

POLYTECH DIJON
Année universitaire 2024-2025
3A FISE I/E

PROBABILITÉS ET STATISTIQUES

ARNAUD ROUSSELLE
arnaud.rouselle@u-bourgogne.fr

Préambule

Ces notes et le cours qui leur est associé ont pour objectifs de fournir aux lecteurs et étudiants qui le suivront un point d'entrée à l'étude des méthodes statistiques courantes et les prérequis probabilistes nécessaires pour comprendre celles-ci. Elles suivent la structure des cours et ouvrages classiques de référence en la matière à l'instar de [1, 4, 5, 7–9, 11, 14, 15]. Ces notes ne prétendent pas à l'exhaustivité et leur lecture à vocation, en y facilitant l'accès, à être complétée par celle d'éléments de ces ouvrages de référence ou plus généralement de l'abondante littérature probabiliste et statisticienne. Ainsi, elles doivent permettre aux lecteurs et étudiants d'acquérir une certaine autonomie pour étoffer leurs connaissances en statistique afin de répondre à des problèmes concrets et pratiques qu'ils pourront rencontrer dans le cadre de leurs futurs stages et emplois.

Les premiers chapitres fournissent des prérequis probabilistes aux études statistiques et sont complétés par une liste de *lois usuelles* en Appendice A. Plus précisément, le Chapitre 1 donne des éléments de la *théorie de la mesure et de l'intégration* nécessaires à l'introduction et à l'étude des variables aléatoires et de leur comportement asymptotique faite dans les Chapitres 2 à 4. Le Chapitre 5 introduit le cadre général de l'échantillonnage statistique. Le Chapitre 6 est dédié à l'*estimation ponctuelle paramétrique*. On y expose les méthodes classiques de construction des *estimateurs* et d'analyse de leur qualité de façon assez détaillée. Le Chapitre 7 introduit les concepts de tests d'hypothèses dans le cadre paramétrique, basés en grande partie sur les estimateurs du chapitre précédant, et présente des outils de comparaison de tests de façon un peu plus succincte. Une ouverture possible est d'étudier également les tests d'hypothèses non paramétriques auxquels il est fait allusion. Le Chapitre 8 est consacré à l'*estimation par intervalle ou région de confiance* pour les paramètres, en se restreignant, par souci de simplicité et de brièveté, au cas unidimensionnel. Les méthodes classiques de construction d'intervalles de confiance et la dualité avec les tests d'hypothèses y sont présentées. Le temps étant limité, les notions de modèles de régression (linéaires ou logistiques) ne pourront être présentées durant ce cours et sont absentes de ces notes. Elles restent néanmoins cruciales et sont certainement celles qui pourrait être étudiées en priorité par un lecteur ou étudiant désireux d'étoffer ces connaissances statistiques après ce cours. Par ailleurs, un des objectifs de ce cours étant de développer les compétences en statistiques à des fins d'applications sur des situations concrètes, avec des données de grande taille, l'utilisation de logiciels ou langages adaptés aux statistiques doit être abordée. Dans le cadre de ce cours, nous utiliserons R (évoqué brièvement en Appendice B), en particulier lors des dernières séances de TD. Aussi, les domaines d'application des lois usuelle décrits dans l'Appendice A se révéleront forts utiles dans l'optique d'application en permettant le choix de modèles cohérents et adaptés au contexte.

Enfin, j'adresse mes plus sincères remerciements à IOANNIS IAKOVOGLOU, KARINE SERIER et MAXIME BERGER (par ordre chronologique) pour leur relecture attentive et leurs

remarques avisées qui ont permis d'améliorer la qualité de ces notes.

Table des matières

1	Éléments de la théorie de la mesure et de l'intégration	1
1.1	Tribus, espaces mesurables	1
1.2	Mesures et probabilités	2
1.3	Ensembles négligeables et la notion de μ -presque partout	4
1.4	Applications mesurables, mesures images	4
1.5	Intégration	5
2	Variables aléatoires	9
2.1	Fonction de répartition	10
2.2	Fonction quantile	11
2.3	Quantiles et médiane	11
2.4	Espérance et moments	12
2.5	Fonction génératrice des moments	15
2.6	Fonction caractéristique	16
2.7	Inégalités classiques	18
2.7.1	Inégalité de Jensen	18
2.7.2	Inégalité de Cauchy-Schwarz	18
2.7.3	Inégalité de Hölder	18
2.7.4	Inégalité de Minkowski	19
2.7.5	Inégalité de Markov	19
2.7.6	Inégalité de Tchebychev	19
2.7.7	Inégalité de Bernstein	19
2.8	Simulation de variables aléatoires	20
3	Couples, n-uplets et familles de variables aléatoires	23
3.1	Fonction de répartition, lois marginales, probabilités et densités conjointes	23
3.2	Lois conditionnelles	24
3.3	Indépendance	25
3.4	Espérance, variance, covariance, corrélation linéaire	27
3.5	Sommes de variables aléatoires	29
3.6	Lois normales multivariés, vecteurs gaussiens	31
3.7	Familles exponentielles de lois	33
3.7.1	Famille des lois de Bernoulli	34
3.7.2	Famille des lois binomiales avec n connu	34
3.7.3	Famille des lois géométriques	34
3.7.4	Famille des lois de Poisson	34

3.7.5	Famille des lois binomiales négatives avec r connu	35
3.7.6	Famille des lois exponentielles	35
3.7.7	Famille des lois gamma avec r connu	35
3.7.8	Famille des lois beta avec p connu	35
3.7.9	Famille des lois beta avec q connu	35
3.7.10	Famille des lois beta	36
3.7.11	Famille des lois normales avec m connu	36
3.7.12	Famille des lois normales avec σ^2 connu	36
3.7.13	Famille des lois normales	36
3.7.14	Famille des lois de Pareto avec a connu	37
4	Modes de convergence de variables aléatoires, théorèmes limites	39
4.1	Modes de convergence de variables aléatoires	39
4.1.1	Convergence presque sûre	39
4.1.2	Convergence en probabilité	39
4.1.3	Convergence dans L^p	39
4.1.4	Convergence en loi	40
4.1.5	Liens entre les modes de convergence	40
4.2	Théorèmes limites	40
4.2.1	Lois des grands nombres	41
4.2.2	Théorème Central Limite	41
5	Principes fondamentaux de l'échantillonnage	43
5.1	Généralités et approche empirique	43
5.2	Statistiques d'ordre	45
5.3	Cas des lois mères gaussiennes	46
6	Estimation paramétrique ponctuelle	49
6.1	Cadre de l'estimation paramétrique ponctuelle	49
6.2	Méthodes classiques de construction d'estimateurs	51
6.2.1	Méthode de substitution	51
6.2.2	Méthode des moments	52
6.2.3	Méthode du maximum de vraisemblance	53
6.2.4	Approche bayésienne	54
6.3	Analyse des estimateurs et choix d'un estimateur	56
6.3.1	Biais	57
6.3.2	Risque quadratique ou erreur quadratique moyenne	57
6.3.3	Modèles et estimateurs réguliers	59
6.3.4	Score, information de Fisher et borne de Cramer-Rao	60
6.3.5	Exhaustivité, minimalité	64
6.3.6	Analyse asymptotique	66
7	Tests d'hypothèses	69
7.1	Cadre et généralités sur tests d'hypothèses	69
7.1.1	Puissance d'un test et erreurs	70
7.1.2	Niveau et seuil d'un test	70
7.1.3	Statistique de test	71

TABLE DES MATIÈRES

7.1.4	p -valeur	71
7.2	Construction de tests	71
7.2.1	Tests du rapport de vraisemblance	71
7.2.2	Tests du rapport de vraisemblance généralisé	73
7.2.3	Tests bayésiens	74
7.3	Comparaison et analyse des tests	75
7.3.1	Tests UPP et UPPSB	75
7.3.2	Cas des tests entre deux hypothèses simples	76
7.3.3	Modèles à rapport de vraisemblance monotone	78
7.3.4	Cas des tests avec hypothèses composites	79
7.4	Mise en œuvre d'un test	82
7.5	Quelques tests usuels	83
7.5.1	Quelques Tests paramétriques	83
7.5.2	Test du Khi-2 d'indépendance	87
7.5.3	Test du Khi-2 d'adéquation à une loi	88
7.5.4	Voir aussi	88
8	Estimation par intervalles ou régions de confiance	91
8.1	Estimation par intervalles de confiance de niveau exact ou par excès	91
8.2	Estimation par intervalles de confiance asymptotiques	93
8.3	Approche bayésienne	94
8.4	Correspondance entre intervalles de confiance et tests	94
8.5	Bases pour quelques intervalles de confiance usuels	95
8.5.1	IC pour une moyenne	95
8.5.2	IC sur la différence des moyennes de deux échantillons gaussiens	96
8.5.3	IC pour la variance d'un échantillon gaussien	97
8.5.4	IC pour le rapport de variances de deux échantillons gaussiens	97
8.5.5	IC pour une proportion	97
A	Lois usuelles	99
A.1	Lois discrètes usuelles	99
A.1.1	Loi uniforme discrète	99
A.1.2	Loi de Bernoulli	99
A.1.3	Loi binomiale	100
A.1.4	Loi multinomiale	101
A.1.5	Loi de Poisson	102
A.1.6	Loi géométrique	103
A.1.7	Loi binomiale négative	103
A.1.8	Loi hypergéométrique	104
A.2	Lois continues usuelles	105
A.2.1	Loi uniforme continue	105
A.2.2	Loi exponentielle	105
A.2.3	Loi gamma	106
A.2.4	Loi beta	107
A.2.5	Loi normale	107
A.2.6	Loi normale multidimensionnelle	108
A.2.7	Loi log-normale	108

A.2.8	Loi de Pareto	109
A.2.9	Loi de Cauchy	109
A.2.10	Loi de Weibull	110
A.2.11	Loi de Gumbel	110
A.2.12	Loi de Fréchet	111
A.2.13	Loi du Khi-2	111
A.2.14	Loi de Student	112
A.2.15	Loi de Fisher-Snedecor	113
B	Quelques mots sur R	115
B.1	Création, lecture et sauvegarde de données	115
B.2	Extraction de donnée	116
B.3	Opération de base	116
B.4	Fonctions mathématiques	116
B.5	Fonctions probabilistes	117
B.6	Graphiques	117
B.7	Programmation	119
	Bibliographie	121

Chapitre 1

Éléments de la théorie de la mesure et de l'intégration

Dans ce chapitre, on donne quelques notions de la théorie de la mesure et de l'intégration nécessaires à la définition formelle des variables aléatoires puis à l'analyse statistique dans ce cours. Ces variables aléatoires (v.a.), généralement représentées par une lettre majuscule (X, Y, \dots), vont permettre d'observer numériquement des phénomènes aléatoires issus d'une expérience aléatoire sur un *ensemble fondamental* (ou des possibles) Ω . Le lecteur désireux de plus amples détails pourra par exemple consulter les premiers chapitres de [4] ou de [9].

1.1 Tribus, espaces mesurables

Définition 1.1. Soit Ω un ensemble et $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω .

Un sous-ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ est appelé tribu sur Ω si :

1. $\Omega \in \mathcal{A}$,
2. \mathcal{A} est stable par passage au complémentaire :

$$A \in \mathcal{A} \implies A^c := \Omega \setminus A \in \mathcal{A},$$

3. \mathcal{A} est stable par réunion finie ou dénombrable : pour I fini ou dénombrable, on a

$$(A_i)_{i \in I} \subset \mathcal{A} \implies \bigcup_{i \in I} A_i \in \mathcal{A}.$$

Le couple (Ω, \mathcal{A}) est alors appelé espace mesurable.

Exemple 1.1. Soit Ω un ensemble.

1. La tribu *triviale* ou *grossière* sur Ω est $\mathcal{A} = \{\emptyset, \Omega\}$.
2. La tribu *discrète* sur Ω est $\mathcal{B} = \mathcal{P}(\Omega)$.
3. Soit $\Omega = \mathbf{R}$. Le sous-ensemble

$$\mathcal{C} = \{]a; b[: -\infty \leq a < b \leq +\infty\}$$

de $\mathcal{P}(\Omega)$ n'est pas une tribu sur \mathbf{R} puisqu'il n'est pas stable par passage au complémentaire.

Remarque 1.1. Dans l'exemple précédent, la tribu \mathcal{B} est plus fine que la tribu \mathcal{A} au sens où $\mathcal{B} \supset \mathcal{A}$. En fait, parmi toutes les tribus sur Ω , la tribu discrète la plus fine alors que la tribu trivial est la moins fine.

Exercice 1.1.

1. Montrer que l'intersection de deux tribus \mathcal{A} et \mathcal{B} sur Ω est une tribu.
2. L'union de deux tribus \mathcal{A} et \mathcal{B} sur Ω est-elle toujours une tribu ? Justifier.

Définition 1.2. Soit \mathcal{E} un sous-ensemble de $\mathcal{P}(\Omega)$.

La tribu engendrée par \mathcal{E} dans Ω est l'intersection de toutes les tribus sur Ω contenant \mathcal{E} . Elle n'est autre que la tribu la moins fine sur Ω contenant \mathcal{E} et est notée $\sigma(\mathcal{E})$.

Lorsque $\Omega = \mathbf{R}$, on appelle *tribu borélienne sur \mathbf{R}* et on note $\mathcal{B}(\mathbf{R})$ la tribu engendrée par

$$\mathcal{C} = \{]a; b[: -\infty \leq a < b \leq +\infty\}$$

(soit encore par les ouverts de \mathbf{R}). De manière analogue, la tribu borélienne $\mathcal{B}(\mathbf{R}^d)$ sur \mathbf{R}^d est la tribu engendrée par les (pavés) ouverts de \mathbf{R}^d . On montre qu'elle coïncide avec la tribu produit $\mathcal{B}(\mathbf{R}) \otimes \cdots \otimes \mathcal{B}(\mathbf{R})$. Sauf mention explicite du contraire, \mathbf{R} et \mathbf{R}^d seront munis de leurs tribus boréliennes dans ce cours.

1.2 Mesures et probabilités

Définition 1.3. Soit (Ω, \mathcal{A}) un espace mesurable.

On appelle *mesure (positive)* toute application μ de \mathcal{A} dans $\mathbf{R}_+ \cup \{+\infty\}$ telle que :

1. $\mu(\emptyset) = 0$,
2. [σ -**additivité**] pour toute famille $(A_i)_{i \in I} \subset \mathcal{A}$, avec I fini ou dénombrable et $A_i \cap A_j = \emptyset$ dès que $i \neq j$:

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i).$$

S'il existe une suite exhaustive $(A_n)_{n \in \mathbf{N}} \subset \mathcal{A}$ de Ω (i.e. telle que $\bigcup_{n \in \mathbf{N}} A_n = \Omega$) vérifiant $\mu(A_n) < +\infty$ pour tout $n \in \mathbf{N}$, μ est dite σ -finie.

Si $\mu(\Omega) = 1$, μ est appelé *mesure de probabilité* ou *probabilité*.

Remarque 1.2.

1. Si μ est une probabilité alors μ est σ -finie.
2. Les mesures de probabilités sont fréquemment (voire généralement) notées \mathbf{P} .

Définition[-Théorème] 1.4. Soient μ une mesure σ -finie sur (Ω, \mathcal{A}) et ν une mesure σ -finie sur (E, \mathcal{B}) . Alors, une unique mesure sur $(\Omega \times E, \mathcal{A} \otimes \mathcal{B})$ notée $\mu \otimes \nu$ vérifie, pour tout $A \in \mathcal{A}$ et $B \in \mathcal{B}$:

$$\mu \otimes \nu(A \times B) = \mu(A) \times \nu(B).$$

Cette mesure est appelée *mesure produit* de μ par ν et est σ -finie.

Remarque 1.3.

CHAPITRE 1. ÉLÉMENTS DE LA THÉORIE DE LA MESURE ET DE L'INTÉGRATION

1. Si μ et ν sont des probabilités, il en est de même pour $\mu \otimes \nu$.
2. La tribu $\mathcal{A} \otimes \mathcal{B}$ apparaissant dans la Définition-Théorème précédente est la *tribu produit* engendrée par les pavés $A \times B$, $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Exemple 1.2.

1. [**Pile ou face**] Soit $\Omega = \{P, F\}$ (P pour « pile » et F pour « face ») muni de la tribu discrète. L'application $\mu : \mathcal{A} \rightarrow \mathbf{R}_+ \cup \{+\infty\}$ définie par $\mu(A) = \text{Card}(A)/2$ est une probabilité. Elle modélise la situation correspondant au cas du tirage d'une pièce équilibrée.
2. [**Masse de Dirac**] Soit (Ω, \mathcal{A}) un espace mesurable et $\omega \in \Omega$. L'application de \mathcal{A} dans $\mathbf{R}_+ \cup \{+\infty\}$

$$\delta_\omega : A \mapsto \mathbf{1}_{\omega \in A}$$

est une probabilité sur Ω appelée *masse de Dirac en ω* .

3. [**Mesure de comptage sur \mathbf{N}**] Munissons \mathbf{N} de sa tribu discrète. L'application définie par

$$\mu_c : A \mapsto \sum_{n \in \mathbf{N}} \delta_n(A)$$

est une mesure σ -finie (mais pas de probabilité) sur \mathbf{N} appelée *mesure de comptage sur \mathbf{N}* .

4. [**Mesure de Lebesgue**] On montre qu'il existe une unique mesure sur \mathbf{R} (muni de sa tribu borélienne) telle que $\mu([a; b]) = b - a$. Cette mesure jouera un rôle important dans la théorie de l'intégration et l'étude des variables aléatoires continues. Sa construction est, pour autant, en dehors des objectifs de ce cours. Elle est notée λ et appelée *mesure de Lebesgue sur \mathbf{R}* . La mesure $\lambda \otimes \dots \otimes \lambda$ (d fois) est la mesure de Lebesgue sur \mathbf{R}^d et donne pour masse à un pavé son volume.

Définition 1.5. Si \mathcal{A} est une tribu sur Ω , le couple (Ω, \mathcal{A}) est appelé un espace mesurable.

Si, de plus, μ est une mesure (resp. probabilité) sur (Ω, \mathcal{A}) le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé un espace mesuré (resp. probabilisé).

La proposition suivante liste les principales propriétés des mesures de probabilité découlant directement de leur définition (et restant valables pour les mesures positive à l'exception du premier point).

Proposition 1.1. Soient $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé.

On a :

1. [**passage au complémentaire**] $\mathbf{P}[A^c] = 1 - \mathbf{P}[A]$;
2. [**croissance**] si $A \subset B$, alors $\mathbf{P}[A] \leq \mathbf{P}[B]$;
3. [**σ -sous-additivité**] pour toute famille $(A_i)_{i \in I} \subset \mathcal{A}$, avec I fini ou dénombrable :

$$\mathbf{P}\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mathbf{P}(A_i);$$

4. [**inclusion-exclusion**]

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k});$$

5. si $(A_i)_{i \in \mathbf{N}} \subset \mathcal{A}$ est croissante (i.e. $A_i \subset A_{i+1}$ pour tout $i \in \mathbf{N}$),

$$\mathbf{P} \left(\bigcup_{i \in \mathbf{N}} A_i \right) = \lim_{i \rightarrow +\infty} \mathbf{P}(A_i);$$

6. si $(A_i)_{i \in \mathbf{N}} \subseteq \mathcal{A}$ est décroissante (i.e. $A_i \supseteq A_{i+1}$ pour tout $i \in \mathbf{N}$),

$$\mathbf{P} \left(\bigcap_{i \in \mathbf{N}} A_i \right) = \lim_{i \rightarrow +\infty} \mathbf{P}(A_i).$$

1.3 Ensembles négligeables et la notion de μ -presque partout

Une mesure μ sur (Ω, \mathcal{A}) peut donner une masse 0 à d'autres ensembles que l'ensemble vide. On dit qu'un (sous-)ensemble de A est μ -négligeable ou plus simplement *négligeable* s'il existe $B \in \mathcal{A}$ tel que $A \subset B$ et $\mu(B) = 0$. On dira qu'une propriété est vraie μ -presque partout (μ -p.p.) si l'ensemble sur lequel elle est fautive est négligeable. Ces notions permettent, en particulier, d'affaiblir les hypothèses de certains résultats en ne les exigeant seulement μ -p.p. et non ponctuellement. Lorsque μ est une mesure de probabilité, on dit qu'un événement $A \in \mathcal{A}$ a lieu μ -presque sûrement (μ -p.s.) lorsque son complémentaire est négligeable, soit encore lorsque $\mu(A) = 1$.

1.4 Applications mesurables, mesures images

Définition 1.6. Soient (Ω, \mathcal{A}) et (E, \mathcal{B}) deux espaces mesurables.

On dit qu'une application f de Ω dans E est mesurable (pour \mathcal{A} et \mathcal{B}) si pour tout $B \in \mathcal{B}$, l'image réciproque de B par f :

$$f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\}$$

appartient à \mathcal{A} .

Exemple 1.3. Soit $A \subset \Omega$ et $\mathbf{1}_A : \Omega \mapsto \mathbf{R}$ l'indicatrice de A . Il apparaît clairement que $\mathbf{1}_A$ est mesurable si, et seulement si, $A \in \mathcal{A}$.

Remarque 1.4. Les fonctions rencontrées dans ce cours seront (comme presque toujours dans la pratique) mesurables. Il est toutefois possible de construire des fonctions non mesurables (par exemple, l'indicatrice de l'ensemble de Vitali) en utilisant notamment l'*axiome du choix*.

Exercice 1.2. [★] Montrer que toute fonction continue de \mathbf{R} dans \mathbf{R} est mesurable.

Définition[-Théorème] 1.7. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, (E, \mathcal{B}) un espace mesurable et f une application mesurable de Ω dans E . Alors, l'application

$$\begin{aligned} \mu_f : \mathcal{B} &\longrightarrow \mathbf{R}_+ \cup \{+\infty\} \\ B &\longmapsto \mu \left(f^{-1}(B) \right) \end{aligned}$$

définit une mesure sur (E, \mathcal{B}) appelée *mesure image* de μ par f .

1.5 Intégration

Une théorie de l'intégration peut être développée à partir des notions introduites précédemment. Elle permet d'intégrer des fonctions mesurables de $(\Omega, \mathcal{A}, \mu)$ dans \mathbf{R} (ou \mathbf{R}^d muni de sa tribu borélienne). Cette approche permet d'intégrer des fonctions qui ne sont pas intégrables au sens de Riemann et de traiter de façon unifiée l'intégration par rapport à des mesures de natures très différentes telles que la mesure de comptage ou la mesure de Lebesgue. Nous référons par exemple au Chapitre 2 de [4] ou au plus technique Chapitre 5 de [2] pour plus de détails.

Globalement, pour définir l'intégrale sur Ω , on commence par imposer que l'intégrale de l'indicatrice d'un ensemble mesurable $A \in \mathcal{A}$ ne soit autre que sa mesure :

$$\int \mathbf{1}_A \, d\mu = \mu(A).$$

Dans un second temps, on étend naturellement cette définition aux fonctions *étagées* positives, *i.e.* de la forme :

$$f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}, \quad A_i \in \mathcal{A}, \quad a_i \geq 0,$$

par

$$\int f \, d\mu = \sum_{i=1}^n a_i \mu(A_i).$$

Cette extension est guidée par la volonté de garantir la linéarité de l'intégrale. Ensuite, pour une fonction mesurable positive f , on prolonge la définition par approximation :

$$\int f \, d\mu = \sup \left\{ \int g \, d\mu : g \text{ étagée positive telle que } g \leq f \right\}.$$

Finalement, pour f mesurable quelconque, on dira que f est intégrable si $\int |f| \, d\mu < +\infty$. Dans ce cas, en écrivant f sous la forme $f = f^+ - f^-$ où $f^+ = \max(f, 0)$ et $f^- = \max(-f, 0)$, on pose :

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

Cette approche se révélant particulièrement fructueuse pour établir des résultats dans cette théorie, elle est appelée par certains auteurs « méthode standard ».

On définit l'intégrale de f sur $B \in \mathcal{A}$ par,

$$\int_B f \, d\mu = \int f \mathbf{1}_B \, d\mu.$$

Il est à noter que pour les fonctions intégrables en ce sens par rapport à la mesure de Lebesgue et au sens de Riemann, les valeurs des intégrales coïncident. Toutefois, l'intégrale de Lebesgue « n'est pas sensible à l'ordre des bornes » lorsque l'on intègre sur un intervalle. Par exemple, pour la fonction constante égale à 1, on a au sens de l'intégrale de Lebesgue

$$\int_{[0,5]} 1 \, dx = 5$$

et au sens de l'intégrale de Riemann

$$\int_0^5 1 \, dx = 5 \quad \text{mais} \quad \int_5^0 1 \, dx = -5$$

CHAPITRE 1. ÉLÉMENTS DE LA THÉORIE DE LA MESURE ET DE
L'INTÉGRATION

L'intégration par rapport à la mesure de comptage μ_c sur un ensemble discret D correspond, quand à elle, à des sommes et séries indexées par cet ensemble :

$$\int_D f \, d\mu_c = \sum_{x \in D} f(x).$$

Par exemple, si μ_c est la mesure de comptage sur \mathbf{N} , l'intégrale $\int_{\mathbf{N}} f \, d\mu_c$ n'est autre que la série

$$\sum_{n \in \mathbf{N}} f(n).$$

Listons maintenant les principales propriétés de cette intégrale.

Proposition 1.2. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f, g intégrables et $\alpha, \beta \in \mathbf{R}$.*

1. [linéarité] *On a :*

$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu.$$

2. *Si $g \leq f$ μ -p.p., alors $\int g \, d\mu \leq \int f \, d\mu$.*

3. *Si $f \geq 0$ μ -p.p. et $A \subset B$, alors $\int_A f \, d\mu \leq \int_B f \, d\mu$.*

4. *Si $\mu(B) = 0$, alors $\int_B f \, d\mu = 0$.*

5. *Si $f = 0$ μ -p.p. sur B , alors $\int_B f \, d\mu = 0$.*

6. *Si $f \geq 0$ μ -p.p. sur $B \in \mathcal{A}$ et $\int_B f \, d\mu = 0$, alors $f = 0$ μ -p.p. sur B .*

7. [Théorème de convergence monotone de Beppo Levi] *Si $(f_n)_n$ est une suite croissante de fonctions mesurables positives convergeant ponctuellement vers f , alors f est mesurable et*

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \int f \, d\mu.$$

8. [Lemme de Fatou] *Si $(f_n)_n$ est une suite de fonctions mesurables positives, alors*

$$\int \liminf_{n \rightarrow \infty} f_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int f_n \, d\mu.$$

9. [Théorème de convergence dominée de Lebesgue] *Si $(f_n)_n$ est une suite de fonctions mesurables convergeant ponctuellement vers f et s'il existe une fonction μ -intégrable g telle que $|f_n| \leq g$ alors, f est μ -intégrable et*

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \int f \, d\mu$$

Mentionnons maintenant un résultat fort utile pour l'étude des variables aléatoires.

Théorème 1.1 (Théorème de transport). *Soit f une application mesurable de $(\Omega, \mathcal{A}, \mu)$ dans (E, \mathcal{B}) et φ mesurable de (E, \mathcal{B}) dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ positive ou μ_f -intégrable. Alors,*

$$\int_E \varphi \, d\mu_f = \int_{\Omega} \varphi \circ f \, d\mu.$$

Pour terminer ce chapitre, donnons deux résultats permettant l'interversion d'intégrales.

CHAPITRE 1. ÉLÉMENTS DE LA THÉORIE DE LA MESURE ET DE L'INTÉGRATION

Théorème 1.2 (Théorème de Fubini-Tonelli (ou Fubini « positif »)). *Soient $(\Omega, \mathcal{A}, \mu)$ et (E, \mathcal{B}, ν) deux espaces mesurés avec μ et ν σ -finies et $(\Omega \times E, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ l'espace produit.*

Si $f : \Omega \times E \rightarrow \mathbf{R}_+ \cup \{+\infty\}$ est $\mathcal{A} \otimes \mathcal{B}$ -mesurable, alors

$$x \mapsto \int_E f(x, y) \, d\nu(y)$$

est \mathcal{A} -mesurable et

$$y \mapsto \int_\Omega f(x, y) \, d\mu(x)$$

est \mathcal{B} -mesurable. Dans ce cas, on a :

$$\int_{\Omega \times E} f(x, y) \, d\mu \otimes \nu(x, y) = \int_\Omega \left(\int_E f(x, y) \, d\nu(y) \right) d\mu(x) = \int_E \left(\int_\Omega f(x, y) \, d\mu(x) \right) d\nu(y).$$

Théorème 1.3 (Théorème de Fubini). *Soient $(\Omega, \mathcal{A}, \mu)$ et (E, \mathcal{B}, ν) deux espaces mesurés avec μ et ν σ -finies et $(\Omega \times E, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ l'espace produit.*

Si $f : \Omega \times E \rightarrow \mathbf{R}$ est $\mu \otimes \nu$ -intégrable, alors

$$x \mapsto \int_E f(x, y) \, d\nu(y)$$

est μ -intégrable et

$$y \mapsto \int_\Omega f(x, y) \, d\mu(x)$$

est ν -intégrable. Dans ce cas, on a :

$$\int_{\Omega \times E} f(x, y) \, d\mu \otimes \nu(x, y) = \int_\Omega \left(\int_E f(x, y) \, d\nu(y) \right) d\mu(x) = \int_E \left(\int_\Omega f(x, y) \, d\mu(x) \right) d\nu(y).$$

Remarque 1.5. Dans les deux résultats précédents, les fonctions

$$x \mapsto \int_E f(x, y) \, d\nu(y) \quad \text{et} \quad y \mapsto \int_\Omega f(x, y) \, d\mu(x)$$

sont à valeurs réelles μ -p.p. et ν -p.p. respectivement.

Chapitre 2

Variables aléatoires

Définition 2.1. Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé.

On appelle variable aléatoire (v.a.) toute application mesurable X de $(\Omega, \mathcal{A}, \mathbf{P})$ dans (E, \mathcal{B}) .

Dans ce cours, l'espace d'arrivé sera, sauf mention du contraire, \mathbf{R} (ou \mathbf{R}^d) muni de sa tribu borélienne. Ces variables aléatoires sont dites réelles (resp. vectorielles). Parmi les v.a. réelles, on distingue principalement les variables *discrètes* pour lesquelles le support de X $\text{Supp } X$ (informellement l'ensemble des valeurs numériques « possibles » pour X) est fini ou dénombrable et les variables *continues* pour lesquelles $\text{Supp } X$ est une réunion d'intervalles de \mathbf{R} . On peut également concevoir des v.a. réelles « hybrides », avec une partie continue et une partie discrète, mais nous ne les rencontrerons que marginalement. Notons que si X est une variable aléatoire et f est mesurable alors $Y = f(X)$ est encore une variable aléatoire.

Exemple 2.1.

1. Si l'on lance un dé classique, la variable aléatoire X donnant le résultat du lancé est une variable aléatoire discrète prenant ses valeurs dans $X(\Omega) = \{1; 2; 3; 4; 5; 6\}$.
2. Si l'on lance simultanément n pièces équilibrées, la variable aléatoire comptant le nombre de faces obtenues est discrète et prend ses valeurs dans $X(\Omega) = \{1; 2; \dots; n\}$.
3. La durée de vie X d'une ampoule est une variable aléatoire continue. Une ampoule donnée fonctionnera un temps $t > 0$ avant de griller et il n'y a *a priori* pas de durée de vie maximale. Ainsi, $X(\Omega) = \mathbf{R}_+^*$.

Les ensembles $\{X = x\} := \{\omega \in \Omega : X(\omega) = x\}$, $\{X \leq x\}$, $\{X \geq x\}$, $\{X < x\}$, $\{X > x\}$ sont des événements. On note généralement $\mathbf{P}[X = x]$, $\mathbf{P}[X \leq x]$, $\mathbf{P}[X \geq x]$, $\mathbf{P}[X < x]$, $\mathbf{P}[X > x]$ au lieu de $\mathbf{P}[\{X = x\}]$, $\mathbf{P}[\{X \leq x\}]$, $\mathbf{P}[\{X \geq x\}]$, $\mathbf{P}[\{X < x\}]$, $\mathbf{P}[\{X > x\}]$.

Définition 2.2. Soit X une variable aléatoire de $(\Omega, \mathcal{A}, \mathbf{P})$ dans (E, \mathcal{B}) .

On appelle loi de X sous \mathbf{P} la mesure de probabilité image \mathbf{P}_X sur (E, \mathcal{B}) (au sens de la Définition 1.7).

Remarque 2.1. On peut maintenant définir plus formellement le support d'une v.a. X comme le support de la mesure. Il s'agit de l'adhérence $\text{Supp}(X)$ de l'ensemble

$$\{x \in \mathbf{R} : \mathbf{P}_X(]x - \varepsilon; x + \varepsilon[) > 0, \forall \varepsilon > 0\} = \{x \in \mathbf{R} : \mathbf{P}[X \in]x - \varepsilon; x + \varepsilon[) > 0, \forall \varepsilon > 0\}.$$

Toute v.a. réelle continue X est *absolument continue* par rapport à la mesure de Lebesgue λ sur \mathbf{R} c'est-à-dire que tout négligeable pour λ est aussi négligeable pour \mathbf{P}_X (en particulier $\mathbf{P}[X = x] = 0$ pour tout x dans le cas continu). Le Théorème de Radon-Nikodym (voir par exemple [4, Théorème II.3.3]) affirme donc qu'il existe une fonction mesurable positive $f_X = \frac{d\mathbf{P}_X}{d\lambda}$, définie λ -p.p., telle que $\mathbf{P}_X(A) = \int_A f_X d\lambda$. Cette fonction f est appelée *densité* de X et permet de mener à bien les calculs. La proposition suivante donne des conditions suffisantes pour qu'une fonction f soit une densité d'une variable aléatoire continue.

Proposition 2.1. *Si $f : \mathbf{R} \rightarrow \mathbf{R}_+$ est une fonction mesurable positive telle que*

$$\int_{\mathbf{R}} f(x) dx = 1$$

alors il existe une v.a.r. dont f est une densité.

Pour une variable aléatoire discrète, la donnée de la *fonction de probabilités* (ou *probabilités élémentaires*) $p_X(x) = \mathbf{P}[X = x]$, $x \in \text{Supp } X$, suffit à caractériser sa loi.

2.1 Fonction de répartition

La fonction de répartition est un outils permettant de caractériser de façon unifiée les lois des variables aléatoires réelles, qu'elles soient discrètes ou continues.

Définition 2.3. *Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé et X une variable aléatoire réelle.*

On appelle fonction de répartition de X la fonction :

$$\begin{aligned} F_X : \mathbf{R} &\longrightarrow [0; 1]. \\ x &\longmapsto F_X(x) = \mathbf{P}[X \leq x] \end{aligned}$$

S'il n'y a pas de confusion possible, on note simplement F la fonction de répartition d'une variable aléatoire X .

Avec les notations précédentes, dans le cas discret, on obtient que

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i)$$

alors que dans le cas continu :

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Proposition 2.2. *Soit X une variable aléatoire réelle et F_X sa fonction de répartition. On a :*

1. $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
2. $\lim_{x \rightarrow +\infty} F_X(x) = 1$;
3. F_X est croissante (donc avec limite à gauche) et continue à droite ;
4. si X est continue alors F_X est continue ;
5. si X est discrète alors F_X est en escalier (constante par morceaux).

Théorème 2.1. *La fonction de répartition caractérise la loi d'une variable aléatoire : $F_X = F_Y$ si, et seulement si, $\mathbf{P}_X = \mathbf{P}_Y$.*

Le résultat suivant relie la fonction de répartition et la densité d'une v.a. continue.

Proposition 2.3. *Soit X un v.a. continue de densité f_X et de fonction de répartition F_X .*

On a :

$$f_X = F'_X \quad \lambda - p.p..$$

Exercice 2.1. Soit X une variable aléatoire continue avec $\text{Supp } X = [1; +\infty[$. On pose $Z = \frac{1}{X} + 2$.

1. Quel est le support de Z ?
2. Exprimer F_Z en fonction de F_X .

2.2 Fonction quantile

Définition 2.4. *Soit X une v.a. réelle et F_X sa fonction de répartition.*

On appelle fonction quantile de X l'inverse généralisée de F_X définie sur $]0; 1[$ par :

$$F_X^{-1}(u) : \inf\{x \in \mathbf{R} : F(x) > u\}.$$

Proposition 2.4. *Toute fonction quantile est càdlag (continue à droite et admet une limite à gauche en tout point).*

Les fonctions quantile se révéleront particulièrement utiles lors des études statistiques et pour la simulation de variables aléatoires.

2.3 Quantiles et médiane

Les quantiles d'une variable aléatoire (ou de sa loi) se révéleront très utiles en statistique, en particulier pour l'estimation par intervalle de confiance et pour la détermination de la zone de rejet dans le cadre des tests d'hypothèses.

Définition 2.5. *Soit $\alpha \in]0; 1[$. On appelle quantile d'ordre α d'une v.a. X toute valeur q_α telle que :*

$$\mathbf{P}[X \leq q_\alpha] \geq \alpha \quad \text{et} \quad \mathbf{P}[X \geq q_\alpha] \geq 1 - \alpha.$$

On appelle médiane de X sont quantile d'ordre $1/2$.

La proposition suivante est immédiate par continuité de la fonction de répartition lorsque la v.a. est continue.

Proposition 2.5. *Si X est une v.a. continue de fonction de répartition F_X , alors, pour tout $\alpha \in]0, 1[$ le quantile d'ordre α de X est unique et est caractérisé par :*

$$F(q_\alpha) = \alpha.$$

2.4 Espérance et moments

Définition 2.6. Soit X une v.a. réelle sur $(\Omega, \mathcal{A}, \mathbf{P})$.

Si X est intégrable, on appelle espérance, espérance mathématique ou encore moyenne mathématique de X la quantité :

$$\mathbf{E}[X] = \int_{\Omega} X \, d\mathbf{P}.$$

Lorsque cette quantité est nulle on dit que X est centrée.

Les propriétés de l'intégrale, et en particulier sa linéarité, se transfèrent immédiatement à l'espérance.

Le Théorème de transport (Théorème 1.1) affirme que si g est mesurable positive ou si $g(X)$ est \mathbf{P} -intégrable,

$$\mathbf{E}[g(X)] = \int_{\Omega} g(X) \, d\mathbf{P} = \int_{\mathbf{R}} g \, d\mathbf{P}_X.$$

Ceci se généralise trivialement aux vecteurs aléatoires dans \mathbf{R}^d . En particulier, si X est réelle, en choisissant g comme étant l'identité, on obtient que :

$$\mathbf{E}[X] = \int_{\mathbf{R}} x \, d\mathbf{P}_X(x).$$

Si X est continue de densité f_X , on obtient que

$$\mathbf{E}[X] = \int_{\mathbf{R}} x f_X(x) \, dx$$

alors que si X est discrète de support S (fini ou dénombrable) :

$$\mathbf{E}[X] = \sum_{x \in S} x p_X(x).$$

Remarque 2.2. On voit que

$$\mathbf{E}[\mathbf{1}_{X \in A}] = \mathbf{P}[X \in A].$$

Définition 2.7. On dit que X admet un moment d'ordre $p > 0$ si $\mathbf{E}[|X|^p] < +\infty$. Le moment d'ordre p de X est alors $\mathbf{E}[X^p]$. Dans ce cas, la norme p de X est la quantité $\|X\|_p = \mathbf{E}[|X|^p]^{\frac{1}{p}}$ et on dit que X appartient à L^p .

Si X admet un moment d'ordre 2, on définit la variance de X comme :

$$\mathbf{V}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

et son écart-type comme $\sigma_X = \sqrt{\mathbf{V}[X]}$. Lorsque $\mathbf{V}[X] = 1$ on dit que X est réduite.

Notons que l'espérance est un paramètre de position pour la v.a. X et que sa variance est un paramètre de dispersion mesurant la facilité pour cette v.a. à s'écartier de sa moyenne. En particulier, $\mathbf{V}[X] = 0$ si, et seulement si, X est presque sûrement constante.

Notons que, de façon analogue à ce que nous avons vu pour l'espérance, si X est une v.a. réelle, on a :

$$\mathbf{V}[X] = \int_{\mathbf{R}} (x - \mathbf{E}[X])^2 \, d\mathbf{P}_X(x).$$

Ainsi, si X est continue de densité f_X , on obtient que

$$\mathbf{V}[X] = \int_{\mathbf{R}} (x - \mathbf{E}[X])^2 f_X(x) \, dx$$

alors que si X est discrète de support S (fini ou dénombrable) :

$$\mathbf{V}[X] = \sum_{x \in S} (x - \mathbf{E}[X])^2 p_X(x).$$

En développant le carré dans cette définition, utilisant la linéarité de l'espérance et le fait que $\mathbf{E}[X]$ est une constante, on obtient que :

$$\begin{aligned} \mathbf{V}[X] &= \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2. \end{aligned}$$

Cette formule est connue sous le nom de formule de *décentrage de la variance*.

Proposition 2.6. *Soient X une variable aléatoire de carré intégrable et $\alpha, \beta \in \mathbf{R}$.*

On a :

$$\mathbf{V}[\alpha X + \beta] = \alpha^2 \mathbf{V}[X].$$

Preuve : Il suffit d'écrire que :

$$\begin{aligned} \mathbf{V}[\alpha X + \beta] &= \mathbf{E}[(\alpha X + \beta - \mathbf{E}[\alpha X + \beta])^2] = \mathbf{E}[(\alpha X + \beta - \alpha \mathbf{E}[X] + \beta)^2] \\ &= \mathbf{E}[(\alpha(X - \mathbf{E}[X]))^2] = \alpha^2 \mathbf{E}[(X - \mathbf{E}[X])^2] = \alpha^2 \mathbf{V}[X]. \end{aligned}$$

□

Dans certains cas, le calcul d'un moment peut être plus aisé à partir de la fonction de répartition (ou plutôt de sa queue de distribution $G_X = 1 - F_X$). On utilise pour cela le résultat suivant.

Proposition 2.7. *Soit X une v.a. réelle positive de fonction de répartition F_X . Alors, pour tout $p > 0$:*

$$\mathbf{E}[X^p] = p \int_0^\infty t^{p-1} (1 - F_X(t)) \, dt.$$

En particulier,

$$\mathbf{E}[X] = \int_0^\infty (1 - F_X(t)) \, dt.$$

Il est aussi possible d'obtenir des formules approchées pour le calcul de moments de variables aléatoires de la forme $Y = g(X)$ sous des conditions de régularité de f et d'intégrabilité. Ceci repose principalement sur l'utilisation de développements limités (D.L.) au voisinage de $\mathbf{E}[X]$. Par exemple, si g admet un D.L. d'ordre 2 au voisinage de $\mathbf{E}[X]$, en prenant l'espérance dans :

$$\begin{aligned} g(x) &= g(\mathbf{E}[X]) + g'(\mathbf{E}[X]) (x - \mathbf{E}[X]) + \frac{g''(\mathbf{E}[X])}{2} (x - \mathbf{E}[X])^2 + o((x - \mathbf{E}[X])^2) \\ &\simeq g(\mathbf{E}[X]) + g'(\mathbf{E}[X]) (x - \mathbf{E}[X]) + \frac{g''(\mathbf{E}[X])}{2} (x - \mathbf{E}[X])^2 \end{aligned}$$

on obtient

$$\begin{aligned} \mathbf{E}[g(X)] &\simeq g(\mathbf{E}[X]) + g'(\mathbf{E}[X]) \mathbf{E}[X - \mathbf{E}[X]] + \frac{g''(\mathbf{E}[X])}{2} \mathbf{E}[(X - \mathbf{E}[X])^2] \\ &\simeq g(\mathbf{E}[X]) + \frac{g''(\mathbf{E}[X])}{2} V[X]. \end{aligned}$$

Mentionnons maintenant une proposition découlant facilement du Théorème de transport (Théorème 1.1) mais néanmoins très utile pour caractériser la loi d'une v.a. au moyen d'espérances. Elle est en particulier utilisée pour déterminer la densité de v.a. ou de fonctionnelles de v.a..

Proposition 2.8. *Deux v.a. X et Y ont même loi si, et seulement si, pour toute fonction mesurable positive φ ,*

$$\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi(Y)].$$

Preuve : Le sens direct découle directement du théorème de transport. La réciproque s'obtient en considérant simplement comme fonctions φ que les indicatrices d'intervalles $] -\infty, x]$ et en remarquant que la condition $\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi(Y)]$ implique alors que $F_X = F_Y$. \square

On voit, en particulier, que si X est une v.a. continue elle admet pour densité f si, et seulement si, pour toute fonction mesurable positive φ (dite « *fonction test* » ou « *fonction muette* »),

$$\mathbf{E}[\varphi(X)] = \int_{\mathbf{R}} \varphi(x) f(x) \, dx.$$

Ce constat conduit à la *méthode de la fonction muette* permettant de déterminer la loi ou la densité de certaines variables aléatoires. L'exemple suivant illustre cette méthode (en anticipant quelque peu sur la notion d'indépendance de variables aléatoires). On admettra pour cet exemple que si X et Y sont deux v.a. indépendantes, pour toutes fonctions mesurables positives f et g , $\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]$.

Exemple 2.2.

Soit $X \sim \mathcal{E}(\lambda)$ pour un certain $\lambda > 0$ et ε , indépendante de X une variable aléatoire de Rademacher, c'est-à-dire telle que :

$$\mathbf{P}[\varepsilon = 1] = \mathbf{P}[\varepsilon = -1] = \frac{1}{2}.$$

On se propose de déterminer la densité de la loi *exponentielle symétrique* définie par $Y = \varepsilon X$.

Soit φ une fonction muette (mesurable positive). On a :

$$\begin{aligned}
 \mathbf{E}[\varphi(Y)] &= \mathbf{E}[\varphi(\varepsilon X)] = \mathbf{E}[\varphi(X)\mathbf{1}_{\varepsilon=1} + \varphi(-X)\mathbf{1}_{\varepsilon=-1}] \\
 &= \mathbf{E}[\varphi(X)\mathbf{1}_{\varepsilon=1}] + \mathbf{E}[\varphi(-X)\mathbf{1}_{\varepsilon=-1}] \quad (\text{par linéarité}) \\
 &= \mathbf{E}[\varphi(X)]\mathbf{E}[\mathbf{1}_{\varepsilon=1}] + \mathbf{E}[\varphi(-X)]\mathbf{E}[\mathbf{1}_{\varepsilon=-1}] \quad (\text{par indépendance}) \\
 &= \mathbf{E}[\varphi(X)]\mathbf{P}[\varepsilon = 1] + \mathbf{E}[\varphi(-X)]\mathbf{P}[\varepsilon = -1] \\
 &= \int_0^{+\infty} \varphi(x)\lambda e^{-\lambda x} dx \times \frac{1}{2} + \int_0^{+\infty} \varphi(-y)\lambda e^{-\lambda y} dy \times \frac{1}{2} \\
 &= \int_0^{+\infty} \varphi(x)\frac{\lambda e^{-\lambda x}}{2} dx - \int_0^{-\infty} \varphi(x)\frac{\lambda e^{\lambda x}}{2} dx \quad (\text{changement de variable } x = -y) \\
 &= \int_0^{+\infty} \varphi(x)\frac{\lambda e^{-\lambda|x|}}{2} dx + \int_{-\infty}^0 \varphi(x)\frac{\lambda e^{-\lambda|x|}}{2} dx \\
 &= \int_{-\infty}^{+\infty} \varphi(x)\frac{\lambda e^{-\lambda|x|}}{2} dx \quad (\text{par la relation de Chasles}).
 \end{aligned}$$

Ainsi, la densité de Y est donnée par

$$f_Y(x) = \frac{\lambda e^{-\lambda|x|}}{2}.$$

Dans certains cas, les moments (et donc la famille de fonction $\varphi_k(x) = x^k$) suffisent à caractériser la loi d'une variable aléatoire comme le montre le résultat suivant.

Proposition 2.9. *Soient X et Y deux v.a. à valeurs dans un intervalle fermé borné $[a; b]$.*

Si, pour tout $k \in \mathbf{N}$, $\mathbf{E}[X^k] = \mathbf{E}[Y^k]$, alors $\mathbf{P}_X = \mathbf{P}_Y$.

2.5 Fonction génératrice des moments

Définition 2.8. *Soit X une v.a. réelle sur $(\Omega, \mathcal{A}, \mathbf{P})$. On appelle fonction génératrice des moments ou transformée de Laplace de X la fonction :*

$$L_X(t) = \mathbf{E}[e^{tX}]$$

définie sur $\{t \in \mathbf{R} : \mathbf{E}[e^{tX}] < +\infty\}$.

Remarque 2.3. Plus généralement, si X est un vecteur aléatoire dans \mathbf{R}^d , sa fonction génératrice des moments est donnée par :

$$L_X(t) = \mathbf{E}[e^{t \cdot X}]$$

où $t \in \mathbf{R}^d$ et \cdot désigne le produit scalaire.

Le résultat suivant justifie la dénomination « fonction génératrice des moments » pour L_X .

Proposition 2.10. *Soit X une v.a. réelle sur $(\Omega, \mathcal{A}, \mathbf{P})$ telle que e^{tX} est P -intégrable sur un voisinage \mathcal{V} de 0 à t fixé. Alors, pour tout $t \in \mathcal{V}$:*

$$L_X(t) = \sum_{n=0}^{+\infty} \frac{t^n}{n!} \mathbf{E}[X^n].$$

En particulier, pour tout $n \in \mathbf{N}$, on a :

$$(L_X)^{(n)}(0) = \mathbf{E}[X^n].$$

Exemple 2.3. Soit X de loi géométrique de paramètre p . On a :

$$\begin{aligned} L_X(t) &= \mathbf{E}[e^{tX}] = \sum_{k=1}^{+\infty} e^{tk} (1-p)^{k-1} p \\ &= \frac{p}{1-p} \sum_{k=1}^{+\infty} (e^t(1-p))^k = \frac{p}{1-p} e^t(1-p) \frac{1}{1 - e^t(1-p)} \\ &= \frac{pe^t}{1 - e^t(1-p)} \end{aligned}$$

Exercice 2.2. Préciser le domaine de définition de la fonction génératrice des moments d'une loi géométrique omis dans l'exemple précédent.

Exemple 2.4. Soit X de loi normale centrée réduite. On a :

$$L_X(t) = \mathbf{E}[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2} + tx} dx$$

S'agissant d'une intégrale gaussienne, il convient de réécrire l'exposant de l'intégrande sous la forme $-\frac{y^2}{2} + c$. On a :

$$-\frac{x^2}{2} + tx = -\frac{1}{2}(x^2 - 2tx) = -\frac{1}{2}(x^2 - 2tx + t^2 - t^2) = -\frac{1}{2}((x-t)^2 - t^2) = -\frac{(x-t)^2}{2} + \frac{t^2}{2},$$

ce qui est la forme recherchée avec $y = x - t$ et $c = \frac{t^2}{2}$. Il vient ensuite en effectuant le changement de variable $y = x - t$ ($dy = dx$) que :

$$\begin{aligned} L_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2} + tx} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-t)^2}{2} + \frac{t^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = e^{\frac{t^2}{2}}. \end{aligned}$$

Exercice 2.3. En imitant la démarche de l'Exemple 2.4, calculer la fonction génératrice des moments d'une loi normale générique $\mathcal{N}(m; \sigma^2)$.

2.6 Fonction caractéristique

Définition 2.9. Soit X une v.a. réelle sur $(\Omega, \mathcal{A}, \mathbf{P})$. On appelle fonction caractéristique ou transformée de Fourier de X la fonction de \mathbf{R} dans \mathbf{C} définie par :

$$\phi_X(t) = \mathbf{E}[e^{itX}].$$

Remarque 2.4.

1. Si X est une v.a.r. continue de densité f , on a pour tout t

$$\phi_X(t) = \int_{\mathbf{R}} e^{itx} f(x) \, dx = \int_{\mathbf{R}} \cos(tx) f(x) \, dx + i \int_{\mathbf{R}} \sin(tx) f(x) \, dx,$$

alors que si X est discrète de fonction de probabilité p et de support $\text{Supp}(X)$, on a pour tout t :

$$\phi_X(t) = \sum_{x \in \text{Supp}(X)} e^{itx} p(x) = \sum_{x \in \text{Supp}(X)} \cos(tx) p(x) + i \sum_{x \in \text{Supp}(X)} \sin(tx) p(x).$$

2. Plus généralement, si X est un vecteur aléatoire dans \mathbf{R}^d , sa fonction caractéristique est donnée par :

$$\phi_X(t) = \mathbf{E}[e^{it \cdot X}]$$

où $t \in \mathbf{R}^d$ et \cdot désigne le produit scalaire.

Le résultat suivant justifie la dénomination « fonction caractéristique » pour ϕ_X .

Proposition 2.11. *Soient X et Y deux variables aléatoires réelles.*

Si $\phi_X = \phi_Y$, alors $\mathbf{P}_X = \mathbf{P}_Y$.

Remarque 2.5. Le résultat précédent reste valable pour les vecteurs aléatoires.

Exemple 2.5. Soit X de loi binomiale de paramètres n et p . On a :

$$\begin{aligned} \phi_X(t) &= \mathbf{E}[e^{itX}] = \sum_{k=0}^n \binom{n}{k} e^{itk} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (e^{it} p)^k (1-p)^{n-k} = (e^{it} p + 1 - p)^n. \end{aligned}$$

Exercice 2.4. Montrer que si $X \sim \mathcal{N}(m, \sigma^2)$, on a :

$$\phi_X(t) = e^{itm - \frac{\sigma^2 t^2}{2}}$$

La fonction caractéristique caractérise la loi d'une v.a. et permet sous certaines conditions d'obtenir effectivement la loi. Le résultat suivant donne un de ces moyens au travers de la *formule d'inversion de Fourier*.

Théorème 2.2. *Soit X une v.a. réelle telle que ϕ_X est intégrable par rapport à la mesure de Lebesgue sur \mathbf{R} . Alors, X admet la densité continue bornée sur \mathbf{R} donnée par :*

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-itx} \phi_X(t) \, dt.$$

Remarque 2.6. Le résultat précédent reste valable pour les vecteurs aléatoires dans \mathbf{R}^d . La densité est alors donnée par :

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} e^{-it \cdot x} \phi_X(t) \, dt.$$

Le résultat suivant est l'analogie de la Proposition 2.10 pour la fonction caractéristique.

Proposition 2.12. Soit X une v.a. réelle sur $(\Omega, \mathcal{A}, \mathbf{P})$ admettant un moment d'ordre n . Alors, ϕ_X est de classe \mathcal{C}^n et pour tout $k \leq n$, pour tout t :

$$\phi_X^{(k)}(t) = i^k \mathbf{E}[X^k e^{itX}].$$

En particulier, on a, pour tout $k \leq n$, :

$$(\phi_X)^{(k)}(0) = i^k \mathbf{E}[X^k].$$

Remarque 2.7. La réciproque partielle suivante est valide. Si ϕ_X est n fois dérivable en 0 pour un certain n pair, alors X admet un moment d'ordre n (donc tout moment d'ordre $k \leq n$) et on a les égalités de la Proposition 2.12 pour $k \leq n$.

2.7 Inégalités classiques

Dans cette section, on liste les inégalités d'usage courant sur les variables aléatoires. Celles-ci sont données pour la plupart sans preuve; les preuves et plus de détails se trouvent dans tous les ouvrages classiques de probabilité (voir [4, 9, 13] par exemple).

2.7.1 Inégalité de Jensen

Théorème 2.3. Soit X une v.a. réelle et $\varphi : \mathbf{R} \rightarrow \mathbf{R}$ une fonction convexe.

Alors, si X et $\varphi(X)$ sont intégrables :

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)].$$

2.7.2 Inégalité de Cauchy-Schwarz

Théorème 2.4. Soit X et Y deux v.a. réelle de carré intégrable.

Alors, XY est intégrable et :

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^2]^{\frac{1}{2}} \mathbf{E}[|Y|^2]^{\frac{1}{2}},$$

autrement dit

$$\|XY\|_1 \leq \|X\|_2 \|Y\|_2.$$

2.7.3 Inégalité de Hölder

Théorème 2.5. Soit X et Y deux v.a. réelle admettant respectivement un moment d'ordre $p \geq 1$ et un moment d'ordre $q \geq 1$ avec $p^{-1} + q^{-1} = 1$.

Alors, XY est intégrable et :

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{\frac{1}{p}} \mathbf{E}[|Y|^q]^{\frac{1}{q}},$$

autrement dit

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

2.7.4 Inégalité de Minkowski

Théorème 2.6. *Soit X et Y deux v.a. réelle admettant un moment d'ordre $p \geq 1$. Alors, $X + Y$ admet un moment d'ordre p et :*

$$\mathbf{E}[|X + Y|^p]^{\frac{1}{p}} \leq \mathbf{E}[|X|^p]^{\frac{1}{p}} + \mathbf{E}[|Y|^p]^{\frac{1}{p}},$$

autrement dit

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

2.7.5 Inégalité de Markov

Théorème 2.7. *Soit X une v.a. réelle intégrable. Alors, pour tout $t > 0$, on a :*

$$\mathbf{P}[X \geq t] \leq \frac{\mathbf{E}[|X|]}{t}.$$

Preuve : On a :

$$\mathbf{1}_{X \geq t} \leq \frac{X}{t} \mathbf{1}_{X \geq t} \leq \frac{|X|}{t} \mathbf{1}_{X \geq t}$$

donc

$$\mathbf{P}[X \geq t] = \mathbf{E}[\mathbf{1}_{X \geq t}] \leq \mathbf{E}\left[\frac{|X|}{t} \mathbf{1}_{X \geq t}\right] \leq \frac{\mathbf{E}[|X|]}{t}.$$

□

2.7.6 Inégalité de Tchebychev

En appliquant l'inégalité de Markov à la v.a. $|X - \mathbf{E}[X]|^2$, on obtient que

Théorème 2.8. *Soit X une v.a. réelle de carré intégrable. Alors, pour tout $t > 0$, on a :*

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq t] \leq \frac{\mathbf{V}[X]}{t^2}.$$

2.7.7 Inégalité de Bernstein

En appliquant l'inégalité de Markov à la v.a. $e^{\lambda X}$, $\lambda > 0$, on obtient que

Théorème 2.9. *Soit X une v.a. réelle telle que $e^{\lambda X}$ est intégrable pour un $\lambda > 0$. Alors, pour tout $t > 0$, on a :*

$$\mathbf{P}[X \geq t] \leq \inf_{\lambda > 0} \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda t}}.$$

2.8 Simulation de variables aléatoires

La plupart des logiciels utilisés pour le traitement statistique des données disposent, à l'instar de R, de fonction permettant de générer des réalisations de variables aléatoires. La question de la génération informatique du hasard reste néanmoins d'intérêt particulier. Sans entrer dans les détails, il est possible de concevoir des méthodes permettant la génération de nombres pseudo-aléatoires selon des lois uniformes (discrètes ou continues). La génération du hasard est donc déterministe et dépend de paramètres dont certains peuvent être accessibles à l'utilisateur. Par exemple, sous R, la *graine* (*seed*) du générateur peut être réglée par l'utilisateur grâce à la commande `set.seed`. Ceci permet d'obtenir une reproductibilité des expériences parfois agréable. À titre d'exemple, le lecteur est invité à comparer les résultats des séquences de commandes :

```
set.seed(111)
rnorm(1,mean=0,sd=1)
set.seed(111)
rnorm(1,mean=0,sd=1)
```

et

```
set.seed(111)
rnorm(1,mean=0,sd=1)
set.seed(111)
rnorm(1,mean=0,sd=1)
```

simulant respectivement deux tirages de loi normale centrée réduite.

Il est instructif de voir comment l'on peut simuler des variables aléatoires de loi quelconque à partir de la loi uniforme sur $]0; 1[$ (accessible grâce à `runif` sous R) ou de copies indépendantes de celles-ci. Une méthode générique repose sur l'utilisation de la fonction quantile.

Proposition 2.13. *Soient X une v.a. réelle, F_X sa fonction de répartition, F_X^{-1} sa fonction quantile, et $U \sim \mathcal{U}(]0; 1[)$.*

Alors, X et $F_X^{-1}(U)$ ont même loi.

Preuve : Il suffit de voir que les fonctions de répartitions de ces deux variables aléatoires coïncident.

Si $F_X^{-1}(u) \leq t$, $u \in]0; 1[$, il existe pour tout $s > t$ un réel $x < s$ tel que $F_X(x) > u$ donc $F_X(s) > u$ par croissance. On en déduit par continuité à droite de F_X que $F_X(t) \geq u$. Ainsi, si $F_X^{-1}(U) \leq t$ alors $F_X(t) \geq U$ et on a :

$$\mathbf{P}[F_X^{-1}(U) \leq t] \leq \mathbf{P}[U \leq F_X(t)] = F_X(t).$$

Il reste à voir que

$$F_X(t) = \mathbf{P}[U \leq F_X(t)] \leq \mathbf{P}[F_X^{-1}(U) \leq t].$$

Pour cela, il suffit d'observer que si $F_X(t) > u$, on a $t \in \{x \in \mathbf{R} : F_X(x) > u\}$ et donc $F_X^{-1}(u) \leq t$. \square

Exemple 2.6. En appliquant ce résultat, on obtient que si $U \sim \mathcal{U}(]0; 1[)$, $X \sim \mathcal{E}(\lambda)$ et $Y = -\frac{\ln(1-U)}{\lambda}$ alors X et Y ont même loi, ce qui permet de simuler des variables exponentielle à partir de variables uniformes sur $]0; 1[$.

Le résultat précédent permet théoriquement de simuler n'importe quelle loi à partir de la loi uniforme sur $]0, 1[$ mais n'est pas toujours applicable dans la pratique car nécessitant l'inversion de la fonction de répartition. La méthode est, par exemple, mise en défaut par la simulation de lois normales pour lesquelles on ne connaît pas d'expression analytique de la fonction de répartition et pour laquelle une inversion numérique serait trop coûteuse. On a dans ce cas recours à la méthode de Box Muller basée sur une transformation en coordonnées polaires.

Proposition 2.14. *Si U_1, U_2 sont deux v.a. indépendantes uniformes sur $]0, 1[$, alors la v.a.*

$$X = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2) \quad \text{et} \quad Y = \sqrt{-2 \ln(U_2)} \cos(2\pi U_1)$$

sont indépendantes de loi normale centrée réduite.

Sachant simuler grâce au résultat précédent une v.a. X de loi $\mathcal{N}(0; 1)$ il est aisé de simuler une v.a. Z de loi $\mathcal{N}(m; \sigma^2)$ en utilisant que

$$Z = \sigma X + m.$$

D'autres méthodes spécifiques pour la simulation de lois usuelles (en particulier discrètes) peuvent être développées. Nous ne les détaillerons pas dans ce cours. Nous ne détaillerons pas non plus les méthodes de rejet, également classiques.

Exercice 2.5. Écrire une fonction permettant la simulation d'une loi binomiale de paramètres n et p à partir de la loi uniforme sur $]0, 1[$.

Chapitre 3

Couples, n -uplets et familles de variables aléatoires

3.1 Fonction de répartition, lois marginales, probabilités et densités conjointes

Définition 3.1. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbf{P})$.

On appelle fonction de répartition (conjointe) de X la fonction définie pour $t = (t_1, \dots, t_d) \in \mathbf{R}^d$ par

$$F_X(t) = \mathbf{P}[X_1 \leq t_1, \dots, X_d \leq t_d].$$

La loi de X_i est appelée i^e marginale et est caractérisée par sa fonction de répartition :

$$F_{X_i}(t_i) = \lim_{t_k \rightarrow +\infty, \forall k \neq i} F_X(t).$$

Dans la suite, nous supposons que toutes les composantes de X sont du même « type », c'est-à-dire toutes discrètes ou toutes continues. Cette restriction est suffisante pour la suite de ce cours puisque, à nos fins statistiques, nous considérerons des échantillons *identiquement distribués*, c'est-à-dire dans lesquels toutes les v.a. suivent la même loi.

Dans le cas où toutes les composantes sont discrètes, un vecteur aléatoire est caractérisé par la donnée des ces probabilités conjointes. En supposant que le support de X_i soit donné par $\{x_j^{(i)}\}_{j \in \{1, \dots, i_k\}}$, $k = k(i)$, $i_k \in \mathbf{N}^* \cup \{+\infty\}$, il s'agit de la donnée des :

$$p_X(x_{j_1}^{(1)}, \dots, x_{j_d}^{(d)}) = \mathbf{P}[X_1 = x_{j_1}^{(1)}, \dots, X_d = x_{j_d}^{(d)}].$$

La i^e loi marginale de X est alors caractérisée par la fonction de probabilité marginale :

$$p_{X_i}(x_j^{(i)}) = \sum_{1 \leq l_k \leq i_k, \forall k \neq i} p_X(x_{l_1}^{(1)}, \dots, x_{l_d}^{(d)}).$$

Exercice 3.1. On lance deux dés classiques et on note X_1 et X_2 leurs résultats. Déterminer la loi de (X_1, X_2) et les lois marginales.

De façon analogue, dans le cas où toutes les composantes sont continues, un vecteur aléatoire est caractérisé par la donnée de sa *densité conjointe* (positive) f_X vérifiant pour

tout $t = (t_1, \dots, t_d)$:

$$F_X(t) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} f_X(s_1, \dots, s_d) \, ds_1 \dots ds_d.$$

On obtient alors la d^{e} densité marginale de X comme :

$$f_{X_d}(t_d) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_X(s_1, \dots, s_{d-1}, t_d) \, ds_1 \dots ds_{d-1}$$

et, plus généralement, la i^{e} densité marginale de X en intégrant f_X selon toutes les coordonnées sauf la i^{e} .

3.2 Lois conditionnelles

Rappelons que si A et B sont deux événements et si $\mathbf{P}[B] \neq 0$, on définit classiquement la probabilité conditionnelle de A sachant B par

$$\mathbf{P}[A|B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]}.$$

Guidés par ceci, pour un couple de v.a. (X, Y) , on peut définir la loi conditionnelle de X sachant $Y \in A$ pourvu que $\mathbf{P}[Y \in A] \neq 0$. Dans le cas discret, la loi conditionnelle de X sachant $Y \in A$ est caractérisée par la donnée des probabilités élémentaires pour x_i dans le support de X :

$$p_{X|Y \in A}(x_i) = \frac{\sum_{j: y_j \in A} p_{(X,Y)}(x_i, y_j)}{\sum_{j: y_j \in A} p_Y(y_j)}.$$

En particulier, lorsque $A = \{y_j\}$ est un singleton dans le support de Y :

$$p_{X|Y=y_j}(x_i) = \frac{p_{(X,Y)}(x_i, y_j)}{p_Y(y_j)}.$$

On définit de façon analogue la loi conditionnelle de Y sachant $X \in B$.

Rappelons que dans le cas continu, $\mathbf{P}[Y = y] = 0$ (puisque l'on intègre la densité de Y sur un singleton qui est un ensemble négligeable par rapport à la mesure de Lebesgue). Il est toutefois clair que l'on peut définir la loi de X sachant $Y \in A$ (dès que $\mathbf{P}[Y \in A] > 0$), pour X et Y continues, et il est même possible de définir, avec quelques précautions, la loi de X sachant $Y = y$. Cet objet représente alors la loi de X lorsque Y est infiniment proche de y et non égal à y . Soit $y \in \mathbf{R}$ et $\varepsilon > 0$ tel que f_Y soit strictement positive pour λ -presque-tout $s \in \mathcal{V}_y(\varepsilon) = [y - \varepsilon, y + \varepsilon]$. On peut définir la fonction de répartition de X sachant $Y \in \mathcal{V}_y(\varepsilon)$ comme :

$$\begin{aligned} F_{X|Y \in \mathcal{V}_y(\varepsilon)}(x) &= \frac{\mathbf{P}[X \leq x, Y \in \mathcal{V}_y(\varepsilon)]}{\mathbf{P}[Y \in \mathcal{V}_y(\varepsilon)]} \\ &= \frac{\mathbf{P}[X \leq x, Y \leq y + \varepsilon] - \mathbf{P}[X \leq x, Y \leq y - \varepsilon]}{\mathbf{P}[Y \leq y + \varepsilon] - \mathbf{P}[Y \leq y - \varepsilon]} \\ &= \frac{F_{(X,Y)}(x, y + \varepsilon) - F_{(X,Y)}(x, y - \varepsilon)}{F_Y(y + \varepsilon) - F_Y(y - \varepsilon)}. \end{aligned}$$

Il est alors naturel de laisser tendre ε vers 0 pour obtenir la fonction de répartition de X sachant $Y = y$.

$$\begin{aligned} F_{X|Y=y}(x) &= \lim_{\varepsilon \rightarrow 0} \frac{F_{(X,Y)}(x, y + \varepsilon) - F_{(X,Y)}(x, y - \varepsilon)}{F_Y(y + \varepsilon) - F_Y(y - \varepsilon)} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{F_{(X,Y)}(x, y + \varepsilon) - F_{(X,Y)}(x, y - \varepsilon)/(2\varepsilon)}{F_Y(y + \varepsilon) - F_Y(y - \varepsilon)/(2\varepsilon)} \\ &= \frac{\int_{-\infty}^x f_{(X,Y)}(s, y) \, ds}{f_Y(y)} \end{aligned}$$

où l'on a utilisé que la densité est la dérivée (partielle selon la deuxième coordonnée au numérateur) de la fonction de répartition d'une v.a. λ -presque partout et une interversion limite/intégrale légitime puisque la densité est positive λ -presque partout. En dérivant selon la première coordonnée, on obtient la densité conditionnelle de X sachant $Y = y$:

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

3.3 Indépendance

Définition 3.2. On dit que deux v.a. X et Y sont indépendantes si pour tous événements $\{X \in A\}$ et $\{Y \in B\}$, on a :

$$\mathbf{P}[X \in A, Y \in B] = \mathbf{P}[X \in A]\mathbf{P}[Y \in B].$$

Remarque 3.1. Pour X, Y des v.a. réelles, il suffit, en fait, de vérifier cette définition pour A, B des intervalles (ou même des singletons dans le cas discret).

Définition 3.3. On dit que des v.a. $X_i, i \in I$ sont deux à deux indépendantes si, pour tous $i \neq j, X_i$ et X_j sont indépendantes.

Définition 3.4. On dit qu'une famille de v.a. $\{X_i\}_{i \in I}$ est (mutuellement) indépendante si pour tout $J \subset I$ fini, on a :

$$\mathbf{P}[X_i \in A_i, i \in J] = \prod_{i \in J} \mathbf{P}[X_i \in A_i].$$

Remarque 3.2. L'indépendance (mutuelle) entraîne l'indépendance 2 à 2.

L'indépendance d'une famille finie de v.a. (ou de leurs lois) se traduit en terme de mesure de probabilité produit comme suit (voir, par exemple, [9, Théorème 5.2.5]).

Théorème 3.1. Les v.a. X_1, \dots, X_n sont indépendantes si, et seulement si :

$$\mathbf{P}_{(X_1, \dots, X_n)} = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}.$$

Voyons maintenant comment se traduit l'indépendance d'une famille finie de v.a. dans les cas discret et continu.

Proposition 3.1. Soient X_1, \dots, X_d des v.a. discrètes. Alors, X_1, \dots, X_d sont indépendantes si, et seulement si,

$$p_{X_1, \dots, X_d}(x_1, \dots, x_d) = \prod_{i=1}^d p_{X_i}(x_i).$$

Proposition 3.2. Soient X_1, \dots, X_d des v.a. continues. Alors, X_1, \dots, X_d sont indépendantes si, et seulement si,

$$F_{X_1, \dots, X_d}(x_1, \dots, x_d) = \prod_{i=1}^d F_{X_i}(x_i).$$

Si ces variables sont à densité, alors, X_1, \dots, X_d sont indépendantes si, et seulement si,

$$f_{X_1, \dots, X_d}(x_1, \dots, x_d) = \prod_{i=1}^d f_{X_i}(x_i).$$

Proposition 3.3. Si X et Y sont indépendantes, alors, pour toutes fonctions g et h , $g(X)$ et $h(Y)$ sont indépendantes.

On montre le résultat plus fort suivant (voir [4, Corollaire IV.1.11]).

Théorème 3.2. Soit X_i , $i \in I$ une famille de v.a. réelles. Alors, les X_i sont indépendantes si, et seulement si, pour tout $J \subset I$ fini, pour toutes fonctions mesurables φ_j telles que $\varphi_j(X_j)$ est intégrable pour tout $j \in J$, on a :

$$\mathbf{E} \left[\prod_{j \in J} \varphi_j(X_j) \right] = \prod_{j \in J} \mathbf{E} [\varphi_j(X_j)].$$

L'indépendance de variables aléatoires peut également être caractérisée à l'aide des fonctions caractéristiques.

Théorème 3.3. Soient X_1, \dots, X_n des v.a. réelles.

Alors, X_1, \dots, X_n sont indépendantes si, et seulement si,

$$\phi_{(X_1, \dots, X_n)} = \prod_{i=1}^n \phi_{X_i}.$$

Preuve :

Le sens direct est une conséquence immédiate du résultat précédent puisque si X_1, \dots, X_n sont indépendantes :

$$\phi_{(X_1, \dots, X_n)}(t) = \mathbf{E} \left[e^{i \sum_{j=1}^n t_j X_j} \right] = \mathbf{E} \left[\prod_{j=1}^n e^{it_j X_j} \right] = \prod_{j=1}^n \mathbf{E} \left[e^{it_j X_j} \right] = \prod_{i=1}^n \phi_{X_i}$$

en choisissant $\varphi_j(u) = e^{it_j u}$. Il entraîne que la fonction caractéristique de la mesure de probabilité produit $\mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}$ se factorise sous la forme :

$$\phi_{\mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}} = \prod_{i=1}^n \phi_{X_i}.$$

Ainsi, si $\phi_{(X_1, \dots, X_n)} = \prod_{i=1}^n \phi_{X_i}$, on a :

$$\phi_{\mathbf{P}_{(X_1, \dots, X_n)}} = \phi_{(X_1, \dots, X_n)} = \prod_{i=1}^n \phi_{X_i} = \phi_{\mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}}.$$

Comme la fonction caractéristique caractérise la loi, $\mathbf{P}_{(X_1, \dots, X_n)} = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}$ ce qui signifie que les X_j sont indépendantes. \square

3.4 Espérance, variance, covariance, corrélation linéaire

Considérons un couple de v.a. (X, Y) . Pour toute fonction (mesurable) g et sous condition d'intégrabilité, l'espérance de $g(X, Y)$:

$$\mathbf{E}[g(X, Y)] = \int g \, d\mathbf{P}_{(X, Y)}$$

se calcule dans le cas discret comme

$$\mathbf{E}[g(X, Y)] = \sum_{i, j} g(x_i, y_j) p_{(X, Y)}(x_i, y_j)$$

et dans le cas continu comme

$$\mathbf{E}[g(X, Y)] = \int_{\mathbf{R}^2} g(x, y) f_{(X, Y)}(x, y) \, dx \, dy.$$

On voit ainsi facilement que :

Proposition 3.4. [Linéarité de l'espérance] Soient X, Y deux variables aléatoires et $\alpha, \beta \in \mathbf{R}$:

$$\mathbf{E}[\alpha X + \beta Y] = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y].$$

Remarque 3.3. Ceci se généralise naturellement au cas de d v.a. :

$$\mathbf{E} \left[\sum_{i=1}^d \alpha_i X_i \right] = \sum_{i=1}^d \alpha_i \mathbf{E}[X_i].$$

Définition 3.5. Soient X et Y deux v.a. réelles admettant des moments d'ordre 2.

La covariance de X et Y est la quantité :

$$\text{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])],$$

et le coefficient de corrélation linéaire de X et Y est la quantité :

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbf{V}[X]\mathbf{V}[Y]}}.$$

Alors que la variance d'une variable aléatoire mesure sa dispersion autour de sa moyenne, les covariance et coefficient de corrélation linéaire mesurent comment deux variables varient ensemble autour de leurs moyennes respectives et donc le lien qu'elles entretiennent mutuellement. Nous verrons que Corr est à valeurs dans $[-1, 1]$. En fait, plus $|\text{Corr}(X, Y)|$ est proche de 1, plus le lien entre X et Y est fort. Il ne faut pas être surpris par le fait qu'une corrélation linéaire proche de -1 reflète un lien fort entre les variables : penser au cas de $Y = -X$.

Proposition 3.5. On a :

1. $\text{Cov}(X, X) = \mathbf{V}[X]$;
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;

3. *Cov est linéaire en chacune des coordonnées :*

$$\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z) = \text{Cov}(Z, aX + bY);$$

4. $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$;

5. $\text{Cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$;

6. *si X et Y sont indépendantes, $\text{Cov}(X, Y) = 0$; on dit alors que X et Y sont décorréliées ;*

7. $\text{Corr}(X, Y) \in [-1, 1]$;

8. $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$.

Preuve : Les quatre premiers points sont évidents ou reposent sur des calculs immédiats ; le cinquième se démontre de façon analogue à la formule de décentrage de la variance vue dans le chapitre précédent. Esquissons la démonstration du point 6. dans le cas de deux variables discrètes (la démonstration dans le cas continue est analogue et laissée en exercice au lecteur).

On a :

$$\begin{aligned} \mathbf{E}[XY] &= \sum_i \sum_j x_i y_j p_{(X,Y)}(x_i, y_j) \\ &= \sum_i \sum_j x_i y_j p_X(x_i) p_Y(y_j) \quad (\text{par indépendance}) \\ &= \sum_i x_i p_X(x_i) \sum_j y_j p_Y(y_j) = \mathbf{E}[X]\mathbf{E}[Y] \end{aligned}$$

donc $\text{Cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 0$.

Pour démontrer le point 7., remarquons que pour tout $\lambda \in \mathbf{R}$, on a :

$$\begin{aligned} 0 \leq \mathbf{V}[X + \lambda Y] &= \mathbf{E} \left[(X + \lambda Y - \mathbf{E}[X + \lambda Y])^2 \right] \\ &= \mathbf{E} \left[(X - \mathbf{E}[X] + \lambda(Y - \mathbf{E}[Y]))^2 \right] \\ &= \mathbf{E} \left[(X - \mathbf{E}[X])^2 + 2\lambda (X - \mathbf{E}[X]) (Y - \mathbf{E}[Y]) + \lambda^2 (Y - \mathbf{E}[Y])^2 \right] \\ &= \mathbf{V}[X] + 2\lambda \text{Cov}(X, Y) + \lambda^2 \mathbf{V}[Y] \end{aligned}$$

et que par conséquent le discriminant du polynôme du second degré en λ apparaissant à la dernière ligne est négatif ou nul. Ainsi,

$$(2 \text{Cov}(X, Y))^2 - 4\mathbf{V}[X]\mathbf{V}[Y] \leq 0,$$

soit

$$\text{Cov}(X, Y)^2 \leq \mathbf{V}[X]\mathbf{V}[Y],$$

puis

$$\frac{\text{Cov}(X, Y)^2}{\mathbf{V}[X]\mathbf{V}[Y]} \leq 1,$$

d'où

$$|\text{Corr}(X, Y)| = \left| \frac{\text{Cov}(X, Y)}{\sqrt{\mathbf{V}[X]\mathbf{V}[Y]}} \right| \leq 1.$$

□

Exercice 3.2. Soit (X, Y) le couple de variables aléatoires définies par :

$$\mathbf{P}[X = 0, Y = 0] = 2\mathbf{P}[X = -1, Y = 1] = 2\mathbf{P}[X = 1, Y = 1] = \frac{1}{2}.$$

1. Calculer $\text{Cov}(X, Y)$.
2. Les v.a. X et Y sont-elles indépendantes ?

3.5 Sommes de variables aléatoires

Il n'est pas difficile de voir qu'en toute généralité :

$$\mathbf{V}[X + Y] = \mathbf{V}[X] + 2\text{Cov}(X, Y) + \mathbf{V}[Y]$$

et par conséquent la variance ne peut pas être linéaire comme l'est l'espérance (aussi puisqu'elle est homogène à un carré). Observons tout de même que si X et Y sont décorréliées (ou mieux, indépendantes), on obtient que

$$\mathbf{V}[X + Y] = \mathbf{V}[X] + \mathbf{V}[Y].$$

Plus généralement, on a :

Proposition 3.6. Si X_1, \dots, X_d sont des v.a. de carré intégrable et deux à deux non corrélées, alors pour tous $\alpha_1, \dots, \alpha_d \in \mathbf{R}$, on a :

$$\mathbf{V}\left[\sum_{i=1}^d \alpha_i X_i\right] = \sum_{i=1}^d \alpha_i^2 \mathbf{V}[X_i].$$

Preuve : Il suffit d'écrire que :

$$\begin{aligned} \mathbf{V}\left[\sum_{i=1}^d \alpha_i X_i\right] &= \mathbf{E}\left[\left(\sum_{i=1}^d \alpha_i (X_i - \mathbf{E}[X_i])\right)^2\right] \\ &= \sum_{1 \leq i, j \leq d} \alpha_i \alpha_j \mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])] \\ &= \sum_{1 \leq i, j \leq d} \alpha_i \alpha_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^d \alpha_i^2 \mathbf{V}[X_i] \end{aligned}$$

où l'on a utilisé dans la dernière égalité le fait que les X_i sont deux à deux non corrélées. \square

Exercice 3.3. [Inégalité de Bienaymé-Tchebychev] Montrer X_1, \dots, X_d sont des v.a. de carré intégrable et deux à deux non corrélées, elles vérifient l'inégalité de Bienaymé-Tchebychev :

$$\mathbf{P}\left[\left|\sum_{i=1}^n (X_i - \mathbf{E}[X_i])\right| \geq t\right] \leq \frac{1}{t^2} \sum_{i=1}^n \mathbf{V}[X_i].$$

Indication : on pourra utiliser l'inégalité de Tchebychev.

Le corollaire suivant jouera un rôle important en statistiques pour l'étude des estimateurs et l'analyse de leur qualité.

Corollaire 3.1. Soient X_1, \dots, X_n des v.a. indépendantes et identiquement distribuées (i.i.d.) de moyenne m et de variance $\sigma^2 < +\infty$. Posons $S_n = \sum_{i=1}^n X_i$ leur somme et $\bar{X}_n = \frac{S_n}{n}$ leur moyenne empirique. On a :

$$E[S_n] = nm \quad \text{et} \quad \mathbf{V}[S_n] = n\sigma^2$$

et par conséquent

$$E[\bar{X}_n] = m \quad \text{et} \quad \mathbf{V}[\bar{X}_n] = \frac{\sigma^2}{n}.$$

Les fonctions génératrices des moments et caractéristiques se révèlent être particulièrement efficaces pour étudier et caractériser la loi de la somme de variables aléatoires indépendantes comme le montrent les deux résultats suivants.

Théorème 3.4. Soient X_1, \dots, X_d des v.a. indépendantes. Alors,

$$L_{X_1+\dots+X_d}(t) = \prod_{i=1}^d L_{X_i}(t).$$

Preuve : On a :

$$\begin{aligned} L_{X_1+\dots+X_d}(t) &= \mathbf{E} \left[e^{t(X_1+\dots+X_d)} \right] \\ &= \mathbf{E} \left[\prod_{i=1}^d e^{tX_i} \right] \\ &= \prod_{i=1}^d \mathbf{E} \left[e^{tX_i} \right] \quad \text{par indépendance} \\ &= \prod_{i=1}^d L_{X_i}(t). \end{aligned}$$

□

On montre de façon analogue que :

Théorème 3.5. Soient X_1, \dots, X_d des v.a. indépendantes. Alors,

$$\phi_{X_1+\dots+X_d}(t) = \prod_{i=1}^d \phi_{X_i}(t).$$

Donnons un exemple de résultat non trivial (et important) dont la preuve est rendu facile par le théorème précédent. Il affirme que toute combinaison linéaire de lois normales indépendantes suit une loi normale.

Proposition 3.7. Soit X_i , $i = 1, \dots, d$, des v.a. indépendantes suivant respectivement la loi normale $\mathcal{N}(m_i, \sigma_i^2)$ et $\alpha_1, \dots, \alpha_n \in \mathbf{R}$.

Alors, $Y = \alpha_1 X_1 + \dots + \alpha_d X_d$ suit la loi normale de moyenne $\sum_{i=1}^d \alpha_i m_i$ et de variance $\sum_{i=1}^d \alpha_i^2 \sigma_i^2$.

CHAPITRE 3. COUPLES, N -UPLETS ET FAMILLES DE VARIABLES ALÉATOIRES

Preuve : Notons que $\alpha_i X_i$ suit la loi normale telle que $\mathbf{E}[\alpha_i X_i] = \alpha_i m_i$ et $\mathbf{V}[\alpha_i X_i] = \alpha_i^2 \sigma_i^2$. Ainsi, en utilisant le résultat de l'Exercice 2.4, on a pour tout t et tout i :

$$\phi_{\alpha_i X_i}(t) = e^{it\alpha_i m_i - \frac{\alpha_i^2 \sigma_i^2 t^2}{2}}.$$

Maintenant, les X_i étant indépendantes, le Théorème 3.5, assure que :

$$\begin{aligned} \phi_{\alpha_1 X_1 + \dots + \alpha_d X_d}(t) &= \prod_{i=1}^d \phi_{\alpha_i X_i}(t) = \prod_{i=1}^d e^{it\alpha_i m_i - \frac{\alpha_i^2 \sigma_i^2 t^2}{2}} \\ &= \exp\left(it \sum_{i=1}^d \alpha_i m_i - \frac{\sum_{i=1}^d \alpha_i^2 \sigma_i^2 t^2}{2}\right). \end{aligned}$$

On reconnaît alors la fonction caractéristique de la loi normale de moyenne $\sum_{i=1}^d \alpha_i m_i$ et de variance $\sum_{i=1}^d \alpha_i^2 \sigma_i^2$. Puisque la fonction caractéristique caractérise la loi, le résultat s'ensuit. \square

3.6 Lois normales multivariés, vecteurs gaussiens

Comme nous venons de le voir toute combinaison linéaire de lois normales indépendantes suit une loi normale. Cette observation guide la définition suivante.

Définition 3.6. On dit qu'un vecteur aléatoire X dans \mathbf{R}^d suit une loi normale (multidimensionnelle) si, pour tout vecteur $a \in \mathbf{R}^d$ la loi de $a \cdot X$ est une loi normale (unidimensionnelle).

Remarque 3.4. On note $\mathbf{E}[X] = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_d])^T$ l'espérance d'un vecteur aléatoire et

$$\Sigma = \mathbf{V}[X] = ((\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d})$$

sa matrice de variances/covariances. On notera $X \sim \mathcal{N}_d(m, \Sigma)$ si X suit la loi normale d -dimensionnelle d'espérance $m \in \mathbf{R}^d$ et de matrice de variances/covariances Σ .

Notons que si X est un vecteur aléatoire de \mathbf{R}^d , A est une matrice $n \times d$ et $b \in \mathbf{R}^n$ (déterministe), alors $Y = AX + b$ est un vecteur aléatoire dans \mathbf{R}^n et on a :

$$\mathbf{E}[Y] = A\mathbf{E}[X] + b \quad \text{et} \quad \mathbf{V}[Y] = A\mathbf{V}[X]A^T.$$

Notons aussi, que si les composantes d'un vecteur aléatoire sont indépendantes alors la matrice de variances/covariances est diagonale (il s'agit de l'identité lorsque les composantes sont réduites).

Le résultat suivant exprime la densité d'une loi normale multidimensionnelle (voir [8, Théorème 2.2]).

Théorème 3.6. Soit Σ une matrice définie positive $d \times d$, $m \in \mathbf{R}^d$ et $X \sim \mathcal{N}_d(m, \Sigma)$. Alors, X admet la densité suivante sur \mathbf{R}^d (par rapport à la mesure de Lebesgue d -dimensionnelle) :

$$x \mapsto \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(x - m)^T \Sigma^{-1}(x - m)\right).$$

Proposition 3.8. *Si $X = (X_1, \dots, X_d) \sim \mathcal{N}_d(m, \Sigma)$ alors sa fonction caractéristique est donnée pour tout $t \in \mathbf{R}^d$ par :*

$$\phi_X(t) = e^{it^T m - \frac{1}{2} t^T \Sigma t}.$$

Preuve : Notons que si puisque $X \sim \mathcal{N}_d(m, \Sigma)$, par définition, pour tout $a \in \mathbf{R}^d$, $t^T X = t_1 X_1 + \dots + t_d X_d$, suit une loi normale unidimensionnelle. Son espérance étant donnée par $t^T m$ et sa variance par $t^T \Sigma t$, on obtient par le résultat de l'Exercice 2.4 que :

$$\phi_{t^T X}(s) = e^{it^T m - \frac{1}{2} t^T \Sigma t s^2}.$$

Il suffit maintenant de remarquer que :

$$\phi_X(t) = \mathbf{E}[e^{it^T X}] = \mathbf{E}[e^{i \times 1 \times (t^T X)}] = \phi_{t^T X}(1) = e^{it^T m - \frac{1}{2} t^T \Sigma t}.$$

□

Proposition 3.9. *Si $X \sim \mathcal{N}_d(m, \Sigma)$, avec Σ définie positive, alors,*

$$\sqrt{\Sigma}^{-1}(X - m) \sim \mathcal{N}_d(0, I_d).$$

Preuve : Puisque Σ définie positive, il existe $A = \sqrt{\Sigma}$ telle que $\Sigma = AA^T$ et l'expression donnée à bien un sens. On montre ensuite que $Y = \sqrt{\Sigma}^{-1}(X - m)$ est bien distribué selon $\mathcal{N}_d(0, I_d)$ par un calcul de fonction caractéristique laissé au lecteur.

□

Proposition 3.10. *Soit $C = (X, Y)$ un vecteur gaussien (au sens de la Définition 3.6). Alors, X et Y sont indépendantes si, et seulement si, $\text{Cov}(X, Y) = 0$.*

Remarque 3.5.

On sera vigilant à ne pas confondre (X, Y) un vecteur (ou couple) gaussien (au sens de la Définition 3.6) et un vecteur (ou couple) de v.a. gaussiennes, sans quoi la conclusion de cette proposition est fautive (ce que l'on verra dans un exercice de TD).

Preuve :

Comme nous l'avons déjà vu le sens direct est vrai en toute généralité (quelque soient les lois de X et Y) contrairement au sens indirect. Pour le sens indirect, d'après le Théorème 3.3, il suffit de voir que, pour tous t_1, t_2 , $\phi_{(X, Y)}(t_1, t_2) = \phi_X(t_1)\phi_Y(t_2)$. Puisque, $\text{Cov}(X, Y) = 0$,

$$\mathbf{V}[(X, Y)] = \begin{pmatrix} \mathbf{V}[X] & 0 \\ 0 & \mathbf{V}[Y] \end{pmatrix}$$

et donc

$$\begin{aligned} \phi_{(X, Y)}(t_1, t_2) &= e^{i(t_1 \mathbf{E}[X] + t_2 \mathbf{E}[Y]) - \frac{1}{2}(t_1^2 \mathbf{V}[X] + t_2^2 \mathbf{V}[Y])} \\ &= e^{it_1 \mathbf{E}[X] - \frac{1}{2} t_1^2 \mathbf{V}[X]} e^{it_2 \mathbf{E}[Y] - \frac{1}{2} t_2^2 \mathbf{V}[Y]} = \phi_X(t_1)\phi_Y(t_2). \end{aligned}$$

□

3.7 Familles exponentielles de lois

Les *familles exponentielles de lois* forment des classes de lois paramétriques, de paramètres θ dans un ensemble de paramètres Θ , jouissant de propriétés communes en statistiques, en particuliers dans les théories des tests et de l'estimation. Ceci découle du fait que, par définition, leurs densités (cas continu) ou fonction de probabilité (cas discret) admettent une *écriture canonique* commune. Notons, qu'en général, $\theta = (\theta_1, \dots, \theta_k)$ est un élément de \mathbf{R}^k et ne contiendra, dans le contexte statistique, que les paramètres inconnus des lois (une partie des paramètres peut être supposée connue). Dans ce cours, nous nous restreindrons au cas continu (dominé par la mesure de Lebesgue λ) et discret (dominé par une mesure de comptage). Un exposé plus général peut être trouvé dans la Section 2.4 de [8] par exemple. Dans le cas continu, nous noterons $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbf{R}^k$ la densité de la loi lorsque le paramètre est θ ; il s'agit d'une fonction d'une variable réelle x . Dans le cas discret, nous noterons, avec un léger abus de notation permettant une plus grande uniformité, $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbf{R}^k$ (au lieu de $p(\cdot, \theta)$) la fonction de probabilité de la loi lorsque le paramètre est θ . La donnée des $f(\cdot, \theta)$, $\theta \in \Theta$, caractérise dans les deux cas la famille paramétrique de lois.

Définition 3.7. *Soit, avec les conventions précédentes, une famille paramétrique \mathcal{F} de lois caractérisée par $\{f(\cdot, \theta), \theta \in \Theta\}$ où $\Theta \subset \mathbf{R}^k$. On dit que \mathcal{F} est une famille exponentielle de lois, s'il existe des fonctions $a, b, c_1, \dots, c_k, d_1, \dots, d_k$ telles que pour tout $x \in \mathbf{R}$:*

$$f(x; \theta) = a(\theta)b(x) \exp \left(\sum_{i=1}^k c_i(\theta)d_i(x) \right).$$

Cette expression est appelée écriture canonique des densités (ou fonctions de probabilité dans le cas discret) de la famille de lois.

Remarque 3.6.

1. Dans le cas continu l'égalité

$$f(x; \theta) = a(\theta)b(x) \exp \left(\sum_{i=1}^k c_i(\theta)d_i(x) \right)$$

λ -p.p. suffit.

2. Insistons sur le fait que, dans la définition précédente k sera le nombre de paramètres inconnus dans le contexte statistique.
3. La séparation des variables x et θ dans l'écriture canonique montre que le support des lois d'une famille exponentielle ne peut pas dépendre des paramètres (inconnus) θ . Ainsi, les lois de Fréchet $\mathcal{F}re(\theta, 1, 1)$, $\theta \in \mathbf{R}$, (voir Section A.2.12) ne forment pas une famille exponentielle. Plus simplement, les lois uniformes $\mathcal{U}([0, \theta])$, $\theta \in \mathbf{R}$, ne forment pas une famille exponentielle. Les lois binomiales $\mathcal{B}in(\theta)$, $\theta = (n, p)$, ne forment pas une famille exponentielle mais si, comme pour nos applications statistiques, le paramètre n est connu, les lois binomiales $\mathcal{B}in(n, \theta)$, $\theta \in [0, 1]$, forment une famille exponentielle pour laquelle on a :

$$a(\theta) = (1 - \theta)^n, \quad b(x) = \binom{n}{x}, \quad c(\theta) = \ln \left(\frac{\theta}{1 - \theta} \right) \quad \text{et} \quad d(x) = x.$$

Donnons maintenant les exemples d'usages les plus courants de familles exponentielles. Pour cela, on précisera le(s) paramètre(s) inconnu(s) θ et l'espace des paramètres Θ et on rappellera dans chaque densités ou fonctions de probabilité de ces famille avant de les réécrire sous forme canonique. On mettra également en valeur, dans chaque cas, les fonctions d (ou d_1 et d_2) qui trouveront des applications en statistiques.

3.7.1 Famille des lois de Bernoulli

Le paramètre inconnu est $\theta = p \in \Theta = [0, 1]$. On a :

$$\begin{aligned} f(x; \theta) &= \theta \mathbf{1}_{x=1} + (1 - \theta) \mathbf{1}_{x=0} \\ &= (1 - \theta) \binom{1}{x} \mathbf{1}_{x \in \{0,1\}} \exp \left(\ln \left(\frac{\theta}{1 - \theta} \right) x \right) \end{aligned}$$

et donc $d(x) = x$.

3.7.2 Famille des lois binomiales avec n connu

Le paramètre inconnu est $\theta = p \in \Theta = [0, 1]$. On a :

$$\begin{aligned} f(x; \theta) &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} = (1 - \theta)^n \binom{n}{x} \left(\frac{\theta}{1 - \theta} \right)^x \\ &= (1 - \theta)^n \binom{n}{x} \mathbf{1}_{x \in \{0, \dots, n\}} \exp \left(\ln \left(\frac{\theta}{1 - \theta} \right) x \right) \end{aligned}$$

et donc $d(x) = x$.

3.7.3 Famille des lois géométriques

Le paramètre inconnu est $\theta = p \in \Theta = [0, 1]$. On a :

$$\begin{aligned} f(x; \theta) &= \mathbf{1}_{x \in \mathbf{N}^*} (1 - \theta)^{x-1} \theta \\ &= \frac{\theta}{1 - \theta} \mathbf{1}_{x \in \mathbf{N}^*} \exp (\ln (1 - \theta) x) \end{aligned}$$

et donc $d(x) = x$.

3.7.4 Famille des lois de Poisson

Le paramètre inconnu est $\theta = \lambda \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= e^{-\theta} \frac{\theta^x}{x!} \\ &= e^{-\theta} \frac{\mathbf{1}_{x \in \mathbf{N}}}{x!} \exp (\ln (\theta) x) \end{aligned}$$

et donc $d(x) = x$.

3.7.5 Famille des lois binomiales négatives avec r connu

Le paramètre inconnu est $\theta = p \in \Theta = [0, 1]$. On a (voir Appendice A) :

$$f(x; \theta) = \frac{x^{r-1} \theta^r e^{-\theta x}}{\Gamma(r)} \mathbf{1}_{x>0}$$

et donc $d(x) = x$.

3.7.6 Famille des lois exponentielles

Le paramètre inconnu est $\theta = \lambda \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \theta e^{-\theta x} \mathbf{1}_{[0; +\infty[}(x) \\ &= \theta \mathbf{1}_{[0; +\infty[}(x) \exp(-\theta x) \end{aligned}$$

et donc $d(x) = x$.

3.7.7 Famille des lois gamma avec r connu

Le paramètre inconnu est $\theta = \lambda \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{x^{r-1} e^{-\frac{x}{\theta}}}{\Gamma(r) \theta^r} \mathbf{1}_{[0; +\infty[}(x) \\ &= \frac{1}{\Gamma(r) \theta^r} x^{r-1} \mathbf{1}_{[0; +\infty[}(x) \exp(-\theta^{-1} x) \end{aligned}$$

et donc $d(x) = x$.

3.7.8 Famille des lois beta avec p connu

Le paramètre inconnu est $\theta = q \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{x^{p-1} (1-x)^{\theta-1}}{B(p, \theta)} \mathbf{1}_{0<x<1} \\ &= \frac{1}{B(p, \theta)} x^{p-1} \mathbf{1}_{0<x<1} \exp((\theta-1) \ln(x-1)) \end{aligned}$$

et donc $d(x) = \ln(x-1)$.

3.7.9 Famille des lois beta avec q connu

Le paramètre inconnu est $\theta = p \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{x^{\theta-1} (1-x)^{q-1}}{B(\theta, q)} \mathbf{1}_{0<x<1} \\ &= \frac{1}{B(\theta, q)} (1-x)^{q-1} \mathbf{1}_{0<x<1} \exp((\theta-1) \ln(x)) \end{aligned}$$

et donc $d(x) = \ln(x)$.

3.7.10 Famille des lois beta

Le couple de paramètres inconnus est $\theta = (\theta_1, \theta_2) = (p, q) \in \Theta = (\mathbf{R}_+^*)^2$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{x^{\theta_1-1}(1-x)^{\theta_2-1}}{B(\theta_1, \theta_2)} \mathbf{1}_{0 < x < 1} \\ &= \frac{1}{B(\theta_1, \theta_2)} \mathbf{1}_{0 < x < 1} \exp((\theta_1 - 1) \ln(x) + (\theta_2 - 1) \ln(1 - x)) \end{aligned}$$

et donc $d_1(x) = \ln(x)$ et $d_2(x) = \ln(1 - x)$.

3.7.11 Famille des lois normales avec m connu

Le paramètre inconnu est $\theta = \sigma^2 \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-m)^2}{2\theta}} \\ &= \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{1}{2\theta}(x-m)^2\right) \end{aligned}$$

et donc $d(x) = (x - m)^2$.

3.7.12 Famille des lois normales avec σ^2 connu

Le paramètre inconnu est $\theta = m \in \Theta = \mathbf{R}$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{1}{\sqrt{2\pi^2\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2} + \frac{\theta x}{\sigma^2} - \frac{\theta^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \exp\left(\frac{\theta}{\sigma^2}x\right) \end{aligned}$$

et donc $d(x) = x$.

3.7.13 Famille des lois normales

Le couple de paramètres inconnus est $\theta = (\theta_1, \theta_2) = (m, \sigma^2) \in \Theta = \mathbf{R} \times \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{1}{\sqrt{2\pi^2\theta_2}} e^{-\frac{(x-\theta_1)^2}{2\theta_2}} \\ &= \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{x^2}{2\theta_2} + \frac{\theta_1 x}{\theta_2} - \frac{\theta_1^2}{2\theta_2}} \\ &= \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{\theta_1^2}{2\theta_2}} \exp\left(-\frac{1}{2\theta_2}x^2 + \frac{\theta_1}{\theta_2}x\right) \end{aligned}$$

et donc $d_1(x) = x^2$ et $d_2(x) = x$.

3.7.14 Famille des lois de Pareto avec a connu

Le paramètre inconnu est $\theta \in \Theta = \mathbf{R}_+^*$. On a :

$$\begin{aligned} f(x; \theta) &= \frac{\theta}{a} \left(\frac{a}{x}\right)^{\theta+1} \mathbf{1}_{[a; +\infty[}(x) \\ &= \theta a^\theta \frac{\mathbf{1}_{[a; +\infty[}(x)}{a} x^{-(\theta+1)} \\ &= \theta a^\theta \frac{\mathbf{1}_{[a; +\infty[}(x)}{a} \exp(-(\theta+1) \ln(x)) \end{aligned}$$

et donc $d(x) = \ln(x)$.

Chapitre 4

Modes de convergence de variables aléatoires, théorèmes limites

4.1 Modes de convergence de variables aléatoires

Dans cette section, on introduit succinctement les modes de convergence de suites de variables aléatoires et les principaux liens qu'ils entretiennent. Pour plus de détails, le lecteur est renvoyé au Chapitre 5 de [4] et aux Chapitres 10 et 14 de [13].

4.1.1 Convergence presque sûre

Définition 4.1. On dit qu'une suite $(X_n)_{n \in \mathbf{N}}$ de v.a. réelles sur $(\Omega, \mathcal{A}, \mathbf{P})$ converge presque sûrement (p.s) vers X si $\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) \neq X(\omega)\}$ est \mathbf{P} -négligeable. On note alors $X_n \xrightarrow{p.s.} X$.

4.1.2 Convergence en probabilité

Définition 4.2. On dit qu'une suite $(X_n)_{n \in \mathbf{N}}$ de v.a. réelles sur $(\Omega, \mathcal{A}, \mathbf{P})$ converge en probabilité vers X si pour tout $\varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} \mathbf{P}[|X_n - X| \geq \varepsilon] = 0.$$

On note alors $X_n \xrightarrow{\mathbf{P}} X$.

4.1.3 Convergence dans L^p

Définition 4.3. Soit $p > 0$. On dit qu'une suite $(X_n)_{n \in \mathbf{N}}$ de v.a. réelles admettant un moment d'ordre p sur $(\Omega, \mathcal{A}, \mathbf{P})$ converge dans L^p vers X (admettant un moment d'ordre p) si :

$$\lim_{n \rightarrow +\infty} \mathbf{E}[|X_n - X|^p] = 0,$$

soit encore si :

$$\lim_{n \rightarrow +\infty} \|X_n - X\|_p = 0,$$

On note alors $X_n \xrightarrow{L^p} X$.

Définition 4.4. Si $X_n \xrightarrow{L^2} X$, on dit que $(X_n)_{n \in \mathbf{N}}$ converge dans en moyenne quadratique vers X . On note alors $X_n \xrightarrow{m.q.} X$.

4.1.4 Convergence en loi

Rappelons que deux v.a. X et Y ont même loi si pour toute fonction continue bornée φ :

$$\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi(Y)],$$

ou encore si leurs fonctions caractéristiques sont égales ou encore si leurs fonctions de répartition sont égales (puisqu'elle caractérisent la loi).

Définition 4.5. On dit qu'une suite $(X_n)_{n \in \mathbf{N}}$ de v.a. réelles sur $(\Omega, \mathcal{A}, \mathbf{P})$ converge en loi vers X si pour toute fonction continue bornée φ :

$$\lim_{n \rightarrow +\infty} \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X)].$$

On note alors $X_n \xrightarrow{\mathcal{L}} X$.

Le résultat suivant donne des caractérisations pratiques de la convergence en loi.

Théorème 4.1. Les assertions suivantes sont équivalentes :

1. $X_n \xrightarrow{\mathcal{L}} X$;
2. F_{X_n} converge vers F_X en tout point de continuité de F_X ;
3. ϕ_{X_n} converge simplement vers ϕ_X sur \mathbf{R} .

Si les X_n et X sont discrètes, $X_n \xrightarrow{\mathcal{L}} X$ si, et seulement si, pour tout x dans le support de X ,

$$\lim_{n \rightarrow \infty} \mathbf{P}[X_n = x] = \mathbf{P}[X = x].$$

4.1.5 Liens entre les modes de convergence

Le schéma suivant résume les liens principaux entre les modes de convergence de variables aléatoires. Des réciproques partielles peuvent être établies dans certains cas particuliers mais ne seront pas utiles pour nos propos.

$$\begin{array}{ccccc} L^p & \xRightarrow{p>q} & L^q & & \\ & & \Downarrow & & \\ \text{p.s.} & \xRightarrow{\quad} & \mathbf{P} & \xRightarrow{\quad} & \mathcal{L} \end{array}$$

4.2 Théorèmes limites

Pour $(X_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées (v.a.i.i.d.) sur $(\Omega, \mathcal{A}, \mathbf{P})$, on notera :

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

la moyenne empirique des n premières de ces v.a..

4.2.1 Lois des grands nombres

Théorème 4.2 (Loi faible des grands nombres). *Soit $(X_n)_{n \in \mathbf{N}}$ une suite de v.a.i.i.d. de même loi que X (loi mère). Si $\mathbf{E}[|X|] < +\infty$, alors $\bar{X}_n \xrightarrow{\mathbf{P}} \mathbf{E}[X]$.*

Ce résultat montre que pour toute suite de v.a.i.i.d. intégrables, la moyenne empirique d'un échantillon de n d'entre elles tend à s'approcher de la moyenne d'une de ces variables individuelles. Une version plus forte de ce résultat est donnée dans le paragraphe suivant.

Théorème 4.3 (Loi forte des grands nombres). *Soit $(X_n)_{n \in \mathbf{N}}$ une suite de v.a.i.i.d. de même loi que X . Alors, $\bar{X}_n \xrightarrow{p.s.} \mathbf{E}[X]$ si, et seulement si, $\mathbf{E}[|X|] < +\infty$.*

4.2.2 Théorème Central Limite

Notons que, si $(X_n)_{n \in \mathbf{N}}$ une suite de v.a.i.i.d. de moyenne m et de variance σ^2 , il est clair que

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}}$$

est centrée et réduite puisque la transformation appliquée à \bar{X}_n consiste précisément en la centrer et réduire. Par ailleurs, la loi forte des grands nombres indique que $\bar{X}_n - m$ converge presque sûrement vers 0. Le résultat suivant précise les fluctuations \bar{X}_n autour de m en identifiant la loi du quotient précédent. Il induit une forme d'universalité de celles-ci, fondamentale pour l'étude statistique, et justifie l'intérêt particulier porté aux lois normales et aux échantillons gaussiens.

Théorème 4.4 (Théorème Central Limite (TCL)). *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de v.a.i.i.d. de moyenne m et de variance $0 < \sigma^2 < +\infty$, alors*

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1).$$

Preuve : Quitte à remplacer X_i par $(X_i - m)/\sigma$, on peut supposer que $m = 0$ et $\sigma = 1$. Rappelons que la fonction caractéristique de $\mathcal{N}(0, 1)$ est donnée par $t \mapsto e^{-\frac{t^2}{2}}$. Par le Théorème 4.1, il suffit de voir que pour tout $t \in \mathbf{R}$:

$$\lim_{n \rightarrow +\infty} \phi_{\sqrt{n}\bar{X}_n}(t) = e^{-\frac{t^2}{2}}.$$

Puisque les X_i sont i.i.d., on a :

$$\begin{aligned} \phi_{\sqrt{n}\bar{X}_n}(t) &= \mathbf{E} \left[e^{it\sqrt{n}\bar{X}_n} \right] = \mathbf{E} \left[e^{i\frac{t}{\sqrt{n}} \sum_{j=1}^n X_j} \right] = \mathbf{E} \left[\prod_{j=1}^n e^{i\frac{t}{\sqrt{n}} X_j} \right] \\ &= \prod_{j=1}^n \mathbf{E} \left[e^{i\frac{t}{\sqrt{n}} X_j} \right] \quad (\text{par indépendance}) \\ &= \mathbf{E} \left[e^{i\frac{t}{\sqrt{n}} X_1} \right]^n \quad (\text{par identique distribution}) \\ &= \left(\phi_X \left(\frac{t}{\sqrt{n}} \right) \right)^n. \end{aligned}$$

CHAPITRE 4. MODES DE CONVERGENCE DE VARIABLES ALÉATOIRES,
THÉORÈMES LIMITES

Comme X admet un moment d'ordre 2, la Proposition 2.12 implique que $\phi'_X(0) = 0$ et $\phi''_X(0) = -1$ et donc, lorsque $u \rightarrow 0$:

$$\phi_X(u) = 1 - \frac{u^2}{2} + o(u^2).$$

Ainsi, pour $n \rightarrow +\infty$

$$\phi_{\sqrt{n}\bar{X}_n} = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n = e^{-\frac{t^2}{2}} + o(1)$$

ce qui termine la preuve. □

Théorème 4.5 (Théorème de Moivre-Laplace). *Soit $(B_n)_{n \in \mathbf{N}^*}$ une suite de v.a. indépendantes avec $B_n \sim \mathcal{Bin}(n; p)$, $p \in]0; 1[$.*

Alors,

$$\frac{B_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z$$

où $Z \sim \mathcal{N}(0; 1)$.

Preuve : Exercice!

Indication : appliquer le TCL à une suite de v.a.i.i.d. bien choisie. □

Remarque 4.1. Dans la pratique, on considère que l'approximation fournie par ce théorème est bonne si $n \geq 30$, $p \geq 0,1$ et $np > 15$.

Chapitre 5

Principes fondamentaux de l'échantillonnage

5.1 Généralités et approche empirique

La statistique a pour objet l'étude de phénomènes ou caractères pouvant varier d'un individu à l'autre dans une population, variation que l'on attribue à un certain hasard sur lequel on souhaite acquérir de l'information, par exemple, en estimant des paramètres ou la loi. L'idée fondamentale est d'obtenir des observations répétées du caractère étudié. Il s'agit d'une approche *empirique*, laquelle consiste en considérer que la connaissance s'acquiert par l'accumulation d'observations, allant du concret à l'abstrait. A titre d'exemple, s'il est possible d'observer un caractère numérique chez un grand nombre d'individus, disons que l'on dispose d'observations x_1, \dots, x_n , la loi des grands nombres (voir Section 4.2.1) donne un moyen simple d'estimer l'espérance de la v.a. régissant ce caractère par la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ sous condition d'intégrabilité et d'indépendance. Cette condition d'indépendance est généralement invérifiable en pratique et même invérifiée. Elle reste pourtant une hypothèse importante permettant de mener à bien l'étude théorique que nous ferons. Discutons brièvement, d'un cadre courant pour lequel cette hypothèse fournit une bonne approximation et est donc tout à fait raisonnable. Lors d'un sondage dans une grande population de taille N , on peut décider de prélever un échantillon *aléatoire et simple*. Ceci consiste à choisir (sans remise) n individus dans la population de façon uniforme. De manière équivalente, on choisit uniformément un individu parmi les N , puis un second parmi les $N - 1$ restants, etc, jusqu'au n^{e} individu sélectionné parmi les $N - n + 1$ restant. Bien entendu, les observations correspondant à ces individus x_1, \dots, x_n ne sont pas indépendantes. On pourrait essayer de contourner cette difficulté en effectuant un tirage avec remise avec des tirages effectivement indépendants, lequel perd de l'efficacité puisque le risque de sélectionner plusieurs fois le même individu est présent. Notons maintenant que si le *taux de sondage* $\frac{n}{N}$ est faible, disons inférieur à 5%, les sondages avec ou sans remise sont moralement proches et l'utilisation d'un échantillon aléatoire et simple avec un tel taux de sondage donne des approximations correctes.

Définition 5.1. *On appelle échantillon aléatoire de taille n toute famille (X_1, \dots, X_n) de v.a.i.i.d.. Si $X_1 \sim \mathcal{L}$, la loi \mathcal{L} est appelée loi mère de l'échantillon.*

Insistons sur le fait qu'un échantillon aléatoire est constitué de v.a. X_1, \dots, X_n . Nous

noterons x_1, \dots, x_n une réalisation d'un tel échantillon qui correspond à des observations « réelles » de ces variables aléatoires.

Définition 5.2. Soit (X_1, \dots, X_n) un échantillon aléatoire de taille n . On appelle statistique toute fonction de X_1, \dots, X_n .

Supposons que X_1, \dots, X_n soient des v.a. réelles. La quantité

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

est une statistique appelée *moyenne empirique*.

De même, la quantité

$$\tilde{S}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2$$

est une statistique appelée *variance empirique*. On définit de manière analogue le *moment empirique d'ordre r* et le *moment empirique centré d'ordre r* par

$$M_r = M_{r,n} = \frac{1}{n} \sum_{k=1}^n X_k^r \quad \text{et} \quad M'_r = M'_{r,n} = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^r.$$

Il découle directement de l'indépendance des v.a. X_1, \dots, X_n et de calculs élémentaires que

$$E[M_r] = E[X_1^r]$$

pourvu que ces moments existent. En particulier, on a :

Proposition 5.1. Supposons que la loi mère soit de moyenne m et de variance σ^2 .

Alors,

$$\mathbf{E}[\bar{X}_n] = m, \quad \mathbf{V}[\bar{X}_n] = \frac{\sigma^2}{n},$$

et

$$\mathbf{E}[\tilde{S}_n^2] = \frac{n-1}{n} \sigma^2.$$

Observons que l'espérance de la moyenne empirique est la moyenne de la loi mère et que sa variance tend vers 0 lorsque n tend vers l'infini. Cette deuxième observation montre que l'estimation de la moyenne par la moyenne empirique est de plus en plus précise et fiable quand la taille de l'échantillon augmente. Observons également que l'espérance de la variance empirique n'est pas la variance de la loi mère (voir notion de biais). Toutefois, l'espérance de la variance empirique tend vers la variance de la loi mère quand n tend vers $+\infty$; on dit que cet estimateur est *asymptotiquement sans biais*. Il est possible de « corriger » cet estimateur en un estimateur sans biais. Pour cela, on définit la *variance (empirique) corrigée* par

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

pour laquelle on a :

$$\mathbf{E}[S_n^2] = \sigma^2.$$

On préférera donc l'utilisation S_n^2 à celle de \tilde{S}_n^2 . Ces deux notions conduisent naturellement à celles d'*écart-type empirique* et *écart-type empirique corrigé* en en prenant la racine carrée.

Notons que les statistiques que nous venons d'introduire présupposent l'existence de moments pour la loi mère et que ceux-ci peuvent ne pas exister (par exemple pour une loi de Cauchy). Une v.a. ou une statistique peut également être *fonctionnelle* c'est-à-dire prendre pour valeurs des fonctions. Un exemple naturel, ayant pour but l'estimation de la fonction de répartition et valide même si la loi mère n'a pas de moment est la *fonction de répartition empirique* dont la valeur en $x \in \mathbf{R}$ est définie par :

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \leq x}.$$

Elle représente la proportion de v.a. dans l'échantillon prenant une valeur inférieure ou égale à x .

5.2 Statistiques d'ordre

Définition 5.3. Soit (X_1, \dots, X_n) un échantillon aléatoire. Considérons une permutation $X_{(1)}, \dots, X_{(n)}$ des X_j telle que $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

On appelle statistique d'ordre k la statistique $X_{(k)}$.

En particulier,

$$X_{(1)} = \min(X_1, \dots, X_n) \quad \text{et} \quad X_{(n)} = \max(X_1, \dots, X_n).$$

Remarquons que les statistiques d'ordre sont exactement les points de discontinuité de la fonction de répartition empirique.

La fonction de répartition de $X_{(k)}$ s'exprime de façon explicite comme suit :

$$\begin{aligned} F_{X_{(k)}}(x) &= \mathbf{P} [X_{(k)} \leq x] \\ &= \mathbf{P} [\text{au moins } k \text{ des } X_j \text{ sont } \leq x] \\ &= \sum_{l=k}^n \mathbf{P} [\text{exactement } l \text{ des } X_j \text{ sont } \leq x] \\ &= \sum_{l=k}^n \binom{n}{l} \mathbf{P} [X_1 \leq x]^l \mathbf{P} [X_1 > x]^{n-l} \quad (X_j \text{ i.i.d.}) \\ &= \sum_{l=k}^n \binom{n}{l} F_{X_1}(x)^l (1 - F_{X_1}(x))^{n-l}. \end{aligned}$$

En particulier,

$$F_{X_{(1)}}(x) = 1 - (1 - F_{X_1}(x))^n \quad \text{et} \quad F_{X_{(n)}}(x) = (F_{X_1}(x))^n.$$

Lorsque la loi mère est à densité, les formules précédentes permettent, en dérivant, de déduire facilement les densités des statistiques d'ordre.

Exercice 5.1. Montrer que si (X_1, X_2) est un couple de v.a. indépendantes de loi continue de densité f , alors $(X_{(1)}, X_{(2)})$ admet pour densité jointe :

$$f_{(X_{(1)}, X_{(2)})}(x_{(1)}, x_{(2)}) = 2f(x_{(1)})f(x_{(2)})\mathbf{1}_{x_{(1)} \leq x_{(2)}}.$$

Montrer, plus généralement, que si (X_1, \dots, X_n) est un échantillon d'une loi continue de densité f , alors $(X_{(1)}, \dots, X_{(n)})$ admet pour densité jointe :

$$f_{(X_{(1)}, \dots, X_{(n)})}(x_{(1)}, \dots, x_{(n)}) = n! \prod_{k=1}^n f(x_{(k)}) \mathbf{1}_{x_{(1)} \leq \dots \leq x_{(n)}}.$$

Indication : On pourra utiliser que, par la Proposition 2.8, il suffit de voir que pour toute fonction mesurable positive φ :

$$\mathbf{E}[\varphi(X_{(1)}, \dots, X_{(n)})] = \int_{\mathbf{R}^n} \varphi(x_{(1)}, \dots, x_{(n)}) n! \prod_{k=1}^n f(x_{(k)}) \mathbf{1}_{x_{(1)} \leq \dots \leq x_{(n)}} dx_{(1)} \dots dx_{(n)}$$

et penser à sommer sur les permutations des indices ordonnant comme il faut les variables.

5.3 Cas des lois mères gaussiennes

Dans cette section, on considère un échantillon (X_1, \dots, X_n) de v.a.i.i.d. de loi mère gaussienne $\mathcal{N}(m, \sigma^2)$. Une application directe de la Proposition 3.7 montre alors que

$$\bar{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right).$$

Le fait que la loi de \bar{X}_n soit accessible facilement dans le cas gaussien est remarquable. Pour une loi mère générique, ce n'est en général pas le cas. Il existe pourtant d'autres lois pour lesquelles la loi de la moyenne empirique est accessible simplement (Bernoulli exponentielle ou Poisson par exemple). Intéressons nous maintenant à la loi de la statistique S_n^2 dans le cas gaussien. Celle-ci s'identifie, après avoir introduit la loi du Khi-2 (voir Section A.2.13), en utilisant le théorème suivant, fondamental pour l'étude d'échantillons gaussiens. Sa démonstration n'est pas l'un des objectifs principaux de ce cours mais indiquée dans un soucis de complétude.

Théorème 5.1 (Théorème de Cochran (simplifié)). *Soit $Y = (Y_1, \dots, Y_n)^T \sim \mathcal{N}(0, I_n)$ et F un sous-espace vectoriel de \mathbf{R}^n de dimension d . Soient π_F et π_{F^\perp} les projections orthogonales sur F et F^\perp respectivement.*

Alors, projections orthogonales $\pi_F Y$ et $\pi_{F^\perp} Y$ de Y sur ces sous espaces sont gaussiennes indépendantes de lois $\mathcal{N}(0, \pi_F)$ et $\mathcal{N}(0, \pi_{F^\perp})$ respectivement et on a $\|\pi_F Y\|^2 \sim \chi^2(d)$ et $\|\pi_{F^\perp} Y\|^2 \sim \chi^2(n-d)$.

Remarque 5.1.

Ce théorème se généralise au cas d'une décomposition de \mathbf{R}^n en p sous-espaces vectoriels orthogonaux. La preuve est alors un peu plus lourde à écrire mais suit la même démarche.

Preuve : Soient (e_1, \dots, e_d) une base orthonormée de F et (e_{d+1}, \dots, e_n) une base orthonormée de F^\perp . Alors, $e = (e_1, \dots, e_n)$ est une base orthonormée de \mathbf{R}^n et la matrice de passage P permettant d'exprimer les coordonnées dans cette nouvelle base à partir des coordonnées dans la base canonique est orthonormale ($P^\perp = P^{-1}$). Notons $I_{n,d}$ la matrice diagonale dont les d premiers coefficients diagonaux sont égaux à 1 et les autres à 0. Notons aussi $J_{n,d} = I_n - I_{n,d}$. Les projections π_F et π_{F^\perp} s'expriment alors simplement dans la base e :

$$\pi_F = P I_{n,d} P^T \quad \text{et} \quad \pi_{F^\perp} = P J_{n,d} P^T.$$

Alors, le vecteur $Z = P^T Y$ des coordonnées de Y dans la base e est gaussien centré réduit (sa matrice de covariance est $PI_n P^T = I_n$). Il s'ensuit que $I_{n,d}Z = (Z_1, \dots, Z_d, 0, \dots, 0)^T$ et $J_{n,d}Z = (0, \dots, 0, Z_{d+1}, \dots, Z_n)^T$ sont gaussiens indépendants de lois $\mathcal{N}(0, I_{n,d})$ et $\mathcal{N}(0, J_{n,d})$ respectivement et que l'on a $\|I_{n,d}Z\|^2 = \sum_{k=1}^d Z_k^2 \sim \chi^2(d)$ et $\|J_{n,d}Z\|^2 = \sum_{k=d+1}^n Z_k^2 \sim \chi^2(n-d)$.

Pour conclure, il reste à remarquer, d'une part, que $\pi_F Y = PI_{n,d}Z$ et $\pi_{F^\perp} Y = PJ_{n,d}Z$ sont gaussiens centrés de matrices de covariance respectives $PI_{n,d}P^T = \pi_F$ et $PJ_{n,d}P^T = \pi_{F^\perp}$ et d'autre part que, puisqu'une transformation orthogonale préserve la norme :

$$\|\pi_F Y\|^2 = \|I_{n,d}Z\|^2 \sim \chi^2(d) \quad \text{et} \quad \|\pi_{F^\perp} Y\|^2 = \|J_{n,d}Z\|^2 \sim \chi^2(n-d).$$

□

Avant la lecture du prochain résultat, un lecteur non familier avec la loi de Student est invité à consulter la Section A.21.

Théorème 5.2. Soit (X_1, \dots, X_n) un échantillon de v.a.i.i.d. de loi mère gaussienne $\mathcal{N}(m, \sigma^2)$.

Alors, \bar{X}_n et S_n^2 sont indépendantes et on a :

$$\bar{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right),$$

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1),$$

et

$$\frac{\bar{X}_n - m}{S_n/\sqrt{n}} \sim \mathcal{T}(n-1).$$

Remarque 5.2. On voit, en particulier, que dans le cas d'un échantillon de loi mère gaussienne $\mathcal{N}(m, \sigma^2)$, on a

$$\mathbf{E}[S_n^2] = \sigma^2 \quad \text{et} \quad \mathbf{V}[S_n^2] = \frac{2\sigma^4}{n-1}.$$

Preuve : Nous avons déjà vu que

$$\bar{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right).$$

Pour montrer que \bar{X}_n et S_n^2 sont indépendantes et identifier la loi de $\frac{n-1}{\sigma^2} S_n^2$, nous allons utiliser le Théorème de Cochran 5.1. Pour tout $k \in \{1, \dots, n\}$, posons $Y_k = \sigma^{-1}(X_k - m)$. Alors, $Y = (Y_1, \dots, Y_n)^T \sim \mathcal{N}(0, I_n)$ et nous pouvons appliquer le Théorème de Cochran avec le sous-espace vectoriel de \mathbf{R}^n (de dimension 1) $F = \text{vect}(\mathbf{1}_n)$ où $\mathbf{1}_n = (1, \dots, 1)$. Notons que $\mathbf{1}_n \cdot Y = \bar{Y}_n = \sum_{k=0}^n Y_k$ et que $\bar{Y}_n \mathbf{1}_n \in F$. Notons également que $(Y - \bar{Y}_n \mathbf{1}_n) \cdot \mathbf{1}_n = 0$ donc $Y - \bar{Y}_n \mathbf{1}_n \in F^\perp$. Ainsi, $\pi_F Y = \bar{Y}_n \mathbf{1}_n$ et $\pi_{F^\perp} Y = Y - \bar{Y}_n \mathbf{1}_n$. Donc, par le Théorème de Cochran 5.1, $\|Y - \bar{Y}_n \mathbf{1}_n\|^2 \sim \chi^2(n-1)$ et est indépendante de \bar{Y}_n . Or, on a $\bar{X}_n = \bar{Y}_n + m$ et

$$\|Y - \bar{Y}_n \mathbf{1}_n\|^2 = \sum_{k=1}^n (Y_k - \bar{Y}_n)^2 = \sum_{k=1}^n \left(\frac{X_k - m}{\sigma} - \frac{\bar{X}_n - m}{\sigma} \right)^2 = \sum_{k=1}^n \left(\frac{X_k - \bar{X}_n}{\sigma} \right)^2 = \frac{n-1}{\sigma^2} S_n^2.$$

Ainsi, \bar{X}_n et S_n^2 sont indépendantes et :

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1).$$

Il reste à voir que :

$$\frac{\bar{X}_n - m}{S_n/\sqrt{n}} \sim \mathcal{T}(n-1).$$

Le reste du théorème étant acquis, ce dernier point découle directement de la définition de la loi de Student (Définition A.21) puisque

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{et} \quad \frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

sont indépendantes et

$$\frac{\frac{\bar{X}_n - m}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{n-1}{\sigma^2} S_n^2}{n-1}}} = \frac{\bar{X}_n - m}{S_n/\sqrt{n}}.$$

□

Terminons cette section par un résultat sur le rapport de variances d'échantillons gaussiens indépendants découlant directement du Théorème 5.2 et de la Définition A.22.

Proposition 5.2. *Soient $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_m)$ deux échantillons gaussiens indépendants ayant des lois mères de même variance (les moyennes peuvent être différentes). Soient S_X^2 et S_Y^2 leurs variances corrigées respectives.*

Alors,

$$\frac{S_X^2}{S_Y^2} \sim \mathcal{F}(n-1, m-1).$$

Preuve : Grâce au Théorème 5.2, on a :

$$\frac{n-1}{\sigma} S_X^2 \sim \chi^2(n-1) \quad \text{et} \quad \frac{m-1}{\sigma} S_Y^2 \sim \chi^2(m-1)$$

et ses v.a. sont indépendantes puisque les échantillons X et Y le sont. □

Chapitre 6

Estimation paramétrique ponctuelle

6.1 Cadre de l'estimation paramétrique ponctuelle

L'objectif de l'*estimation paramétrique ponctuelle* est de fournir une valeur « plausible » (une *estimation*) de certaines caractéristiques d'une loi inconnue, ou « partiellement inconnue » régissant un phénomène aléatoire. Elle est à distinguer de l'*estimation par intervalle de confiance* que nous aborderons au Chapitre 8. Dans ce cours, même si des généralisations restent possibles, nous nous restreindrons au cadre dans lequel l'estimation est obtenue à partir d'échantillons i.i.d. (X_1, \dots, X_n) que l'on choisira de taille suffisamment importante (disons $n \geq 30$) afin de fournir des estimations convenables. On notera qu'un tel échantillon, noté avec des lettres majuscules, est constitué de variables aléatoires et que la valeur numérique de l'estimation sera *in fine* calculée à partir d'une observation (x_1, \dots, x_n) , notée en minuscules, qui n'est autre qu'une réalisation particulière de l'échantillon observée « sur le terrain », autrement dit, la répétition de n observations indépendantes de la loi mère. Nous indiquons que la loi mère de l'échantillon est inconnue (ou l'est partiellement). Nous sommes donc amenés à considérer des familles de lois \mathbf{P}_θ indexées par les paramètres inconnus θ prenant leurs valeurs dans un *espace de paramètres* Θ . Ceci traduit un certain *a priori* sur la loi mère. L'*ensemble des observations* \mathcal{X} est l'union sur tous les $\theta \in \Theta$ des supports des lois \mathbf{P}_θ . On supposera nos modèles identifiables c'est-à-dire tels que $\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2}$ dès que $\theta_1 \neq \theta_2$.

Notons qu'il est également possible de fournir des estimations lorsque la loi mère ne fait pas partie d'une famille paramétrable de lois lorsque, comme souvent dans la réalité, l'information sur cette loi mère est plus vague ; il s'agit du cadre de l'estimation *non paramétrique* et de l'estimation *fonctionnelle*. Même si l'étude est dans ce cas liée aux concepts développés dans ce cours, nous ne la traiterons pas ici. Le lecteur intéressé par cette question est renvoyé vers [7, Chapitre 8].

Exemple 6.1. Si la loi mère est supposée $\mathcal{B}er(p)$, $\theta = p \in \Theta = [0, 1]$ et $\mathcal{X} = \{0, 1\}$.

Exemple 6.2. Si la loi mère est supposée gaussienne (non dégénérée), sans plus d'information sur ses paramètres, on a alors $\theta = (m, \sigma^2)$, l'espace des paramètres est $\Theta = \mathbf{R} \times \mathbf{R}_+^*$ et $\mathcal{X} = \mathbf{R}$.

Exemple 6.3. Si la loi mère est supposée gaussienne de moyenne m (non dégénérée), mais sans d'information sur sa variance, on a alors $\theta = \sigma^2$, l'espace des paramètres est $\Theta = \mathbf{R}_+^*$ et $\mathcal{X} = \mathbf{R}$.

Exemple 6.4. Si la loi mère est supposée uniforme sur $[1, \theta]$ pour un $\theta > 1$, on a $\Theta = [1, +\infty[$ et $\mathcal{X} = [1, +\infty[$.

Dans certains cas, il pourra être utile d'estimer une fonction de θ , disons $g(\theta)$, plutôt que θ lui-même, par exemple, si $g(\theta)$ est la moyenne, ou plus généralement, un moment de la loi \mathbf{P}_θ . Notons également qu'il est possible de donner plusieurs paramétrisations de la même famille de lois (voir suite de l'Exemple 6.4).

Définition 6.1. Avec les conventions précédentes le couple $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ est appelé modèle statistique. Un estimateur de $g(\theta)$ est alors une fonction de X_1, \dots, X_n , indépendante de θ (donc une statistique) et à valeurs dans $g(\Theta)$.

Les Exemples 6.1 à 6.4 correspondent donc aux modèles statistiques $(\{0, 1\}, (\mathcal{B}(\theta))_{\theta \in [0,1]})$, $(\mathbf{R}, (\mathcal{N}(\theta))_{\theta \in \mathbf{R} \times \mathbf{R}_+^*})$, $(\mathbf{R}, (\mathcal{N}(m, \theta))_{\theta \in \mathbf{R}_+^*})$ et $([1, +\infty[, (\mathcal{U}([1, \theta]))_{\theta \in [1, +\infty[})$ respectivement. La moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et la v.a. p.s. constante égale à 1 (ou (1, 1) dans l'Exemple 6.2) sont des estimateurs mais la v.a. p.s. constante égale à θ n'en est pas un. Dans le modèle gaussien avec moyenne connue de l'Exemple 6.3, $n^{-1} \sum_{i=1}^n (X_i - m)^2$ est un estimateur θ mais dans le modèle gaussien avec moyenne inconnue de l'Exemple 6.2, $n^{-1} \sum_{i=1}^n (X_i - \theta_1)^2$ n'est pas un estimateur $\theta = (\theta_1, \theta_2) = (m, \sigma^2)$. Le « graal » serait d'identifier θ afin de caractériser la loi mère de l'échantillon. Ceci n'est pas possible mais on cherchera à déterminer des estimateurs pertinents pour chacun des paramètres (ou le paramètre vu comme un objet à éventuellement plusieurs dimensions). Ainsi, pour estimer la moyenne de la loi mère, on préférera certainement la moyenne empirique à la v.a. p.s. constante égale à 1. Notons que les estimateurs seront généralement désignés par des lettres majuscules, par exemple T_n , pour insister sur le fait qu'il s'agit de v.a. alors que les estimations, valeurs prises par un estimateur au cours d'une observation donnée, seront désignées par des lettres minuscules. Lorsque des lettres grecques sont utilisées, nous n'adopterons pas cette convention et n'utiliserons que les minuscules. Par exemple, $\hat{\theta}$ pourra être un estimateur de θ ou une estimation de θ , ce qui se décidera grâce au contexte.

Une exigence minimale sur un estimateur (ou plutôt une suite d'estimateurs) T_n de $g(\theta)$ est qu'il s'approche en un certain sens de $g(\theta)$ lorsque n tend vers l'infini. Si T_n converge vers $g(\theta)$ pour l'un des modes de convergence présentés dans le Chapitre 4, on dit qu'il est *convergent* pour ce mode. Plus rigoureusement on devrait dire la suite d'estimateurs est convergente. Pour deux de ces modes de convergences, certains auteurs utilisent les dénominations de la définition suivante.

Définition 6.2. On dit qu'un estimateur T_n de $g(\theta)$ est convergent (ou consistant) si $T_n \xrightarrow{p.s.} g(\theta)$ et qu'il est faiblement convergent si $T_n \xrightarrow{\mathbf{P}} g(\theta)$

Exemple 6.4 (suite).

Reprenons le modèle uniforme sur $[1, \theta]$ et choisissons de le reparamétriser en rappelant que si $X \sim \mathcal{U}([1, \theta])$, alors $\mathbf{E}_\theta[X] = (\theta + 1)/2 = \theta' \in \Theta' = [1, +\infty[$. Ici et dans la suite, \mathbf{E}_θ désigne l'espérance sous \mathbf{P}_θ . La loi forte des grands nombres (LGN) assure alors que la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur convergent de θ' . La moyenne empirique est, en fait, un estimateur convergent pour la moyenne pour tout modèle pour lequel la LGN s'applique.

Dans ce cours, nous considérerons des familles de v.a. continues ou discrètes. Dans le premier cas, la loi \mathbf{P}_θ est caractérisée par la (une en fait) densité $f(\cdot; \theta)$ dépendant du paramètre

à estimer. Dans le second cas, elle l'est par la fonction de probabilité $p(\cdot; \theta)$ que nous pourrions noter, avec un léger abus unificateur, $f(\cdot; \theta)$. Ces observations conduisent à la définition de la vraisemblance d'un modèle statistique.

Définition 6.3. On appelle vraisemblance du modèle statistique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ toute fonction L définie sur $\mathcal{X}^n \times \Theta$ dont les applications partielles $L(\cdot, \theta)$ sont la densité (ou la fonction de probabilité) de $\mathbf{P}_\theta^{\otimes n} = \mathbf{P}_\theta \otimes \cdots \otimes \mathbf{P}_\theta$ (n fois).

Remarque 6.1. On écrira abusivement « la » vraisemblance au lieu d'une vraisemblance (les vraisemblances ne peuvent différer que sur des ensembles négligeables). Cette définition s'étend plus généralement à des v.a. qui ne sont ni continues ni discrètes pour des modèles dominés (voir par exemple le premier chapitre de [8]). Nous ne nous attarderons pas sur ces généralisations hors des objectifs du cours. Pour la légèreté des notations, nous désignons les vraisemblances L sans insister que le fait que l'échantillon soit de taille n comme on le ferait avec la notation L_n . Le contexte nous guidera donc dans l'interprétation : $L(x; \theta)$ est à comprendre comme $L_1(x; \theta) = f_{X_1}(x; \theta)$ et $L(x_1, \dots, x_n; \theta)$ est à comprendre comme $L_n(x_1, \dots, x_n; \theta) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n; \theta)$.

Rappelons que la loi jointe d'un échantillon est caractérisée par sa densité (ou fonction de probabilité) jointe donnée. Lorsque l'on considère un échantillon i.i.d. (X_1, \dots, X_n) de loi mère supposée être \mathbf{P}_θ , par indépendance, celle-ci est donnée, pour $(x_1, \dots, x_n) \in \mathcal{X}$, par :

$$L(x_1, \dots, x_n; \theta) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{X_1}(x_i; \theta) = \prod_{i=1}^n L(x_i; \theta).$$

Par exemple, la vraisemblance du modèle gaussien de moyenne m de l'Exemple 6.3 est donnée par

$$L(x_1, \dots, x_n; \theta) = \frac{1}{(2\pi\theta)^{\frac{n}{2}}} \exp\left(-\frac{\sum_{i=1}^n (x_i - m)^2}{2\theta}\right).$$

Nous allons maintenant présenter les méthodes les plus classiques de construction d'estimateurs avant de donner des outils permettant d'analyser la qualité des estimateurs et donc de les discriminer.

6.2 Méthodes classiques de construction d'estimateurs

6.2.1 Méthode de substitution

Supposons que l'on dispose d'un estimateur convergent T_n de θ et que l'on souhaite estimer $g(\theta)$ pour une certaine fonction g continue. Une idée très simple fournit alors un estimateur convergent de $g(\theta)$: substituer T_n à θ . L'estimateur de $g(\theta)$ obtenu est $G_n = g(T_n)$.

Exemple 6.4 (suite).

Nous avons vu que la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur convergent de θ' avec la seconde paramétrisation proposée du modèle uniforme sur $[1, \theta']$ et que θ' est relié à θ dans la première paramétrisation présentée dans ces notes par la relation $\theta' = (\theta + 1)/2$ soit $\theta = g(\theta') := 2\theta' - 1$. Ainsi, $T_n = g(\bar{X}_n) = 2\bar{X}_n - 1$ est un estimateur convergent de θ (d'intérêt pour la première paramétrisation).

6.2.2 Méthode des moments

La *méthode des moments* présente l'avantage d'être très intuitive mais présuppose l'existence d'au moins autant de moments pour la loi \mathbf{P}_θ que de dimensions dans le paramètre à estimer et ne donne pas toujours des résultats satisfaisants. Elle n'est, par exemple, pas applicable pour la loi de Cauchy qui est sans moment (voir Section A.2.9).

Cette méthode repose sur le fait que, si \mathbf{P}_θ admet des moments jusqu'à l'ordre k $\mu_1(\theta), \dots, \mu_k(\theta)$, la LGN assure que, pour tout $r \in \{1, \dots, k\}$ le r^{e} moment empirique $M_r = \frac{1}{n} \sum_{i=1}^n X_i^r$ est un estimateur convergent de $\mu_r(\theta)$. Ainsi, si le paramètre θ est de dimension k , pour une réalisation (m_1, \dots, m_k) des moments empiriques (M_1, \dots, M_k) , la résolution en θ du système :

$$\begin{cases} \mu_1(\theta) = m_1 \\ \mu_2(\theta) = m_2 \\ \dots\dots\dots \\ \mu_k(\theta) = m_k \end{cases}$$

fournit une estimation de θ par la méthode des moments. L'estimateur de la méthode des moments (EM) est obtenu en remplaçant la réalisation (m_1, \dots, m_k) par les v.a. « moments empiriques » (M_1, \dots, M_k) . S'il est nécessaire de mettre en valeur le fait qu'un estimateur a été obtenu par la méthode des moments, on le notera $\hat{\theta}^M$. Bien sûr, cette estimation n'est valable que si le système précédent admet une unique solution dans l'espace des paramètres Θ .

Exemple 6.2 (suite).

Reprenons l'exemple de l'estimation du paramètre de dimension 2, $\theta = (\theta_1, \theta_2) = (m, \sigma^2)$, dans le modèle gaussien. Le premier moment de la loi $\mathcal{N}(\theta) = \mathcal{N}(m, \sigma^2)$ est $\mu_1(\theta) = m = \theta_1$ et son deuxième moment est $\mu_2(\theta) = m^2 + \sigma^2 = \theta_1^2 + \theta_2$. On est donc amenés, en écrivant avec un léger abus directement les v.a. M_1 et M_2 , à résoudre le système

$$\begin{cases} \theta_1 = M_1 \\ \theta_1^2 + \theta_2 = M_2 \end{cases} .$$

On obtient

$$\begin{cases} \theta_1 = M_1 = \bar{X}_n \\ \theta_2 = M_2 - M_1^2 = \tilde{S}_n^2 \end{cases}$$

où \tilde{S}_n^2 désigne la variance empirique (voir 5.1). La méthode des moments conduit donc dans le cadre gaussien à estimer la moyenne et la variance par la moyenne empirique et la variance empirique respectivement.

Exemple 6.5. Considérons le modèle de lois beta $([0, 1], (\beta(\theta))_{\theta \in (\mathbf{R}_+^*)^2})$ (voir A.2.4). La méthode des moments conduit au système :

$$\begin{cases} \frac{\theta_1}{\theta_1 + \theta_2} = M_1 = \bar{X}_n \\ \frac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2 (\theta_1 + \theta_2 + 1)} + \left(\frac{\theta_1}{\theta_1 + \theta_2} \right)^2 = M_2 = \tilde{S}_n^2 + \bar{X}_n^2 \end{cases} .$$

La résolution de celui-ci donne les estimateurs convergents de θ_1 et θ_2 :

$$\hat{\theta}_1^M = \bar{X}_n \frac{\bar{X}_n - M_2}{\tilde{S}_n^2} \quad \text{et} \quad \hat{\theta}_2^M = (1 - \bar{X}_n) \frac{\bar{X}_n - M_2}{\tilde{S}_n^2} .$$

6.2.3 Méthode du maximum de vraisemblance

La méthode d'estimation par *maximum de vraisemblance* est certainement celle d'usage le plus courant et c'est imposée comme référence. Elle est notamment celle utilisée par la plus part des logiciels de statistique à l'instar de R. L'idée sous-jacente est de chercher, dans l'espace des paramètres Θ , en disposant d'une observation (x_1, \dots, x_n) un paramètre rendant la plus probable possible l'observation. Ainsi, sous réserve qu'il existe, une *estimation du maximum de vraisemblance*, basé sur l'observation (x_1, \dots, x_n) est :

$$\hat{\theta}^{MV} = \operatorname{argmax}_{\theta \in \Theta} L(x_1, \dots, x_n; \theta),$$

où L est la vraisemblance du modèle. Il s'agit alors d'une fonction de (x_1, \dots, x_n) , disons $h(x_1, \dots, x_n)$. L'*estimateur du maximum de vraisemblance* (EMV) est alors la statistique $h(X_1, \dots, X_n)$ et est fréquemment lui aussi noté $\hat{\theta}^{MV}$.

Remarque 6.2. Même si, pour la plupart des modèles statistiques, l'EMV existe et est unique, ceci n'est pas toujours le cas. Pour s'en convaincre, le lecteur peut considérer l'estimation de la moyenne $\theta \in \mathbf{R}$ dans famille des lois de Laplace \mathbf{P}_θ de densité

$$f(x) = \frac{1}{2} e^{-|x-\theta|}, \quad x \in \mathbf{R}.$$

S'agissant d'un problème de maximisation en θ d'une fonction positive (la vraisemblance $L(x_1, \dots, x_n; \dots)$), il est équivalent à celui de la maximisation de son logarithme, la *Log-vraisemblance* $\ln L(x_1, \dots, x_n; \dots)$. Ce second problème peut conduire, dans certains cas, à des calculs plus simples. C'est en particulier le cas lorsque la vraisemblance est produit de puissances ou d'exponentielles.

Remarque 6.3. On montre que, sous certaines hypothèses sur le modèle statistique, l'EMV est un estimateur convergent (voir par exemple [7, Théorème VIII.9.]).

Exemple 6.2 (suite). Reprenons le modèle gaussien avec pour paramètre à estimer $\theta = (\theta_1, \theta_2) = (m, \sigma^2)$. La vraisemblance s'écrit

$$L(x_1, \dots, x_n; \theta_1, \theta_2) = \frac{1}{(2\pi\theta_2)^{\frac{n}{2}}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \theta_1)^2}{2\theta_2}\right),$$

et la Log-vraisemblance s'écrit

$$\ln L(x_1, \dots, x_n; \theta_1, \theta_2) = -\frac{1}{2\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2 - \frac{n}{2} \ln(2\pi\theta_2).$$

En dérivant en θ_1 et θ_2 , on obtient les système :

$$\begin{cases} \sum_{i=1}^n (x_i - \theta_1) = 0 \\ \frac{1}{\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2 - \frac{n}{\theta_2} = 0 \end{cases}$$

dont la solution est $(\theta_1, \theta_2) = (\bar{x}_n, \tilde{s}_n^2)$. Ainsi, l'EMV est pour ce modèle : $(\bar{X}_n, \tilde{S}_n^2)$.

Exemple 6.4 (suite). Reprenons l'exemple de l'estimation du paramètre $\theta > 1$ du modèle uniforme sur $[1, \theta]$. La vraisemblance s'écrit

$$L(x_1, \dots, x_n; \theta) = \frac{1}{(\theta - 1)^n} \prod_{i=1}^n \mathbf{1}_{x_i \in [1, \theta]} = (\theta - 1)^{-n} \mathbf{1}_{x_{(1)} \geq 1} \mathbf{1}_{x_{(n)} \leq \theta}$$

où $x_{(1)} = \min(x_1, \dots, x_n)$ et $x_{(n)} = \max(x_1, \dots, x_n)$. Cette vraisemblance est donc nulle si $x_{(n)} < \theta$ et strictement positive sinon. La fonction $\theta \mapsto (\theta - 1)^{-n}$ étant décroissante en θ , il vient que le maximum de vraisemblance est atteint en $x_{(n)}$. Ainsi, l'EMV est pour ce modèle la n^{e} statistique d'ordre $X_{(n)}$.

Remarque 6.4. On voit que dans l'Exemple 6.4, l'EMV $X_{(n)}$ diffère de l'estimateur obtenu par la méthode des moments (ou de substitution) $2\bar{X}_n - 1$. Nous développerons dans la suite des outils permettant de décider lequel privilégier.

La proposition suivante, dont la démonstration est immédiate, montre que la méthode d'estimation par maximum de vraisemblance est invariante par reparamétrisation.

Proposition 6.1. Soit $h : \Theta \rightarrow \Theta'$ bijective induisant une reparamétrisation du modèle $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ en $(\mathcal{X}, (\mathbf{P}_{\theta'})_{\theta' \in \Theta'})$. Alors, l'EMV de $\theta' = h(\theta)$ est

$$\hat{\theta}'^{MV} = h(\hat{\theta}^{MV}).$$

Remarque 6.5. Plus généralement, que h soit bijective ou non, on appellera *estimateur du maximum de vraisemblance* de $h(\theta)$ la statistique $h(\hat{\theta}^{MV})$.

6.2.4 Approche bayésienne

L'*approche bayésienne* est une approche particulière en statistique consistant à considérer le paramètre θ comme une variable aléatoire sur Θ . L'espace des paramètres Θ étant continu dans les exemples que nous traiterons, on supposera la loi ν le régissant continue de densité f_ν . Cette loi est appelée *loi a priori*. Elle doit refléter une connaissance acquise sur le paramètre par le passé, par exemple lors d'expériences précédentes. On cherchera pour autant généralement à utiliser des lois *a priori* peu ou non informatives pour limiter leur impact sur les résultats finaux. L'information sur le paramètre sera *in fine* accessible au travers de la *loi a posteriori* que nous allons définir. Notons que, dans le cadre bayésien, la vraisemblance du modèle doit s'entendre comme une vraisemblance conditionnelle par rapport à la v.a. θ . Il serait donc plus approprié de la noter $L(x_1, \dots, x_n | \theta)$ mais nous continuerons pourtant à la noter $L(x_1, \dots, x_n; \theta)$.

Rappelons la classique *formule de Bayes* permettant, pour deux événements A et B de probabilité strictement positive, d'« échanger le conditionnement » :

$$\mathbf{P}[B|A] = \frac{\mathbf{P}[A|B]\mathbf{P}[B]}{\mathbf{P}[A]}.$$

L'idée est maintenant d'utiliser cette formule pour passer de la loi $\mathbf{x} = (x_1, \dots, x_n)$ sachant notre *a priori* sur θ à la loi de θ sachant que l'on a effectivement observé \mathbf{x} ; il s'agit bien d'une

information *a posteriori* puisque obtenu après l'observation. La densité de la loi *a posteriori* $\mathbf{P}_{\mathbf{x}}$, $\mathbf{x} \in \mathcal{X}^n$, sachant $\mathbf{X} = (X_1, \dots, X_n) = \mathbf{x}$ s'écrit alors :

$$f_{\nu|\mathbf{X}=\mathbf{x}}(\theta) = \frac{L(\mathbf{x}; \theta) f_{\nu}(\theta)}{\int_{\Theta} L(\mathbf{x}; \theta) f_{\nu}(\theta) d\theta} = c(\mathbf{x}, \nu) L(\mathbf{x}; \theta) f_{\nu}(\theta).$$

Insistons sur le fait qu'ici $L(\mathbf{x}; \theta)$ se comprend comme la densité (ou fonction de probabilité) de \mathbf{x} sachant θ . Dans l'équation précédente, $c(\mathbf{x}, \nu)$ ne dépend pas de θ et est la constante de normalisation de la densité conditionnelle $f_{\nu|\mathbf{X}=\mathbf{x}}$. L'estimation bayésienne étant souvent assez gourmande en calculs, ne pas expliciter cette constante (mais utiliser que c'est la « bonne valeur » pour obtenir à la fin une densité en θ) peut être intéressant pour limiter le nombre de calculs.

Ayant observé $\mathbf{x} = (x_1, \dots, x_n)$, une *estimation bayésienne* de θ est alors simplement :

$$\hat{\theta}^B = \mathbf{E}_{\mathbf{x}}[\theta]$$

où $\mathbf{E}_{\mathbf{x}}$ désigne l'espérance sous $\mathbf{P}_{\mathbf{x}}$. Si l'on désigne par h la fonction de \mathcal{X}^n dans Θ $\mathbf{x} \mapsto \hat{\theta}^B$ l'*estimateur bayésien* de θ est $h(\mathbf{X})$ que l'on notera également $\hat{\theta}^B$.

Remarque 6.6.

1. Ici, on a choisi comme estimation bayésienne de θ la moyenne de la loi *a posteriori* ce qui correspond à la minimisation du *risque quadratique* (voir Section 6.3.2). Il s'agit du critère de choix d'estimateur retenu dans ce cours puisque le plus courant. Il est possible de considérer d'autres critères comme un risque L^1 (donné, pour un estimateur T par $\mathbf{E}_{\theta}[|T - \theta|]$). Sa minimisation conduirait à préférer la médiane de la loi *a posteriori*.
2. Il est possible de montrer de nombreuses propriétés de ces estimateurs dont leur convergence pour toute loi *a priori*.
3. Ces estimateurs seront utilisés dans le chapitre d'estimation par intervalle de confiance (voir Chapitre 8) et peuvent également l'être dans la *théorie de la décision*. Nous ne la développerons pas ici et renvoyons, par exemple, à [8, Chapitre 7].

Exemple 6.6. Considérons le problème de l'estimation de θ dans la famille de lois $\mathbf{P}_{\theta} = \mathcal{N}(\theta, 1)$, $\theta \in \mathbf{R}$ et choisissons comme loi *a priori* sur $\Theta = \mathbf{R}$ la loi $\nu = \mathcal{N}(m, \sigma^2)$ (m et σ^2 connus).

On a, pour $\mathbf{x} = (x_1, \dots, x_n)$:

$$L(\mathbf{x}; \theta) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right)$$

et la densité de la loi *a priori* est au point θ

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - m)^2}{2\sigma^2}\right).$$

Ainsi, la densité de la loi *a posteriori* est :

$$\begin{aligned}
 f_{\nu|\mathbf{X}=\mathbf{x}}(\theta) &= \frac{L(\mathbf{x}; \theta) f_{\nu}(\theta)}{\int_{\Theta} L(\mathbf{x}; \theta) f_{\nu}(\theta) d\theta} \\
 &= \frac{\frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta-m)^2}{2\sigma^2}\right)}{\int_{\mathbf{R}} \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta-m)^2}{2\sigma^2}\right) d\theta} \\
 &= c(\mathbf{x}) \exp\left(-\frac{1}{2} \left(\sum_{i=1}^n (x_i - \theta)^2 + \frac{(\theta - m)^2}{\sigma^2}\right)\right)
 \end{aligned}$$

où $c(\mathbf{x})$ ne dépend pas de θ et dont la valeur pouvant changer d'une ligne à l'autre se retrouvera à la fin en utilisant que $f_{\nu|\mathbf{X}=\mathbf{x}}(\cdot)$ est une densité (en θ). Après des calculs assez fastidieux mais simples consistant en la réduction de la forme quadratique dans l'exponentielle, on obtient que

$$f_{\nu|\mathbf{X}=\mathbf{x}}(\theta) = c(\mathbf{x}) \exp\left(-\frac{1}{2} \frac{1+n\sigma^2}{\sigma^2} \left(\theta - \left(\frac{1}{1+n\sigma^2}m + \frac{n\sigma^2}{1+n\sigma^2}\bar{x}_n\right)\right)^2\right).$$

En reconnaissant dans l'exponentielle un noyau gaussien (et déduisant la valeur de $c(\mathbf{x})$), on obtient que la loi *a posteriori* $\mathbf{P}_{\mathbf{x}}$ n'est autre que :

$$\mathcal{N}\left(\frac{1}{1+n\sigma^2}m + \frac{n\sigma^2}{1+n\sigma^2}\bar{x}_n; \frac{\sigma^2}{1+n\sigma^2}\right).$$

L'estimation bayésienne de θ est donc

$$\hat{\theta}^B = \frac{1}{1+n\sigma^2}m + \frac{n\sigma^2}{1+n\sigma^2}\bar{x}_n.$$

Notons que lorsque la variance σ^2 tend vers $+\infty$, $\hat{\theta}^B$ tend vers \bar{x}_n en « oubliant » les paramètres de la loi *a priori*. Ceci traduit le fait que l'information apportée par la loi *a priori* est alors de plus en plus vague. Notons également que lorsque n tend vers $+\infty$ la variance de la loi *a posteriori* tend vers 0.

6.3 Analyse des estimateurs et choix d'un estimateur

Dans cette section, on introduit des outils permettant d'analyser la qualité des estimateurs et, par suite, de choisir quel estimateur préférer. Ces outils sont à voir comme fonction du paramètre inconnu θ . On considérera un modèle d'échantillonnage (i.i.d.) pour lequel les variables X_1, \dots, X_n sont continues à valeurs dans $\mathcal{X} = \mathbf{R}^d$. Ce choix est fait par soucis de concision et de simplicité. Un exposé tout aussi concis mais un peu plus formel reste possible en profitant de la théorie de la mesure et de l'intégration de Lebesgue. Il s'agirait de considérer, comme dans [8] par exemple, des modèles statistiques dominés par une mesure σ -finie : la mesure de Lebesgue pour le cas continu et une mesure de comptage pour le cas discret. On peut étendre les notions abordées au cas discret en remplaçant le symbole \int par \sum . Dans la suite et sans le rappeler, on désignera par \mathbf{E}_{θ} et \mathbf{V}_{θ} l'espérance et la variance sous la loi \mathbf{P}_{θ} .

6.3.1 Biais

La première mesure de la qualité d'un estimateur T est son écart-moyen au paramètre θ sous la loi \mathbf{P}_θ .

Définition 6.4. Soit T un estimateur de θ intégrable. Le biais de T pour θ est

$$\mathbf{b}(\theta) = \mathbf{b}(T, \theta) := \mathbf{E}_\theta[T] - \theta.$$

On dit qu'un estimateur est sans biais si $\mathbf{b}(\theta) = 0$ quelque soit $\theta \in \Theta$.

Remarque 6.7. Attention : si T est un estimateur sans biais de θ , $g(T)$ n'est pas nécessairement un estimateur sans biais de $g(\theta)$.

Il est clair qu'on attend d'un estimateur d'être de biais faible voir sans biais. Nous verrons cependant, que l'on peut préférer un estimateur avec un léger biais à un estimateur sans biais si ce premier a une moins forte variabilité autour de θ . La recherche de bons estimateurs sans biais reste une question importante en statistique.

Remarque 6.8. Dans le cadre asymptotique, certaines suites d'estimateurs (on dira simplement estimateur par brièveté) sont de biais non nuls mais tendant vers 0 lorsque la taille de l'échantillon tend vers l'infini. On dit qu'ils sont *asymptotiquement sans biais*.

6.3.2 Risque quadratique ou erreur quadratique moyenne

On choisit ici, pour mesurer la variabilité d'un estimateur autour de la quantité à estimer, un risque en norme 2. Si d'autres choix sont possibles, il s'agit du plus couramment utilisé. De ce fait, nous supposerons dans la suite l'estimateur de carré intégrable. Notons que la variance de l'estimateur n'est pas l'outil le plus adéquat d'analyse puisqu'elle mesure la dispersion de l'estimateur autour de sa moyenne et non autour du paramètre ou de la fonction du paramètre à estimer, ce qui est différent en présence de biais.

Définition 6.5. Soit T un estimateur de θ de carré intégrable. La fonction de risque quadratique ou erreur quadratique moyenne de T pour l'estimation de θ est

$$\mathcal{R}(T, \theta) := \mathbf{E}_\theta[(T - \theta)^2], \quad \text{si } d = 1,$$

et

$$\mathcal{R}(T, \theta) := \mathbf{E}_\theta[(T - \theta)(T - \theta)^T], \quad \text{si } d > 1.$$

Remarque 6.9. Si $d > 1$, $\mathcal{R}(T, \theta)$ est une matrice symétrique et semi-définie positive. On utilise donc, dans ce cas, pour comparer les risques d'estimateurs, l'ordre partiel sur l'ensemble des matrices symétriques semi-définies positives défini par $A \geq B$ si $A - B$ est semi-définie positive (c'est-à-dire si ces valeurs propres sont positives ou nulles).

Si la variance de l'estimateur n'est pas une mesure pertinente de sa qualité, son risque se ré-exprime au travers d'elle comme le montre la proposition suivante dont la démonstration similaire à celle de la formule de décentrage de la variance est laissée en exercice.

Proposition 6.2 (décomposition biais/variance). On a :

$$\mathcal{R}(T, \theta) = \mathbf{V}_\theta[T] + \mathbf{b}(T, \theta)^2,$$

en particulier, si T est sans biais :

$$\mathcal{R}(T, \theta) = \mathbf{V}_\theta[T].$$

Il est naturel de chercher des estimateurs de risque le plus faible possible.

Définition 6.6. On dit qu'un estimateur $T^{(1)}$ est préférable à $T^{(2)}$, si pour tout $\theta \in \Theta$:

$$\mathcal{R}(T^{(1)}, \theta) \leq \mathcal{R}(T^{(2)}, \theta).$$

On dit que $T^{(2)}$ est inadmissible, s'il existe un estimateur $T^{(1)}$ préférable à $T^{(2)}$ pour lequel l'inégalité ci dessus est stricte pour au moins une valeur de θ .

Remarque 6.10. Pour les modèles d'intérêt, il n'existe malheureusement pas d'estimateur préférable à tous les autres. En effet, si tel est le cas, les supports des lois \mathbf{P}_θ , $\theta \in \Theta$, sont nécessairement disjoints et une unique observation permet l'identification exacte de θ . Il sera toutefois possible de rechercher des estimateurs de risque minimal dans des classes d'estimateurs. Par exemple, on pourra rechercher un estimateur minimisant le risque quadratique parmi les estimateurs sans biais ; par la Proposition 6.2, ceci est, dans cette classe, équivalent à la minimisation de la variance. Le lecteur désireux de plus de détails pourra consulter, par exemple, la Section 6.6 de [11] ou la Section 4.3.2 de [8].

Exemple 6.7. Considérons le modèle $(\{0, 1\}, (\mathcal{B}er(\theta))_{\theta \in [0,1]})$ dans lequel on veut estimer θ au moyen d'un échantillon de taille n . On propose les estimateurs :

$$T^{(1)} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad T^{(2)} = \frac{1}{n+2} \left(1 + \sum_{i=1}^n X_i \right).$$

On a

$$\mathbf{E}_\theta [T^{(1)}] = \mathbf{E}_\theta \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \theta,$$

donc $T^{(1)}$ est sans biais et on a

$$\mathcal{R}(T^{(1)}, \theta) = \mathbf{V}_\theta [T^{(1)}] = \mathbf{V}_\theta \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{\theta(1-\theta)}{n}.$$

Par ailleurs,

$$\mathbf{E}_\theta [T^{(2)}] = \mathbf{E}_\theta \left[\frac{1}{n+2} \left(1 + \sum_{i=1}^n X_i \right) \right] = \frac{1+n\theta}{n+2},$$

donc

$$\mathbf{b}(T^{(2)}, \theta) = \frac{1+n\theta}{n+2} - \theta = \frac{1-2\theta}{n+2}$$

et en utilisant la décomposition biais/variance du risque, on a

$$\mathcal{R}(T^{(2)}, \theta) = \mathbf{V}_\theta [T^{(2)}] + \mathbf{b}(T^{(2)}, \theta)^2 = \frac{n\theta(1-\theta)}{(n+2)^2} + \left(\frac{1-2\theta}{n+2} \right)^2 = \frac{n\theta(1-\theta) + (1-2\theta)^2}{(n+2)^2}.$$

Pour $\theta = 1$, on a

$$\mathcal{R}(T^{(1)}, 1) = 0 \leq \mathcal{R}(T^{(2)}, 1) = \frac{1}{(n+2)^2},$$

alors que pour $\theta = \frac{1}{2}$, on a

$$\mathcal{R}\left(T^{(1)}, \frac{1}{2}\right) = \frac{1}{4n} \geq \mathcal{R}\left(T^{(2)}, \frac{1}{2}\right) = \frac{n}{4(n+2)^2}.$$

Ainsi, aucun des deux estimateurs $T^{(1)}$ et $T^{(2)}$ n'est préférable à l'autre.

6.3.3 Modèles et estimateurs réguliers

Si $\theta = (\theta_1, \dots, \theta_d)$ et si les dérivées partielles de f existent, on note :

$$\frac{\partial f}{\partial \theta} = \left(\frac{\partial f}{\partial \theta_1}, \dots, \frac{\partial f}{\partial \theta_d} \right)$$

et

$$\frac{\partial^2 f}{\partial \theta^2} = \left(\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \right)_{1 \leq i, j \leq d}.$$

Définition 6.7. On dit qu'un modèle statistique est régulier si :

1. Θ est un ouvert ;
2. le support de \mathbf{P}_θ ne dépend pas de θ (il s'agit donc de \mathcal{X}) ;
3. $\frac{\partial f_{X_1}}{\partial \theta}(x, \theta)$ et $\frac{\partial^2 f_{X_1}}{\partial \theta^2}(x, \theta)$ existent pour tout $x \in \mathcal{X}$ et $\theta \in \Theta$;
4. $\frac{\partial f_{X_1}}{\partial \theta}$ et $\frac{\partial^2 f_{X_1}}{\partial \theta^2}$ sont intégrable et on peut les effectuer sous le signe somme :

$$\frac{\partial}{\partial \theta} \int_A f_{X_1}(x, \theta) \, dx = \int_A \frac{\partial}{\partial \theta} f_{X_1}(x, \theta) \, dx$$

et

$$\frac{\partial^2}{\partial \theta^2} \int_A f_{X_1}(x, \theta) \, dx = \int_A \frac{\partial^2}{\partial \theta^2} f_{X_1}(x, \theta) \, dx$$

pour tout A ;

5. $\frac{\partial \ln f_{X_1}}{\partial \theta}$ est de carré intégrable.

Exercice 6.1. Montrer que le modèle gaussien de l'Exemple 6.2 est régulier.

Exercice 6.2. Le modèle uniforme sur $[1, \theta]$ de l'Exemple 6.4 est-il régulier ?

Définition 6.8. Un estimateur T de carré intégrable sur un modèle statistique régulier $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ est dit régulier si, pour tout $\theta \in \Theta$, $T(\cdot) \frac{\partial}{\partial \theta} L(\cdot, \theta)$ est intégrable sur \mathcal{X}^n et

$$\frac{\partial}{\partial \theta} \int_{\mathcal{X}^n} T(\mathbf{x}) L(\mathbf{x}, \theta) \, d\mathbf{x} = \int_{\mathcal{X}^n} T(\mathbf{x}) \frac{\partial}{\partial \theta} L(\mathbf{x}, \theta) \, d\mathbf{x}.$$

6.3.4 Score, information de Fisher et borne de Cramer-Rao

On considère un modèle régulier de vraisemblance

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{X_1}(x_i, \theta).$$

Définition 6.9. On appelle score de (X_1, \dots, X_n) le vecteur :

$$V_n = \frac{\partial \ln L}{\partial \theta}(X_1, \dots, X_n; \theta) = \sum_{i=1}^n \frac{\partial \ln f_{X_1}}{\partial \theta}(X_i, \theta).$$

Remarque 6.11. Si la (Log-)vraisemblance atteint son maximum en un point de Θ , l'estimateur du maximum de vraisemblance annule sa dérivée donc le score. Notons que si le paramètre à estimer θ est de dimension 1, le score n'est pas un vecteur mais simplement un réel, à voir comme fonction de θ .

Remarquons que :

$$\begin{aligned} \mathbf{E}_\theta \left[\frac{\partial \ln f_{X_1}}{\partial \theta}(X_1, \theta) \right] &= \int_{\mathcal{X}} \frac{\partial \ln f_{X_1}}{\partial \theta}(x, \theta) f_{X_1}(x, \theta) dx \\ &= \int_{\mathcal{X}} \frac{\partial f_{X_1}}{\partial \theta}(x, \theta) dx \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{X}} f_{X_1}(x, \theta) dx = \frac{\partial}{\partial \theta} 1 = 0. \end{aligned}$$

Ainsi le score V_n est centré.

Définition 6.10. On appelle information de Fisher de (X_1, \dots, X_n) la matrice de covariance/variance de son score :

$$I_n(\theta) = \mathbf{V}_\theta[V_n].$$

Remarque 6.12. Si la dimension du paramètre à estimer est $d = 1$, l'information de Fisher est simplement la variance du score.

Posons, si $d = 1$:

$$I(\theta) = \mathbf{E}_\theta \left[\left(\frac{\partial \ln f_{X_1}}{\partial \theta}(X_1, \theta) \right)^2 \right]$$

et si $d > 1$

$$I(\theta) = \mathbf{E}_\theta \left[\left(\frac{\partial \ln f_{X_1}}{\partial \theta}(X_1, \theta) \right) \left(\frac{\partial \ln f_{X_1}}{\partial \theta}(X_1, \theta) \right)^T \right].$$

Le résultat suivant relie $I_n(\theta)$ et $I(\theta)$ est immédiat par indépendance.

Proposition 6.3.

$$I_n(\theta) = nI(\theta).$$

Si le modèle est régulier, on a la reformulation suivante de $I(\theta)$.

Proposition 6.4. Si le modèle est régulier, on a :

$$I(\theta) = -\mathbf{E}_\theta \left[\frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(X_1, \theta) \right].$$

Preuve : On se place dans le cas $d = 1$ pour alléger la démonstration. On a :

$$\begin{aligned}
 \frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(x, \theta) &= \frac{\partial}{\partial \theta} \left(\frac{\partial \ln f_{X_1}}{\partial \theta}(x, \theta) \right) \\
 &= \frac{\partial}{\partial \theta} \left(\frac{1}{f_{X_1}(x, \theta)} \frac{\partial f_{X_1}}{\partial \theta}(x, \theta) \right) \\
 &= - \left(\frac{1}{f_{X_1}(x, \theta)} \frac{\partial f_{X_1}}{\partial \theta}(x, \theta) \right)^2 + \frac{1}{f_{X_1}(x, \theta)} \frac{\partial^2 f_{X_1}}{\partial \theta^2}(x, \theta) \\
 &= - \left(\frac{\partial \ln f_{X_1}}{\partial \theta}(x, \theta) \right)^2 + \frac{1}{f_{X_1}(x, \theta)} \frac{\partial^2 f_{X_1}}{\partial \theta^2}(x, \theta)
 \end{aligned}$$

donc

$$\left(\frac{\partial \ln f_{X_1}}{\partial \theta}(x, \theta) \right)^2 = \frac{1}{f_{X_1}(x, \theta)} \frac{\partial^2 f_{X_1}}{\partial \theta^2}(x, \theta) - \frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(x, \theta).$$

Puis, en prenant l'espérance contre \mathbf{P}_θ :

$$\begin{aligned}
 I(\theta) &= \mathbf{E}_\theta \left[\left(\frac{\partial \ln f_{X_1}}{\partial \theta}(X_1, \theta) \right)^2 \right] \\
 &= \mathbf{E}_\theta \left[\frac{1}{f_{X_1}(X_1, \theta)} \frac{\partial^2 f_{X_1}}{\partial \theta^2}(X_1, \theta) \right] - \mathbf{E}_\theta \left[\frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(X_1, \theta) \right] \\
 &= \int_{\mathcal{X}} \frac{\partial^2 f_{X_1}}{\partial \theta^2}(x, \theta) dx - \mathbf{E}_\theta \left[\frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(X_1, \theta) \right] \\
 &= \frac{\partial^2}{\partial \theta^2} \int_{\mathcal{X}} f_{X_1}(x, \theta) dx - \mathbf{E}_\theta \left[\frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(X_1, \theta) \right] \quad (\text{car le modèle est régulier}) \\
 &= \frac{\partial^2}{\partial \theta^2} 1 - \mathbf{E}_\theta \left[\frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(X_1, \theta) \right] = -\mathbf{E}_\theta \left[\frac{\partial^2 \ln f_{X_1}}{\partial \theta^2}(X_1, \theta) \right].
 \end{aligned}$$

□

Exercice 6.3. Vérifier que pour le modèle gaussien de l'Exemple 6.2, on a :

$$I_n(\theta_1, \theta_2) = \frac{n}{2\theta_2^2} \begin{pmatrix} 2\theta_2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Un intérêt majeur de l'information de Fisher est qu'elle permet de formuler une borne absolue sur le risque des estimateurs dans des modèles réguliers.

Théorème 6.1. (borne de Cramer-Rao, $d = 1$). Soit $T = T(X_1, \dots, X_n)$ un estimateur régulier et sans biais de θ de dimension 1 dans un modèle régulier et construit à partir d'un échantillon $\mathbf{X} = (X_1, \dots, X_n)$ de taille n . Alors, pour tout $\theta \in \Theta$:

$$\mathcal{R}(T, \theta) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}.$$

La borne $\frac{1}{nI(\theta)}$ est appelée borne de Cramer-Rao du modèle.

Preuve : Puisque T est sans biais, pour tout $\theta \in \Theta$, on a :

$$\theta = \mathbf{E}_\theta[T]$$

donc

$$\begin{aligned} 1 &= \frac{\partial}{\partial \theta} \theta = \frac{\partial}{\partial \theta} \mathbf{E}_\theta[T] \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{X}^n} T(\mathbf{x}) L(\mathbf{x}, \theta) \, d\mathbf{x} \\ &= \int_{\mathcal{X}^n} T(\mathbf{x}) \frac{\partial}{\partial \theta} L(\mathbf{x}, \theta) \, d\mathbf{x} \\ &= \int_{\mathcal{X}^n} T(\mathbf{x}) \left(\frac{\partial}{\partial \theta} \ln L(\mathbf{x}, \theta) \right) L(\mathbf{x}, \theta) \, d\mathbf{x} \\ &= \mathbf{E}_\theta \left[T(\mathbf{X}) \frac{\partial}{\partial \theta} \ln L(\mathbf{X}, \theta) \right] \\ &= \mathbf{E}_\theta \left[T(\mathbf{X}) \frac{\partial}{\partial \theta} \ln L(\mathbf{X}, \theta) \right] - \theta \mathbf{E}_\theta \left[\frac{\partial}{\partial \theta} \ln L(\mathbf{X}, \theta) \right] \quad (\text{car le score est centré}) \\ &= \mathbf{E}_\theta \left[(T(\mathbf{X}) - \theta) \frac{\partial}{\partial \theta} \ln L(\mathbf{X}, \theta) \right] \\ &\leq \mathbf{E}_\theta \left[(T(\mathbf{X}) - \theta)^2 \right]^{\frac{1}{2}} \mathbf{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \ln L(\mathbf{X}, \theta) \right)^2 \right]^{\frac{1}{2}} \quad (\text{par l'inégalité de Cauchy-Schwarz}) \\ &= \sqrt{\mathcal{R}(T, \theta) I_n(\theta)}. \end{aligned}$$

Le résultat s'ensuit. □

Ce résultat se généralise au cas d'un paramètre de dimension d quelconque :

Théorème 6.2. (borne de Cramer-Rao, $d > 1$). Soit $T = T(X_1, \dots, X_n)$ un estimateur régulier et sans biais de θ de dimension $d > 1$ dans un modèle régulier et construit à partir d'un échantillon $\mathbf{X} = (X_1, \dots, X_n)$ de taille n . Si $I(\theta)$ est inversible, on a, pour tout $\theta \in \Theta$:

$$\mathcal{R}(T, \theta) \geq \frac{1}{n} I(\theta)^{-1}.$$

La borne $\frac{1}{n} I(\theta)^{-1}$ est appelée borne de Cramer-Rao du modèle.

Définition 6.11. Un estimateur sans-biais atteignant la borne de Cramer-Rao est dit efficace.

Par définition, un estimateur efficace est préférable à tout autre estimateur sans biais. Il existe pourtant, sauf dans des cas particuliers, des estimateurs biaisés préférables aux estimateurs efficaces.

Exemple 6.3 (suite).

Reprenons le modèle gaussien à moyenne connue de l'Exemple 6.3 : $(\mathbf{R}, (\mathcal{N}(m, \theta))_{\theta \in \mathbf{R}_+^*})$. On a :

$$I(\theta) = -\mathbf{E} \left[\frac{\partial^2}{\partial \theta^2} \ln f_{X_1}(X_1; \theta) \right] = \mathbf{E} \left[-\frac{1}{2\theta^2} + \frac{(X_1 - m)^2}{\theta^3} \right] = \frac{1}{2\theta^2}.$$

La borne de Cramer-Rao de ce modèle est donc $2\theta^2/n$.

On a vu que la variance empirique corrigée S_n^2 est un estimateur sans biais de θ . Il est également régulier. De plus, on sait que $\frac{n-1}{\theta}S_n^2 \sim \chi^2(n-1)$ (voir Théorème 5.2). On a donc :

$$\begin{aligned}\mathcal{R}(S_n^2, \theta) &= \mathbf{V}_\theta [S_n^2] = \left(\frac{\theta}{n-1}\right)^2 \mathbf{V}_\theta \left[\frac{n-1}{\theta}S_n^2\right] \\ &= \left(\frac{\theta}{n-1}\right)^2 2(n-1) = \frac{2\theta^2}{n-1}.\end{aligned}$$

Cet estimateur n'atteint donc pas la borne de Cramer-Rao.

En profitant du fait que la moyenne m est connue, il est naturel dans ce cas de proposer l'estimateur :

$$\widehat{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2.$$

On a :

$$\mathbf{E}_\theta[\widehat{S}_n^2] = \mathbf{E}_\theta \left[\frac{1}{n} \sum_{i=1}^n (X_i - m)^2 \right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}_\theta [(X_i - m)^2] = \theta.$$

Il s'agit donc d'un estimateur sans biais ; il est également régulier. On a cette fois :

$$\begin{aligned}\mathcal{R}(\widehat{S}_n^2, \theta) &= \mathbf{V}_\theta [\widehat{S}_n^2] = \mathbf{V}_\theta \left[\frac{1}{n} \sum_{i=1}^n (X_i - m)^2 \right] = \frac{1}{n} \mathbf{V}_\theta [(X_1 - m)^2] \\ &= \frac{1}{n} \left(\mathbf{E}_\theta [(X_1 - m)^4] - \mathbf{E}_\theta [(X_1 - m)^2]^2 \right) = \frac{1}{n} (3\theta^2 - \theta^2) = \frac{2\theta^2}{n}\end{aligned}$$

où l'on a utilisé que le moment d'ordre 4 de $Z = X_1 - m \sim \mathcal{N}(0, \theta)$ est $3\theta^2$ (ce qui s'obtient facilement avec la Proposition 2.10). Ainsi, cet estimateur sans biais atteint la borne de Cramer-Rao. Il est donc efficace.

Nous allons maintenant voir que l'on peut trouver un estimateur biaisé qui lui est préférable. Pour cela considérons les estimateurs de la forme $T^{(\alpha)} = \alpha \widehat{S}_n^2$, $\alpha \in \mathbf{R}$. Son biais est :

$$\mathbf{b}_\alpha(\theta) = \mathbf{E}_\theta[T^{(\alpha)}] - \theta = \mathbf{E}_\theta[\alpha \widehat{S}_n^2] - \theta = (\alpha - 1)\theta.$$

et son risque quadratique :

$$\begin{aligned}\mathcal{R}(T^{(\alpha)}, \theta) &= \mathbf{V}_\theta [\alpha \widehat{S}_n^2] + \mathbf{b}_\alpha(\theta)^2 = \alpha^2 \mathbf{V}_\theta [\widehat{S}_n^2] + (\alpha - 1)^2 \theta^2 = \alpha^2 \mathcal{R}(\widehat{S}_n^2, \theta) + (\alpha - 1)^2 \theta^2 \\ &= \frac{2\theta^2}{n} \alpha^2 + \alpha^2 \theta^2 - 2\alpha \theta^2 + \theta^2 = \frac{(n+2)\theta^2}{n} \alpha^2 - 2\theta^2 \alpha + \theta^2.\end{aligned}$$

Une simple optimisation en α montre que le risque quadratique est minimal pour $\alpha_0 = \frac{n}{n+2}$. Ainsi, $T^{(\alpha_0)}$ est préférable à l'estimateur efficace \widehat{S}_n^2 .

Exemple 6.2 (suite). On a vu dans l'Exercice 6.3 que :

$$I_n(\theta_1, \theta_2) = \frac{n}{2\theta_2^2} \begin{pmatrix} 2\theta_2 & 0 \\ 0 & 1 \end{pmatrix}.$$

La borne de Cramer-Rao de ce modèle est donc :

$$\frac{\theta_2}{n} \begin{pmatrix} 1 & 0 \\ 0 & 2\theta_2 \end{pmatrix}.$$

On a vu que $T = (\bar{X}_n, S_n^2)$ est un estimateur sans biais (et régulier) de $\theta = (\theta_1, \theta_2)$. De plus, par le Théorème 5.2, on sait que $(\bar{X}_n$ et $S_n^2)$ sont indépendante, que $\bar{X}_n \sim \mathcal{N}(\theta_1, \theta_2/n)$ et que $\frac{n-1}{\theta} S_n^2 \sim \chi^2(n-1)$. Ainsi,

$$\mathcal{R}(T, \theta) = \mathbf{V}_\theta[T] = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2}{n-1} \end{pmatrix}$$

et $T = (\bar{X}_n, S_n^2)$ n'est pas un estimateur efficace. On montre qu'il est pourtant préférable à tout autre estimateur sans biais.

Exercice 6.4. Montrer que pour le modèle gaussien à variance connue σ^2 , $(\mathbf{R}, (\mathcal{N}(\theta, \sigma^2))_{\theta \in \mathbf{R}}$, la moyenne empirique \bar{X}_n est un estimateur efficace.

Remarquons que la borne de Cramer-Rao n'est pas forcément atteinte par un estimateur régulier en général ce qui constitue une limite de ce résultat. Nous admettrons le résultat suivant :

Proposition 6.5. *La borne de Cramer-Rao ne peut être atteinte que si $(\mathbf{P}_\theta)_{\theta \in \Theta}$ est une famille exponentielle de lois.*

6.3.5 Exhaustivité, minimalité

Il est intuitif qu'un « bon » estimateur doit retenir d'un échantillon la totalité de l'information utile pour l'estimation mais pas d'information superflue. Ceci conduit aux notions d'*exhaustivité* et d'*exhaustivité minimale* présentées dans cette section.

Définition 6.12. *On dit qu'une statistique S est exhaustive si la loi conditionnelle de $\mathbf{X} = (X_1, \dots, X_n)$ sachant S ne dépend pas de θ .*

Cette définition correspond intuitivement au fait que toute l'information sur θ contenue dans \mathbf{X} est déjà contenue dans S .

Exemple 6.8. Considérons le modèle statistique $(\mathbf{N}, (\mathcal{P}(\theta))_{\theta \in \mathbf{R}_+^*})$ et vérifions que la statistique $S = \sum_{i=1}^n X_i$ est exhaustive. Puisque X_1, \dots, X_n sont indépendante de loi $\mathcal{P}(\theta)$, $S \sim \mathcal{P}(n\theta)$. On a donc pour $k_1, \dots, k_n \in \mathbf{N}$ tels que $k_1 + \dots + k_n = k$:

$$\mathbf{P}_\theta[X_1 = k_1, \dots, X_n = k_n | S = k] = \frac{\prod_{i=1}^n \left(e^{-\theta} \frac{\theta^{k_i}}{k_i!} \right)}{e^{-n\theta} \frac{(n\theta)^k}{k!}} = \frac{e^{-n\theta} k! \prod_{i=1}^n \theta^{k_i}}{e^{-n\theta} n^k \theta^k \prod_{i=1}^n k_i!} = \frac{k!}{n^k \prod_{i=1}^n k_i!}$$

ce qui montre que S est exhaustive.

Le résultat suivant, que nous admettrons, permet de caractériser les statistiques exhaustives.

Théorème 6.3 (de factorisation). *Une statistique $S = S(X_1, \dots, X_n)$ est exhaustive si, et seulement, si la vraisemblance $L(\cdot; \theta)$ de $\mathbf{X} = (X_1, \dots, X_n)$ admet, pour tout $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$, une factorisation de la forme :*

$$L(\mathbf{x}; \theta) = \varphi(S(\mathbf{x}), \theta)h(\mathbf{x}).$$

Exemple 6.8 (suite). Dans le cadre du modèle statistique $(\mathbf{N}, (\mathcal{P}(\theta))_{\theta \in \mathbf{R}_+^*})$, en considérant la statistique $S = \sum_{i=1}^n X_i$, on a :

$$\begin{aligned} L(k_1, \dots, k_n; \theta) &= \prod_{i=1}^n \left(e^{-\theta} \frac{\theta^{k_i}}{k_i!} \right) = e^{-n\theta} \theta^{\sum_{i=1}^n k_i} \times \prod_{i=1}^n (k_i!)^{-1} \\ &= e^{-n\theta} \theta^{S(k_1, \dots, k_n)} \times \prod_{i=1}^n (k_i!)^{-1} \end{aligned}$$

ce qui est une factorisation de la forme souhaitée et montre, à nouveau, l'exhaustivité de S pour ce modèle.

Ce que nous avons observé dans l'exemple précédent s'étend à toute loi de la famille exponentielle.

Proposition 6.6. *Soit un modèle de la famille exponentielle avec un paramètre de dimension k . Alors, avec les notations de la Définition 3.7, la statistique :*

$$\left(\sum_{i=1}^n d_1(X_i), \dots, \sum_{i=1}^n d_k(X_i) \right)$$

est exhaustive.

Preuve : On a pour une telle loi :

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \prod_{j=1}^n f_{X_1}(x_j, \theta) = \prod_{j=1}^n a(\theta)b(x_j) \exp \left(\sum_{i=1}^k c_i(\theta)d_i(x_j) \right) \\ &= a(\theta)^n \prod_{j=1}^n b(x_j) \exp \left(\sum_{i=1}^k c_i(\theta) \sum_{j=1}^n d_i(x_j) \right) \end{aligned}$$

ce qui permet de conclure par le Théorème de factorisation 6.3. □

Remarque 6.13. On peut montrer que pour les modèles réguliers, l'existence d'une statistique exhaustive de même dimension que le paramètre inconnu et appartenant à la classe des familles exponentielles.

Définition 6.13. *Une statistique S est dite totale si $h(S) = 0$ \mathbf{P}_θ -p.s pour tout $\theta \in \Theta$ dès que $\mathbf{E}_\theta[|h(S)|] < +\infty$ et $\mathbf{E}_\theta[h(S)] = 0$ \mathbf{P}_θ -p.s pour tout $\theta \in \Theta$.*

Une statistique exhaustive S est dite minimale si pour toute statistique exhaustive T , il existe une fonction mesurable g telle que $S = g(T)$.

Une statistique exhaustive minimale est donc un bon résumé de l'information contenue dans un échantillon puisqu'elle est exhaustive donc contient toute l'information nécessaire et qu'elle la résume puisque toute statistique exhaustive suffit à la retrouver. Ainsi, si un estimateur de θ est une statistique exhaustive minimale, nous le considérerons pertinent. Le fait qu'une statistique soit minimale est généralement délicat à montrer en utilisant directement la définition puisque ceci présuppose de connaître toutes les statistiques exhaustives du modèle. Le résultat suivant donne un moyen plus simple pour montrer qu'une statistique exhaustive est minimale.

Proposition 6.7. *Si une statistique exhaustive est totale alors elle est minimale.*

Exercice 6.5. Dans le cadre du modèle statistique $(\mathbf{N}, (\mathcal{P}(\theta))_{\theta \in \mathbf{R}_+^*})$ de l'Exemple 6.8, montrer que la statistique $S = \sum_{i=1}^n X_i$ est minimale.

Pour conclure cette section, notons qu'il est possible d'améliorer (au sens du risque quadratique) un estimateur T en utilisant une statistique exhaustive S . L'approche classique repose sur l'utilisation de l'espérance conditionnelle sachant S et le Théorème de Rao-Blackwell et dépasse le cadre de ce cours. Elle affirme que si T est de plus de carré intégrable, $\mathbf{E}[T|S]$ est préférable à T . L'intuition est que la statistique exhaustive S a résumé l'information et toute l'information contenue dans l'échantillon nécessaire à l'estimation et l'estimateur résultant de ce conditionnement est « moins parasité » par des informations superflues donc possède une moins grande variabilité que T . Nous ne développerons pas plus ce sujet et renvoyons le lecteur à la Section VIII.5.2 de [7] ou à la Section 4.3.2 de [8].

6.3.6 Analyse asymptotique

Pour terminer ce chapitre, nous donnons quelques éléments sur l'analyse asymptotique de suites d'estimateurs T_n basés sur un échantillon (X_1, \dots, X_n) dont la taille n tend vers $+\infty$. Par abus de langage, on parlera d'analyse asymptotique de l'estimateur T_n au lieu de celle de la suite d'estimateurs $(T_n)_{n \in \mathbf{N}^*}$. Insistons sur le fait que cette approche sera utile pour l'estimation par intervalle de confiance du Chapitre 8 et ne sera valable que si n est suffisamment grand.

Dans le cadre d'un estimateur obtenu par la méthode des moments, la proposition suivante est une conséquence de la LGN et du TCL.

Proposition 6.8. *Soit φ une fonction telle que, pour tout $\theta \in \Theta$, $\varphi(X_1)$ est de carré intégrable sous \mathbf{P}_θ et $\mathbf{E}_\theta[\varphi(X_1)] = \theta$. Alors, l'estimateur des moments $\hat{\theta}_n^M = n^{-1} \sum_{i=1}^n \varphi(X_i)$ est convergent et vérifie, pour tout θ , sous \mathbf{P}_θ :*

$$\sqrt{n} (\hat{\theta}_n^M - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{V}_\theta[\varphi(X_1)]).$$

Remarque 6.14. On dit que $\hat{\theta}^M$ est asymptotiquement normal de (matrice de covariance/variance $\mathbf{V}_\theta[\varphi(X_1)]$).

Ce résultat admet dans le cadre d'un estimateur obtenu par la méthode du maximum de vraisemblance, un analogue plus délicat à démontrer.

Théorème 6.4. *Supposons le modèle régulier et identifiable et notons $I(\theta)$ l'information de Fisher du modèle. Alors, l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV}$ vérifie, pour tout θ , sous \mathbf{P}_θ :*

$$\sqrt{n} (\hat{\theta}_n^{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\theta)^{-1}).$$

Dans le deux cas, on pourra considérer que $\sqrt{n}(\hat{\theta}_n^{MV} - \theta)$ est proche d'une loi normale si n est suffisamment grand.

Chapitre 7

Tests d'hypothèses

Comme nous l'avons fait dans le cadre de l'estimation, nous nous concentrons ici principalement sur des tests *paramétriques*; en fait, même sur les tests paramétriques pour un paramètre de dimension 1 ($\Theta \subset \mathbf{R}$ pour ce chapitre). Toutefois, certains tests *non-paramétriques*, à l'instar des tests du Khi-2, sont d'un usage si courant qu'il est impossible de ne pas les présenter. Nous le ferons brièvement dans la Section 7.5.

7.1 Cadre et généralités sur tests d'hypothèses

On considère un modèle statistique paramétrique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ et un échantillon (i.i.d.) $\mathbf{X} = (X_1, \dots, X_n)$. Le paramètre θ de la loi mère étant dans $\Theta \subset \mathbf{R}$, on choisit deux sous-ensembles Θ_0 et Θ_1 non vides et disjoints de Θ ; quitte à modifier Θ on peut supposer que Θ_0 et Θ_1 forment une partition non triviale de Θ . À partir de Θ_0 et Θ_1 , on formule deux hypothèses : l'hypothèse nulle $H_0 : \theta \in \Theta_0$ et l'hypothèse alternative $H_1 : \theta \in \Theta_1$. Lorsque Θ_i est un singleton, on dit que l'hypothèse H_i est *simple*; sinon, on dit quelle est *composite*. Un test est une procédure, qui à partir d'une réalisation $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ de $\mathbf{X} = (X_1, \dots, X_n)$ (une observation), a pour objectif de déterminer laquelle des hypothèses H_0 et H_1 est vérifiée.

Définition 7.1. Une fonction de test pur ou fonction de test déterministe est une application $\varphi : \mathcal{X}^n \rightarrow \{0, 1\} \equiv \{H_0, H_1\}$. La région $R^c = \varphi^{-1}(\{0\}) = \{\mathbf{x} \in \mathcal{X}^n : \varphi(\mathbf{x}) = 0\}$ est la région d'acceptation de H_0 du test fondé sur φ . La région $R = \varphi^{-1}(\{1\})$ sur laquelle on rejette H_0 (donc « accepte » H_1) est appelée région critique ou de rejet du test.

Définition 7.2. Une fonction de test aléatoire ou fonction de test stochastique est une application $\varphi : \mathcal{X}^n \rightarrow [0, 1]$. La région $\varphi^{-1}(\{0\})$ est la région d'acceptation de H_0 du test fondé sur φ . La région $R = \varphi^{-1}(\{1\})$ est la région critique ou de rejet du test. La région $H = \varphi^{-1}(]0, 1])$ est la région d'hésitation du test. Pour $\mathbf{x} \in \mathcal{X}^n$, $\varphi(\mathbf{x})$ s'interprète alors comme la probabilité de rejeter H_0 .

Un test aléatoire fait donc appel à un aléa extérieur (*via* la valeur de $\varphi(\mathbf{x})$) à l'observation pour rendre sa décision; il s'agirait de regarder la réalisation d'une variable de Bernoulli de paramètre $\varphi(\mathbf{x})$ pour décider de rejeter H_0 si, et seulement si, celle-ci vaut 1. Il paraît donc par nature hasardeux et peu fondé. L'intérêt de ces tests est essentiellement théorique, en particulier pour l'analyse de la qualité des tests d'hypothèses simples. Notons qu'un test pur est un test stochastique particulier.

7.1.1 Puissance d'un test et erreurs

Définition 7.3. On appelle fonction puissance d'un test (aléatoire) φ la fonction :

$$\begin{aligned} \rho_\varphi : \Theta &\longrightarrow [0, 1] \\ \theta &\longmapsto \mathbf{E}_\theta[\varphi] = \mathbf{E}_\theta[\varphi(X_1, \dots, X_n)] \end{aligned}$$

donnant la probabilité de rejeter l'hypothèse nulle H_0 .

Remarque 7.1. Si φ est un test déterministe, on a $\varphi(\mathbf{x}) = \mathbf{1}_{\mathbf{x} \in R}$ et donc

$$\rho_\varphi(\theta) = \mathbf{E}_\theta[\varphi(\mathbf{X})] = \mathbf{E}_\theta[\mathbf{1}_{\mathbf{X} \in R}] = \mathbf{P}_\theta[\mathbf{X} \in R] = \mathbf{P}_\theta[\varphi = 1].$$

Lors de la réalisation d'un test, il est possible de commettre une erreur de deux façons différentes : soit on rejette à tort H_0 alors qu'elle était vraie, soit on accepte à tort H_0 alors qu'elle était fausse. Ceci conduit à la définition suivante :

Définition 7.4. On appelle erreur ou risque de première espèce la restriction de ρ_φ à Θ_0 . On appelle erreur ou risque de deuxième espèce la restriction de $1 - \rho_\varphi$ à Θ_1 .

Remarque 7.2.

1. Si φ est un test pur (déterministe), l'erreur de première espèce est donc la probabilité $\mathbf{P}_\theta[\mathbf{X} \in R]$ que l'observation $\mathbf{X} = (X_1, \dots, X_n)$ soit dans la région critique (donc de rejeter H_0) alors que $\theta \in \Theta_0$. L'erreur de deuxième espèce est donc la probabilité $1 - \rho_\varphi(\theta) = \mathbf{P}_\theta[\mathbf{X} \notin R]$ que \mathbf{X} soit hors de la région critique (donc d'« accepter » – ou plutôt de pas rejeter – H_0) alors que $\theta \in \Theta_1$.
2. Dans la définition de l'erreur de deuxième espèce, certains auteurs préfèrent restreindre $1 - \rho_\varphi$ à Θ_0^c plutôt qu'à Θ_1 sans supposer que Θ_0 et Θ_1 forment une partition de Θ . L'erreur de deuxième espèce s'interprète alors pour un test pur comme la probabilité d'« accepter » H_0 alors que $\theta \notin \Theta_0$.

On cherchera naturellement à minimiser les erreurs de première et deuxième espèce. La minimisation simultanée des deux erreurs étant en général impossible, on choisira par convention de minimiser en priorité l'erreur de première espèce. Bien que H_0 et H_1 jouent des rôles symétriques dans la définition d'un test, la définition des erreurs et cette convention dissymétrisent la situation. Ainsi, dans la pratique, cette dissymétrie induit un choix des hypothèses adapté au contexte.

7.1.2 Niveau et seuil d'un test

Définition 7.5. Soit $\alpha \in [0, 1]$. On dit qu'un test φ est de niveau (resp. seuil) α pour H_0 si

$$\sup_{\theta \in \Theta_0} \rho_\varphi(\theta) = \sup_{\theta \in \Theta_0} \mathbf{E}_\theta[\varphi] = \alpha \quad (\text{resp. } \leq \alpha).$$

On dit qu'il est α -semblable si pour tout $\theta \in \Theta_0$, on a $\rho_\varphi(\theta) = \alpha$.

Les seuils ou niveaux des tests permettent, en particulier, de choisir quel test et région critique utiliser, ce que nous développerons dans la Section 7.3.

7.1.3 Statistique de test

Définition 7.6. Soit $A \subset \mathbf{R}$. Supposons que la région critique d'un test pur s'écrive sous la forme :

$$R = \{\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n : \zeta(\mathbf{x}) \in A\}$$

pour un certaine fonction $\zeta : \mathcal{X}^n \rightarrow \mathbf{R}$. Alors, la statistique $\zeta = \zeta(X_1, \dots, X_n)$ est appelée une statistique de test. On parle de test basé sur la statistique ζ .

Remarque 7.3. On choisira les statistiques de test de façon à ce que la région critique s'écrive simplement, par exemple de façon unilatérale :

$$R = \{\zeta \geq a\} := \{\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n : \zeta(\mathbf{x}) \in [a, +\infty[]\}$$

ou $R = \{\zeta \leq a\}$ ou bilatérale : $R = \{a \leq \zeta \leq b\}$ ou $R = \{a \leq \zeta \text{ ou } \geq b\}$.

7.1.4 p -valeur

La p -valeur permet de bien quantifier le risque pris et rejetant H_0 . En effet, elle correspond à la probabilité de rejeter H_0 à tort en évaluant la probabilité pour que, sous H_0 , la statistique de test ζ dépasse la valeur observée ζ^{obs} pour un test pur de région critique de la forme $R = \{\zeta \geq a\}$, plus généralement la probabilité pour que ζ^{obs} soit « aberrante » sous H_0 . Elle se définit plus formellement comme suit.

Définition 7.7. Supposons que, pour une statistique de test ζ , pour tout $\alpha \in]0, 1[$, la région critique s'écrivant sous la forme $R = \{\zeta \in A_\alpha\}$, pour un certain $A_\alpha \subset \mathbf{R}$ définisse un test pur de niveau α .

Ayant observé $\mathbf{x}^{\text{obs}} = (x_1^{\text{obs}}, \dots, x_n^{\text{obs}})$ et évalué ζ en \mathbf{x}^{obs} ($\zeta^{\text{obs}} := \zeta(\mathbf{x}^{\text{obs}})$), la p -valeur (ou p -value) est plus petit niveau

$$p - \text{val} = \inf\{\alpha > 0 : \zeta^{\text{obs}} \in A_\alpha\}$$

pour lequel on rejette H_0 .

Si la région critique est de la forme $R = \{\zeta \geq c_\alpha\}$, notons que par les définitions, on a :

$$p - \text{val} = \inf\{\alpha > 0 : \zeta^{\text{obs}} \in A_\alpha\} = \inf\left\{\sup_{\theta \in \Theta_0} \mathbf{P}_\theta[\zeta \geq c_\alpha] : \zeta^{\text{obs}} \geq c_\alpha\right\} = \sup_{\theta \in \Theta_0} \mathbf{P}_\theta[\zeta \geq \zeta^{\text{obs}}].$$

Si on rejette H_0 au seuil α , $\zeta^{\text{obs}} \geq c_\alpha$ donc $p - \text{val} \leq \alpha$. Réciproquement, si $p - \text{val} = \alpha$ la région $\{\zeta \geq \zeta^{\text{obs}}\} \ni \zeta^{\text{obs}}$ est de niveau α et si $p - \text{val} < \alpha$, $\zeta^{\text{obs}} > c_\alpha$; dans les deux cas on rejette H_0 au seuil α .

On retiendra que l'on rejette H_0 au seuil α ssi $p - \text{val} \leq \alpha$.

7.2 Construction de tests

7.2.1 Tests du rapport de vraisemblance

Comme dans le cadre de l'estimation, les vraisemblances et leurs propriétés conduisent à des méthodes de construction de tests très généralement applicables et efficaces. Nous l'introduisons tout d'abord dans le cadre très simple du test d'une hypothèse simple contre une hypothèse simple.

Test du rapport de vraisemblance pour deux hypothèses simples

On considère les hypothèses $H_0 : \theta = \theta_0$ ($\theta \in \Theta_0 = \{\theta_0\}$) et $H_1 : \theta = \theta_1$ ($\theta \in \Theta_1 = \{\theta_1\}$), $\theta_1 \neq \theta_0$ et une vraisemblance L dans un modèle identifiable ($\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_0}$). Il est intuitivement clair que si H_0 est fautive, une observation « typique » \mathbf{x}^{obs} devrait conduire à de « petites » valeurs de $L(\mathbf{x}^{\text{obs}}; \theta_0)$ et de « grandes » valeurs de $L(\mathbf{x}^{\text{obs}}; \theta_1)$ donc à un rapport de vraisemblance $l(\mathbf{x}^{\text{obs}}) = L(\mathbf{x}^{\text{obs}}; \theta_0)/L(\mathbf{x}^{\text{obs}}; \theta_1)$ faible. Ainsi, on choisira l comme statistique de test et on obtiendra la région critique sous la forme $R = \{l \leq c_\alpha\}$ pour un certain c_α à déterminer en fonction du risque de première espèce. Un tel test est appelé *test du rapport de vraisemblance pour deux hypothèses simples* (RV).

Exemple 7.1. Considérons le modèle $(\mathbf{R}_+, (\mathcal{E}(\theta))_{\theta > 0})$ et supposons que l'on souhaite tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ ($\theta_1 > \theta_0$) en utilisant un échantillon de taille n .

Les vraisemblances s'écrivant sous la forme :

$$L(x_1, \dots, x_n; \theta_j) = \prod_{i=1}^n (\theta_j e^{-\theta_j x_i} \mathbf{1}_{x_i \geq 0}) = \theta_j^n \exp\left(-\theta_j \sum_{i=1}^n x_i\right) \mathbf{1}_{x_1, \dots, x_n \geq 0}$$

on obtient que le rapport de vraisemblance est donné par

$$l(x_1, \dots, x_n) = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left(-(\theta_0 - \theta_1) \sum_{i=1}^n x_i\right) \mathbf{1}_{x_1, \dots, x_n \geq 0}.$$

Ainsi, en posant $g(s) = \left(\frac{\theta_0}{\theta_1}\right)^n \exp(-(\theta_0 - \theta_1)s)$ et $S = S(X_1, \dots, X_n) = \sum_{i=1}^n X_i$, la statistique de test s'écrit :

$$l(\mathbf{X}) = l(X_1, \dots, X_n) = g(S) \mathbf{1}_{X_1, \dots, X_n \geq 0}.$$

Remarquons que, sous H_0 ($\theta = \theta_0$), on a $\mathbf{1}_{X_1, \dots, X_n \geq 0} = 1$ p.s. et que $S \sim \Gamma(n, \theta_0)$ (voir Exercice A.2). Remarquons également que g est bijective et que sa réciproque (facilement explicitable) est croissante puisque $\theta_0 < \theta_1$. Il s'ensuit que pour que le test soit de niveau $\alpha \in]0, 1[$, la région critique $R_\alpha =]-\infty, c_\alpha]$ doit vérifier pour $\theta = \theta_0$:

$$\mathbf{P}_{\theta_0} [l(\mathbf{X}) \in R_\alpha] = \mathbf{P}_{\theta_0} [g(S) \leq c_\alpha] = \mathbf{P}_{\theta_0} [S \leq g^{-1}(c_\alpha)].$$

En notant q_α le quantile d'ordre α de $\Gamma(n, \theta_0)$ (facilement accessible avec un logiciel de statistique comme R) vérifiant puisque cette loi est continue :

$$\mathbf{P}_{\theta_0} [S \leq q_\alpha] = \alpha$$

on obtient par identification que $g^{-1}(c_\alpha) = q_\alpha$ soit $c_\alpha = g(q_\alpha)$. On conclue ainsi que la région critique du test de niveau α est $R_\alpha = \{\mathbf{x} \in \mathcal{X}^n : S(\mathbf{x}) \in]-\infty, g(q_\alpha)]\}$.

Remarque 7.4.

1. Dans l'exemple précédent, on voit que l'on a déterminé la loi de la statistique de test (via celle de la statistique exhaustive S) sous H_0 pour contrôler l'erreur de première espèce. Ce contrôle se passe toujours de cette façon. Celui de l'erreur de deuxième espèce passe par la détermination de la loi de la statistique de test sous H_1 . La simplicité de la mise en œuvre du test de rapport de vraisemblance est partagée par de nombreux modèles, en particulier ceux basés sur une famille de lois de la classe exponentielle comme nous le verrons dans la Section 7.3 (voir aussi Théorème 7.1).

2. Par abus de langage permettant d'alléger les choses, on confondra parfois la zone rejet et l'ensemble permettant de la déterminer grâce à la statistique de test. Dans l'exemple précédent, on dira qu'en utilisant la statistique de test S la zone de rejet est $]-\infty, g(c_\alpha)]$.

7.2.2 Tests du rapport de vraisemblance généralisé

Lorsque les hypothèses sont composites, c'est-à-dire que Θ_0 et Θ_1 ne sont pas des singletons, la comparaison établie dans la Sous-section précédente n'est plus valable mais pourrait s'étendre en utilisant la statistique $l = l(\mathbf{X})$ définie par

$$l(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta_1} L(\mathbf{x}; \theta)}.$$

Le même raisonnement conduirait à rejeter H_0 pour des petites valeurs de l . Observons que l'on retrouverait même exactement la même statistique que plus haut lorsque les hypothèses sont simples. Toutefois, on va préférer l'usage de la statistique $\ell = \ell(\mathbf{X})$ définie pour $\mathbf{x} \in \mathcal{X}^n$ par

$$\ell(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta} L(\mathbf{x}; \theta)}.$$

On appellera cette statistique *statistique du rapport de vraisemblance généralisé* (RVG). Il y a deux raisons principales à préférer ℓ à l . La première est que $\ell \leq 1$ p.s. et la seconde qu'elle est liée à l'estimateur du maximum de vraisemblance (non restreint) $\hat{\theta}^{MV}$. Plus précisément, on voit que le dénominateur $\sup_{\theta \in \Theta} L(\mathbf{x}; \theta)$ n'est autre que $L(\mathbf{x}; \hat{\theta}^{MV})$. De même, le numérateur est lié à l'estimateur du maximum de vraisemblance restreint à Θ_0 , $\hat{\theta}_0^{MV}$, puisque l'on a $\sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta) = L(\mathbf{x}; \hat{\theta}_0^{MV})$.

Si l'on dispose d'une statistique exhaustive, le résultat suivant, conséquence du Théorème de factorisation 6.3 permet de simplifier le calcul de la statistique du RVG.

Théorème 7.1. *Si S est une statistique exhaustive pour θ dans le modèle $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ de vraisemblance L , alors, pour tout $\Theta_0 \subset \Theta$ fixé, la statistique du RVG ℓ se factorise au travers de S de la manière suivante : il existe une fonction λ telle que pour tout $\mathbf{x} \in \mathcal{X}^n$, on a*

$$\ell(\mathbf{x}) = \lambda(S(\mathbf{x})).$$

Preuve : Le Théorème de factorisation 6.3 assure que l'existence de fonctions φ et h telles que pour tout $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$, on a :

$$L(\mathbf{x}; \theta) = \varphi(S(\mathbf{x}), \theta)h(\mathbf{x}).$$

Ainsi, pour tout $\mathbf{x} \in \mathcal{X}^n$, on a :

$$\ell(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta} L(\mathbf{x}; \theta)} = \frac{\sup_{\theta \in \Theta_0} \varphi(S(\mathbf{x}), \theta)h(\mathbf{x})}{\sup_{\theta \in \Theta} \varphi(S(\mathbf{x}), \theta)h(\mathbf{x})} = \frac{\sup_{\theta \in \Theta_0} \varphi(S(\mathbf{x}), \theta)}{\sup_{\theta \in \Theta} \varphi(S(\mathbf{x}), \theta)},$$

ce qui donne le résultat en posant :

$$\lambda(s) = \frac{\sup_{\theta \in \Theta_0} \varphi(s, \theta)}{\sup_{\theta \in \Theta} \varphi(s, \theta)}.$$

□

Exemple 7.2. Considérons le modèle $(\mathbf{R}_+, (\mathcal{E}(\theta))_{\theta>0})$ et supposons que l'on souhaite tester $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$ en utilisant un échantillon de taille n .

Vérifier que la statistique du RVG est donnée par :

$$\ell(\mathbf{X}) = \left(\frac{\theta_0 S(\mathbf{X})}{n} \right)^n e^{n - \theta_0 S(\mathbf{X})} \mathbf{1}_{n - \theta_0 S > 0} + \mathbf{1}_{n - \theta_0 S \leq 0}$$

où $S = S(\mathbf{X})$ est la statistique exhaustive pour θ dans ce modèle définie par $S(\mathbf{X}) = \sum_{i=1}^n X_i$.

7.2.3 Tests bayésiens

Considérons sur $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ deux hypothèses à tester $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_0^c$ en utilisant la loi *a priori* ν sur Θ (voir Section 6.2.4). Rappelons que dans le contexte bayésien, l'inférence se fait, après avoir observé $\mathbf{x} = (x_1, \dots, x_n)$, *via* la loi *a posteriori* $\mathbf{P}_\mathbf{x}$ contenant l'information apportée par \mathbf{x} et celle apportée par ν (généralement de faible impact). Dans ce contexte $\mathbf{P}_\mathbf{x}[\theta \in \Theta_i]$ représente la probabilité pour que H_i soit vraie sachant que l'on a observé \mathbf{x} . Elles induisent naturellement des tests, appelés *test bayésiens*, dans lesquels le contrôle de l'erreur de première espèce, et donc la détermination de la zone de rejet, se fait à travers de la probabilité $\mathbf{P}_\mathbf{x}[\theta \in \Theta_0^c]$. Ainsi, pour obtenir un test de seuil $\alpha \in]0, 1[$, il faut fonder le test sur la région critique :

$$R_\alpha = \{\mathbf{x} \in \mathcal{X}^n : \mathbf{P}_\mathbf{x}[\theta \in \Theta_0^c] \geq 1 - \alpha\}.$$

Exemple 6.6 (suite).

Reprenons le contexte de l'Exemple 6.6, c'est-à-dire du modèle $(\mathbf{R}, (\mathcal{N}(\theta, 1)_{\theta \in \mathbf{R}}))$ avec comme loi *a priori* sur $\Theta = \mathbf{R}$ la loi $\nu = \mathcal{N}(m, \sigma^2)$ (m et σ^2 connus). Rappelons que la loi *a posteriori* $\mathbf{P}_\mathbf{x}$ régissant θ (qui est dans ce contexte bayésien vu comme une variable aléatoire) est, en ayant observé $\mathbf{x} = (x_1, \dots, x_n)$, la loi normale de moyenne

$$\mathbf{m} = g(\bar{x}_n) := \frac{1}{1 + n\sigma^2}m + \frac{n\sigma^2}{1 + n\sigma^2}\bar{x}_n \quad \text{avec } \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

et de variance

$$\mathbf{v} = \frac{\sigma^2}{1 + n\sigma^2}.$$

Notons que la dépendance de $\mathbf{P}_\mathbf{x}$ en l'observation \mathbf{x} ne se fait qu'au travers de \mathbf{m} *via* \bar{x}_n et que g est bijective et de réciproque croissante.

Notons également, que sous $\mathbf{P}_\mathbf{x}$,

$$Z = \frac{\theta - \mathbf{m}}{\sqrt{\mathbf{v}}} \sim \mathcal{N}(0, 1).$$

Ainsi, si l'on souhaite tester l'hypothèse $H_0 : \theta > \theta_0$ au seuil $\alpha \in]0, 1[$, le test bayésien fonde, avec les notations précédentes, la décision sur la zone de rejet :

$$\begin{aligned} R_\alpha &= \{\mathbf{x} \in \mathcal{X}^n : \mathbf{P}_\mathbf{x}[\theta \leq \theta_0] \geq 1 - \alpha\} \\ &= \left\{ \mathbf{x} \in \mathcal{X}^n : \mathbf{P}_\mathbf{x} \left[\frac{\theta - \mathbf{m}}{\sqrt{\mathbf{v}}} \leq \frac{\theta_0 - \mathbf{m}}{\sqrt{\mathbf{v}}} \right] \geq 1 - \alpha \right\} \\ &= \left\{ \mathbf{x} \in \mathcal{X}^n : \mathbf{P} \left[Z \leq \frac{\theta_0 - \mathbf{m}}{\sqrt{\mathbf{v}}} \right] \geq 1 - \alpha \text{ pour } Z \sim \mathcal{N}(0, 1) \right\}. \end{aligned}$$

En notant $q_{1-\alpha}$ le quantile d'ordre $1 - \alpha$ de $\mathcal{N}(0, 1)$ (facilement accessible avec un logiciel de statistique comme R) vérifiant puisque cette loi est continue :

$$\mathbf{P} [Z \leq q_{1-\alpha}] = 1 - \alpha$$

on obtient que

$$\begin{aligned} R_\alpha &= \left\{ \mathbf{x} \in \mathcal{X}^n : \mathbf{P} \left[Z \leq \frac{\theta_0 - \mathbf{m}}{\sqrt{\mathbf{v}}} \right] \geq \mathbf{P} [Z \leq q_{1-\alpha}] \text{ pour } Z \sim \mathcal{N}(0, 1) \right\} \\ &= \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{\theta_0 - \mathbf{m}}{\sqrt{\mathbf{v}}} \geq q_{1-\alpha} \right\} \quad (\text{par croissance de la fonction de répartition}) \\ &= \{ \mathbf{x} \in \mathcal{X}^n : \mathbf{m} \leq \theta_0 - \sqrt{\mathbf{v}} q_{1-\alpha} \} \\ &= \{ \mathbf{x} \in \mathcal{X}^n : g(\bar{x}_n) \leq \theta_0 - \sqrt{\mathbf{v}} q_{1-\alpha} \} \\ &= \left\{ \mathbf{x} \in \mathcal{X}^n : \bar{x}_n \leq g^{-1} (\theta_0 - \sqrt{\mathbf{v}} q_{1-\alpha}) \right\} \\ &= \left\{ \mathbf{x} \in \mathcal{X}^n : \bar{x}_n \leq \frac{1 + n\sigma^2}{n\sigma^2} \left(\theta_0 - \sqrt{\mathbf{v}} q_{1-\alpha} \frac{1}{1 + n\sigma^2} m \right) \right\}. \end{aligned}$$

7.3 Comparaison et analyse des tests

7.3.1 Tests UPP et UPPSB

Puisque l'on cherche en priorité à contrôler l'erreur de première espèce, on ne retient, par convention consistant à suivre le *principe de Neyman*, que des test φ de niveau inférieur à un certain seuil $\alpha \in]0, 1[$ choisi *a priori* par l'utilisateur. Ensuite, on a intérêt à chercher parmi ces tests celui (ou ceux) d'erreur de seconde espèce la plus faible possible, autrement dit de puissance maximale.

Définition 7.8. On dit qu'un test φ est uniformément plus puissant au seuil α que φ' (de seuil α) s'il est de seuil α et si φ' est de puissance supérieure à celle de φ sur Θ_1 :

$$\rho_\varphi(\theta) = \mathbf{E}_\theta [\varphi] \geq \mathbf{E}_\theta [\varphi'] = \rho_{\varphi'}(\theta), \quad \text{pour tout } \theta \in \Theta_1,$$

autrement dit si son erreur de deuxième espèce est inférieure à celle de φ' .

On dit qu'un test φ est uniformément plus puissant au seuil α (UPP ou UMP - *uniformly most powerful*) s'il est de seuil α et s'il est uniformément plus puissant que tout autre test φ' de seuil α , autrement dit si son erreur de deuxième espèce est inférieure à celle de tout autre test de même seuil.

Remarque 7.5. On se rappellera que si φ est un test pur, alors $\rho_\varphi(\theta) = \mathbf{P}_\theta[X \in R]$. Ainsi, un test pur UPP au seuil α , rejette H_0 lorsqu'elle est fausse avec la probabilité la plus grande possible parmi les tests de seuils α .

La proposition suivant montre qu'un test UPP au seuil α est nécessairement de niveau α .

Proposition 7.1. Soit $0 \leq \alpha' < \alpha \leq 1$ et φ' un test de niveau α' . Alors, il existe un test φ de niveau α uniformément plus puissant que φ' .

Preuve :

Il suffit de considérer le test défini par :

$$\varphi(\mathbf{x}) = \varphi'(\mathbf{x}) + \frac{\alpha - \alpha'}{1 - \alpha'}(1 - \varphi'(\mathbf{x})).$$

Les détails sont laissés au lecteur en exercice. □

Nous verrons qu'il n'existe pas toujours de tests UPP – en fait, ils n'existent que dans des situations particulières que nous étudierons plus loin. De façon analogue à ce que nous avons fait dans le cadre de l'estimation ponctuelle, on peut se restreindre à chercher des tests de puissance maximale dans des classe plus restreintes de tests, par exemple les tests *sans biais*.

Définition 7.9. *On dit qu'un test φ est sans biais au seuil α*

$$\sup_{\theta \in \Theta_0} \mathbf{E}_\theta [\varphi] \leq \alpha \leq \inf_{\theta \in \Theta_1} \mathbf{E}_\theta [\varphi].$$

On dit qu'un test φ est uniformément plus puissant sans biais au seuil α (UPPSB) s'il est sans biais au seuil α et s'il est UPP que tout autre test φ' sans biais au seuil α .

7.3.2 Cas des tests entre deux hypothèses simples

On considère dans toute cette section un modèle pour lequel $\Theta = \{\theta_0, \theta_1\}$ et pour lequel on souhaite tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$.

Définition 7.10. *On appelle test de Neyman-Pearson tout test φ tel que :*

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } L(\mathbf{x}; \theta_1) > \kappa L(\mathbf{x}; \theta_0) \\ \gamma(x) & \text{si } L(\mathbf{x}; \theta_1) = \kappa L(\mathbf{x}; \theta_0) \\ 0 & \text{si } L(\mathbf{x}; \theta_1) < \kappa L(\mathbf{x}; \theta_0) \end{cases}$$

pour une certaine constante $\kappa > 0$ et $\gamma : \mathcal{X}^n \rightarrow [0, 1]$.

Remarque 7.6.

1. Si $L(\mathbf{x}; \theta_1) \neq \kappa L(\mathbf{x}; \theta_0)$ \mathbf{P}_{θ_0} -p.s. ou γ est constante égale à 0 ou 1, alors le test de Neyman-Pearson est pur.
2. En réécrivant, par exemple, $L(\mathbf{x}; \theta_1) < \kappa L(\mathbf{x}; \theta_0)$ sous la forme $L(\mathbf{x}; \theta_1)/L(\mathbf{x}; \theta_0) < \kappa$, on voit que les tests de Neyman-Pearson sont basé sur un rapport de vraisemblances et sont, en fait, essentiellement des tests du rapport de vraisemblance. On formalisera cette remarque par la suite. Notons que ce rapport est l'inverse de celui proposé dans la Section 7.2.1 ; il s'agit des choix standards faits pour conserver l'intuition d'une part et simplifier certaines écritures d'autre part.

Proposition 7.2 (Existence). *Pour tout $\alpha \in]0, 1[$, il existe un test de Neyman-Pearson au seuil α avec γ constante.*

Plus précisément, considérons la statistique du rapport de vraisemblance (simple) ℓ , définie sur \mathcal{X}^n par

$$\ell(\mathbf{x}) = \frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta_0)}.$$

Soient F sa fonction de répartition sous \mathbf{P}_{θ_0} et $q_{1-\alpha}$ son quantile d'ordre $1 - \alpha$ sous \mathbf{P}_{θ_0} . On a les cas suivants :

- si $F(q_{1-\alpha}) = 1 - \alpha$, on choisit $\kappa = q_{1-\alpha}$, $\gamma = 1$ et le test du rapport de vraisemblance de région critique $R = \{\ell \geq \kappa\}$ est un test de Neyman-Pearson (pur) de niveau α ;
- sinon, on choisit $\kappa = q_{1-\alpha}$,

$$\gamma = \frac{F(\kappa) - 1 + \alpha}{F(\kappa) - \lim_{t \rightarrow \kappa, t < \kappa} F(t)} \in]0, 1]$$

et le test du rapport de vraisemblance de région critique $R = \{\ell > \kappa\}$ est un test de Neyman-Pearson (aléatoire) de niveau α ;

Preuve :

Avec les notations de la proposition, on pose $\varphi(\mathbf{x}) = \mathbf{1}_{l(\mathbf{x}) > \kappa} + \gamma \mathbf{1}_{l(\mathbf{x}) = \kappa}$. Il est facile de voir que $\gamma \in [0, 1]$ par définition des quantiles et croissance de la fonction de répartition. Il reste à vérifier que φ est de niveau α :

$$\begin{aligned} \mathbf{E}_{\theta_0}[\varphi(\mathbf{X})] &= \mathbf{P}_{\theta_0}[l(\mathbf{X}) > \kappa] + \gamma \mathbf{P}_{\theta_0}[l(\mathbf{X}) = \kappa] \\ &= 1 - F(\kappa) + \gamma(F(\kappa) - \lim_{t \rightarrow \kappa, t < \kappa} F(t)) = \alpha. \end{aligned}$$

□

Le Théorème suivant caractérise les test UPP pour le cas de deux hypothèses simples. Combiné avec la proposition précédente, il explique pourquoi les tests du rapports de vraisemblance se sont imposés dans ce cas.

Théorème 7.2 (Lemme de Neyman-Pearson). *Soit $\alpha \in]0, 1[$ et le problème du test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$.*

Alors, un test est un test de Neyman-Pearson de niveau α si, et seulement si, il est UPP au seuil α .

Preuve :

Soit φ un test de Neyman-Pearson de niveau α avec $\kappa > 0$ et $c \in [0, 1]$ et φ' un test de seuil α .

Remarquons que si $L(\mathbf{x}, \theta_1) > \kappa L(\mathbf{x}, \theta_0)$ alors, $\varphi(\mathbf{x}) = 1 \geq \varphi'(\mathbf{x})$ alors que si $L(\mathbf{x}, \theta_1) < \kappa L(\mathbf{x}, \theta_0)$, on a $\varphi(\mathbf{x}) = 0 \leq \varphi'(\mathbf{x})$. Ainsi, pour tout $\mathbf{x} \in \mathcal{X}^n$, on a $(L(\mathbf{x}, \theta_1) - \kappa L(\mathbf{x}, \theta_0))(\varphi(\mathbf{x}) - \varphi'(\mathbf{x})) \geq 0$ et donc

$$\int_{\mathcal{X}^n} (L(\mathbf{x}, \theta_1) - \kappa L(\mathbf{x}, \theta_0))(\varphi(\mathbf{x}) - \varphi'(\mathbf{x})) d \lambda_{\mathbf{x}} \geq 0.$$

Il s'ensuit que

$$\begin{aligned} \rho_{\varphi}(\theta_1) - \rho_{\varphi'}(\theta_1) &= \mathbf{E}_{\theta_1} [\varphi - \varphi'] \geq \kappa \mathbf{E}_{\theta_0} [\varphi - \varphi'] \\ &= \kappa (\mathbf{E}_{\theta_0} [\varphi] - \mathbf{E}_{\theta_0} [\varphi']) = \kappa (\alpha - \mathbf{E}_{\theta_0} [\varphi']) \\ &\geq 0. \end{aligned}$$

Réciproquement, si φ' est UPP il vient que les dernières inégalités du sens direct sont en fait des égalités. Or, $(L(\cdot, \theta_1) - \kappa L(\cdot, \theta_0))(\varphi(\cdot) - \varphi'(\cdot))$ est positive donc la Proposition 1.2 : (6) implique que $(L(\cdot, \theta_1) - \kappa L(\cdot, \theta_0))(\varphi(\cdot) - \varphi'(\cdot))$ est nulle p.p. puis que $\varphi = \varphi'$ p.p. sur l'ensemble $\{\mathbf{x} : L(\mathbf{x}, \theta_1) \neq \kappa L(\mathbf{x}, \theta_0)\}$. Il s'ensuit que φ' est un test de Neyman-Pearson. □

7.3.3 Modèles à rapport de vraisemblance monotone

Définition 7.11. *Un modèle statistique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta \subset \mathbf{R}})$ est dit à rapport de vraisemblance (strictement) monotone en une statistique S si pour tout $\theta < \theta'$ le rapport de vraisemblance se factorise en S via une fonction $h_{\theta, \theta'} : \mathbf{R} \rightarrow \mathbf{R} \cup \{\pm\infty\}$ (strictement) monotone :*

$$\frac{L(\mathbf{x}; \theta')}{L(\mathbf{x}; \theta)} = h_{\theta, \theta'}(S(\mathbf{x})), \quad \text{pour tout } \mathbf{x} \in \mathcal{X}^n.$$

Il est dit à rapport de vraisemblance croissant (resp. strictement croissant, décroissant, strictement décroissant) si $h_{\theta, \theta'}$ l'est.

Remarque 7.7. En fait, un test à rapport de vraisemblance décroissant est aussi un test à rapport de vraisemblance croissant et vice versa. Pour le voir, il suffit de changer S en $-S$ et $h_{\theta, \theta'}$ en $h_{\theta, \theta'} \circ (-\text{id})$. Dans la suite on ne considèrera donc que le cas à rapport de vraisemblance croissant.

Exemple 7.3. Si $(\mathbf{P}_\theta)_{\theta \in \Theta}$ est une famille exponentielle à paramètre de dimension un telle que la densité de \mathbf{P}_θ s'écrive sous la forme :

$$f(x, \theta) = a(\theta)b(x) \exp(c(\theta)d(x)), \quad \text{pour tous } \theta \in \Theta, x \in \mathcal{X},$$

avec a et b positives, on a que le rapport de vraisemblance s'écrit pour $\theta < \theta'$ et $\mathbf{x} \in \mathcal{X}^n$:

$$\begin{aligned} \frac{L(\mathbf{x}; \theta')}{L(\mathbf{x}; \theta)} &= \frac{a(\theta')^n \prod_{i=1}^n b(x_i) \exp(c(\theta') \sum_{i=1}^n d(x_i))}{a(\theta)^n \prod_{i=1}^n b(x_i) \exp(c(\theta) \sum_{i=1}^n d(x_i))} \\ &= \left(\frac{a(\theta')}{a(\theta)} \right)^n \exp((c(\theta') - c(\theta)) S(\mathbf{x})) \end{aligned}$$

avec $S = S(\mathbf{x}) = \sum_{i=1}^n d(x_i)$ la statistique *canonique* (ou *privilégiée*) de ce modèle. Ainsi, un tel modèle est à vraisemblance monotone pour S si, et seulement si, c est monotone.

La proposition suivante, immédiate, donne la forme de tests de Neyman-Pearson pour des modèles à rapport de vraisemblance croissant.

Proposition 7.3. *Soit un modèle statistique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta \subset \mathbf{R}})$ à rapport de vraisemblance croissant pour la statistique S .*

Si $\theta_0 < \theta_1$, on pose :

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } S(\mathbf{x}) > k \\ \gamma & \text{si } S(\mathbf{x}) = k \\ 0 & \text{si } S(\mathbf{x}) < k \end{cases}$$

et si $\theta_0 > \theta_1$, on pose :

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } S(\mathbf{x}) < k \\ \gamma & \text{si } S(\mathbf{x}) = k \\ 0 & \text{si } S(\mathbf{x}) > k \end{cases}$$

Alors φ est un test de Neyman-Pearson.

7.3.4 Cas des tests avec hypothèses composites

Extension du Lemme de Neyman-Pearson

Bien que ne s'appliquant qu'au cas de deux hypothèses simples, le Lemme de Neyman-Pearson admet l'extension suivante

Proposition 7.4. *Soient les hypothèses à tester $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$ dans le modèle statistique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ et φ un test de niveau α pour H_0 contre H_1 . S'il existe $\theta_0 \in \Theta_0$ tel que $\rho_\varphi(\theta_0) = \alpha$ et si pour tout $\theta_1 \in \Theta_1$, il existe $\kappa = \kappa_{\theta_1}$ tel que*

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } L(\mathbf{x}; \theta_1) > \kappa L(\mathbf{x}; \theta_0) \\ 0 & \text{si } L(\mathbf{x}; \theta_1) < \kappa L(\mathbf{x}; \theta_0) \end{cases}$$

alors, φ est UPP au seuil α pour H_0 contre H_1 .

Preuve : Soient un test φ' un test de niveau α pour H_0 contre H_1 et $\theta_1 \in \Theta_1$ arbitraire. Remarquons que φ un test de Neyman-Pearson de niveau α pour $H'_0 : \theta = \theta_0$ contre $H'_1 : \theta = \theta_1$. Ainsi, par le Lemme de Neyman-Pearson (Théorème 7.2), il est UPP au seuil α pour H'_0 contre H'_1 . Or, $\rho_{\varphi'}(\theta_0) \leq \sup_{\theta \in \Theta_0} \rho_{\varphi'}(\theta) \leq \alpha$ donc φ' est de seuil α pour H'_0 contre H'_1 . En particulier, φ est UPP que φ' au seuil α pour H'_0 contre H'_1 et donc $\rho_\varphi(\theta_1) \geq \rho_{\varphi'}(\theta_1)$. Le choix de $\theta_1 \in \Theta_1$ étant arbitraire, le résultat s'ensuit. □

Cas des tests unilatéraux

Test unilatéral de $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$

Théorème 7.3. *Soit un modèle statistique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ à rapport de vraisemblance strictement croissant en la statistique S . Alors, pour tout $\alpha \in]0, 1[$, il existe un test UPP au seuil α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$ de la forme*

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } S(\mathbf{x}) > k \\ \gamma & \text{si } S(\mathbf{x}) = k \\ 0 & \text{si } S(\mathbf{x}) < k \end{cases} .$$

Remarque 7.8. On peut écrire un théorème analogue pour le test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$ en intervertissant les symboles $<$ et $>$ dans la définition de φ .

Preuve : Pour $\theta_1 > \theta_0$ et on considère le test de Neyman-Pearson au seuil α

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } S(\mathbf{x}) > k \\ \gamma & \text{si } S(\mathbf{x}) = k \\ 0 & \text{si } S(\mathbf{x}) < k \end{cases}$$

de H_0 contre $H'_1 : \theta = \theta_1$. Comme $L(\mathbf{x}, \theta_1)/L(\mathbf{x}, \theta_0) = h_{\theta_0, \theta_1}(S(\mathbf{x}))$ avec h_{θ_0, θ_1} strictement croissante, il existe κ tel que

$$S(\mathbf{x}) > k \iff L(\mathbf{x}; \theta_1) > \kappa L(\mathbf{x}; \theta_0) \quad \text{et} \quad S(\mathbf{x}) < k \iff L(\mathbf{x}; \theta_1) < \kappa L(\mathbf{x}; \theta_0).$$

On conclue avec la Proposition 7.4. □

Test unilatéral de $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$

Théorème 7.4 (de Karlin-Rubin ou de Lehmann). *Soit un modèle statistique $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ à rapport de vraisemblance strictement croissant en la statistique S . Alors, pour tout $\alpha \in]0, 1[$, il existe un test UPP au seuil α de $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$ de la forme*

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } S(\mathbf{x}) > k \\ \gamma & \text{si } S(\mathbf{x}) = k \\ 0 & \text{si } S(\mathbf{x}) < k \end{cases} .$$

De plus, on a :

$$\sup_{\theta \leq \theta_0} \rho_\varphi(\theta) = \rho_\varphi(\theta_0) = \alpha .$$

Remarque 7.9. L'équation du niveau $\sup_{\theta \leq \theta_0} \rho_\varphi(\theta) = \rho_\varphi(\theta_0) = \alpha$ détermine k .

Preuve : On sait qu'il existe un test de φ tel que $\rho_\varphi(\theta_0) = \alpha$ de la forme

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } S(\mathbf{x}) > k \\ \gamma & \text{si } S(\mathbf{x}) = k \\ 0 & \text{si } S(\mathbf{x}) < k \end{cases} .$$

Si $\theta' < \theta''$, comme $L(\mathbf{x}, \theta'')/L(\mathbf{x}, \theta') = h_{\theta', \theta''}(S(\mathbf{x}))$ avec $h_{\theta', \theta''}$ strictement croissante, il s'agit d'après la Proposition 7.3 d'un test de Neyman-Pearson pour $H'_0 : \theta = \theta'$ contre $H'_1 : \theta = \theta''$. Par le Lemme de Neyman-Pearson (Théorème 7.2), il est UPP que tout test φ' de même seuil de H'_0 contre H'_1 . Il vient donc que pour tout test φ' de même seuil :

$$\rho_{\varphi'}(\theta') \leq \rho_{\varphi'}(\theta'') \implies \rho_{\varphi'}(\theta') \leq \rho_\varphi(\theta'') .$$

Choisissons $\theta' = \theta_0$ et $\theta'' > \theta_0$. Si φ' est un test de niveau α pour H_0 contre H_1 , alors $\rho_{\varphi'}(\theta_0) \leq \sup_{\theta} \rho_{\varphi'}(\theta) \leq \alpha = \rho_\varphi(\theta_0)$ donc $\rho_{\varphi'}(\theta'') \leq \rho_\varphi(\theta'')$ et φ est UPP que φ' pour H_0 contre H_1 .

Il reste à voir que φ de niveau α pour H_0 contre H_1 . Pour cela, on choisit $\theta' < \theta'' = \theta_0$ et φ' le test constant égal à $\rho_{\varphi'}(\theta')$. On a alors que $\rho_{\varphi'}(\theta') = \rho_{\varphi'}(\theta')$ (donc $\rho_{\varphi'}(\theta') \leq \rho_\varphi(\theta')$) donc $\rho_\varphi(\theta') = \rho_{\varphi'}(\theta_0) \leq \rho_\varphi(\theta_0)$. Ceci étant vrai pour tout $\theta' < \theta_0$ conclue en prenant le supremum sur $\theta' \leq \theta_0$:

$$\sup_{\theta \leq \theta_0} \rho_\varphi(\theta) \leq \rho_\varphi(\theta_0) = \alpha .$$

□

Remarque 7.10. Il découle de cette preuve, que si le modèle est identifiable, la fonction puissance d'un test de Neyman-Pearson entre deux hypothèses simples est strictement croissante.

Cas des tests bilatéraux

Il n'existe en général pas de test UPP au seuil α pour des hypothèses bilatérales. On se restreindra donc essentiellement aux modèles dans la classe exponentielle.

Cas $H_0 : \theta \leq \theta_1$ ou $\theta \geq \theta_2$ contre $H_1 : \theta \in]\theta_1, \theta_2[$ et $H_0 : \theta \neq \theta_0$ contre $H_1 : \theta = \theta_0$ On admet le résultat suivant.

Théorème 7.5. Soit $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta \subset \mathbf{R}})$ un modèle exponentielle (à paramètre de dimension 1) de vraisemblance

$$L(\mathbf{X}; \theta) = a(\theta)b(\mathbf{x}) \exp(c(\theta)S(\mathbf{x}))$$

avec c strictement croissante de sorte à ce que le modèle soit à rapport de vraisemblance strictement croissant en $T(\mathbf{x})$. On considère vouloir tester $H_0 : \theta \leq \theta_1$ ou $\theta \geq \theta_2$ contre $H_1 : \theta \in]\theta_1, \theta_2[$. Pour tout α , il existe un test UPP au seuil α de la forme :

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } k_1 < S(\mathbf{x}) < k_2 \\ \gamma_1 & \text{si } S(\mathbf{x}) = k_1 \\ \gamma_2 & \text{si } S(\mathbf{x}) = k_2 \\ 0 & \text{sinon} \end{cases} .$$

De plus, on a $\rho_\varphi(\theta_1) = \rho_\varphi(\theta_2) = \alpha$.

Remarque 7.11.

1. La difficulté pratique est la détermination de k_1 et k_2 tels que $\rho_\varphi(\theta_1) = \rho_\varphi(\theta_2) = \alpha$.
2. On peut écrire un résultat analogue pour tester $H_0 : \theta \neq \theta_0$ contre $H_1 : \theta = \theta_0$ l'équation du seuil permettant de déterminer γ_1, γ_2, k_1 et k_2 devient alors :

$$\begin{cases} \rho_\varphi(\theta_0) = \alpha \\ \mathbf{E}_{\theta_0} [S(\mathbf{X})\varphi(\mathbf{X})] = \alpha \mathbf{E}_{\theta_0} [S(\mathbf{X})] \end{cases} .$$

3. Comme nous le verrons, ces résultats font figure d'exception dans l'analyse des test bilatéraux et on ne peut pas les obtenir en intervertissant les formes de H_0 et H_1 .

Cas $H_0 : \theta \in [\theta_1, \theta_2]$ contre $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$ et $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ En général, il n'existe pas de test UPP permettant de tester $H_0 : \theta = \theta_0$ contre l'hypothèse alternative bilatérale $H_1 : \theta \neq \theta_0$, même pour un modèle dans dans la classe exponentielle. En effet, considérons vouloir tester ces hypothèses dans le modèle $(\mathbf{R}, (\mathcal{N}(\theta, 1))_{\theta \in \mathbf{R}})$ et supposons que φ soit un test UPP au seuil α . Alors, φ est aussi un test UPP au seuil α pour tester H_0 contre $H_1' : \theta = \theta_1$ pour un certain $\theta_1 > \theta_0$ fixé. Par le Lemme de Neyman-Pearson, il s'agit d'un test de Neyman-Pearson et, en fait, d'un test pur de région critique de la forme $\{\sum_{i=1}^n x_i \geq c_\alpha\}$. De même, φ est aussi un test UPP au seuil α pour tester H_0 contre $H_1'' : \theta = \theta_2$ pour un certain $\theta_2 < \theta_0$ fixé. Par le Lemme de Neyman-Pearson, il s'agit d'un test de Neyman-Pearson et, en fait, d'un test pur de région critique de la forme $\{\sum_{i=1}^n x_i \leq c'_\alpha\}$. Ceci étant absurde, on conclue qu'il n'existe pas de test UPP permettant de tester $H_0 : \theta = \theta_0$ contre l'hypothèse alternative bilatérale $H_1 : \theta \neq \theta_0$ dans ce modèle (pourtant simple, régulier et dans la famille exponentielle).

La même observation peut être faite pour tester $H_0 : \theta \in [\theta_1, \theta_2]$ contre $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$. Du fait de cette observation, on est amenés à rechercher dans ces cas des test optimaux parmi des classes restreintes de test (par exemple les tests sans biais) même dans des classes restreintes de modèles (comme ceux des familles exponentielles).

On admet le résultat suivant.

Théorème 7.6. Soit $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta \subset \mathbf{R}})$ un modèle exponentielle (à paramètre de dimension 1) de vraisemblance

$$L(\mathbf{X}; \theta) = a(\theta)b(\mathbf{x}) \exp(c(\theta)S(\mathbf{x}))$$

avec c strictement croissante de sorte à ce que le modèle soit à rapport de vraisemblance strictement croissant en $T(\mathbf{x})$. On considère vouloir tester $H_0 : \theta \in [\theta_1, \theta_2]$ contre $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$. Pour tout α , il existe un test UPPSB au seuil α de la forme :

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } k_1 < S(\mathbf{x}) < k_2 \\ \gamma_1 & \text{si } S(\mathbf{x}) = k_1 \\ \gamma_2 & \text{si } S(\mathbf{x}) = k_2 \\ 0 & \text{sinon} \end{cases} .$$

De plus, on a $\rho_\varphi(\theta_1) = \rho_\varphi(\theta_2) = \alpha$.

Remarque 7.12. On peut écrire un résultat analogue pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ l'équation du seuil permettant de déterminer γ_1, γ_2, k_1 et k_2 devient alors :

$$\begin{cases} \rho_\varphi(\theta_0) = \alpha \\ \mathbf{E}_{\theta_0} [S(\mathbf{X})\varphi(\mathbf{X})] = \alpha \mathbf{E}_{\theta_0} [S(\mathbf{X})] \end{cases} .$$

7.4 Mise en œuvre d'un test

La mise en œuvre complète d'un test passe par le schéma général suivant.

1. Modélisation : Choix d'un modèle statistique en accord avec le contexte du problème concret considéré.
2. Choix des hypothèses : Ce choix doit être fait en adéquation avec la dissymétrie des rôles joués par H_0 et H_1 : on cherche à contrôler en priorité l'erreur de première espèce, à un seuil fixé *a priori*, on évite en premier lieu de rejeter à tort H_0 .
3. Choix de la statistique de test S : Celui-ci doit être fait de façon à obtenir un zone de rejet agréable (unilatérale ou bilatérale) en général et de sorte à ce que celle-ci ait un comportement différent sous H_0 (sous laquelle sa loi doit être parfaitement connue) et H_1 .
4. Comportement de S sous H_0 : la loi de S doit être parfaitement identifiée sous H_0 ; il peut être asymptotique.
5. Comportement de S sous H_1 : la loi de S sous H_1 doit être différente de celle sous H_0 ; il peut être asymptotique.
6. Détermination de la région critique : elle correspond aux valeurs raisonnable de S sous H_0 et aberrantes sous H_1 ; elle se fait sous H_0 .
7. Analyse de l'erreur de seconde espèce : Contrôle de la puissance et/ou tracé de la puissance ; elle se fait sous H_1 .
8. Conclusion : à partir d'une observation grâce à la région critique ou calcul de la p -valeur.

7.5 Quelques tests usuels

Dans cette section, on ne fait pas une zoologie complète des tests d'hypothèses ni une analyse poussée de ceux qui seront présentés. On se contente de donner les grandes lignes des tests d'usage le plus courant qu'ils soient paramétriques ou non, asymptotiques ou non. On sera vigilant à vérifier que la taille de l'échantillon est suffisante pour les tests asymptotiques (disons $n \geq 50$ pour fixer les idées).

Pour plus de détails sur ces tests ou d'autres tests le lecteur est renvoyé par exemple vers le Chapitre VI de [14] ou les Chapitres 9 et 10 de [11] (voir certains renvois spécifiques plus loin).

7.5.1 Quelques Tests paramétriques

On suppose disposer d'un échantillon de taille n $\mathbf{X} = (X_1, \dots, X_n)$ i.i.d. de la loi mère. Pour une raison de concision, l'expression « loi sous H_0 » est parfois employée légèrement abusivement, ci-dessous, lorsque Θ_0 n'est pas un singleton. Rigoureusement, une discussion de supremum est nécessaire (voir Définition 7.5 en particulier) mais les calculs se font *in fine* avec la loi mentionnée.

Test pour la moyenne dans un échantillon gaussien avec variance connue

Cadre : $(\mathbf{R}, (\mathcal{N}(\theta, \sigma^2))_{\theta \in \mathbf{R}})$, $\sigma^2 > 0$ connu.

Cas 1 :

$$H_0 : \theta = \theta_0 \text{ ou } \theta \leq \theta_0$$

$$H_1 : \theta = \theta_1 \text{ } (\theta_1 > \theta_0) \text{ ou } \theta > \theta_0$$

Statistique de test :

$$S = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}$$

Loi sous H_0 :

$$S \sim \mathcal{N}(0, 1).$$

Forme de la région critique :

$$R = \{\mathbf{x} : S(\mathbf{x}) \geq c\}$$

Cas 2 :

$$H_0 : \theta = \theta_0 \text{ ou } \theta \geq \theta_0$$

$$H_1 : \theta = \theta_1 \text{ } (\theta_1 < \theta_0) \text{ ou } \theta < \theta_0$$

Statistique de test :

$$S = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}$$

Loi sous H_0 :

$$S \sim \mathcal{N}(0, 1).$$

Forme de la région critique :

$$R = \{\mathbf{x} : S(\mathbf{x}) \leq c\}$$

Cas 3 :

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

Statistique de test :

$$S = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}$$

Loi sous H_0 :

$$S \sim \mathcal{N}(0, 1).$$

Forme de la région critique :

$$R = \{|S(\mathbf{x})| \geq c\}$$

Test pour la moyenne dans un échantillon gaussien avec variance inconnue

Cadre : $(\mathbf{R}, (\mathcal{N}(\theta, \sigma^2))_{\theta \in \mathbf{R}, \sigma^2 > 0})$, ($\sigma^2 > 0$ inconnu).

Cas 1 :

$$H_0 : \theta = \theta_0 \text{ ou } \theta \leq \theta_0$$

$$H_1 : \theta = \theta_1 \text{ } (\theta_1 > \theta_0) \text{ ou } \theta > \theta_0$$

Statistique de test :

$$S(\mathbf{X}) = \frac{\bar{X}_n - \theta_0}{\sqrt{S_n^2(\mathbf{X})/n}}$$

où $S_n^2(\mathbf{X})$ est l'estimateur sans biais de la variance :

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

Loi sous H_0 :

$$S \sim \mathcal{T}(n-1) \quad (\text{loi de Student à } n-1 \text{ d.d.l.}).$$

Forme de la région critique :

$$R = \{\mathbf{x} : S(\mathbf{x}) \geq c\}$$

Cas 2 :

$$H_0 : \theta = \theta_0 \text{ ou } \theta \geq \theta_0$$

$$H_1 : \theta = \theta_1 \text{ } (\theta_1 < \theta_0) \text{ ou } \theta < \theta_0$$

Statistique de test :

$$S(\mathbf{X}) = \frac{\bar{X}_n - \theta_0}{\sqrt{S_n^2(\mathbf{X})/n}}$$

où $S_n^2(\mathbf{X})$ est l'estimateur sans biais de la variance :

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

Loi sous H_0 :

$$S \sim \mathcal{T}(n-1) \quad (\text{loi de Student à } n-1 \text{ d.d.l.}).$$

Forme de la région critique :

$$R = \{\mathbf{x} : S(\mathbf{x}) \leq c\}$$

Cas 3 :

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

Statistique de test :

$$S(\mathbf{X}) = \frac{\bar{X}_n - \theta_0}{\sqrt{S_n^2(\mathbf{X})/n}}$$

où $S_n^2(\mathbf{X})$ est l'estimateur sans biais de la variance :

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

Loi sous H_0 :

$$S \sim \mathcal{T}(n-1) \quad (\text{loi de Student à } n-1 \text{ d.d.l.}).$$

Forme de la région critique :

$$R = \{|S(\mathbf{x})| \geq c\}.$$

Test pour la variance dans un échantillon gaussien avec moyenne connue

Cadre : $(\mathbf{R}, (\mathcal{N}(m, \theta))_{\theta > 0})$, m connu.

$$H_0 : \theta = \theta_0$$

Statistique de test :

$$S(\mathbf{X}) = \frac{1}{\theta_0} \sum_{i=1}^n (X_i - m)^2.$$

Loi sous H_0 :

$$S \sim \chi^2(n).$$

Forme de la région critique : Selon la forme de H_1 , similaire aux cas précédents pour les cas unilatéraux. Dans les cas bilatéraux, du fait de la dissymétrie de la loi du Khi-2, la région critique est de la forme $] -\infty, a_\alpha] \cup [b_\alpha, +\infty[$ que l'on détermine de façon à ce que, pour $\chi^2 \sim \chi^2(n)$:

$$\mathbf{P}[\chi^2 \leq a_\alpha] = \mathbf{P}[\chi^2 \geq b_\alpha] = \frac{\alpha}{2}.$$

Test pour la variance dans un échantillon gaussien avec moyenne inconnue

Cadre : $(\mathbf{R}, (\mathcal{N}(m, \theta))_{m \in \mathbf{R}, \theta > 0})$, (m inconnu).

$$H_0 : \theta = \theta_0$$

Statistique de test :

$$S(\mathbf{X}) = \frac{1}{\theta_0} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Loi sous H_0 :

$$S \sim \chi^2(n-1).$$

Forme de la région critique : Selon la forme de H_1 , similaire aux cas précédents pour les cas unilatéraux. Dans les cas bilatéraux, du fait de la dissymétrie de la loi du Khi-2, la région critique est de la forme $] -\infty, a_\alpha] \cup [b_\alpha, +\infty[$ que l'on détermine de façon à ce que, pour $\chi^2 \sim \chi^2(n-1)$:

$$\mathbf{P}[\chi^2 \leq a_\alpha] = \mathbf{P}[\chi^2 \geq b_\alpha] = \frac{\alpha}{2}.$$

Test de comparaison de moyennes dans des échantillons gaussiens

Cadre : $\mathbf{X} = (X_1, \dots, X_{n_1})$ de loi mère $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$ de loi mère $\mathcal{N}(m_2, \sigma_2^2)$ indépendants.

Cas des variances connues

$H_0 : m_1 = m_2$

Statistique de test :

$$S(\mathbf{X}) = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Loi sous H_0 :

$$S \sim \mathcal{N}(0, 1).$$

Forme de la région critique : Selon la forme de H_1 , similaire aux cas précédents.

Cas des variances inconnues mais supposées égales

$H_0 : m_1 = m_2$

Statistique de test :

$$S(\mathbf{X}) = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{S^2(\mathbf{X}, \mathbf{Y}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

où

$$S^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_i - \bar{X}_n)^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y}_n)^2 \right)$$

Loi sous H_0 :

$$S \sim \mathcal{T}(n_1 + n_2 - 2).$$

Forme de la région critique : Selon la forme de H_1 , similaire aux cas précédents.

Remarque : Pour le cas général à variance inconnue, on peut construire un test asymptotique de comparaison de moyennes (voir *problème de Behrens-Fisher*).

Test de comparaison de variances dans des échantillons gaussiens

Cadre : $\mathbf{X} = (X_1, \dots, X_{n_1})$ de loi mère $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$ de loi mère $\mathcal{N}(m_2, \sigma_2^2)$ indépendants.

$H_0 : \sigma_1^2 = \sigma_2^2$

Cas des moyennes connues

Statistique de test :

$$S(\mathbf{X}) = \frac{\widehat{S}_{n_1}^2(\mathbf{X})}{\widehat{S}_{n_2}^2(\mathbf{Y})}$$

où

$$\widehat{S}_n^2(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n (Z_i - \mathbf{E}[Z])^2.$$

Loi sous H_0 :

$$S \sim \mathcal{F}(n_1, n_2) \quad (\text{loi de Fisher-Snedecor}).$$

Cas des moyennes inconnues

Statistique de test :

$$S(\mathbf{X}) = \frac{S_{n_1}^2(\mathbf{X})}{S_{n_2}^2(\mathbf{Y})}$$

Loi sous H_0 :

$$S \sim \mathcal{F}(n_1 - 1, n_2 - 1).$$

Test pour une proportion

Cadre : X_1, \dots, X_n i.i.d. de loi $\mathcal{Ber}(\theta)$.

$H_0 : \theta = \theta_0$

Statistique de test :

$$S(\mathbf{X}) = \sum_{i=1}^n X_i.$$

Loi sous H_0 :

$$S(\mathbf{X}) \sim \mathcal{Bin}(n, \theta_0).$$

Remarque : Si n est assez grand, il est courant de plutôt utiliser un test asymptotique en approchant convenablement la loi $\mathcal{Bin}(n, \theta_0)$ par une loi de Poisson ou une loi normale.

7.5.2 Test du Khi-2 d'indépendance

Cadre : Pour deux caractères étudiés sur une même population, X et Y , de modalités x_1, \dots, x_l et y_1, \dots, y_r , on dispose des effectifs observés $O_{i,j}$ du couple (x_i, y_j) présentés généralement sous la forme d'un tableau à double entrée. On note N l'effectif total de la population et $N_{i,\cdot} = \sum_{j=1}^r O_{i,j}$ les effectifs marginaux en X et $N_{\cdot,j} = \sum_{i=1}^l O_{i,j}$ les effectifs marginaux en Y .

H_0 : X et Y sont indépendantes. On note

$$T_{i,j} = \frac{N_{i,\cdot} N_{\cdot,j}}{N}$$

l'effectif théorique de la modalité (x_i, y_j) sous H_0 . On suppose que tous les effectifs théoriques sont ≥ 5 , sinon on effectue des regroupements de lignes ou de colonnes.

Statistique de test : pseudo-distance du Khi-2 :

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - T_{i,j})^2}{T_{i,j}}$$

Loi sous H_0 : Asymptotiquement, sous H_0 :

$$\chi^2 \underset{\text{approx.}}{\sim} \chi^2((r-1)(c-1))$$

où r et c sont les nombres de lignes et de colonnes *après éventuels regroupements*.

Forme de la région critique :

$$R_\alpha = \{\chi_{\text{obs}}^2 \geq \chi_{\text{crit}}^2\}$$

où χ_{obs}^2 est la valeur observée de la statistique de test et χ_{crit}^2 est déterminé en fonction du seuil du test et du nombre de d.d.l.

Remarque : Il s'agit d'un test asymptotique. Voir par exemple [7] Sections IX.9.3. ou [14] p. 204-206.

7.5.3 Test du Khi-2 d'adéquation à une loi

Cadre : On dispose des effectifs observés O_1, \dots, O_l , dans un échantillon de taille $N = \sum_i O_i$, d'un caractère discret X de modalités x_1, \dots, x_l et l'on souhaite tester si X se distribue conformément à une loi discrète spécifiée μ . S'il est nécessaire d'estimer, des paramètres pour cette loi, on note r le nombre de paramètres estimés.

H_0 : $X \sim \mu$. On note

$$T_i = N\mu(x_i)$$

l'effectif théorique de la modalité x_i sous H_0 . *On suppose que tous les effectifs théoriques sont ≥ 5 , sinon on effectue des regroupements en classes.*

Statistique de test : pseudo-distance du Khi-2 :

$$\chi^2 = \sum_i \frac{(O_i - T_i)^2}{T_i}$$

Loi sous H_0 : Asymptotiquement, sous H_0 :

$$\chi^2 \underset{\text{approx.}}{\sim} \chi^2(l - r - 1)$$

où l est le nombre de modalités *après éventuels regroupements* et r le nombre de paramètres estimés (éventuellement nul).

Forme de la région critique :

$$R_\alpha = \{\chi_{\text{obs}}^2 \geq \chi_{\text{crit}}^2\}$$

où χ_{obs}^2 est la valeur observée de la statistique de test et χ_{crit}^2 est déterminé en fonction du seuil du test et du nombre de d.d.l.

Remarque : Il s'agit d'un test asymptotique. Voir par exemple [7] Sections IX.9.2.

7.5.4 Voir aussi

Test de comparaison de deux proportions

[11] Section 9.7.6., [14] p.191-192.

Test de comparaison de deux échantillons gaussiens de Fisher-Snedecor et Student

[14] p.182-185.

Test de comparaison de deux échantillons

[7] Section IX.10.3

Test du Khi-2 d'homogénéité de plusieurs échantillons

[11] Section 10.2., [14] p.186-190 (Smirnov, Wilcoxon, ...)

Test de corrélation dans un couple gaussien

[11] Section 9.7.7.

Test de corrélation de Spearman

[11] Section 10.5.5., [14] p.198-200

Test exact de Fisher

test d'indépendance non asymptotique [11] Section 10.3.2

Test de Kolmogorov-Smirnov

Adéquation à une loi continue, [11] p. 266-267 et 270-271, [7] Sections IX.10.1 et IX.10.2, [14] p. 176.

Test de Cramer-von Mises

test d'ajustement, en particulier applicable pour des lois normales ou exponentielles [14] p. 177-178.

Test de localisation de deux lois

[11] Section 10.5.4.

Test d'utilité des régresseurs

[7] Sections IX.7.2., [14] p.218

Chapitre 8

Estimation par intervalles ou régions de confiance

Alors que dans le cadre de l'estimation paramétrique ponctuelle l'objectif était de donner une valeur unique pour approcher le paramètre inconnu θ , l'idée sous-jacente à l'estimation par intervalle ou région de confiance, est de donner un ensemble de valeur plausibles pour le paramètre à estimer telle que la probabilité pour que le paramètre appartienne effectivement à cette région est prescrit. Il est clair qu'une estimation ponctuelle du paramètre inconnu doit être un bon point de départ pour construire de tels intervalles ou régions et que le contrôle de la variance de l'estimateur doit permettre de contrôler la taille de la région, que l'on souhaite petite par soucis de précision.

Remarquons que cette approche est naturelle puisque même si l'estimation ponctuelle $\hat{\theta}$ de θ est convenable la probabilité pour que θ soit effectivement égale à $\hat{\theta}$ est faible, et est même nulle dès que la loi de $\hat{\theta}$ est continue.

Dans ce chapitre, nous nous restreindrons au cas de l'estimation d'un paramètre de dimension 1 par intervalle de confiance et n'explorerons pas le cas des dimensions supérieures et des régions de confiance. Nous décrirons les méthodes de construction d'intervalle et donnerons des exemples classiques. Nous ne nous intéresserons pas à la qualité et à l'optimalité de tels intervalles de confiance. Le lecteur intéressé par ces questions est renvoyé par exemple à la Section 7.7 de [11] ou au plus complet Chapitre 7 de [15].

8.1 Estimation par intervalles de confiance de niveau exact ou par excès

Définition 8.1. Soit $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ un modèle statistique avec $\Theta \subset \mathbf{R}$ sur lequel on observe un échantillon (X_1, \dots, X_n) de taille n , $g : \Theta \rightarrow \mathbf{R}$ une application (mesurable) et $\alpha \in]0, 1[$.

Un intervalle aléatoire $I_{n,\alpha} = I_{n,\alpha}(X_1, \dots, X_n)$ est appelé intervalle de confiance (IC) de niveau exact (resp. par excès) $1 - \alpha$ pour $g(\theta)$ si pour tout $\theta \in \Theta$:

$$\mathbf{P}_\theta [g(\theta) \in I_{n,\alpha}] = 1 - \alpha \quad (\text{resp. } \geq 1 - \alpha).$$

Remarque 8.1.

1. Il est fréquent de prendre $g = \text{id}$.

2. Dans la pratique, ayant donné la garantie du niveau au moyen de l'étude de l'intervalle aléatoire $I_{n,\alpha}(X_1, \dots, X_n)$, on fournira l'estimation par intervalle de confiance $I_{n,\alpha}(x_1, \dots, x_n)$ pour une observation (x_1, \dots, x_n) .
3. On utilisera les IC par excès en particulier pour les lois discrètes pour lesquels l'IC de niveau exact n'est en général pas accessible en raison de la non continuité de la fonction de répartition.

Définition 8.2. Dans le cadre de la définition précédente, soit $\hat{g}_n = \hat{g}_n(X_1, \dots, X_n)$ un estimateur de $g(\theta)$. On appelle fonction pivot toute fonction u définie sur $g(\Theta)^2$ telle que la loi de $u(\hat{g}_n, g(\theta))$ est indépendante de θ .

Remarque 8.2. Les limites de cette méthode sont de déterminer une fonction pivot (s'il en existe une) et de déterminer la loi de $u(\hat{g}_n, g(\theta))$.

Proposition 8.1. S'il existe une fonction pivot, alors, pour tout α , il existe un IC de niveau exact $1 - \alpha$ basé sur l'estimateur \hat{g}_n .

Preuve : Dans ce cas, pour α fixé arbitrairement dans $]0, 1[$, il existe I tel que, indépendamment de θ ,

$$\mathbf{P}_\theta [u(\hat{g}_n, g(\theta)) \in I] = 1 - \alpha$$

et un intervalle de confiance de niveau exact $1 - \alpha$ pour $g(\theta)$ est donné par :

$$I_{n,\alpha}(X_1, \dots, X_n) = \{\mathbf{x} \in \mathcal{X}^n : u(\hat{g}_n(\mathbf{x}), g(\theta)) \in I\}.$$

□

Exemple 8.1. Soit le modèle statistique $(\mathbf{R}_+, (\mathcal{E}(\theta))_{\theta>0})$ sur lequel on souhaite estimer la moyenne, inverse du paramètre, $1/\theta$ au moyen d'un échantillon $\mathbf{X} = (X_1, \dots, X_n)$. Rappelons que la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est dans ce cadre un estimateur de $1/\theta$. Nous allons voir que la fonction u définie par

$$u(\bar{X}_n, 1/\theta) = \theta n \bar{X}_n$$

est une fonction pivot. Pour cela, rappelons que par l'Exercice A.2, $n\bar{X}_n \sim \Gamma(n; \theta)$. Ainsi, on a :

$$F_{u(\bar{X}_n, 1/\theta)}(t) = F_{\theta n \bar{X}_n}(t) = \mathbf{P}_\theta [\theta n \bar{X}_n \leq t] = \mathbf{P}_\theta \left[n \bar{X}_n \leq \frac{t}{\theta} \right] = F_{n \bar{X}_n} \left(\frac{t}{\theta} \right)$$

et donc, en dérivant,

$$\begin{aligned} f_{u(\bar{X}_n, \theta)}(t) &= \frac{1}{\theta} f_{n \bar{X}_n} \left(\frac{t}{\theta} \right) = \frac{1}{\theta} \frac{\left(\frac{t}{\theta}\right)^{n-1} \theta^n e^{-\theta \frac{t}{\theta}}}{(n-1)!} \mathbf{1}_{t>0} \\ &= \frac{t^{n-1} e^{-t}}{(n-1)!} \mathbf{1}_{t>0}. \end{aligned}$$

On obtient donc que $u(\bar{X}_n, \theta) \sim \Gamma(n, 1)$ indépendamment de θ . Par suite, pour tout choix de $0 < i^- < i^+ < +\infty$ tel que pour $Y \sim \Gamma(n, 1)$

$$\mathbf{P} [Y \in [i^-, i^+]] = 1 - \alpha$$

on peut trouver un intervalle de confiance de niveau $1 - \alpha$ pour $1/\theta$. En effet,

$$\begin{aligned} u(\bar{X}_n, \theta) \in [i^-, i^+] &\iff i^- \leq \theta n \bar{X}_n \leq i^+ \\ &\iff \frac{n \bar{X}_n}{i^+} \leq \theta^{-1} \leq \frac{n \bar{X}_n}{i^-} \\ &\iff \theta^{-1} \in \left[\frac{n \bar{X}_n}{i^+}, \frac{n \bar{X}_n}{i^-} \right]. \end{aligned}$$

L'intervalle de confiance recherché est donc de la forme $\left[\frac{n \bar{X}_n}{i^+}, \frac{n \bar{X}_n}{i^-} \right]$.

Remarque 8.3.

1. On voit dans l'exemple précédent que l'intervalle de confiance n'est pas unique et d'amplitude aléatoire. Un courant est de prendre i^- le quantile d'ordre $\frac{\alpha}{2}$ de la loi de $u(\hat{g}_n, g(\theta))$ (ici de la loi $\Gamma(n; 1)$) et i^+ son quantile d'ordre $1 - \frac{\alpha}{2}$. L'IC est toujours d'amplitude aléatoire. Ce choix permet lorsque la loi de $u(\hat{g}_n, g(\theta))$ est symétrique et unimodale – par exemple gaussienne centrée – d'obtenir un IC d'amplitude minimale (ce n'est pas le cas ici). Les quantiles nécessaires seront accessible *via* les fonctions adéquates de tout logiciel de traitement statistique comme R, ou de manière plus ancestrale *via* des tables.
2. Dans certains contextes particuliers, on peut préférer des IC unilatéraux de la forme $] - \infty, a]$ ou $[a, +\infty[$.

Exercice 8.1. Soit le modèle statistique $(\mathbf{R}, (\mathcal{N}(\theta, \sigma^2))_{\theta \in \mathbf{R}})$, $\sigma^2 > 0$ fixé, sur lequel on souhaite estimer la moyenne θ au moyen d'un échantillon $\mathbf{X} = (X_1, \dots, X_n)$. Construire un IC de niveau exact $1 - \alpha$.

8.2 Estimation par intervalles de confiance asymptotiques

L'estimation par IC asymptotiques est de portée plus large, mais nécessite que l'échantillon utilisé soit de taille suffisante, disons $n \geq 30$ pour fixer les idées. Sa légitimité s'accroît avec la taille de l'échantillon.

Définition 8.3. Soit $(\mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ un modèle statistique avec $\Theta \subset \mathbf{R}$ sur lequel on observe un échantillon (X_1, \dots, X_n) de taille $n \geq 1$, $g : \Theta \rightarrow \mathbf{R}$ une application (mesurable) et $\alpha \in]0, 1[$.

Un intervalle aléatoire $I_{n,\alpha} = I_{n,\alpha}(X_1, \dots, X_n)$ est appelé intervalle de confiance (IC) de niveau asymptotique $1 - \alpha$ pour $g(\theta)$ si pour tout $\theta \in \Theta$:

$$\lim_{n \rightarrow +\infty} \mathbf{P}_\theta [g(\theta) \in I_{n,\alpha}] = 1 - \alpha.$$

Cette approche est en particulier valable lorsque, avec un échantillon de grande taille, on dispose d'un estimateur $\hat{\theta}_n$ de $g(\theta)$ asymptotiquement normal (voir Section 6.3.6). Nous avons vu dans la Section 6.3.6 des conditions garantissant une telle normalité asymptotique pour l'EM et l'EMV (voir Proposition 6.8 et Théorème 6.4).

Dans ce cadre, supposons que l'on ait pour tout θ :

$$\frac{\hat{\theta}_n - g(\theta)}{s_n(\theta)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

avec s_n positive p.p.. En recherchant $I_{n,\alpha} = I_{n,\alpha}(X_1, \dots, X_n)$ de la forme $I_{n,\alpha} = [i^-, i^+]$, il vient :

$$\begin{aligned} \mathbf{P}_\theta [g(\theta) \in I_{n,\alpha}] &= 1 - \alpha \\ \iff \mathbf{P}_\theta [i^- \leq g(\theta) \leq i^+] &= 1 - \alpha \\ \iff \mathbf{P}_\theta [g(\theta) \leq i^- \text{ ou } g(\theta) \geq i^+] &= \alpha \\ \iff \mathbf{P}_\theta \left[\frac{\hat{\theta}_n - g(\theta)}{s_n(\theta)} \geq \frac{\hat{\theta}_n - i^-}{s_n(\theta)} \text{ ou } \frac{\hat{\theta}_n - g(\theta)}{s_n(\theta)} \leq \frac{\hat{\theta}_n - i^+}{s_n(\theta)} \right] &= \alpha. \end{aligned}$$

Or, en notant $q_{\frac{\alpha}{2}} = 1 - q_{1-\frac{\alpha}{2}}$ et $q_{1-\frac{\alpha}{2}}$ les quantiles d'ordre $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ de $\mathcal{N}(0, 1)$, on a, pour $Z \sim \mathcal{N}(0, 1)$:

$$\mathbf{P} [Z \geq q_{1-\frac{\alpha}{2}} \text{ ou } Z \leq q_{\frac{\alpha}{2}}] = \alpha.$$

Il s'ensuit en identifiant lorsque la taille de l'échantillon est suffisante pour permettre l'approximation gaussienne que

$$\frac{\hat{\theta}_n - i^-}{s_n(\theta)} \simeq q_{1-\frac{\alpha}{2}} \quad \text{et} \quad \frac{\hat{\theta}_n - i^+}{s_n(\theta)} \simeq q_{\frac{\alpha}{2}}.$$

On obtient ainsi l'IC asymptotique (approché) :

$$I_{n,\alpha} \simeq [\hat{\theta}_n - q_{1-\alpha} s_n(\theta), \hat{\theta}_n - q_\alpha s_n(\theta)] = [\hat{\theta}_n - q_{1-\alpha} s_n(\theta), \hat{\theta}_n + q_{1-\alpha} s_n(\theta)].$$

8.3 Approche bayésienne

Rappelons que dans le cadre de l'approche bayésienne, l'inférence est faite *via* la loi *a posteriori* (voir Section 6.2.4). Dans ce contexte, on substitue à la notion d'intervalle de confiance celle d'*intervalle de crédibilité* que l'on continue à noter IC. On encadrera simplement, pour déterminer un intervalle de crédibilité de niveau α , θ ou $g(\theta)$ par les quantiles d'ordre $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ de la loi *a posteriori*.

Exercice 8.2. Dans le contexte de l'Exemple 6.6, écrire l'IC de niveau $1 - \alpha$.

8.4 Correspondance entre intervalles de confiance et tests

Si $I_{n,\alpha} = I_{n,\alpha}(x_1, \dots, x_n)$ est un IC de niveau $1 - \alpha$ pour $g(\theta)$ basé sur un estimateur $\hat{\theta}_n$, on a :

$$g(\theta) \notin I_{n,\alpha} \iff \mathbf{x} = (x_1, \dots, x_n) \in R := \{\mathbf{x} \in \mathcal{X}^n : g(\theta) \notin I_{n,\alpha}\},$$

et donc pour tout θ :

$$P_\theta [\mathbf{X} \in R] = P_\theta [g(\theta) \notin I_{n,\alpha}] = \alpha.$$

Il en résulte que, pour tout $\theta_0 \in \Theta$, R est la région critique d'un test de risque de première espèce α de l'hypothèse $H_0 : g(\theta) = g(\theta_0)$ contre l'alternative $H_1 : g(\theta) \neq g(\theta_0)$.

Réciproquement, si pour tout $\theta_0 \in \Theta$, R_α est la région critique d'un test de risque de première espèce α de l'hypothèse $H_0 : g(\theta) = g(\theta_0)$ contre l'alternative $H_1 : g(\theta) \neq g(\theta_0)$, alors $I_{n,\alpha} = \{g(\theta) : \mathbf{x} \notin R_\alpha\}$ définit un IC de niveau $1 - \alpha$ pour $g(\theta)$.

Remarque 8.4. Même si nous ne développons pas ces notions, notons que cette correspondance permet de construire des intervalles de confiance optimaux, selon des critères intrinsèques, par dualité avec les tests UPP et UPPSB.

8.5 Bases pour quelques intervalles de confiance usuels

Dans cette section, on rappelle les éléments théoriques permettant de construire quelques IC d'usage courant soit dans leurs versions bilatérales soit dans leurs versions unilatérales, ce choix étant fait selon le contexte de la situation concrète traitée.

8.5.1 IC pour une moyenne

Cas d'un échantillon gaussien de variance connue

Si $\mathbf{X} = (X_1, \dots, X_n)$ est un échantillon de loi $\mathcal{N}(\theta, \sigma^2)$, $\sigma^2 > 0$ fixé, on utilise que d'après la Proposition 3.7 et sous \mathbf{P}_θ :

$$\sum_{i=1}^n X_i \sim \mathcal{N}(n\theta, n\sigma^2),$$

soit encore

$$\frac{\bar{X}_n - \theta}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1).$$

Remarque 8.5. Bien qu'il puisse sembler peu réaliste (comment connaître la variance si l'on ignore la moyenne ?), ce cas peut se rencontrer dans des contextes spécifiques, typiquement, lorsque sur une machine dont la précision induit des fluctuations gaussiennes de variance fixée sur les mesures des objets fabriqués mais un réglage influe sur la moyenne indépendamment de la variance.

Cas d'un échantillon gaussien de variance inconnue

Si $\mathbf{X} = (X_1, \dots, X_n)$ est un échantillon de loi $\mathcal{N}(\theta_1, \theta_2)$, de variance $\theta_2 > 0$ inconnue, on utilise que d'après le Théorème 5.2 et sous $\mathbf{P}_{(\theta_1, \theta_2)}$:

$$\frac{\bar{X}_n - \theta_1}{S_n/\sqrt{n}} \sim \mathcal{T}(n-1),$$

avec $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne empirique de l'échantillon et $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$ son écart-type corrigé.

Cas asymptotique pour un échantillon de loi de carré intégrable

Si $\mathbf{X} = (X_1, \dots, X_n)$ est un échantillon d'une loi de carré intégrable de moyenne θ et d'écart-type σ , le TCL (Théorème 4.4) assure que

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

et l'approximation de la loi de ce quotient par $\mathcal{N}(0, 1)$ pour n grand est bonne dès que $n \geq 30$. En général, σ est inconnu et il est naturel de vouloir l'estimer par S_n . Pour assurer que cette

estimation est convenable, il est nécessaire de considérer des échantillons de taille bien plus importante (disons $n \geq 100$) pour fixer les idées. Par ailleurs, la loi de S_n peut, en général, différer sensiblement de celle de son analogue dans le cas gaussien (donnée dans le Théorème 5.2). On admettra cependant, que si $n \geq 100$, un IC pour la moyenne θ dans ce cadre plus général, peut être obtenu en utilisant que :

$$\frac{\bar{X}_n - \theta}{S_n/\sqrt{n}} \underset{\text{approx.}}{\sim} \mathcal{T}(n-1),$$

avec $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne empirique de l'échantillon et $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$ son écart-type corrigé.

8.5.2 IC sur la différence des moyennes de deux échantillons gaussiens

Cas d'échantillons de même variance σ^2

On suppose que $\mathbf{X} = (X_1, \dots, X_{n_1})$ est un échantillon de loi mère $\mathcal{N}(\theta_1, \sigma^2)$ et $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$ est un échantillon de loi mère $\mathcal{N}(\theta_2, \sigma^2)$ et que les deux échantillons sont indépendants. On s'intéresse à la différence $\theta_1 - \theta_2$. Il vient que

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \sim \mathcal{N}\left(\theta_1 - \theta_2, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

La difficulté est, en fait, l'estimation de σ^2 mais l'on sait (voir Théorème 5.2) que les variances corrigées $S_{\mathbf{X}}^2$ et $S_{\mathbf{Y}}^2$ des deux échantillons satisfont :

$$\frac{n_1 - 1}{\sigma^2} S_{\mathbf{X}}^2 \sim \chi^2(n_1 - 1) \quad \text{et} \quad \frac{n_2 - 1}{\sigma^2} S_{\mathbf{Y}}^2 \sim \chi^2(n_2 - 1)$$

d'où l'on déduit par indépendance des deux échantillons que

$$S_{\text{pond}}^2 = \frac{n_1 - 1}{\sigma^2} S_{\mathbf{X}}^2 + \frac{n_2 - 1}{\sigma^2} S_{\mathbf{Y}}^2 \sim \chi^2(n_1 + n_2 - 2).$$

En définissant l'estimateur pondéré de la variance de ces deux échantillons :

$$\frac{(n_1 - 1)S_{\mathbf{X}}^2 + (n_2 - 1)S_{\mathbf{Y}}^2}{n_1 + n_2 - 2},$$

on obtient que

$$\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - (\theta_1 - \theta_2)}{S_{\text{pond}}^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{T}(n_1 + n_2 - 2).$$

Remarque 8.6.

On peut montrer que cette approche reste applicable lorsque les variances des deux échantillons ne sont pas égales mais sont proches.

Cas d'échantillons de variances significativement différentes

Si les variances sont significativement différentes, l'approche précédente est mise en défaut. Toutefois, on obtient que

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \sim \mathcal{N}\left(\theta_1 - \theta_2, \frac{\sigma_{\mathbf{X}}^2}{n_1} + \frac{\sigma_{\mathbf{Y}}^2}{n_2}\right).$$

Si les tailles des échantillons est importante (disons supérieures à 100), on peut conduire un calcul approché raisonnable en substituant aux variances les variances corrigées $S_{\mathbf{X}}^2$ et $S_{\mathbf{Y}}^2$ des deux échantillons.

Remarque 8.7.

Cette approche est applicable asymptotiquement dans le cadre de deux échantillons de même loi, non nécessairement gaussienne.

8.5.3 IC pour la variance d'un échantillon gaussien

Il suffit d'utiliser que par le Théorème 5.2,

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1).$$

Remarque 8.8. Contrairement à la construction d'IC pour la moyenne, même pour des échantillons de très grande taille, cette approche n'est pas applicable hors du cadre gaussien.

8.5.4 IC pour le rapport de variances de deux échantillons gaussiens

Il suffit d'utiliser la Proposition 5.2.

8.5.5 IC pour une proportion

Estimer une proportion revient à estimer le paramètre $\theta \in [0, 1]$ dans le modèle d'échantillonnage basé sur la famille des lois de Bernoulli. Pour un échantillon de taille n , sous \mathbf{P}_θ , on a :

$$\sum_{i=1}^n X_i \sim \mathcal{B}in(n, \theta),$$

ce qui permet d'obtenir des IC exact. Il est naturel de vouloir estimer la proportion θ par la fréquence empirique $\hat{f}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Lorsque l'échantillon est de taille suffisante (disons $n\hat{f}_n, n(1-\hat{f}_n) > 5$ pour fixer les idées), on peut utiliser l'approximation gaussienne fournie par le TCL et raisonner comme pour l'IC pour la moyenne.

Annexe A

Lois usuelles

A.1 Lois discrètes usuelles

A.1.1 Loi uniforme discrète

Définition A.1. On appelle loi uniforme sur $\{1, 2, \dots, n\}$ la loi dont le support est $\{1, 2, \dots, n\}$ et dont les probabilités des événements élémentaires sont identiques, c'est-à-dire telle qu'une variable aléatoire (v.a.) suivant cette loi vérifie :

$$\mathbf{P}[X = k] = \frac{1}{n}, \quad \text{pour tout } k \in \{1, 2, \dots, n\}.$$

Cette loi est notée $\mathcal{U}(\{1, 2, \dots, n\})$.

Remarque A.1.

1. Cette loi modélise l'équiprobabilité.
2. On note $X \sim \mathcal{U}(\{1, 2, \dots, n\})$ pour dire que la v.a. X suit la loi $\mathcal{U}(\{1, 2, \dots, n\})$.

Exemple A.1. Considérons une urne contenant 50 boules indiscernables numérotées de 1 à 50. Si l'on choisit une boule au hasard dans l'urne, la variable aléatoire indiquant le numéro de la boule suit une loi uniforme sur $\{1, 2, \dots, 50\}$.

Proposition A.1. Soit $X \sim \mathcal{U}(\{1, 2, \dots, n\})$.

On a :

$$\mathbf{E}[X] = \frac{n+1}{2}, \quad \mathbf{V}[X] = \frac{n^2-1}{12},$$
$$L_X(t) = \frac{1}{n} \sum_{k=1}^n e^{tk} \quad \text{et} \quad \phi_X(t) = \frac{1}{n} \sum_{k=1}^n e^{itk}.$$

A.1.2 Loi de Bernoulli

Considérons une expérience aléatoire possédant (exactement) deux issues alternatives du type « succès » ou « échec », « vrai » ou « faux », « pile » ou « face », ... Une telle expérience est appelée *épreuve de Bernoulli*. On modélise une telle expérience par une v.a. X telle que l'événement $\{X = 1\}$ représente un succès et l'événement $\{X = 0\}$ un échec. On a alors $X(\Omega) = \{0; 1\}$ et si $\mathbf{P}[X = 1] = p \in [0; 1]$, $\mathbf{P}[X = 0] = 1 - \mathbf{P}[X = 1] = 1 - p$.

Définition A.2. On dit qu'une v.a. X telle que $X(\Omega) = \{0; 1\}$, $\mathbf{P}[X = 1] = p \in [0; 1]$, $\mathbf{P}[X = 0] = 1 - p$ suit une loi de Bernoulli de paramètre p .

On note alors $X \sim \mathcal{Ber}(p)$.

Exemple A.2. Considérons un tirage à pile ou face d'une pièce bien équilibrée. La variable aléatoire définie par :

$$X = \begin{cases} 1 & \text{si la pièce tombe sur face} \\ 0 & \text{si la pièce tombe sur pile} \end{cases}$$

suit la loi de Bernoulli de paramètre $\frac{1}{2}$.

Proposition A.2. Si $X \sim \mathcal{Ber}(p)$, on a :

$$\begin{aligned} \mathbf{E}[X] &= p, & \mathbf{V}[X] &= p(1 - p), \\ L_X(t) &= 1 - p + pe^t & \text{et} & \quad \phi_X(t) = 1 - p + pe^{it}. \end{aligned}$$

A.1.3 Loi binomiale

Lorsqu'une épreuve de Bernoulli est répétée plusieurs fois, disons n fois, **indépendamment**, on peut s'intéresser au nombre de succès obtenus lors de ces n expériences ; ce nombre est simplement la somme des variables de Bernoulli servant à modéliser ces expériences répétées.

Définition A.3. On dit qu'une v.a. X suit une loi binomiale de paramètres $n \in \mathbf{N}$ et $p \in [0; 1]$ si X s'écrit sous la forme :

$$X = \sum_{k=1}^n X_k,$$

où X_1, \dots, X_n sont **indépendantes et identiquement distribuées (i.i.d.)** de loi de Bernoulli de paramètre p .

On note alors $X \sim \mathcal{Bin}(n, p)$.

Remarque A.2. Le support de la loi $\mathcal{Bin}(n, p)$ est $\{0, \dots, n\}$.

Exemple A.3. Considérons 10 tirages successifs à pile ou face d'une pièce tombant sur face avec probabilité p . Pour $k = 1, \dots, n$, la variable aléatoire définie par :

$$X_k = \begin{cases} 1 & \text{si la } k^{\text{e}} \text{ pièce tombe sur face} \\ 0 & \text{si la } k^{\text{e}} \text{ pièce tombe sur pile} \end{cases}$$

suit la loi de Bernoulli de paramètre p . De plus, X_1, \dots, X_n sont indépendantes. Ainsi, le nombre de faces obtenu lors de ces tirages est :

$$X = \sum_{k=1}^n X_k$$

et suit la loi $\mathcal{Bin}(10, p)$.

Proposition A.3. Soit $X \sim \mathcal{Bin}(n, p)$.

1. Pour tout $k \in \{0, \dots, n\}$, on a :

$$\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

2. On a :

$$\begin{aligned} \mathbf{E}[X] &= np, & \mathbf{V}[X] &= np(1-p), \\ L_X(t) &= (1-p+pe^t)^n & \text{et} & \quad \phi_X(t) = (1-p+pe^{it})^n. \end{aligned}$$

Exemple A.3 (suite). Reprenons l'Exemple A.3 et supposons que $p = 0,4$. On a alors :

$$\begin{aligned} \mathbf{P}[3 \leq X \leq 5] &= \mathbf{P}[X = 3] + \mathbf{P}[X = 4] + \mathbf{P}[X = 5] \\ &= \binom{10}{3} 0,4^3 (1-0,4)^7 + \binom{10}{4} 0,4^4 (1-0,4)^6 + \binom{10}{5} 0,4^5 (1-0,4)^5 \\ &\simeq 0,6665. \end{aligned}$$

On a 66,65% de chances d'observer entre 3 et 5 faces lors d'une série de 10 lancés. L'espérance de la variable X est $\mathbf{E}[X] = 10 \times 0,4 = 4$: si on répète un grand nombre de fois l'expérience, on s'attend à observer en moyenne 4 faces par série de 10 lancés.

A.1.4 Loi multinomiale

Il s'agit d'une généralisation de la loi binomiale au cas où l'on répète indépendamment n fois une expérience aléatoire n'ayant pas 2 mais K issues alternatives de probabilités de réalisations respectives p_1, \dots, p_K (avec $p_1 + \dots + p_K = 1$). On peut, par exemple, simplement penser à des lancés successifs d'un dé plutôt que d'une pièce.

Définition A.4. Si on note X_i le nombre de fois que la i^{e} alternative a été réalisés parmi les n essais de l'expérience aléatoire décrite ci-dessus, on dit que $X = (X_1, \dots, X_K)$ suit la loi multinomiale de paramètres n et p_1, \dots, p_K et on note $X \sim \text{Mult}(n; p_1, \dots, p_K)$.

Notons que la i^{e} marginale X_i suit la loi $\mathcal{B}in(n, p_i)$ et que celles-ci ne sont clairement pas indépendantes. Par exemple, les liens qu'elles entretiennent ont des conséquences sur son support : chacune des marginales peut prendre une valeur k_i entre 0 et n avec la contraintes $k_1 + \dots + k_K = n$.

Proposition A.4. Soit $X \sim \text{Mult}(n; p = (p_1, \dots, p_K))$.

1. Pour tout $k = (k_1, \dots, k_K) \in \{\{0, \dots, n\}^K : k_1 + \dots + k_K = n\}$, on a :

$$\mathbf{P}[X = k] = \frac{n!}{k_1! \dots k_K!} p_1^{k_1} \dots p_K^{k_K}.$$

2. On a :

$$\begin{aligned} \mathbf{E}[X] &= np, & \mathbf{V}[X] &= n \text{diag}(p_1, \dots, p_K) - n(p_i p_j)_{1 \leq i, j \leq K}, \\ L_X(t) &= \left(\sum_{j=1}^K p_j e^{t_j} \right)^n & \text{et} & \quad \phi_X(t) = \left(\sum_{j=1}^K p_j e^{it_j} \right)^n. \end{aligned}$$

A.1.5 Loi de Poisson

Lorsque le nombre d'épreuves n devient très important, la manipulation de la loi binomiale devient fastidieuse voire impossible. On peut alors remplacer son utilisation par celle de la *loi de Poisson* sous certaines conditions (Voir Théorème A.1). Celle-ci évalue le nombre aléatoire d'événements (rares) de même probabilité pendant une durée donnée comme, par exemple, le nombre d'appels reçus par un standard téléphonique en une heure, le nombre de voitures se présentant à un péage dans une journée, ...

Définition A.5. On dit qu'une v.a. X suit une loi de Poisson de paramètre $\lambda > 0$, si son support est $X(\Omega) = \mathbf{N}$ et, pour tout $k \in \mathbf{N}$:

$$\mathbf{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On note alors $X \sim \mathcal{P}(\lambda)$.

Proposition A.5. Si $X \sim \mathcal{P}(\lambda)$, on a :

$$\mathbf{E}[X] = \mathbf{V}[X] = \lambda,$$

$$L_X(t) = \exp(\lambda(e^t - 1)) \quad \text{et} \quad \phi_X(t) = \exp(\lambda(e^{it} - 1)).$$

Exemple A.4. Si, en moyenne, 10 voitures se présentent à un péage donné en une heure, on modélise le nombre de voiture se présentant au péage en une heure par une v.a. de loi $\mathcal{P}(10)$. Cette modélisation sera justifiée dans la suite.

Proposition A.6. Soient X_1 et X_2 deux v.a. indépendantes de lois de Poisson de paramètres respectifs λ_1 et λ_2 .

On a :

$$X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2).$$

Preuve : Pour tout $t \in \mathbf{R}$, on a :

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t) = \exp(\lambda_1(e^{it} - 1)) \exp(\lambda_2(e^{it} - 1)) = \exp((\lambda_1 + \lambda_2)(e^{it} - 1))$$

et on reconnaît la fonction caractéristique de $\mathcal{P}(\lambda_1 + \lambda_2)$. □

Approximation binomiale/Poisson Dans cette sous-section nous allons justifier le fait, évoqué ci-dessus, que l'on peut approcher certaines lois binomiales par une loi de Poisson.

Théorème A.1. Soit $\lambda > 0$, $X_n \sim \mathcal{B}in(n, \frac{\lambda}{n})$ et $Y \sim \mathcal{P}(\lambda)$.

On a :

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y.$$

Remarque A.3. On peut donc approcher la loi $\mathcal{B}in(n, p)$ par la loi $\mathcal{P}(n \times p)$. Dans la pratique, on ne fait une telle approximation que si p est proche de 0, $n \geq 30$, $p \leq 0,1$, $np \leq 10$, sans quoi l'approximation est mauvaise.

Exemple A.4 (suite). Justifions l'utilisation de la loi de Poisson $\mathcal{P}(10)$ dans l'Exemple A.4. Supposons que l'on observe s'il y a eu une arrivée de voiture seulement à n instants fixés dans l'heure (n est voué à être grand). Alors, la variable X_n comptant le nombre d'instants où l'on a observé une arrivée de voiture suit une loi binomiale $\mathcal{B}in(n, p)$. Puisqu'en moyenne on observe 10 arrivées de voitures en une heure, on a $np = 10$, soit $p = \frac{10}{n}$. Plus n est grand, plus la discrétisation du temps est fine et s'approche de la réalité. L'idée est donc de faire tendre n vers l'infini. Le Théorème A.1 affirme que la loi limite est la loi $\mathcal{P}(10)$.

A.1.6 Loi géométrique

Considérons une expérience de Bernoulli dont la probabilité de succès est $p \in]0; 1[$. Si l'on répète plusieurs fois indépendamment cette expérience, on peut s'intéresser au nombre aléatoire de répétition de l'expérience nécessaires pour obtenir un premier succès.

Définition A.6. *Considérons une expérience de Bernoulli dont la probabilité de succès est $p \in]0; 1[$ que l'on répète jusqu'au premier succès. On appelle loi géométrique de paramètre $p \in]0; 1[$ la loi du rang du premier succès. Cette loi est notée $\mathcal{G}(p)$.*

Proposition A.7. *Soit $X \sim \mathcal{G}(p)$.*

1. *Le support de X est $X(\Omega) = \mathbf{N}^*$ et pour tout $k \in \mathbf{N}^*$, on a :*

$$\mathbf{P}[X = k] = (1 - p)^{k-1}p.$$

2. *On a :*

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

$$L_X(t) = \frac{pe^t}{1 - (1-p)e^t} \quad \text{et} \quad \phi_X(t) = \frac{pe^{it}}{1 - (1-p)e^{it}}.$$

Exemple A.5. Une personne rentre ivre chez elle et prélève au hasard une clef dans son trousseau, en contenant 5, pour tenter d'ouvrir la porte. Si elle échoue, elle remet la clef dans son trousseau et recommence. Le nombre de tentatives X jusqu'à l'ouverture de la porte suit alors une loi géométrique de paramètre $\frac{1}{5} = 0,2$. En moyenne, la porte sera ouverte après 5 tentative et la probabilité qu'elle soit ouverte après k tentatives est :

$$\mathbf{P}[X = k] = 0,2 \times 0,8^{k-1}.$$

A.1.7 Loi binomiale négative

Considérons une expérience de Bernoulli dont la probabilité de succès est $p \in]0; 1[$ que l'on répète indépendamment jusqu'à l'obtention du n^e succès. Alors, la loi binomiale négative de paramètre n et p est la loi de la v.a. donnant le nombre X d'échecs observés avant l'obtention du n^e succès. En utilisant le caractère i.i.d. des essais et en choisissant les places des échecs parmi les $n + k - 1$ possibles (le $n + k^e$ sera un succès), on obtient la probabilité pour que $X = k \in \mathbf{N}$:

$$\mathbf{P}[X = k] = \binom{n+k-1}{k} (1-p)^k p^n.$$

Notons que l'on peut étendre la formule précédente en remplaçant le paramètre $n \in \mathbf{N}^*$ par un réel $r > 0$ et définir cette loi de façon plus générale (même si l'interprétation de la loi est alors moins immédiate).

Définition A.7. On dit que X suit la loi binomiale négative ou de Pólya de paramètres $r > 0$ et $p \in]0; 1[$ si son support est \mathbf{N} et si pour tout $k \in \mathbf{N}$:

$$\mathbf{P}[X = k] = \frac{\Gamma(r+k)}{k!\Gamma(r)}(1-p)^k p^r.$$

Cette loi est notée $\mathcal{BN}(r, p)$.

Remarque A.4. Si $X \sim \mathcal{BN}(1, p)$, alors $X + 1 \sim \mathcal{G}(p)$ (et réciproquement).

Proposition A.8. Soit $X \sim \mathcal{BN}(r, p)$.

$$\mathbf{E}[X] = \frac{r(1-p)}{p}, \quad \mathbf{V}[X] = \frac{r(1-p)}{p^2},$$

$$L_X(t) = \left(\frac{p}{1 - (1-p)e^t} \right)^r \quad \text{et} \quad \phi_X(t) = \left(\frac{p}{1 - (1-p)e^{it}} \right)^r.$$

Le résultat suivant (basé sur un calcul explicite) justifie la terminologie « binomiale négative » en montrant que si $X \sim \mathcal{BN}(n, p)$, $\mathbf{P}[X \leq k]$ n'est autre que la probabilité pour qu'il y ait eu au moins n succès après $n+k$ épreuves de Bernoulli indépendantes de paramètre p .

Proposition A.9. Soit $X \sim \mathcal{BN}(n, p)$ et $Y \sim \mathcal{Bin}(n+k, p)$, $n \in \mathbf{N}^*$, $k \in \mathbf{N}$, $p \in [0, 1]$. Alors,

$$\mathbf{P}[X \leq k] = \mathbf{P}[Y \geq n].$$

Exercice A.1. Soit $X_1 \sim \mathcal{BN}(n_1, p)$ et $X_2 \sim \mathcal{BN}(n_2, p)$ deux v.a. indépendantes.

Déterminer la loi de $X_1 + X_2$.

Indication : On pourra utiliser les fonctions caractéristiques.

A.1.8 Loi hypergéométrique

Considérons une urne contenant $A \in \mathbf{N}^*$ boules dont pA boules dites gagnantes (pA étant supposé entier, $p \in [0, 1]$).

Définition A.8. On tire dans cette urne simultanément $n \leq A$ boules. On dit que X suit la loi hypergéométrique de paramètres n , p et A si X donne le nombre de boules gagnantes ainsi tirées. On note alors $X \sim \mathcal{H}(n, p, A)$.

Proposition A.10. Soit $X \sim \mathcal{H}(A, p, n)$.

1. Pour tout $k \in \{0, \dots, n\}$, on a :

$$\mathbf{P}[X = k] = \frac{\binom{pA}{k} \binom{(1-p)A}{n-k}}{\binom{A}{n}}.$$

2. On a :

$$\mathbf{E}[X] = np, \quad \mathbf{V}[X] = np(1-p) \frac{A-n}{A-1},$$

$$L_X(t) = \frac{\binom{(1-p)A}{n}}{\binom{A}{n}} {}_2F_1(-n, -p; (1-p)A - n + 1; e^t)$$

et

$$\phi_X(t) = \frac{\binom{(1-p)A}{n}}{\binom{A}{n}} {}_2F_1(-n, -p; (1-p)A - n + 1; e^{it}),$$

où ${}_2F_1$ désigne la fonction hypergéométrique de Gauss

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$$

avec $(a)_0 = 1$ et $(a)_n = a(a+1)\dots(a+n-1)$, $n \geq 1$.

A.2 Lois continues usuelles

A.2.1 Loi uniforme continue

Définition A.9. Soient $a < b$ deux réels.

On dit qu'une v.a. X suit la loi uniforme sur $[a; b]$ si sa densité est donnée par :

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

On note alors $X \sim \mathcal{U}([a; b])$.

Remarque A.5. Il s'agit de l'analogie continue de la loi uniforme discrète.

Proposition A.11. Soit $X \sim \mathcal{U}([a; b])$.

1. La fonction de répartition F de X est donnée par :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x \geq b \end{cases}.$$

2. On a :

$$\mathbf{E}[X] = \frac{a+b}{2}, \quad \mathbf{V}[X] = \frac{(b-a)^2}{12},$$

$$L_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)} \quad \text{et} \quad \phi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

A.2.2 Loi exponentielle

Définition A.10. Soit $\lambda > 0$.

On dit qu'une v.a. X suit la loi exponentielle de paramètre λ si sa densité est donnée par :

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0;+\infty[}(x), \quad x \in \mathbf{R}.$$

On note alors $X \sim \mathcal{E}(\lambda)$.

Remarque A.6. Elle est utilisée pour modéliser des phénomènes *sans mémoire* ou *sans vieillissement* tels que le temps d'attente avant le prochain tremblement de terre ou la prochaine désintégration dans un réacteur nucléaire ou encore la durée de vie de certains appareils comme des ampoules. Ceci est justifié par le deuxième point de la Proposition A.12.

Proposition A.12. Soit $X \sim \mathcal{E}(\lambda)$, $\lambda > 0$.

1. La fonction de répartition F_X de X est donnée par :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases}.$$

2. [Perte de mémoire] Pour tous $s, t \geq 0$:

$$\mathbf{P}[X > s + t | X > t] = \mathbf{P}[X > s].$$

3. On a :

$$\mathbf{E}[X] = \frac{1}{\lambda} \quad \mathbf{V}[X] = \frac{1}{\lambda^2},$$

$$L_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} \quad \text{et} \quad \phi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$

Preuve du point 2. : On a :

$$\begin{aligned} \mathbf{P}[X > s + t | X > t] &= \frac{\mathbf{P}[X > s + t, X > t]}{\mathbf{P}[X > t]} = \frac{\mathbf{P}[X > s + t]}{\mathbf{P}[X > t]} \\ &= \frac{1 - \mathbf{P}[X \leq s + t]}{1 - \mathbf{P}[X \leq t]} = \frac{1 - F_X(s + t)}{1 - F_X(t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} = 1 - F_X(s) = \mathbf{P}[X > s]. \end{aligned}$$

□

Exemple A.6. On a observé que la durée de vie d'une ampoule d'un modèle donné est d'en moyenne 1000 heures. Considérons une ampoule de ce modèle et intéressons nous à sa durée de vie X (exprimée en heures). La v.a. X est continue et sans mémoire. On considère donc que X suit une loi exponentielle. Puisque l'on s'attend à avoir une durée de vie moyenne de 1000 heures, le paramètre de cette loi exponentielle est $\lambda = \frac{1}{1000}$ de sorte que $\mathbf{E}[X] = \frac{1}{\lambda} = 1000$.

Ainsi, la probabilité pour que l'ampoule fonctionne au plus 100h est :

$$\mathbf{P}[X \leq 100] = 1 - e^{-\frac{1}{1000} \times 100} = 1 - e^{-\frac{1}{10}} \simeq 0,01.$$

De même, la probabilité pour que l'ampoule fonctionne plus de 4500 heures est :

$$\mathbf{P}[X > 4500] = 1 - \mathbf{P}[X \leq 4500] = e^{-\frac{1}{1000} \times 4500} = 1 - e^{-4,5} \simeq 0,01.$$

A.2.3 Loi gamma

Définition A.11. On dit qu'une v.a. X suit la loi gamma de paramètres $r > 0$ et $\lambda > 0$ si sa densité est donnée par :

$$f_X(x) = \frac{x^{r-1} \lambda^r e^{-\lambda x}}{\Gamma(r)} \mathbf{1}_{[0, +\infty[}(x), \quad x \in \mathbf{R},$$

où $\Gamma(\cdot)$ désigne la fonction gamma d'Euler :

$$\Gamma(r) = \int_0^{+\infty} t^{r-1} e^{-t} dt.$$

On note alors $X \sim \Gamma(r; \lambda)$.

L'exercice suivant donne l'interprétation pratique de cette loi.

Exercice A.2. Montrer que si X_1, \dots, X_k sont k v.a.i.i.d. de loi $\mathcal{E}(\lambda)$, alors $X = X_1 + \dots + X_k$ suit la loi $\Gamma(k; \lambda)$.

Proposition A.13. Soit $X \sim \Gamma(r; \lambda)$, $r, \lambda > 0$.

On a :

$$\mathbf{E}[X] = \frac{r}{\lambda}, \quad \mathbf{V}[X] = \frac{r}{\lambda^2},$$

$$L_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}, \quad (\text{si } t \leq \lambda^{-1}) \quad \text{et} \quad \phi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-r}.$$

Remarque A.7. La fonction de répartition de cette loi n'a pas de forme explicite agréable et nécessite l'utilisation de tables ou d'un logiciel (voir `pgamma` sous R).

A.2.4 Loi beta

Définition A.12. On dit qu'une v.a. X suit la loi beta de paramètres $p, q > 0$ si sa densité est donnée par :

$$f_X(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p, q)} \mathbf{1}_{0 < x < 1}, \quad x \in \mathbf{R},$$

où $B(\cdot, \cdot)$ désigne l'intégrale beta :

$$B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

On note alors $X \sim \beta(p; q)$.

Proposition A.14. Soit $X \sim \beta(p; q)$, $p, q > 0$.

On a :

$$\mathbf{E}[X] = \frac{p}{p+q} \quad \text{et} \quad \mathbf{V}[X] = \frac{pq}{(p+q)^2(p+q+1)}.$$

Proposition A.15. Soient $X \sim \Gamma(p; \lambda)$ et $Y \sim \Gamma(q; \lambda)$ deux v.a. indépendantes.

Alors,

1. $X/(X+Y) \sim \beta(p; q)$;
2. $X/(X+Y)$ et $X+Y$ sont indépendantes.

A.2.5 Loi normale

Définition A.13. On dit qu'une v.a. X suit la loi normale (ou gaussienne) de moyenne m et de variance σ^2 si sa densité est donnée par :

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

On note alors $X \sim \mathcal{N}(m; \sigma^2)$.

Remarque A.8.

1. La fonction de répartition d'une loi normale n'a pas de forme analytique close (autre que son expression intégrale. On a donc recours à des tables (de la loi normale centrée réduite) et à un changement de variable ou à l'utilisation de logiciels pour le calcul de ses valeurs (voir `pnorm` sous R).
2. Soit $X \sim \mathcal{N}(0; 1)$.

- (a) La fonction densité f_X de X est paire *i.e.* $f_X(-x) = f_X(x)$ pour tout réel x . En particulier, on a :

$$F_X(-x) = 1 - F_X(x).$$

- (b) $Y = \sigma X + m \sim \mathcal{N}(m; \sigma^2)$ et réciproquement si $Y \sim \mathcal{N}(0; 1)$, $Z = \sigma^{-1}(Y - m) \sim \mathcal{N}(0; 1)$.

Proposition A.16. Soit $X \sim \mathcal{N}(m; \sigma^2)$.

Alors,

$$\mathbf{E}[X] = m, \quad \mathbf{V}[X] = \sigma^2.$$

$$L_X(t) = \exp\left(mt + \frac{\sigma^2 t^2}{2}\right) \quad \text{et} \quad \phi_X(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right).$$

Proposition A.17 (Proposition 3.7). Soit $X_i, i = 1, \dots, d$, des v.a. indépendantes suivant respectivement la loi normale $\mathcal{N}(m_i, \sigma_i^2)$ et $\alpha_1, \dots, \alpha_n \in \mathbf{R}$.

Alors, $Y = \alpha_1 X_1 + \dots + \alpha_d X_d$ suit la loi normale de moyenne $\sum_{i=1}^d \alpha_i m_i$ et de variance $\sum_{i=1}^d \alpha_i^2 \sigma_i^2$.

En vertu du TCL, la concentration de la masse autour de la moyenne pour des v.a. normales est d'intérêt particulier. Donnons quelques valeurs d'usage courant :

$$\mathbf{P}[m - \sigma \leq X \leq m + \sigma] \simeq 0,6827,$$

$$\mathbf{P}[m - 2\sigma \leq X \leq m + 2\sigma] \simeq 0,9545$$

et

$$\mathbf{P}[m - 3\sigma \leq X \leq m + 3\sigma] \simeq 0,9973.$$

A.2.6 Loi normale multidimensionnelle

Voir Section 3.6.

A.2.7 Loi log-normale

Définition A.14. On dit qu'une v.a. X suit la loi log-normale de moyenne m et de variance σ^2 si $X = \ln(Y)$ pour $Y \sim \mathcal{N}(m; \sigma^2)$.

On note alors $X \sim \text{Log-}\mathcal{N}(m; \sigma^2)$.

Remarque A.9.

1. Dans le contexte multidimensionnel, on définit de manière analogue la loi log-normale $X \sim \text{Log-}\mathcal{N}(m; \Sigma)$ comme celle de $X = \ln(Y)$ pour $Y \sim \mathcal{N}_d(m; \Sigma)$.

2. Cette loi fournit de bon modèles pour les v.a. strictement positives asymétriques à queues lourdes.

Proposition A.18. Soit $X \sim \text{Log} -\mathcal{N}(m; \sigma^2)$.

Alors,

$$f_X(x) = \frac{1}{\sqrt{x}2\pi\sigma^2} e^{-\frac{(\ln(x)-m)^2}{2\sigma^2}}, \quad x \in \mathbf{R}.$$

$$\mathbf{E}[X] = e^{m+\frac{\sigma^2}{2}} \quad \text{et} \quad \mathbf{V}[X] = e^{2m+\sigma^2}(e^{\sigma^2} - 1).$$

A.2.8 Loi de Pareto

Définition A.15. Soit $a, \theta > 0$.

On dit qu'une v.a. X suit la loi de Pareto de paramètres a et θ si sa densité est donnée par :

$$f_X(x) = \frac{\theta}{a} \left(\frac{a}{x}\right)^{\theta+1} \mathbf{1}_{[a;+\infty[}(x), \quad x \in \mathbf{R}.$$

On note alors $X \sim \mathcal{P}ar(a, \theta)$.

Remarque A.10. Elle est par exemple utilisée pour modéliser des la distribution de revenus supérieurs à un seuil donné, la performance de réseaux, mais aussi en gestion de qualité ou en réassurance.

Proposition A.19. Soit $X \sim \mathcal{P}ar(a, \theta)$, $a, \theta > 0$.

1. La fonction de répartition F_X de X est donnée par :

$$F_X(x) = \left(1 - \left(\frac{a}{x}\right)^\theta\right) \mathbf{1}_{[a;+\infty[}(x).$$

2. On a :

$$\mathbf{E}[X] = \frac{a\theta}{\theta-1} \quad (\theta > 1) \quad \mathbf{V}[X] = \frac{a^2\theta}{(\theta-1)^2(\theta-2)} \quad (\theta > 2),$$

et

$$\phi_X(t) = \theta(-iat)^\theta \Gamma(-\theta, -iat).$$

Remarque A.11. Sa fonction génératrice des moments n'est pas définie.

A.2.9 Loi de Cauchy

Il s'agit de l'exemple typique de loi sans moment, en particulier les lois des grands nombres et le TCL ne sont pas valables pour cette loi.

Définition A.16. Soit $a > 0$ et $m \in \mathbf{R}$.

On dit qu'une v.a. X suit la loi de Cauchy de paramètres a et m si sa densité est donnée par :

$$f_X(x) = \frac{a}{\pi((x-m)^2 + a^2)}, \quad x \in \mathbf{R}.$$

On note alors $X \sim \mathcal{C}(a, m)$.

Proposition A.20. Soit $X \sim \mathcal{C}(a, m)$, $a > 0$, $m \in \mathbf{R}$.

1. La fonction de répartition F_X de X est donnée par :

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-m}{a}\right).$$

2. On a :

$$\phi_X(t) = \exp(imt - a|t|).$$

Remarque A.12. Son espérance et *a fortiori* sa variance ainsi que sa fonction génératrice des moments ne sont pas définis. Le paramètre m est un paramètre de position ; plus précisément, il s'agit de la médiane de cette loi.

A.2.10 Loi de Weibull

Avec les lois de Gumbel et de Fréchet, il s'agit d'une loi des valeurs extrêmes. Elle généralise la loi exponentielle pour modéliser les durées de vie et s'obtient comme transformation de celle-ci.

Définition A.17. Soit $\alpha, \lambda > 0$.

On dit qu'une v.a. X suit la loi de Weibull de paramètres λ et α si sa densité est donnée par :

$$f_X(x) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} \mathbf{1}_{[0;+\infty[}(x), \quad x \in \mathbf{R}.$$

On note alors $X \sim \mathcal{W}(\lambda, \alpha)$.

Proposition A.21. Soit $X \sim \mathcal{W}(\lambda, \alpha)$, $\alpha, \lambda > 0$.

1. La fonction de répartition F_X de X est donnée par :

$$F_X(x) = \left(1 - e^{-\lambda x^\alpha}\right) \mathbf{1}_{[0;+\infty[}(x).$$

2. On a :

$$\mathbf{E}[X] = \frac{\Gamma(1 + \alpha^{-1})}{\lambda^{\frac{1}{\alpha}}} \quad \text{et} \quad \mathbf{V}[X] = \frac{\Gamma(1 + 2\alpha^{-1}) - \Gamma(1 + \alpha^{-1})^2}{\lambda^{\frac{2}{\alpha}}}.$$

3. Si $Y \sim \mathcal{E}(\lambda)$ alors $Y^{\alpha^{-1}} \sim \mathcal{W}(\lambda, \alpha)$.

A.2.11 Loi de Gumbel

Il s'agit d'une autre loi des valeurs extrêmes apparaissant notamment dans l'étude du maximum de n observations de v.a. lorsque n tend vers l'infini.

Définition A.18. Soient $\mu \in \mathbf{R}$ et $\beta > 0$.

On dit qu'une v.a. X suit la loi de Gumbel de paramètres μ et β si sa densité est donnée par :

$$f_X(x) = \frac{\exp\left(-\frac{x-\mu}{\beta}\right) \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)}{\beta}, \quad x \in \mathbf{R}.$$

On note alors $X \sim \mathcal{Gum}(\mu; \beta)$.

Proposition A.22. Soit $X \sim \text{Gum}(\mu; \beta)$, $\mu \in \mathbf{R}$, $\beta > 0$.

1. La fonction de répartition F_X de X est donnée par :

$$F_X(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right).$$

2. On a :

$$\mathbf{E}[X] = \mu + \beta\gamma,$$

avec $\gamma = \lim_{n \rightarrow \infty} \sum_{k=1}^n k^{-1} - \ln(n)$ la constante d'Euler,

$$\mathbf{V}[X] = \frac{\pi^2\beta^2}{6},$$

$$L_X(t) = \Gamma(1 - \beta t)e^{\mu t} \quad \text{et} \quad \phi_X(t) = \Gamma(1 - i\beta t)e^{i\mu t}.$$

A.2.12 Loi de Fréchet

Il s'agit de la troisième loi classique des valeurs extrêmes.

Définition A.19. Soient $\mu \in \mathbf{R}$ et $\alpha, s > 0$.

On dit qu'une v.a. X suit la loi de Fréchet de paramètres μ , α et s si sa densité est donnée par :

$$f_X(x) = \frac{\alpha}{s} \left(\frac{x-\mu}{s}\right)^{-\alpha-1} \exp\left(-\left(\frac{x-\mu}{s}\right)^{-\alpha}\right) \mathbf{1}_{x \geq \mu}.$$

On note alors $X \sim \mathcal{F}re(\mu, \alpha, s)$.

Proposition A.23. Soit $X \sim \mathcal{F}re(\mu; \beta)$, $\mu \in \mathbf{R}$, $\beta > 0$.

1. La fonction de répartition F_X de X est donnée par :

$$F_X(x) = \exp\left(-\left(\frac{x-\mu}{s}\right)^{-\alpha}\right) \mathbf{1}_{x \geq \mu}.$$

2. On a :

$$\mathbf{E}[X] = \mu + s\Gamma(1 - \alpha^{-1}) \quad (\text{si } \alpha > 1)$$

et

$$\mathbf{V}[X] = s^2(\Gamma(1 - 2\alpha^{-1}) - \Gamma(1 - \alpha^{-1})^2) \quad (\text{si } \alpha > 2).$$

Remarque A.13. Le k^{e} moment de cette loi existe ssi $\alpha > k$.

A.2.13 Loi du Khi-2

Commençons par établir un résultat dont va découler la définition de la loi du Khi-2 (χ^2) à ν degrés de liberté et qui conduira aux nombreuses applications de cette loi en statistique.

Proposition A.24. Soient Z_1, \dots, Z_ν i.i.d. de loi $\mathcal{N}(0, 1)$.

Alors,

$$X = \sum_{k=1}^{\nu} Z_k^2 \sim \Gamma(\nu/2, 2).$$

Preuve : On a :

$$\begin{aligned}
 L_{Z_1^2}(t) &= \mathbf{E} \left[e^{tZ_1^2} \right] \\
 &= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \exp \left(-\frac{1}{2}(1-2t)z^2 \right) dz \\
 &= \frac{1}{\sqrt{1-2t}} \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \exp \left(-\frac{s^2}{2} \right) ds \quad (s := \sqrt{1-2t}z, ds = \sqrt{1-2t} dz) \\
 &= \frac{1}{\sqrt{1-2t}}.
 \end{aligned}$$

Comme les Z_k sont indépendantes, les Z_k^2 le sont, et on a par le Théorème 3.4 :

$$L_X(t) = \left(\frac{1}{\sqrt{1-2t}} \right)^\nu = (1-2t)^{-\frac{\nu}{2}}.$$

On conclue en reconnaissant ici la fonction génératrice des moments de la loi $\Gamma(\nu/2, 2)$ (voir Proposition A.13). \square

Définition A.20. On appelle loi du Khi-2 à ν degrés de liberté la loi $\Gamma(\nu/2, 2)$. On la note $\chi^2(\nu)$.

Le résultat suivant liste des conséquences immédiates des Définitions A.11 et A.20 et des Propositions A.13 et A.24.

Proposition A.25. Soit $X \sim \chi^2(\nu)$, $\nu \in \mathbf{N}^*$.

On a :

1. La densité de X est donnée par :

$$f_X(x) = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{x>0}.$$

2. On a :

$$\mathbf{E}[X] = \nu \quad \mathbf{V}[X] = 2\nu,$$

$$L_X(t) = (1-2t)^{-\frac{\nu}{2}} \quad (t < 1/2) \quad \text{et} \quad \phi_X(t) = (1-2it)^{-\frac{\nu}{2}}.$$

3. Si $X_1 \sim \chi^2(\nu_1)$ et $X_2 \sim \chi^2(\nu_2)$, $\nu_1, \nu_2 \in \mathbf{N}^*$, sont indépendantes alors $X_1 + X_2 \sim \chi^2(\nu_1 + \nu_2)$.

Remarque A.14. La fonction de répartition de cette loi s'exprime en terme de la fonction gamma incomplète et est peu maniable. On a donc recours à l'utilisation de tables ou de logiciels. Dans la pratique, sous R, nous utiliserons la fonction `pchisq`.

A.2.14 Loi de Student

La loi de Student a été introduite, ainsi que le test du même nom, par William Gosset qui, ingénieur chez Guinness, n'a pu publier à ce sujet que sous le nom d'emprunt de Student.

Définition A.21. Soient $Z \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(\nu)$ deux v.a. indépendantes.

On appelle loi de Student à ν degrés de liberté (d.d.l) et on note $\mathcal{T}(\nu)$ la loi de :

$$T = \frac{Z}{\sqrt{\frac{Y}{\nu}}}.$$

Proposition A.26. Soit $T \sim \mathcal{T}(\nu)$, $\nu \in \mathbf{N}^*$.

On a :

1. La densité de T est donnée par :

$$f_T(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

2. On a :

$$\mathbf{E}[T] = 0 \quad (\nu \geq 2) \quad \text{et} \quad \mathbf{V}[T] = \frac{\nu}{\nu-2} \quad (\nu \geq 3).$$

Remarque A.15.

1. La fonction de répartition de cette loi ou ses fonctions génératrices des moments ou caractéristique ne s'exprime pas de façon agréable. On a donc recours à l'utilisation de tables ou de logiciels. Dans la pratique, sous R, nous utiliserons les fonctions `pt`, `dt`,...
2. Lorsque $\nu = 1$, on retrouve la loi de Cauchy $\mathcal{Cau}(0, 1)$ qui n'admet pas de moment d'ordre 1.

A.2.15 Loi de Fisher-Snedecor

La loi définition même de la loi de Fisher-Snedecor laisse pressentir son intérêt en statistique lorsque l'on souhaite étudier le rapport de sommes de gaussiennes indépendantes.

Définition A.22. Soient $X_1 \sim \chi^2(\nu_1)$ et $X_2 \sim \chi^2(\nu_2)$ deux v.a. indépendantes.

On appelle loi de Fisher-Snedecor à ν_1 d.d.l. au numérateur et ν_2 d.d.l. au dénominateur et on note $\mathcal{F}(\nu_1, \nu_2)$ la loi de :

$$F = \frac{\nu_2 X_1}{\nu_1 X_2}.$$

La proposition suivante découle directement de la définition de la loi de Fisher-Snedecor.

Proposition A.27. Soit $F \sim \mathcal{F}(\nu_1, \nu_2)$. Alors, $1/F \sim \mathcal{F}(\nu_2, \nu_1)$.

Remarque A.16.

1. Les fonctions de densité, de répartition, génératrices des moments ou caractéristique de cette loi ne s'exprime pas de façon très agréable. Dans la pratique, sous R, nous utiliserons les fonctions `pf`, `df`,...
2. Cette loi admet un moment d'ordre 1 ssi $\nu_2 \geq 3$. Son espérance est alors $\nu_2(\nu_2 - 2)^{-1}$. Sa variance est définie ssi $\nu_2 \geq 5$.

Annexe B

Quelques mots sur R

R est à la fois un environnement de manipulation et de traitement de données particulièrement adapté aux statistiques (et d'usage important) et un langage de programmation autonome et interprété. Il peut donc être utilisé comme une « grosse calculatrice » ou à des fins de programmation. Basé sur une écriture vectorielle, on peut très souvent éviter l'emploi de boucles et produire des codes assez courts. La programmation sous R est souple, en particulier les objets ne sont pas typés (il n'y a pas de déclaration de variables à faire). Notons que de nombreux packages R ont été développés et permettent une utilisation facile et rapide des méthodes statistiques en général, même les plus récentes.

Nous ne donnons ici qu'un aperçu minimal des différentes commandes et fonctionnalités de R utiles pour illustrer les propos de ce cours. Il existe une large littérature et des aides en ligne sur R qu'il est possible de consulter pour de plus amples détails. Par exemple, on peut citer [10] qui constitue une introduction à la programmation en R ou [1, 6] qui sont des ouvrages dédiés à la statistique sous R.

B.1 Création, lecture et sauvegarde de données

`c(argument)` : combine les arguments pour former un vecteur

`n:m` : crée un vecteur d'entiers allant de n à m ; opération prioritaire

`seq(a,b)` : génère une séquence pour laquelle on peut spécifier l'incrément (`by=`) et la longueur (`length=`)

`"chaîne"` : crée une chaîne de caractères

`matrix(x,nrow=,ncol=)` : crée une matrice ; les éléments se répètent s'ils sont trop courts

`rbind(arguments)` : combine les arguments par ligne

`cbind(arguments)` : combine les arguments par colonne

`data.frame(argument)` crée un data frame (tableau dont les colonnes peuvent être de types différents) avec les arguments

`list(arguments)` : crée une liste avec les arguments (nommés ou non, qui peuvent être de longueur différente)

`save("fichier", x,y)` enregistre les objets x et y dans le fichier (format propre à R)

`scan("fichier")` : lit le contenu de `"fichier"` et le transcrit en un vecteur

`load()` : charge le jeu de données écrit avec `save`

`data(x)` : charge le jeu de données x

`read.table(file)` : lit un fichier au format tabulaire et en fait un data frame; séparateur de colonne par défaut `sep=""`; pour prendre la première ligne comme titre de colonne : `header=TRUE`; pour empêcher les vecteurs de caractères d'être transformés en facteurs (`as.is=TRUE`); ...

`read.csv2("filename",header=TRUE)` : idem mais avec des options pré-définies pour lire les fichiers CSV

`save.image("fichier")` : enregistre tous les objets

B.2 Extraction de donnée

Pour les vecteurs

`x[n]` : n^{e} élément de x

`x[-n]` : supprime le n^{e} élément de x

`x[1:n]` : n premiers éléments de x

`x[-(1:n)]` : supprime les n premiers éléments de x

Pour les matrices

`x[i,j]` : élément de la ligne i colonne j

`x[i,]` : ligne i

`x[,j]` : colonne j

B.3 Opération de base

`<-`, `->` : affectation dans le sens des flèches

`+`, `-`, `*`, `/` : opérations terme à terme; si x est de longueur nm et y de longueur n , répète m fois y pour l'effectuer

`%*%` : produit matriciel

`rev(x)` : renverse l'ordre des éléments de x

`sort(x)` : trie les éléments de x par ordre croissant

`t(x)` : transposée de x

`solve(A)` : inverse de la matrice A

`solve(A,B)` : résout en X l'équation $AX = B$

B.4 Fonctions mathématiques

`sin`, `cos`, `tan`, `log`, `log10`, `exp`, `max`, `min`, `abs`... : ce que l'on imagine

`round(x,d)` : arrondi les éléments de x à d décimales

`sum(x)` : somme de tous les éléments de x

`rowSums(x)` : sommes par ligne

`colSums(x)` : sommes par colonne

`prod(x)` : produit des éléments de x

`median(x)` : médiane des éléments de x

`mean(x)` : moyenne des éléments de x

`rowMeans(x)` : moyennes par ligne de x
`colMeans(x)` : moyennes par colonne de x
`weighted.mean(x,p)` : moyenne des éléments de x pondérés par p
`var(x)` ou `cov(x)` : variance *corrigée* des éléments de x (division par $n - 1$)
`sd(x)` : écart-type *corrigée* des éléments de x
`var(x,y)` ou `cov(x,y)` : covariance de x et y
`cor(x)` : coefficient de corrélation linéaire de x et y

B.5 Fonctions probabilistes

`d"suffixe de loi"` : densité de la loi
`p"suffixe de loi"` : fonction de répartition de la loi
`q"suffixe de loi"` : quantiles de la loi
`r"suffixe de loi"` : génération de nombres aléatoires selon la loi

Loi	Suffixe	Paramètre(s)
$\beta(p, q)$	beta	shape1= p , shape2= q
$\text{Bin}(n, p)$	binom	size= n , prob= p
$\mathcal{C}(a, m)$	cauchy	location= m , scale= a
$\mathcal{E}(\lambda)$	exp	rate= λ^{-1}
$\mathcal{F}(\nu_1, \nu_2)$	f	df1= ν_1 , df2= ν_2
$\mathcal{G}(p)$	geom	prob= p
$\Gamma(r, \lambda)$	gamma	shape= r , scale= λ
Log — $\mathcal{N}(m, \sigma^2)$	lnorm	mean= m , sd= σ
$\text{Mult}(n, p)$	multinom	size= n , prob= p
$\mathcal{N}(m, \sigma^2)$	norm	mean= m , sd= σ
$\mathcal{P}(\lambda)$	pois	mean= λ
$\mathcal{T}(\nu)$	t	df= ν
$\mathcal{U}([a, b])$	unif	min= a , max= b
$\mathcal{W}(\lambda, \alpha)$	weibull	shape= α , scale= λ
$\chi^2(\nu)$	chisq	df= ν

B.6 Graphiques

`windows()` : ouvre une fenêtre graphique sous Windows
`x11()` : ouvre une fenêtre graphique sous GNU/linux ou MacOSX
`pdf(file)`, `png(file)`, `jpeg(file)`, `bmp(file)`, `tiff(file)` : se prépare à écrire les instructions graphiques qui suivront dans le fichier `file`, au format désigné (pdf ou png recommandés); `width=` et `height=` fixent les dimensions
`dev.off()` : ferme la fenêtre graphique ou le fichier graphique spécifié (par défaut : celui en cours)
`plot(x)` : graphique de x (différents effets selon l'objet)
`plot(x,y)` : nuage de points
`hist(x)` : histogramme des fréquences de x
`barplot(x)` : diagramme en barres
`pie(x)` : diagramme circulaire

`boxplot(x)` : boîte à moustaches ;
`sunflowerplot(x, y)` : comme `plot(x,y)` mais les points qui se superposent exactement sont représentés avec des fleurs (un pétale par valeur répétée)
`coplot(y~x | a)` : nuage des points de coordonnées x, y pour chaque valeur ou intervalle de valeur de a

Paramètres de fonctions graphiques

`add=TRUE` ajoute sur le graphique précédent `axes=FALSE` ne trace pas les axes
`type="p"` : type de représentation des coordonnées ; "p" : points, "l" : lignes, "b" : (both) points et lignes, ...
`xlim=`, `ylim=` : limites des zones du graphique,
`xlab=`, `ylab=` : titre des axes
`main=` titre du graphique
`sub=` : sous-titre
`par(...)` : définit les paramètres suivants pour les graphiques à venir
`col=` :couleur(s) des symboles et lignes
`lty` : type de ligne

Ajout d'éléments à un graphique existant

`points(x, y)` : ajoute des points
`lines(x, y)` : ajoute des lignes
`curve(f(x),add=T)` : ajoute la courbe de f
`text(x, y, "texte", ...)` : ajoute du texte aux coordonnées (x, y) ;
`segments(x0, y0, x1, y1)` : trace le segment $[(x_0, y_0); (x_1, y_1)]$
`abline(a,b)` : trace une droite $y = a + bx$
`legend(x, y, legend)` : ajoute une légende au point (x, y) avec les symboles donnés par `legend`
`box()` : encadre le graphique

Faire plusieurs graphiques dans la même fenêtre

`layout(matrix(1:nm,n,m))` : découpe la fenêtre en nm blocs de graphiques (n nombre de lignes, m nombre de colonnes)
`par()` : permette de combiner les graphiques avec différentes surfaces ; `fig =` permet de définir la région où sera le graphique ; `new = T` indique qu'on pourra tracer ensuite un nouveau graphique là où l'on veut ; `mar =` définit les marges autour du graphique

Superposer des graphiques

`par(new=T)` et on peut ajouter les axes, ...

B.7 Programmation

`function(arglist) expr` : définition de fonction ; `arglist` est une liste d'arguments, `expr` est une expression exécutée ;

`return(value)` : mis dans `expr` lors d'une définition de fonction, indique que la fonction doit renvoyer ce résultat (sinon la fonction renvoie la dernière valeur calculée dans `expr`)

`if(cond) expr` : ce que l'on imagine ; opérateurs de comparaison : `==` `!=` `<` `>` `<=` `>=`

`if(cond) cons.expr else alt.expr`, `for(var in seq) expr`, `while(cond) expr`, `repeat expr`, `if(...)` `break` : ce que l'on imagine

Bibliographie

- [1] BERTRAND, F., *Initiation à la statistique avec R*, Dunod (2018).
- [2] BOUZIAD, A. ET CALBRIX, J., *Théorie de la mesure et de l'intégration*, Publications de l'Université de Rouen (1995).
- [3] BERTRAND, F., CLAEYS, E. ET MAUMY-BERTRAND, M., *Modélisation statistique par la pratique avec R*, Dunod (2019).
- [4] BARBE, P. ET LEDOUX, M., *Probabilité*, EDP Sciences (2007).
- [5] CHESNAU, C., *Statistiques : Méthodes et applications avec le logiciel R*, Spartacus IDH (2020).
- [6] DALGAARD, P., *Introductory statistics with R*, Springer (2002).
- [7] DELMAS, J.-F., *Introduction au calcul des probabilités et à la statistique*, Les presses de l'ENSTA (2010).
- [8] FOURDRINIER, D., *Statistique inférentielle*, Dunod (2002).
- [9] GARET, O. ET KURTZMANN, A., *De l'intégration aux probabilités*, Ellipse (2011).
- [10] GOULET, V., *Introduction à la programmation en R*, disponible en ligne : https://cran.r-project.org/doc/contrib/Goulet_introduction_programmation_R.pdf (2016).
- [11] LEJEUNE, M., *Statistique : La théorie et ses applications*, Springer (2010).
- [12] MADSEN, B., *Statistics for Non-Statistician*, Springer (2011).
- [13] OUVRARD, J.-Y., *Probabilité (2 Tomes)*, Cassini (1998).
- [14] SAPORTA, G., *Théories et méthodes de la statistique*, Publications de l'Institut Français du pétrole, Technip (1978).
- [15] SHAO, J., *Mathematical Statistics*, Springer (2003).
- [16] WILCOX, R., *Fundamentals of Modern Statistical Methods*, Springer (2010).