UNIVERSITÉ LIBRE DE BRUXELLES

FACULTÉ DES SCIENCES

Département de Mathématique

MATH-F-207 – Statistique mathématique 1

Davy Paindaveine

Version 2024–2025

"What do I know, or think I know, from my own experience
and not by literary osmosis ? An honest answer would be :
'Not much, and I am not too sure of most of it'. "

Dean Acheson
in *J. Cornfield's "A statistician's apology"*.

Table des matières

Table des matières **i**

I Introduction **1**

I.1 Le lien entre Probabilités et Statistique 1

I.2 La notion de modèle statistique 2

I.3 Trois exemples récurrents 3

I.3.1 L'exemple du verre de bière 4

I.3.2 L'exemple du bus 4

I.3.3 L'exemple électoral 4

II La théorie de l'échantillonnage **5**

II.1 Terminologie et définitions 5

II.1.1 Fonction de répartition 6

II.1.2 Moments 7

II.1.3 Quantiles 9

II.2 Lois échantillonnées 10

II.3 Lemme de Fisher 12

III Estimation ponctuelle **15**

III.1 Introduction 15

III.2 Critères d'estimation 16

III.2.1 Estimateurs convergents 16

III.2.2 Estimateurs exhaustifs 18
III.2.3 Biais 24
III.2.4 Estimateurs à erreur quadratique minimum 26
III.2.5 Estimateurs efficaces 31
III.3 Méthodes d'estimation 36
III.3.1 Méthode des moments 37
III.3.2 Méthode du maximum de vraisemblance 38
III.3.3 Comportement asymptotique des estimateurs du maximum de vraisemblance 41

IV Tests d'hypothèses **45**
IV.1 Terminologie et concepts de base 45
IV.1.1 Problèmes de test 45
IV.1.2 Tests, erreurs, risques, puissance 46
IV.1.3 Le principe de Neyman 48
IV.2 Le lemme fondamental de Neyman–Pearson 49
IV.3 Tests unilatéraux à puissance uniformément maximale 52
IV.4 Tests bilatéraux 58
IV.5 Tests de rapport de vraisemblance 60
IV.6 Tests χ^2 62
IV.6.1 Loi multinomiale 62
IV.6.2 Tests d'ajustement / tests d'adéquation 63
IV.6.3 Tests d'homogénéité 66
IV.6.4 Tests d'indépendance 68

V Intervalles de confiance **71**
V.1 Introduction 71
V.2 Intervalles exacts 72
V.2.1 Sans paramètre de nuisance 72
V.2.2 Avec paramètres de nuisance 75

V.3 Intervalles asymptotiques 76

V.4 Lien avec les tests d'hypothèses 78

V.5 Zones de confiance 80

A Quelques résultats de théorie des probabilités **82**

A.1 Convergence stochastique 82

A.2 Moments de vecteurs aléatoires 84

B Tables de probabilités **86**

B.1 Tables associées à la loi normale 86

B.1.1 Fonction de répartition de la loi normale 86

B.1.2 Quantiles de la loi normale 86

B.2 Tables associées à la loi de Student 88

B.3 Tables associées à la loi de chi-carré 88

B.4 Tables associées à la loi de Fisher 89

Bibliographie **95**

Chapitre I

Introduction

I.1 Le lien entre Probabilités et Statistique

Considérons, en guise d'exemple initial, l'expérience qui consiste à remplir un verre de bière. La quantité de liquide servie est influencée par différents facteurs, tels que la vitesse du mouvement, l'inclinaison du verre, etc. Ces facteurs rendent impossible de prédire avec certitude la quantité de bière versée (mesurée en cl) et celle-ci peut donc être considérée comme une variable aléatoire, que nous noterons X, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$. Il est raisonnable de supposer que la distribution P^X associée, qui est définie à travers le fait que $\mathrm{P}^X[B] = \mathrm{P}[X^{-1}(B)]$ pour tout B dans la sigma-algèbre de Borel $\mathcal{B}$ sur $\mathbb{R}$, est continue. Sur la base de la fonction de densité associée f^X, nous pouvons par exemple calculer l'espérance de la quantité de bière versée, donnée par

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f^X(x)\, dx,$$

ou la probabilité que X soit supérieure à 20 cl, qui vaut

$$\mathrm{P}[X > 20] \;=\; \int_{20}^{\infty} f^X(x)\, dx = 1 - \int_{-\infty}^{20} f^X(x)\, dx = 1 - F^X(20),$$

où F^X désigne la fonction de répartition de X. La *théorie des probabilités* permet d'évaluer les quantités précédentes sur la base de f^X (ou de F^X), c'est-à-dire sur la base de la distribution de X. En pratique, il est cependant impossible d'évaluer $\mathrm{E}[X]$ ou $\mathrm{P}[X > 20]$, car la distribution de X est inconnue.

Dans ce cadre, la *statistique* permet d'obtenir des informations sur $\mathrm{E}[X]$ et $\mathrm{P}[X > 20]$, ou plus généralement sur la distribution de X. Pour ce faire, on disposera d'observations $x_1, \ldots, x_n$ (pour un certain naturel non nul n), qui sont des réalisations de variables aléatoires $X_1, \ldots, X_n$ indépendantes ayant toutes la même distribution que X (il peut s'agir, par exemple, de n capacités de bières obtenues dans les mêmes conditions). Puisque les variables aléatoires X_i sont toutes de densité f^X, elles portent chacune de l'information sur la distribution de X. Intuitivement, lorsque n est grand, cette

information est suffisamment riche pour pouvoir "reconstruire" de façon satisfaisante la distribution de X, et donc également $\mathrm{E}[X]$ ou $\mathrm{P}[X > 20]$. Les probabilités et la statistique sont donc des processus de nature antagoniste : les probabilités font usage d'une distribution donnée/connue pour décrire (à travers leur plausibilité) les réalisations aléatoires correspondantes, tandis que la statistique, à l'inverse, fait usage des réalisations pour pouvoir appendre des choses sur la distribution inconnue. On parlera de *processus inférentiel* ou d'*inférence statistique*.

I.2 La notion de modèle statistique

En statistique, l'"observation" prend généralement la forme d'un vecteur aléatoire réel

$$X^{(n)} = (X_1, \ldots, X_n),$$

où les X_i sont des variables aléatoires à valeurs réelles. Puisque la distribution jointe de $X^{(n)}$, qui sera notée $\mathrm{P}^{(n)}$, est inconnue, on considérera un *modèle statistique*

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)}\right),$$

où $\mathcal{P}^{(n)}$ est une collection de mesures de probabilité sur $(\mathbb{R}^n, \mathcal{B}^{(n)})$, dont une seule constitue la "vraie" distribution de $X^{(n)}$. On peut par exemple adopter un modèle paramétrique d'échantillonnage, ce qui signifie qu'on suppose que $X_1, \ldots, X_n$ sont *i.i.d.* (*i*ndépendantes et *i*dentiquement *d*istribuées) avec une distribution commune appartenant à

$$\mathcal{P} = \left\{\mathrm{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\right\} \tag{I.1}$$

(dans ce cas, la collection $\mathcal{P}^{(n)}$ qui en résulte rassemble alors toutes les mesures produit $\mathrm{P}_\theta^{(n)} = \mathrm{P}_\theta \times \ldots \times \mathrm{P}_\theta$, n fois). Le terme "*échantillonnage*" indique qu'on considère des observations i.i.d., tandis que le qualificatif "*paramétrique*" désigne le fait que la famille de distributions (I.1) est indicée par un paramètre fini-dimensionnel.

En revenant à l'exemple du verre de bière, un modèle statistique possible est le modèle d'échantillonnage paramétrique gaussien

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)} : \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_0^+\right\}\right),$$

où $\mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ (la loi normale de moyenne μ et de variance σ^2). Pour ce modèle statistique paramétrique, on a

$$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \quad \Theta = \mathbb{R} \times \mathbb{R}_0^+ \quad \text{et} \quad k = 2.$$

Lorsqu'on adopte un modèle statistique paramétrique, il est rarement le cas en pratique que la distribution jointe de $X^{(n)}$ appartienne à ce modèle. Dans l'exemple considéré, il est clair qu'il n'existe pas de valeurs de μ et σ^2 pour lesquelles X_1 (la quantité de bière servie dans le premier

verre) soit exactement de loi $\mathcal{N}(\mu, \sigma^2)$. En effet, quelles que soient ces valeurs, la loi $\mathcal{N}(\mu, \sigma^2)$ prévoit une probabilité strictement positive que X_1 prenne une valeur strictement négative, ce qui est incompatible avec la nature de X_1. Néanmoins, pour certaines valeurs raisonnables des paramètres ($\mu = 15$ et $\sigma^2 = 1$?), la probabilité que X_1 soit négative est extrêmement proche de zéro, de sorte qu'il n'est pas exclu que la vraie distribution de X_1 puisse être très bien approximée par une loi normale. L'adoption d'un modèle paramétrique s'accompagne donc de l'hypothèse que l'une des distributions du modèle fournit une approximation raisonnable de la vraie loi sous-jacente (cette hypothèse pourra en fait être "testée" sur la base de $X^{(n)}$; voir la section IV.6.2).

Bien entendu, d'autres modèles peuvent être considérés, en choisissant d'autres familles de distributions que la famille des lois normales. Mais le choix d'une famille paramétrique de lois reste souvent assez arbitraire. C'est la raison pour laquelle on considère parfois des modèles d'échantillonnage *non paramétriques*

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_f^{(n)} : f \in \mathcal{F}\right\}\right),$$

où $\mathrm{P}_f^{(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. avec une loi commune admettant la fonction de densité f. L'ensemble $\mathcal{F}$ des densités sur $\mathbb{R}$ est un (sous-ensemble convexe [1] d'un) espace de fonctions qui, si on ne fait pas d'hypothèses sur la forme de f, est de dimension infinie. Une autre direction dans laquelle il est possible d'étendre les modèles d'échantillonnage paramétriques consiste à laisser tomber la structure d'échantillonnage. Des phénomènes aléatoires temporels (comme une suite de températures moyennes journalières) sont incompatibles avec le fait que les observations sont indépendantes (puisque, par exemple, les journées chaudes ont tendance à être groupées). Pour de tels modèles, on ne pourra se contenter de définir un ensemble de lois pour chaque observation X_i et on devra directement adopter un modèle paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\right\}\right)$$

ou un modèle non paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{\mathrm{P}_f^{(n)} : f \in \mathcal{F}^{(n)}\right\}\right)$$

(où $\mathcal{F}^{(n)}$ désigne ici l'ensemble de toutes les densités sur $\mathbb{R}^n$), qui prévoient dans chaque cas un ensemble de lois possibles pour la distribution jointe de $X^{(n)}$.

Dans ce cours, nous nous restreindrons le plus souvent aux modèles paramétriques. Certains des résultats présentés ne tiendront que pour les modèles d'échantillonnage correspondants, tandis que d'autres ne requerront pas la structure d'échantillonnage.

I.3 Trois exemples récurrents

Pour illustrer les concepts et méthodes considérés dans la suite, nous aurons souvent recours aux trois exemples (et modèles statistiques associés) suivants.

1. Ceci est lié aux contraintes voulant que $f \geq 0$ et $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

I.3.1 L'exemple du verre de bière

Nous avons déjà rencontré cet exemple plus haut, dans lequel la variable aléatoire générique d'intérêt est la quantité de bière X (en cl) servie dans un verre. Nous adopterons le modèle d'échantillonnage paramétrique gaussien

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)} : \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_0^+\right\}\right),$$

où $\mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. Parfois, nous considérerons le sous-modèle pour lequel la variance prend une valeur σ_0^2 connue. Dans ce cas, le paramètre est alors $\theta = \mu \in \Theta = \mathbb{R}$.

I.3.2 L'exemple du bus

Dans ce second exemple, nous supposons qu'un bus se présente de façon parfaitement régulière toutes les $\theta(> 0)$ minutes à l'arrêt. Le temps d'attente d'une personne qui arrive à un moment arbitraire à cet arrêt est alors une variable aléatoire X, que l'on supposera de loi uniforme sur $[0, \theta]$ ($X \sim \mathrm{Unif}([0, \theta])$). Si cette personne monte systématiquement dans le bus dès qu'un bus se présente, la valeur de θ reste inconnue. Des répétitions indépendantes de cette expérience mènent alors au modèle d'échantillonnage paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_{\theta}^{(n)} : \theta \in \mathbb{R}_0^+ \subset \mathbb{R}\right\}\right),$$

où $\mathrm{P}_{\theta}^{(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathrm{Unif}([0, \theta])$. Contrairement à l'exemple précédent, le paramètre est scalaire plutôt que vectoriel.

I.3.3 L'exemple électoral

Enfin, nous considérons, dans un scrutin électoral entre deux candidats, la proportion p (inconnue) de la population en faveur du candidat A. Si on sonde un individu choisi au hasard, sa préférence est bien représentée par une distribution de Bernoulli de paramètre p ($\mathrm{Bern}(p)$) : cet individu va répondre 1 (être en faveur du candidat A) avec probabilité p et 0 (être en faveur du candidat B) avec probabilité $1 - p$. Si on sonde n personnes de façon indépendante dans la rue sur leurs intentions de vote, on obtient alors le modèle d'échantillonnage paramétrique de Bernoulli

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_{p}^{(n)} : p \in [0, 1] \subset \mathbb{R}\right\}\right),$$

où $\mathrm{P}_{p}^{(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathrm{Bern}(p)$. Contrairement aux deux exemples précédents, (i) la distribution commune des observations X_i est discrète plutôt que continue et (ii) l'espace paramétrique ($\Theta = [0, 1]$) est ici compact.

Chapitre II

La théorie de l'échantillonnage

Ce chapitre préliminaire inclut les notions de base nécessaires pour aborder les trois grands problèmes d'inférence statistique faisant l'objet des chapitres suivants.

II.1 Terminologie et définitions

Considérons le modèle statistique d'échantillonnage paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\right\}\right),$$

où l'observation $X^{(n)} = (X_1, \ldots, X_n)$ est constituée d'observations i.i.d. (la distribution commune des X_i sera notée P_θ). Dans cette section, nous introduisons un certain nombre de concepts, que nous déclinerons tous dans deux versions : une version "théorique" ou "population" qui fait référence à la distribution P_θ inconnue, et une version "empirique" ou "échantillon", qui est associée aux observations $X_1, \ldots, X_n$.

Alors que les concepts population dépendront du paramètre θ, leur équivalent échantillon seront des *statistiques* $T(X^{(n)}) = T(X_1, \ldots, X_n)$, c'est-à-dire des fonctions (mesurables) des seules quantités à disposition du statisticien, à savoir les observations $X_1, \ldots, X_n$. Par exemple, les quantités

$$X_1, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i, \text{ et} \qquad \arctan\left(X_{n-1} + X_n^3\right),$$

sont des statistiques. Une statistique peut aussi être à valeurs vectorielles ou ensemblistes. Ainsi,

$$\begin{pmatrix} X_1 \\ \bar{X} \end{pmatrix}, \quad X^{(n)} = (X_1, \ldots, X_n) \quad \text{et} \quad \left[\bar{X} - X_1, \bar{X} + X_1\right]$$

sont également des statistiques. Lorsque, sous toutes les distributions $\mathrm{P}_\theta^{(n)}$ du modèle considéré, les observations X_i sont deux à deux différentes avec probabilité 1, on peut considérer la *statistique*

d'ordre

$$(X_{(1)}, \ldots, X_{(n)}),$$

où la *ième statistique d'ordre* $X_{(i)}$ est la $i$ème plus petite valeur parmi $X_1, \ldots, X_n$. En particulier, on a $X_{(1)} = \min(X_1, \ldots, X_n)$ et $X_{(n)} = \max(X_1, \ldots, X_n)$.

II.1.1 Fonction de répartition

Que la loi P_θ soit discrète ou continue, elle peut être décrite par sa fonction de répartition, définie par

$$\begin{aligned} F_\theta : \mathbb{R} &\to [0,1] \\ x &\mapsto F_\theta(x) = P_\theta[X_1 \leq x]. \end{aligned}$$

Pour tout $\theta \in \Theta$, la fonction de répartition F_θ satisfait les propriétés caractéristiques suivantes : (i) F_θ est non décroissante, (ii) F_θ est continue à droite, (iii) $\lim_{x\to-\infty} F_\theta(x) = 0$ et $\lim_{x\to+\infty} F_\theta(x) = 1$. Pour tout $a < b$, on a

$$\mathrm{P}_\theta[a < X_1 \leq b] = F_\theta(b) - F_\theta(a),$$

ce qui, en faisant tendre a vers b par valeurs plus petites, mène en particulier à

$$\mathrm{P}_\theta[X_1 = b] = F_\theta(b) - F_\theta(b-0), \quad \text{où } F_\theta(b-0) \overset{\text{not}}{=} \lim_{x \overset{<}{\to} b} F_\theta(x).$$

Si on désigne par $\mathbb{I}[A]$ la fonction indicatrice associée à la condition A (qui prend la valeur 1 si A est vraie et 0 sinon), l'équivalent empirique de la fonction de répartition F_θ est la *fonction de répartition empirique*

$$\begin{aligned} F_n : \mathbb{R} &\to [0,1] \\ x &\mapsto F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[X_i \leq x]. \end{aligned}$$

Il faut noter que, puisqu'elle dépend des variables aléatoires X_i, la fonction F_n est elle-même aléatoire. Quelles que soient les valeurs prises par les X_i, la fonction de répartition empirique est une fonction constante par morceaux, satisfaisant les propriétés caractéristiques (i)–(iii) ci-dessus. Il s'agit donc d'une fonction de répartition. En guise d'illustration, la figure II.1 présente la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$ ainsi qu'une instance de la fonction de répartition empirique résultant d'un tirage aléatoire de $n = 15$ observations.

Une application directe de la loi des grands nombres (le théorème A.9) montre que, pour tout $\theta \in \Theta$ et tout $x \in \mathbb{R}$,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[X_i \leq x] \overset{\text{p.s.}}{\to} \mathrm{E}_\theta[\mathbb{I}[X_1 \leq x]] = \mathrm{P}_\theta[X_1 \leq x] = F_\theta(x) \quad \text{sous } \mathrm{P}_\theta^{(n)}.$$

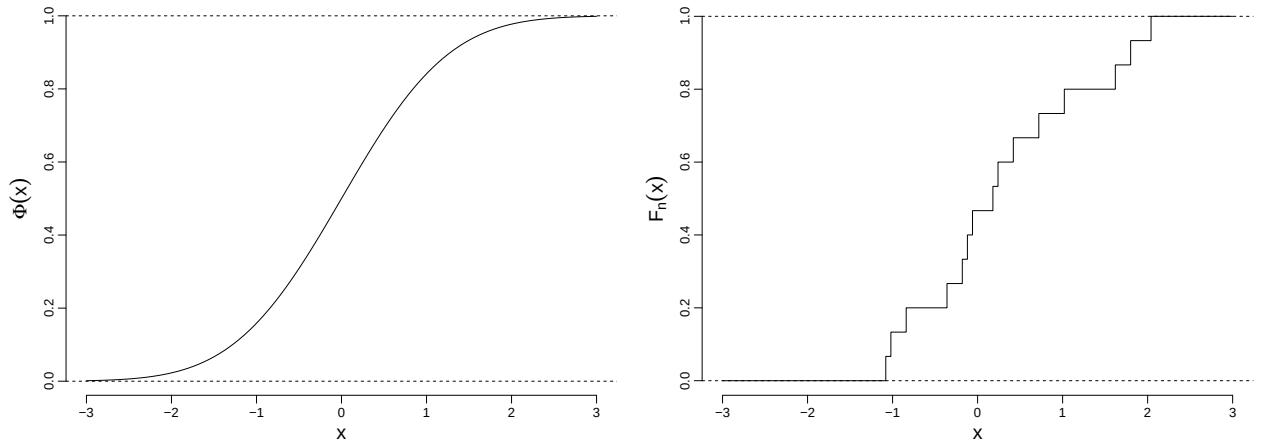


Figure II.1 – *Gauche : fonction de répartition Φ de la loi normale centrée réduite. Droite : fonction de répartition empirique F_n d'un échantillon de données i.i.d. $\mathcal{N}(0,1)$ de taille 15.*

Pour n grand, la fonction de répartition empirique fournit donc une très bonne approximation de la fonction de répartition F_θ inconnue. Ceci est d'autant plus vrai que ce résultat de convergence presque sûre ponctuelle peut être renforcé en le résultat de convergence presque sûre uniforme affirmant que, pour tout $\theta \in \Theta$,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F_\theta(x)| \overset{\text{p.s.}}{\to} 0 \quad \text{sous } \mathrm{P}_\theta^{(n)}.$$

Ce résultat est connu sous le nom de *théorème de Glivenko–Cantelli.*

II.1.2 Moments

Considérons d'abord le niveau population. On dira que la distribution de X_1 (ou, par extension, la variable aléatoire X_1 elle-même) admet des moments finis d'ordre $r(>0)$ sous P_θ si et seulement si

$$\mathrm{E}_\theta[|X_1|^r] = \int_{-\infty}^{+\infty} |x|^r \, dF_\theta(x)$$

existe et est finie. L'intégrale ci-dessus est une intégrale de Lebesgue–Stieltjes : dans le cas continu où la loi P_θ admet la densité f_θ, on évaluera une telle intégrale suivant la règle

$$\int_{-\infty}^{+\infty} g(x)\, dF_\theta(x) = \int_{-\infty}^{+\infty} g(x) f_\theta(x)\, dx,$$

tandis que si la loi P_θ est discrète, avec les valeurs possibles $x_i(\theta)$ et probabilités correspondantes $p_i(\theta)$, où i parcourt un ensemble au plus dénombrable d'indices $\mathcal{I}$, on utilisera plutôt l'expression

$$\int_{-\infty}^{+\infty} g(x)\, dF_\theta(x) = \sum_{i\in\mathcal{I}} g(x_i(\theta)) p_i(\theta).$$

On peut montrer que si une variable aléatoire admet des moments finis d'ordre $r(>0)$, alors elle admet aussi des moments finis d'ordre s pour tout $s \in (0, r)$. Ainsi, imposer des moments finis

d'ordre 2 est plus fort qu'imposer seulement des moments finis d'ordre 1 (en ce sens, les hypothèses du théorème central limite sont plus fortes que celle de la loi des grands nombres).

Considérons alors un modèle statistique d'échantillonnage paramétrique et un entier strictement positif r tel que, pour tout $\theta \in \Theta$, X_1 admet des moments finis d'ordre r sous P_θ. La quantité

$$\mu'_r(\theta) = \mathrm{E}_\theta[X_1^r]$$

est appelée *moment population non centré d'ordre r* (sous P_θ). Le moment non centré d'ordre 1 est simplement l'espérance de X_1, et est souvent noté μ. La quantité

$$\mu_r(\theta) = \mathrm{E}_\theta[(X_1 - \mu)^r]$$

est appelée *moment population centré d'ordre r* (sous P_θ). Si on se restreint aux ordres entiers, le premier moment centré intéressant est le moment d'ordre 2 (pourquoi ?), qui est la variance de X_1, laquelle sera souvent notée σ^2.

Il est bien connu que μ (le moment non centré d'ordre 1) est une mesure de position de la distribution de X_1, et que σ^2 (le moment centré d'ordre 2) est une mesure de dispersion de cette distribution. Parfois, on fonde sur les moments d'ordres 3 et 4 des mesures d'*asymétrie* et d'*aplatissement*. Plus précisément,

— on appelle *coefficient d'asymétrie de Fisher* la quantité

$$\gamma_1 = \frac{\mu_3(\theta)}{(\mu_2(\theta))^{3/2}}. \tag{II.1}$$

Des valeurs positives de γ_1 sont caractéristiques de déviations plus marquées à droite de μ qu'à gauche de μ (ce qui serait le cas, par exemple, pour le salaire mensuel X_1 d'un individu pris au hasard dans la population belge). Au contraire, des valeurs négatives de γ_1 sont typiquement obtenues lorsqu'on s'écarte plus fortement à gauche de μ qu'à droite, tandis que des distributions symétriques par rapport à leur moyenne μ livreront toujours $\gamma_1 = 0$ (par accident, une valeur nulle de γ_1 peut cependant être obtenue pour certaines distributions asymétriques !)

— on appelle *coefficient d'aplatissement de Fisher* la quantité

$$\gamma_2 = \frac{\mu_4(\theta)}{(\mu_2(\theta))^2} - 3. \tag{II.2}$$

Les lois normales livrent $\gamma_2 = 0$. Des valeurs positives de γ_2 sont associées à des distributions qui permettent plus d'observations extrêmes (c'est-à-dire situées loin de la moyenne) que pour les lois normales (à variances égales). De telles distributions "à queues lourdes" sont rencontrées notamment dans les données financières, lesquelles font intervenir plus souvent que les lois normales le prévoient, des bonnes et mauvaises surprises (des returns positifs et négatifs importants). À l'inverse, une valeur négative de γ_2 est associée à des queues plus

légères que celles des lois normales.

En (II.1)–(II.2), c'est un souci d'invariance qui explique la présence du dénominateur : quels que soient $a \in \mathbb{R}_0^+$ et $b \in \mathbb{R}$, γ_1 et γ_2 ne changent pas s'ils sont calculés sur $aX_1 + b$ plutôt que sur X_1. Ceci garantit notamment que ces quantités ne soient pas sensibles à des changements d'unité de mesure.

Au niveau échantillon, les moments empiriques non centrés et centrés d'ordre r sont respectivement les statistiques

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \quad \text{et} \quad m_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - m'_1)^r.$$

Contrairement aux moments population correspondants, ceux-ci peuvent être considérés indépendamment de toute condition de moments finis. Par contre, ces conditions sont requises pour s'assurer que les moments empiriques constituent des approximations raisonnables des moments populations associés. Par exemple, pour autant que X_1 admette des moments finis d'ordre r sous P_θ, la loi forte des grands nombres assure directement que, pour $n \to \infty$,

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \overset{\text{p.s.}}{\to} \mathrm{E}_\theta[X_1^r] = \mu'_r(\theta).$$

Bien entendu, les équivalents empiriques des moments populations considérés ci-dessus sont la moyenne empirique et la variance empirique

$$\bar{X} \overset{\text{not}}{=} m'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

et

$$s^2 \overset{\text{not}}{=} m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

ainsi que les coefficients d'asymétrie et d'aplatissement de Fisher empiriques

$$\frac{m_3}{m_2^{3/2}} \quad \text{et} \quad \frac{m_4}{m_2^2} - 3\cdot$$

II.1.3 Quantiles

Au niveau population, la *fonction quantile* est essentiellement l'inverse de la fonction de répartition F_θ. Si $F_\theta : \mathbb{R} \to [0,1]$ est continue et strictement croissante, le quantile d'ordre $\alpha(\in (0,1))$ de la loi P_θ est défini comme

$$x_\alpha(\theta) \overset{\text{not}}{=} F_\theta^{-1}(\alpha),$$

où l'existence de la fonction réciproque est assurée par les hypothèses considérées sur F_θ. On lira donc $x_\alpha(\theta)$ comme le nombre (unique) x tel que $X_1(\sim \mathrm{P}_\theta)$ ait une probabilité α de prendre une valeur inférieure ou égale à $x_\alpha(\theta)$: $F_\theta(x_\alpha(\theta)) = \alpha$.

Si F_θ n'est pas continue, il se pourrait que l'équation $F_\theta(x) = \alpha$ n'ait pas de solution en x. De même, si F_θ n'est pas strictement croissante, il se pourrait que le nombre de solutions de cette équation soit infini non dénombrable ! Pour définir le quantile d'ordre α de façon générale, on a recours a un concept d'inverse généralisé. Plus précisément, on appellera *quantile d'ordre* $\alpha(\in (0,1))$ *de* X_1 *sous* P_θ le nombre

$$x_\alpha(\theta) \stackrel{\text{def}}{=} \inf\Big\{x \in \mathbb{R} : F_\theta(x) \geq \alpha\Big\}. \tag{II.3}$$

Lorsqu'il y a une unique solution à l'équation $F_\theta(x) = \alpha$, on vérifie que cette solution coïncide avec le quantile général défini en (II.3), de sorte que cette définition générale englobe la définition particulière. Points de terminologie : $x_{1/2}(\theta)$ est appelée la *médiane*, $x_{1/4}(\theta)$ et $x_{3/4}(\theta)$ sont appelés les *1er et 3ème quartiles*. Les quantiles $x_{m/10}(\theta)$, $m = 1, \ldots, 9$, sont les *déciles*, tandis que les quantiles $x_{m/100}(\theta)$, $m = 1, \ldots, 99$, sont les *percentiles*. Chacun de ces quantiles constitue une mesure de position, tandis que la différence entre deux quantiles peut être considérée comme une mesure de dispersion, la plus couramment utilisée étant l'*intervalle interquartile* $x_{3/4}(\theta) - x_{1/4}(\theta)$.

Au niveau empirique, on définira de même *le quantile empirique d'ordre* $\alpha(\in (0,1))$ comme le nombre

$$x_\alpha^{(n)} \stackrel{\text{def}}{=} \inf\Big\{x \in \mathbb{R} : F_n(x) \geq \alpha\Big\},$$

où F_n désigne la fonction de répartition empirique associée à l'échantillon $X^{(n)} = (X_1, \ldots, X_n)$. Sur cette base, on peut considérer la médiane empirique, les 1er et 3ème quartiles empiriques, etc.

II.2 Lois échantillonnées

Considérons le modèle statistique paramétrique

$$\Big(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \Big\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\Big\}\Big),$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Soit $T(X^{(n)})$ une statistique à valeurs dans $\mathbb{R}^m$. Dans ce cadre, la *distribution échantillonnée de* $T(X^{(n)})$ *sous* $\mathrm{P}_\theta^{(n)}$ est la mesure de probabilité définie par

$$\mathrm{P}_\theta^{T(n)}[B] = \mathrm{P}_\theta^{(n)}[T(X^{(n)}) \in B] = \mathrm{P}_\theta^{(n)}\big[\{x^{(n)} \in \mathbb{R}^n : T(x^{(n)}) \in B\}\big]$$

pour tout B dans la sigma-algèbre de Borel $\mathcal{B}^m$ sur $\mathbb{R}^m$. Nous illustrons maintenant ce concept à l'aide des trois exemples de la Section I.3 :

(1) Exemple électoral : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $p \in [0,1]$. Alors la loi échantillonnée de la statistique $T(X^{(n)}) = \sum_{i=1}^n X_i$ est la loi Bin(n,p).

(2) Exemple du verre de bière : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+$. Alors la loi échantillonnée de la statistique $T(X^{(n)}) = \sum_{i=1}^n X_i$ est la loi $\mathcal{N}(n\mu, n\sigma^2)$, et celle de $T(X^{(n)}) = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ est la loi $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

(3) Exemple du bus : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif$([0,\theta])$, où $\theta \in \mathbb{R}_0^+$. Alors la fonction de répartition associée à la loi échantillonnée de la statistique $T(X^{(n)}) =$

$X_{(n)} = \max(X_1, \ldots, X_n)$ est donnée par

$$F_\theta^{X_{(n)}}(x) = \mathrm{P}_\theta^{(n)}[X_{(n)} \leq x] = \mathrm{P}_\theta^{(n)}[X_1 \leq x, \ldots, X_n \leq x]$$

$$= \prod_{i=1}^{n} \mathrm{P}_\theta^{(n)}[X_i \leq x] = \begin{cases} 0 & \text{si } x < 0 \\ (\frac{x}{\theta})^n & \text{si } 0 < x < \theta \\ 1 & \text{si } x \geq \theta. \end{cases} \tag{II.4}$$

De façon équivalente, on peut décrire la distribution échantillonnée de $X_{(n)}$ par sa densité

$$f_\theta^{X_{(n)}}(x) = \frac{n x^{n-1}}{\theta^n} \mathbb{I}[0 \leq x \leq \theta]. \tag{II.5}$$

Dans les exemples qui précèdent, on a toujours pu déterminer la distribution échantillonnée des statistiques considérées. Ceci reste cependant exceptionnel. Quand il n'est pas possible de déterminer la distribution échantillonnée de $T(X^{(n)})$, on peut

(i) déterminer les moments (ou certains moments) de la distribution échantillonnée. Si une collection finie de moments ne détermine pas la distribution échantillonnée, elle peut parfois décrire de façon utile cette distribution. Sous certaines conditions, on peut montrer que la collection de tous les moments $\mathrm{E}_\theta[(T(X^{(n)}))^r]$, $r \in \mathbb{N}_0$, détermine la distribution échantillonnée de $T(X^{(n)})$, ce qui motive de considérer les moments pour étudier cette distribution ;

(ii) déterminer la distribution échantillonnée *asymptotique*, c'est-à-dire la limite en loi de la distribution échantillonnée *exacte* (celle, ci-dessus, qui est associée à une valeur fixée n de la taille d'échantillon). En pratique, si n est "grand", cette distribution échantillonnée asymptotique fournira une excellente approximation de la distribution échantillonnée exacte incalculable.

Exemple pour (i) : considérons le modèle statistique pour lequel l'observation $X^{(n)} = (X_1, \ldots, X_n)$ regroupe des variables aléatoires i.i.d. dont la loi commune admet des moments finis d'ordre 2, mais est pour le reste non spécifiée. Pour ce modèle d'échantillonnage de nature non paramétrique, la collection des lois associées à $X^{(n)}$ peut par exemple être indicée par la fonction de répartition F commune des X_i. Posons $\mu = \mathrm{E}_F[X_1]$ et $\sigma^2 = \mathrm{Var}_F[X_1]$, et considérons la statistique $T(X^{(n)}) = \bar{X}$. Alors le moment non centré d'ordre 1 de $T(X^{(n)})$ est

$$\mathrm{E}_F[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_F[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu$$

et son moment centré d'ordre 2 est

$$\mathrm{Var}_F[\bar{X}] = \frac{1}{n^2} \mathrm{Var}_F\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}_F[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.$$

Notons que ceci est parfaitement compatible avec la distribution exacte de $\bar{X}$ que nous avons obtenue ci-dessus dans le modèle d'échantillonnage paramétrique gaussien.

Exemple pour (ii) : pour le même modèle que ci-dessus, le théorème central limite (le théorème A.10) livre directement que, quand $n \to \infty$,

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad \text{sous } \mathrm{P}_F^{(n)}.$$

Ceci décrit complètement la distribution échantillonnée asymptotique de $\bar{X}$ sous $\mathrm{P}_F^{(n)}$. En écrivant "$\approx$" pour "est approximativement de loi", on pourra donc considérer pour n grand que $\sqrt{n}(\bar{X} - \mu) \approx \mathcal{N}(0, \sigma^2)$ sous $\mathrm{P}_F^{(n)}$, ou encore que

$$\bar{X} \approx \mathcal{N}(\mu, \tfrac{\sigma^2}{n})$$

sous $\mathrm{P}_F^{(n)}$. À nouveau, ceci est en ligne avec la distribution exacte de $\bar{X}$ que nous avons obtenue ci-dessus dans le modèle d'échantillonnage paramétrique gaussien.

II.3 Lemme de Fisher

Le but de cette section est d'établir un résultat classique, qui précise la distribution échantillonnée (exacte) de la statistique

$$T(X^{(n)}) = \begin{pmatrix} \bar{X} \\ s^2 \end{pmatrix}$$

dans le modèle d'échantillonnage paramétrique gaussien. Comme déterminé dans l'exemple (2) ci-dessus, on a $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$. Il reste donc à déterminer la distribution échantillonnée de s^2, ainsi que la structure de dépendance de $\bar{X}$ et s^2. Pour ce faire, nous devons introduire les lois suivantes.

Definition II.1 *La variable aléatoire Q est de loi chi-carré à $k(\in \mathbb{N}_0)$ degrés de liberté (notation : $Q \sim \chi^2_k$) $\Leftrightarrow$ Q a la même distribution que $\sum_{i=1}^k Z_i^2$, où $Z_1, \ldots, Z_k$ sont i.i.d. de loi commune $\mathcal{N}(0,1)$.*

On peut montrer que si $Q \sim \chi^2_k$, alors Q admet la densité

$$f^Q(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} \exp(-x/2)\, \mathbb{I}[x > 0],$$

où

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t)\, dt$$

est la fonction Gamma d'Euler. On peut aussi vérifier que $\mathrm{E}[Q] = k$ et $\mathrm{Var}[Q] = 2k$. Bien entendu, il découle directement de la définition ci-dessus que, si $Q_1 \sim \chi^2_{k_1}$ et $Q_2 \sim \chi^2_{k_2}$ sont des variables aléatoires indépendantes, alors $Q_1 + Q_2 \sim \chi^2_{k_1+k_2}$.

Nous pouvons maintenant présenter le résultat principal de cette section, qui est connu sous le nom de *Lemme de Fisher*.

Théorème II.2 *Soient $X_1, \ldots, X_n$ ($n \geq 2$) i.i.d. de loi commune $\mathcal{N}(\mu, \sigma^2)$. Alors,*

(i) $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$,

(ii) $ns^2/\sigma^2 \sim \chi^2_{n-1}$, *et*

(iii) $\bar{X}$ *et* s^2 *sont indépendantes.*

PREUVE. Si on pose $Z_i = (X_i - \mu)/\sigma$ pour $i = 1, \ldots, n$, on vérifie directement que

$$\bar{X} = \sigma\bar{Z} + \mu \quad \text{et} \quad ns_z^2 = \frac{ns^2}{\sigma^2},$$

où on a écrit $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ et $s_z^2 = \frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z})^2$. Par conséquent, pour démontrer le résultat, il suffit de prouver que (i) $\bar{Z} \sim \mathcal{N}(0, \frac{1}{n})$, (ii) $ns_z^2 \sim \chi^2_{n-1}$ et que (iii) $\bar{Z}$ et s_z^2 sont mutuellement indépendantes.

Pour ce faire, notons que les Z_i sont i.i.d. de loi commune $\mathcal{N}(0,1)$. La loi jointe de $Z^{(n)} = (Z_1, \ldots, Z_n)$ admet donc la densité

$$\begin{aligned} z^{(n)} = (z_1, \ldots, z_n) \mapsto f^{Z^{(n)}}(z^{(n)}) &= \prod_{i=1}^n \left\{ \left(\frac{1}{2\pi}\right)^{1/2} \exp\Big(-\frac{1}{2}z_i^2\Big) \right\} \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\Big(-\frac{1}{2}\sum_{i=1}^n z_i^2\Big) \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\Big(-\frac{1}{2}\|z^{(n)}\|^2\Big). \end{aligned} \tag{II.6}$$

Soit O une matrice orthogonale $n \times n$ arbitraire telle que $O_{1j} = 1/\sqrt{n}$ pour tout $j = 1, \ldots, n$. Alors la densité de

$$Y^{(n)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = O \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

est

$$y^{(n)} = (y_1, \ldots, y_n) \mapsto f^{Y^{(n)}}(y^{(n)}) = |\det O^{-1}|\, f^{Z^{(n)}}(O^{-1}y^{(n)}) = f^{Z^{(n)}}(O'y^{(n)}) = f^{Z^{(n)}}(y^{(n)}),$$

où on a utilisé l'expression (II.6). On en conclut que $Y^{(n)}$ et $Z^{(n)}$ ont la même loi, ce qui implique que les Y_i sont également i.i.d. de loi commune $\mathcal{N}(0,1)$.

En particulier,

$$\sqrt{n}\bar{Z} = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i = Y_1 \sim \mathcal{N}(0,1),$$

ce qui montre (i). D'autre part, la loi jointe de $Y^{(n)}$ implique aussi que

$$\begin{aligned} ns_z^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 &= \sum_{i=1}^n (Z_i^2 - 2Z_i\bar{Z} + \bar{Z}^2) \\ &= \Big(\sum_{i=1}^n Z_i^2\Big) - 2\bar{Z}\Big(\sum_{i=1}^n Z_i\Big) + n\bar{Z}^2 = \Big(\sum_{i=1}^n Z_i^2\Big) - n\bar{Z}^2 \\ &= \|Z^{(n)}\|^2 - \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\Big)^2 = \|Y^{(n)}\|^2 - Y_1^2 = \sum_{i=2}^n Y_i^2 \end{aligned}$$

est de loi χ^2_{n-1}, ce qui établit (ii). Enfin, (iii) découle du fait que $\bar{Z}$ est une fonction de Y_1, que s_z^2 est une fonction de $(Y_2, \ldots, Y_n)$ et que les Y_i sont mutuellement indépendants. □

Pour pouvoir énoncer un corollaire du lemme de Fisher, nous introduisons une nouvelle famille de lois.

Definition II.3 *La variable aléatoire V est de loi de Student (ou de loi t) à $k(\in \mathbb{N}_0)$ degrés de liberté (notation : $V \sim t_k$) $\Leftrightarrow$ V a la même distribution que*

$$\frac{Z}{\sqrt{Q/k}},$$

où $Z \sim \mathcal{N}(0,1)$ et $Q \sim \chi^2_k$ sont indépendantes.

Il découle de cette définition, de la loi des grands nombres et du Lemme de Slutsky (Théorème A.8) que si $k \to \infty$, la loi t_k tend vers la loi $\mathcal{N}(0,1)$, au sens de la convergence en loi (exercice). Ceci peut aussi être vérifié directement sur la base de la fonction de densité d'une variable aléatoire de loi t_k, qui satisfait

$$f^V(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})}\left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} \qquad \text{pour tout } x \in \mathbb{R}.$$

De cette fonction de densité, il découle que $V \sim t_k$ a des moments finis d'ordre r si et seulement si $r < k$ (exercice). Les queues de la loi t_k sont donc d'autant plus lourdes que k est petit. En particulier, la loi t_1 (ou *loi de Cauchy*) n'admet pas de moment fini d'ordre 1 et on ne peut donc pas parler de l'espérance d'une variable aléatoire possédant cette loi !

Corollaire II.4 *Soient $X_1, \ldots, X_n$ $(n \geq 2)$ i.i.d. de loi commune $\mathcal{N}(\mu, \sigma^2)$. Alors*

$$\frac{\sqrt{n-1}(\bar{X} - \mu)}{s} \sim t_{n-1},$$

où s désigne la racine carrée de la variance empirique s^2.

Preuve. Par le lemme de Fisher,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim \mathcal{N}(0,1)$$

et

$$\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

sont indépendantes. L'égalité

$$\frac{\sqrt{n-1}(\bar{X} - \mu)}{s} = \frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{ns^2}{\sigma^2}/(n-1)}}$$

livre alors le résultat. □

Chapitre III

Estimation ponctuelle

III.1 Introduction

Considérons le modèle statistique paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{ \mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k \right\}\right), \tag{III.1}$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Le paramètre θ étant inconnu, le premier objectif naturel de l'inférence statistique consiste à parier sur la *vraie* valeur de θ, c'est-à-dire sur l'unique valeur du paramètre correspondant à la distribution jointe de $X^{(n)}$. Ce pari ne peut être fondé que sur la seule chose à disposition du statisticien, à savoir $X^{(n)}$, et doit viser une valeur de θ, que l'on sait appartenir à Θ. Ceci mène à la définition suivante.

Definition III.1 *Considérons le modèle statistique en (III.1) et une fonction mesurable $g : \Theta \to \mathbb{R}^m$. Un* estimateur $T(X^{(n)})$ *de $g(\theta)$ est une statistique à valeurs dans $g(\Theta)$. Dans le cas où on estime θ lui-même (pour lequel g est la fonction identité sur Θ), on notera parfois un estimateur par $\hat{\theta}$.*

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0, 1] \subset \mathbb{R}$ (exemple électoral). Un estimateur naturel de p est $T(X^{(n)}) = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. De même, un estimateur naturel de $g(p) = 3p^2 + 1$ est $T(X^{(n)}) = 3\bar{X}^2 + 1$.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $\theta = (\mu, \sigma^2)$ est un élement de $\Theta = \mathbb{R} \times \mathbb{R}_0^+ \subset \mathbb{R}^2$ (exemple du verre de bière). L'estimateur le plus classique de θ est celui qui apparaît dans le lemme de Fisher, c'est-à-dire

$$T(X^{(n)}) = \begin{pmatrix} \bar{X} \\ s^2 \end{pmatrix}.$$

Dans ce cadre où le paramètre θ est multidimensionnel, la fonction g permet de considérer le problème d'une des composantes de θ seulement. Par exemple, la fonction $g(\theta) = g(\mu, \sigma^2) = \mu$ mène à

l'estimation de la composante μ de θ ; on dira que μ est le *paramètre d'intérêt*, tandis que σ^2 joue le rôle de *paramètre de nuisance*. Pour ce problème, $\bar{X}$ et $\mathrm{Med}(X_1, \ldots, X_n)$ sont deux estimateurs naturel de $g(\theta) = \mu$. Bien entendu, le cas pour lequel σ^2 est le paramètre d'intérêt et μ est le paramètre de nuisance peut aussi être envisagé et est associé à la fonction $g(\theta) = \sigma^2$.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathrm{Unif}([0, \theta])$, où $\theta \in \Theta = \mathbb{R}_0^+$ (exemple du bus). Alors les statistiques $X_{(n)}$, $2\bar{X}$, $nX_{(1)}$, $X_{(1)}+X_{(n)}$ sont toutes des estimateurs raisonnables de θ. Bien entendu, il y a bien d'autres estimateurs de θ, dont certains sont beaucoup moins raisonnables, comme X_1, X_1^2, $\exp(\mathrm{Med}(X_1, \ldots, X_n))$, etc.

Ce dernier exemple rend clair que la classe de tous les estimateurs possibles est en fait gigantesque (tous les coups sont permis !). Ceci est excitant, mais soulève également un grand nombre de questions, parmi lesquelles : *comment comparer deux estimateurs ? Quelles propriétés un "bon" estimateur doit-il avoir ? Y a-t-il une limite à la précision d'un estimateur ? Comment construire de façon "systématique" des estimateurs ? Comment construire un estimateur atteignant la limite de précision si elle existe ? ...*

Le but de ce chapitre est d'apporter des réponses à ces questions.

III.2 Critères d'estimation

En pratique, le statisticien reçoit les données observées, qui prennent la forme d'un n-uple de réels $x^{(n)} = (x_1, \ldots, x_n)$. Si on dispose d'un estimateur $T(X^{(n)})$ de $g(\theta)$, ceci mène à *l'estimation de* $g(\theta)$ donnée par $T(x^{(n)})$, laquelle est une valeur fixée dans $g(\Theta)$. On distinguera cette valeur fixée de *l'estimateur* $T(X^{(n)})$, qui est une variable aléatoire (un vecteur aléatoire, dans le cas où $m > 1$). Puisqu'un estimateur est en particulier une statistique, on peut considérer sa distribution échantillonnée. Les *critères d'estimation* désirables que nous allons présenter ci-dessous sont des critères qui concernent cette distribution échantillonnée.

III.2.1 Estimateurs convergents

La situation où la distribution échantillonnée de l'estimateur $T(X^{(n)}) = T(X_1, \ldots, X_n)$ est dégénérée en $g(\theta)$ (c'est-à-dire où, pour tout θ, $\mathrm{P}_\theta^{(n)}[T(X^{(n)}) = g(\theta)] = 1$) est le cas où on peut estimer $g(\theta)$ sans erreur, un cas idéal qui est malheureusement tout à fait irréaliste. Par contre, on peut raisonnablement espérer que plus la taille d'échantillon n croît, plus l'estimation est précise, et que quand n diverge vers l'infini, on se retrouve finalement dans ce cas idéal. Ceci mène au concept d'estimation *convergente*.

Definition III.2 *Considérons le modèle statistique en (III.1) et une fonction mesurable $g : \Theta \to \mathbb{R}^m$. (i) L'estimateur $T(X^{(n)})$ de $g(\theta)$ est faiblement convergent si et seulement si, pour tout $\theta \in \Theta$, $T(X^{(n)}) \xrightarrow{\mathrm{P}} g(\theta)$ sous $\mathrm{P}_\theta^{(n)}$ quand $n \to \infty$. (ii) L'estimateur $T(X^{(n)})$ de $g(\theta)$ est fortement convergent si et seulement si, pour tout $\theta \in \Theta$, $T(X^{(n)}) \xrightarrow{\mathrm{p.s.}} g(\theta)$ sous $\mathrm{P}_\theta^{(n)}$ quand $n \to \infty$.*

La convergence forte d'un estimateur peut parfois suivre d'un résultat limite classique en pro-

babilité, comme la loi forte des grands nombres. Dans le cas du modèle d'échantillonnage gaussien, par exemple, c'est le cas, trivialement, pour la convergence forte de l'estimateur $\hat{\mu} = \bar{X}$ de μ. Mais c'est aussi le cas pour l'estimateur $\hat{\sigma}^2 = s^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ de σ^2, puisqu'on a

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^n \Big\{(X_i - \mu) + (\mu - \bar{X})\Big\}^2 \\ &= \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X})\frac{1}{n}\sum_{i=1}^n (X_i - \mu) + (\mu - \bar{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2 \\ &\overset{\text{p.s.}}{\to} \sigma^2, \end{aligned}$$

où on a utilisé la loi forte des grands nombres et le théorème A.7. C'est encore le cas, dans l'exemple du bus, pour l'estimateur $2\bar{X}$ de θ, puisque la loi forte des grands nombres implique que, sous $\mathrm{P}_\theta^{(n)}$, $2\bar{X} \overset{\text{p.s.}}{\to} 2\mathrm{E}_\theta[X_1] = 2 \times \frac{\theta}{2} = \theta$, quand $n \to \infty$. Dans tous les exemples ci-dessus, les estimateurs sont aussi faiblement convergents, puisque la convergence presque sûre implique la convergence en probabilité (Théorème A.5(i)).

Si nous verrons à la section III.2.4 une autre méthode pour établir la convergence (faible) d'un estimateur, on devra parfois recourir directement à la définition pour établir (ou écarter !) la convergence. Nous illustrons ceci sur l'exemple du bus en considérant l'estimateur $X_{(n)}$ de θ. En utilisant la fonction de répartition en (II.4), on obtient que, pour tout $\theta > 0$ et tout $\varepsilon > 0$,

$$\begin{aligned} \mathrm{P}_\theta^{(n)}[|X_{(n)} - \theta| > \varepsilon] &= \mathrm{P}_\theta^{(n)}[X_{(n)} < \theta - \varepsilon] + \mathrm{P}_\theta^{(n)}[X_{(n)} > \theta + \varepsilon] \\ &= F^{X_{(n)}}(\theta - \varepsilon) + 0 \\ &= \begin{cases} (\frac{\theta - \varepsilon}{\theta})^n & \text{si } 0 < \varepsilon < \theta \\ 0 & \text{si } \varepsilon \geq \theta \end{cases} \\ &\to 0 \end{aligned}$$

quand $n \to \infty$. Cet estimateur est donc faiblement convergent pour θ. On montre de la même façon (ou en utilisant le théorème A.7) que $\frac{n}{n+1}X_{(n)}$ est également faiblement convergent (nous verrons plus tard ce qui rend cet estimateur intéressant). On peut en fait aussi montrer que $X_{(n)}$ et $\frac{n}{n+1}X_{(n)}$ sont fortement convergents pour θ.

Toujours dans l'exemple du bus, considérons l'estimateur $\hat{\theta}^{(n)} = (n+1)X_{(1)}$ de θ. Sous $\mathrm{P}_\theta^{(n)}$, sa

fonction de répartition est donnée par

$$F^{\hat{\theta}^{(n)}}(x) = \mathrm{P}_\theta^{(n)}[(n+1)X_{(1)} \le x] = 1 - \mathrm{P}_\theta^{(n)}[X_{(1)} > \tfrac{x}{n+1}] = 1 - \textstyle\prod_{i=1}^n \mathrm{P}_\theta[X_i > \tfrac{x}{n+1}]$$

$$= 1 - (1 - F_\theta^{X_1}(\tfrac{x}{n+1}))^n = \begin{cases} 0 & \text{si } x < 0 \\ 1 - (1 - \frac{x}{(n+1)\theta})^n & \text{si } 0 \le x < (n+1)\theta \\ 1 & \text{si } x \ge (n+1)\theta. \end{cases}$$

En utilisant le fait que $(1 - 1/n)^n \to e^{-1}$ quand $n \to \infty$, on en déduit que, pour tout $x \in \mathbb{R}$,

$$F^{\hat{\theta}^{(n)}}(x) \to F(x) \stackrel{\text{def}}{=} (1 - e^{-x/\theta})\mathbb{I}[x \ge 0]$$

quand $n \to \infty$. En d'autres termes, $\hat{\theta}^{(n)}$ converge en loi vers la loi exponentielle de paramètre θ. Il en découle que $\hat{\theta}^{(n)}$ ne converge pas, sous $\mathrm{P}_\theta^{(n)}$, vers θ en probabilité, car, si c'était le cas, cette convergence tiendrait aussi en loi (Théorème A.5). On conclut que cet estimateur de θ n'est pas faiblement (donc pas non plus fortement) convergent.

On considère généralement la convergence comme la qualité la plus fondamentale dont doit bénéficier un estimateur, au sens où tout estimateur non-convergent est à écarter définitivement. On a surtout à l'esprit qu'on écarte ainsi des estimateurs fantaisistes qui n'ont aucune chance d'être convergents, tels que des estimateurs de la forme $T(X_1, \ldots, X_r)$ qui, indépendamment de n, n'utilisent qu'un nombre fini r d'observations (penser aux cas extrêmes où $r = 0$ ou $r = 1$, comme, dans l'exemple du bus, l'estimateur θ_0 de θ, où $\theta_0 > 0$ est fixé, ou l'estimateur X_1 de θ). Cependant, des estimateurs a priori raisonnables, comme l'est (on le verra plus bas) l'estimateur $(n+1)X_{(1)}$ de θ dans l'exemple du bus, peuvent également être écartés grâce à ce critère.

III.2.2 Estimateurs exhaustifs

Considérons de nouveau le modèle statistique paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\right\}\right),$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Comme on vient de le voir, un estimateur qui, indépendamment de n, n'utilise qu'un nombre fini d'observations X_i, ne peut pas être convergent. Ceci est lié au fait que de tels estimateurs ne font pas usage de toute l'information pertinente sur θ qui réside dans $X^{(n)}$. Nous présentons maintenant un concept qui permettra de préciser mathématiquement ce qu'est une statistique (et donc en particulier, un estimateur) qui exploite toute l'information pertinente.

Definition III.3 *Dans le cadre du modèle paramétrique ci-dessus, la statistique $T(X^{(n)})$ est* exhaustive *si et seulement si pour tout $B \in \mathcal{B}^n$ (la sigma-algèbre de Borel sur $\mathbb{R}^n$) et pour tout $t \in \mathcal{T}^{(n)}$ (l'ensemble des valeurs possibles de $T(X^{(n)})$),*

$$\mathrm{P}_\theta^{(n)}[X^{(n)} \in B \,|\, T(X^{(n)}) = t]$$

ne dépend pas de θ.

Ce concept peut être interprété de la façon suivante.

Tacitement, on a bien entendu supposé que des valeurs différentes de θ mènent à des mesures de probabilité $\mathrm{P}_\theta^{(n)}$ différentes (si au contraire, la paramétrisation n'est pas injective, alors il est impossible d'estimer θ). Il en découle qu'il existe toujours $B \in \mathcal{B}^n$ tel que $\mathrm{P}_\theta^{(n)}[X^{(n)} \in B]$ dépend de θ. Et c'est sur la comparaison de ces probabilités que s'appuie l'inférence statistique pour discerner entre des valeurs de θ qui sont "vraisemblables" et d'autres qui le sont moins. C'est dans cette dépendance que se matérialise l'information sur θ résidant dans $X^{(n)}$. La définition ci-dessus précise qu'une fois donnée la valeur t prise par une statistique exhaustive $T(X^{(n)})$, les probabilités des événements $[X^{(n)} \in B]$ ne dépendent plus de θ, de sorte qu'il est inutile, pour apprendre quelque chose de plus sur θ, de savoir si un tel événement s'est réalisé ou pas. En d'autres termes, lorsqu'on donne la valeur qu'une statistique exhaustive a prise sur un échantillon $X^{(n)}$, on donne toute l'information résidant dans $X^{(n)}$ qui est pertinente pour faire de l'inférence sur θ.

Par exemple, quel que soit le modèle paramétrique général considéré ci-dessus, la statistique $T(X^{(n)}) = X^{(n)}$ est exhaustive car, pour tout $B \in \mathcal{B}^n$ et pour toute valeur possible $x^{(n)}$ de $X^{(n)}$,

$$\mathrm{P}_\theta^{(n)}[X^{(n)} \in B \,|\, X^{(n)} = x^{(n)}] = \mathbb{I}[x^{(n)} \in B],$$

qui ne dépend pas de θ. Cet exemple est cependant dégénéré, car, au vu de l'interprétation ci-dessus, on vient seulement d'apprendre que "toute l'information pertinente sur θ et qui réside dans $X^{(n)}$... se trouve dans $X^{(n)}$". Clairement, le concept d'exhaustivité n'aura d'intérêt que quand $T(X^{(n)})$ n'est pas en bijection avec $X^{(n)}$, de sorte qu'on aura réduit l'information en passant de $X^{(n)}$ à $T(X^{(n)})$, mais sans, si $T(X^{(n)})$ est effectivement exhaustive, perdre de l'information pertinente sur θ.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0,1] \subset \mathbb{R}$ (exemple électoral). Nous allons montrer que la statistique $T(X^{(n)}) = \sum_{i=1}^n X_i$ est exhaustive. L'ensemble des valeurs possibles de $X^{(n)}$ est $\mathcal{X}^{(n)} = \{0,1\}^n$ et celui de $T(X^{(n)})$ est $\mathcal{T}^{(n)} = \{0, 1, \ldots, n\}$. Prenons d'abord $x^{(n)} = (x_1, \ldots, x_n) \in \mathcal{X}^{(n)}$ et $t \in \mathcal{T}^{(n)}$. Alors,

$$\begin{aligned}
\mathrm{P}_p^{(n)}[X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t] &= \frac{\mathrm{P}_p^{(n)}[X^{(n)} = x^{(n)}, T(X^{(n)}) = t]}{\mathrm{P}_p^{(n)}[T(X^{(n)}) = t]} \\
&= \frac{\mathrm{P}_p^{(n)}[X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^n X_i = t]}{\mathrm{P}_p^{(n)}[\sum_{i=1}^n X_i = t]}.
\end{aligned}$$

Si $t \neq \sum_{i=1}^n x_i$, $\mathrm{P}_p^{(n)}[X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^n X_i = t] = 0$. Si $t = \sum_{i=1}^n x_i$, on obtient alors, en

utilisant le fait que $T(X^{(n)}) \sim \text{Bin}(n,p)$ sous $\mathrm{P}_p^{(n)}$,

$$\mathrm{P}_p^{(n)}[X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t] = \frac{\mathrm{P}_p^{(n)}[X_1 = x_1, \ldots, X_n = x_n]}{\mathrm{P}_p^{(n)}[\sum_{i=1}^n X_i = t]}$$

$$= \frac{\prod_{i=1}^n \mathrm{P}_p[X_i = x_i]}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{p^t (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{1}{\binom{n}{t}}.$$

En résumé,

$$\mathrm{P}_p^{(n)}[X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t] = \frac{1}{\binom{n}{t}} \mathbb{I}\big[t = \textstyle\sum_{i=1}^n x_i\big]. \tag{III.2}$$

Donc, pour $B \in B^{(n)}$ et $t \in \mathcal{T}^{(n)}$, on obtient que

$$\mathrm{P}_p^{(n)}[X^{(n)} \in B \,|\, T(X^{(n)}) = t] = \sum_{x^{(n)} \in B} \mathrm{P}_p^{(n)}[X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t]$$

ne dépend pas de p, puisque c'est le cas des probabilités individuelles en (III.2).

Malheureusement, il est rarement possible de pouvoir utiliser, comme on l'a fait ci-dessus, la définition d'exhaustivité pour déterminer si une statistique est ou n'est pas exhaustive. Dans ce contexte, le théorème III.5 ci-dessous, qui fournit une condition nécessaire et suffisante d'exhaustivité, se révèle très utile. Ce résultat requiert le concept de fonction de vraisemblance.

Definition III.4 *Considérons le modèle paramétrique général ci-dessus. Si la distribution* $\mathrm{P}_\theta^{(n)}$ *de* $X^{(n)}$ *est discrète (respectivement, continue) pour tout* $\theta \in \Theta$*, la* fonction de vraisemblance *est définie par*

$$\theta \mapsto L_\theta^{(n)}(x^{(n)}) = \mathrm{P}_\theta^{(n)}[X^{(n)} = x^{(n)}] \quad \Big(\textit{respectivement, } \theta \mapsto L_\theta^{(n)}(x^{(n)}) = f_\theta^{X^{(n)}}(x^{(n)})\Big),$$

où $f_\theta^{X^{(n)}}(\cdot)$ *désigne la densité de* $X^{(n)}$ *sous* $\mathrm{P}_\theta^{(n)}$*. Ici,* $x^{(n)}$ *est fixé (dans l'ensemble* $\mathcal{X}_\theta^{(n)}$ *des valeurs possibles de* $X^{(n)}$ *sous* $\mathrm{P}_\theta^{(n)}$*).*

Intuitivement, la vraisemblance $L_\theta^{(n)}(x^{(n)})$ mesure l'adéquation de l'ensemble d'observations $x^{(n)}$ à la loi sous-jacente $P_\theta^{(n)}$. Plus cette valeur est grande, plus $x^{(n)}$ est en accord avec (i.e. est vraisemblable pour la loi) $P_\theta^{(n)}$.

Exemple 1 : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0,1] \subset \mathbb{R}$ (exemple électoral). La fonction de vraisemblance est alors déterminée par

$$L_p^{(n)}(x^{(n)}) = \mathrm{P}_p^{(n)}[X^{(n)} = x^{(n)}] = \mathrm{P}_p^{(n)}[X_1 = x_1, \ldots, X_n = x_n]$$

$$= \prod_{i=1}^n \mathrm{P}_p[X_i = x_i] = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \tag{III.3}$$

Exemple 2 : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $\theta = (\mu, \sigma^2)$ appartient à $\Theta = \mathbb{R} \times \mathbb{R}_0^+ \subset \mathbb{R}^2$ (exemple du verre de bière). Alors la fonction de vraisemblance est déterminée

par

$$\begin{aligned} L_\theta^{(n)}(x^{(n)}) = f_\theta^{X^{(n)}}(x^{(n)}) &= \prod_{i=1}^n f_\theta^{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\tfrac{1}{2\sigma^2}(x_i-\mu)^2\right) \\ &= \Big(\frac{1}{\sqrt{2\pi}\sigma}\Big)^n \exp\left(-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (x_i-\mu)^2\right). \end{aligned} \tag{III.4}$$

Exemple 3 : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif($[0, \theta]$), où $\theta \in \Theta = \mathbb{R}_0^+$ (exemple du bus). Dans ce contexte, la fonction de vraisemblance est donnée par

$$\begin{aligned} L_\theta^{(n)}(x^{(n)}) = f_\theta^{X^{(n)}}(x^{(n)}) &= \prod_{i=1}^n f_\theta^{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\theta}\mathbb{I}[0 \le x_i \le \theta] = \frac{1}{\theta^n}\mathbb{I}[0 \le x_1, \ldots, x_n \le \theta] \\ &= \frac{1}{\theta^n}\mathbb{I}[0 \le x_{(1)}, \ldots, x_{(n)} \le \theta] = \frac{1}{\theta^n}\mathbb{I}[x_{(1)} \ge 0]\mathbb{I}[x_{(n)} \le \theta]. \end{aligned} \tag{III.5}$$

La figure III.1 ci-dessous illustre le calcul de la vraisemblance dans l'exemple 2. Les $n = 15$ observations $x^{(n)}$ proviennent du modèle d'échantillonage paramétrique gaussien, de moyenne nulle et de variance unitaire. La vraisemblance sous $\theta = (\mu, \sigma^2) = (0, 1)$ (à gauche dans la figure) y est clairement plus élevée que sous $\theta = (\mu, \sigma^2) = (2.5, 1.3)$.

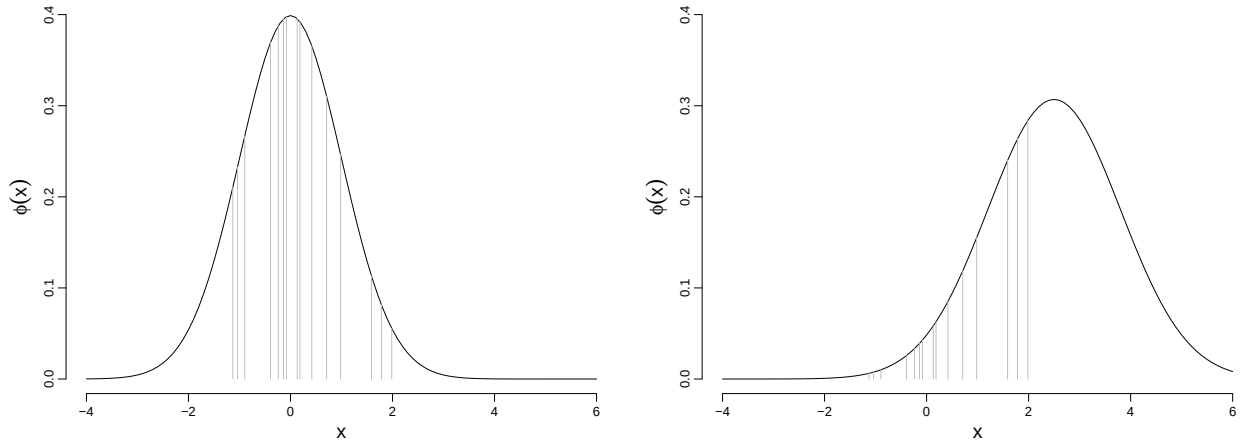


Figure III.1 – *Calcul des vraisemblances sous $\theta = (\mu, \sigma^2) = (0, 1)$ (gauche) et sous $\theta = (\mu, \sigma^2) = (2.5, 1.3)$ (droite) d'un jeu de données i.i.d. $\mathcal{N}(0, 1)$. La vraisemblance s'obtient en multipliant les longueurs des segments gris.*

Nous pouvons maintenant énoncer la condition nécessaire et suffisante d'exhaustivité annoncée.

Théorème III.5 (critère de factorisation de Neyman-Fisher) *Considérons le cadre du modèle paramétrique général considéré ci-dessus. Alors la statistique $T(X^{(n)})$ est exhaustive si et seulement si pour tout $\theta \in \Theta$, la vraisemblance $L_\theta^{(n)}$ se factorise sous la forme*

$$L_\theta^{(n)}(X^{(n)}) = g_\theta(T(X^{(n)}))\, h(X^{(n)})$$

presque sûrement sous $\mathrm{P}_\theta^{(n)}$ (au sens où $\mathrm{P}_\theta^{(n)}[\{x^{(n)} : L_\theta^{(n)}(x^{(n)}) = g_\theta(T(x^{(n)}))\, h(x^{(n)})\}] = 1$).

PREUVE. La démonstration de ce résultat dans le cas général est complexe. Considérons seulement ici le cas d'un modèle paramétrique discret.

Preuve de ($\Rightarrow$) : Soit $x^{(n)} \in \mathcal{X}^{(n)} = \{x^{(n)} \in \mathbb{R}^n : L_\theta^{(n)}(x^{(n)}) > 0\}$ un échantillon fixé. Puisque $T(X^{(n)})$ est une statistique exhaustive,

$$\begin{aligned} L_\theta^{(n)}(x^{(n)}) &= \mathrm{P}_\theta^{(n)}[X^{(n)} = x^{(n)}] = \mathrm{P}_\theta^{(n)}[[X^{(n)} = x^{(n)}] \cap [T(X^{(n)}) = T(x^{(n)}]] \\ &= \underbrace{\mathrm{P}_\theta^{(n)}[X^{(n)} = x^{(n)} | T(X^{(n)}) = T(x^{(n)})]}_{=h(x^{(n)})} \underbrace{\mathrm{P}_\theta^{(n)}[T(X^{(n)}) = T(x^{(n)})]}_{=g_\theta(T(x^{(n)}))}, \end{aligned}$$

où la fonction h ne dépend pas de θ. Les égalités étant vérifiées pour tout $x^{(n)} \in \mathcal{X}^{(n)}$, il tient que $L_\theta^{(n)}(X^{(n)}) = g_\theta(T(X^{(n)}))\, h(X^{(n)})$ avec probabilité 1.

Preuve de ($\Leftarrow$) : Supposons maintenant que $L_\theta^{(n)}(X^{(n)}) = g_\theta(T(X^{(n)}))\, h(X^{(n)})$ avec probabilité 1, i.e. il existe un ensemble $B \subset \mathcal{X}^{(n)}$ tel que $\mathrm{P}_\theta^{(n)}(B) = 1$ et, pour tout $x^{(n)} \in B$,

$$L_\theta^{(n)}(x^{(n)}) = g_\theta(T(x^{(n)}))\, h(x^{(n)}).$$

Soit $x^{(n)} \in B$ et soit $A_t = \{y^{(n)} \in \mathcal{X}^{(n)} : T(y^{(n)}) = t\}$. Alors,

$$\mathrm{P}_\theta^{(n)}[X^{(n)} = x^{(n)} | T(X^{(n)}) = t] = \frac{\mathrm{P}_\theta^{(n)}[[X^{(n)} = x^{(n)}] \cap [T(X^{(n)}) = t]]}{\mathrm{P}_\theta^{(n)}[T(X^{(n)}) = t]} = (*).$$

Si $T(x^{(n)}) \neq t$, alors cette quantité est nulle (et donc indépendante de θ). De plus, si $T(x^{(n)}) = t$, alors

$$\begin{aligned} (*) &= \frac{\mathrm{P}_\theta^{(n)}[X^{(n)} = x^{(n)}]}{\mathrm{P}_\theta^{(n)}[T(X^{(n)}) = t]} = \frac{\mathrm{P}_\theta^{(n)}[X^{(n)} = x^{(n)}]}{\sum_{y^{(n)} \in A_t} \mathrm{P}_\theta^{(n)}[X^{(n)} = y^{(n)}]} = \frac{L_\theta^{(n)}(x^{(n)})}{\sum_{y^{(n)} \in A_t} L_\theta^{(n)}(y^{(n)})} \\ &= \frac{g_\theta(T(x^{(n)}))h(x^{(n)})}{\sum_{y^{(n)} \in A_t} g_\theta(T(y^{(n)}))h(y^{(n)})} = \frac{g_\theta(t)h(x^{(n)})}{\sum_{y^{(n)} \in A_t} g_\theta(t)h(y^{(n)})} = \frac{h(x^{(n)})}{\sum_{y^{(n)} \in A_t} h(y^{(n)})}, \end{aligned}$$

qui est indépendant de θ également. Ceci prouve l'exhaustivité de la statistique $T(X^{(n)})$. □

Exemple 1 : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0, 1] \subset \mathbb{R}$ (exemple électoral). La fonction de vraisemblance en (III.3) se factorise en

$$L_p^{(n)}(x^{(n)}) = \underbrace{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}_{g_p(\sum_{i=1}^n x_i)} \times \underbrace{1}_{h(x^{(n)})},$$

ce qui montre que $T(X^{(n)}) = \sum_{i=1}^n X_i$ est exhaustive.

Exemple 2 : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $\theta = (\mu, \sigma^2)$ appartient à $\Theta = \mathbb{R} \times \mathbb{R}_0^+ \subset \mathbb{R}^2$ (exemple du verre de bière). Alors la fonction de vraisemblance en (III.4) se

réécrit

$$\begin{aligned} L_\theta^{(n)}(x^{(n)}) &= \Big(\frac{1}{\sqrt{2\pi}\sigma}\Big)^n \exp\big(-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (x_i-\mu)^2\big) \\ &= \underbrace{\Big(\frac{1}{\sqrt{2\pi}\sigma}\Big)^n \exp\big(-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n x_i^2 + \tfrac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \tfrac{n\mu^2}{2\sigma^2}\big)}_{g_{(\mu,\sigma^2)}\left(\left(\begin{array}{c}\sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2\end{array}\right)\right)} \times \underbrace{1}_{h(x^{(n)})}, \end{aligned}$$

ce qui montre que

$$T(X^{(n)}) = \begin{pmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i^2 \end{pmatrix}$$

est exhaustive. On en déduit que

$$T(X^{(n)}) = \begin{pmatrix} \bar{X} \\ s^2 \end{pmatrix}$$

est également une statistique exhaustive (en effet, il découle directement du critère de factorisation de Neyman-Fisher que si $T(X^{(n)})$ est une statistique exhaustive et H est une bijection, alors $H(T(X^{(n)}))$ est encore une statistique exhaustive).

Exemple 3 : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif$([0,\theta])$, où $\theta \in \Theta = \mathbb{R}_0^+$ (exemple du bus). Dans ce contexte, les différentes expressions obtenues pour la fonction de vraisemblance en (III.5) fournissent diverses statistiques exhaustives. Par exemple, la factorisation

$$L_\theta^{(n)}(x^{(n)}) = \underbrace{\frac{1}{\theta^n}\mathbb{I}[0 \le x_{(1)}, \ldots, x_{(n)} \le \theta]}_{g_\theta(x_{(1)},\ldots,x_{(n)})} \times \underbrace{1}_{h(x^{(n)})}$$

montre que la statistique d'ordre $T(X^{(n)}) = (X_{(1)}, \ldots, X_{(n)})$ est exhaustive (notons que celle-ci sera une statistique exhaustive dans tout modèle d'échantillonnage). Par ailleurs,

$$L_\theta^{(n)}(x^{(n)}) = \underbrace{\frac{1}{\theta^n}\mathbb{I}[x_{(1)} \ge 0]\mathbb{I}[x_{(n)} \le \theta]}_{g_\theta(x_{(1)},x_{(n)})} \times \underbrace{1}_{h(x^{(n)})}$$

indique que

$$T(X^{(n)}) = \begin{pmatrix} X_{(1)} \\ X_{(n)} \end{pmatrix}$$

est encore une statistique exhaustive. En fait, puisque

$$L_\theta^{(n)}(x^{(n)}) = \underbrace{\frac{1}{\theta^n}\mathbb{I}[x_{(n)} \le \theta]}_{g_\theta(x_{(n)})} \times \underbrace{\mathbb{I}[x_{(1)} \ge 0]}_{h(x^{(n)})},$$

$T(X^{(n)}) = X_{(n)}$ est une autre statistique exhaustive. Ceci livre donc une hiérarchie de statistiques exhaustives — à savoir, $X^{(n)}$, $(X_{(1)}, \ldots, X_{(n)})$, $(X_{(1)}, X_{(n)})$, $X_{(n)}$ — qui contiennent de moins en

moins d'information, mais portent pourtant chacune toute l'information pertinente pour conduire l'inférence sur θ. On peut montrer qu'on ne peut pas aller plus loin que $T(X^{(n)}) = X_{(n)}$ (au sens où on ne peut pas continuer à diminuer l'information tout en gardant une statistique exhaustive), et cette statistique exhaustive peut donc être considérée comme *minimale*. Plus rigoureusement, une statistique exhaustive $T(X^{(n)})$ est minimale si, pour toute autre statistique exhaustive $S(X^{(n)})$, on a $T(X^{(n)}) = \ell(S(X^{(n)}))$ pour une certaine fonction ℓ.

III.2.3 Biais

Considérons encore l'exemple du bus, c'est-à-dire celui pour lequel l'observation est donnée par $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif$([0, \theta])$, où $\theta \in \Theta = \mathbb{R}_0^+$. Nous venons de voir que l'estimateur $X_{(n)}$ est exhaustif. Néanmoins, cet estimateur livre systématiquement une sous-estimation de θ, au sens où, pour tout $\theta \in \Theta$, on a $\mathrm{P}_\theta^{(n)}[X_{(n)} < \theta] = 1$, ce qui, bien entendu, implique $\mathrm{E}_\theta^{(n)}[X_{(n)}] < \theta$. On dira que l'estimateur est *biaisé*.

Definition III.6 *Considérons le modèle statistique paramétrique général donné en (III.1). (i) L'estimateur $T(X^{(n)})$ de $g(\theta)$ est* sans biais *(ou* non biaisé*)* $\Leftrightarrow$ *pour tout $\theta \in \Theta$, $\mathrm{E}_\theta^{(n)}[T(X^{(n)})] = g(\theta)$. (ii) L'estimateur $T(X^{(n)})$ de $g(\theta)$ est* asymptotiquement sans biais *(ou* asymptotiquement non biaisé*)* $\Leftrightarrow$ *pour tout $\theta \in \Theta$, $\mathrm{E}_\theta^{(n)}[T(X^{(n)})] \to g(\theta)$ quand $n \to \infty$.*

Dans l'exemple ci-dessus, la densité de $X_{(n)}$ en (II.5) permet d'obtenir

$$\mathrm{E}_\theta^{(n)}[X_{(n)}] = \int_{-\infty}^{+\infty} x f_\theta^{X_{(n)}}(x)\, dx = \int_0^\theta x \times \frac{n x^{n-1}}{\theta^n}\, dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n\theta}{n+1}. \tag{III.6}$$

Ceci confirme que $\mathrm{E}_\theta^{(n)}[X_{(n)}] < \theta$, de sorte que l'estimateur $X_{(n)}$ de θ est effectivement biaisé. Par contre, $\mathrm{E}_\theta^{(n)}[X_{(n)}] \to \theta$ quand $n \to \infty$, ce qui montre que cet estimateur est asymptotiquement non biaisé. Dans le cas présent, on peut "débiaiser" l'estimateur en le remplaçant par

$$T(X^{(n)}) = \frac{n+1}{n} X_{(n)}.$$

À titre d'exercice, on peut vérifier que $X_{(1)} + X_{(n)}$, $(n+1)X_{(i)}/i$ et $2\bar{X}$ sont des estimateurs sans biais de θ. En particulier, l'estimateur $(n+1)X_{(1)}$ de θ est sans biais alors qu'il s'agit d'un mauvais estimateur (puisqu'on a vu qu'il n'était pas faiblement convergent !)

Definition III.7 *Dans le cadre de la définition III.6, le* biais *d'un estimateur $T(X^{(n)})$ de $g(\theta)$, sous $\mathrm{P}_\theta^{(n)}$, est le réel $b_\theta^{(n)}(T(X^{(n)})) = \mathrm{E}_\theta^{(n)}[T(X^{(n)})] - g(\theta)$.*

Clairement, l'estimateur $T(X^{(n)})$ de $g(\theta)$ est sans biais si et seulement si $b_\theta^{(n)} = 0$ pour tout $\theta \in \Theta$, et il est asymptotiquement sans biais si et seulement si $b_\theta^{(n)} \to 0$ pour tout $\theta \in \Theta$. Dans le cas de l'exemple ci-dessus, le biais de $X_{(n)}$ en θ vaut

$$b_\theta^{(n)}(X^{(n)}) = \mathrm{E}_\theta[X_{(n)}] - \theta = \frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}. \tag{III.7}$$

Le signe négatif de ce biais matérialise la sous-estimation de θ en moyenne par l'estimateur considéré. Le fait que $b_\theta^{(n)}$ tende clairement vers 0 si $n \to \infty$ confirme le non-biais asymptotique.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $\theta = (\mu, \sigma^2)$ appartient à $\Theta = \mathbb{R} \times \mathbb{R}_0^+ \subset \mathbb{R}^2$ (exemple du verre de bière). Nous avons déjà vu que $\mathrm{E}_\theta^{(n)}[\bar{X}] = \mu$, ce qui montre que $\bar{X}$ est un estimateur sans biais pour μ. Qu'en est-il de s^2 pour le problème de l'estimation de σ^2 ? Puisque

$$s^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - 2\bar{X}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) + \bar{X}^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \bar{X}^2,$$

on a $\mathrm{E}_\theta^{(n)}[s^2] = \mathrm{E}_\theta^{(n)}[X_1^2] - \mathrm{E}_\theta^{(n)}[\bar{X}^2]$. En se rappelant que $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2$ pour toute variable aléatoire Y et en utilisant le fait que $\mathrm{Var}_\theta^{(n)}[\bar{X}] = \sigma^2/n$, on obtient donc

$$\mathrm{E}_\theta^{(n)}[s^2] = \left(\sigma^2 + \mu^2\right) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{(n-1)\sigma^2}{n}\cdot \tag{III.8}$$

Par conséquent, le biais de s^2 en $\theta = (\mu, \sigma^2)$ est

$$b_\theta^{(n)} = \mathrm{E}_\theta^{(n)}[s^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = -\frac{\sigma^2}{n}\ (< 0)\cdot \tag{III.9}$$

Donc s^2 est un estimateur biaisé, mais asymptotiquement sans biais. Clairement, il découle de (III.8) que

$$S^2 \stackrel{\text{def}}{=} \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

est un estimateur sans biais de σ^2.

Nous concluons cette section avec trois remarques sur le non-biais.

1) *Le non-biais est une propriété qui ne résiste pas de façon générale aux transformations non-affines.* À titre d'illustration, considérons le problème de l'estimation de $\sigma = \sqrt{\sigma^2}$ dans l'exemple du verre de bière. Puisque S^2 est un estimateur sans biais de σ^2, on peut s'interroger sur le fait que $S = \sqrt{S^2}$ soit sans biais pour $\sigma = \sqrt{\sigma^2}$. Puisque S est une variable aléatoire non dégénérée,

$$0 < \mathrm{Var}_\theta^{(n)}[S] = \mathrm{E}_\theta^{(n)}[S^2] - \left(\mathrm{E}_\theta^{(n)}[S]\right)^2 = \sigma^2 - \left(\mathrm{E}_\theta^{(n)}[S]\right)^2,$$

ce qui livre $\left(\mathrm{E}_\theta^{(n)}[S]\right)^2 < \sigma^2$, et par suite $\mathrm{E}_\theta^{(n)}[S] < \sigma$. Nous en concluons que S est un estimateur biaisé de σ, et donc que le non-biais n'a pas résisté à la transformation non-affine $z \mapsto \sqrt{z}$. Il est par contre clair que le non-biais est maintenu sous des transformations affines, au sens où, si $T(X^{(n)})$ est un estimateur sans biais de $g(\theta)$, alors $AT(X^{(n)}) + b$ est un estimateur sans biais de $Ag(\theta) + b$.

2) *Il n'existe pas toujours un estimateur sans biais.* Considérons le problème de l'estimation de $g(p) = p^2$ dans l'exemple électoral fondé sur une unique observation X_1 de loi Bern(p), où $\theta =$

$p \in \Theta = [0,1] \subset \mathbb{R}$. Alors un estimateur $T(X_1)$ de p^2 est non biaisé si et seulement si

$$p^2 = \mathrm{E}_p[T(X_1)] = T(1)p + T(0)(1-p) \tag{III.10}$$

pour tout $p \in [0,1]$. Quels que soient $T(0)$ et $T(1)$, l'équation (III.10) en p a au plus deux solutions, ce qui exclut que cette équation soit satisfaite pour tout $p \in [0,1]$. Dans cette situation, il n'existe donc pas d'estimateur sans biais.

3) *À elle seule, la propriété de non-biais est loin de garantir que l'estimateur considéré est satisfaisant.* Considérons un modèle paramétrique d'échantillonnage gaussien à variance unitaire, pour lequel on observe $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, 1)$, avec $\theta = \mu \in \Theta = \mathbb{R}$. Considérons l'estimateur $T(X^{(n)}) = \bar{X} + Y$, où Y est une variable aléatoire indépendante de $X^{(n)}$ et prenant les valeurs possibles 10^9 et -10^9 avec probabilité $1/2$. Cet estimateur est clairement non biaisé, mais commet, avec une probabilité proche de 1, une erreur d'estimation énorme !

III.2.4 Estimateurs à erreur quadratique minimum

La dernière remarque ci-dessus montre que le biais n'est pas la seule qualité à considérer. En effet, un estimateur non-biaisé pourrait être arbitrairement loin de son objectif. Il est donc désirable qu'un estimateur soit proche de la quantité $g(\theta)$ qu'il estime (attention : un estimateur étant une quantité aléatoire, la notion d'"être proche" est elle aussi aléatoire). Le concept suivant, qui mesure l'erreur (carrée) d'estimation attendue, permet d'apprécier cet aspect.

Definition III.8 *Considérons le modèle statistique paramétrique général en (III.1) et une fonction mesurable $g : \Theta \to \mathbb{R}$ à valeurs scalaires. Alors l'*erreur quadratique moyenne*, sous $\mathrm{P}_\theta^{(n)}$, de l'estimateur $T(X^{(n)})$ de $g(\theta)$ est* $\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] = \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))^2]$.

En anglais, on parle de *mean square(d) error*, ce qui explique la notation utilisée ci-dessus. Il aurait peut-être été plus naturel de considérer l'*erreur absolue moyenne* (ou *mean absolute error*) $\mathrm{MAE}_\theta^{(n)} = \mathrm{E}_\theta^{(n)}[|T(X^{(n)}) - g(\theta)|]$. L'une des raisons pour lesquelles on préfère le MSE est l'existence de la décomposition suivante : en utilisant le fait que $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2$ pour toute variable aléatoire Y, on obtient

$$\begin{aligned}
\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] &= \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))^2] \\
&= \big(\mathrm{E}_\theta^{(n)}[T(X^{(n)}) - g(\theta)]\big)^2 + \mathrm{Var}_\theta^{(n)}[T(X^{(n)}) - g(\theta)] \\
&= \big(\mathrm{E}_\theta^{(n)}[T(X^{(n)})] - g(\theta)\big)^2 + \mathrm{Var}_\theta^{(n)}[T(X^{(n)})] \\
&= (b_\theta^{(n)}(T(X^{(n)})))^2 + \mathrm{Var}_\theta^{(n)}[T(X^{(n)})],
\end{aligned} \tag{III.11}$$

ce qui montre que le MSE d'un estimateur se décompose en termes de biais (au carré) et de variance. Puisque les deux composantes du MSE en (III.11) sont positives ou nulles, il faut, pour avoir un MSE faible, avoir à la fois un petit biais et une petite variance.

Notons que si l'estimateur $T(X^{(n)})$ (i) est asymptotiquement sans biais et (ii) a asymptotiquement une variance nulle (au sens où pour tout $\theta \in \Theta$, $\mathrm{Var}_\theta^{(n)}[T(X^{(n)})] \to 0$ quand $n \to \infty$),

alors $T(X^{(n)})$ est faiblement convergent. En effet, il découle alors directement de (III.11) que $\mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))^2] = \mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] \to 0$ quand $n \to \infty$, ou, autrement dit, que $T(X^{(n)})$ converge vers $g(\theta)$ en moyenne quadratique quand $n \to \infty$. Puisque la convergence en moyenne quadratique implique la convergence en probabilité (Théorème A.5), ceci montre bien que $T(X^{(n)})$ est faiblement convergent.

Nous avons vu ci-dessus qu'un MSE faible résulte d'un petit biais et d'une petite variance. Minimiser le biais et la variance sont cependant souvent des objectifs antagonistes. Il est classique que lorsqu'on diminue le biais, la variance augmente, et que lorsqu'on diminue la variance, le biais augmente. Dans ce cadre, minimiser le MSE permet de réaliser un équilibre entre biais et variance. Pour illustrer ceci, considérons le modèle d'échantillonnage paramétrique uniforme (exemple du bus) et les estimateurs $T_1(X^{(n)}) = X_{(n)}$ et sa version non biaisée $T_2(X^{(n)}) = (n+1)X_{(n)}/n$. Le carré du biais diminue lorsqu'on remplace $T_1(X^{(n)})$ par $T_2(X^{(n)})$ (puisque ce dernier est non-biaisé), mais la variance augmente puisque

$$\begin{aligned}
\mathrm{Var}_\theta^{(n)}[T_2(X^{(n)})] &= \mathrm{Var}_\theta^{(n)}\left[\frac{n+1}{n}T_1(X^{(n)})\right] \\
&= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}_\theta^{(n)}[T_1(X^{(n)})] \\
&> \mathrm{Var}_\theta^{(n)}[T_1(X^{(n)})].
\end{aligned}$$

C'est la valeur précise du MSE, que nous allons calculer maintenant, qui permettra d'évaluer si on a finalement gagné ou perdu en remplaçant $T_1(X^{(n)})$ par $T_2(X^{(n)})$. Puisque

$$\begin{aligned}
\mathrm{E}_\theta^{(n)}[(T_1(X^{(n)}))^2] = \mathrm{E}_\theta^{(n)}[X_{(n)}^2] &= \int_{-\infty}^{\infty} x^2 f_\theta^{X_{(n)}}(x)\,dx = \int_0^\theta x^2 \times \frac{nx^{n-1}}{\theta^n}\,dx \\
&= \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^2}{n+2},
\end{aligned}$$

on en déduit, en utilisant (III.6), que

$$\mathrm{Var}_\theta^{(n)}[T_1(X^{(n)})] = \mathrm{E}_\theta^{(n)}[X_{(n)}^2] - \left(\mathrm{E}_\theta^{(n)}[X_{(n)}]\right)^2 = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

En utilisant l'expression du biais de $T_1(X^{(n)}) = X_{(n)}$ obtenu en (III.7), on obtient donc

$$\begin{aligned}
\mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})] &= (b_\theta^{(n)}(T_1(X^{(n)})))^2 + \mathrm{Var}_\theta^{(n)}[T_1(X^{(n)})] \\
&= \left(\frac{-\theta}{n+1}\right)^2 + \frac{n\theta^2}{(n+1)^2(n+2)} \\
&= \frac{2\theta^2}{(n+1)(n+2)}.
\end{aligned} \tag{III.12}$$

Ceci est à comparer à

$$\mathrm{MSE}_\theta^{(n)}[T_2(X^{(n)})] = (b_\theta^{(n)}(T_2(X^{(n)})))^2 + \mathrm{Var}_\theta^{(n)}[T_2(X^{(n)})] = 0^2 + \left(\frac{n+1}{n}\right)^2 \mathrm{Var}_\theta^{(n)}[T_1(X^{(n)})]$$
$$= \left(\frac{n+1}{n}\right)^2 \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}. \tag{III.13}$$

Il est facile de montrer que pour tout $n \in \mathbb{N}_0$, on a

$$\mathrm{MSE}_\theta^{(n)}[T_2(X^{(n)})] \leq \mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})] \quad \text{pour tout } \theta \in \Theta \tag{III.14}$$

(et l'inégalité est stricte dès que $n \geq 2$). Dans cet exemple, on a donc gagné, en termes de MSE, à débiaiser l'estimateur $T_1(X^{(n)}) = X_{(n)}$. Autrement dit, en remplaçant $T_1(X^{(n)})$ par $T_2(X^{(n)})$, on a moins perdu en variance qu'on a gagné en biais.

Il n'en va cependant pas toujours ainsi, comme on peut le voir en considérant le problème de l'estimation de $g(\theta) = \sigma^2$ dans le modèle d'échantillonnage paramétrique gaussien. En utilisant le fait que $ns^2/\sigma^2 \sim \chi^2_{n-1}$ (Théorème II.2(ii)), on obtient

$$\mathrm{Var}_\theta^{(n)}[s^2] = \left(\frac{\sigma^2}{n}\right)^2 \mathrm{Var}_\theta^{(n)}\left[\frac{ns^2}{\sigma^2}\right] = \frac{\sigma^4}{n^2} \times 2(n-1) = \frac{2(n-1)\sigma^4}{n^2},$$

ce qui livre, au vu du biais de s^2 en (III.9),

$$\mathrm{MSE}_\theta^{(n)}[s^2] = (b_\theta^{(n)}(s^2))^2 + \mathrm{Var}_\theta^{(n)}[s^2] = \left(\frac{-\sigma^2}{n}\right)^2 + \frac{2(n-1)\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}.$$

On en conclut donc que

$$\mathrm{MSE}_\theta^{(n)}[S^2] = \mathrm{Var}_\theta^{(n)}[S^2] = \left(\frac{n}{n-1}\right)^2 \mathrm{Var}_\theta^{(n)}[s^2] = \frac{2\sigma^4}{n-1} > \mathrm{MSE}_\theta^{(n)}[s^2] \tag{III.15}$$

pour tout $n \in \mathbb{N}_0$ et tout $\theta \in \Theta$. Dans ce cas, éliminer le biais n'a donc pas été bénéfique puisque cela a résulté en une augmentation du MSE.

Lorsqu'on veut comparer des estimateurs $T_1(X^{(n)})$ et $T_2(X^{(n)})$ en termes de MSE, on ne peut le faire de façon définitive que si on a $\mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})] \leq \mathrm{MSE}_\theta^{(n)}[T_2(X^{(n)})]$ pour tout $\theta \in \Theta$ (auquel cas on préfèrera $T_1(X^{(n)})$ à $T_2(X^{(n)})$) ou si on a $\mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})] \geq \mathrm{MSE}_\theta^{(n)}[T_2(X^{(n)})]$ pour tout $\theta \in \Theta$ (auquel cas on préférera $T_2(X^{(n)})$ à $T_1(X^{(n)})$. Nous nous trouvions dans une telle situation favorable tant dans l'exemple du bus, où on a préféré $(n+1)X_{(n)}/n$ à $X_{(n)}$ (voir (III.14)), que dans l'exemple gaussien, où on a préféré s^2 à S^2 (voir (III.15)). À titre d'illustration, les deux courbes de MSE de chacun des exemples ci-dessus sont données dans la figure III.2 ci-dessous.

Si, par contre, les graphes des courbes de MSE associées aux deux estimateurs se croisent, on ne sera pas en mesure de comparer les deux estimateurs en termes de MSE. Ceci mène à la définition suivante.

Definition III.9 *Considérons le modèle statistique paramétrique général en (III.1) et une fonction*

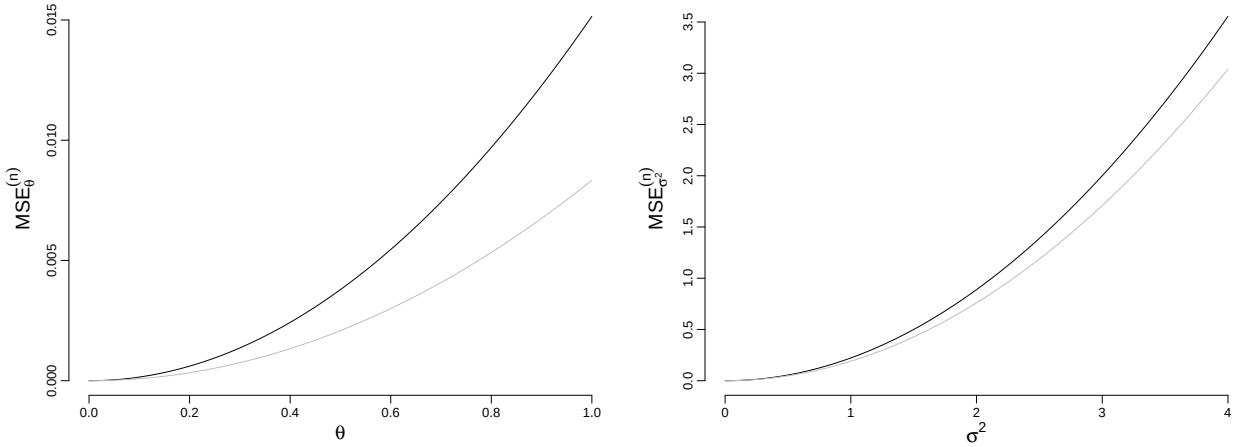


Figure III.2 – *Courbes des MSE dans l'exemple du bus (gauche) et du verre de bière (droite). Dans le premier, les courbes sont associées aux estimateurs $T_2(X^{(n)})$ (gris) et $T_1(X^{(n)})$ (noir). Dans le second, les estimateurs sont s^2 (gris) et S^2 (noir). Dans ces deux exemples, n est fixé à* 10.

mesurable $g : \Theta \to \mathbb{R}$ à valeurs scalaires. Soit $\mathcal{C}$ une classe d'estimateurs de $g(\theta)$. Alors $T_*(X^{(n)})$ est à erreur quadratique minimum dans $\mathcal{C}$ *$\Leftrightarrow$ (i) $T_*(X^{(n)}) \in \mathcal{C}$ et (ii) pour tout $T(X^{(n)}) \in \mathcal{C}$, $\mathrm{MSE}_\theta^{(n)}[T_*(X^{(n)})] \leq \mathrm{MSE}_\theta^{(n)}[T(X^{(n)})]$ pour tout $\theta \in \Theta$ (où les MSE sont relatifs à l'estimation de $g(\theta)$).*

La nécessité de se restreindre à une classe d'estimateurs est expliquée par le fait que, dans tous les cas intéressants, il n'existe pas d'estimateur à erreur quadratique minimum dans la classe de tous les estimateurs (ou en tout cas dans la classe $\mathcal{C}_0$ des estimateurs admettant un MSE fini en toute valeur du paramètre θ). Pour expliquer ceci, supposons qu'il existe un estimateur $T_*(X^{(n)})$ à erreur quadratique minimum dans $\mathcal{C}_0$. Soit alors θ_0 arbitraire dans Θ et considérons l'estimateur associé à la statistique $T_{\theta_0}(x^{(n)}) = g(\theta_0)$ pour tout $x^{(n)}$. Dans ce cas, on doit avoir

$$\mathrm{E}_\theta^{(n)}[(T_*(X^{(n)}) - g(\theta))^2] \leq \mathrm{E}_\theta^{(n)}[(T_{\theta_0}(X^{(n)}) - g(\theta))^2] = (g(\theta_0) - g(\theta))^2$$

pour tout $\theta \in \Theta$, ce qui en particulier implique $\mathrm{E}_{\theta_0}^{(n)}[(T_*(X^{(n)}) - g(\theta))^2] = 0$. Puisque θ_0 est arbitraire, on doit donc avoir

$$\mathrm{E}_\theta^{(n)}[(T_*(X^{(n)}) - g(\theta))^2] = 0$$

pour tout $\theta \in \Theta$, ce qui signifie que, pour tout $\theta \in \Theta$, on doit avoir $T_*(X^{(n)}) = g(\theta)$ presque sûrement sous $\mathrm{P}_\theta^{(n)}$ (au sens où $\mathrm{P}_\theta^{(n)}[T_*(X^{(n)}) = g(\theta)] = 1$). L'estimation de $g(\theta)$ par $T_*(X^{(n)})$ se fait donc sans erreur, ce qui est une situation idéale qui est malheureusement impossible dans tous les cas intéressants (comme le précise, dans certaines classes de modèles statistiques, la "borne de Cramér–Rao" dans la section suivante).

La raison pour laquelle il n'existe pas d'estimateurs à erreur quadratique minimum dans $\mathcal{C}_0$ est que celle-ci comprend des estimateurs déraisonnables (les estimateurs de la forme $T_{\theta_0}(X^{(n)})$), qui sont très bons pour certaines valeurs du paramètre mais très mauvais pour d'autres. La solution consiste alors à adopter un "principe statistique", qui mène à se restreindre à une classe $\mathcal{C}(\subset \mathcal{C}_0)$ d'estimateurs

qui exclut en particulier les estimateurs déraisonnables $T_{\theta_0}(X^{(n)})$ ci-dessus. Par exemple, le *principe du non-biais* préconise de se restreindre à la classe $\mathcal{C}$ des estimateurs sans biais pour $g(\theta)$.

Nous concluons cette section en étendant les concepts de cette section au cas où la quantité $g(\theta)$ à estimer est multidimensionnelle (voir par exemple l'annexe A.2 pour des détails sur les moments d'un vecteur aléatoire).

Definition III.10 *Considérons le modèle statistique paramétrique général en (III.1) et une fonction mesurable $g : \Theta \to \mathbb{R}^m$. L'*erreur quadratique moyenne*, sous $\mathrm{P}_\theta^{(n)}$, de l'estimateur $T(X^{(n)})$ de $g(\theta)$ est*

$$\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] = \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))(T(X^{(n)}) - g(\theta))'],$$

où l'espérance est calculée composante par composante.

Dans ce cadre général, le MSE d'un estimateur est une matrice réelle $m \times m$. Un estimateur est à préférer à un autre s'il a un MSE inférieur au MSE de l'autre (uniformément en θ), ce qui nous amène à considérer une relation d'ordre sur les matrices. Dans le reste du cours, on dira que la matrice réelle $m \times m$ A est inférieure ou égale à la matrice réelle $m \times m$ B ($A \leq B$) si et seulement si $B - A \geq 0$, au sens où $B - A$ est semi-définie positive (c'est-à-dire, au sens où $v'(B - A)v \geq 0$ pour tout $v \in \mathbb{R}^m$). Nous adoptons alors la définition suivante.

Definition III.11 *Considérons le modèle statistique paramétrique général en (III.1) et une fonction mesurable $g : \Theta \to \mathbb{R}^m$. Soit $\mathcal{C}$ une classe d'estimateurs de $g(\theta)$. Alors $T_*(X^{(n)})$ est à* erreur quadratique minimum dans $\mathcal{C} \Leftrightarrow$ *(i) $T_*(X^{(n)}) \in \mathcal{C}$ et (ii) pour tout $T(X^{(n)}) \in \mathcal{C}$, $\mathrm{MSE}_\theta^{(n)}[T_*(X^{(n)})] \leq \mathrm{MSE}_\theta^{(n)}[T(X^{(n)})]$ pour tout $\theta \in \Theta$ (où ces MSE sont relatifs à l'estimation de $g(\theta)$).*

Appréhender ce concept d'optimalité requiert d'interpréter l'ordre sur les MSE introduit plus haut. Pour ce faire, considérons des estimateurs $T_1(X^{(n)})$ et $T_2(X^{(n)})$ tels que

$$\mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})] \leq \mathrm{MSE}_\theta^{(n)}[T_2(X^{(n)})] \tag{III.16}$$

pour une certaine valeur de θ. Pour tout $v \in \mathbb{R}^m$, ceci signifie que

$$v'(\mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})] - \mathrm{MSE}_\theta^{(n)}[T_2(X^{(n)})])v \leq 0.$$

Puisque

$$\begin{aligned} v'\mathrm{MSE}_\theta^{(n)}[T_1(X^{(n)})]v &= \mathrm{E}_\theta^{(n)}[v'(T(X^{(n)}) - g(\theta))(T(X^{(n)}) - g(\theta))'v] \\ &= \mathrm{E}_\theta^{(n)}[(v'(T(X^{(n)}) - g(\theta)))^2] \\ &= \mathrm{E}_\theta^{(n)}[(v'T(X^{(n)}) - v'g(\theta))^2], \end{aligned}$$

ceci se réécrit

$$\mathrm{MSE}_\theta^{(n)}[v'T_1(X^{(n)})] \leq \mathrm{MSE}_\theta^{(n)}[v'T_2(X^{(n)})]$$

pour tout $v \in \mathbb{R}^m$, où les MSE sont relatifs à l'estimation de $v'g(\theta)$. La domination en termes de

MSE multivarié en (III.16) implique donc que, quel que soit $v \in \mathbb{R}^m$, l'estimateur $v'T_1(X^{(n)})$ de $v'g(\theta)$ domine l'estimateur $v'T_2(X^{(n)})$ en termes de MSE univarié.

III.2.5 Estimateurs efficaces

Puisqu'on a vu qu'il est désirable d'utiliser des estimateurs qui ont une erreur quadratique moyenne aussi petite que possible, on peut se demander jusqu'où on peut aller en termes de MSE. On peut répondre à cette question dans le cadre des modèles statistiques paramétriques qui sont réguliers au sens de la définition suivante.

Definition III.12 *Considérons le modèle statistique paramétrique général en (III.1) et notons $L_\theta^{(n)}(\cdot)$ sa fonction de vraisemblance. Alors ce modèle est* régulier *si et seulement si les hypothèses suivantes sont satisfaites :*

(H1) Θ *est ouvert.*
(H2) $\mathcal{X}^{(n)} = \{x^{(n)} \in \mathbb{R}^n : L_\theta^{(n)}(x^{(n)}) > 0\}$ *ne dépend pas de* θ*.*
(H3) Pour tout $x^{(n)} \in \mathcal{X}^{(n)}$*, la fonction* $\theta \mapsto L_\theta^{(n)}(x^{(n)})$ *est différentiable sur* Θ*.*
(H4) L'expression $\int_{\mathcal{X}^{(n)}} L_\theta^{(n)}(x^{(n)})\,dx^{(n)}$ *est dérivable sous le signe, au sens où*

$$\nabla_\theta \int_{\mathcal{X}^{(n)}} L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \int_{\mathcal{X}^{(n)}} \nabla_\theta L_\theta^{(n)}(x^{(n)})\,dx^{(n)}.$$

(H5) Pour tout $\theta \in \Theta$*, la matrice réelle* $k \times k$

$$I^{(n)}(\theta) \stackrel{\text{def}}{=} \int_{\mathcal{X}^{(n)}} \nabla_\theta \ln L_\theta^{(n)}(x^{(n)})(\nabla_\theta \ln L_\theta^{(n)}(x^{(n)}))'\, L_\theta^{(n)}(x^{(n)})\,dx^{(n)}$$

existe, est finie et est inversible.

Dans cette définition, nous avons utilisé la notation intégrale correspondant au cas continu, mais la définition couvre aussi le cas discret, pour lequel il faudra lire les intégrales comme des sommes finies ou des séries, où on somme sur toutes les valeurs possibles $x^{(n)}$ de $X^{(n)}$.

Notons que l'hypothèse (H2) écarte l'exemple du modèle d'échantillonnage uniforme (exemple du bus), puisque le support de la vraisemblance $L_\theta^{(n)}(\cdot)$ est $\mathcal{X}_\theta^{(n)} = [0,\theta] \times \ldots \times [0,\theta]$ (n fois), qui dépend de θ. On pourra par contre vérifier que les modèles d'échantillonnage concernant l'exemple électoral et l'exemple du verre de bière sont réguliers.

Remarquons qu'il découle de (H4) que

$$\begin{aligned}\mathrm{E}_\theta^{(n)}[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})] &= \int_{\mathcal{X}^{(n)}} \big(\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\big)\, L_\theta^{(n)}(x^{(n)})\,dx^{(n)} \\ &= \int_{\mathcal{X}^{(n)}} \nabla_\theta L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \nabla_\theta \int_{\mathcal{X}^{(n)}} L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \nabla_\theta 1 = 0. \qquad \text{(III.17)}\end{aligned}$$

La matrice $I^{(n)}(\theta)$ apparaissant dans (H5), qui est connue sous le nom d'*information de Fisher en* θ,

se réécrit dès lors

$$I^{(n)}(\theta) = \mathrm{E}_\theta^{(n)}[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}))'] = \mathrm{Var}_\theta^{(n)}[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})]. \tag{III.18}$$

Si le modèle considéré est un modèle d'échantillonnage, on a $L_\theta^{(n)}(x^{(n)}) = \prod_{i=1}^n L_\theta^{(1)}(x_i)$ (où $x^{(n)} = (x_1, \ldots, x_n)$), ce qui mène à

$$\begin{aligned} I^{(n)}(\theta) &= \mathrm{Var}_\theta^{(n)}[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})] = \mathrm{Var}_\theta^{(n)}\left[\sum_{i=1}^n \nabla_\theta \ln L_\theta^{(1)}(X_i)\right] \\ &= \sum_{i=1}^n \mathrm{Var}_\theta^{(n)}[\nabla_\theta \ln L_\theta^{(1)}(X_i)] = n\mathrm{Var}_\theta^{(n)}[\nabla_\theta \ln L_\theta^{(1)}(X_1)] = nI^{(1)}(\theta), \end{aligned} \tag{III.19}$$

ce qu'on lira en disant que *l'information contenue dans un échantillon de taille n est égale à n fois l'information contenue dans une unique observation.* Pour interpréter $I^{(n)}(\theta)$ comme une mesure d'information, considérons la situation particulière pour laquelle $k = 1$ et $I^{(n)}(\theta) = 0$ pour tout θ (ce qui est bien entendu incompatible avec l'hypothèse (H5)). Dans ce cas, il découle de (III.17)–(III.18) que, pour tout θ,

$$\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}) = 0 \text{ presque sûrement sous } \mathrm{P}_\theta^{(n)},$$

ce qui implique que

$$\nabla_\theta L_\theta^{(n)}(X^{(n)}) = 0 \text{ presque sûrement sous } \mathrm{P}_\theta^{(n)}.$$

Par conséquent, la fonction de vraisemblance $\theta \mapsto L_\theta^{(n)}(x^{(n)})$ est constante (presque partout en $x^{(n)}$ pour tout θ), ce qui implique que les probabilités

$$\mathrm{P}_\theta^{(n)}[B] = \int_B L_\theta^{(n)}(x^{(n)})\, dx^{(n)}, \quad B \in \mathcal{B}^n,$$

ne dépendent pas de θ, de sorte que les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ coïncident. Dans une pareille situation (violant l'hypothèse d'injectivité de la paramétrisation), il est impossible de conduire l'inférence sur θ, et on peut donc considérer que $X^{(n)}$ ne porte pas d'information sur θ.

Nous considérons maintenant le concept d'estimateur régulier.

Definition III.13 *Considérons le modèle statistique paramétrique général en (III.1) et notons $L_\theta^{(n)}(\cdot)$ sa fonction de vraisemblance. Soit $T(X^{(n)})$ un estimateur de $g(\theta)$, où $g : \Theta \to \mathbb{R}^m$ est mesurable. Alors cet estimateur est* régulier *si et seulement si les hypothèses suivantes sont satisfaites :*

(H1) $\mathrm{E}_\theta^{(n)}[\|T(X^{(n)})\|^2]$ *existe et est finie pour tout $\theta \in \Theta$.*

(H2) L'expression $\psi(\theta) \overset{\text{def}}{=} \mathrm{E}_\theta^{(n)}[T(X^{(n)})] = \int_{\mathcal{X}^{(n)}} T(x^{(n)})\, L_\theta^{(n)}(x^{(n)})\, dx^{(n)}$ *est dérivable sous le signe, au sens où*

$$\nabla_\theta \int_{\mathcal{X}^{(n)}} T(x^{(n)})\, L_\theta^{(n)}(x^{(n)})\, dx^{(n)} = \int_{\mathcal{X}^{(n)}} T(x^{(n)})\, \nabla_\theta L_\theta^{(n)}(x^{(n)})\, dx^{(n)}.$$

Nous pouvons maintenant énoncer le résultat principal de cette section.

Théorème III.14 (Borne de Cramér–Rao) *Supposons que le modèle statistique paramétrique donné en (III.1) soit régulier. Soit $T(X^{(n)})$ un estimateur régulier de $g(\theta)$, où $g : \Theta \to \mathbb{R}^m$ est mesurable. Alors, pour tout $\theta \in \Theta$,*

$$\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] = \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))(T(X^{(n)}) - g(\theta))'] \geq \Delta_\theta (I^{(n)}(\theta))^{-1} \Delta_\theta',$$

où $\Delta_\theta = (\frac{\partial \psi_i}{\partial \theta_j}(\theta))_{i=1,\ldots,m,\, j=1,\ldots,k}$ est la matrice jacobienne de $\psi(\cdot)$ et où "$\geq$" est l'ordre associé aux matrices semi-définies positives.

Preuve. Notons d'abord que

$$\begin{aligned}
(\Delta_\theta)_{ij} &= \frac{\partial}{\partial \theta_j} \int_{\mathcal{X}^{(n)}} T_i(x^{(n)})\, L_\theta^{(n)}(x^{(n)})\, dx^{(n)} \\
&= \int_{\mathcal{X}^{(n)}} T_i(x^{(n)})\, \frac{\partial}{\partial \theta_j} L_\theta^{(n)}(x^{(n)})\, dx^{(n)} \\
&= \int_{\mathcal{X}^{(n)}} T_i(x^{(n)}) \left(\frac{\partial}{\partial \theta_j} \ln L_\theta^{(n)}(x^{(n)}) \right) L_\theta^{(n)}(x^{(n)})\, dx^{(n)} \\
&= \mathrm{E}_\theta[T_i(X^{(n)}) \tfrac{\partial}{\partial \theta_j} \ln L_\theta^{(n)}(X^{(n)})],
\end{aligned}$$

ce qui livre

$$\Delta_\theta = \mathrm{E}_\theta[T(X^{(n)})(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}))']. \tag{III.20}$$

Nous allons établir le résultat en montrant que

$$\begin{aligned}
\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] &= \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))(T(X^{(n)}) - g(\theta))'] \\
&= \mathrm{Var}_\theta^{(n)}[T(X^{(n)})] + (\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))' && \text{(III.21)} \\
&\geq \mathrm{Var}_\theta^{(n)}[T(X^{(n)})] && \text{(III.22)} \\
&\geq \Delta_\theta (I^{(n)}(\theta))^{-1} \Delta_\theta'. && \text{(III.23)}
\end{aligned}$$

L'égalité (III.21) découle du fait que

$$\begin{aligned}
&\mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))(T(X^{(n)}) - g(\theta))'] \\
&\quad = \mathrm{E}_\theta^{(n)}[\{(T(X^{(n)}) - \psi(\theta)) + (\psi(\theta) - g(\theta))\}\{(T(X^{(n)}) - \psi(\theta)) + (\psi(\theta) - g(\theta))\}'] \\
&\quad = \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - \psi(\theta))(T(X^{(n)}) - \psi(\theta))'] + \mathrm{E}_\theta^{(n)}[T(X^{(n)}) - \psi(\theta)](\psi(\theta) - g(\theta))' \\
&\qquad + (\psi(\theta) - g(\theta))(\mathrm{E}_\theta^{(n)}[T(X^{(n)}) - \psi(\theta)])' + \mathrm{E}_\theta^{(n)}[(\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))'] \\
&\quad = \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - \psi(\theta))(T(X^{(n)}) - \psi(\theta))'] + (\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))' \\
&\quad = \mathrm{Var}_\theta^{(n)}[T(X^{(n)})] + (\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))',
\end{aligned}$$

où on a utilisé l'identité $\mathrm{E}_\theta^{(n)}[T(X^{(n)})] = \psi(\theta)$ et le fait que $(\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))'$ est une matrice constante.

Par définition de l'ordre considéré sur les matrices, l'inégalité (III.22) revient à montrer que $(\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))'$ est une matrice semi-définie positive, ce qui est bien le cas puisque, pour tout $v \in \mathbb{R}^m$, $v'\big[(\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))'\big]v = \big[v'(\psi(\theta) - g(\theta))\big]^2 \geq 0$.

Il ne reste donc qu'à prouver (III.23). Pour ce faire, posons

$$S_\theta(X^{(n)}) = T(X^{(n)}) - \Delta_\theta(I^{(n)}(\theta))^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}).$$

Notons qu'au vu de (III.17), $\mathrm{E}_\theta^{(n)}[S_\theta(X^{(n)})] = \mathrm{E}_\theta^{(n)}[T(X^{(n)})] = \psi(\theta)$. Donc

$$\begin{aligned}
\mathrm{Var}_\theta^{(n)}[S_\theta(X^{(n)})] &= \mathrm{E}_\theta^{(n)}[S_\theta(X^{(n)})(S_\theta(X^{(n)}))'] - \psi(\theta)(\psi(\theta))' \\
&= \mathrm{E}_\theta^{(n)}[\{T(X^{(n)}) - \Delta_\theta(I^{(n)}(\theta))^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\} \\
&\qquad \{(T(X^{(n)}))' - (\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}))'(I^{(n)}(\theta))^{-1}\Delta_\theta'\}] - \psi(\theta)(\psi(\theta))' \\
&= \mathrm{E}_\theta^{(n)}[T(X^{(n)})(T(X^{(n)}))'] - \psi(\theta)(\psi(\theta))' \\
&\quad -\mathrm{E}_\theta^{(n)}[T(X^{(n)})(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}))'(I^{(n)}(\theta))^{-1}\Delta_\theta'] \\
&\quad -\mathrm{E}_\theta^{(n)}[\Delta_\theta(I^{(n)}(\theta))^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})(T(X^{(n)}))'] \\
&\quad +\mathrm{E}_\theta^{(n)}[\Delta_\theta(I^{(n)}(\theta))^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}))'(I^{(n)}(\theta))^{-1}\Delta_\theta'].
\end{aligned}$$

En utilisant (III.20), on obtient alors

$$\begin{aligned}
\mathrm{Var}_\theta^{(n)}[S_\theta(X^{(n)})] &= \mathrm{Var}_\theta^{(n)}[T(X^{(n)})] \\
&\quad -\Delta_\theta(I^{(n)}(\theta))^{-1}\Delta_\theta' \\
&\quad -\Delta_\theta(I^{(n)}(\theta))^{-1}\Delta_\theta' \\
&\quad +\Delta_\theta(I^{(n)}(\theta))^{-1}I^{(n)}(\theta)(I^{(n)}(\theta))^{-1}\Delta_\theta' \\
&= \mathrm{Var}_\theta^{(n)}[T(X^{(n)})] - \Delta_\theta(I^{(n)}(\theta))^{-1}\Delta_\theta'.
\end{aligned}$$

L'inégalité (III.23) découle donc de l'inégalité $\mathrm{Var}_\theta^{(n)}[S_\theta(X^{(n)})] \geq 0$. □

Dans un modèle d'échantillonnage régulier, (III.19) montre que la borne de Cramér–Rao se réécrit

$$\frac{1}{n}\,\Delta_\theta(I^{(1)}(\theta))^{-1}\Delta_\theta',$$

ce qui montre que l'erreur quadratique moyenne d'un estimateur (régulier) tend vers 0 au mieux à la vitesse $1/n$. Ceci confirme que le modèle d'échantillonnage uniforme (exemple du bus) et l'estimateur $X_{(n)}$ de θ ne rentrent pas dans le cadre régulier considéré ci-dessus puisque nous avons vu

en (III.12) que

$$\mathrm{MSE}_\theta^{(n)}[X_{(n)}] = \frac{2\theta^2}{(n+1)(n+2)},$$

qui tend vers 0 à la vitesse $1/n^2$. Si on considère l'estimation de $g(\theta) = \theta$ (ce qui mène à $m = k$) dans un modèle paramétrique régulier général (pas nécessairement un modèle d'échantillonnage) et que l'estimateur $T(X^{(n)})$ est sans biais pour θ ($\psi(\theta) = \theta$), alors la borne de Cramér–Rao prend la forme

$$(I^{(n)}(\theta))^{-1},$$

ce qui s'interprète naturellement : plus grande est l'information, meilleure est la précision d'un estimateur $T(X^{(n)})$ optimal.

Bien entendu, la borne de Cramér–Rao fournit une cible naturelle de performance d'estimation dans les modèles réguliers.

Definition III.15 *Dans la situation considérée au théorème III.14, on dira que l'estimateur* $T(X^{(n)})$ *est* efficace *pour* $g(\theta)$ *si et seulement si* $\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] = \mathrm{E}_\theta^{(n)}[(T(X^{(n)}) - g(\theta))(T(X^{(n)}) - g(\theta))'] = \Delta_\theta (I^{(n)}(\theta))^{-1} \Delta_\theta'$ *pour tout* $\theta \in \Theta$.

Un estimateur est donc efficace si son erreur quadratique moyenne coïncide avec la borne de Cramér–Rao pour tout $\theta \in \Theta$. Clairement, ceci a lieu si et seulement si les inégalités en (III.22)–(III.23) sont des égalités, c'est-à-dire si et seulement si (i) $\psi(\cdot) = g(\cdot)$ et (ii) $\mathrm{Var}_\theta^{(n)}[T(X^{(n)})] = \Delta_\theta (I^{(n)}(\theta))^{-1} \Delta_\theta'$ pour tout $\theta \in \Theta$. On en conclut qu'un estimateur $T(X^{(n)})$ ne peut être efficace que pour l'estimation de $g(\theta) = \psi(\theta)$ (autrement dit, le non-biais est une condition nécessaire d'efficacité). Ceci permet de reformuler la définition ci-dessus.

Definition III.16 *Dans la situation considérée au théorème III.14, on dira que l'estimateur* $T(X^{(n)})$ *est* efficace (*automatiquement pour sa moyenne* $\psi(\theta)$) *si et seulement si*

$$\mathrm{Var}_\theta^{(n)}[T(X^{(n)})] = \Delta_\theta (I^{(n)}(\theta))^{-1} \Delta_\theta'$$

pour tout $\theta \in \Theta$.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0,1] \subset \mathbb{R}$ (exemple électoral). Nous avons vu que la vraisemblance est donnée par $L_p^{(n)}(x^{(n)}) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$. Par conséquent, $\ln L_p^{(n)}(x^{(n)}) = \big(\sum_{i=1}^n x_i\big) \ln p + \big(n - \sum_{i=1}^n x_i\big) \ln(1-p)$, ce qui mène à

$$\frac{d}{dp} \ln L_p^{(n)}(x^{(n)}) = \big(\textstyle\sum_{i=1}^n x_i\big) \dfrac{1}{p} - \big(n - \textstyle\sum_{i=1}^n x_i\big) \dfrac{1}{1-p}.$$

Donc l'information de Fisher $I^{(n)}(p)$ est donnée par

$$\begin{aligned} I^{(n)}(p) &= \mathrm{Var}_p^{(n)}\Big[\frac{d}{dp}\ln L_p^{(n)}(X^{(n)})\Big] = \mathrm{Var}_p^{(n)}\Big[\big(\textstyle\sum_{i=1}^n X_i\big)\frac{1}{p} - \big(n - \sum_{i=1}^n X_i\big)\frac{1}{1-p}\Big] \\ &= \mathrm{Var}_p^{(n)}\Big[\Big(\frac{1}{p} + \frac{1}{1-p}\Big)\big(\textstyle\sum_{i=1}^n X_i\big)\Big] = \frac{1}{p^2(1-p)^2}\mathrm{Var}_p^{(n)}\Big[\textstyle\sum_{i=1}^n X_i\Big] \\ &= \frac{n}{p^2(1-p)^2}\mathrm{Var}_p^{(n)}\big[X_1\big] = \frac{n}{p(1-p)}. \end{aligned}$$

Un estimateur efficace de p doit donc être un estimateur de p qui est sans biais et dont la variance est égale à $(I^{(n)}(p))^{-1} = p(1-p)/n$. On vérifiera facilement que $T(X^{(n)}) = \bar{X}$ est donc efficace pour p.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma_0^2)$, où σ_0^2 est un réel strictement positif fixé, de sorte que le paramètre est $\theta = \mu \in \Theta = \mathbb{R}$ (version à variance connue de l'exemple du verre de bière). La vraisemblance est donnée par

$$L_\mu^{(n)}(x^{(n)}) = \Big(\frac{1}{\sqrt{2\pi}\sigma_0}\Big)^n \exp\big(-\tfrac{1}{2\sigma_0^2}\textstyle\sum_{i=1}^n (x_i-\mu)^2\big),$$

ce qui mène à

$$\frac{d}{d\mu}\ln L_\mu^{(n)}(x^{(n)}) = \frac{d}{d\mu}\left[n\ln\Big(\frac{1}{\sqrt{2\pi}\sigma_0}\Big) - \frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu)^2\right] = \frac{1}{\sigma_0^2}\sum_{i=1}^n (x_i-\mu).$$

Donc l'information de Fisher est

$$\begin{aligned} I^{(n)}(\mu) &= \mathrm{Var}_\mu^{(n)}\Big[\frac{d}{d\mu}\ln L_\mu^{(n)}(X^{(n)})\Big] = \mathrm{Var}_\mu^{(n)}\left[\frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i-\mu)\right] \\ &= \frac{1}{\sigma_0^4}\mathrm{Var}_\mu^{(n)}\left[\sum_{i=1}^n (X_i-\mu)\right] = \frac{n}{\sigma_0^4}\mathrm{Var}_\mu^{(n)}\big[X_1-\mu\big] = \frac{n}{\sigma_0^2}. \end{aligned}$$

Un estimateur efficace de μ doit donc être un estimateur de μ qui est sans biais et dont la variance est égale à $(I^{(n)}(\mu))^{-1} = \sigma_0^2/n$. Par conséquent, $T(X^{(n)}) = \bar{X}$ est efficace pour μ.

III.3 Méthodes d'estimation

Jusqu'ici, nous avons considéré des critères nous permettant de comparer différents estimateurs pour un modèle statistique donné. Néanmoins, en pratique, personne ne fournit ces estimateurs et il est donc nécessaire de pouvoir construire de tels estimateurs. Dans cette section, nous présentons deux méthodes générales permettant d'obtenir des estimateurs raisonnables.

III.3.1 Méthode des moments

Considérons de nouveau un modèle statistique paramétrique d'échantillonnage

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{ \mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k \right\}\right),$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. La méthode des moments permet d'estimer le paramètre θ pour autant que la fonction

$$\begin{array}{rcl} M : \Theta & \to & M(\Theta) \\ \theta & \mapsto & \begin{pmatrix} \mu'_1(\theta) \\ \vdots \\ \mu'_k(\theta) \end{pmatrix}, \end{array}$$

où $\mu'_r(\theta) = \mathrm{E}_\theta^{(n)}[X_1^r]$ est le moment non centré d'ordre r, soit bien définie (moments d'ordre k finis) et *inversible* ; voir la section II.1.2. Si la valeur des moments $\mu'_r(\theta)$, $r = 1, \ldots, k$, était connue, l'hypothèse d'inversibilité permettrait alors d'en déduire l'unique valeur de θ qui correspond à ces moments.

En pratique, la valeur des moments $\mu'_r(\theta)$, $r = 1, \ldots, k$, est inconnue, mais peut être approximée par les moments empiriques correspondants, c'est-à-dire par les quantités $m'_r = \frac{1}{n}\sum_{i=1}^n X_i^r$, $r = 1, \ldots, k$. Rappelons en effet que dans un modèle d'échantillonnage, et sous les hypothèses de moments finis d'ordre k considérées, la loi forte des grands nombres assure que $m'_r \overset{\text{p.s.}}{\to} \mu'_r(\theta)$ sous $\mathrm{P}_\theta^{(n)}$ pour tout $r = 1, \ldots, k$, de sorte qu'à moins que n soit petit, les moments empiriques m'_r fournissent une bonne approximation des moments population $\mu'_r(\theta)$ correspondants.

La méthode des moments consiste alors à considérer l'estimateur $T(X^{(n)})$ qui est la solution en θ du système

$$\left\{ \begin{array}{c} \mu'_1(\theta) = m'_1 \\ \vdots \\ \mu'_k(\theta) = m'_k. \end{array} \right.$$

L'hypothèse d'inversibilité assure l'existence d'une solution unique, pour autant que $(m'_1, \ldots, m'_k) \in M(\Theta)$, ce qui, si $\theta \in \mathrm{int}(\Theta)$, sera le cas avec une probabilité arbitrairement proche de 1 pour n suffisamment grand, au vu de la loi (faible) des grands nombres.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0, 1] \subset \mathbb{R}$ (exemple électoral). Puisque $\mu'_1(p) = \mathrm{E}_p^{(n)}[X_1] = p$, l'estimateur $T(X^{(n)})$ des moments de p est la solution en p de l'équation

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

Il s'agit donc de l'estimateur $T(X^{(n)}) = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ (que nous avons vu être exhaustif et efficace pour p).

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $\theta = (\mu, \sigma^2)$ appartient à $\Theta = \mathbb{R} \times \mathbb{R}_0^+ \subset \mathbb{R}^2$ (exemple du verre de bière). Puisque $\mu_1'(\theta) = \mathrm{E}_\theta^{(n)}[X_1] = \mu$ et $\mu_2'(\theta) = \mathrm{E}_\theta^{(n)}[X_1^2] = \mathrm{Var}_\theta^{(n)}[X_1] + (\mathrm{E}_\theta^{(n)}[X_1])^2 = \sigma^2 + \mu^2$, l'estimateur $T(X^{(n)})$ des moments de θ est la solution en θ du système d'équations

$$\left\{ \begin{array}{c} \mu = \frac{1}{n}\sum_{i=1}^n X_i \\ \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 \end{array} \right. ,$$

laquelle est donnée par

$$T(X^{(n)}) = \left(\begin{array}{c} \bar{X} \\ \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \bar{X}^2 \end{array} \right) = \left(\begin{array}{c} \bar{X} \\ s^2 \end{array} \right).$$

Cet estimateur de θ est exhaustif et *asymptotiquement efficace* (en un sens qui sera précisé à la fin de ce chapitre).

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathrm{Unif}([0, \theta])$, où $\theta \in \Theta = \mathbb{R}_0^+$ (exemple du bus). Comme $\mu_1'(\theta) = \mathrm{E}_\theta^{(n)}[X_1] = \theta/2$, l'estimateur $T(X^{(n)})$ des moments de θ est la solution en θ de l'équation

$$\frac{\theta}{2} = \frac{1}{n}\sum_{i=1}^n X_i.$$

Il s'agit donc de l'estimateur $T(X^{(n)}) = 2\bar{X}$. Celui-ci n'est pas exhaustif... Et est en fait loin d'être à erreur quadratique minimum pour n grand. Ceci résulte du fait que

$$\mathrm{MSE}_\theta^{(n)}[T(X^{(n)})] = \frac{4}{n}\mathrm{Var}_\theta^{(n)}[X_1] = \frac{c_\theta}{n}$$

(pour une certaine constante $c_\theta > 0$) tend vers 0 à la vitesse $1/n$, alors que, comme on l'a vu, l'erreur quadratique moyenne associée à l'estimateur $X_{(n)}$ tend vers 0 à la vitesse $1/n^2$.

D'une manière générale, les estimateurs des moments sont fortement convergents, sous les hypothèses considérées. Par contre, il n'y a aucune garantie qu'ils soient (même pas asymptotiquement) efficaces. Le dernier exemple va dans ce sens, même si, à strictement parler, on ne peut pas parler d'*efficacité* dans cet exemple (puisque le modèle considéré n'est pas régulier).

III.3.2 Méthode du maximum de vraisemblance

La seconde méthode d'estimation considérée est celle du *maximum de vraisemblance*. Considérons le modèle statistique paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\right\}\right),$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$, et notons $L_\theta^{(n)}(\cdot)$ la fonction de vraisemblance associée. Essentiellement, la méthode du maximum de vraisemblance consiste à estimer θ par "la" valeur $T(X^{(n)})$ de θ qui, à $X^{(n)}$ fixé, maximise la vraisemblance $L_\theta^{(n)}(X^{(n)})$. Parce qu'il n'y a aucune garantie quant à l'unicité, nous adoptons la

définition suivante.

Definition III.17 *Dans le cadre du modèle paramétrique général considéré ci-dessus, la statistique $T(X^{(n)})$ à valeurs dans Θ est un estimateur du maximum de vraisemblance de θ si et seulement si, pour tout $\theta \in \Theta$,*

$$L^{(n)}_{T(X^{(n)})}(X^{(n)}) \geq L^{(n)}_{\theta}(X^{(n)}) \quad \textit{presque sûrement sous } \mathrm{P}^{(n)}_{\theta},$$

au sens où $\mathrm{P}^{(n)}_{\theta}[\{x^{(n)} : L^{(n)}_{T(x^{(n)})}(x^{(n)}) \geq L^{(n)}_{\theta}(x^{(n)})\}] = 1$.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif$([0, \theta])$, où $\theta \in \Theta = \mathbb{R}^+_0$ (exemple du bus). Nous avons vu que la vraisemblance correspondante est déterminée par

$$L^{(n)}_{\theta}(x^{(n)}) = \frac{1}{\theta^n} \mathbb{I}[x_{(1)} \geq 0] \mathbb{I}[x_{(n)} \leq \theta].$$

Par conséquent, on vérifie facilement que $T(X^{(n)}) = X_{(n)}$ est un estimateur du maximum de vraisemblance de θ.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif$([\theta - \frac{1}{2}, \theta + \frac{1}{2}])$, où $\theta \in \Theta = \mathbb{R}$. La vraisemblance associée est telle que

$$\begin{aligned} L^{(n)}_{\theta}(x^{(n)}) &= \prod_{i=1}^{n} f^{X_i}_{\theta}(x_i) = \prod_{i=1}^{n} \mathbb{I}[\theta - \tfrac{1}{2} \leq x_i \leq \theta + \tfrac{1}{2}] = \mathbb{I}[\theta - \tfrac{1}{2} \leq x_1, \ldots, x_n \leq \theta + \tfrac{1}{2}] \\ &= \mathbb{I}[\theta - \tfrac{1}{2} \leq x_{(1)}, x_{(n)} \leq \theta + \tfrac{1}{2}] = \mathbb{I}[x_{(n)} - \tfrac{1}{2} \leq \theta \leq x_{(1)} + \tfrac{1}{2}]. \end{aligned}$$

Par conséquent, toute statistique $T(X^{(n)})$ qui, avec probabilité 1 (sous n'importe quel $\mathrm{P}^{(n)}_{\theta}$), prend ses valeurs dans l'intervalle $[X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}]$ est un estimateur du maximum de vraisemblance pour θ. En particulier, $X_{(n)} - \frac{1}{2}$, $X_{(1)} + \frac{1}{2}$ et $(X_{(1)} + X_{(n)})/2$ sont tous les trois des estimateurs du maximum de vraisemblance pour θ, ce qui confirme qu'il n'y a pas toujours unicité. Notons cependant que cette non-unicité ne pose pas réellement problème puisque pour tout θ, la longueur de l'intervalle $[X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}]$ tend vers 0 en probabilité sous $\mathrm{P}^{(n)}_{\theta}$ quand $n \to \infty$.

Revenons au cadre général de la définition III.17 et supposons que $S(X^{(n)})$ est une statistique exhaustive. Par le critère de factorisation de Neyman-Fisher (théorème III.5), on a que, pour tout $\theta \in \Theta$,

$$L^{(n)}_{\theta}(X^{(n)}) = g_{\theta}(S(X^{(n)}))\, h(X^{(n)})$$

presque sûrement sous $\mathrm{P}^{(n)}_{\theta}$. Par conséquent, une statistique $T(X^{(n)})$ maximise la vraisemblance $L^{(n)}_{\theta}(X^{(n)})$ (au sens de la définition III.17 ci-dessus) si et seulement si $T(X^{(n)})$ maximise $g_{\theta}(S(X^{(n)}))$ (au sens correspondant). On en déduit qu'un estimateur du maximum du vraisemblance $T(X^{(n)})$ ne fait intervenir $X^{(n)} = (X_1, \ldots, X_n)$ qu'à travers une statistique exhaustive $S(X^{(n)})$: plus précisément, quelle que soit la statistique exhaustive $S(X^{(n)})$, il existe une fonction m telle que $T(X^{(n)}) = m(S(X^{(n)}))$. Ceci est illustré par l'exemple du bus ci-dessus, pour lequel on a vu que $X_{(n)}$ est une statistique exhaustive.

Une situation privilégiée intéressante est celle où, pour tout $x^{(n)}$, la vraisemblance $\theta \mapsto L_\theta^{(n)}(x^{(n)})$ est différentiable. Dans ce cas, on peut utiliser les outils habituels du calcul différentiel pour identifier l'estimateur du maximum de vraisemblance. En particulier, toute valeur du paramètre θ (appartenant à l'intérieur de Θ) maximisant la vraisemblance doit être solution du système d'équations

$$\nabla_\theta L_\theta^{(n)}(X^{(n)}) = 0.$$

Parce qu'il est plus commode de dériver des sommes que des produits (dans un modèle d'échantillonnage, toute vraisemblance est un produit de n facteurs), on considèrera plutôt le système d'équations

$$\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}) = 0, \tag{III.24}$$

qui admet les mêmes solutions que le système ci-dessus. Les équations en (III.24) forment un système de k équations à k inconnues et sont appelées les *équations de vraisemblance*. En pratique, on veillera à s'assurer que les solutions de ce système correspondent bien à des maxima globaux de la vraisemblance et, au besoin, à étudier séparément le bord de Θ.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0,1] \subset \mathbb{R}$ (exemple électoral). Puisque $L_p^{(n)}(X^{(n)}) = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i}$, nous obtenons

$$\frac{d}{dp} \ln L_p^{(n)}(X^{(n)}) = \big(\textstyle\sum_{i=1}^n X_i\big)\dfrac{1}{p} - \big(n - \textstyle\sum_{i=1}^n X_i\big)\dfrac{1}{1-p}.$$

Les équations de vraisemblance prennent donc la forme

$$\big(\textstyle\sum_{i=1}^n X_i\big)\dfrac{1}{p} - \big(n - \textstyle\sum_{i=1}^n X_i\big)\dfrac{1}{1-p} = 0,$$

et on vérifie facilement que leur seule solution en p est donnée par

$$T(X^{(n)}) = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

À titre d'exercice, on peut vérifier (soigneusement, en tenant compte notamment du fait que cet estimateur est parfois à valeur dans le bord de Θ) que $T(X^{(n)}) = \bar{X}$ est un estimateur du maximum de vraisemblance au sens de la définition III.17.

Ci-dessus, nous n'avons considéré l'estimation par maximum de vraisemblance que du paramètre θ lui-même et pas d'une transformation $g(\theta)$ du paramètre. On peut étendre l'estimation du maximum de vraisemblance à ce cadre en procédant de la façon suivante. Dans un premier temps, supposons que la fonction $g : \Theta \mapsto g(\Theta)$ est bijective. Dans ce cas, θ et $\tilde{\theta} = g(\theta)$ fournissent deux "adressages" valides du modèle statistique considéré et les deux familles de vraisemblances qui en résultent sont liées par

$$L_\theta^{(n)}(X^{(n)}) = L_{g^{-1}(g(\theta))}^{(n)}(X^{(n)}) = L_{g^{-1}(\tilde{\theta})}^{(n)}(X^{(n)}) = \tilde{L}_{\tilde{\theta}}^{(n)}(X^{(n)}),$$

où la notation $\tilde{L}^{(n)}$ est utilisée pour la vraisemblance associée à la paramétrisation par $\tilde{\theta}$. Si la

vraisemblance $L_\theta^{(n)}(X^{(n)})$ est maximisée pour la valeur $T(X^{(n)})$ de θ (au sens de la définition III.17), la vraisemblance $\tilde{L}_{\tilde{\theta}}^{(n)}(X^{(n)})$ le sera pour la valeur $g(T(X^{(n)}))$ de $\tilde{\theta}$. Par conséquent, l'estimateur du maximum de vraisemblance de $g(\theta)$ est $g(T(X^{(n)}))$, où $T(X^{(n)})$ est l'estimateur du maximum de vraisemblance de θ. Considérons maintenant le cas où $g : \Theta \to g(\Theta)$ n'est pas bijective. Puisque la fonction

$$\begin{array}{rcl} G : \Theta & \to & G(\Theta) \\ \theta & \mapsto & \begin{pmatrix} g(\theta) \\ \theta \end{pmatrix} \end{array}$$

est bijective, on en déduit que l'estimateur du maximum de vraisemblance de $G(\theta)$ est $G(T(X^{(n)}))$, où $T(X^{(n)})$ désigne encore l'estimateur du maximum de vraisemblance de θ. Se restreindre à la première composante de $G(\cdot)$ mène alors à considérer que l'estimateur du maximum de vraisemblance de $g(\theta)$ est ici aussi $g(T(X^{(n)}))$. On retiendra que l'estimateur du maximum de vraisemblance d'un paramètre transformé est la transformation de l'estimateur du maximum de vraisemblance du paramètre original. On parle d'*invariance du maximum de vraisemblance.*

III.3.3 Comportement asymptotique des estimateurs du maximum de vraisemblance

Comme nous l'avons vu, il n'y a aucune garantie que la méthode des moments mène à des estimateurs qui soient bons en termes d'efficacité. À l'inverse, sous certaines hypothèses de régularité, la méthode du maximum de vraisemblance livre des estimateurs qui sont *asymptotiquement efficaces* (en un sens qui sera précisé plus bas).

Théorème III.18 *Considérons un modèle statistique paramétrique d'échantillonnage pour lequel le paramètre est scalaire* ($\Theta \subset \mathbb{R}$) *et notons* $L_\theta^{(1)}(\cdot)$ *la fonction de vraisemblance associée à un échantillon de taille* 1. *Supposons que les hypothèses suivantes sont satisfaites :*

(H1) Θ *est ouvert.*
(H2) $\mathcal{X} = \{x \in \mathbb{R} : L_\theta^{(1)}(x) > 0\}$ *ne dépend pas de* θ.
(H3) Pour tout $x \in \mathcal{X}$, *la fonction* $\theta \mapsto L_\theta^{(1)}(x)$ *est dérivable deux fois sur* Θ.
(H4) L'expression $\int_{\mathcal{X}} L_\theta^{(1)}(x)\,dx$ *est dérivable deux fois sous le signe, au sens où*

$$\frac{d^r}{d\theta^r}\int_{\mathcal{X}} L_\theta^{(1)}(x)\,dx = \int_{\mathcal{X}} \frac{d^r}{d\theta^r} L_\theta^{(1)}(x)\,dx, \quad r = 1, 2.$$

(H5) La fonction $\theta \mapsto B_\theta^{(1)}(x) \stackrel{\text{def}}{=} \frac{d^2}{d\theta^2} \ln L_\theta^{(1)}(x)$ *est continue uniformément en* x, *au sens où, pour toute suite* $(\theta_n) \to \theta$, *on a* $\sup_{x\in\mathcal{X}} |B_{\theta_n}^{(1)}(x) - B_\theta^{(1)}(x)| \to 0$.
(H6) Pour tout $\theta \in \Theta$, *le réel*

$$I^{(1)}(\theta) \stackrel{\text{def}}{=} \int_{\mathcal{X}} \Big(\frac{d}{d\theta} \ln L_\theta^{(1)}(x)\Big)^2 L_\theta^{(1)}(x)\,dx$$

existe et appartient à $(0, \infty)$.

Alors (i) il existe une suite $(T(X^{(n)}))$ de solutions des équations de vraisemblance telle que, pour tout $\theta \in \Theta$, $T(X^{(n)}) \overset{\text{p.s.}}{\to} \theta$ sous $\mathrm{P}_\theta^{(n)}$ quand $n \to \infty$. (ii) Pour toute telle suite $(T(X^{(n)}))$, on a que, pour tout $\theta \in \Theta$,

$$\sqrt{n}(T(X^{(n)}) - \theta) \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \frac{1}{I^{(1)}(\theta)}\right)$$

sous $\mathrm{P}_\theta^{(n)}$ quand $n \to \infty$.

Pour n grand, il découle de ce résultat que, sous $\mathrm{P}_\theta^{(n)}$, $T(X^{(n)})$ est à peu près d'espérance θ et de variance $1/(nI^{(1)}(\theta)) = 1/I^{(n)}(\theta)$. Puisqu'un estimateur efficace pour θ, à n fixé, est exactement d'espérance θ et de variance $1/I^{(n)}(\theta)$, on peut donc considérer que ce résultat établit que l'estimateur du maximum de vraisemblance est *asymptotiquement efficace*.

La preuve de (i) est assez technique et dépasse le cadre de ce cours. Nous allons donc nous concentrer sur la preuve de (ii), en supposant le résultat en (i) acquis.

Preuve de (ii). Dans cette preuve, nous fixons une valeur $\theta \in \Theta$ et toutes les convergences stochastiques seront sous $\mathrm{P}_\theta^{(n)}$ quand $n \to \infty$.

Rappelons que la structure d'échantillonnage implique que $L_\theta^{(n)}(x^{(n)}) = L_\theta^{(n)}(x_1, \ldots, x_n) = \prod_{i=1}^n L_\theta^{(1)}(x_i)$. Par conséquent,

$$A_\theta^{(n)}(x^{(n)}) \overset{\text{def}}{=} \frac{1}{n}\frac{d}{d\theta} \ln L_\theta^{(n)}(x^{(n)}) = \frac{1}{n}\sum_{i=1}^n \frac{d}{d\theta} \ln L_\theta^{(1)}(x_i) = \frac{1}{n}\sum_{i=1}^n A_\theta^{(1)}(x_i)$$

et, de façon similaire,

$$B_\theta^{(n)}(x^{(n)}) \overset{\text{def}}{=} \frac{1}{n}\frac{d^2}{d\theta^2} \ln L_\theta^{(n)}(x^{(n)}) = \frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2} \ln L_\theta^{(1)}(x_i) = \frac{1}{n}\sum_{i=1}^n B_\theta^{(1)}(x_i).$$

Puisque $(T^{(n)}) \overset{\text{not}}{=} (T(X^{(n)}))$ est une suite de solutions des équations de vraisemblance, on a que, pour tout n,

$$A_{T^{(n)}}^{(n)}(X^{(n)}) = 0$$

presque sûrement sous $\mathrm{P}_\theta^{(n)}$. La formule des accroissement finis assure alors que

$$\begin{aligned} 0 = A_{T^{(n)}}^{(n)}(X^{(n)}) &= A_\theta^{(n)}(X^{(n)}) + \frac{d}{d\theta}A_\theta^{(n)}(X^{(n)})\Big|_{T_*^{(n)}}\big(T^{(n)} - \theta\big) \\ &= A_\theta^{(n)}(X^{(n)}) + B_{T_*^{(n)}}^{(n)}(X^{(n)})\big(T^{(n)} - \theta\big), \end{aligned}$$

où $T_*^{(n)}$ est entre θ et $T^{(n)}$. Par conséquent, on a

$$\sqrt{n}\big(T^{(n)} - \theta\big) = \frac{-\sqrt{n}A_\theta^{(n)}(X^{(n)})}{B_{T_*^{(n)}}^{(n)}(X^{(n)})}.$$

Comme $|T_*^{(n)} - \theta| \leq |T^{(n)} - \theta|$, le point (i) du théorème implique que $T_*^{(n)} \overset{\text{p.s.}}{\to} \theta$. Donc l'hypothèse (H5) implique que

$$\begin{aligned}
\left|B^{(n)}_{T_*^{(n)}}(X^{(n)}) - B^{(n)}_{\theta}(X^{(n)})\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left(B^{(1)}_{T_*^{(n)}}(X_i) - B^{(1)}_{\theta}(X_i)\right)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|B^{(1)}_{T_*^{(n)}}(X_i) - B^{(1)}_{\theta}(X_i)\right| \\
&\leq \sup_{x\in\mathcal{X}}\left|B^{(1)}_{T_*^{(n)}}(x) - B^{(1)}_{\theta}(x)\right| \\
&\overset{\text{p.s.}}{\to} 0.
\end{aligned}$$

Par conséquent, la loi forte des grands nombres et le théorème A.7 livrent

$$\begin{aligned}
B^{(n)}_{T_*^{(n)}}(X^{(n)}) &= B^{(n)}_{\theta}(X^{(n)}) + \left(B^{(n)}_{T_*^{(n)}}(X^{(n)}) - B^{(n)}_{\theta}(X^{(n)})\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} B^{(1)}_{\theta}(X_i) + \left(B^{(n)}_{T_*^{(n)}}(X^{(n)}) - B^{(n)}_{\theta}(X^{(n)})\right) \\
&\overset{\text{p.s.}}{\to} \mathrm{E}_\theta\big[B^{(1)}_{\theta}(X_1)\big] + 0 = \mathrm{E}_\theta\big[B^{(1)}_{\theta}(X_1)\big].
\end{aligned}$$

En particulier, par le théorème A.5, on a montré que

$$B^{(n)}_{T_*^{(n)}}(X^{(n)}) \overset{\mathcal{L}}{\to} \mathrm{E}_\theta\big[B^{(1)}_{\theta}(X_1)\big]. \tag{III.25}$$

Dériver par rapport à θ les deux membres de l'égalité $1 = \int_{\mathcal{X}} L^{(1)}_{\theta}(x)\,dx$ (qui tient pour tout θ) fournit

$$0 = \int_{\mathcal{X}} \frac{d}{d\theta} L^{(1)}_{\theta}(x)\,dx = \int_{\mathcal{X}} \left(\frac{d}{d\theta}\ln L^{(1)}_{\theta}(x)\right) L^{(1)}_{\theta}(x)\,dx = \mathrm{E}_\theta[A^{(1)}_{\theta}(X_1)].$$

Le théorème central limite implique alors directement que

$$-\sqrt{n}A^{(n)}_{\theta}(X^{(n)}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(-A^{(1)}_{\theta}(X_i)\big) \overset{\mathcal{L}}{\to} \mathcal{N}\Big(0, I^{(1)}(\theta)\Big), \tag{III.26}$$

où on a utilisé l'identité $\mathrm{Var}_\theta[A^{(1)}_{\theta}(X_i)] = I^{(1)}(\theta)$, qui découle de la définition de $I^{(1)}(\theta)$. Au vu

de (III.25)–(III.26), le lemme de Slutzky (le théorème A.8) livre

$$\begin{aligned}\sqrt{n}\big(T^{(n)}-\theta\big) &= \frac{-\sqrt{n}A_\theta^{(n)}(X^{(n)})}{B^{(n)}_{T_*^{(n)}}(X^{(n)})}\\ &\xrightarrow{\mathcal{L}} \frac{1}{\mathrm{E}_\theta\big[B_\theta^{(1)}(X_i)\big]}\,\mathcal{N}\Big(0, I^{(1)}(\theta)\Big)\\ &= \mathcal{N}\bigg(0, \frac{I^{(1)}(\theta)}{\big(\mathrm{E}_\theta\big[B_\theta^{(1)}(X_1)\big]\big)^2}\bigg).\end{aligned} \tag{III.27}$$

En dérivant une nouvelle fois par rapport à θ l'égalité

$$0=\int_{\mathcal{X}}\left(\frac{d}{d\theta}\ln L_\theta^{(1)}(x)\right)L_\theta^{(1)}(x)\,dx$$

qui vient d'être obtenue (et qui tient aussi pour tout θ), on obtient

$$\begin{aligned}0 &= \int_{\mathcal{X}}\left(\left(\frac{d^2}{d\theta^2}\ln L_\theta^{(1)}(x)\right)L_\theta^{(1)}(x)+\left(\frac{d}{d\theta}\ln L_\theta^{(1)}(x)\right)\frac{d}{d\theta}L_\theta^{(1)}(x)\right)dx\\ &= \mathrm{E}_\theta\left[\frac{d^2}{d\theta^2}\ln L_\theta^{(1)}(X_1)\right]+\int_{\mathcal{X}}\left(\frac{d}{d\theta}\ln L_\theta^{(1)}(x)\right)^2 L_\theta^{(1)}(x)\,dx\\ &= \mathrm{E}_\theta\big[B_\theta^{(1)}(X_1)\big]+I^{(1)}(\theta),\end{aligned}$$

ce qui montre que $\mathrm{E}_\theta\big[B_\theta^{(1)}(X_1)\big]=-I^{(1)}(\theta)$. Nous avons donc montré que

$$\big(\mathrm{E}_\theta\big[B_\theta^{(1)}(X_1)\big]\big)^2=\big(I^{(1)}(\theta)\big)^2,$$

ce qui, au vu de (III.27), termine la démonstration. □

Chapitre IV

Tests d'hypothèses

IV.1 Terminologie et concepts de base

Considérons un modèle statistique paramétrique de la forme

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\right\}\right),$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Dans ce chapitre, on ne veut plus parier sur la valeur inconnue du paramètre θ, mais seulement décider si $\theta \in \Theta_0$ ou si $\theta \in \Theta_1$, où Θ_0 et $\Theta_1 (= \Theta \setminus \Theta_0)$ forment une partition de l'espace paramétrique Θ. Bien entendu, cette décision doit être prise sur la base de $X^{(n)}$.

IV.1.1 Problèmes de test

La terminologie, dans ce cadre, est très structurée. Le problème considéré ci-dessus est le *problème de test d'hypothèses*

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1. \end{cases}$$

On dira qu'il s'agit de tester l'*hypothèse nulle* $\mathcal{H}_0$ contre la *contre-hypothèse* $\mathcal{H}_1$. Deux décisions seulement sont possibles : *rejeter l'hypothèse nulle* $\mathcal{H}_0$ et *ne pas rejeter l'hypothèse nulle* $\mathcal{H}_0$, que nous noterons respectivement par $R\mathcal{H}_0$ et $\not{R}\mathcal{H}_0$ dans la suite. On ne dira pas *accepter* $\mathcal{H}_0$, ce qui est en ligne avec le fait que, comme nous le verrons, $\mathcal{H}_0$ et $\mathcal{H}_1$ jouent des rôles asymétriques. Habituellement, on prend pour $\mathcal{H}_1$ la partie de l'espace paramétrique dont on veut éventuellement établir la "véracité" (car une procédure de test ne sera vraiment conclusive que lorsqu'on prend la décision $R\mathcal{H}_0$).

Dans l'exemple du verre de bière, l'étudiant-consommateur pourra tenter de savoir si, comme le promet le barman, la quantité de bière moyenne servie dans un verre est bien supérieure ou égale à 15 cl, ou si au contraire ce n'est pas le cas (auquel cas les autorités devront être contactées au plus vite !). Comme discuté dans l'introduction, le modèle d'échantillonnage paramétrique gaussien (où

on observe les quantités de bière $X_1, \ldots, X_n$ qui sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, avec $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_0^+$) constitue un modèle raisonnable pour cette situation. La question, à laquelle on répondra sur base de $X^{(n)} = (X_1, \ldots, X_n)$, mène alors au problème de test

$$\begin{cases} \mathcal{H}_0 : \theta = (\mu, \sigma^2) \in \Theta_0 = \{(\mu, \sigma^2) \in \Theta : \mu \geq 15\} \\ \mathcal{H}_1 : \theta = (\mu, \sigma^2) \in \Theta_1 = \{(\mu, \sigma^2) \in \Theta : \mu < 15\}, \end{cases}$$

qu'on écrira de façon plus informelle

$$\begin{cases} \mathcal{H}_0 : \mu \geq 15 \\ \mathcal{H}_1 : \mu < 15. \end{cases} \tag{IV.1}$$

Considérons l'exemple électoral, où l'observation $X^{(n)} = (X_1, \ldots, X_n)$ a des composantes X_i qui sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0, 1] \subset \mathbb{R}$. Pour une valeur fixée $p_0 \in (0, 1)$, les problèmes de test

$$\begin{cases} \mathcal{H}_0 : p \leq p_0 \\ \mathcal{H}_1 : p > p_0 \end{cases} \quad \text{et} \quad \begin{cases} \mathcal{H}_0 : p \geq p_0 \\ \mathcal{H}_1 : p < p_0 \end{cases} \tag{IV.2}$$

seront dits *unilatéraux* car la partie de l'espace paramétrique associée à $\mathcal{H}_1$ est située d'un seul côté de celle correspondant à $\mathcal{H}_0$. On peut bien entendu aussi considérer le problème *bilatéral* correspondant, qui est donné par

$$\begin{cases} \mathcal{H}_0 : p = p_0 \\ \mathcal{H}_1 : p \neq p_0 \end{cases} \tag{IV.3}$$

(ce dernier problème serait considéré, par exemple, si on voulait tester qu'une pièce de monnaie est équilibrée, auquel cas on prendrait $p_0 = 0.5$).

On dira que l'hypothèse nulle $\mathcal{H}_0$ est *simple* si et seulement si elle est associée à une unique mesure de probabilité du modèle statistique considéré (dans le cas paramétrique, ceci sera le cas si et seulement si Θ_0 est un singleton). Dans le cas inverse, elle sera dite *composée*. De la même façon, on parlera aussi de contre-hypothèse $\mathcal{H}_1$ simple ou composée.

Dans le cas général où le modèle statistique considéré n'est pas nécessairement paramétrique, c'est la collection de lois $\mathcal{P}^{(n)} = \{\mathrm{P}^{(n)}\}$ qu'il convient de partitionner pour définir une hypothèse nulle $\mathcal{H}_0$ et une contre-hypothèse $\mathcal{H}_1$. Bien entendu, dans le cas paramétrique, ceci est strictement équivalent à partitionner l'espace paramétrique Θ lui-même, comme on l'a fait ci-dessus (mais les puristes considéreront dans tous les cas une partition de la collection de lois $\mathcal{P}^{(n)}$, même dans le cas paramétrique).

IV.1.2 Tests, erreurs, risques, puissance

Formellement, la décision de rejeter ou de ne pas rejeter l'hypothèse nulle sera prise au moyen d'une procédure statistique appelée un *test*. Puisqu'un test doit fonder sa décision sur les observations $x^{(n)} = (x_1, \ldots, x_n)$, on adopte la définition suivante.

Definition IV.1 *Un* test pur ϕ *est une fonction (mesurable) de* $\mathcal{X}^{(n)}$ *dans* $\{0, 1\}$*, où* $\mathcal{X}^{(n)}$ *désigne l'ensemble des valeurs possibles pour* $x^{(n)}$*.*

Le codage retenu classiquement est que si $\phi(x^{(n)}) = 0$, alors le test conduit à la décision $\not{R}\mathcal{H}_0$ lorsque $x^{(n)}$ est observé. Si au contraire, le $x^{(n)}$ qui est observé est tel que $\phi(x^{(n)}) = 1$, alors le test conduit à la décision $R\mathcal{H}_0$. Dans ce cadre, la collection de tous les $x^{(n)}$ menant au rejet, c'est-à-dire $\phi^{-1}(\{1\})$, est appelée la *zone critique* du test. Par exemple, des tests raisonnables pour le premier problème en (IV.2) sont les tests tels que

$$\phi_c(x^{(n)}) = \mathbb{I}\big[\textstyle\sum_{i=1}^n x_i > c\big],$$

où c est un nombre fixé. Dans la suite, nous serons aussi amenés à considérer des tests *randomisés*.

Definition IV.2 *Un* test randomisé ϕ *est une fonction (mesurable) de* $\mathcal{X}^{(n)}$ *dans* $[0,1]$.

La lecture de la valeur prise par un test randomisé est la suivante. Si $x^{(n)}$ est observé, alors le test ϕ conduit à la décision $R\mathcal{H}_0$ avec probabilité $\phi(x^{(n)})$ sur la base d'un aléa indépendant du modèle statistique considéré (en d'autres termes, on rejette si une pièce de monnaie donnant face avec probabilité $\phi(x^{(n)})$ donne effectivement face). Notons que si $\phi(x^{(n)}) = 1$, on est donc amené à prendre la décision $R\mathcal{H}_0$ (avec probabilité 1), tandis que si $\phi(x^{(n)}) = 0$, on est amené à prendre la décision $\not{R}\mathcal{H}_0$ (avec probabilité 1) ; lorsque $\phi(x^{(n)}) \in \{0,1\}$, la décision prise par un test randomisé est donc en ligne avec celle prise par un test pur. Ce n'est que quand $x^{(n)}$ appartient à la *zone de randomisation* $\phi^{-1}((0,1))$ qu'on sera amené à fonder la décision finale sur le jet d'une pièce de monnaie. Pour un test randomisé, on continuera à appeler $\phi^{-1}(\{1\})$ la *zone critique* du test.

Indépendamment du test utilisé, des erreurs de deux types peuvent être commises lorsqu'on applique [1] un test. La première erreur, connue sous le nom d'*erreur de première espèce*, consiste à prendre la décision $R\mathcal{H}_0$ alors que $\mathcal{H}_0$ est vraie. La seconde, connue sous le nom d'*erreur de seconde espèce*, consiste à prendre la décision $\not{R}\mathcal{H}_0$ alors que $\mathcal{H}_0$ est fausse. Les différents cas de figure sont représentés de façon synthétique dans la table suivante.

	$\not{R}\mathcal{H}_0$	$R\mathcal{H}_0$
$\mathcal{H}_0$ est vraie	OK !	Erreur de première espèce
$\mathcal{H}_0$ est fausse	Erreur de seconde espèce	OK !

La qualité d'un test ϕ est mesurée par son *risque de première espèce* et son *risque de seconde espèce*, qui sont les probabilités respectives que ce test commette une erreur de première espèce ou de seconde espèce. Plus précisément, nous adoptons la définition suivante.

Definition IV.3 *Soit* ϕ *un test (pur ou randomisé) pour le problème de test* $\mathcal{H}_0 : \theta \in \Theta_0$ *contre* $\mathcal{H}_1 : \theta \in \Theta_1$. *Alors (i) le* risque de première espèce *de* ϕ *en* $\theta(\in \Theta_0)$ *est* $\mathrm{P}_\theta^{(n)}[\phi$ *mène à* $R\mathcal{H}_0] = \mathrm{E}_\theta^{(n)}[\phi]$. *(ii) Le* risque de seconde espèce *de* ϕ *en* $\theta(\in \Theta_1)$ *est* $\mathrm{P}_\theta^{(n)}[\phi$ *mène à* $\not{R}\mathcal{H}_0] = 1 - \mathrm{E}_\theta^{(n)}[\phi]$.

La réécriture de ces risques en termes d'espérance découle du fait que la formule des probabilités

1. Notons au passage ici l'utilisation du verbe "appliquer" plutôt que, par exemple, effectuer. Un test d'hypothèses est une fonction.

totales livre

$$\mathrm{P}_\theta^{(n)}[\phi \text{ mène à } R\mathcal{H}_0] = \int_{\mathcal{X}^{(n)}} \mathrm{P}[\phi \text{ mène à } R\mathcal{H}_0 \,|\, X^{(n)} = x^{(n)}]\, L_\theta^{(n)}(x^{(n)})\, dx^{(n)}$$

$$= \int_{\mathcal{X}^{(n)}} \phi(x^{(n)})\, L_\theta^{(n)}(x^{(n)})\, dx^{(n)} = \mathrm{E}_\theta^{(n)}[\phi(X^{(n)})] \overset{\text{not}}{=} \mathrm{E}_\theta^{(n)}[\phi],$$

où $L_\theta^{(n)}(x^{(n)})$ désigne la vraisemblance associée à $\mathrm{P}_\theta^{(n)}$. En lien avec les risques de première et seconde espèces d'un test, nous introduisons aussi la terminologie suivante.

Definition IV.4 *Soit ϕ un test (pur ou randomisé) pour le problème de test $\mathcal{H}_0 : \theta \in \Theta_0$ contre $\mathcal{H}_1 : \theta \in \Theta_1$. Alors la* dimension *de ϕ en $\theta(\in \Theta)$ est $\mathrm{P}_\theta^{(n)}[\phi$ mène à $R\mathcal{H}_0] = \mathrm{E}_\theta^{(n)}[\phi]$. Dans le cas particulier où $\theta \in \Theta_1$, on parlera de* puissance *de ϕ en θ.*

Un test satisfaisant est donc un test qui a de faibles risques de première et seconde espèces, ou de façon équivalente, un faible risque de première espèce et une grande puissance. Néanmoins, il est clair que minimiser le risque de première espèce et maximiser la puissance sont des objectifs antagonistes : pour minimiser le risque de première espèce, il faut rejeter moins souvent $\mathcal{H}_0$, ce qui s'obtient en définissant une zone critique réduite. Au contraire, pour maximiser la puissance, il faut que le test considéré rejette plus souvent $\mathcal{H}_0$, ce qui est obtenu en prenant une zone critique étendue.

IV.1.3 Le principe de Neyman

La manière la plus classique de résoudre cet antagonisme (et d'obtenir un problème mathématique pour lequel il est possible de trouver un test optimal) est d'adopter le *principe de Neyman*, qui consiste à maximiser la puissance (de manière équivalente, à minimiser le risque de seconde espèce) sous la contrainte que le risque de première espèce vaut au plus α, pour un certain nombre $\alpha \in (0, 1)$ fixé. Plus précisément, on tentera de maximiser la puissance (uniformément en $\theta(\in \Theta_1)$) dans la classe des tests *de niveau α*.

Definition IV.5 *Soit ϕ un test (pur ou randomisé) pour le problème de test $\mathcal{H}_0 : \theta \in \Theta_0$ contre $\mathcal{H}_1 : \theta \in \Theta_1$. Alors ϕ est* de niveau $\alpha(\in (0, 1))$ *si et seulement si $\mathrm{E}_\theta^{(n)}[\phi] \leq \alpha$ pour tout $\theta \in \Theta_0$.*

Les *niveaux* les plus classiquement utilisés en pratique sont $\alpha = 0.05$ et $\alpha = 0.01$ (les plus timorés peuvent considérer des niveaux encore plus petits). Le principe de Neyman énoncé ci-dessus mène alors à des tests qui sont optimaux en le sens suivant.

Definition IV.6 *Le test ϕ_* est à* puissance uniformément maximale (PUM) dans la classe $\mathcal{C}_\alpha$ des tests de niveau $\alpha \Leftrightarrow$ *(i) $\phi_* \in \mathcal{C}_\alpha$ et (ii) pour tout $\phi \in \mathcal{C}_\alpha$, $\mathrm{E}_\theta^{(n)}[\phi_*] \geq \mathrm{E}_\theta^{(n)}[\phi]$ pour tout $\theta \in \Theta_1$.*

Notre objectif est de construire des tests optimaux en le sens ci-dessus pour divers problèmes de test d'intérêt. Dans de nombreuses situations, cette construction s'appuie sur le résultat fondamental présenté dans la section suivante.

IV.2 Le lemme fondamental de Neyman–Pearson

Considérons le cadre le plus simple, où l'hypothèse nulle et la contre-hypothèse sont toutes deux *simples*, au sens défini plus haut. Plus spécifiquement, nous considérons le modèle statistique associé à une observation $X^{(n)} = (X_1, \ldots, X_n)$ de loi $\mathrm{P}_\theta^{(n)}$, où $\theta \in \Theta = \{\theta_0, \theta_1\}$, ainsi que le problème de test

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1. \end{cases} \tag{IV.4}$$

Ici, la classe $\mathcal{C}_\alpha$ des tests de niveau α est donc celle des tests ϕ tels que $\mathrm{E}_{\theta_0}^{(n)}[\phi] \leq \alpha$. Le résultat suivant décrit explicitement comment construire un test à puissance (uniformément) maximale dans $\mathcal{C}_\alpha$.

Théorème IV.7 (Lemme fondamental de Neyman–Pearson) *Considérons le problème de test (IV.4) et fixons $\alpha \in (0,1)$. Désignons par $L_\theta^{(n)}(x^{(n)})$ la vraisemblance associée à $\mathrm{P}_\theta^{(n)}$. Alors (i) il existe $k_\alpha \in \mathbb{R}^+$ et $\gamma_\alpha \in [0,1]$ tels que le test ϕ_* défini par*

$$\phi_*(x^{(n)}) = \begin{cases} 1 & si \quad L_{\theta_1}^{(n)}(x^{(n)}) > k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ \gamma_\alpha & si \quad L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ 0 & si \quad L_{\theta_1}^{(n)}(x^{(n)}) < k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \end{cases} \tag{IV.5}$$

vérifie $\mathrm{E}_{\theta_0}[\phi_*] = \alpha$; (ii) ce test ϕ_* est à puissance (uniformément) maximale dans $\mathcal{C}_\alpha$.

Preuve. (i) Désignons par $F_{\theta_0}(\cdot)$ la fonction de répartition, sous $\mathrm{P}_{\theta_0}^{(n)}$, de la variable aléatoire $L_{\theta_1}^{(n)}(X^{(n)})/L_{\theta_0}^{(n)}(X^{(n)})$, où on prend les conventions $c/0 = +\infty$ pour tout $c > 0$ et $0/0 = 0$ (on pourrait en fait donner n'importe quelle autre valeur à $0/0$). En d'autres termes, $F_{\theta_0}(\cdot)$ est définie par

$$F_{\theta_0}(z) = \mathrm{P}_{\theta_0}^{(n)}\left[\frac{L_{\theta_1}^{(n)}(X^{(n)})}{L_{\theta_0}^{(n)}(X^{(n)})} \leq z\right]$$

pour tout $z \in \mathbb{R}$. Ceci permet d'écrire

$$\begin{aligned} \mathrm{E}_{\theta_0}^{(n)}[\phi_*] &= \mathrm{P}_{\theta_0}^{(n)}\big[L_{\theta_1}^{(n)}(X^{(n)}) > k_\alpha\, L_{\theta_0}^{(n)}(X^{(n)})\big] + \gamma_\alpha\, \mathrm{P}_{\theta_0}^{(n)}\big[L_{\theta_1}^{(n)}(X^{(n)}) = k_\alpha\, L_{\theta_0}^{(n)}(X^{(n)})\big] \\ &= 1 - F_{\theta_0}(k_\alpha) + \gamma_\alpha\,(F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - 0)), \end{aligned}$$

où $F_{\theta_0}(k_\alpha - 0) \stackrel{\text{def}}{=} \lim_{z \stackrel{<}{\to} k_\alpha} F_{\theta_0}(z)$. Nous devons considérer deux cas.

(1) Supposons qu'il existe z_α tel que $F_{\theta_0}(z_\alpha) = 1 - \alpha$. Dans ce cas, si on prend $k_\alpha \stackrel{\text{def}}{=} z_\alpha$, qui est bien positif ou nul (pourquoi ?) et $\gamma_\alpha \stackrel{\text{def}}{=} 0$, on a bien $\mathrm{E}_{\theta_0}^{(n)}[\phi_*] = 1 - F_{\theta_0}(k_\alpha) = \alpha$.

(2) Supposons que $F_{\theta_0}(z) \neq 1 - \alpha$ pour tout $z \in \mathbb{R}$. Dans ce cas, posons

$$k_\alpha \stackrel{\text{def}}{=} \inf\Big\{z \in \mathbb{R} : F_{\theta_0}(z) \geq 1 - \alpha\Big\}.$$

Alors on a $k_\alpha \geq 0$ (pourquoi ?) et, par continuité à droite de $F_{\theta_0}(\cdot)$, $F_{\theta_0}(k_\alpha - 0) \leq 1 - \alpha < F_{\theta_0}(k_\alpha)$.

Par conséquent,

$$\gamma_\alpha \stackrel{\text{def}}{=} \frac{F_{\theta_0}(k_\alpha) - (1-\alpha)}{F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - 0)} \in (0,1].$$

On vérifie alors trivialement que $\mathrm{E}_{\theta_0}^{(n)}[\phi_*] = 1 - F_{\theta_0}(k_\alpha) + \gamma_\alpha \left(F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - 0)\right) = \alpha$.

(ii) Soit $\phi \in \mathcal{C}_\alpha$ (i.e. ϕ satisfait $\mathrm{E}_{\theta_0}^{(n)}[\phi] \leq \alpha$). En notant $\mathcal{X}^{(n)}$ l'ensemble des valeurs possibles de $x^{(n)}$, considérons la fonction $g : \mathcal{X}^{(n)} \to \mathbb{R}$ telle que

$$g(x^{(n)}) = \big(\phi_*(x^{(n)}) - \phi(x^{(n)})\big)\big(L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)})\big),$$

où k_α est le réel intervenant dans ϕ_*.

Si $x^{(n)}$ est tel que $L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) > 0$, alors $\phi_*(x^{(n)}) - \phi(x^{(n)}) = 1 - \phi(x^{(n)}) \geq 0$, tandis que si $x^{(n)}$ est tel que $L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) < 0$, alors $\phi_*(x^{(n)}) - \phi(x^{(n)}) = 0 - \phi(x^{(n)}) \leq 0$. Dans les deux cas, on a donc $g(x^{(n)}) \geq 0$. Puisqu'on a bien entendu aussi $g(x^{(n)}) \geq 0$ si $L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) = 0$, on conclut que $g(x^{(n)}) \geq 0$ pour tout $x^{(n)} \in \mathcal{X}^{(n)}$.

Par conséquent,

$$\begin{aligned}
0 &\leq \int_{\mathcal{X}^{(n)}} g(x^{(n)})\, dx^{(n)} \\
&= \int_{\mathcal{X}^{(n)}} \phi_*(x^{(n)}) L_{\theta_1}^{(n)}(x^{(n)})\, dx^{(n)} - \int_{\mathcal{X}^{(n)}} \phi(x^{(n)}) L_{\theta_1}^{(n)}(x^{(n)})\, dx^{(n)} \\
&\qquad -k_\alpha \left(\int_{\mathcal{X}^{(n)}} \phi_*(x^{(n)}) L_{\theta_0}^{(n)}(x^{(n)})\, dx^{(n)} - \int_{\mathcal{X}^{(n)}} \phi(x^{(n)}) L_{\theta_0}^{(n)}(x^{(n)})\, dx^{(n)} \right) \\
&= \mathrm{E}_{\theta_1}^{(n)}[\phi_*] - \mathrm{E}_{\theta_1}^{(n)}[\phi] - k_\alpha(\alpha - \mathrm{E}_{\theta_0}^{(n)}[\phi]),
\end{aligned}$$

où on a utilisé $\mathrm{E}_{\theta_0}^{(n)}[\phi_*] = \alpha$. Puisque $\mathrm{E}_{\theta_0}^{(n)}[\phi] \leq \alpha$ par hypothèse, on en conclut que $\mathrm{E}_{\theta_1}^{(n)}[\phi_*] \geq \mathrm{E}_{\theta_1}^{(n)}[\phi] + k_\alpha(\alpha - \mathrm{E}_{\theta_0}^{(n)}[\phi]) \geq \mathrm{E}_{\theta_1}^{(n)}[\phi]$. □

Exemple : soit X de loi Unif($[0,1]$) ou Exp(1). Désignons par P_{θ_0} et P_{θ_1} les mesures de probabilité respectives et considérons le problème consistant à tester $\mathcal{H}_0 : X \sim \text{Unif}([0,1])$ contre $\mathcal{H}_1 : X \sim \text{Exp}(1)$ au niveau $\alpha (\in (0,1))$.

On vérifie directement que

$$\frac{L_{\theta_1}(x)}{L_{\theta_0}(x)} = \frac{e^{-x}\mathbb{I}[x \geq 0]}{\mathbb{I}[0 \leq x \leq 1]},$$

de sorte que la fonction de répartition $F_{\theta_0}(\cdot)$ de $L_{\theta_1}(X)/L_{\theta_0}(X)$ sous $\mathcal{H}_0$ est donnée par (ci-dessous,

U désigne une variable aléatoire uniformément distribuée sur $[0,1]$)

$$F_{\theta_0}(z) = \mathrm{P}_{\theta_0}\left[\frac{L_{\theta_1}(X)}{L_{\theta_0}(X)} \le z\right] = \mathrm{P}\left[\frac{L_{\theta_1}(U)}{L_{\theta_0}(U)} \le z\right] = \mathrm{P}\big[e^{-U} \le z\big] = \mathrm{P}\big[U \ge -\ln z\big]$$

$$= 1 - \mathrm{P}\big[U \le -\ln z\big] = \begin{cases} 0 & \text{si } z < 1/e \\ 1 + \ln z & \text{si } 1/e \le z < 1 \\ 1 & \text{si } z \ge 1. \end{cases}$$

Donc $k_\alpha = e^{-\alpha}$ satisfait $F_{\theta_0}(k_\alpha) = 1 - \alpha$, et on peut prendre $\gamma_\alpha = 0$. Puisque

$$\frac{L_{\theta_1}(x)}{L_{\theta_0}(x)} > e^{-\alpha} \quad \Leftrightarrow \quad e^{-x} > e^{-\alpha} \text{ ou } x > 1 \quad \Leftrightarrow \quad x < \alpha \text{ ou } x > 1,$$

le test de Neyman–Pearson au niveau α pour le problème considéré est donc $\phi_*(x) = \mathbb{I}[x < \alpha \text{ ou } x > 1]$. Bien entendu, on vérifie directement qu'on a bien $\mathrm{E}_{\theta_0}[\phi_*] = \alpha$.

Théorème IV.8 *Dans le cadre du théorème IV.7, (i) si ϕ est de niveau α ($\mathrm{E}^{(n)}_{\theta_0}[\phi] \le \alpha$) et est aussi puissant que ϕ_* ($\mathrm{E}^{(n)}_{\theta_1}[\phi] = \mathrm{E}^{(n)}_{\theta_1}[\phi_*]$), alors il existe une fonction $\gamma_\alpha(\cdot) : \mathcal{X}^{(n)} \to [0,1]$ telle que, pour presque tout $x^{(n)} \in \mathcal{X}^{(n)}$, on a*

$$\phi(x^{(n)}) = \begin{cases} 1 & si \quad L^{(n)}_{\theta_1}(x^{(n)}) > k_\alpha L^{(n)}_{\theta_0}(x^{(n)}) \\ \gamma_\alpha(x^{(n)}) & si \quad L^{(n)}_{\theta_1}(x^{(n)}) = k_\alpha L^{(n)}_{\theta_0}(x^{(n)}) \\ 0 & si \quad L^{(n)}_{\theta_1}(x^{(n)}) < k_\alpha L^{(n)}_{\theta_0}(x^{(n)}), \end{cases} \tag{IV.6}$$

où k_α est la constante intervenant dans ϕ_* ; (ii) De plus, $\mathrm{E}^{(n)}_{\theta_1}[\phi_*] > \alpha$.

Preuve. (i) Pour un test ϕ quelconque, on a établi dans la preuve du théorème IV.7(ii) que

$$0 \le \int_{\mathcal{X}^{(n)}} g(x^{(n)})\, dx^{(n)} = \mathrm{E}^{(n)}_{\theta_1}[\phi_*] - \mathrm{E}^{(n)}_{\theta_1}[\phi] - k_\alpha(\alpha - \mathrm{E}^{(n)}_{\theta_0}[\phi]). \tag{IV.7}$$

Si ϕ est tel que décrit dans l'énoncé du théorème, on a donc $0 \le -k_\alpha(\alpha - \mathrm{E}^{(n)}_{\theta_0}[\phi]) \le 0$. Il en découle que $k_\alpha(\alpha - \mathrm{E}^{(n)}_{\theta_0}[\phi]) = 0$ et $\mathrm{E}^{(n)}_{\theta_1}[\phi] = \mathrm{E}^{(n)}_{\theta_1}[\phi_*]$, ce qui implique que

$$\int_{\mathcal{X}^{(n)}} g(x^{(n)})\, dx^{(n)} = 0.$$

Comme $g(x^{(n)}) \ge 0$ pour tout $x^{(n)} \in \mathcal{X}^{(n)}$, on en déduit

$$g(x^{(n)}) = \big(\phi_*(x^{(n)}) - \phi(x^{(n)})\big)\big(L^{(n)}_{\theta_1}(x^{(n)}) - k_\alpha L^{(n)}_{\theta_0}(x^{(n)})\big) = 0$$

pour tout $x^{(n)} \in B$, où $\mathcal{X}^{(n)} \setminus B$ est de mesure nulle. Donc, si $x^{(n)} \in B$ est tel que $L^{(n)}_{\theta_1}(x^{(n)}) > k_\alpha L^{(n)}_{\theta_0}(x^{(n)})$, alors $\phi(x^{(n)}) = \phi_*(x^{(n)}) = 1$, et si $x^{(n)} \in B$ est tel que $L^{(n)}_{\theta_1}(x^{(n)}) < k_\alpha L^{(n)}_{\theta_0}(x^{(n)})$, alors $\phi(x^{(n)}) = \phi_*(x^{(n)}) = 0$. Ceci établit le résultat.

(ii) Considérons le test ϕ_α défini par $\phi_\alpha(x^{(n)}) = \alpha$ pour tout $x^{(n)} \in \mathcal{X}^{(n)}$. Etant donné que

$\mathrm{E}_{\theta_0}^{(n)}[\phi_\alpha] = \mathrm{E}_{\theta_0}^{(n)}[\alpha] = \alpha$, ce test est de niveau α. Puisque ϕ_* est à puissance maximale dans la classe des tests de niveau α, on doit donc avoir

$$\mathrm{E}_{\theta_1}^{(n)}[\phi_*] \geq \mathrm{E}_{\theta_1}^{(n)}[\phi_\alpha] = \mathrm{E}_{\theta_1}^{(n)}[\alpha] = \alpha.$$

Il reste donc à écarter la possibilité qu'on ait $\mathrm{E}_{\theta_1}^{(n)}[\phi_*] = \alpha$. Pour ce faire, supposons par l'absurde qu'on ait effectivement $\mathrm{E}_{\theta_1}^{(n)}[\phi_*] = \alpha$. Dans ce cas, ϕ_α est un test de niveau α qui est aussi puissant que ϕ_*. Donc la partie (i) du théorème implique que, pour presque tout $x^{(n)} \in \mathcal{X}^{(n)}$, on a

$$\alpha = \phi_\alpha(x^{(n)}) = \begin{cases} 1 & \text{si} \quad L_{\theta_1}^{(n)}(x^{(n)}) > k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ \gamma_\alpha(x^{(n)}) & \text{si} \quad L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ 0 & \text{si} \quad L_{\theta_1}^{(n)}(x^{(n)}) < k_\alpha L_{\theta_0}^{(n)}(x^{(n)}). \end{cases}$$

Puisque $\alpha \in (0,1)$, on doit donc avoir que

$$L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \quad \text{pour presque tout } x^{(n)} \in \mathcal{X}^{(n)}.$$

En intégrant les deux membres de cette égalité sur $\mathcal{X}^{(n)}$, on obtient que

$$1 = \int_{\mathcal{X}^{(n)}} L_{\theta_1}^{(n)}(x^{(n)})\, dx^{(n)} = k_\alpha \int_{\mathcal{X}^{(n)}} L_{\theta_0}^{(n)}(x^{(n)})\, dx^{(n)} = k_\alpha,$$

de sorte que $L_{\theta_1}^{(n)}(x^{(n)}) = L_{\theta_0}^{(n)}(x^{(n)})$ pour presque tout $x^{(n)} \in \mathcal{X}^{(n)}$. Pour tout $B \in \mathcal{B}^n$, on a donc

$$\mathrm{P}_{\theta_1}^{(n)}[B] = \int_B L_{\theta_1}^{(n)}(x^{(n)})\, dx^{(n)} = \int_B L_{\theta_0}^{(n)}(x^{(n)})\, dx^{(n)} = \mathrm{P}_{\theta_0}^{(n)}[B],$$

ce qui montre que les mesures de probabilité $\mathrm{P}_{\theta_0}^{(n)}$ et $\mathrm{P}_{\theta_1}^{(n)}$ coïncident. Ceci étant une contradiction, le théorème est démontré. □

Dans l'exemple considéré à la page 50, la puissance du test de Neyman–Pearson peut être évaluée de la façon suivante. En notant V une variable aléatoire de loi Exp(1), on a

$$\begin{aligned} \mathrm{E}_{\theta_1}[\phi_*] &= \mathrm{E}_{\theta_1}[\mathbb{I}[X \leq \alpha \text{ ou } X > 1]] = \mathrm{P}[V \leq \alpha \text{ ou } V > 1] = 1 - \mathrm{P}[\alpha < V \leq 1] \\ &= 1 - \int_\alpha^1 e^{-x}\, dx = 1 - \Big[-e^{-x} \Big]_\alpha^1 = 1 + e^{-1} - e^{-\alpha}. \end{aligned}$$

Clairement, la puissance est une fonction strictement croissante de α, et on peut vérifier que, comme le prévoit le théorème IV.8(ii), on a bien $\mathrm{E}_{\theta_1}[\phi_*] > \alpha$.

IV.3 Tests unilatéraux à puissance uniformément maximale

Le lemme fondamental de Neyman–Pearson fournit une solution élégante et complète au problème de la construction d'un test optimal lorsque l'hypothèse nulle $\mathcal{H}_0 : \theta \in \Theta_0$ et la contre-hypothèse $\mathcal{H}_1$:

$\theta \in \Theta_1$ sont simples. Dans cette section, nous considérons des cas plus complexes et plus pertinents en pratique.

Nous considérons des modèles paramétriques à paramètre scalaire de la forme

$$\Big(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \Big\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}\Big\}\Big), \tag{IV.8}$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Dans ce cadre, nous allons nous intéresser à des problèmes de test unilatéraux de la forme

$$\begin{cases} \mathcal{H}_0 : \theta \leq \theta_0 \\ \mathcal{H}_1 : \theta > \theta_0, \end{cases}$$

où θ_0 est une valeur fixée dans l'intérieur de θ. Dans ce contexte où la contre-hypothèse $\mathcal{H}_1$ est composée, le mieux que l'on puisse espérer d'un test est qu'il soit *à puissance uniformément maximale (PUM)* dans la classe des tests de niveau α, au sens de la définition IV.6. Comme nous allons le voir, il est possible de construire un tel test optimal au moins dans la situation où la propriété suivante tient.

Definition IV.9 *Le modèle statistique paramétrique ci-dessus est à rapport de vraisemblance monotone en la statistique $T(x^{(n)})$ $\Leftrightarrow$ pour tout $\theta', \theta'' \in \Theta$ avec $\theta' < \theta''$, il existe une fonction $h_{\theta',\theta''}(\cdot)$ strictement croissante telle que le rapport de vraisemblance*

$$\frac{L_{\theta''}^{(n)}(x^{(n)})}{L_{\theta'}^{(n)}(x^{(n)})}$$

s'écrit sous la forme $h_{\theta',\theta''}(T(x^{(n)}))$.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Bern(p), où $\theta = p \in \Theta = [0,1] \subset \mathbb{R}$ (exemple électoral). Alors, pour tout p', p'' avec $p' < p''$,

$$\frac{L_{p''}^{(n)}(x^{(n)})}{L_{p'}^{(n)}(x^{(n)})} = \frac{p''^{\sum_{i=1}^n x_i}(1-p'')^{n-\sum_{i=1}^n x_i}}{p'^{\sum_{i=1}^n x_i}(1-p')^{n-\sum_{i=1}^n x_i}} = \Big(\frac{p''}{p'}\Big)^{\sum_{i=1}^n x_i}\Big(\frac{1-p''}{1-p'}\Big)^{n-\sum_{i=1}^n x_i}$$

est une fonction strictement croissante de $T(x^{(n)}) = \sum_{i=1}^n x_i$.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma_0^2)$, avec σ_0^2 un réel strictement positif fixé, de sorte que le paramètre est $\theta = \mu \in \Theta = \mathbb{R}$ (version à variance connue de l'exemple

du verre de bière). Alors, pour tout μ', μ'' avec $\mu' < \mu''$,

$$\begin{aligned}\frac{L^{(n)}_{\mu''}(x^{(n)})}{L^{(n)}_{\mu'}(x^{(n)})} &= \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu'')^2\right)}{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu')^2\right)} \\ &= \frac{\exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2 + \frac{\mu''}{\sigma_0^2}\sum_{i=1}^n x_i - \frac{n(\mu'')^2}{2\sigma_0^2}\right)}{\exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2 + \frac{\mu'}{\sigma_0^2}\sum_{i=1}^n x_i - \frac{n(\mu')^2}{2\sigma_0^2}\right)} \\ &= \exp\left(\frac{n}{2\sigma_0^2}((\mu')^2-(\mu'')^2)\right)\exp\left(\frac{(\mu''-\mu')}{\sigma_0^2}\sum_{i=1}^n x_i\right)\end{aligned}$$

est une fonction strictement croissante de $T(x^{(n)}) = \sum_{i=1}^n x_i$ ou de $T(x^{(n)}) = \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

Montrons qu'en effet cette propriété est suffisante pour l'existence d'un test à PUM dans les problèmes unilatéraux sur θ. Pour ce faire, considérons, dans le cadre du modèle statistique (IV.8), le problème de test

$$\begin{cases} \mathcal{H}_0 : \theta \leq \theta_0 \\ \mathcal{H}_1 : \theta > \theta_0 \end{cases} \tag{IV.9}$$

au niveau $\alpha \in (0,1)$ fixé (θ_0 est une valeur arbitraire fixée dans l'intérieur de Θ). Nous avons le théorème suivant.

Théorème IV.10 *Supposons que le modèle statistique (IV.8) soit à rapport de vraisemblance monotone en la statistique $T(x^{(n)})$ et fixons $\alpha \in (0,1)$. Alors (i) il existe $t_\alpha^+ \in \mathbb{R}$ et $\gamma_\alpha \in [0,1]$ tels que le test ϕ^+_{*T} défini par*

$$\phi^+_{*T}(x^{(n)}) = \begin{cases} 1 & si \quad T(x^{(n)}) > t_\alpha^+ \\ \gamma_\alpha & si \quad T(x^{(n)}) = t_\alpha^+ \\ 0 & si \quad T(x^{(n)}) < t_\alpha^+ \end{cases} \tag{IV.10}$$

vérifie $\mathrm{E}^{(n)}_{\theta_0}[\phi^+_{*T}] = \alpha$; (ii) la fonction $\theta \mapsto \mathrm{E}^{(n)}_{\theta}[\phi^+_{*T}]$ est strictement croissante ; (iii) le test ϕ^+_{*T} est à PUM dans la classe des tests de niveau α pour le problème (IV.9).

Preuve. (i) En désignant par $F^T_{\theta_0}(\cdot)$ la fonction de répartition de $T(X^{(n)})$ sous $\mathrm{P}^{(n)}_{\theta_0}$, on a

$$\begin{aligned}\mathrm{E}^{(n)}_{\theta_0}[\phi^+_{*T}] &= \mathrm{P}^{(n)}_{\theta_0}\left[T(X^{(n)}) > t_\alpha^+\right] + \gamma_\alpha \,\mathrm{P}^{(n)}_{\theta_0}\left[T(X^{(n)}) = t_\alpha^+\right] \\ &= 1 - F^T_{\theta_0}(t_\alpha^+) + \gamma_\alpha\,(F^T_{\theta_0}(t_\alpha^+) - F^T_{\theta_0}(t_\alpha^+ - 0)),\end{aligned}$$

où $F^T_{\theta_0}(t_\alpha^+ - 0) \stackrel{\text{def}}{=} \lim_{z \nearrow t_\alpha^+} F^T_{\theta_0}(z)$. Nous procédons strictement comme dans la preuve du théorème IV.7(i) et considérons deux cas.

(1) Supposons qu'il existe z_α tel que $F^T_{\theta_0}(z_\alpha) = 1-\alpha$. Dans ce cas, si on prend $t_\alpha^+ \stackrel{\text{def}}{=} z_\alpha$ et $\gamma_\alpha \stackrel{\text{def}}{=} 0$, on a bien $\mathrm{E}^{(n)}_{\theta_0}[\phi^+_{*T}] = 1 - F^T_{\theta_0}(t_\alpha^+) = \alpha$.

(2) Supposons que $F^T_{\theta_0}(z) \neq 1-\alpha$ pour tout $z \in \mathbb{R}$. Dans ce cas, posons

$$t_\alpha^+ \stackrel{\text{def}}{=} \inf\left\{z \in \mathbb{R} : F^T_{\theta_0}(z) \geq 1-\alpha\right\}.$$

Par continuité à droite de $F^T_{\theta_0}(\cdot)$, on a $F_{\theta_0}(t^+_\alpha - 0) \leq 1-\alpha < F^T_{\theta_0}(t^+_\alpha)$. Par conséquent,

$$\gamma_\alpha \stackrel{\text{def}}{=} \frac{F^T_{\theta_0}(t^+_\alpha) - (1-\alpha)}{F^T_{\theta_0}(t^+_\alpha) - F^T_{\theta_0}(t^+_\alpha - 0)} \in (0,1).$$

On vérifie alors trivialement que $\mathrm{E}^{(n)}_{\theta_0}[\phi^+_{*T}] = 1 - F^T_{\theta_0}(t^+_\alpha) + \gamma_\alpha \left(F^T_{\theta_0}(t^+_\alpha) - F^T_{\theta_0}(t^+_\alpha - 0)\right) = \alpha$.

(ii) Prenons $\theta', \theta'' \in \Theta$, avec $\theta' < \theta''$. Puisque la famille considérée est à rapport de vraisemblance monotone en $T(x^{(n)})$, le test ϕ^+_{*T} se réécrit

$$\begin{aligned}
\phi^+_{*T}(x^{(n)}) &= \begin{cases} 1 & \text{si} \quad h_{\theta',\theta''}(T(x^{(n)})) > h_{\theta',\theta''}(t^+_\alpha) \\ \gamma_\alpha & \text{si} \quad h_{\theta',\theta''}(T(x^{(n)})) = h_{\theta',\theta''}(t^+_\alpha) \\ 0 & \text{si} \quad h_{\theta',\theta''}(T(x^{(n)})) < h_{\theta',\theta''}(t^+_\alpha) \end{cases} \\
&= \begin{cases} 1 & \text{si} \quad L^{(n)}_{\theta''}(x^{(n)}) > h_{\theta',\theta''}(t^+_\alpha) L^{(n)}_{\theta'}(x^{(n)}) \\ \gamma_\alpha & \text{si} \quad L^{(n)}_{\theta''}(x^{(n)}) = h_{\theta',\theta''}(t^+_\alpha) L^{(n)}_{\theta'}(x^{(n)}) \\ 0 & \text{si} \quad L^{(n)}_{\theta''}(x^{(n)}) < h_{\theta',\theta''}(t^+_\alpha) L^{(n)}_{\theta'}(x^{(n)}). \end{cases}
\end{aligned} \tag{IV.11}$$

Clairement, le test en (IV.11) est le test de Neyman–Pearson pour le problème de test

$$\begin{cases} \mathcal{H}_0 : \theta = \theta' \\ \mathcal{H}_1 : \theta = \theta'' \end{cases} \tag{IV.12}$$

au niveau $\mathrm{E}^{(n)}_{\theta'}[\phi^+_{*T}]$. Le théorème IV.8(ii) montre donc que $\mathrm{E}^{(n)}_{\theta''}[\phi^+_{*T}] > \mathrm{E}^{(n)}_{\theta'}[\phi^+_{*T}]$.

(iii) Fixons une valeur θ_1 associée à la contre-hypothèse ($\theta_1 > \theta_0$). L'argument utilisé dans la preuve de (ii) (avec $\theta' = \theta_0$ et $\theta'' = \theta_1$) montre que ϕ^+_{*T} est le test de Neyman–Pearson pour le problème de test

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1 \end{cases} \tag{IV.13}$$

au niveau $\mathrm{E}^{(n)}_{\theta_0}[\phi^+_{*T}] = \alpha$ (voir le point (i) du théorème). Le théorème IV.7(ii) implique donc que ϕ^+_{*T} est à puissance maximale au niveau α pour le problème (IV.13). Puisque $\theta_1 (> \theta_0)$ est arbitraire ci-dessus, on en déduit que ϕ^+_{*T} est à puissance uniformément maximale au niveau α pour le problème

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta > \theta_0. \end{cases} \tag{IV.14}$$

Soit alors ϕ un test de niveau α pour le problème (IV.9). Alors ϕ est également de niveau α pour le problème (IV.14), de sorte qu'il doit être uniformément dominé par ϕ^+_{*T} en termes de puissance ($\mathrm{E}^{(n)}_\theta[\phi^+_{*T}] \geq \mathrm{E}^{(n)}_\theta[\phi]$ pour tout $\theta > \theta_0$). Puisque, par ailleurs, (i)–(ii) impliquent que ϕ^+_{*T} est de niveau α pour le problème (IV.9), on en déduit que ϕ^+_{*T} est bien à PUM au niveau α pour le problème (IV.9). □

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2_0)$, avec σ^2_0 un réel strictement

positif fixé, de sorte que le paramètre est $\theta = \mu \in \Theta = \mathbb{R}$ (version à variance connue de l'exemple du verre de bière). Nous considérons le problème de test

$$\begin{cases} \mathcal{H}_0 : \mu \leq \mu_0 \\ \mathcal{H}_1 : \mu > \mu_0 \end{cases}$$

au niveau $\alpha \in (0,1)$, où μ_0 est un nombre réel donné. Puisque la famille est à rapport de vraisemblance monotone en $T(x^{(n)}) = \bar{x}^{(n)}$, le test à PUM pour le problème considéré est de la forme

$$\phi^+_{*T}(x^{(n)}) = \begin{cases} 1 & \text{si} \quad \bar{x} > t^+_\alpha \\ \gamma_\alpha & \text{si} \quad \bar{x} = t^+_\alpha \\ 0 & \text{si} \quad \bar{x} < t^+_\alpha. \end{cases} \tag{IV.15}$$

Les constantes γ_α et t^+_α doivent être déterminées en imposant la contrainte $\mathrm{E}^{(n)}_{\mu_0}[\phi^+_{*T}] = \alpha$. Sous $\mathrm{P}^{(n)}_{\mu_0}$, $\bar{X}$ est de loi normale de moyenne μ_0 et de variance σ_0^2/n, de sorte que la contrainte se réécrit (ci-dessous, Z est une variable aléatoire de loi normale standard)

$$\begin{aligned} \alpha = \mathrm{E}^{(n)}_{\mu_0}[\phi_*] &= \mathrm{P}^{(n)}_{\mu_0}[\bar{X} > t^+_\alpha] + \gamma_\alpha \mathrm{P}^{(n)}_{\mu_0}[\bar{X} = t^+_\alpha] \\ &= \mathrm{P}^{(n)}_{\mu_0}\Big[\frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2/n}} > \frac{t^+_\alpha - \mu_0}{\sqrt{\sigma_0^2/n}}\Big] + \gamma_\alpha \mathrm{P}^{(n)}_{\mu_0}\Big[\frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2/n}} = \frac{t^+_\alpha - \mu_0}{\sqrt{\sigma_0^2/n}}\Big] \\ &= \mathrm{P}\left[Z > \frac{t^+_\alpha - \mu_0}{\sqrt{\sigma_0^2/n}}\right] + \gamma_\alpha \mathrm{P}\left[Z = \frac{t^+_\alpha - \mu_0}{\sqrt{\sigma_0^2/n}}\right] \\ &= 1 - \Phi\left(\frac{t^+_\alpha - \mu_0}{\sqrt{\sigma_0^2/n}}\right), \end{aligned}$$

où Φ désigne la fonction de répartition de la loi normale standard. Par conséquent, on peut choisir arbitrairement γ_α dans $[0,1]$ (l'usage est de prendre $\gamma_\alpha = 0$) et on doit prendre

$$t^+_\alpha = \mu_0 + \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha).$$

En résumé, le test à PUM au niveau α est le test

$$\phi^+_{*T}(x^{(n)}) = \begin{cases} 1 & \text{si} \quad \bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha) \\ 0 & \text{si} \quad \bar{x} \leq \mu_0 + \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha), \end{cases} \tag{IV.16}$$

qui rejette l'hypothèse nulle $\mathcal{H}_0 : \mu \leq \mu_0$ en faveur $\mathcal{H}_1 : \mu > \mu_0$ si et seulement si $\bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \Phi^{-1}(1-\alpha)$.

La "fonction de puissance" ($\mu \mapsto \mathrm{E}^{(n)}_\mu[\phi^+_{*T}]$) peut être déterminée de la façon suivante. En

désignant de nouveau par Z une variable aléatoire de loi normale standard, on a

$$\begin{aligned}
\mathrm{E}_\mu^{(n)}[\phi_{*T}^+] &= \mathrm{P}_\mu^{(n)}\left[\bar{X} > \mu_0 + \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right] \\
&= \mathrm{P}_\mu^{(n)}\left[\frac{\bar{X}-\mu}{\sqrt{\sigma_0^2/n}} > \Phi^{-1}(1-\alpha) + \frac{\mu_0-\mu}{\sqrt{\sigma_0^2/n}}\right] \\
&= \mathrm{P}\left[Z > \Phi^{-1}(1-\alpha) - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\right] \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\right),
\end{aligned}$$

où on a utilisé le fait que, sous $\mathrm{P}_\mu^{(n)}$, $\bar{X}$ est de loi normale de moyenne μ et de variance σ_0^2/n. Clairement, ceci confirme que $\mu \mapsto \mathrm{E}_\mu^{(n)}[\phi_{*T}^+]$ est strictement croissante. Si μ tend vers μ_0 par valeurs plus grandes, la puissance correspondante tend vers α, ce qui montre que, de façon générale, on ne peut pas être plus précis dans le théorème IV.8(ii). Notons aussi que, pour $\mu_0, \mu(> \mu_0)$ et $\alpha(\in (0,1))$ fixés, la puissance $\mathrm{E}_\mu^{(n)}[\phi_{*T}^+]$ tend vers 1 si $n \to \infty$. On parlera de *test convergent*.

Bien entendu, l'approche considérée dans cette section peut aussi être appliquée au problème de test

$$\begin{cases} \mathcal{H}_0 : \theta \geq \theta_0 \\ \mathcal{H}_1 : \theta < \theta_0. \end{cases} \tag{IV.17}$$

À titre d'exercice, on établira le résultat suivant.

Théorème IV.11 *Supposons que le modèle statistique (IV.8) soit à rapport de vraisemblance monotone en la statistique $T(x^{(n)})$ et fixons $\alpha \in (0,1)$. Alors (i) il existe $t_\alpha^- \in \mathbb{R}$ et $\gamma_\alpha \in [0,1]$ tels que le test ϕ_{*T}^- défini par*

$$\phi_{*T}^-(x^{(n)}) = \begin{cases} 1 & si \quad T(x^{(n)}) < t_\alpha^- \\ \gamma_\alpha & si \quad T(x^{(n)}) = t_\alpha^- \\ 0 & si \quad T(x^{(n)}) > t_\alpha^- \end{cases} \tag{IV.18}$$

vérifie $\mathrm{E}_{\theta_0}[\phi_{*T}^-] = \alpha$; (ii) la fonction $\theta \mapsto \mathrm{E}_\theta[\phi_{*T}^-]$ est strictement décroissante ; (iii) le test ϕ_{*T}^- est à PUM dans la classe des tests de niveau α pour le problème (IV.17).

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma_0^2)$, avec σ_0^2 un réel strictement positif fixé, de sorte que le paramètre est $\theta = \mu \in \Theta = \mathbb{R}$ (version à variance connue de l'exemple du verre de bière). Pour le problème de test

$$\begin{cases} \mathcal{H}_0 : \mu \geq \mu_0 \\ \mathcal{H}_1 : \mu < \mu_0, \end{cases}$$

on vérifiera que le test à PUM au niveau α est le test

$$\phi_{*T}^-(x^{(n)}) = \begin{cases} 1 & \text{si} \quad \bar{x} < \mu_0 - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\alpha) \\ 0 & \text{si} \quad \bar{x} \geq \mu_0 - \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}(1-\alpha), \end{cases}$$

dont la fonction de puissance est

$$\mu \mapsto \mathrm{E}_{\mu}^{(n)}[\phi_{*T}^{-}] = \Phi\bigg(-\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\bigg).$$

IV.4 Tests bilatéraux

Dans la section précédente, nous avons montré que, si le modèle paramétrique à paramètre scalaire

$$\Big(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \Big\{\mathrm{P}_{\theta}^{(n)} : \theta \in \Theta \subset \mathbb{R}\Big\}\Big) \tag{IV.19}$$

(où les mesures de probabilité $\mathrm{P}_{\theta}^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$) est à rapport de vraisemblance monotone, il est possible de construire des tests à PUM au niveau α pour les problèmes de test unilatéraux de la forme

$$\left\{ \begin{array}{l} \mathcal{H}_0 : \theta \leq \theta_0 \\ \mathcal{H}_1 : \theta > \theta_0 \end{array} \right. \quad \text{et} \quad \left\{ \begin{array}{l} \mathcal{H}_0 : \theta \geq \theta_0 \\ \mathcal{H}_1 : \theta < \theta_0, \end{array} \right.$$

où θ_0 est une valeur fixée dans l'intérieur de θ. Un autre problème classique dans ce contexte est le problème de test bilatéral

$$\left\{ \begin{array}{l} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0. \end{array} \right. \tag{IV.20}$$

Nous avons, dans le cas des familles à rapport de vraisemblance monotone, le résultat négatif suivant.

Théorème IV.12 *Supposons que le modèle statistique (IV.8) soit à rapport de vraisemblance monotone en la statistique $T(x^{(n)})$. Quel que soit $\alpha \in (0,1)$, il n'existe pas de test à puissance uniformément maximale (PUM) au niveau α pour le problème (IV.20).*

Preuve. Par l'absurde, supposons que ϕ_{*T} soit un test à PUM au niveau α pour le problème considéré. Fixons alors $\theta_1 \in \Theta$ avec $\theta_1 > \theta_0$. Puisque ϕ_{*T}^{+} est de niveau α pour le problème (IV.20), on doit avoir $\mathrm{E}_{\theta_1}[\phi_{*T}] \geq \mathrm{E}_{\theta_1}[\phi_{*T}^{+}]$. Comme on a vu à la section précédente que ϕ_{*T}^{+} est le test de Neyman–Pearson pour

$$\left\{ \begin{array}{l} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1 \end{array} \right. \tag{IV.21}$$

au niveau α, on en déduit que $\mathrm{E}_{\theta_1}[\phi_{*T}] = \mathrm{E}_{\theta_1}[\phi_{*T}^{+}]$. Mais puisque le test ϕ_{*T} est de niveau α pour le problème (IV.21) et est aussi puissant que le test de Neyman–Pearson pour ce problème, le théorème IV.8(i) livre que

$$\phi_{*T}(x^{(n)}) = \left\{ \begin{array}{ccl} 1 & \text{si} & L_{\theta_1}^{(n)}(x^{(n)}) > k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ \gamma_\alpha(x^{(n)}) & \text{si} & L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ 0 & \text{si} & L_{\theta_1}^{(n)}(x^{(n)}) < k_\alpha L_{\theta_0}^{(n)}(x^{(n)}), \end{array} \right.$$

pour une certaine constante $k_\alpha \in \mathbb{R}^+$ et une certaine fonction $\gamma_\alpha(\cdot)$. Puisque le modèle considéré est à rapport de vraisemblance monotone, on peut alors montrer de la même manière que dans la preuve du théorème IV.10(ii) que $\theta \mapsto \mathrm{E}_\theta^{(n)}[\phi_{*T}]$ est strictement croissante. Par conséquent, $\mathrm{E}_\theta^{(n)}[\phi_{*T}] < \mathrm{E}_{\theta_0}^{(n)}[\phi_{*T}] = \alpha = \mathrm{E}_{\theta_0}^{(n)}[\phi_{*T}^-] < \mathrm{E}_\theta^{(n)}[\phi_{*T}^-]$ pour tout $\theta < \theta_0$. Ceci montre que ϕ_{*T} n'est pas à PUM pour le problème (IV.20) (puisque ϕ_{*T}^- est de niveau α pour ce problème et a une puissance strictement plus élevée que celle de ϕ_{*T} en tout $\theta < \theta_0$). □

Comme dans le cas de l'estimation ponctuelle, on peut espérer pouvoir identifier une procédure optimale si on se restreint à une plus petite classe de procédures. Dans le cas qui nous occupe, la classe habituellement considérée est celle des *tests sans biais*.

Definition IV.13 *Le test ϕ de niveau α pour le problème $\mathcal{H}_0 : \theta \in \Theta_0$ contre $\mathcal{H}_1 : \theta \in \Theta_1$ est* sans biais $\Leftrightarrow \mathrm{E}_\theta^{(n)}[\phi] \geq \alpha$ *pour tout $\theta \in \Theta_1$.*

Pour le problème bilatéral considéré, un test ϕ admissible est donc tel que $\mathrm{E}_{\theta_0}^{(n)}[\phi] \leq \alpha$ et $\mathrm{E}_\theta^{(n)}[\phi] \geq \alpha$ pour tout $\theta \neq \theta_0$ [2]. Ceci exclut clairement les compétiteurs ϕ_{*T}^- et ϕ_{*T}^+ qui ont permis ci-dessus d'écarter l'existence d'un test à PUM au niveau α. Le concept d'optimalité associé aux tests de niveau α sans biais est le suivant.

Definition IV.14 *Le test ϕ_* est à puissance uniformément maximale (PUM) dans la classe des tests sans biais au niveau $\alpha \Leftrightarrow$ (i) ϕ_* est sans biais au niveau α et (ii) pour tout test ϕ sans biais au niveau α, on a $\mathrm{E}_\theta[\phi_*] \geq \mathrm{E}_\theta[\phi]$ pour tout $\theta \in \Theta_1$.*

Pour certaines classes importantes de problèmes bilatéraux, il existe des tests à PUM dans la classe des tests sans biais au niveau α. La construction des tests est cependant plus délicate que pour la construction des tests à PUM au niveau α de la section précédente. Nous énonçons le résultat suivant, sans démonstration.

Théorème IV.15 *Supposons que le modèle statistique (IV.19) mène à des vraisemblances de la forme $L_\theta(x^{(n)}) = c_\theta h(x^{(n)}) \exp(\theta\, T(x^{(n)}))$ (donc est à rapport de vraisemblance monotone en $T(x^{(n)})$). Soit $\alpha \in (0,1)$. Alors (i) il existe $t_{\alpha 1}, t_{\alpha 2} \in \mathbb{R}$ et $\gamma_{\alpha 1}, \gamma_{\alpha 2} \in [0,1]$ tels que le test ϕ_{*T} défini par*

$$\phi_{*T}(x^{(n)}) = \begin{cases} 1 & si \quad T(x^{(n)}) \notin [t_{\alpha 1}, t_{\alpha 2}] \\ \gamma_{\alpha 1} & si \quad T(x^{(n)}) = t_{\alpha 1} \\ \gamma_{\alpha 2} & si \quad T(x^{(n)}) = t_{\alpha 2} \\ 0 & si \quad T(x^{(n)}) \in (t_{\alpha 1}, t_{\alpha 2}) \end{cases}$$

vérifie $\mathrm{E}_{\theta_0}[\phi_{*T}] = \alpha$ et $\mathrm{E}_{\theta_0}[\phi_{*T} T] = \alpha \mathrm{E}_{\theta_0}[T]$ (où on a écrit $T = T(X^{(n)})$) ; (ii) ϕ_{*T} est à PUM dans la classe des tests sans biais au niveau α pour le problème de test bilatéral (IV.20).

À titre d'illustration, nous traitons rapidement l'analogue bilatéral de l'exemple gaussien considéré dans la section précédente. Nous considérons donc $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d.

2. Bien entendu, si la fonction de puissance $\theta \mapsto \mathrm{E}_\theta^{(n)}[\phi]$ est continue, alors un tel test satisfait automatiquement $\mathrm{E}_{\theta_0}^{(n)}[\phi] = \alpha$.

$\mathcal{N}(\mu, \sigma_0^2)$, avec σ_0^2 un réel strictement positif fixé, de sorte que le paramètre est $\theta = \mu \in \Theta = \mathbb{R}$ (version à variance connue de l'exemple du verre de bière). Le problème bilatéral correspondant est de la forme

$$\begin{cases} \mathcal{H}_0 : \mu = \mu_0 \\ \mathcal{H}_1 : \mu \neq \mu_0, \end{cases}$$

où μ_0 est un nombre réel donné.

Puisque

$$\begin{aligned} L_\mu^{(n)}(x^{(n)}) &= \Big(\frac{1}{\sqrt{2\pi}\sigma_0}\Big)^n \exp\Big(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu)^2\Big) \\ &= \underbrace{\Big(\frac{1}{\sqrt{2\pi}\sigma_0}\Big)^n \exp\Big(-\frac{n\mu^2}{2\sigma_0^2}\Big)}_{c_\mu} \underbrace{\exp\Big(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2\Big)}_{h(x^{(n)})} \underbrace{\exp\Big(\frac{\mu}{\sigma_0^2}\sum_{i=1}^n x_i\Big)}_{\exp(\mu T(x^{(n)}))}, \end{aligned}$$

la vraisemblance est de la forme mentionnée dans le théorème, avec $T(x^{(n)}) = \frac{1}{\sigma_0^2}\sum_{i=1}^n x_i$. À titre d'exercice, on vérifiera alors que le test ϕ_{*T} optimal est alors donné par

$$\phi_{*T}(x^{(n)}) = \begin{cases} 1 & \text{si} \quad \big|\bar{X} - \mu_0\big| > \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}\big(1-\frac{\alpha}{2}\big) \\ 0 & \text{si} \quad \big|\bar{X} - \mu_0\big| \leq \frac{\sigma_0}{\sqrt{n}}\Phi^{-1}\big(1-\frac{\alpha}{2}\big) \end{cases}$$

et que sa fonction de puissance est

$$\begin{aligned} \mu \mapsto \mathrm{E}_\mu^{(n)}[\phi] &= \mathrm{P}_\mu^{(n)}[\phi = 1] = \mathrm{P}_\mu^{(n)}\Big[\frac{\bar{X}-\mu_0}{\sqrt{\sigma_0^2/n}} \notin \Big[-\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big), \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big]\Big] \\ &= \mathrm{P}_\mu^{(n)}\Big[\frac{\bar{X}-\mu}{\sqrt{\sigma_0^2/n}} \notin \Big[-\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + \frac{\mu_0-\mu}{\sqrt{\sigma_0^2/n}}, \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + \frac{\mu_0-\mu}{\sqrt{\sigma_0^2/n}}\Big]\Big] \\ &= 1 - \Phi\Big(\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + \frac{\sqrt{n}}{\sigma_0}(\mu_0-\mu)\Big) + \Phi\Big(-\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + \frac{\sqrt{n}}{\sigma_0}(\mu_0-\mu)\Big). \end{aligned} \tag{IV.22}$$

L'expression (IV.22) confirme que le risque de première espèce de ce test vaut α et permet de vérifier (exercice) la fonction de puissance est symétrique par rapport à μ_0. On pourra aussi vérifier que le test est convergent (au sens où en chaque $\mu \neq \mu_0$, sa puissance tend vers 1 lorsque $n \to \infty$) et que la puissance est d'autant plus grande que σ_0 est petit.

IV.5 Tests de rapport de vraisemblance

Jusqu'ici, nous avons considéré des tests qui ont de bonnes propriétés d'optimalité, mais qui sont restreints à un contexte particulier (modèle à rapport de vraisemblance monotone, paramètre d'intérêt scalaire, pas de paramètre de nuisance, etc.). Dans cette section, nous présentons une méthode classique qui permet de traiter des modèles paramétriques et des problèmes de test assez généraux.

Considérons un modèle statistique paramétrique de la forme

$$\Big(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \Big\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k\Big\}\Big),$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$, et notons $L_\theta^{(n)}(x^{(n)})$ la vraisemblance associée à $\mathrm{P}_\theta^{(n)}$. Nous nous intéressons au problème de test général

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1, \end{cases}$$

où Θ_0 et $\Theta_1 (= \Theta \backslash \Theta_0)$ forment une partition de Θ. On notera r le nombre de paramètres indépendants dans Θ_0 (en particulier, si Θ_0 est un espace affin, il s'agit de la dimension correspondante).

Dans ce contexte, la méthode du rapport de vraisemblance consiste à rejeter $\mathcal{H}_0$ en faveur de $\mathcal{H}_1$ lorsque la valeur de

$$\Lambda(x^{(n)}) = \frac{\sup_{\theta\in\Theta_0} L_\theta^{(n)}(x^{(n)})}{\sup_{\theta\in\Theta} L_\theta^{(n)}(x^{(n)})}$$

est trop petite. Notons que $\Lambda(x^{(n)}) \in [0,1]$ pour tout $x^{(n)}$ et qu'une petite valeur de $\Lambda(x^{(n)})$ est en effet une indication que l'observation $x^{(n)}$ apparaît moins exceptionnelle pour une valeur du paramètre appartenant à Θ_1 que pour la meilleure valeur du paramètre appartenant à Θ_0.

Le test de rapport de vraisemblance est alors formellement

$$\phi_{\mathrm{RV}}(x^{(n)}) = \mathbb{I}\big[\Lambda(x^{(n)}) < c_\alpha\big],$$

où c_α est choisi de telle façon que

$$\sup_{\theta\in\Theta_0} \mathrm{E}_\theta^{(n)}[\phi_{\mathrm{RV}}] = \alpha.$$

En pratique, il est rarement possible de pouvoir déterminer ce c_α. Heureusement, on peut conduire ce test de façon asymptotique grâce au résultat suivant.

Théorème IV.16 *Sous certaines conditions de régularité, on a que, pour tout $\theta \in \Theta_0$,*

$$-2\ln\Lambda(X^{(n)}) \xrightarrow{\mathcal{L}} \chi^2_{k-r}$$

sous $\mathrm{P}_\theta^{(n)}$ quand $n \to \infty$.

Il découle de ce théorème que le test de rapport de vraisemblance asymptotique

$$\phi_{\mathrm{RV}}^{\mathrm{as}}(x^{(n)}) = \mathbb{I}\Big[-2\ln\Lambda(x^{(n)}) > \chi^2_{k-r,1-\alpha}\Big],$$

où $\chi^2_{k-r,1-\alpha}$ désigne le quantile d'ordre $1-\alpha$ de la loi χ^2_{k-r}, vérifie

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_0} \mathrm{E}_\theta^{(n)}[\phi_{\mathrm{RV}}] \leq \alpha$$

(et est donc "asymptotiquement de niveau α"). À l'instar de la procédure correspondante pour l'estimation ponctuelle (l'estimateur du maximum de vraisemblance), ce test jouit de certaines propriétés d'optimalité asymptotiques pour peu que certaines conditions de régularité soient satisfaites.

IV.6 Tests χ^2

IV.6.1 Loi multinomiale

Considérons une expérience aléatoire E qui donne lieu à un nombre fini de résultats possibles $A_1, \ldots, A_I$ et notons les probabilités correspondantes par $p_1, \ldots, p_I$.

Definition IV.17 *Dans la situation où on répète n fois indépendamment l'expérience E ci-dessus, la loi du vecteur aléatoire $N = (N_1, \ldots, N_I)$, où N_i compte le nombre d'occurences du résultat A_i parmi les n expériences réalisées, est dite* loi multinomiale de paramètres $n, p_1, \ldots, p_{I-1}$ *(notation : $N \sim \text{Mult}(n, p_1, \ldots, p_{I-1})$).*

Notons qu'il n'est pas nécessaire d'inclure p_I dans les paramètres de la loi ci-dessus, puisque $p_I = 1 - \sum_{i=1}^{I-1} p_i$. Par ailleurs, si $N \sim \text{Mult}(n, p_1, \ldots, p_{I-1})$, on a clairement que $N_i \sim \text{Bin}(n, p_i)$ pour tout i. Par conséquent,

$$\text{E}[N_i] = np_i \tag{IV.23}$$

et $\text{Var}[N_i] = np_i(1-p_i)$. Puisque $N_i + N_j \sim \text{Bin}(n, p_i + p_j)$ pour tout $i \neq j$, on a $n(p_i+p_j)(1-(p_i+p_j)) = \text{Var}[N_i+N_j] = \text{Var}[N_i]+\text{Var}[N_j]+2\text{Cov}[N_i, N_j] = np_i(1-p_i)+np_j(1-p_j)+2\text{Cov}[N_i, N_j]$, ce qui livre $\text{Cov}[N_i, N_j] = -np_ip_j < 0$ pour tout $i \neq j$. Les N_i sont donc dépendants (et négativement corrélés), ce qui est compatible avec la contrainte $N_1 + \ldots + N_I = n$.

On peut montrer les deux résultats asymptotiques suivants.

Théorème IV.18 *Soit $N \sim \text{Mult}(n, p_1, \ldots, p_{I-1})$. Alors on a que*

$$\sum_{i=1}^{I} \frac{(N_i - np_i)^2}{np_i} \xrightarrow{\mathcal{L}} \chi^2_{I-1}$$

quand $n \to \infty$.

Théorème IV.19 *Soit $N \sim \text{Mult}(n, p_1(\theta), \ldots, p_{I-1}(\theta))$, avec $\theta \in \Theta \subset \mathbb{R}^k$, où Θ est d'intérieur non vide. Alors, en désignant par $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ, on a que, pour tout $\theta \in \Theta$,*

$$\sum_{i=1}^{I} \frac{(N_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} \xrightarrow{\mathcal{L}} \chi^2_{I-1-k}$$

sous $\text{P}_\theta^{(n)}$ quand $n \to \infty$.

Il est coutume de considérer que les approximations χ^2 des distributions des variables aléatoires considérées dans ces deux théorèmes sont satisfaisantes pourvu que $n \geq 30$, que $np_i \geq 5$ pour au moins 80% des valeurs de i, et que $np_i \geq 1$ pour toutes les valeurs de i.

Ces deux résultats permettent de définir, en présence de comptages, des tests classiques connus sous le nom de *tests* χ^2 ("tests chi-carré").

IV.6.2 Tests d'ajustement / tests d'adéquation

Considérons une expérience aléatoire E comme ci-dessus donnant lieu à des résultats possibles $A_1, \ldots, A_I$ et notons les probabilités correspondantes par $p_1, \ldots, p_I$. Supposons qu'on observe $N \sim \text{Mult}(n, p_1, \ldots, p_{I-1})$ et qu'on s'intéresse au problème de test

$$\begin{cases} \mathcal{H}_0 : p_i = p_{i0} \text{ pour tout } i = 1, \ldots, I \\ \mathcal{H}_1 : p_i \neq p_{i0} \text{ pour au moins un } i = 1, \ldots, I, \end{cases} \tag{IV.24}$$

où p_{i0}, $i = 1, \ldots, I$, sont des nombres fixés dans $(0, 1)$ tels que $\sum_{i=1}^{I} p_{i0} = 1$.

Puisque, sous $\mathcal{H}_0$, on a $\mathrm{E}[N_i] = np_{i0}$ pour tout $i = 1, \ldots, I$, il est naturel d'adopter un test qui rejette $\mathcal{H}_0$ pour les grandes valeurs de

$$Q^{(n)} = \sum_{i=1}^{I} \frac{(N_i - np_{i0})^2}{np_{i0}},$$

qui peut-être vue comme une "distance" entre les effectifs observés N_i et les effectifs attendus np_{i0} sous $\mathcal{H}_0$. Bien entendu, c'est le théorème IV.18 qui motive de regarder cette distance particulière. Il découle en effet directement de ce résultat que le test qui rejette $\mathcal{H}_0$ si et seulement si

$$\sum_{i=1}^{I} \frac{(N_i - np_{i0})^2}{np_{i0}} > \chi^2_{I-1,1-\alpha}$$

est asymptotiquement de niveau α.

La notation $\chi^2_{k,\alpha}$ désignera dorénavant le quantile d'ordre α d'une loi chi-carré à k degrés de liberté. De manière similaire, on notera $t_{k,\alpha}$ le quantile d'ordre α d'une loi de Student à k degrés de liberté. Généralement, toute loi "connue" (qu'on ne manquera pas de découvrir par la suite) se verra associer une notation compacte pour ses quantiles de la même manière. Pour des raisons historiques (et pour cause de symétrie), cependant, on ne dénotera pas le quantile d'une loi gaussienne (centrée et réduite) par $\mathcal{N}_\alpha$. Plutôt, on utilisera soit la notation $\Phi^{-1}(\alpha)$, où Φ désigne historiquement la fonction de répartition d'une loi normale centrée réduite, soit la notation z_α, qui désigne le quantile d'ordre $1 - \alpha$ de cette même loi. De nombreuses tables de ces quantiles existent. Voir par exemple l'annexe B.

Exemple : les lois de Mendel prévoient lors de croisements de pois de certains types d'obtenir des descendants jaunes lisses, jaunes ridés, verts lisses et verts ridés en proportions respectives 9/16, 3/16, 3/16 et 1/16. Pour confronter la théorie (l'hypothèse $\mathcal{H}_0 : (p_1, p_2, p_3, p_4) = (9/16, 3/16, 3/16, 1/16)$) aux données, Mendel effectue $n = 556$ tels croisements et observe les effectifs $n_1 = 315$, $n_2 = 101$,

$n_3 = 108$ et $n_4 = 32$. La statistique de test $Q^{(n)}$ ci-dessus prend la valeur

$$\begin{aligned} q^{(n)} &= \frac{(315 - 556 \times \frac{9}{16})^2}{556 \times \frac{9}{16}} + \frac{(101 - 556 \times \frac{3}{16})^2}{556 \times \frac{3}{16}} + \frac{(108 - 556 \times \frac{3}{16})^2}{556 \times \frac{3}{16}} + \frac{(32 - 556 \times \frac{1}{16})^2}{556 \times \frac{1}{16}} \\ &\approx 0.47. \end{aligned}$$

Au niveau $\alpha = 1\%$, on ne rejette pas $\mathcal{H}_0$ puisque

$$0.47 \approx q^{(n)} \leq \chi^2_{3,0.99} \approx 11.34.$$

On ne rejette pas non plus l'hypothèse nulle au niveau $\alpha = 5\%$ ou au niveau $\alpha = 10\%$, puisque

$$0.47 \approx q^{(n)} \leq \chi^2_{3,0.95} \approx 7.81 \quad \text{et} \quad 0.47 \approx q^{(n)} \leq \chi^2_{3,0.90} \approx 6.25.$$

Cet exemple met en évidence qu'effectuer le test à différents niveaux α demande d'évaluer dans chaque cas la valeur critique $\chi^2_{3,1-\alpha}$ correspondante. Pour éviter ceci, on considère souvent la *p-valeur*, qui est définie comme

$$\text{p-valeur} = \mathrm{P}^{(n)}_{\mathcal{H}_0}[Q^{(n)} > q^{(n)}],$$

où $q^{(n)}$ représente la valeur observée de la statistique $Q^{(n)}$. Clairement, le test ci-dessus rejette l'hypothèse nulle si et seulement si la p-valeur est strictement inférieure à α, ce qui permet, sur la base de la seule p-valeur, d'effectuer sans effort le test pour toute une gamme de valeurs de α. Dans l'exemple ci-dessus, la p-valeur vaut

$$\text{p-valeur} \approx \mathrm{P}^{(n)}_{\mathcal{H}_0}[Q^{(n)} > 0.47] = \mathrm{P}[\chi^2_3 > 0.47] \approx 0.93.$$

Ceci montre que dès qu'on conduit le test à un niveau inférieur à 93% (donc à tous les niveaux usuels, puisqu'on n'utilise jamais un niveau supérieur à $\alpha = 10\%$ en pratique), le test ne conduit pas à rejeter l'hypothèse nulle. On conclut donc que les données, à tous les niveaux usuels, ne prennent pas la théorie de Mendel en défaut.

Dans le problème de test (IV.24), la valeur de chaque p_i est fixée sous l'hypothèse nulle. Il est cependant plus courant d'avoir un "modèle" qui spécifie ces probabilités à un ou plusieurs paramètre(s) près. Dans ce cas, le problème prend la forme

$$\begin{cases} \mathcal{H}_0 : \exists \theta \in \Theta \text{ tel que } p_i = p_i(\theta) \text{ pour tout } i = 1, \ldots, I \\ \mathcal{H}_1 : \forall \theta \in \Theta,\ p_i \neq p_i(\theta) \text{ pour au moins un } i = 1, \ldots, I, \end{cases} \tag{IV.25}$$

où $\Theta \subset \mathbb{R}^k$ est d'intérieur non vide. En notant $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ sous l'hypothèse nulle, il découle du théorème IV.19 que le test qui rejette $\mathcal{H}_0$ si et seulement si

$$\sum_{i=1}^{I} \frac{(N_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} > \chi^2_{I-1-k,1-\alpha}$$

est asymptotiquement de niveau α pour le problème considéré.

Exemple : Le groupe sanguin MN se compose de trois phénotypes distincts : M, MN, et N. Une population d'individus appartenant à ce groupe est dite *stable dans sa composition* si les proportions relatives à ces trois phénotypes sont de la forme

$$\mathrm{P}[M] = p_1(\theta) = \theta^2, \quad \mathrm{P}[MN] = p_2(\theta) = 2\theta(1-\theta), \quad \mathrm{P}[N] = p_3(\theta) = (1-\theta)^2,$$

pour un certain $\theta \in (0,1)$. Un échantillon de taille 500 est prélevé dans une population de ce type, et donne lieu à $n_1 = 125$ individus présentant le phénotype M, $n_2 = 225$ individus présentant le phénotype MN, et $n_3 = 150$ individus présentant le phénotype N. Nous désirons tester au niveau $\alpha = 5\%$ si cet échantillon est compatible avec l'hypothèse que la population est stable dans sa composition.

L'hypothèse nulle $\mathcal{H}_0$ à tester est donc celle pour laquelle il existe une valeur de $\theta \in (0,1)$ pour laquelle les "vraies" proportions $\mathrm{P}[M]$, $\mathrm{P}[MN]$, et $\mathrm{P}[N]$ de la population considérée sont données par $\mathrm{P}[M] = p_1(\theta) = \theta^2$, $\mathrm{P}[MN] = p_2(\theta) = 2\theta(1-\theta)$, $\mathrm{P}[N] = p_3(\theta) = (1-\theta)^2$.

Déterminons d'abord l'estimateur $\hat{\theta}$ du maximum de vraisemblance pour θ. Sous $\mathcal{H}_0$, la vraisemblance est donnée par

$$L_\theta(n_1, n_2, n_3) = \frac{n!}{(n_1!)(n_2!)(n_3!)} (p_1(\theta))^{n_1} (p_2(\theta))^{n_2} (p_3(\theta))^{n_3},$$

où $n = n_1 + n_2 + n_3$. On a donc

$$\begin{aligned}
\ln L_\theta(n_1, n_2, n_3) &= C + n_1 \ln p_1(\theta) + n_2 \ln p_2(\theta) + n_3 \ln p_3(\theta) \\
&= C + 2n_1 \ln \theta + n_2 \ln(2\theta(1-\theta)) + 2n_3 \ln(1-\theta) \\
&= D + 2n_1 \ln \theta + n_2 \ln \theta + n_2 \ln(1-\theta) + 2n_3 \ln(1-\theta) \\
&= D + (2n_1 + n_2) \ln \theta + (n_2 + 2n_3) \ln(1-\theta),
\end{aligned}$$

où C et D ne dépendent pas de θ. Il est facile de vérifier que la valeur de θ qui maximise $\ln L_\theta(n_1, n_2, n_3)$ est donnée par

$$\hat{\theta} = \frac{2n_1 + n_2}{2n} = \frac{19}{40}.$$

La statistique de test prend donc la valeur

$$\begin{aligned}
\sum_{i=1}^{3} \frac{(n_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} &= \frac{(125 - 500p_1(\frac{19}{40}))^2}{500p_1(\frac{19}{40})} + \frac{(225 - 500p_2(\frac{19}{40}))^2}{500p_2(\frac{19}{40})} + \frac{(150 - 500p_3(\frac{19}{40}))^2}{500p_3(\frac{19}{40})} \\
&\approx 4.78,
\end{aligned}$$

qui, au niveau $\alpha = 5\%$, est à comparer à $\chi^2_{3-1-1;0.95} = \chi^2_{1;0.95} \approx 3.84$. Puisque $4.78 > 3.84$, il convient donc de rejeter l'hypothèse nulle au niveau 5%. On peut vérifier que ceci est compatible avec la p-valeur qui vaut ici à peu près 0.03 (ce qui indique que l'hypothèse nulle ne serait pas rejetée au niveau $\alpha = 1\%$).

Comme on vient de le voir, l'adéquation d'un modèle statistique peut être testée au moyen du théorème IV.19. Ceci s'étend à des problèmes où on veut tester que le modèle statistique d'échantillonnage paramétrique adopté est adéquat. Ceci couvre aussi les cas où la loi commune des observations à tester est discrète avec un nombre infini de valeurs possibles ou est carrément continue, mais il faut bien entendu alors constituer des classes pour se ramener au schéma multinomial dans lequel on peut considérer les résultats ci-dessus.

Pour illustrer rapidement ceci, imaginons qu'on observe $X_1, \ldots, X_n$, qui sont des variables aléatoires i.i.d. représentant le nombre d'accidents de voiture dans une rue. On peut penser à tester l'hypothèse nulle que la loi commune des X_i est la loi de Poisson de paramètre λ pour un certain $\lambda > 0$ (l'hypothèse nulle spécifie donc qu'un modèle d'échantillonnage paramétrique de Poisson convient à la situation considérée). Tronquer le nombre d'accidents à $I-1$ permet de rendre ce problème de test compatible avec le schéma multinomial. On peut en effet alors considérer $N = (N_1, \ldots, N_I)$, où

$$N_i = \sum_{j=1}^{n} \mathbb{I}[X_j = i-1], \quad i = 1, \ldots, I-1$$

et

$$N_I = \sum_{j=1}^{n} \mathbb{I}[X_j \geq I-1].$$

Sous l'hypothèse nulle, ceci est associé à des probabilités de type Poisson données par

$$p_i(\lambda) = e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!}, \quad i = 1, \ldots, I-1$$

et

$$p_I(\lambda) = \sum_{i=I-1}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} = 1 - \sum_{i=1}^{I-1} p_i(\lambda).$$

Le test consistera à rejeter l'hypothèse nulle (au niveau asymptotique α) si

$$\sum_{i=1}^{I} \frac{(N_i - np_i(\hat{\lambda}))^2}{np_i(\hat{\lambda})} > \chi^2_{I-2,1-\alpha},$$

où $\hat{\lambda}$ est l'estimateur de maximum de vraisemblance de λ.

IV.6.3 Tests d'homogénéité

Considérons une situation dans laquelle on observe J multinomiales mutuellement indépendantes

$$N_j = (N_{1j}, \ldots, N_{Ij}) \sim \text{Mult}(n_j, p_{1j}, \ldots, p_{I-1,j}), \qquad j = 1, \ldots, J.$$

Le nombre de modalités (I) est le même pour chaque multinomiale, mais les autres paramètres peuvent dépendre de j. Par exemple, les multinomiales N_j peuvent représenter les nombres de personnes qui, parmi les n_j personnes interrogées dans le pays j, se sont prononcées en faveur de

l'extrême gauche ($i = 1$), de la gauche ($i = 2$), ..., de l'extrême droite ($i = I$). Dans ce cadre, un problème de test naturel est

$$\begin{cases} \mathcal{H}_0 : \forall i \in \{1, \dots, I\},\ p_{i1} = \dots = p_{iJ} \\ \mathcal{H}_1 : \exists i \in \{1, \dots, I\} \text{ tels que } p_{ij} \neq p_{ij'} \text{ pour certains } j \neq j'. \end{cases} \tag{IV.26}$$

Dans le cadre politique ci-dessus, l'hypothèse nulle considérée signifie que les poids des différents courants politiques sont exactement les mêmes dans les J pays considérés.

Un test pour ce problème peut être obtenu de la façon suivante. Le théorème IV.18 livre que, pour tout $j = 1, \dots, J$,

$$\sum_{i=1}^{I} \frac{(N_{ij} - n_j p_{ij})^2}{n_j p_{ij}} \xrightarrow{\mathcal{L}} \chi^2_{I-1}$$

quand $n \to \infty$. Puisque les multinomiales sont mutuellement indépendantes, on en déduit que

$$\sum_{j=1}^{J} \sum_{i=1}^{I} \frac{(N_{ij} - n_j p_{ij})^2}{n_j p_{ij}} \xrightarrow{\mathcal{L}} \chi^2_{J(I-1)}$$

quand $n \to \infty$. Sous l'hypothèse nulle, ceci peut se réécrire

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - n_j p_i)^2}{n_j p_i} \xrightarrow{\mathcal{L}} \chi^2_{J(I-1)},$$

où p_i est la valeur commune des probabilités $p_{i1}, \dots, p_{iJ}$. Sous l'hypothèse nulle, les estimateurs de maximum de vraisemblance de p_i, $i = 1, \dots, I$ sont donnés par

$$\hat{p}_i = \frac{N_{i\cdot}}{n} \overset{\text{def}}{=} \frac{1}{n} \sum_{j=1}^{J} N_{ij}, \qquad i = 1, \dots, I.$$

En notant qu'on n'a estimé que $I - 1$ paramètres indépendants (puisque $p_I = 1 - \sum_{i=1}^{I-1} p_i$), on peut montrer que, en ligne avec le théorème IV.19, on a

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - n_j \hat{p}_i)^2}{n_j \hat{p}_i} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - \frac{N_{i\cdot} n_j}{n})^2}{\frac{N_{i\cdot} n_j}{n}} \xrightarrow{\mathcal{L}} \chi^2_{J(I-1)-(I-1)} = \chi^2_{(I-1)(J-1)}$$

sous $\mathcal{H}_0$ quand $n \to \infty$. Le test qui en résulte consiste donc à rejeter l'hypothèse nulle d'homogénéité (au niveau asymptotique α) si et seulement si

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - \frac{N_{i\cdot} n_j}{n})^2}{\frac{N_{i\cdot} n_j}{n}} > \chi^2_{(I-1)(J-1),1-\alpha}. \tag{IV.27}$$

Exemple : une enquête a été réalisée, lors d'une période de récession économique, auprès de trois groupes d'acteurs de la vie économique : les chefs d'entreprises, les économistes et les hommes politiques. Dans chaque groupe, 100 individus ont été interrogés sur les décisions qui, selon eux,

seraient les plus favorables à une sortie de la récession. Les résultats sont les suivants.

	Chefs d'entreprise	Economistes	Hommes politiques
Accroître le déficit budgétaire	10	15	29
Réduire les impôts	37	37	33
Réduire les taux d'intérêts	24	34	25
Offrir des avantages fiscaux aux entreprises	29	14	13
	$n_1 = 100$	$n_2 = 100$	$n_3 = 100$

Peut-on conclure, au niveau $\alpha = 5\%$, que les opinions sont réparties de façon différente dans ces trois groupes ? Dans cette situation où $I = 4$, $J = 3$ et $n = 300$, on vérifie que la statistique

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij} - \frac{N_{i\cdot}n_j}{n})^2}{\frac{N_{i\cdot}n_j}{n}}$$

prend la valeur 21.877 et que la valeur critique est donnée par

$$\chi^2_{(I-1)(J-1),1-\alpha} = \chi^2_{6,0.95} \approx 12.59.$$

Au niveau 5%, on rejette donc l'hypothèse nulle que les opinions sont réparties de la même façon dans les trois groupes. Le rejet aurait en fait lieu à tous les niveaux usuels (sauf $\alpha = 0.001$) car la p-valeur correspondante est donnée par $\mathrm{P}[\chi^2_6 > 21.877] \approx 0.0013$.

IV.6.4 Tests d'indépendance

Considérons un vecteur aléatoire (X, Y), où X ne peut prendre que I valeurs possibles $x_1, \ldots, x_I$ et Y ne peut prendre que J valeurs possibles $y_1, \ldots, y_J$ ($I, J \in \mathbb{N} \setminus \{0, 1\}$). La distribution jointe de (X, Y) est déterminée par les probabilités

$$p_{ij} = \mathrm{P}[X = x_i, Y = y_j] \quad i = 1, \ldots, I,\ j = 1, \ldots, J,$$

dont on peut déduire les distributions marginales de X et de Y, associées respectivement aux probabilités

$$p_{i\cdot} = \mathrm{P}[X = x_i] = \sum_{j=1}^{J} p_{ij} \quad i = 1, \ldots, I$$

et

$$p_{\cdot j} = \mathrm{P}[Y = y_j] = \sum_{i=1}^{I} p_{ij} \quad j = 1, \ldots, J.$$

Dans cette section, nous désirons tester l'hypothèse nulle que X et Y sont des variables aléatoires indépendantes sur la base de n observations i.i.d. $(X_1, Y_1), \ldots, (X_n, Y_n)$. Ceci mène au problème de

test

$$\begin{cases} \mathcal{H}_0 : \forall i, j,\ p_{ij} = p_{i.}p_{.j} \\ \mathcal{H}_1 : \exists i, j \text{ tel que } p_{ij} \neq p_{i.}p_{.j}. \end{cases} \tag{IV.28}$$

Si on note $N_{ij} = \sum_{\ell=1}^{n} \mathbb{I}[(X_\ell, Y_\ell) = (x_i, y_j)]$ pour tout i, j, $N = (N_{11}, N_{12}, \ldots, N_{IJ})$ est de loi multinomiale de paramètres $n, p_{11}, p_{12}, \ldots, p_{IJ-1}$. Par suite, le théorème IV.18 assure que

$$\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(N_{ij} - np_{ij})^2}{np_{ij}} \xrightarrow{\mathcal{L}} \chi^2_{IJ-1}$$

quand $n \to \infty$. Sous $\mathcal{H}_0$, ce résultat se réécrit

$$\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(N_{ij} - np_{i.}p_{.j})^2}{np_{i.}p_{.j}} \xrightarrow{\mathcal{L}} \chi^2_{IJ-1}.$$

Sous $\mathcal{H}_0$, les estimateurs du maximum de vraisemblance des $p_{i.}$ et des $p_{.j}$ sont donnés par

$$\hat{p}_{i.} = \frac{N_{i.}}{n} \overset{\text{def}}{=} \frac{1}{n}\sum_{j=1}^{J} N_{ij}, \qquad i = 1, \ldots, I$$

et

$$\hat{p}_{.j} = \frac{N_{.j}}{n} \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{I} N_{ij}, \qquad j = 1, \ldots, J.$$

Puisque ceci est associé à l'estimation de $(I-1)+(J-1)$ paramètres indépendants, on peut montrer qu'en ligne avec le théorème IV.19, on a

$$\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(N_{ij} - n\hat{p}_{i.}\hat{p}_{.j})^2}{n\hat{p}_{i.}\hat{p}_{.j}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(N_{ij} - \frac{N_{i.}N_{.j}}{n})^2}{\frac{N_{i.}N_{.j}}{n}} \xrightarrow{\mathcal{L}} \chi^2_{IJ-1-(I-1)-(J-1)} = \chi^2_{(I-1)(J-1)}$$

sous $\mathcal{H}_0$ quand $n \to \infty$. Le test qui en résulte consiste donc à rejeter l'hypothèse nulle d'indépendance (au niveau asymptotique α) si et seulement si

$$\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(N_{ij} - \frac{N_{i.}N_{.j}}{n})^2}{\frac{N_{i.}N_{.j}}{n}} > \chi^2_{(I-1)(J-1),1-\alpha}. \tag{IV.29}$$

Exemple : un ensemble de $n = 2100$ certificats de décès d'adultes, enregistrés auprès des hôpitaux d'une grande ville, ont été classés de la façon suivante (X : cause du décès ; Y : habitudes tabagiques).

	Gros fumeurs	Fumeurs légers	Non-fumeurs
Décès d'origine respiratoire	55	120	162
Décès d'origine cardio-vasculaire	49	388	315
Autres causes de décès	61	300	650

Ces données permettent-elles de rejeter (au niveau 5%) l'hypothèse d'indépendance entre habitudes tabagiques et causes de décès ? On vérifie que la statistique

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij}-\frac{N_{i.}N_{.j}}{N})^2}{\frac{N_{i.}N_{.j}}{N}}$$

prend la valeur 134.12 et que la valeur critique est donnée par

$$\chi^2_{(I-1)(J-1),1-\alpha}=\chi^2_{4,0.95}\approx 9.49.$$

Au niveau 5%, on rejette donc l'hypothèse nulle d'indépendance. Ce rejet ne pourrait pas être plus clair car la p-valeur est en fait inférieure à 10^{-15}.

Il est nécessaire de faire remarquer que, à notation près, les expressions (IV.27) et (IV.29) sont rigoureusement identiques. Il s'agit du même test, faisant intervenir les mêmes quantités. La différence entre le test d'homogénétié et le test d'indépendance se joue donc sur l'interprétation du problème de test. Dans le premier cas, le statisticien se demande si différentes *populations* se répartissent de la même manière dans un facteur donné. Le second test cherche à mesure l'indépendance entre deux facteurs (qui jouent, là, un rôle symétrique).

Chapitre V

Intervalles de confiance

V.1 Introduction

Dans un premier temps, considérons un modèle statistique paramétrique à paramètre scalaire de la forme

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{ \mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R} \right\}\right), \tag{V.1}$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Dans ce dernier chapitre, nous désirons construire, sur la base de $X^{(n)}$, un sous-ensemble aléatoire de Θ qui contienne la vraie valeur du paramètre avec une certaine probabilité cible (typiquement élevée). Ce sous-ensemble sera appelé *zone de confiance* (ou *intervalle de confiance* dans le cas où il s'agit d'un intervalle) et la probabilité cible *niveau de confiance*. Plutôt que de fournir une estimation ponctuelle $T(x^{(n)})(\in \Theta)$ basée sur la valeur $x^{(n)} = (x_1, \ldots, x_n)$ prise par $X^{(n)}$, l'idée est donc ici de donner une collection $S(x^{(n)})(\subset \Theta)$ de valeurs du paramètre θ qui soient plausibles au vu de $x^{(n)}$.

Nous adoptons la définition suivante (dans tout ce chapitre, on supposera tacitement que Θ est connexe).

Definition V.1 *Considérons le modèle statistique paramétrique à paramètre scalaire en (V.1) et fixons $\alpha \in (0,1)$. Alors $[L(X^{(n)}), U(X^{(n)})]$ est un* intervalle de confiance *pour θ au* niveau de confiance $1-\alpha$ *si et seulement si les statistiques $L(X^{(n)})$ et $U(X^{(n)})$ vérifient les deux conditions suivantes pour tout $\theta \in \Theta$: (i) $L(X^{(n)}) \leq U(X^{(n)})$ p.s. sous $\mathrm{P}_\theta^{(n)}$ et (ii) $\mathrm{P}_\theta^{(n)}\left[\theta \in [L(X^{(n)}), U(X^{(n)})]\right] \geq 1 - \alpha$.*

Lorsqu'on a observé la valeur $x^{(n)}$ de $X^{(n)}$, on peut bien entendu considérer l'intervalle de confiance $[L(x^{(n)}), U(x^{(n)})]$. Celui-ci est souvent interprété de façon erronée, en disant que "la vraie valeur de θ (celle qui est telle que $X^{(n)} \sim \mathrm{P}_\theta^{(n)}$) a une probabilité au moins $1-\alpha$ d'appartenir à l'intervalle $[L(x^{(n)}), U(x^{(n)})]$". Ceci n'a cependant pas de sens : $[\theta \in [L(x^{(n)}), U(x^{(n)})]]$ n'est pas un événement aléatoire (car ni θ, ni $[L(x^{(n)}), U(x^{(n)})]$ n'est aléatoire), mais une proposition logique qui est soit vraie soit fausse (si on voulait regarder cette proposition comme un événement aléatoire, celui-ci recevrait donc une probabilité 0 ou 1, et donc pas $1-\alpha$).

L'interprétation correcte est de considérer $[L(x^{(n)}), U(x^{(n)})]$ comme la réalisation d'un intervalle aléatoire $[L^{(n)}(X^{(n)}), U^{(n)}(X^{(n)})]$, lequel contient la vraie valeur de θ avec une probabilité d'au moins $1-\alpha$. Par conséquent, il faut comprendre que si on répète m fois (de façon indépendante) l'expérience aléatoire consistant à observer $X^{(n)} = (X_1, \ldots, X_n)$, la proportion des m intervalles de confiance $[L(X^{(n)}), U(X^{(n)})]$ contenant la vraie valeur de θ tendra vers un nombre plus grand ou égal à $1-\alpha$ quand $m \to \infty$.

Dans ce chapitre, nous allons notamment étudier comment on peut construire de tels intervalles de confiance.

V.2 Intervalles exacts

V.2.1 Sans paramètre de nuisance

Nous construisons d'abord un intervalle de confiance dans un exemple. Soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma_0^2)$, avec σ_0^2 un réel strictement positif fixé, de sorte que le paramètre est $\theta = \mu \in \Theta = \mathbb{R}$ (version à variance connue de l'exemple du verre de bière). Sous $\mathrm{P}_\mu^{(n)}$ (la mesure de probabilité associée à $X^{(n)}$ lorsque la vraie valeur du paramètre est μ), la variable aléatoire

$$g(X^{(n)}, \mu) = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_0}$$

est de loi $\mathcal{N}(0,1)$. Donc, en notant Φ la fonction de répartition de cette loi, on a que

$$\forall \mu \in \mathbb{R}, \quad \mathrm{P}_\mu^{(n)}\left[\Phi^{-1}\Big(\frac{\alpha}{2}\Big) \le \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_0} \le \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\right] = 1 - \alpha. \tag{V.2}$$

En isolant μ dans l'événement considéré, ceci se réécrit

$$\forall \mu \in \mathbb{R}, \quad \mathrm{P}_\mu^{(n)}\left[\bar{X} + \Phi^{-1}\Big(\frac{\alpha}{2}\Big)\frac{\sigma_0}{\sqrt{n}} \le \mu \le \bar{X} + \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\frac{\sigma_0}{\sqrt{n}}\right] = 1 - \alpha.$$

Par conséquent,

$$\begin{aligned}\big[L(X^{(n)}), U(X^{(n)})\big] &\stackrel{\text{def}}{=} \left[\bar{X} + \Phi^{-1}\Big(\frac{\alpha}{2}\Big)\frac{\sigma_0}{\sqrt{n}}, \bar{X} + \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\frac{\sigma_0}{\sqrt{n}}\right] \\ &= \left[\bar{X} - \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\frac{\sigma_0}{\sqrt{n}}, \bar{X} + \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\frac{\sigma_0}{\sqrt{n}}\right] \end{aligned} \tag{V.3}$$

est un intervalle de confiance au niveau $1-\alpha$ au sens de la définition ci-dessus.

Ce développement appelle quelques commentaires. Tout d'abord, plutôt que (V.2), nous aurions pu écrire

$$\forall \mu \in \mathbb{R}, \quad \mathrm{P}_\mu^{(n)}\left[\Phi^{-1}\big(\alpha - u\big) \le \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_0} \le \Phi^{-1}\big(1 - u\big)\right] = 1 - \alpha,$$

pour un $u \in (0, \alpha)$ arbitraire. Par le même raisonnement, ceci aurait mené à l'intervalle de confiance

$$\begin{aligned}\big[L_u(X^{(n)}), U_u(X^{(n)})\big] &\stackrel{\text{def}}{=} \left[\bar{X} - \Phi^{-1}\big(1-u\big)\frac{\sigma_0}{\sqrt{n}}, \bar{X} - \Phi^{-1}\big(\alpha-u\big)\frac{\sigma_0}{\sqrt{n}}\right] \\ &= \left[\bar{X} - \Phi^{-1}\big(1-u\big)\frac{\sigma_0}{\sqrt{n}}, \bar{X} + \Phi^{-1}\big(1-(\alpha-u)\big)\frac{\sigma_0}{\sqrt{n}}\right].\end{aligned}$$

Parmi tous les intervalles de confiance $\big[L_u(X^{(n)}), U_u(X^{(n)})\big]$, l'intervalle associé à $u = \alpha/2$, qui est l'intervalle en (V.3), est le seul à respecter les propriétés de symétrie du problème, mais est aussi celui pour lequel la longueur de l'intervalle de confiance (c'est-à-dire $U_u(X^{(n)}) - L_u(X^{(n)})$) est la plus petite !

Pour le reste, on constate directement que la longueur de l'intervalle de confiance en (V.3) est une fonction croissante de $1-\alpha$ et σ_0, mais décroissante de n, ce qui s'interprète facilement.

Cet exemple, qui peut paraître simpliste, porte en lui l'essence de la méthode la plus classique pour construire des intervalles de confiance, à savoir la *méthode des fonctions pivotales*. Décrivons cette méthode dans le cadre général du modèle statistique paramétrique à paramètre scalaire en (V.1). Pour ce faire, le concept important est le suivant.

Definition V.2 *Désignons par $\mathcal{X}^{(n)}$ l'ensemble des valeurs possibles de $x^{(n)}$. Alors la fonction $g : \mathcal{X}^{(n)} \times \Theta \to \mathbb{R}$ est pivotale pour θ si et seulement si les trois conditions suivantes sont satisfaites : (i) pour tout $\theta \in \Theta$, la fonction $x^{(n)} \mapsto g(x^{(n)}, \theta)$ est mesurable ; (ii) pour tout $x^{(n)} \in \mathcal{X}^{(n)}$ (l'ensemble des valeurs possibles de $x^{(n)}$), la fonction $\theta \mapsto g(x^{(n)}, \theta)$ est strictement monotone (croissante ou décroissante) ; (iii) la distribution de $g(X^{(n)}, \theta)$ sous $\mathrm{P}_\theta^{(n)}$ ne dépend pas de θ.*

Dans le cadre de l'exemple ci-dessus, la fonction définie par

$$g(x^{(n)}, \mu) = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma_0}$$

est clairement pivotale pour μ.

Revenons à la description de la méthode des fonctions pivotales. Puisque la distribution de $g(X^{(n)}, \theta)$ sous $\mathrm{P}_\theta^{(n)}$ ne dépend pas de θ, nous pouvons trouver deux réels a, b ($a < b$) *ne dépendant pas de θ* et tels que

$$\forall \theta \in \Theta, \quad \mathrm{P}_\theta^{(n)}\Big[a \leq g(X^{(n)}, \theta) \leq b\Big] \geq 1 - \alpha \tag{V.4}$$

(on choisira idéalement a et b de telle sorte l'égalité ait lieu en (V.4), mais cela ne sera pas toujours possible lorsque la distribution de la fonction pivotale est discrète). Notons que, comme dans l'exemple, ces nombres a et b ne sont bien entendu pas uniques.

Puisque la fonction $\theta \mapsto g(x^{(n)}, \theta)$ est strictement monotone, elle est inversible, ce qui permet de réécrire (V.4) sous la forme

$$\forall \theta \in \Theta, \quad \mathrm{P}_\theta^{(n)}\Big[\theta \in g^{-1}(X^{(n)}, \cdot)([a, b])\Big] \geq 1 - \alpha,$$

où $g^{-1}(X^{(n)}, \cdot)([a,b])$ désigne l'image de $[a,b]$ par la fonction réciproque de la fonction $t \mapsto g(X^{(n)}, t)$. Il est facile de vérifier que cette fonction réciproque est continue, de sorte que $g^{-1}(X^{(n)}, \cdot)([a,b])$ est bien un *intervalle* de confiance pour θ au niveau $1-\alpha$. Souvent, on pourra lever le degré de liberté sur a et b mentionné ci-dessus en imposant que l'intervalle de confiance soit de longueur minimale.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. Unif$([0,\theta])$, où $\theta \in \Theta = \mathbb{R}_0^+$ (exemple du bus). Comme nous l'avons vu, l'estimateur du maximum de vraisemblance pour θ dans ce modèle est $X_{(n)}$. Il découle directement de (II.4) que la distribution de $X_{(n)}/\theta$ a pour fonction de répartition

$$x \mapsto F_\theta^{X_{(n)}/\theta}(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^n & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1. \end{cases} \tag{V.5}$$

Par conséquent, on vérifiera facilement que

$$g(x^{(n)}, \theta) = \frac{x_{(n)}}{\theta}$$

définit une fonction pivotale pour θ. Puisque le quantile d'ordre $\beta (\in [0,1])$ de la loi de fonction de répartition en (V.5) est donné par $q_\beta = \beta^{1/n}$, la méthode générale décrite ci-dessus conduit à écrire

$$\forall \theta > 0, \quad \mathrm{P}_\theta^{(n)}\left[(\alpha - u)^{1/n} \leq \frac{X_{(n)}}{\theta} \leq (1-u)^{1/n}\right] = 1 - \alpha,$$

pour un $u \in [0, \alpha]$ arbitraire Les intervalles de confiance qui en résultent sont obtenus en réécrivant cette dernière relation sous la forme

$$\forall \theta > 0, \quad \mathrm{P}_\theta^{(n)}\left[\frac{X_{(n)}}{(1-u)^{1/n}} \leq \theta \leq \frac{X_{(n)}}{(\alpha - u)^{1/n}}\right] = 1 - \alpha;$$

pour $u = \alpha$, ceci mène à l'intervalle de confiance non borné $[X_{(n)}/(1-\alpha)^{1/n}, +\infty)$. Parmi cette famille d'intervalles de confiance indicée par $u \in [0, \alpha]$, on vérifiera à titre d'exercice que c'est celui associé à $u = 0$ qui est le plus court. L'intervalle de confiance "optimal" fondé sur $X_{(n)}$ au niveau de confiance $1 - \alpha$ est donc donné par

$$\left[X_{(n)}\,, \frac{X_{(n)}}{\alpha^{1/n}}\right].$$

En particulier, dans cet exemple, l'intervalle de confiance

$$\left[\frac{X_{(n)}}{(1-\alpha/2)^{1/n}}\,, \frac{X_{(n)}}{(\alpha/2)^{1/n}}\right]$$

associé aux quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ de la loi de $g(X^{(n)}, \theta)$ n'est pas de longueur minimale (et on se méfiera donc de la pratique courante qui veut qu'*on coupe systématiquement une probabilité* $\alpha/2$ *à gauche et à droite*).

V.2.2 Avec paramètres de nuisance

Considérons le modèle statistique paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{ \mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k \right\}\right), \tag{V.6}$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$. Nous désirons ici construire un intervalle de confiance au niveau $1-\alpha$ pour l'une des composantes de $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. Sans perte de généralité, nous pouvons nous restreindre au problème de la construction d'un intervalle de confiance pour θ_1, pour lequel θ_1 joue le rôle de paramètre d'intérêt et $\theta_2, \ldots, \theta_k$ jouent le rôle de paramètres de nuisance. Commençons par préciser la notion d'intervalle de confiance pour θ_1 dans ce cadre.

Definition V.3 *Considérons le modèle statistique paramétrique (V.6) et fixons $\alpha \in (0,1)$. Alors $[L(X^{(n)}), U(X^{(n)})]$ est un* intervalle de confiance *pour θ_1 au* niveau de confiance $1-\alpha$ *si et seulement si les statistiques $L(X^{(n)})$ et $U(X^{(n)})$ vérifient les deux conditions suivantes pour tout $\theta \in \Theta$: (i) $L(X^{(n)}) \leq U(X^{(n)})$ p.s. sous $\mathrm{P}_\theta^{(n)}$ et (ii) $\mathrm{P}_\theta^{(n)}\big[\theta_1 \in [L(X^{(n)}), U(X^{(n)})]\big] \geq 1-\alpha$.*

Il est facile d'étendre à cette situation le concept de fonction pivotale.

Definition V.4 *Dans le cadre du modèle paramétrique ci-dessus, notons Θ_1 l'ensemble des valeurs que peut prend le paramètre θ_1 lorsque θ parcourt Θ. Alors la fonction $g : \mathbb{R}^n \times \Theta_1 \to \mathbb{R}$ est pivotale pour θ_1 si et seulement si les trois conditions suivantes sont satisfaites : (i) pour tout $\theta_1 \in \Theta_1$, la fonction $x^{(n)} \mapsto g(x^{(n)}, \theta_1)$ est mesurable ; (ii) pour tout $x^{(n)} \in \mathcal{X}^{(n)}$, la fonction $\theta_1 \mapsto g(x^{(n)}, \theta_1)$ est strictement monotone (croissante ou décroissante) ; (iii) la distribution de $g(X^{(n)}, \theta_1)$ sous $\mathrm{P}_\theta^{(n)}$ ne dépend pas de θ.*

Pour peu que nous disposions d'une telle fonction pivotale, la construction d'un intervalle de confiance pour θ_1 au niveau de confiance $1-\alpha$ se fait exactement suivant les mêmes principes qu'à la section précédente.

Exemple : soit $X^{(n)} = (X_1, \ldots, X_n)$, où les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, où $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+$ (exemple du verre de bière). Une fonction pivotale pour μ est définie par

$$g(x^{(n)}, \mu) = \frac{\sqrt{n-1}(\bar{x} - \mu)}{s},$$

où s est la racine carrée de la variance empirique $s^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. Rappelons que le corollaire (II.4) implique en effet que, quels que soient μ et σ^2, la loi de $g(X^{(n)}, \mu)$ sous $\mathrm{P}_{\mu,\sigma^2}^{(n)}$ est la loi de Student t_{n-1}. Par conséquent, si on désigne par $t_{n-1,\alpha}$ le quantile d'ordre α d'une loi de Student à $n-1$ degrés de liberté, on a

$$\forall (\mu, \sigma^2) \in \Theta, \quad \mathrm{P}_{\mu,\sigma^2}^{(n)}\left[t_{n-1,\alpha/2} \leq \frac{\sqrt{n-1}(\bar{X} - \mu)}{s} \leq t_{n-1,1-\alpha/2} \right] = 1-\alpha,$$

ce qui se réécrit

$$\forall(\mu,\sigma^2)\in\Theta,\quad \mathrm{P}_{\mu}^{(n)}\left[\bar{X}+t_{n-1,\alpha/2}\frac{s}{\sqrt{n-1}}\leq\mu\leq\bar{X}+t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n-1}}\right]=1-\alpha.$$

On en conclut que

$$\begin{aligned}&\left[\bar{X}+t_{n-1,\alpha/2}\frac{s}{\sqrt{n-1}},\bar{X}+t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n-1}}\right]\\&\qquad=\left[\bar{X}-t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n-1}},\bar{X}+t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n-1}}\right]\end{aligned}\tag{V.7}$$

est un intervalle de confiance pour μ au niveau de confiance $1-\alpha$.

Si on désire plutôt un intervalle de confiance pour σ^2, on peut utiliser la fonction pivotale pour σ^2 définie par

$$g(x^{(n)},\sigma^2)=\frac{ns^2}{\sigma^2}.$$

Rappelons que le lemme de Fisher (Théorème II.2) affirme en particulier que, quels que soient μ et σ^2, la loi de $g(X^{(n)},\sigma^2)$ sous $\mathrm{P}_{\mu,\sigma^2}^{(n)}$ est la loi χ^2_{n-1}. On vérifiera alors à titre d'exercice que

$$\left[\frac{ns^2}{\chi^2_{n-1,1-\alpha/2}},\frac{ns^2}{\chi^2_{n-1,\alpha/2}}\right]$$

est un intervalle de confiance pour σ^2 au niveau de confiance $1-\alpha$. Comme précédemment, on aura noté $\chi^2_{n-1,\alpha}$ le quantile d'ordre α d'une loi chi-carré à $(n-1)$ degrés de liberté.

V.3 Intervalles asymptotiques

La construction des intervalles de confiance repose crucialement sur l'existence d'une fonction pivotale pour le problème considéré. Dans de nombreux cas, il est cependant difficile (voire impossible) d'identifier une telle fonction. Tout n'est pas perdu pour autant car, pour peu que la taille d'échantillon ne soit pas trop petite (en fait, soit suffisamment grande pour que les théorèmes de convergence en loi livrent une approximation raisonnable), on peut dans la plupart des cas construire des intervalles de confiance *approchés* qui, en pratique, sont aussi satisfaisants que leurs équivalents *exacts*.

Considérons le modèle statistique paramétrique

$$\left(\mathbb{R}^n,\mathcal{B}^n,\mathcal{P}^{(n)}=\left\{\mathrm{P}_{\theta}^{(n)}:\theta\in\Theta\subset\mathbb{R}^k\right\}\right),\tag{V.8}$$

où les mesures de probabilité $\mathrm{P}_{\theta}^{(n)}$ représentent les distributions possibles pour $X^{(n)}=(X_1,\ldots,X_n)$. Nous désirons ici construire un intervalle de confiance *asymptotique* pour θ_1 au niveau $1-\alpha$ au sens suivant.

Definition V.5 *Considérons le modèle statistique paramétrique (V.8) et fixons $\alpha\in(0,1)$. Alors*

$[L(X^{(n)}), U(X^{(n)})]$ *est un* intervalle de confiance asymptotique *pour* θ_1 *au* niveau de confiance $1-\alpha$ *si et seulement si les statistiques* $L(X^{(n)})$ *et* $U(X^{(n)})$ *vérifient les deux conditions suivantes et pour tout* $\theta \in \Theta$ *: (i) pour tout* n, $L(X^{(n)}) \leq U(X^{(n)})$ *presque sûrement sous* $\mathrm{P}_\theta^{(n)}$ *et (ii)* $\mathrm{P}_\theta^{(n)}\big[\theta_1 \in [L(X^{(n)}), U(X^{(n)})]\big] \to 1-\alpha$ *quand* $n \to \infty$.

On pourrait en fait seulement demander en (ii) que $\mathrm{P}_\theta^{(n)}\big[\theta_1 \in [L(X^{(n)}), U(X^{(n)})]\big]$ soit plus grand ou égal à α pour n suffisamment grand, mais nous adoptons la définition ci-dessus car il est en général possible d'avoir asymptotiquement un contenu de probabilité égal à $1-\alpha$ (ce qui se traduit par des intervalles de confiance asymptotiques plus courts).

Les situations pour lesquelles on peut construire un tel intervalle sont nombreuses. L'une d'elle est celle où on dispose d'un estimateur $T(X^{(n)}) = (T_1(X^{(n)}), \ldots, T_k(X^{(n)}))$ de $\theta = (\theta_1, \ldots, \theta_k)$ qui est faiblement convergent et dont la composante $T_1(X^{(n)})$ satisfait un résultat de normalité asymptotique affirmant que, pour tout $\theta \in \Theta$,

$$\sqrt{n}(T_1(X^{(n)}) - \theta_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta)) \quad \text{sous } \mathrm{P}_\theta^{(n)}$$

quand $n \to \infty$, où $\sigma^2(\cdot)$ est une fonction continue. En pareil cas, il est clair que, pour tout $\theta \in \Theta$,

$$\frac{\sqrt{n}(T_1(X^{(n)}) - \theta_1)}{\sigma(\theta)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{sous } \mathrm{P}_\theta^{(n)} \tag{V.9}$$

quand $n \to \infty$. Il découle alors des théorèmes A.6 et A.8 que, pour tout $\theta \in \Theta$,

$$\frac{\sqrt{n}(T_1(X^{(n)}) - \theta_1)}{\sigma(T(X^{(n)}))} = \frac{\sigma(\theta)}{\sigma(T(X^{(n)}))} \times \frac{\sqrt{n}(T_1(X^{(n)}) - \theta_1)}{\sigma(\theta)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{sous } \mathrm{P}_\theta^{(n)} \tag{V.10}$$

quand $n \to \infty$. On peut donc considérer que la fonction définie par

$$g(x^{(n)}, \theta_1) = \frac{\sqrt{n}(T_1(X^{(n)}) - \theta_1)}{\sigma(T(X^{(n)}))}$$

est une fonction "*asymptotiquement pivotale*" pour θ_1.

Quel que soit $\alpha \in (0,1)$, on aura donc que, pour tout $\theta \in \Theta$,

$$\mathrm{P}_\theta^{(n)}\Big[-\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) \leq \frac{\sqrt{n}(T_1(X^{(n)}) - \theta_1)}{\sigma(T(X^{(n)}))} \leq \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\Big] \to 1-\alpha.$$

quand $n \to \infty$, ce qui livre que, pour tout $\theta \in \Theta$,

$$\mathrm{P}_\theta^{(n)}\left[T_1(X^{(n)}) - \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\frac{\sigma(T(X^{(n)}))}{\sqrt{n}} \leq \theta_1 \leq T_1(X^{(n)}) + \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\frac{\sigma(T(X^{(n)}))}{\sqrt{n}}\right] \to 1-\alpha.$$

En d'autres termes,

$$\left[T_1(X^{(n)}) - \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\frac{\sigma(T(X^{(n)}))}{\sqrt{n}}, T_1(X^{(n)}) + \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\frac{\sigma(T(X^{(n)}))}{\sqrt{n}}\right]$$

est un intervalle de confiance asymptotique pour θ_1 au niveau $1-\alpha$.

Exemple : considérons le modèle statistique d'échantillonnage pour lequel l'observation $X^{(n)} = (X_1, \ldots, X_n)$ regroupe des variables aléatoires dont la loi commune admet des moments finis d'ordre 2, mais est pour le reste non spécifiée. Pour ce modèle d'échantillonnage de nature non paramétrique, la collection des lois associées à $X^{(n)}$ peut par exemple être indicée par la fonction de répartition F commune des X_i. Posons $\mu = \mathrm{E}_F[X_1]$ et $\sigma^2 = \mathrm{Var}_F[X_1]$. Nous désirons construire un intervalle de confiance asymptotique pour μ au niveau $1-\alpha$ (la définition et la construction d'un tel intervalle s'étendent aisément au cas non paramétrique considéré).

Puisque le théorème central limite livre que, pour tout F,

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

quand $n \to \infty$ et que s^2 est un estimateur faiblement convergent de σ^2, on peut vérifier que la fonction définie par

$$g(x^{(n)}, \mu) = \frac{\sqrt{n}(\bar{X} - \mu)}{s}$$

est asymptotiquement pivotale pour μ. À titre d'exercice, on déterminera que l'intervalle asymptotique pour μ (au niveau de confiance $1-\alpha$) qui en résulte est

$$\left[\bar{X} - \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\frac{s}{\sqrt{n}}\,, \bar{X} + \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)\frac{s}{\sqrt{n}}\right].$$

Notons que dans le cas particulier où les observations sont i.i.d. de loi normale, cet intervalle de confiance aymptotique est équivalent, pour n grand, à l'intervalle de confiance exact obtenu en (V.7).

Pour terminer cette section, notons que, dans certains cas précis (l'exemple électoral en est un), il se peut que $\sigma^2(\theta)$ ne dépende que de θ_1 et que la fonction définie par

$$g(x^{(n)}, \theta_1) = \frac{\sqrt{n}(T(X^{(n)}) - \theta_1)}{\sigma(\theta_1)}$$

soit une fonction asymptotiquement pivotale en le sens ci-dessus (ce qui requiert que la fonction $\sigma(\theta_1)$ soit suffisamment simple pour qu'on puisse procéder à l'inversion de $\theta_1 \mapsto g(x^{(n)}, \theta_1)$). À titre d'exercice, on déterminera l'intervalle de confiance asymptotique pour p qui en résulte dans l'exemple électoral.

V.4 Lien avec les tests d'hypothèses

Considérons le modèle statistique paramétrique à paramètre scalaire

$$\Big(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \Big\{\mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}\Big\}\Big), \tag{V.11}$$

où les mesures de probabilité $\mathrm{P}_\theta^{(n)}$ représentent les distributions possibles pour $X^{(n)} = (X_1, \ldots, X_n)$ (les considérations de cette section s'étendent directement au cas où le modèle comprend aussi des

paramètres de nuisance).

Supposons que nous disposions d'un intervalle de confiance $[L(X^{(n)}), U(X^{(n)})]$ pour θ au niveau de confiance $1-\alpha$. Si nous considérons le problème de test

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0, \end{cases} \tag{V.12}$$

on peut penser à adopter la règle de décision qui consiste à rejeter $\mathcal{H}_0$ si et seulement si $\theta_0 \notin [L(X^{(n)}), U(X^{(n)})]$, puisque cet intervalle de confiance regroupe les valeurs "plausibles" de θ à un certain niveau de confiance. En d'autres termes, le test considéré ici est défini par

$$\phi(x^{(n)}) = \begin{cases} 1 & \text{si} \quad \theta_0 \notin [L(x^{(n)}), U(x^{(n)})] \\ 0 & \text{si} \quad \theta_0 \in [L(x^{(n)}), U(x^{(n)})]. \end{cases}$$

Le risque de première espèce de ce test est donné par

$$\begin{aligned} \mathrm{E}_{\theta_0}^{(n)}[\phi] &= \mathrm{P}_{\theta_0}^{(n)}\big[\theta_0 \notin [L(X^{(n)}), U(X^{(n)})]\big] \\ &= 1 - \mathrm{P}_{\theta_0}^{(n)}\big[\theta_0 \in [L(X^{(n)}), U(X^{(n)})]\big] \leq 1 - (1-\alpha) = \alpha, \end{aligned}$$

de sorte que le test ϕ est de niveau α pour le problème (V.12).

À l'inverse, supposons que ϕ soit un test (pur) de niveau α pour le problème (V.12). Pour insister sur la dépendance en θ_0 de ce test, nous le noterons ϕ_{θ_0}. Définissons alors

$$A(x^{(n)}) \stackrel{\text{def}}{=} \Big\{\theta_0 \in \Theta : \phi_{\theta_0}(x^{(n)}) = 0\Big\}.$$

L'ensemble $A(x^{(n)})$ est donc la collection des valeurs θ_0 du paramètre qui ne sont pas rejetées par le test correspondant lorsqu'on observe $x^{(n)}$. Il est alors naturel de prendre $A(X^{(n)})$ pour "intervalle" de confiance (il n'y a pas de garantie qu'il s'agisse toujours d'un intervalle, même si c'est le plus souvent le cas). Le niveau de confiance associé est $1-\alpha$ puisque

$$\mathrm{P}_{\theta}^{(n)}[\theta \in A(X^{(n)})] = \mathrm{P}_{\theta}^{(n)}[\phi_\theta(X^{(n)}) = 0] = 1 - \mathrm{P}_{\theta}^{(n)}[\phi_\theta(X^{(n)}) = 1] \geq 1 - \alpha.$$

Il existe donc un lien très fort entre tests d'hypothèses (bilatéraux) et intervalles de confiance. Citons deux implications intéressantes de ce lien. (a) Tout d'abord, il permet de corriger le fait que jusqu'ici, nous ne disposons pas de méthodes universelles pour construire un intervalle de confiance. Par contre, nous avons vu que des tests (comme ceux du rapport de vraisemblance) peuvent être construits dans des situations générales. Le lien entre tests d'hypothèses et intervalles de confiance permet donc de construire les intervalles associés à ces tests généraux, et par là, de construire de façon systématique des intervalles de confiance (les tests de rapport de vraisemblance peuvent typiquement être conduits de façon asymptotique, ce qui, par le même raisonnement que ci-dessus, pourra mener à des intervalles de confiance asymptotiques). (b) Les intervalles de confiance que nous avons construits ne bénéficient pas de propriété d'optimalité particulière (outre celle de minimiser,

dans certaines classes d'intervalles, la longueur). Ceci est en contraste avec les tests d'hypothèses du chapitre précédent, qui étaient typiquement à puissance uniformément maximale à un niveau α donné (parfois avec une restriction de non-biais). Le lien ci-dessus permet de construire des intervalles optimaux en un certain sens, en considérant les intervalles associés aux tests jouissant de ces propriétés d'optimalité.

V.5 Zones de confiance

Dans ce chapitre, lorsque les modèles statistiques étaient indicés par un paramètre vectoriel θ, nous avons supposé que *l'une* des composantes de θ seulement était un paramètre d'intérêt et que les autres composantes jouaient le rôle de paramètres de nuisance. Il est cependant parfois désirable de construire une *zone (ou région) de confiance* pour un paramètre vectoriel. Nous précisons maintenant cette notion.

Definition V.6 *Considérons le modèle statistique paramétrique (V.6) et fixons $\alpha \in (0,1)$. Alors la statistique $R^{(n)}(X^{(n)})$, à valeurs dans l'ensemble des parties de Θ, est une* zone de confiance *pour θ au* niveau de confiance $1-\alpha$ *si et seulement si $\mathrm{P}_\theta^{(n)}\big[\theta \in R(X^{(n)})\big] \geq 1-\alpha$ pour tout $\theta \in \Theta$.*

Il est à noter qu'il n'est pas demandé ici que $R^{(n)}$ soit un ensemble connexe. À strictement parler, la version univariée ($k=1$) du concept de zone de confiance est donc un peu plus générale que le concept d'intervalle de confiance, pour lequel $R(X^{(n)}) = [L(X^{(n)}), U(X^{(n)})]$ doit être connexe. Les zones de confiance univariées peuvent donc très bien avoir plusieurs composantes connexes (ou être de la forme $[L(X^{(n)}), \infty)$ ou $(-\infty, U(X^{(n)})]$).

Si la construction de zones de confiance dans le cas $k \geq 2$ singe celle fondée sur la méthode des fonctions pivotales dans le cas univarié, il est excessivement rare que les zones qui en résultent soient des hyper-rectangles (c'est-à-dire des produits d'intervalles). Il s'agit en fait souvent d'ellipsoïdes ou de zones plus complexes, qui sont typiquement plus difficiles à représenter pour k grand.

Pour cette raison, tenter de construire des zones de confiance qui soient des hyper-rectangles est une pratique courante. Sans vouloir rentrer ici dans les détails, nous faisons la remarque suivante. Supposons que nous obtenions, via la méthode décrite dans la Section V.2.2, et pour chaque $i = 1, \ldots, k$, un intervalle de confiance $[L_i(X^{(n)}), U_i(X^{(n)})]$ au niveau $1-\alpha$ fixé pour la $i$ème composante θ_i du paramètre θ. La zone de confiance hyper-rectangulaire qui en résulte est

$$\begin{aligned} R(X^{(n)}) &= [L_1(X^{(n)}), U_1(X^{(n)})] \times \ldots \times [L_k(X^{(n)}), U_k(X^{(n)})] \\ &= \Big\{\theta \in \Theta : \theta_i \in [L_i(X^{(n)}), U_i(X^{(n)})] \ \forall i = 1, \ldots, k\Big\}. \end{aligned} \tag{V.13}$$

Il est crucial de remarquer qu'en général, on a

$$\mathrm{P}_\theta^{(n)}[\theta \in R(X^{(n)})] < \mathrm{P}_\theta^{(n)}[\theta_i \in [L_i(X^{(n)}), U_i(X^{(n)})]]$$

pour tout i, et que le niveau de confiance associé à $R(X^{(n)})$ n'est donc pas clair. Par contre, il est clair que si les intervalles de confiance individuels ont été construits de sorte à recouvrir θ_i avec

une probabilité exactement égale à $1-\alpha$, $R^{(n)}$ ne sera pas une zone de confiance au niveau $1-\alpha$ (pourquoi ?)

Il est cependant possible de dire à quel niveau de confiance (plus élevé) il faut construire chaque intervalle de confiance individuel $[L_i(X^{(n)}), U_i(X^{(n)})]$ pour garantir que l'hyper-rectangle $R(X^{(n)})$ en (V.13) soit une zone de confiance pur θ au niveau de confiance $1-\alpha$. En effet, on a toujours que

$$\begin{aligned} \mathrm{P}_\theta^{(n)}[\theta \in R(X^{(n)})] &= 1 - \mathrm{P}_\theta^{(n)}[\theta \notin R(X^{(n)})] \\ &= 1 - \mathrm{P}_\theta^{(n)}\left[\bigcup_{i=1}^{k}\left[\theta_i \notin \left[L_i(X^{(n)}), U_i(X^{(n)})\right]\right]\right] \\ &\geq 1 - \sum_{i=1}^{k} \mathrm{P}_\theta^{(n)}\left[\theta_i \notin [L_i(X^{(n)}), U_i(X^{(n)})]\right]. \end{aligned}$$

Par conséquent, si chaque intervalle de confiance $[L_i(X^{(n)}), U_i(X^{(n)})]$ a un niveau de confiance de $1-\frac{\alpha}{k}$, on aura alors

$$\mathrm{P}_\theta^{(n)}[\theta \in R(X^{(n)})] \geq 1 - \sum_{i=1}^{k}\left(1 - \left(1 - \frac{\alpha}{k}\right)\right) = 1 - \alpha,$$

de sorte que l'hyper-rectangle $R(X^{(n)})$ en (V.13) sera effectivement une zone de confiance au niveau de confiance $1-\alpha$.

Annexe A

Quelques résultats de théorie des probabilités

A.1 Convergence stochastique

La *convergence stochastique* désigne la convergence d'une suite de variables aléatoires vers une variable aléatoire limite. Contrairement à la convergence des suites de nombres réels, il existe plusieurs façons naturelles de définir une convergence stochastique. Dans cette section, nous introduisons les concepts de convergence stochastique les plus célèbres et énonçons quelques résultats associés.

Considérons une suite (X_n) de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$ et une variable aléatoire X définie sur le même espace. Les concepts les plus classiques de convergence stochastique sont présentés dans les définitions suivantes.

Definition A.1 (X_n) *converge vers* X presque sûrement *(notation :* $X_n \overset{\text{p.s.}}{\to} X$*)* $\Leftrightarrow \mathrm{P}[\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}] = 1$.

Definition A.2 (X_n) *converge vers* X en probabilité *(notation :* $X_n \overset{\mathrm{P}}{\to} X$*)* $\Leftrightarrow$ *Pour tout* $\varepsilon > 0$, $\mathrm{P}[|X_n - X| > \varepsilon] \to 0$.

Definition A.3 (X_n) *converge vers* X en norme L_r*, où* $r > 0$ *(notation :* $X_n \overset{L_r}{\to} X$*)* $\Leftrightarrow \mathrm{E}[|X_n - X|^r] \to 0$. *Dans le cas particulier* $r = 2$*, on parlera souvent de* convergence en moyenne quadratique.

Definition A.4 (X_n) *converge vers* X en loi *(notation :* $X_n \overset{\mathcal{L}}{\to} X$*)* $\Leftrightarrow$ *En tout point de continuité* x *de* F^X*, on a* $F^{X_n}(x) \to F^X(x)$*, où* F^{X_n} *et* F^X *désignent respectivement les fonctions de répartition de* X_n *et* X.

Comme le montre le résultat suivant, certains sens de convergence stochastique sont plus forts que d'autres.

Théorème A.5 *Soient une suite* (X_n) *de variables aléatoires définies sur un espace probabilisé* $(\Omega, \mathcal{A}, \mathrm{P})$ *et une variable aléatoire* X *définie sur le même espace. Alors*

(i) $X_n \overset{\text{p.s.}}{\to} X \Rightarrow X_n \overset{\text{P}}{\to} X$;
(ii) $X_n \overset{L_r}{\to} X$ *pour un certain* $r > 0 \Rightarrow X_n \overset{\text{P}}{\to} X$;
(iii) $X_n \overset{\text{P}}{\to} X \Rightarrow X_n \overset{\mathcal{L}}{\to} X$.

Les autres implications ne tiennent pas de façon générale, mais tiennent si on ajoute des hypothèses convenables. À titre d'exemple, l'implication inverse en (iii) tient si la limite est constante : autrement dit, si $c \in \mathbb{R}$, alors $X_n \overset{\mathcal{L}}{\to} c \Rightarrow X_n \overset{\text{P}}{\to} c$.

Le résultat suivant énonce que certaines convergences stochastiques sont préservées sous des transformations continues.

Théorème A.6 *Soient une suite* (X_n) *de variables aléatoires (à valeurs dans* $C(\subset \mathbb{R})$ *avec probabilité 1) définies sur un espace probabilisé* $(\Omega, \mathcal{A}, \text{P})$ *et une variable aléatoire* X *(aussi à valeurs dans* $C(\subset \mathbb{R})$ *avec probabilité 1) définie sur le même espace. Soit* $g : C \to \mathbb{R}$ *une fonction continue. Alors*
(i) $X_n \overset{\text{p.s.}}{\to} X \Rightarrow g(X_n) \overset{\text{p.s.}}{\to} g(X)$;
(ii) $X_n \overset{\text{P}}{\to} X \Rightarrow g(X_n) \overset{\text{P}}{\to} g(X)$;
(iii) $X_n \overset{\mathcal{L}}{\to} X \Rightarrow g(X_n) \overset{\mathcal{L}}{\to} g(X)$.

Les résultats suivants détaillent les convergences de sommes et produits de suites de variables aléatoires et se révèlent souvent utiles. Pour la convergence en loi, de tels résultats ne tiennent de façon générale que si l'une des deux limites est dégénérée et portent le nom de *Lemme de Slutzky.*

Théorème A.7 *Soient deux suites* (X_n) *et* (Y_n) *de variables aléatoires définies sur un espace probabilisé* $(\Omega, \mathcal{A}, \text{P})$ *et deux variables aléatoires* X *et* Y *définies sur le même espace. Alors*
(i) $X_n \overset{\text{p.s.}}{\to} X$ *et* $Y_n \overset{\text{p.s.}}{\to} Y \Rightarrow X_n + Y_n \overset{\text{p.s.}}{\to} X + Y$;
(ii) $X_n \overset{\text{p.s.}}{\to} X$ *et* $Y_n \overset{\text{p.s.}}{\to} Y \Rightarrow X_n Y_n \overset{\text{p.s.}}{\to} XY$;
(iii) $X_n \overset{\text{P}}{\to} X$ *et* $Y_n \overset{\text{P}}{\to} Y \Rightarrow X_n + Y_n \overset{\text{P}}{\to} X + Y$;
(iv) $X_n \overset{\text{P}}{\to} X$ *et* $Y_n \overset{\text{P}}{\to} Y \Rightarrow X_n Y_n \overset{\text{P}}{\to} XY$;
(v) $X_n \overset{L_r}{\to} X$ *et* $Y_n \overset{L_r}{\to} Y$ *pour un certain* $r > 0 \Rightarrow X_n + Y_n \overset{L_r}{\to} X + Y$.

Théorème A.8 (Lemme de Slutzky) *Soient deux suites* (X_n) *et* (Y_n) *de variables aléatoires définies sur un espace probabilisé* $(\Omega, \mathcal{A}, \text{P})$ *et une variable aléatoire* X *définie sur le même espace. Soit* $c \in \mathbb{R}$. *Alors*
(i) $X_n \overset{\mathcal{L}}{\to} X$ *et* $Y_n \overset{\mathcal{L}}{\to} c \Rightarrow X_n + Y_n \overset{\mathcal{L}}{\to} X + c$;
(ii) $X_n \overset{\mathcal{L}}{\to} X$ *et* $Y_n \overset{\mathcal{L}}{\to} c \Rightarrow X_n Y_n \overset{\mathcal{L}}{\to} Xc$;
(iii) $X_n \overset{\mathcal{L}}{\to} X$ *et* $Y_n \overset{\mathcal{L}}{\to} c (\neq 0) \Rightarrow X_n / Y_n \overset{\mathcal{L}}{\to} X/c$.

Nous concluons avec les deux résultats limites classiques suivants.

Théorème A.9 (Lois forte et faible des grands nombres) *Soient* $X_1, X_2, \ldots$ *des variables aléatoires i.i.d. définies sur un espace probabilisé* $(\Omega, \mathcal{A}, \text{P})$. *Supposons que la loi commune des* X_i *admette des moments finis d'ordre 1. Posons* $\mu \overset{\text{def}}{=} \text{E}[X_1]$ *et* $\bar{X}_n \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} X_i$. *Alors*
(i) $\bar{X}_n \overset{\text{p.s.}}{\to} \mu$ *(loi forte) ;*
(ii) $\bar{X}_n \overset{\text{P}}{\to} \mu$ *(loi faible).*

Théorème A.10 (Théorème central limite) *Soient $X_1, X_2, \ldots$ des variables aléatoires i.i.d. définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$. Supposons que la loi commune des X_i admette des moments finis d'ordre 2. Posons $\mu \stackrel{\text{def}}{=} \mathrm{E}[X_1]$, $\sigma^2 \stackrel{\text{def}}{=} \mathrm{Var}[X_1]$ et $\bar{X}_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n X_i$. Alors*

$$Z_n \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) = \sqrt{n}(\bar{X}_n - \mu) \stackrel{\mathcal{L}}{\to} \mathcal{N}(0, \sigma^2),$$

au sens où $Z_n \stackrel{\mathcal{L}}{\to} Z$, avec $Z \sim \mathcal{N}(0, \sigma^2)$.

A.2 Moments de vecteurs aléatoires

Soit

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix}$$

un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Si les marginales Z_i possèdent des moments finis d'ordre 1, on peut considérer le *vecteur moyen* associé à Z, qui est le d-vecteur

$$\mu_Z = \mathrm{E}[Z] \stackrel{\text{def}}{=} \begin{pmatrix} \mathrm{E}[Z_1] \\ \vdots \\ \mathrm{E}[Z_d] \end{pmatrix}.$$

Dans tous les cas, on appliquera l'opérateur d'espérance composante par composante, que ce soit lorsqu'il est appliqué à des vecteurs (comme ci-dessus) ou à des matrices. Il en découle que pour toute matrice $m \times d$ et tout m-vecteur b, on a

$$\mathrm{E}[AZ + b] = A\mathrm{E}[Z] + b.$$

Si les marginales Z_i possèdent en outre des moments finis d'ordre 2, on peut définir la *matrice de variance-covariance* de Z, qui est la matrice $d \times d$

$$\begin{aligned} \Sigma_Z = \mathrm{Var}[Z] &\stackrel{\text{def}}{=} \mathrm{E}[(Z - \mu_Z)(Z - \mu_Z)'] \\ &= \begin{pmatrix} \mathrm{Var}[Z_1] & \mathrm{Cov}[Z_1, Z_2] & \ldots & \mathrm{Cov}[Z_1, Z_d] \\ \mathrm{Cov}[Z_1, Z_2] & \mathrm{Var}[Z_2] & \ldots & \mathrm{Cov}[Z_2, Z_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[Z_1, Z_d] & \mathrm{Cov}[Z_2, Z_d] & \ldots & \mathrm{Var}[Z_d] \end{pmatrix}. \end{aligned}$$

Cette matrice est bien entendu symétrique. Elle est aussi semi-définie positive au sens où $v'\Sigma_Z v \geq 0$ pour tout d-vecteur v. Ceci découle du fait que

$$v'\Sigma_Z v = \mathrm{E}[v'(Z - \mu_Z)(Z - \mu_Z)'v] = \mathrm{E}[(v'(Z - \mu_Z))^2] \geq 0.$$

La matrice de variance-covariance se réécrit également sous la forme

$$\begin{aligned}\Sigma_Z &= \mathrm{E}[(Z-\mu_Z)(Z-\mu_Z)'] \\ &= \mathrm{E}[ZZ' - Z\mu_Z' - \mu_Z Z' + \mu_Z\mu_Z'] \\ &= \mathrm{E}[ZZ'] - \mathrm{E}[Z]\mu_Z' - \mu_Z(\mathrm{E}[Z])' + \mu_Z\mu_Z' \\ &= \mathrm{E}[ZZ'] - \mu_Z\mu_Z'.\end{aligned}$$

En outre, on vérifiera facilement que

$$\mathrm{Var}[AZ+b] = A\mathrm{Var}[Z]A'$$

pour toute matrice $m \times d$ et tout m-vecteur b.

Enfin, si

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_r \end{pmatrix} \quad \text{et} \quad Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix}$$

sont des vecteurs aléatoires dont les marginales possèdent des moments finis d'ordre 2, on peut définir la covariance entre Y et Z comme la matrice $r \times d$

$$\begin{aligned}\mathrm{Cov}[Y,Z] &\stackrel{\text{def}}{=} \mathrm{E}[(Y-\mu_Y)(Z-\mu_Z)'] \\ &= \begin{pmatrix} \mathrm{Cov}[Y_1,Z_1] & \mathrm{Cov}[Y_1,Z_2] & \dots & \mathrm{Cov}[Y_1,Z_d] \\ \mathrm{Cov}[Y_2,Z_1] & \mathrm{Cov}[Y_2,Z_2] & \dots & \mathrm{Cov}[Y_2,Z_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_r,Z_1] & \mathrm{Cov}[Y_r,Z_2] & \dots & \mathrm{Cov}[Y_r,Z_d] \end{pmatrix}.\end{aligned}$$

Notons que $\mathrm{Cov}[Z,Y] = (\mathrm{Cov}[Y,Z])'$. En procédant comme ci-dessus, on peut établir que

$$\mathrm{Cov}[Y,Z] = \mathrm{E}[YZ'] - \mu_Y\mu_Z'.$$

Enfin, on vérifiera facilement que

$$\mathrm{Cov}[A_1Y+b_1, A_2Z+b_2] = A_1\mathrm{Cov}[Y,Z]A_2'$$

pour toute matrice $m \times r$ A_1, tout m-vecteur b_1, toute matrice $n \times d$ A_2 et tout n-vecteur b_2.

Annexe B

Tables de probabilités

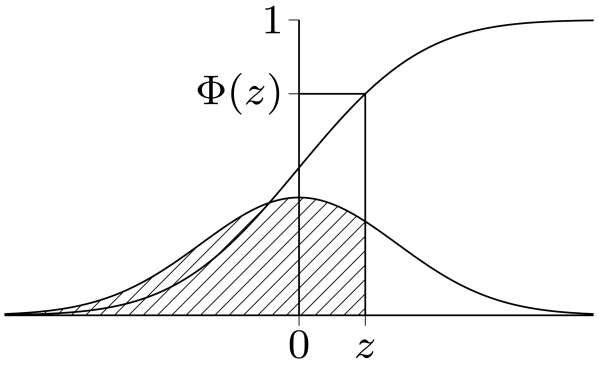
B.1 Tables associées à la loi normale

B.1.1 Fonction de répartition de la loi normale

La fonction de répartition Φ de la loi normale centrée réduite $\mathcal{N}(0,1)$ est définie par

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du,$$

pour tout $z \in \mathbb{R}$. Par symétrie de la loi, il tient que on a $\Phi(-z) = 1 - \Phi(z)$ pour tout $z \in \mathbb{R}$. Cette dernière remarque permet, en particulier, de ne détailler la fonction de répartition que pour des valeurs positives de z. La table B.1 donne les valeurs de $\Phi(z)$ pour z positif.



Par exemple, la table B.1 ci-dessous fournit $\Phi(0.25) \approx 0.5987$, $\Phi(-0.32) = 1 - \Phi(0.32) \approx 1 - 0.6255 = 0.3745$.

B.1.2 Quantiles de la loi normale

Pour $\alpha \in (0,1)$, le quantile d'ordre α de la loi normale est $\Phi^{-1}(\alpha)$. Pour tout $\alpha \in (0,1)$, on a $\Phi^{-1}(-\alpha) = -\Phi^{-1}(1-\alpha)$. La table B.2 donne les quantiles associés à différentes valeurs de α.

Table B.1 – *Valeurs de la fonction de répartition de la loi normale centrée réduite $\Phi(z)$. La table se lit de la manière suivante : pour $z \in \{0, 0.01, 0.02, \ldots, 2.99\}$, identifier la ligne correspondant à l'unité et la première décimale. La valeur de la fonction de répartition se lit ensuite dans la colonne correspondant au centièmes de z.*

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Par exemple, on y trouve que $\Phi^{-1}(0.75) \approx 0.6745$, $\Phi^{-1}(0.995) \approx 2.5758$, $\Phi^{-1}(0.9995) \approx 3.2905$; ainsi que $\Phi^{-1}(0.25) \approx -0.6745$, $\Phi^{-1}(0.005) \approx -2.5758$, $\Phi^{-1}(0.0005) \approx -3.2905$.

De plus, si $Z \sim \mathcal{N}(0, 1)$, la table B.3 donne, pour α fixé, la valeur $z_{\alpha/2}$ telle que

$$P(|Z| > z_{\alpha/2}) = 1 - \alpha.$$

Ainsi, z_α est le quantile d'ordre $1-\alpha$ de Z. Par exemple, pour $\alpha = 0.5$; on trouve $z_{\alpha/2} \approx 0.6745$; pour $\alpha = 10^{-6}$, on trouve $z_{\alpha/2} \approx 4.8916$.

Remarquons que ces deux dernières tables peuvent être obtenues directement de la table B.1. Elles sont cependant présentées par souci de simplicité.

Table B.2 – *Valeurs des quantiles $\Phi^{-1}(\alpha)$ pour différentes valeurs de α.*

α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.0000	0.0251	0.0502	0.0753	0.1004	0.1257	0.1510	0.1764	0.2019	0.2275
0.6	0.2533	0.2793	0.3055	0.3319	0.3585	0.3853	0.4125	0.4399	0.4677	0.4959
0.7	0.5244	0.5534	0.5828	0.6128	0.6433	0.6745	0.7063	0.7388	0.7722	0.8064
0.8	0.8416	0.8779	0.9154	0.9542	0.9945	1.0364	1.0803	1.1264	1.1750	1.2265
0.9	1.2816	1.3408	1.4051	1.4758	1.5548	1.6449	1.7507	1.8808	2.0537	2.3263

α	0.99	0.991	0.992	0.993	0.994	0.995	0.996	0.997	0.998	0.999
$\Phi^{-1}(\alpha)$	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902
α	0.999	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999
$\Phi^{-1}(\alpha)$	3.0902	3.1214	3.1559	3.1947	3.2389	3.2905	3.3528	3.4316	3.5401	3.7190

Table B.3 – *Valeurs $z_{\alpha/2}$ pour différentes valeurs de α.*

α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	$+\infty$	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

α	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$z_{1-\alpha/2}$	3.2905	3.8906	4.4172	4.8916	5.3267	5.7307	6.1094

B.2 Tables associées à la loi de Student

Si T est une variable aléatoire suivant une loi de Student à k degrés de liberté, la table B.4 donne, pour différentes valeurs de α, les quantiles $t_{k,\alpha}$ tels que $P[T < t_{k,\alpha}] = \alpha$. Par exemple, on y lit que $t_{4,0.8} = 0.941$, $t_{10,0.99} = 2.764$, $t_{40,0.6} = 0.255$ et $\Phi^{-1}(0.9) = 1.282$. Cette dernière valeur est obtenue en regardant la dernière ligne de la table.

Comme pour la table B.3, la table B.4 permet de trouver les valeurs $t_{k,1-\alpha/2}$ telle que

$$P(|T| > t_{k,1-\alpha/2}) = 1 - \alpha.$$

B.3 Tables associées à la loi de chi-carré

Si X est une variable aléatoire suivant une loi chi-carré à k degrés de liberté, les tables B.5 et B.6 donnent, pour différentes valeurs de α (resp. inférieures et supérieures à 50%), les quantiles $\chi^2_{k,\alpha}$ tels que $P[X < \chi^2_{k,\alpha}] = \alpha$. Par exemple, on y lit que $\chi^2_{3,0.1} = 0.584$, $\chi^2_{10,0.33} = 7.612$ et $\chi^2_{40,0.6} = 41.622$. Dans un souci d'économie de place, ces tables sont réduites en taille.

Table B.4 – *Table des quantiles des lois de Student. Pour chaque k et chaque α, la table renvoie $t_{k,\alpha}$, le quantile d'ordre α de la loi t_k. Lorsque $k = \infty$, les quantiles sont ceux de la loi $\mathcal{N}(0,1)$.*

k	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

B.4 Tables associées à la loi de Fisher

Si F est une variable aléatoire suivant la loi de Fisher–Snedecor (parfois appelée loi F) à (k_1, k_2) degrés de liberté, les tables B.7 à B.14 donnent les valeurs $F_{k_1,k_2,\alpha}$ des quantiles d'ordre α tels que $P[F \leq F_{k_1,k_2,\alpha}] = \alpha$. Dans un souci d'économie de place, ces tables sont réduites en taille.

Table B.5 – *Table des quantiles inférieurs des lois chi-carré. Pour k et $\alpha \leq 0.5$, la table renvoie $\chi^2_{k,\alpha}$, le quantile d'ordre α de la loi χ^2_k.*

k	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 B.6 – *Table des quantiles supérieurs des lois chi-carré. Pour chaque k et $\alpha > 0.5$, la table renvoie $\chi^2_{k,\alpha}$, le quantile d'ordre α de la loi χ^2_k.*

k	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 B.7 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
1	0.500	1.50	1.71	1.82	1.89	1.94	1.98	2.00	2.04	2.07	2.09	2.12	2.15	2.17	2.20
	0.600	2.63	2.93	3.09	3.20	3.27	3.32	3.36	3.41	3.45	3.48	3.52	3.56	3.59	3.64
	0.667	4.00	4.42	4.64	4.78	4.88	4.95	5.00	5.08	5.13	5.18	5.24	5.29	5.33	5.39
	0.750	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.32	9.41	9.50	9.58	9.67	9.74	9.85
	0.800	12.0	13.1	13.6	14.0	14.3	14.4	14.6	14.8	14.9	15.0	15.2	15.3	15.4	15.6
2	0.500	1.00	1.13	1.21	1.25	1.28	1.30	1.32	1.35	1.36	1.38	1.39	1.41	1.42	1.44
	0.600	1.50	1.64	1.72	1.76	1.80	1.82	1.84	1.86	1.88	1.89	1.91	1.92	1.94	1.96
	0.667	2.00	2.15	2.22	2.27	2.30	2.33	2.34	2.37	2.38	2.40	2.42	2.43	2.45	2.47
	0.750	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.38	3.39	3.41	3.43	3.44	3.46	3.48
	0.800	4.00	4.16	4.24	4.28	4.32	4.34	4.36	4.38	4.40	4.42	4.43	4.45	4.47	4.48
3	0.500	0.88	1.00	1.06	1.10	1.13	1.15	1.16	1.18	1.20	1.21	1.23	1.24	1.25	1.27
	0.600	1.26	1.37	1.43	1.47	1.49	1.51	1.52	1.54	1.55	1.56	1.57	1.58	1.59	1.60
	0.667	1.62	1.72	1.77	1.80	1.82	1.83	1.84	1.86	1.87	1.88	1.89	1.90	1.90	1.91
	0.750	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.45	2.46	2.46	2.47	2.47	2.47
	0.800	2.89	2.94	2.96	2.97	2.97	2.97	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98
4	0.500	0.83	0.94	1.00	1.04	1.06	1.08	1.09	1.11	1.13	1.14	1.15	1.16	1.18	1.19
	0.600	1.16	1.26	1.31	1.34	1.36	1.37	1.38	1.40	1.41	1.42	1.43	1.43	1.44	1.45
	0.667	1.46	1.55	1.58	1.61	1.62	1.63	1.64	1.65	1.65	1.66	1.67	1.67	1.68	1.68
	0.750	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
	0.800	2.47	2.48	2.48	2.48	2.47	2.47	2.47	2.46	2.46	2.45	2.44	2.44	2.43	2.43
5	0.500	0.80	0.91	0.96	1.00	1.02	1.04	1.05	1.07	1.09	1.10	1.11	1.12	1.13	1.15
	0.600	1.11	1.20	1.24	1.27	1.29	1.30	1.31	1.32	1.33	1.34	1.34	1.35	1.36	1.37
	0.667	1.38	1.45	1.48	1.50	1.51	1.52	1.53	1.53	1.54	1.54	1.54	1.55	1.55	1.55
	0.750	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.87
	0.800	2.26	2.25	2.24	2.23	2.22	2.21	2.20	2.19	2.18	2.18	2.17	2.16	2.15	2.13
6	0.500	0.78	0.89	0.94	0.98	1.00	1.02	1.03	1.05	1.06	1.07	1.08	1.10	1.11	1.12
	0.600	1.07	1.16	1.20	1.22	1.24	1.25	1.26	1.27	1.28	1.29	1.29	1.30	1.31	1.31
	0.667	1.33	1.39	1.42	1.44	1.44	1.45	1.45	1.46	1.46	1.47	1.47	1.47	1.47	1.47
	0.750	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.76	1.76	1.75	1.75	1.74
	0.800	2.13	2.11	2.09	2.08	2.06	2.05	2.04	2.03	2.02	2.01	2.00	1.98	1.97	1.95
7	0.500	0.77	0.87	0.93	0.96	0.98	1.00	1.01	1.03	1.04	1.05	1.07	1.08	1.09	1.10
	0.600	1.05	1.13	1.17	1.19	1.21	1.22	1.23	1.24	1.24	1.25	1.26	1.26	1.27	1.27
	0.667	1.29	1.35	1.38	1.39	1.40	1.40	1.41	1.41	1.41	1.41	1.41	1.42	1.42	1.42
	0.750	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.68	1.68	1.67	1.66	1.66	1.65
	0.800	2.04	2.02	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.89	1.88	1.86	1.85	1.83
8	0.500	0.76	0.86	0.91	0.95	0.97	0.99	1.00	1.02	1.03	1.04	1.05	1.07	1.07	1.09
	0.600	1.03	1.11	1.15	1.17	1.19	1.20	1.20	1.21	1.22	1.22	1.23	1.24	1.24	1.25
	0.667	1.26	1.32	1.35	1.36	1.36	1.37	1.37	1.37	1.37	1.38	1.38	1.38	1.37	1.37
	0.750	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.62	1.62	1.61	1.60	1.59	1.58
	0.800	1.98	1.95	1.92	1.90	1.88	1.87	1.86	1.84	1.83	1.81	1.80	1.78	1.76	1.74

Table B.8 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
9	0.500	0.75	0.85	0.91	0.94	0.96	0.98	0.99	1.01	1.02	1.03	1.04	1.05	1.06	1.08
	0.600	1.02	1.10	1.13	1.15	1.17	1.18	1.18	1.19	1.20	1.21	1.21	1.22	1.22	1.22
	0.667	1.24	1.30	1.32	1.33	1.34	1.34	1.34	1.34	1.35	1.35	1.35	1.34	1.34	1.34
	0.750	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.58	1.57	1.56	1.55	1.54	1.53
	0.800	1.93	1.90	1.87	1.85	1.83	1.81	1.80	1.78	1.76	1.75	1.73	1.71	1.70	1.67
10	0.500	0.74	0.85	0.90	0.93	0.95	0.97	0.98	1.00	1.01	1.02	1.03	1.05	1.06	1.07
	0.600	1.01	1.08	1.12	1.14	1.15	1.16	1.17	1.18	1.18	1.19	1.19	1.20	1.20	1.21
	0.667	1.23	1.28	1.30	1.31	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.31
	0.750	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.48
	0.800	1.90	1.86	1.83	1.80	1.78	1.77	1.75	1.73	1.72	1.70	1.68	1.66	1.65	1.62
11	0.500	0.74	0.84	0.89	0.93	0.95	0.96	0.98	0.99	1.01	1.02	1.03	1.04	1.05	1.06
	0.600	1.00	1.07	1.11	1.13	1.14	1.15	1.16	1.17	1.17	1.18	1.18	1.18	1.19	1.19
	0.667	1.22	1.27	1.29	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.29
	0.750	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.45
	0.800	1.87	1.83	1.80	1.77	1.75	1.73	1.72	1.69	1.68	1.66	1.64	1.62	1.60	1.57
12	0.500	0.73	0.84	0.89	0.92	0.94	0.96	0.97	0.99	1.00	1.01	1.02	1.03	1.04	1.06
	0.600	0.99	1.07	1.10	1.12	1.13	1.14	1.15	1.16	1.16	1.17	1.17	1.17	1.18	1.18
	0.667	1.21	1.26	1.27	1.28	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.28	1.28	1.27
	0.750	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.45	1.44	1.42
	0.800	1.85	1.80	1.77	1.74	1.72	1.70	1.69	1.66	1.65	1.63	1.61	1.59	1.57	1.54
13	0.500	0.73	0.83	0.88	0.92	0.94	0.96	0.97	0.98	1.00	1.01	1.02	1.03	1.04	1.05
	0.600	0.98	1.06	1.09	1.11	1.13	1.13	1.14	1.15	1.15	1.16	1.16	1.16	1.17	1.17
	0.667	1.20	1.25	1.26	1.27	1.28	1.28	1.28	1.28	1.28	1.28	1.27	1.27	1.27	1.26
	0.750	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.43	1.42	1.40
	0.800	1.83	1.78	1.75	1.72	1.69	1.68	1.66	1.64	1.62	1.60	1.58	1.56	1.54	1.51
14	0.500	0.73	0.83	0.88	0.91	0.94	0.95	0.96	0.98	0.99	1.00	1.01	1.03	1.04	1.05
	0.600	0.98	1.05	1.09	1.11	1.12	1.13	1.13	1.14	1.14	1.15	1.15	1.16	1.16	1.16
	0.667	1.19	1.24	1.26	1.26	1.27	1.27	1.27	1.27	1.27	1.26	1.26	1.26	1.25	1.24
	0.750	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.41	1.40	1.38
	0.800	1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.62	1.60	1.58	1.56	1.53	1.51	1.48
15	0.500	0.73	0.83	0.88	0.91	0.93	0.95	0.96	0.98	0.99	1.00	1.01	1.02	1.03	1.05
	0.600	0.97	1.05	1.08	1.10	1.11	1.12	1.13	1.13	1.14	1.14	1.15	1.15	1.15	1.15
	0.667	1.18	1.23	1.25	1.25	1.26	1.26	1.26	1.26	1.26	1.25	1.25	1.25	1.24	1.23
	0.750	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.38	1.36
	0.800	1.80	1.75	1.71	1.68	1.66	1.64	1.62	1.60	1.58	1.56	1.54	1.51	1.49	1.46
16	0.500	0.72	0.82	0.88	0.91	0.93	0.95	0.96	0.97	0.99	1.00	1.01	1.02	1.03	1.04
	0.600	0.97	1.04	1.08	1.10	1.11	1.12	1.12	1.13	1.13	1.14	1.14	1.14	1.14	1.14
	0.667	1.18	1.22	1.24	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.24	1.24	1.23	1.22
	0.750	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.38	1.37	1.34
	0.800	1.78	1.74	1.70	1.67	1.64	1.62	1.61	1.58	1.56	1.54	1.52	1.49	1.47	1.43
17	0.500	0.72	0.82	0.87	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.01	1.02	1.03	1.04
	0.600	0.97	1.04	1.07	1.09	1.10	1.11	1.12	1.12	1.13	1.13	1.13	1.14	1.14	1.14
	0.667	1.17	1.22	1.23	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.23	1.23	1.22	1.21
	0.750	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.39	1.37	1.36	1.33
	0.800	1.77	1.72	1.68	1.65	1.63	1.61	1.59	1.57	1.55	1.53	1.50	1.48	1.46	1.42

Table B.9 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
18	0.500	0.72	0.82	0.87	0.90	0.93	0.94	0.95	0.97	0.98	0.99	1.00	1.02	1.02	1.04
	0.600	0.96	1.04	1.07	1.09	1.10	1.11	1.11	1.12	1.12	1.13	1.13	1.13	1.13	1.13
	0.667	1.17	1.21	1.23	1.24	1.24	1.24	1.24	1.24	1.23	1.23	1.23	1.22	1.22	1.21
	0.750	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.40	1.39	1.38	1.36	1.34	1.32
	0.800	1.76	1.71	1.67	1.64	1.62	1.60	1.58	1.55	1.53	1.51	1.49	1.46	1.44	1.40
19	0.500	0.72	0.82	0.87	0.90	0.92	0.94	0.95	0.97	0.98	0.99	1.00	1.01	1.02	1.04
	0.600	0.96	1.03	1.07	1.09	1.10	1.10	1.11	1.12	1.12	1.12	1.13	1.13	1.13	1.13
	0.667	1.16	1.21	1.22	1.23	1.23	1.23	1.23	1.23	1.23	1.23	1.22	1.22	1.21	1.20
	0.750	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.40	1.38	1.37	1.35	1.33	1.30
	0.800	1.75	1.70	1.66	1.63	1.61	1.58	1.57	1.54	1.52	1.50	1.48	1.45	1.43	1.39
20	0.500	0.72	0.82	0.87	0.90	0.92	0.94	0.95	0.97	0.98	0.99	1.00	1.01	1.02	1.03
	0.600	0.96	1.03	1.06	1.08	1.09	1.10	1.11	1.11	1.12	1.12	1.12	1.12	1.12	1.12
	0.667	1.16	1.21	1.22	1.23	1.23	1.23	1.23	1.23	1.22	1.22	1.22	1.21	1.20	1.19
	0.750	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.40	1.39	1.37	1.36	1.34	1.32	1.29
	0.800	1.75	1.70	1.65	1.62	1.60	1.58	1.56	1.53	1.51	1.49	1.47	1.44	1.41	1.37
21	0.500	0.72	0.81	0.87	0.90	0.92	0.94	0.95	0.96	0.98	0.99	1.00	1.01	1.02	1.03
	0.600	0.96	1.03	1.06	1.08	1.09	1.10	1.10	1.11	1.11	1.12	1.12	1.12	1.12	1.12
	0.667	1.16	1.20	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.21	1.20	1.20	1.19
	0.750	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.39	1.38	1.37	1.35	1.33	1.32	1.28
	0.800	1.74	1.69	1.65	1.61	1.59	1.57	1.55	1.52	1.50	1.48	1.46	1.43	1.40	1.36
22	0.500	0.72	0.81	0.87	0.90	0.92	0.93	0.95	0.96	0.97	0.99	1.00	1.01	1.02	1.03
	0.600	0.96	1.03	1.06	1.08	1.09	1.10	1.10	1.11	1.11	1.11	1.12	1.12	1.12	1.12
	0.667	1.16	1.20	1.21	1.22	1.22	1.22	1.22	1.22	1.21	1.21	1.21	1.20	1.19	1.18
	0.750	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.37	1.36	1.34	1.32	1.31	1.28
	0.800	1.73	1.68	1.64	1.61	1.58	1.56	1.54	1.51	1.49	1.47	1.45	1.42	1.39	1.35
23	0.500	0.71	0.81	0.86	0.90	0.92	0.93	0.95	0.96	0.97	0.98	1.00	1.01	1.02	1.03
	0.600	0.95	1.02	1.06	1.07	1.09	1.09	1.10	1.10	1.11	1.11	1.11	1.11	1.11	1.11
	0.667	1.15	1.20	1.21	1.22	1.22	1.22	1.22	1.21	1.21	1.21	1.20	1.19	1.19	1.17
	0.750	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.38	1.37	1.35	1.34	1.32	1.30	1.27
	0.800	1.73	1.68	1.63	1.60	1.57	1.55	1.53	1.51	1.49	1.46	1.44	1.41	1.38	1.34
24	0.500	0.71	0.81	0.86	0.90	0.92	0.93	0.94	0.96	0.97	0.98	0.99	1.01	1.01	1.03
	0.600	0.95	1.02	1.06	1.07	1.08	1.09	1.10	1.10	1.10	1.11	1.11	1.11	1.11	1.11
	0.667	1.15	1.19	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.20	1.20	1.19	1.18	1.17
	0.750	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.36	1.35	1.33	1.31	1.29	1.26
	0.800	1.72	1.67	1.63	1.59	1.57	1.55	1.53	1.50	1.48	1.46	1.43	1.40	1.38	1.33
25	0.500	0.71	0.81	0.86	0.89	0.92	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.03
	0.600	0.95	1.02	1.05	1.07	1.08	1.09	1.09	1.10	1.10	1.11	1.11	1.11	1.11	1.11
	0.667	1.15	1.19	1.21	1.21	1.21	1.21	1.21	1.21	1.20	1.20	1.19	1.19	1.18	1.16
	0.750	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.37	1.36	1.34	1.33	1.31	1.29	1.25
	0.800	1.72	1.66	1.62	1.59	1.56	1.54	1.52	1.49	1.47	1.45	1.42	1.39	1.37	1.32
26	0.500	0.71	0.81	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.03
	0.600	0.95	1.02	1.05	1.07	1.08	1.09	1.09	1.10	1.10	1.10	1.10	1.11	1.11	1.10
	0.667	1.15	1.19	1.20	1.21	1.21	1.21	1.21	1.20	1.20	1.20	1.19	1.18	1.18	1.16
	0.750	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.35	1.34	1.32	1.30	1.28	1.25
	0.800	1.71	1.66	1.62	1.58	1.56	1.53	1.52	1.49	1.47	1.44	1.42	1.39	1.36	1.31

Table B.10 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
27	0.500	0.71	0.81	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.ga	0.99	1.00	1.01	1.03
	0.600	0.95	1.02	1.05	1.07	1.08	1.08	1.09	1.10	1.10	1.10	1.10	1.10	1.10	1.10
	0.667	1.14	1.19	1.20	1.21	1.21	1.21	1.20	1.20	1.20	1.19	1.19	1.18	1.17	1.16
	0.750	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.36	1.35	1.33	1.32	1.30	1.28	1.24
	0.800	1.71	1.66	1.61	1.58	1.55	1.53	1.51	1.48	1.46	1.44	1.41	1.3a	1.35	1.30
28	0.500	0.71	0.81	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.02
	0.600	0.95	1.02	1.05	1.07	1.08	1.08	1.09	1.09	1.10	1.10	1.10	1.10	1.10	1.10
	0.667	1.14	1.18	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.19	1.19	1.18	1.17	1.15
	0.750	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.36	1.34	1.33	1.31	1.29	1.27	1.24
	0.800	1.71	1.65	1.61	1.57	1.55	1.52	1.51	1.48	1.46	1.43	1.41	1.37	1.35	1.30
29	0.500	0.71	0.81	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.02
	0.600	0.95	1.02	1.05	1.06	1.08	1.08	1.09	1.09	1.10	1.10	1.10	1.10	1.10	1.10
	0.667	1.14	1.18	1.20	1.20	1.20	1.20	1.20	1.20	1.19	1.19	1.18	1.17	1.17	1.15
	0.750	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.35	1.34	1.32	1.31	1.29	1.27	1.23
	0.800	1.70	1.65	1.60	1.57	1.54	1.52	1.50	1.47	1.45	1.43	1.40	1.37	1.34	1.29
30	0.500	0.71	0.al	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.02
	0.600	0.94	1.01	1.05	1.06	1.07	1.08	1.08	1.09	1.09	1.10	1.10	1.10	1.10	1.09
	0.667	1.14	1.18	1.19	1.20	1.20	1.20	1.20	1.19	1.19	1.19	1.18	1.17	1.16	1.15
	0.750	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.35	1.34	1.32	1.30	1.28	1.26	1.23
	0.800	1.70	1.64	1.60	1.57	1.54	1.52	1.50	1.47	1.45	1.42	1.39	1.36	1.34	1.28
60	0.500	0.70	0.80	0.85	0.88	0.90	0.92	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01
	0.600	0.93	1.00	1.03	1.04	1.05	1.06	1.06	1.07	1.07	1.07	1.07	1.07	1.07	1.06
	0.667	1.12	1.16	1.17	1.17	1.17	1.17	1.17	1.16	1.16	1.15	1.14	1.13	1.12	1.10
	0.750	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.30	1.29	1.27	1.25	1.22	1.20	1.15
	0.800	1.65	1.59	1.55	1.51	1.48	1.46	1.44	1.41	1.38	1.35	1.32	1.29	1.25	1.18
80	0.500	0.70	0.80	0.85	0.88	0.90	0.91	0.93	0.94	0.95	0.96	0.97	0.99	1.00	1.01
	0.600	0.93	0.99	1.02	1.04	1.05	1.06	1.06	1.06	1.07	1.07	1.07	1.06	1.06	1.05
	0.667	1.11	1.15	1.16	1.17	1.17	1.16	1.16	1.16	1.15	1.14	1.13	1.12	1.11	1.08
	0.750	1.41	1.40	1.38	1.36	1.34	1.32	1.31	1.29	1.27	1.26	1.23	1.21	1.18	1.12
	0.800	1.64	1.58	1.53	1.50	1.47	1.44	1.42	1.39	1.37	1.34	1.31	1.27	1.23	1.16
100	0.500	0.70	0.79	0.84	0.88	0.90	0.91	0.92	0.94	0.95	0.96	0.97	0.98	0.99	1.01
	0.600	0.92	0.99	1.02	1.04	1.05	1.05	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.04
	0.667	1.11	1.15	1.16	1.16	1.16	1.16	1.16	1.15	1.15	1.14	1.13	1.12	1.10	1.07
	0.750	1.41	1.39	1.37	1.35	1.33	1.32	1.30	1.28	1.27	1.25	1.23	1.20	1.17	1.11
	0.800	1.64	1.58	1.53	1.49	1.46	1.43	1.41	1.38	1.36	1.33	1.30	1.26	1.22	1.14
∞	0.500	0.69	0.79	0.84	0.87	0.89	0.91	0.92	0.93	0.95	0.96	0.97	0.98	0.99	1.00
	0.600	0.92	0.98	1.01	1.03	1.04	1.04	1.04	1.05	1.05	1.05	1.05	1.04	1.04	1.00
	0.667	1.10	1.13	1.14	1.15	1.14	1.14	1.14	1.13	1.13	1.12	1.11	1.09	1.07	1.00
	0.750	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.25	1.24	1.22	1.19	1.16	1.13	1.00
	0.800	1.61	1.55	1.50	1.46	1.43	1.40	1.38	1.34	1.32	1.29	1.25	1.21	1.16	1.00

Table B.11 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
1	0.900	49.5	53.6	55.8	57.2	58.2	59.1	59.7	60.5	61.0	61.5	62.0	62.6	63.0	63.3
	0.950	200	216	225	230	234	237	239	242	244	246	248	250	252	254
	0.975	800	864	900	922	937	948	957	969	977	985	993	1.0e3	1.0e3	1.0e3
	0.990	5.4e3	5.6e3	5.8e3	5.9e3	5.9e3	6.0e3	6.0e3	6.1e3	6.1e3	6.2e3	6.2e3	6.3e3	6.3e3	6.4e3
	0.999	5.4e5	5.6e5	5.8e5	5.9e5	5.9e5	6.0e5	6.0e5	6.1e5	6.1e5	6.2e5	6.2e5	6.3e5	6.3e5	6.4e5
2	0.900	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.39	9.41	9.43	9.44	9.46	9.47	9.49
	0.950	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
	0.975	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5
	0.990	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5
	0.999	999	999.2	999.2	999.3	999.4	999.4	999.4	999.4	999.4	999.4	999.4	999.5	999.5	999.5
3	0.900	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.23	5.22	5.20	5.18	5.17	5.15	5.13
	0.950	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.70	8.66	8.62	8.58	8.53
	0.975	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.4	14.3	14.3	14.2	14.1	14.0	13.9
	0.990	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.2	27.1	26.9	26.7	26.5	26.4	26.1
	0.999	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.2	129.3	127.4	126.4	125.4	124.6	123.5
4	0.900	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.92	3.90	3.87	3.84	3.82	3.79	3.76
	0.950	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.86	5.80	5.75	5.70	5.63
	0.975	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.84	8.75	8.66	8.56	8.46	8.38	8.26
	0.990	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.5	14.4	14.2	14.0	13.8	13.7	13.5
	0.999	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48.0	47.4	46.8	46.1	45.4	44.9	44.1
5	0.900	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.30	3.27	3.24	3.21	3.17	3.15	3.10
	0.950	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.62	4.56	4.50	4.44	4.36
	0.975	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.62	6.52	6.43	6.33	6.23	6.14	6.02
	0.990	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.1	9.89	9.72	9.55	9.38	9.24	9.02
	0.999	37.1	33.2	31.1	29.8	28.8	28.2	27.6	26.9	26.4	25.9	25.4	24.9	24.4	23.8
6	0.900	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.94	2.90	2.87	2.84	2.80	2.77	2.72
	0.950	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.94	3.87	3.81	3.75	3.67
	0.975	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.46	5.37	5.27	5.17	5.07	4.98	4.85
	0.990	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.87	7.72	7.56	7.40	7.23	7.09	6.88
	0.999	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.4	18.0	17.6	17.1	16.7	16.3	15.7
7	0.900	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.70	2.67	2.63	2.59	2.56	2.52	2.47
	0.950	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.51	3.44	3.38	3.32	3.23
	0.975	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.76	4.67	4.57	4.47	4.36	4.28	4.14
	0.990	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.62	6.47	6.31	6.16	5.99	5.86	5.65
	0.999	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.1	13.7	13.3	12.9	12.5	12.2	11.7
8	0.900	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.54	2.50	2.46	2.42	2.38	2.35	2.29
	0.950	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.22	3.15	3.08	3.02	2.93
	0.975	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.29	4.20	4.10	4.00	3.89	3.81	3.67
	0.990	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.81	5.67	5.52	5.36	5.20	5.07	4.86
	0.999	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.5	11.2	10.8	10.5	10.1	9.80	9.33

Table B.12 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
9	0.900	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.42	2.38	2.34	2.30	2.25	2.22	2.16
	0.950	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	3.01	2.94	2.86	2.80	2.71
	0.975	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.96	3.87	3.77	3.67	3.56	3.47	3.33
	0.990	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.26	5.11	4.96	4.81	4.65	4.52	4.31
	0.999	16.4	13.9	12.6	11.7	11.1	10.7	10.4	9.89	9.57	9.24	8.90	8.55	8.26	7.81
10	0.900	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.32	2.28	2.24	2.20	2.16	2.12	2.06
	0.950	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.84	2.77	2.70	2.64	2.54
	0.975	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.72	3.62	3.52	3.42	3.31	3.22	3.08
	0.990	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.85	4.71	4.56	4.41	4.25	4.11	3.91
	0.999	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.75	8.45	8.13	7.80	7.47	7.19	6.76
11	0.900	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.25	2.21	2.17	2.12	2.08	2.04	1.97
	0.950	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.72	2.65	2.57	2.51	2.40
	0.975	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.53	3.43	3.33	3.23	3.12	3.03	2.88
	0.990	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.54	4.40	4.25	4.10	3.94	3.81	3.60
	0.999	13.8	11.6	10.3	9.58	9.05	8.66	8.35	7.92	7.63	7.32	7.01	6.68	6.42	6.00
12	0.900	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.19	2.15	2.10	2.06	2.01	1.97	1.90
	0.950	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.62	2.54	2.47	2.40	2.30
	0.975	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.37	3.28	3.18	3.07	2.96	2.87	2.72
	0.990	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.30	4.16	4.01	3.86	3.70	3.57	3.36
	0.999	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.29	7.00	6.71	6.40	6.09	5.83	5.42
13	0.900	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.14	2.10	2.05	2.01	1.96	1.92	1.85
	0.950	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.53	2.46	2.38	2.31	2.21
	0.975	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.25	3.15	3.05	2.95	2.84	2.74	2.60
	0.990	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.10	3.96	3.82	3.66	3.51	3.37	3.17
	0.999	12.3	10.2	9.07	8.35	7.86	7.49	7.21	6.80	6.52	6.23	5.93	5.63	5.37	4.97
14	0.900	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.10	2.05	2.01	1.96	1.91	1.87	1.80
	0.950	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.46	2.39	2.31	2.24	2.13
	0.975	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.15	3.05	2.95	2.84	2.73	2.64	2.49
	0.990	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.94	3.80	3.66	3.51	3.35	3.22	3.00
	0.999	11.8	9.73	8.62	7.92	7.44	7.08	6.80	6.40	6.13	5.85	5.56	5.25	5.00	4.60
15	0.900	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.06	2.02	1.97	1.92	1.87	1.83	1.76
	0.950	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.40	2.33	2.25	2.18	2.07
	0.975	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.06	2.96	2.86	2.76	2.64	2.55	2.40
	0.990	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.80	3.67	3.52	3.37	3.21	3.08	2.87
	0.999	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.08	5.81	5.53	5.25	4.95	4.70	4.31
16	0.900	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.03	1.99	1.94	1.89	1.84	1.79	1.72
	0.950	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.35	2.28	2.19	2.12	2.01
	0.975	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.99	2.89	2.79	2.68	2.57	2.47	2.32
	0.990	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.69	3.55	3.41	3.26	3.10	2.97	2.75
	0.999	11.0	9.01	7.94	7.27	6.80	6.46	6.19	5.81	5.55	5.27	4.99	4.70	4.45	4.06
17	0.900	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.00	1.96	1.91	1.86	1.81	1.76	1.69
	0.950	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.31	2.23	2.15	2.08	1.96
	0.975	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.92	2.82	2.72	2.62	2.50	2.41	2.25
	0.990	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.59	3.46	3.31	3.16	3.00	2.87	2.65
	0.999	10.7	8.73	7.68	7.02	6.56	6.22	5.96	5.58	5.32	5.05	4.77	4.48	4.24	3.85

Table B.13 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
18	0.900	2.62	2.42	2.29	2.20	2.13	2.08	2.04	1.98	1.93	1.89	1.84	1.78	1.74	1.66
	0.950	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.34	2.27	2.19	2.11	2.04	1.92
	0.975	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.87	2.77	2.67	2.56	2.44	2.35	2.19
	0.990	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.51	3.37	3.23	3.08	2.92	2.78	2.57
	0.999	10.4	8.49	7.46	6.81	6.35	6.02	5.76	5.39	5.13	4.87	4.59	4.30	4.06	3.67
19	0.900	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.96	1.91	1.86	1.81	1.76	1.71	1.63
	0.950	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.23	2.16	2.07	2.00	1.88
	0.975	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.82	2.72	2.62	2.51	2.39	2.30	2.13
	0.990	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.43	3.30	3.15	3.00	2.84	2.71	2.49
	0.999	10.2	8.28	7.27	6.62	6.18	5.85	5.59	5.22	4.97	4.70	4.43	4.14	3.90	3.51
20	0.900	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.94	1.89	1.84	1.79	1.74	1.69	1.61
	0.950	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.20	2.12	2.04	1.97	1.84
	0.975	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.77	2.68	2.57	2.46	2.35	2.25	2.09
	0.990	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.37	3.23	3.09	2.94	2.78	2.64	2.42
	0.999	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.08	4.82	4.56	4.29	4.00	3.76	3.38
21	0.900	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.92	1.87	1.83	1.78	1.72	1.67	1.59
	0.950	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.32	2.25	2.18	2.10	2.01	1.94	1.81
	0.975	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.73	2.64	2.53	2.42	2.31	2.21	2.04
	0.990	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.31	3.17	3.03	2.88	2.72	2.58	2.36
	0.999	9.77	7.94	6.95	6.32	5.88	5.56	5.31	4.95	4.70	4.44	4.17	3.88	3.64	3.26
22	0.900	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.90	1.86	1.81	1.76	1.70	1.65	1.57
	0.950	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.30	2.23	2.15	2.07	1.98	1.91	1.78
	0.975	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.70	2.60	2.50	2.39	2.27	2.17	2.00
	0.990	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.26	3.12	2.98	2.83	2.67	2.53	2.31
	0.999	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.83	4.58	4.33	4.06	3.78	3.54	3.15
23	0.900	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.89	1.84	1.80	1.74	1.69	1.64	1.55
	0.950	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.27	2.20	2.13	2.05	1.96	1.88	1.76
	0.975	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.67	2.57	2.47	2.36	2.24	2.14	1.97
	0.990	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.21	3.07	2.93	2.78	2.62	2.48	2.26
	0.999	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.73	4.48	4.23	3.96	3.68	3.44	3.05
24	0.900	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.88	1.83	1.78	1.73	1.67	1.62	1.53
	0.950	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.18	2.11	2.03	1.94	1.86	1.73
	0.975	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.64	2.54	2.44	2.33	2.21	2.11	1.94
	0.990	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.17	3.03	2.89	2.74	2.58	2.44	2.21
	0.999	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.64	4.39	4.14	3.87	3.59	3.36	2.97
25	0.900	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.87	1.82	1.77	1.72	1.66	1.61	1.52
	0.950	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.24	2.16	2.09	2.01	1.92	1.84	1.71
	0.975	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.61	2.51	2.41	2.30	2.18	2.08	1.91
	0.990	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.13	2.99	2.85	2.70	2.54	2.40	2.17
	0.999	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.56	4.31	4.06	3.79	3.52	3.28	2.89
26	0.900	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.86	1.81	1.76	1.71	1.65	1.59	1.50
	0.950	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	2.15	2.07	1.99	1.90	1.82	1.69
	0.975	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.59	2.49	2.39	2.28	2.16	2.05	1.88
	0.990	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.09	2.96	2.81	2.66	2.50	2.36	2.13
	0.999	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.48	4.24	3.99	3.72	3.44	3.21	2.82

Table B.14 – *Table des quantiles des lois F, pour différentes valeurs de k_1, k_2 et α.*

$k_2 \backslash k_1$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
	α														
27	0.900	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.85	1.80	1.75	1.70	1.64	1.58	1.49
	0.950	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.20	2.13	2.06	1.97	1.88	1.81	1.67
	0.975	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.57	2.47	2.36	2.25	2.13	2.03	1.85
	0.990	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.06	2.93	2.78	2.63	2.47	2.33	2.10
	0.999	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.41	4.17	3.92	3.66	3.38	3.14	2.75
28	0.900	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.84	1.79	1.74	1.69	1.63	1.57	1.48
	0.950	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.19	2.12	2.04	1.96	1.87	1.79	1.65
	0.975	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.55	2.45	2.34	2.23	2.11	2.01	1.83
	0.990	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.03	2.90	2.75	2.60	2.44	2.30	2.06
	0.999	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.35	4.11	3.86	3.60	3.32	3.09	2.69
29	0.900	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.83	1.78	1.73	1.68	1.62	1.56	1.47
	0.950	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.18	2.10	2.03	1.94	1.85	1.77	1.64
	0.975	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.53	2.43	2.32	2.21	2.09	1.99	1.81
	0.990	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.00	2.87	2.73	2.57	2.41	2.27	2.03
	0.999	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.29	4.05	3.80	3.54	3.27	3.03	2.64
30	0.900	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.82	1.77	1.72	1.67	1.61	1.55	1.46
	0.950	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	2.01	1.93	1.84	1.76	1.62
	0.975	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.51	2.41	2.31	2.20	2.07	1.97	1.79
	0.990	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.98	2.84	2.70	2.55	2.39	2.25	2.01
	0.999	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.24	4.00	3.75	3.49	3.22	2.98	2.59
60	0.900	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.71	1.66	1.60	1.54	1.48	1.41	1.29
	0.950	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.84	1.75	1.65	1.56	1.39
	0.975	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.27	2.17	2.06	1.94	1.82	1.70	1.48
	0.990	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.63	2.50	2.35	2.20	2.03	1.88	1.60
	0.999	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.54	3.32	3.08	2.83	2.55	2.32	1.89
80	0.900	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.44	1.38	1.24
	0.950	3.11	2.72	2.49	2.33	2.21	2.13	2.06	1.95	1.88	1.79	1.70	1.60	1.51	1.32
	0.975	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.21	2.11	2.00	1.88	1.75	1.63	1.40
	0.990	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.55	2.42	2.27	2.12	1.94	1.79	1.49
	0.999	7.54	5.97	5.12	4.58	4.20	3.92	3.70	3.39	3.16	2.93	2.68	2.41	2.16	1.72
100	0.900	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.66	1.61	1.56	1.49	1.42	1.35	1.21
	0.950	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.93	1.85	1.77	1.68	1.57	1.48	1.28
	0.975	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.18	2.08	1.97	1.85	1.71	1.59	1.35
	0.990	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.50	2.37	2.22	2.07	1.89	1.74	1.43
	0.999	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.30	3.07	2.84	2.59	2.32	2.08	1.62
∞	0.900	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.60	1.55	1.49	1.42	1.34	1.26	1.00
	0.950	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.83	1.75	1.67	1.57	1.46	1.35	1.00
	0.975	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.05	1.94	1.83	1.71	1.57	1.43	1.00
	0.990	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.32	2.18	2.04	1.88	1.70	1.52	1.00
	0.999	6.91	5.42	4.62	4.10	3.74	3.47	3.27	2.96	2.74	2.51	2.27	1.99	1.73	1.00

Bibliographie

[1] Bartoszyński et Niewiadomska-Bugaj (1996). *Probability and Statistical Inference*. John Wiley, New York.

[2] Bickel, P.J., et Doksum, K.A. (2001). *Mathematical Statistics. Basic Ideas and Selected Topics*. Prentice Hall, New Jersey.

[3] Billingsley, P. (1995) *Probability and Measure*. John Wiley, New York.

[4] Bilodeau, M., and D. Brenner (1999). *Theory of Multivariate Statistics*. Springer-Verlag, New York.

[5] Cornfield, J. (1975). A statistician's apology. *J. Amer. Statist. Assoc., 70*(349), 7-14.

[6] Knight, K. (2000). *Mathematical Statistics*. Chapman & Hall/CRC.

[7] Roussas, G.G. (1997). *A Course in Mathematical Statistics*. Academic Press, San Diego.

[8] Shao, J. (2003). *Mathematical Statistics*. Springer-Verlag, New York.

[9] Van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

[10] Young, G.A., et Smith, R.L. (2005). *Essentials of Statistical Inference*. Cambridge University Press, New York.