loi commune la loi de Y_1.

1. Montrer que pour $k \in \{2, \dots, 6\}$

   $$p_1 = 1 - \exp(-4\lambda), \quad p_k = \exp\left(-\lambda \min C_k\right)\{1 - \exp(-2\lambda)\} \quad \text{et} \quad p_7 = \exp(-14\lambda).$$

2. Rappeler les hypothèses du test considéré.
3. Donner l'expression, en fonction des données du problème, de la statistique de test $T(X)$ permettant de réaliser le test d'ajustement souhaité.
4. Pour les données pluviométriques du mois de mars 2022, on obtient $T(x) = 10.22$. À l'aide de la table du χ^2, donner la p-valeur associée à ce test. Quelle est la conclusion du test pour un test asymptotiquement de taille $\alpha = 5\%$? $\alpha = 20\%$?

Exercice 2 : Soit X la variable aléatoire qui modélise la quantité de pluie (en centimètres) tombée pendant une journée. On introduit alors Z une variable de Bernoulli de paramètre $p \in]0;1[$ et Y une variable exponentielle de paramètre $\lambda > 0$ indépendante de Z dont la densité est donnée en (C.1). On pose

$$X = YZ$$

1. Montrer que $\mathbb{E}[X] = \frac{p}{\lambda}$.
2. Montrer que
$$X = \begin{cases} Y & \text{si} \quad Z = 1, \\ 0 & \text{si} \quad Z = 0. \end{cases} \tag{C.2}$$
Donner une interprétation de l'écriture (C.2) et des paramètres p et λ.
3. Déterminer l'ensemble des valeurs prises par X. Démontrer que X n'est ni discrète, ni absolument continue.

   On effectue des relevés pluviométriques durant n jours que l'on suppose suffisamment espacés dans le temps pour que les n mesures, notées $X_1, \dots, X_n$, puissent être considérées comme indépendantes et identiquement distribuées et de même lois que X.
4. On souhaite estimer les paramètres p et λ à l'aide de $X_1, \dots, X_n$.
   (a) Démontrer que pour tout $i \in \{1, \dots, n\}$,
   $$p = 1 - \mathbb{P}(X_i = 0).$$
   (b) En déduire un estimateur $\widetilde{p}_n$ de p grâce à la méthode des moments et qui soit asymptotiquement normal.
   (c) Donner un estimateur fortement consistant du paramètre λ qui soit fonction uniquement des X_i.
5. On suppose toujours disposer du même schéma d'observations $X_1, \dots, X_n$ que précédemment. On pose
$$\overline{N}_n = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i > 0\}}, \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$
   (a) Démontrer que la vraisemblance du modèle s'écrit
   $$V_{X_1,\dots,X_n}(p,\lambda) = (1-p)^{n(1-\overline{N}_n)}(p\lambda)^{n\overline{N}_n} \exp\Big(-\lambda \sum_{i=1}^n X_i\Big), \quad (p,\lambda) \in]0;1[\times \mathbb{R}_+^*.$$
   (b) Calculer soigneusement $(\widehat{p}_n, \widehat{\lambda}_n)$ l'estimateur du maximum de vraisemblance du vecteur (p, λ) en fonction de $\overline{N}_n$ et $\overline{X}_n$.

(c) Démontrer que

$$\operatorname{var}(1_{\{X_1>0\}}) = p(1-p), \quad \operatorname{var}(X_1) = \frac{p(2-p)}{\lambda^2}, \quad \operatorname{cov}(1_{\{X_1>0\}}, X_1) = \frac{p(1-p)}{\lambda}.$$

(d) En déduire que

$$\sqrt{n}\Big(\widehat{\lambda}_n - \lambda\Big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \frac{\lambda^2}{p}\right) \quad \text{en loi.} \tag{C.3}$$

(e) Du résultat précédent, bâtir un intervalle de confiance asymptotique de niveau de confiance 95% dépendant uniquement des observations X_i.

6. On suppose dans cette question que $p = 1$. Soit $\lambda_0 > 0$ et $\alpha \in]0;1[$. On considère le test d'hypothèses suivant

$$H_0 \,:\, \lambda = \lambda_0 \quad \text{contre} \quad H_1 \,:\, \lambda \neq \lambda_0.$$

A l'aide de l'estimateur $\widehat{\lambda}_n$ et de la convergence (C.3) (qui reste vraie lorsque $p = 1$), proposer le test de Wald de niveau asymptotique α permettant de tester les hypothèses H_0 et H_1.

7. On pose pour tout $i \in \{1, \dots, n\}$, $Y_i = 1_{\{X_i>0\}}$.

(a) Soient $p_0 \in]0;1[$ et $\alpha \in]0;1[$. On souhaite réaliser le test d'hypothèses suivant

$$H_0 \,:\, p = p_0 \quad \text{contre} \quad H_1 \,:\, p > p_0.$$

Quelle est la loi de $\sum_{i=1}^n Y_i$ lorsque H_0 est vraie ?

(b) Construire alors un test de niveau α qui réponde à la question posée.

(c) Soient $\alpha \in]0;1[$ et $(p_0, p_1) \in]0;1[^2$ avec $p_1 > p_0$. On considère le test d'hypothèses suivant

$$H_0 \,:\, p = p_0 \quad \text{contre} \quad H_1 \,:\, p = p_1.$$

Donner la forme du test de Neyman-Pearson fondé uniquement sur l'observation des Y_i. A quelle condition est-il UPP(α) ?

ddl	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

TABLE C.3 – Quantiles de la loi du χ^2 en fonction du nombre de degré de libertés (ddl)

ddl	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

TABLE C.4 – Quantiles de la loi du χ^2 en fonction du nombre de degré de libertés (ddl)

C.7 Partiel 2022-2023

Statistique mathématique

PARTIEL – 14 mars 2023 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Question de cours : Enoncer le théorème de convergence fourni par la méthode delta. Appliquer alors la méthode delta pour montrer que si on observe $(X_1, \dots, X_n)$ un n-échantillon de loi exponentielle de paramètre $\theta \in \mathbb{R}_+^*$ de densité f_θ définie par

$$f_\theta(x) = \theta e^{-\theta x} 1_{\mathbb{R}_+}(x), \quad x \in \mathbb{R}$$

alors, avec

$$T_n = \sqrt{\frac{2}{\frac{1}{n}\sum_{i=1}^n X_i^2}}$$

on a la convergence en loi suivante :

$$\sqrt{n}\big(T_n - \theta\big) \overset{n \to +\infty}{\rightsquigarrow} \mathcal{N}(0, 5\theta^2/4).$$

<u>Indication :</u> On rappelle que pour tout $k \in \mathbb{N}^*$, $\mathbb{E}_\theta[X_1^k] = k!\,\theta^{-k}$.

Exercice 1 : On note h et ϕ les fonctions définies sur $\mathbb{R}_+$ par

$$h(x) = (1+x)\log(1+x) - x, \quad \phi(x) = \exp(x) - 1 - x.$$

Soit X une variable aléatoire suivant une loi de Poisson de paramètre $\theta \in \mathbb{R}_+^*$.

1. Pour tout $\lambda \in \mathbb{R}$, exprimer $\mathbb{E}[\exp(\lambda(X-\theta))]$ à l'aide de la fonction ϕ.
2. Montrer que pour tout $\lambda \in \mathbb{R}_+$ et pour tout $x > 0$,

$$\mathbb{P}(X - \theta \geq x) \leq \exp(\theta\phi(\lambda) - \lambda x).$$

3. Soit $\mu > 0$. Calculer $\inf_{\lambda \in \mathbb{R}_+}\{\phi(\lambda) - \mu\lambda\}$.
4. En déduire que pour tout $x > 0$,

$$\mathbb{P}(X - \theta \geq x) \leq \exp(-\theta h(x/\theta)).$$

5. Lorsque $\theta = 1$ et $x = 4$, on obtient $\exp(-h(4)) = 0.017$. Comparer à la majoration de $\mathbb{P}(X - \theta \geq x)$ obtenue à l'aide de l'inégalité de Bienaymé-Tchebychev.

Exercice 2 : Soit X une variable aléatoire de loi $\mathbb{P}_\theta$ admettant pour densité, par rapport à la mesure de Lebesgue, la fonction f_θ définie pour $x \in \mathbb{R}$ par

$$f_\theta(x) = \theta^{-1} \exp\big(-\theta^{-1}(x-\theta)\big) 1_{[\theta;+\infty[}(x),$$

où $\theta \in \mathbb{R}_+^*$ est inconnu. On observe $(X_1, \ldots, X_n)$ un n-échantillon de même loi que X.

1. Calculer $\mathbb{E}_\theta[X]$ et $\mathbb{E}_\theta[X^2]$, les deux premiers moments de X.
2. On considère la méthode des moments.
   (a) Déterminer $\widetilde{\theta}_n^{(1)}$, un estimateur sans biais de θ, par la méthode des moments.
   (b) Montrer que $\widetilde{\theta}_n^{(1)}$ est asymptotiquement normal. Donner sa variance asymptotique en fonction de θ.
   (c) Soit $\alpha \in]0;1[$. A l'aide de $\widetilde{\theta}_n^{(1)}$, proposer un intervalle de confiance asymptotique de niveau $1-\alpha$.
   (d) Proposer $\widetilde{\theta}_n^{(2)}$, un second estimateur estimateur de θ par la méthode des moments, biaisé ou non, asymptotiquement normal (on ne calculera pas la variance asymptotique).
3. On considère la méthode du maximum de vraisemblance.
   (a) Calculer la vraisemblance du modèle.
   (b) Montrer soigneusement que l'estimateur du maximum de vraisemblance, noté $\widehat{\theta}_n$, vaut
   $$\widehat{\theta}_n = \min_{1 \leq i \leq n} X_i.$$
   (c) Calculer pour tout $t \in \mathbb{R}$,
   $$\mathbb{P}_\theta\big(\widehat{\theta}_n > t\big).$$
   (d) En déduire que $\widehat{\theta}_n$ est consistant.
   (e) Calculer le biais de l'estimateur $\widehat{\theta}_n$.
   Indication : On utilisera la formule suivante valable pour toute variable aléatoire positive Y :
   $$\mathbb{E}[Y] = \int_0^{+\infty} \mathbb{P}(Y > t)\,dt.$$

C.8 Examen 2022-2023

Statistique mathématique

EXAMEN – 24 mai 2023 – Durée : 2h00

Documents, calculatrices et smartphones interdits.

On attachera un grand soin à la rigueur mathématique et à la qualité de la rédaction

Exercice 1 Pour $n \geq 1$, on considère un n-échantillon $X = (X_1, \ldots, X_n)$ de loi exponentielle de paramètre $\theta^{-1} \in \mathbb{R}_+^*$ dont la densité sur $\mathbb{R}_+^*$ est :

$$f_\theta(x) = \theta^{-1} \exp(-\theta^{-1} x), \quad x \in \mathbb{R}_+^*.$$

On note $V_X(\theta)$ la vraisemblance du modèle au point $\theta \in \mathbb{R}_+^*$.

1. Donner l'expression de $V_X(\theta)$.
2. On cherche à estimer θ.
   (a) Calculer l'estimateur du maximum de vraisemblance, noté $\widehat{\theta}_n$.
   (b) Montrer que $\widehat{\theta}_n$ est asymptotiquement normal.
   (c) Pour $\alpha \in]0; 1[$, en déduire un intervalle de confiance asymptotique bilatère de niveau de confiance égal à $1 - \alpha$ pour l'estimation de θ.

<u>Dans toute la suite, on suppose que $n = 1$.</u>

3. Soient θ_1 et θ_2 deux réels strictement positifs avec $\theta_1 < \theta_2$.
   (a) Démontrer que pour tout $k > 0$, il existe $c \in \mathbb{R}$ tel que

   $$\frac{V_X(\theta_2)}{V_X(\theta_1)} > k \iff X_1 > c.$$

   Donner l'expression de c en fonction de k.
   (b) Soit $\alpha \in]0; 1[$. Calculer $c_\alpha \in \mathbb{R}$ de sorte que le test $\phi_\alpha(X) = 1_{\{X_1 > c_\alpha\}}$ soit de taille α pour tester

   $$H_0 : \theta = \theta_1 \quad \text{contre} \quad H_1 : \theta = \theta_2.$$

   Etablir que c_α ne dépend pas de θ_2.
   (c) Démontrer que $\phi_\alpha(X)$ est UPP(α) pour le test précédent.
4. On considère toujours $\theta_1 \in \mathbb{R}_+^*$ et le test $\phi_\alpha(X)$ de la question 3b).

(a) Soit $\theta_0 < \theta_1$. On note
$$\alpha_0 = \mathbb{E}_{\theta_0}[\phi_\alpha(X)].$$
Montrer que $\phi_\alpha(X)$ est UPP(α_0) pour tester
$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta = \theta_1,$$

(b) En déduire que $\alpha_0 \leq \alpha$.

(c) A l'aide des questions précédentes, conclure que $\phi_\alpha(X)$ est UPP(α) pour tester
$$H_0 : \theta \leq \theta_1 \quad \text{contre} \quad H_1 : \theta > \theta_1. \tag{C.4}$$

(d) Utiliser un théorème du cours pour justifier à nouveau que $\phi_\alpha(X)$ est UPP(α) pour (C.4).

(e) Calculer la p-valeur $\widehat{\alpha}(X_1)$ du test $\phi_\alpha(X)$ pour le problème (C.4).

Exercice 2 Pour $n \geq 1$, on considère un n-échantillon $X = (X_1, \ldots, X_n)$ de loi Beta de paramètres $p > 0$ et $q > 0$ dont la densité sur $]0;1[$ est :
$$f_{p,q}(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \quad x \in]0;1[,$$
avec pour $u > 0$,
$$\Gamma(u) = \int_0^{+\infty} x^{u-1} \exp(-x) dx.$$
Dans cet exercice, on admet que
$$\mathbb{E}[X_1] = \frac{p}{p+q} \quad \text{et} \quad \text{var}(X_1) = \frac{pq}{(p+q)^2(p+q+1)}.$$

1. On suppose dans cette question que $q = 2$.

   (a) Proposer $\widehat{p}_n$, un estimateur de p obtenu par la méthode des moments.

   (b) Démontrer que $\widehat{p}_n$ est fortement consistant.

   (c) Montrer que $\widehat{p}_n$ est asymptotiquement normal. Donner la valeur de sa variance asymptotique en fonction de p.

   (d) Soit $p_0 > 0$. Donner le test de Wald pour tester
   $$H_0 : p = p_0 \quad \text{contre} \quad H_1 : p \neq p_0.$$
   On exprimera soigneusement la statistique de test à l'aide de X.

2. On revient au cas général ou p et q sont inconnus. On veut tester

$$H_0 : p = q \quad \text{contre} \quad H_1 : p > q.$$

Pour k une constante, on propose un test de la forme

$$\phi_n(X) = 1_{\{\overline{X}_n > k\}}, \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   (a) Justifier la forme de la région de rejet.

   (b) Soit $\alpha \in]0;1[$. On pose
   $$k = \frac{1}{2} + \frac{1}{2\sqrt{n\alpha}}.$$
   En appliquant l'inégalité de Bienaymé-Tchebychev à $\overline{X}_n$, démontrer que $\phi_n(X)$ est de niveau α.

   (c) Montrer que la suite de tests $(\phi_n(X))_n$ est consistante.

   (d) Utiliser l'inégalité de Hoeffding pour obtenir une valeur $\widetilde{k}$ de sorte que le test $\widetilde{\phi}_n(X) = 1_{\{\overline{X}_n > \widetilde{k}\}}$ soit de niveau α. Comparer les valeurs de k et de $\widetilde{k}$ lorsque α prend de très petites valeurs.

Bibliographie

[1] Borovkov, A. A. (1998) *Mathematical statistics.* Gordon and Breach science publishers.

[2] Dacunha-Castelle, D. et Duflo, M. (1982) *Probabilités et statistiques, Tome 1 : Problèmes à temps fixe.* Masson.

[3] Hoffmann, M. (2014) *Introduction aux méthodes statistiques.* Polycopié de l'Ecole Polytechnique.

[4] Rivoirard, V. et Stoltz, G. (2009) *Statistique en action.* Vuibert.

[5] van der Vaart, A. (1998) *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, 3. Cambridge University Press, Cambridge.

[6] Wasserman, L. (2005) *All of statistics. A concise course in statistical inference.* Springer.