



INSTITUT
POLYTECHNIQUE
DE PARIS

MATHÉMATIQUES

Théorie des Probabilités et Statistiques

Auteur :

Gabriel FRANÇOIS

28 mai 2025

Table des matières

1	Introduction	4
2	Lois de Probabilités à Connaître	5
2.1	Lois Discrètes	5
2.2	Lois Continues	5
3	Glossaire des Notations	7
4	Dénombrement	8
4.1	Principes de Base du Dénombrement	8
4.1.1	Définitions et Propriétés Élémentaires	8
4.1.2	Principe d'Addition	9
4.1.3	Principe de Multiplication	10
4.1.4	Partitions d'un Ensemble	11
4.2	Permutations	12
4.2.1	Permutations sans répétitions	12
4.2.2	Permutations avec Répétition	13
4.3	Arrangements	15
4.3.1	Arrangements sans Répétitions	16
4.3.2	Arrangements avec Répétitions	16
4.4	Combinaisons	17
4.4.1	Combinaisons sans Répétition	18
4.4.2	Combinaisons avec répétition	19
5	Espace probabilisé et mesure de probabilité	21
5.1	Définitions et Exemples	21
5.2	Expériences Aléatoires et Modélisation	24
5.3	Probabilités Conditionnelles	25
6	Variables Aléatoires	27
6.1	Indépendances	27
6.2	Fonctions de Répartitions et Densités	32
6.3	Espérance	37
6.4	Moments, Variance, Covariance	39

7	Notions de Convergence	42
7.1	Tout d'abord, les Égalités	42
7.2	Convergence de Variable et Vecteur Aléatoire	43
7.2.1	Convergence Presque Sure	43
7.2.2	Convergence dans L^p	45
7.2.3	Convergence en Probabilité	47
7.2.4	Convergence en Loi	49
8	Théorèmes de Convergences dans le Cas iid	56
8.1	Convergences en Loi	58
8.2	Convergence Presque Sûre	64
9	Fonction Caractéristique et Fonction Gamma	69
9.1	Rappel sur la Loi Gamma	69
9.2	Fonction Caractéristique	69
10	Vecteurs Aléatoires	71
10.1	Règles de Calcul à Connaître	71
10.2	Vecteur Gaussiens	72
10.3	Théorèmes de Convergences pour les Vecteurs Aléatoires	75
11	Espérances Conditionnelles	76
11.1	Espérances Conditionnelles pour les Variables Aléatoires de Carré Intégrable	76
11.2	Généralisation aux Variable Intégrables	82
12	Lois Conditionnelles	87
12.1	Définitions et Premières Propriétés	87
12.2	Densités Conditionnelles	88
12.3	Cas Discret	91
12.4	Caractérisation de l'Indépendance via les Lois Gaussiennes	92
12.5	Théorème de Fubini Généralisé	93
12.6	Liens avec l'Espérance Conditionnelle	94
12.6.1	Théorème de Transfert Conditionnel	94
12.6.2	Outils pour le Calcul des Lois Conditionnelles	95
13	Inégalités	97
13.1	Inégalités de Concentration	101
13.1.1	Variables Aléatoires bornées	102
13.1.2	Le Cas des Variables Aléatoires Non Bornées	113

14 Estimation Statistique	120
14.1 Principes d'Échantillonnage	120
14.2 Principe d'un Estimateur Statistique	120
14.3 Modèle Statistique	123
14.3.1 Définitions	124
14.4 Estimateurs	125
14.4.1 Comparer les Estimateurs	126
14.4.2 Estimateur de la Méthode des Moments	128
14.4.3 Estimateur du Maximum de Vraisemblance	130
14.4.4 Estimateurs de Bayes	137
15 Test Statistique	140
15.1 Notion de Test	140
15.1.1 Régions de Confiance	143
15.2 Tests Uniformément Plus Puissants	144
15.2.1 Dans le Cas de Deux Hypothèses Simples	145
15.2.2 Test UPP Cas Général	148
15.2.3 Test de Rapport de Vraisemblance	152
15.2.4 Cas des Modèles à Rapports de Vraisemblances Monotones	153
15.3 Test Asymptotiques	154
15.3.1 Tests de Wald	157
15.3.2 Lois Discrètes : Test du χ^2	159
15.4 Méthodes non Asymptotiques et Non Paramétriques	161
15.4.1 Bandes de Confiance Uniformes sur la Fonction de Répartition	161
15.4.2 Test de Permutation	162
15.5 Test d'Hypothèses Multiples	165
15.5.1 FWER et Correction de Bonferroni	166
15.5.2 La FDR et la Correction de Benjamini-Hochberg	167

1 Introduction

L'objectif de ce polycopié est de synthétiser les connaissances essentielles en théorie des probabilités et en statistiques pour un niveau de Master 2. Il est structuré de manière à ce que chaque section puisse servir d'aide-mémoire pour les connaisseurs, ce qui explique que certains éléments des premiers chapitres sur la théorie des probabilités soient utilisés avant d'être formellement introduits. Les lecteurs souhaitant utiliser ce cours comme base initiale devront se référer aux sections introductives des concepts en question. Il s'adresse à toute personne ayant des bases solides en théorie de la mesure et en analyse. Les preuves sont mises en évidence en bleu pour les distinguer des propositions et théorèmes. Cependant, les martingales et la théorie des processus stochastiques ne sont pas abordées ici ; l'accent est mis sur la théorie des probabilités essentielle et les statistiques, en vue de leur application dans l'apprentissage automatique ou la recherche en général.

Cet enseignement s'adresse principalement aux étudiants de niveau Master 2 en mathématiques, statistiques, ou dans des domaines connexes nécessitant une solide compréhension des probabilités et des statistiques. Il est aussi conçu pour les chercheurs et professionnels souhaitant disposer d'un référentiel pratique pour se remémorer rapidement les concepts clés, ainsi que les preuves associées. Les divers chapitres couvrent un large éventail de sujets, allant des lois de probabilités fondamentales aux notions avancées de convergence et aux vecteurs aléatoires, en passant par les espérances conditionnelles et les tests statistiques. Mon objectif est de fournir une ressource complète qui rassemble tout ce qu'il est essentiel de savoir en théorie des probabilités et en statistiques, en incluant un maximum de preuves détaillées. Ces preuves, mises en évidence, visent à approfondir la compréhension des concepts et à servir d'aide-mémoire et de gain de temps pour les lecteurs, qu'ils soient novices ou experts dans le domaine. Ainsi, ce polycopié se veut un outil pratique et accessible, permettant à chacun de se familiariser avec ces sujets complexes et de les appliquer dans leurs travaux de recherche ou leurs projets professionnels.

Ce polycopié est structuré pour offrir une présentation rigoureuse et concise des concepts essentiels en théorie des probabilités et en statistiques. Chaque chapitre est conçu pour être à la fois autonome et complémentaire, permettant une consultation efficace et ciblée. L'objectif est de fournir un outil exhaustif où chaque notion est expliquée avec précision et accompagnée de preuves détaillées. Cette approche vise à aller droit au but, en privilégiant la rigueur mathématique pour servir d'aide-mémoire fiable aux étudiants et professionnels. Les chapitres sont organisés de manière logique, facilitant ainsi une compréhension progressive et approfondie des sujets abordés.

2 Lois de Probabilités à Connaître

2.1 Lois Discrètes

$$\boxed{X \sim \mathcal{P}(\lambda), \lambda > 0} \quad \mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}; \mathbb{E}(X) = \lambda; \mathbb{V}(X) = \lambda; \varphi_X(t) = e^{\lambda(e^{it} - 1)}$$

$$\boxed{X \sim \mathcal{U}(\{1; \dots; n\})} \quad \mathbb{P}(X = k) = \frac{1}{n}; \mathbb{E}(X) = \frac{n+1}{2}; \mathbb{V}(X) = \frac{n^2-1}{12}; \varphi_X(t) = \frac{e^{it} e^{itn} - 1}{n e^{it} - 1}$$

$$\boxed{X \sim \mathcal{B}(n, p)} \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; \mathbb{E}(X) = np; \mathbb{V}(X) = np(1-p); \varphi_X(t) = (1 + p(e^{it} - 1))^n$$

$$\boxed{X \sim \mathcal{G}_{\mathbb{N}^*}(p)} \quad \mathbb{P}(X = k) = (1-p)^{k-1} p; \mathbb{E}(X) = \frac{1}{p}; \mathbb{V}(X) = \frac{1-p}{p^2}; \varphi_X(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$$

2.2 Lois Continues

$$\boxed{X \sim \mathcal{E}(\lambda), \lambda > 0} \quad f_X(x) = \lambda e^{-\lambda x}; \mathbb{E}(X) = \frac{1}{\lambda}; \mathbb{V}(X) = \frac{1}{\lambda^2}; \varphi_X(t) = \frac{\lambda}{\lambda - it}; \text{ Soit } X_1, \dots, X_n \sim \mathcal{E}(\lambda)$$

$$\text{Alors } \sum_{i=1}^n X_i \sim \Gamma(n, \frac{1}{\lambda}) \text{ et } \sum_{i=1}^n X_i \sim \Gamma^{-1}(n, \lambda); \text{ Soit } a \in \mathbb{R}, aX \sim \mathcal{E}(\frac{\lambda}{a}); \forall k \in \mathbb{N}, \mathbb{E}[X^k] = \frac{k!}{\lambda^k}$$

$$\boxed{X \sim \Gamma(n, \lambda), n, \lambda > 0} \quad f_X(x) = \frac{1}{\lambda^n \Gamma(n)} x^{n-1} e^{-\frac{x}{\lambda}}; \mathbb{E}(X) = n\lambda; \mathbb{V}(X) = n\lambda^2; \text{ si } X \sim \Gamma(a, \lambda) \text{ et}$$

$$Y \sim \Gamma(b, \lambda), X \perp\!\!\!\perp Y \text{ alors } X + Y \sim \Gamma(a + b, \lambda); \varphi_X(t) = (\frac{1}{1 - \lambda it})^n; \text{ si } X \sim \Gamma(n, \lambda) \text{ alors } \forall c \in \mathbb{R}$$

$$cX \sim \Gamma(n, \lambda c); \forall k \in \mathbb{N}, \mathbb{E}[X^k] = \lambda^k \prod_{i=1}^k (n + i - 1)$$

$$\boxed{X \sim \Gamma^{-1}(n, \lambda), n, \lambda > 0} \quad f_X(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} \mathbb{1}_{[0, +\infty)}(x); \mathbb{E}[X] = \frac{n}{\lambda}; \mathbb{V}(X) = \frac{n}{\lambda^2}; c \in \mathbb{R},$$

$$cX \sim \gamma^{-1}(n, \frac{\lambda}{c}); X \sim \Gamma^{-1}(a, \lambda) \text{ et } Y \sim \Gamma^{-1}(b, \lambda), X \perp\!\!\!\perp Y \text{ alors } X + Y \sim \Gamma^{-1}(a + b, \lambda)$$

$$\varphi_X(t) = (\frac{\lambda}{\lambda - it})^n, \Gamma^{-1}(n, \lambda) = \Gamma(n, \frac{1}{\lambda})$$

$$\boxed{X \sim \mathcal{U}([a, b])} \quad f_X(x) = \frac{1}{b-a} \mathbb{1}_{(b-a)}(x); F(x) = \frac{x-a}{b-a} \mathbb{1}_{[a, b]}(x) + \mathbb{1}_{(a, +\infty)}(x); \mathbb{E}[X] = \frac{a+b}{2}; \mathbb{V}(X) = \frac{(b-a)^2}{12}$$

$$\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

$$\boxed{X \sim \mathcal{N}(\mu, \sigma^2)}; f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}}; X_1, \dots, X_n \sim \mathcal{N}(0, 1) \implies \sum_{i=1}^n X_i^2 \sim \chi^2(n)$$

$$\varphi_X(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}; \text{ Si } (X_i)_{1 \leq i \leq n} \stackrel{\perp}{\sim} \mathcal{N}(\mu_i, \sigma_i^2), t_1 \dots t_n \in \mathbb{R}, \sum_{i=1}^n t_i X_i \sim \mathcal{N}(\sum_{i=1}^n t_i \mu_i, \sum_{i=1}^n t_i^2 \sigma_i^2)$$

$$\boxed{X \sim \mathcal{N}_d(\mu, \Sigma), \mu \in \mathbb{R}^d, \Sigma \in \mathcal{M}_{d \times d}(\mathbb{R})}; f_X(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}; \mathbb{E}[X] = \mu$$

$$\mathbb{V}(X) = \Sigma; \varphi_X(t) = e^{it^T \mu - \frac{1}{2} t^T \Sigma t} \text{ (Voir chapitre vecteur gaussiens pour plus de détails)}$$

$$\boxed{X \sim \chi^2(n)}; X \sim \chi^2(n) \iff X \sim \Gamma(\frac{n}{2}, 2) = \Gamma^{-1}(\frac{n}{2}, \frac{1}{2}); f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} \mathbb{1}_{[0;+\infty)}(x)$$

$$\mathbb{E}[\overline{X}_n] = n; \mathbb{V}(\overline{X}_n) = 2n$$

$$\boxed{X \sim \text{Cauchy}(\lambda, a), \lambda \in \mathbb{R}, a > 0} f_X(x) = \frac{a}{\pi((x-\lambda)^2 + a^2)}; \text{ d'où, si } \lambda = 0, x = 1 \text{ on a } f_X(x) = \frac{1}{\pi(x^2+1)}$$

$$\mathbb{E}[X] = +\infty; \varphi_X(t) = e^{i\lambda t - a|t|}; F_X(t) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\lambda}{a}\right)$$

$$\boxed{X \sim \text{Beta}(\alpha, \beta)} \alpha, \beta > 0 f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbb{1}_{(0,1)}(x) \text{ avec } B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$$

$$\mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \text{ Si } Y \sim \mathcal{U}([0, 1]), \forall r > 0, Y^r \sim \text{Beta}(\frac{1}{r}, 1)$$

$$\text{Si } Z \sim \text{Beta}(\alpha, 1) \text{ Alors } \log(Z) \sim \mathcal{E}(\alpha)$$

3 Glossaire des Notations

- Les scalaires ou variables aléatoires notés en **gras** correspondent à des **vecteurs** $\mathbf{x}, \mathbf{X}, \mathbf{u}$ etc.
- Les matrices sont notées en majuscule.

- Soit $f : \mathbb{R}^p \mapsto \mathbb{R}^d$ une fonction dérivable en un point $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p$. On note alors

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial f_d}{\partial x_1}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_p}(\mathbf{x}) & \dots & \frac{\partial f_d}{\partial x_p}(\mathbf{x}) \end{pmatrix} \in \mathcal{M}_{p \times d}(\mathbb{R})$$

Avec

$$\forall i \in \llbracket 1, p \rrbracket, \quad f_i(\mathbf{x}) : \mathbb{R}^p \mapsto \mathbb{R} \mid \quad f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_d(\mathbf{x}) \end{pmatrix}$$

On notera aussi $J_f(\mathbf{x}) = (\nabla f(\mathbf{x}))^T \in \mathcal{M}_{d \times p}(\mathbb{R})$

- Soit $f : \mathbb{R}^p \mapsto \mathbb{R}$ une fonction dérivable en un point $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p$. On note alors

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_p}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_p \partial x_1}(\mathbf{x}) & \dots & \frac{\partial^2 f}{\partial x_p^2}(\mathbf{x}) \end{pmatrix} \in \mathcal{M}_{p \times p}(\mathbb{R})$$

- " $X \sim Q$ " signifie que la loi de X notée P_X est Q .
- λ définira ici la mesure de Lebesgue
- \log représente sans autres précisions le logarithme naturel *i.e.* : $\log(e^x) = x, \forall x \in \mathbb{R}$
- $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$
- Pour un modèle statistique $\{\mathbb{P}_\theta, \theta \in \Theta\}$ donné, on a $l(x, \theta) = \log(f(x, \theta))$ avec $f(x, \theta) = \frac{d\mathbb{P}_\theta}{d\mu}(x)$ et on écrira $l'(x, \theta) = \frac{\partial l(x, \theta)}{\partial \theta}$.
- On notera le cardinal d'un ensemble fini E : $\#E$ ou bien $|E|$.
- L'ensemble E^N représente l'ensemble de tous les N -uplets (x_1, x_2, \dots, x_N) où chaque $x_i \in E$ pour $i = 1, 2, \dots, N$, où N peut être un ensemble indénombrable.

4 Dénombrement

Avant de définir les variables aléatoires et les espaces probabilisés, nous allons rappeler les bases du dénombrement. Le dénombrement est une branche des mathématiques qui se concentre sur le comptage des objets. Il s'agit de déterminer le nombre d'éléments dans un ensemble fini ou de calculer le nombre de configurations possibles dans un certain cadre. C'est un domaine très utile pour répondre à des questions telles que "combien de façons différentes peut-on arranger n objets distincts?" ou "combien de sous-ensembles peut-on former à partir d'un ensemble de n éléments?". Son utilisation se retrouve partout, et notamment en probabilités et en statistiques. C'est pourquoi nous aborderons ces notions, qui sont pour certains élémentaires mais fondamentales. On s'intéressera ici au Dénombrement dans des ensembles finis.

Cette partie peut être allègrement sautée par ceux ayant déjà les bases dans ce domaine, cependant, un peu de rafraîchissement ne peut jamais faire de mal.

4.1 Principes de Base du Dénombrement

4.1.1 Définitions et Propriétés Élémentaires

Définition 4|4.1|1

Soit E un ensemble. On définit l'ensemble $N_n = \{x \in \mathbb{N} \mid x < n\}$. On dit que E est fini de **cardinal** $n \in \mathbb{N}$ i.e. : $|E| = n$ (ou bien $\#E = n$) si E est en bijection avec N_n . C'est-à-dire qu'il existe une application bijective f de E dans N_n .

Théorème 4|4.1|1

Soit A une partie d'un ensemble fini E . Alors A est elle-même finie et $|A| \leq |E|$.

Si en outre, $|A| = |E|$ alors $A = E$ (La preuve est immédiate).

Proposition 4|4.1|1

Soit E un ensemble fini, et F un ensemble et $f : E \mapsto F$ une application. On a :

$$(i) \quad |f(E)| \leq |E|$$

$$(ii) \quad f \text{ est injective} \iff |f(E)| = |E|$$

Preuve :

① On précise que comme f est une application, elle est définie sur tout l'espace E . On a que $f : E \mapsto f(E)$ avec $f(E) = \{f(x) : x \in E\}$ est une surjection. Ainsi,

$$|f(E)| \leq |E|$$

② Si f est injective, alors tout les éléments de $f(E)$ ont un antécédent unique. Ainsi, l'application $f : E \mapsto f(E)$ est bijective, et donc $|f(E)| = |E|$.

Réciproquement, si $|f(E)| = |E|$, alors $f : E \mapsto f(E)$ est injective (par construction) et par conséquent bijective, puisque $|f(E)| = |E|$. Donc, sachant que $E \subset F$, on a $f : E \mapsto F$ est injective.

4.1.2 Principe d'Addition

Le principe d'addition stipule que si un événement peut se produire de n manières différentes et un autre événement, qui ne peut pas se produire en même temps que le premier, peut se produire de m manières différentes, alors le nombre total de façons dont l'un ou l'autre événement peut se produire est donné par $n + m$. En d'autres termes, si deux événements sont mutuellement exclusifs, le nombre total de résultats possibles est la somme des résultats possibles pour chaque événement.

Proposition 4|4.1|2

Soit E et F deux ensembles finis disjoints avec $|E| = k \in \mathbb{N}$ et $|F| = n \in \mathbb{N}$. Alors on a

$$|E \cup F| = |E| + |F| = k + n$$

Preuve :

Soit f une bijection de E dans $\llbracket 1, k \rrbracket$ ($\llbracket 1, k \rrbracket = \{1, 2, \dots, k\}$) et g une bijection de F dans $\llbracket 1+k, n+k \rrbracket$. On peut alors construire h une application de $E \cup F$ dans $\llbracket 1, n+k \rrbracket$ telle que sa restriction à E est f et sa restriction F est g . Comme h est une bijection, c'est une injection, et donc $|h(E \cup F)| = |\llbracket 1, n+k \rrbracket| = n+k$ par la proposition précédente.

corollaire

Par récurrence on a bien que pour $(E_i)_{1 \leq i \leq n}$ une suite d'ensemble finis disjoints. Alors

$$|\bigcup_{i=1}^n E_i| = \sum_{i=1}^n |E_i|$$

Proposition 4|4.1|3

Soit $(E_n)_{n \in \mathbb{N}}$ une famille dénombrable d'ensemble disjoints. Alors

$$|\bigcup_{i=1}^{+\infty} E_i| = \sum_{i=1}^{+\infty} |E_i|$$

Preuve :

Chaque E_i est fini, donc on peut dire que pour tout $i \in \mathbb{N}$, il existe $n_i \in \mathbb{N}$ tel que $|E_i| = |\{0, \dots, n_i\}| = n_i$. Pour chaque $i \in \mathbb{N}$, $E_i = \{e_{i,1}, \dots, e_{i,n_i}\}$. Comme les E_i sont disjoints, on peut étiqueter chaque élément dans $\bigcup_{i=1}^{+\infty} E_i$ par le couple (i, j) avec $j \in \{1, \dots, n_i\}$. Pour $A = \{(i, j) \in \mathbb{N} \times \mathbb{N} \mid 1 \leq j \leq n_i\}$, on définit l'application par :

$$f : \bigcup_{i=1}^{+\infty} E_i \mapsto A$$

avec pour tout $e_{i,j} \in \bigcup_{i=1}^{+\infty} E_i$, $f(e_{i,j}) = (i, j)$. Cette fonction est clairement injective car les $(E_i)_{i \in \mathbb{N}}$ sont disjoints et surjective car tout élément de A correspond à un élément de (E_i) .

On termine puisque, par définition, $|A| = \sum_{i=1}^{+\infty} n_i = \sum_{i=1}^{+\infty} |E_i|$ donc il existe une bijection entre $\bigcup_{i=1}^{+\infty} E_i$ et A et $|A| = \sum_{i=1}^{+\infty} |E_i|$. Donc finalement on a

$$\left| \bigcup_{i=1}^{+\infty} E_i \right| = \sum_{i=1}^{+\infty} |E_i|$$

4.1.3 Principe de Multiplication

En combinatoire, le principe de multiplication stipule que si une tâche peut être accomplie de n manières différentes et qu'une autre tâche peut être accomplie de m manières différentes, alors les deux tâches peuvent être accomplies ensemble de $n \times m$ manières différentes.

Proposition 4|4.1|4

Soit E et F deux ensembles finis de cardinaux respectifs n et k . Alors $E \times F$ est fini de cardinal

$$|E \times F| = nk$$

Preuve :

On a $E \times F = \{(e, f) \mid e \in E, f \in F\}$ avec $E = \{e_1, \dots, e_n\}$ et $F = \{f_1, \dots, f_k\}$ or donc

$$E \times F = \{(e_i, f_j) \mid 1 \leq i \leq n, 1 \leq j \leq k\}$$

Chaque paires (i, j) donne une paire unique (e_i, f_j) , et comme il y a n choix pour i et k pour j , alors

$$|E \times F| = |\{(e_i, f_j) \mid 1 \leq i \leq n, 1 \leq j \leq k\}| = nk$$

corollaire

Soit $n \in \mathbb{N}$ et $(E_i)_{1 \leq i \leq n}$ une famille d'ensembles finis. Alors :

$$|E \times \dots \times E_n| = \prod_{i=1}^n |E_i|$$

Remarque :

Contrairement au principe d'addition, ce dernier ne se généralise pas pour un produit cartésien infini.

4.1.4 Partitions d'un Ensemble

Définition 4|4.1|2

Soit S un ensemble. Une **partition** d'un ensemble S est une collection (pas nécessairement finie ni même dénombrable) de sous-ensembles non vides de S tels que :

1) **Recouvrement** : L'union de tous les sous-ensembles de la partition est égale à S .

2) **Disjoints** : Les sous-ensembles sont deux à deux disjoints, c'est-à-dire que l'intersection de deux sous-ensembles distincts est vide.

Remarque :

Il est à noter que lorsque l'on introduira une partition, elle sera, sans informations supplémentaire au plus dénombrable, si elle ne l'est pas, ça sera précisé.

Proposition 4|4.1|5

Soit E un ensemble fini tel que $|E| = k$. Alors si on définit $\mathcal{P}(E)$ comme l'ensemble des parties de E , on a

$$|\mathcal{P}(E)| = 2^k$$

Preuve :

Soit une application $f : E \mapsto \{0, 1\}$. On définit l'ensemble $A_f = \{x \in E \mid f(x) = 1\}$ et on obtient une bijection entre l'ensemble $\mathcal{F} = \{f \mid f : E \mapsto \{0, 1\}\}$ et $\mathcal{P}(E)$. L'application en question associe à chaque $f \in \mathcal{F}$ l'ensemble $A_f \in \mathcal{P}(E)$, nous allons démontrer qu'elle est bijective.

En effet, on a l'injection car si $f \neq g$ alors $A_f \neq A_g$. De plus, pour tout $A \in \mathcal{P}(E)$ on définit l'application

$$f_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

Donc pour tout $x \in \mathcal{P}(E)$, on peut lui associer une fonction $f \in \mathcal{F}$ par l'application définie plus haut. Ainsi on a la bijection et donc $|\mathcal{P}(E)| = |\mathcal{F}|$. Maintenant, regardons \mathcal{F} : chaque application $f : E \mapsto \{0, 1\}$ associe à un $x \in E$ 0 ou 1. Il y a donc 2 choix pour chaque x . Comme $|E| = k$ on a donc 2^k f possible. Ainsi :

$$|\mathcal{F}| = |\mathcal{P}(E)| = 2^k$$

4.2 Permutations

Une situation fréquente en combinatoire consiste à ordonner tous les éléments d'un ensemble donné. Une **permutation** désigne une réorganisation (ou réarrangement) complète des éléments d'un ensemble fini, c'est-à-dire une bijection de cet ensemble sur lui-même.

Définition 4|4.2|1

Soit E un ensemble. On appelle **permutations de E** l'ensemble des bijections de E dans lui-même.

4.2.1 Permutations sans répétitions

Proposition 4|4.2|1

Soit E un ensemble fini à n élément. Alors

$$|E| = n!$$

Preuve : Soit $\sigma : E \mapsto E$ une bijection (donc une permutation). Notons $E = \{x_1, \dots, x_n\}$ alors pour construire une permutation f de E dans E :

- pour x_1 on a n choix d'image pour σ .
- pour x_2 , on a plus que $n - 1$ choix (pour éviter l'image de x_1).

⋮

- Pour le dernier choix, il ne reste plus qu'une possibilité.

Ainsi, on a bien $n(n-1)\dots 1 = n!$ nombre de permutations possibles.

Exemple :

Si on considère un ensemble $E = \{1, 2, 3\}$, alors les différentes manières de réorganiser tous ses éléments correspondent aux permutations de E . Ces permutations sont : $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$, et $(3, 2, 1)$ soit 6 au total.

4.2.2 Permutations avec Répétition

Dans certaines situations, on souhaite permuter tous les éléments d'un ensemble, mais certains éléments sont identiques. Cela donne lieu à ce qu'on appelle des **permutations avec répétition**.

Contrairement aux permutations classiques (où tous les éléments sont distincts), ici, certaines configurations deviennent indiscernables du fait de la répétition d'éléments.

Définition 4|4.2|2

Soit $E = \{x_1, \dots, x_n\}$ un ensemble fini de taille n . Soit aussi k_1, \dots, k_n des entiers naturels tel que $k_1 + \dots + k_n = k \in \mathbb{N}$. Une **permutation** de k éléments de E avec k_1, \dots, k_n **répétitions** est un n -uplet (une collection ordonnée et finie) d'éléments de E dans lequel chacun des éléments x_1, \dots, x_n apparaissent respectivement k_1, \dots, k_n fois.

Exemple :

Soit $E = \{x_1, \dots, x_n\}$ un ensemble fini de taille n . Et soit k_1, \dots, k_n des entiers naturels tel que $k_1 + \dots + k_n = k \in \mathbb{N}$. Le n -uplet

$$\underbrace{(x_1, \dots, x_1)}_{k_1 \text{--fois}}, \dots, \underbrace{(x_n, \dots, x_n)}_{k_n \text{--fois}}$$

est une permutation avec répétition particulière.

Proposition 4|4.2|2

Le nombre de permutations de n éléments avec k_1, \dots, k_n répétitions est égal à

$$\frac{k!}{k_1! k_2! \dots k_n!}$$

Ce nombre est connu sous le nom de **coefficient multinomial** et se note habituellement

$$\binom{k}{k_1, k_2, \dots, k_n} = \frac{k!}{k_1! k_2! \dots k_n!}$$

Preuve :

Considérons qu'il y a k positions distinctes à remplir dans le k -uplet. Chaque position contiendra un élément de E , et l'élément x_i doit apparaître exactement k_i fois.

Pour le premier élément x_1 , on doit choisir k_1 positions parmi les k restantes. Il s'agit donc d'une combinaison car l'ordre ne compte pas, de k_1 parmi k (voir combinaison en 4.4). Ainsi, on a :

$$C_k^{k_1} = \binom{k}{k_1} = \frac{k!}{k_1!(k - k_1)!}$$

Après cette étape, pour placer x_2 , on doit choisir k_2 parmi $k - k_1$ positions donc $C_{k-k_1}^{k_2}$. On généralise donc à $k_i, i \in \llbracket 1, n \rrbracket$, et on obtient que pour placer x_i on a donc

$$C_{k - \sum_{j=1}^{i-1} k_j}^{k_i} = \binom{k - \sum_{j=1}^{i-1} k_j}{k_i} = \frac{(k - \sum_{j=1}^{i-1} k_j)!}{k_i!(k - \sum_{j=1}^i k_j)!} \quad \text{possibilités}$$

Pour le dernier éléments, toutes les places sont prises donc il n'y a qu'une seule façon de les arranger.

Ainsi, on a un nombre total de possibilités est le produit des choix pour chaque étape, car les étapes sont séquentielles et indépendantes (le choix des positions pour un élément n'affecte pas les choix ultérieurs, si l'ordre des éléments est fixé). On a donc un nombre de possibilités égales à :

$$\frac{k!}{k_1!(k - k_1)!} \times \frac{(k - k_1)!}{k_2!(k - k_1 - k_2)!} \times \dots \times \frac{(k - k_1 - \dots - k_{n-1})!}{k_n!(0)!} = \frac{k!}{k_1! k_2! \dots k_n!}$$

Exemple :

Considérons le mot **ABA**. Il contient trois lettres, dont certaines sont identiques. On veut déterminer combien de mots distincts on peut former en permutant les lettres de ce mot.

- Il y a au total $n = 3$ lettres.
- La lettre **A** apparaît 2 fois.
- La lettre **B** apparaît 1 fois.

Si toutes les lettres étaient distinctes, le nombre de permutations serait $3! = 6$. Mais ici, échanger les deux lettres **A** ne produit pas un nouveau mot distinct. Il faut donc corriger ce comptage en divisant par le nombre de permutations internes possibles des lettres répétées.

Le nombre de permutations distinctes est donné par le **coefficient multinomial** :

$$\frac{3!}{2! \cdot 1!} = \frac{6}{2} = 3$$

On peut les lister pour vérifier :

1. AAB
2. ABA
3. BAA

On retrouve bien 3 arrangements distincts.

Formule générale. Si un mot de n lettres contient :

n_1 lettres de type a_1 , n_2 lettres de type a_2 , \dots , n_k lettres de type a_k

avec $n_1 + n_2 + \dots + n_k = n$, alors le nombre de permutations distinctes est donné par :

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

4.3 Arrangements

Dans de nombreux problèmes de dénombrement, on s'intéresse à la manière de sélectionner et d'ordonner un certain nombre d'éléments à partir d'un ensemble donné. Lorsque **l'ordre des éléments choisis a de l'importance**, on parle alors d'**arrangements**.

Autrement dit, un arrangement de k éléments parmi n correspond à un choix ordonné de k éléments distincts (ou non, selon le contexte) pris dans un ensemble de n éléments.

Par exemple, dans une course de 10 coureurs, combien de podiums différents (1^{er}, 2^e, 3^e) peut-on former ? Ici, l'ordre dans lequel les coureurs montent sur le podium est crucial, et le problème consiste donc à compter des arrangements. Les arrangements permettent de modéliser des situations où la position, la hiérarchie ou l'ordre chronologique jouent un rôle déterminant.

Dans ce qui suit, nous étudierons deux types d'arrangements :

- les **arrangements sans répétition**, où chaque élément de l'ensemble ne peut apparaître qu'une seule fois dans l'arrangement ;
- les **arrangements avec répétition**, où un même élément peut être utilisé plusieurs fois.

4.3.1 Arrangements sans Répétitions

Proposition 4|4.3|1

Soit $E = \{x_1, \dots, x_n\}$ un ensemble fini à n éléments et $k \in \mathbb{N}$ tel que $k \leq n$. Le nombre de manières de choisir k objets parmi les n objets de E , chaque objet ne pouvant être choisi qu'une fois et l'ordre du choix étant pris en compte, est donné par

$$A_n^k = \frac{n!}{(n-k)!}$$

Preuve :

Analysons les choix possibles à chaque étape. La construction d'un tel k -uplet se fait en choisissant successivement les éléments pour chaque position, de la première à la k -ième position, sans remise. Analysons les choix possibles à chaque étape :

Pour le premier élément x_1 , on a n choix possible pour le placer.

Pour le second élément x_2 on a $n - 1$ possibilités

On déduit donc que pour le i -ième élément, $i \in \llbracket 1, k \rrbracket$, on a $n - k + 1$ possibilités.

Ainsi, on en déduit qu'on a donc

$$n \times (n-1) \times \dots \times (n-k+1)! = \frac{n!}{(n-k)!} = A_n^k \quad \text{Possibilités}$$

Ainsi, le nombre d'arrangement possibles sans répétitions de k éléments parmi n est de $\frac{n!}{(n-k)!}$

Remarque :

① L'arrangement de n éléments parmi n , noté A_n^n correspond donc à l'ensemble des permutations sans répétitions d'un ensemble E à n éléments.

② On en déduit que le nombre d'injections d'un ensemble à k éléments dans un ensemble à n éléments est égale à A_n^k si $k \leq n$ et 0 sinon.

4.3.2 Arrangements avec Répétitions

Proposition 4|4.3|2

Soit E un ensemble fini de taille n et $k \leq n$. Le nombre de façons de choisir k élément de E parmi ses n éléments, chaque élément pouvant être choisi plusieurs fois et l'ordre du choix étant pris en compte est noté B_n^k et est égal à n^k .

Preuve :

Posons $E = \{x_1, \dots, x_n\}$. La preuve est en réalité élémentaire.

Un choix ordonné de k éléments de E avec répétitions autorisées correspond à une séquence ordonnée de longueur k (un k -uplet) où chaque composante est un élément de E . Ainsi, Pour construire une telle séquence, nous avons k positions à remplir (position 1, position 2, ..., position k). À chaque position, nous sélectionnons un élément de E . Comme les répétitions sont autorisées, le choix à chaque position est indépendant des choix précédents et il n'y a aucune restriction sur le nombre de fois qu'un élément peut apparaître. Ainsi, à chaque fois on a n possibilités pour chaque positions $i \in \llbracket 1, k \rrbracket$ ce qui donne un total de

$$\overbrace{n \times \dots \times n}^{n \text{ fois}} = n^k \quad \text{possibilités}$$

Exemple :

Considérons l'ensemble $E = \{a, b\}$ de $n = 2$ éléments, et supposons que nous voulons former des mots de longueur $k = 3$ en utilisant les éléments de E , avec **répétition autorisée**.

Chaque position dans le mot peut être remplie par n'importe quel élément de E . Il y a donc :

$$\text{nombre d'arrangements avec répétition} = n^k = 2^3 = 8$$

On peut vérifier cela en listant toutes les possibilités :

- | | |
|--------|--------|
| 1. aaa | 5. bab |
| 2. aab | 6. bba |
| 3. aba | 7. bbb |
| 4. abb | 8. baa |

On obtient bien 8 arrangements, ce qui confirme le résultat donné par la formule générale :

$$n^k \quad \text{où } n = \text{nombre de symboles possibles, et } k = \text{nombre de positions.}$$

4.4 Combinaisons

Dans un grand nombre de problèmes de dénombrement, il est essentiel de savoir compter combien de sous-ensembles de taille donnée peuvent être formés à partir d'un ensemble initial.

Lorsque l'ordre des éléments choisis "n'a pas d'importance", on parle alors de **combinaisons**. Autrement dit, choisir une combinaison de k éléments parmi n revient à sélectionner un sous-ensemble de

k éléments dans un ensemble de n éléments, sans se soucier de l'ordre dans lequel ces éléments sont choisis.

Les combinaisons jouent un rôle fondamental en probabilité, en statistiques, et dans l'analyse de situations où seule la composition d'un groupe importe, et non la façon dont ses éléments sont arrangés. Dans ce qui suit, nous distinguerons deux cas importants :

- les **combinaisons sans répétition**, où chaque élément de l'ensemble de départ ne peut être choisi qu'une seule fois ;
- les **combinaisons avec répétition**, où chaque élément peut être choisi plusieurs fois.

4.4.1 Combinaisons sans Répétition

Les combinaisons sans répétitions concernent le nombre de façons de choisir k éléments d'un ensemble de n éléments, où les répétitions ne sont pas autorisées et l'ordre n'est pas important. Cela signifie que chaque élément ne peut être choisi qu'une seule fois. Par exemple, si vous avez un ensemble de fruits {Pomme, Banane, Cerise} et vous voulez choisir 2 fruits, les combinaisons possibles sont {Pomme, Banane}, {Pomme, Cerise}, et {Cerise, Banane}. Les combinaisons sans répétitions sont couramment utilisées dans les problèmes de sélection où chaque élément ne peut être choisi qu'une seule fois, comme dans les tirages sans remise ou les problèmes de sous-ensembles.

Proposition 4|4.4|1

Soit E un ensemble fini à n éléments et $k \in \mathbb{N}$ tel que $k \leq n$. Le nombre de combinaisons (sans répétition) de k éléments de E noté C_n^k correspond au nombre de possibilité de sélectionner de manière non ordonnée k éléments de E sans répétitions. Il est égal à

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Preuve :

Une combinaison d'un ensemble E est un sous-ensemble de k éléments distincts de E , où l'ordre n'a pas d'importance. Ainsi, chaque combinaison de k éléments correspond à $k!$ arrangements différents.

En effet, puisque que l'ensemble des permutations possible dans un ensemble à k élément est égal à $k!$, il y a $k!$ possibilités d'ordonner les k éléments. Donc un arrangement de k éléments parmi n équivaut à $k!$ combinaisons de k parmi n . Ainsi, on a

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

Exemple :

① Si vous avez un ensemble de 3 éléments distincts $\{A, B, C\}$ et vous voulez choisir 2 éléments sans répétition, le nombre de combinaisons possibles est donné par :

$$C_3^2 = \binom{3}{2} = \frac{3!}{2!(3-2)!} = 3$$

Les combinaisons possibles sont en effet au nombre de 3 : $\{A, B\}, \{A, C\}, \{B, C\}$.

② Calcul du nombre de paires (i, j) tels que $1 \leq i < j \leq n, n \in \mathbb{N}$.

Il s'agit d'un problème d'ordre combinatoire. En effet, le 1 calcul du nombre de paires (i, j) tels que $1 \leq i < j \leq n, n \in \mathbb{N}$ revient à calculer la somme

$$\sum_{1 \leq i < j}^n 1$$

Ce problème revient à compter le nombre de façons de choisir 2 indices distincts parmi n , où l'ordre n'a pas d'importance. L'ordre n'a pas d'importance puisqu'on a pour tout $(i, j), i < j$ l'ensemble (j, i) qui est identique à (i, j) donc on ne compte qu'une seule fois les paires $(i, j), i < j$. Ainsi, chaque paire $(i, j), i < j$ correspond à un sous-ensemble non ordonné de 2 éléments choisis parmi n indices. on a donc

$$\binom{n}{2} = \frac{n!}{2(n-2)!}$$

4.4.2 Combinaisons avec répétition

Les combinaisons avec répétitions permettent de déterminer le nombre de façons de choisir k éléments d'un ensemble de n éléments, où les répétitions sont autorisées et l'ordre n'est pas important. Cela signifie que vous pouvez choisir le même élément plus d'une fois. Par exemple, si vous avez un ensemble de fruits $\{\text{Pomme}, \text{Banane}, \text{Cerise}\}$, et vous voulez choisir 2 fruits, vous pouvez avoir des combinaisons comme $\{\text{Pomme}, \text{Banane}\}, \{\text{Pomme}, \text{Pomme}\}, \{\text{Banane}, \text{Banane}\}$ etc. Les combinaisons avec répétitions sont utiles dans des situations où les éléments peuvent être sélectionnés plus d'une fois, comme dans les problèmes de distribution ou de sélection avec remise.

Proposition 4|4.4|2

Soit E un ensemble fini de taille n et $k \leq n$. Le nombre de possibilités différentes de sélectionner de

manière ordonnée k éléments de E sans répétitions est égal à

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Preuve :

On pose $E = \{x_1, \dots, x_n\}$, Nous voulons choisir k éléments de E avec répétitions autorisées, sans tenir compte de l'ordre. Cela revient à déterminer le nombre de d'ensemble différents de taille k formés à partir des éléments de E , ou de manière équivalente, le nombre de solutions en entiers naturels de l'équation :

$$\sum_{i=1}^n X_i = k, \quad X_i \geq 0 \text{ est le nombre de fois où } x_i \text{ est choisi}$$

Considérons $k*$ représentant les k éléments choisis. Pour séparer ces étoiles en n groupes (un groupe par élément de E), plaçons $n-1$ barres "|" entre les étoiles. Chaque groupe correspond au nombre d'occurrences d'un élément de E .

Par exemple, pour $n=3$ et $k=2$, la solution $X_1=1, X_2=1, X_3=0$ est représentée $*|*|$.

Chaque séquence de $k*$ et de $n-1$ "|" correspond à une unique solution de l'équation $\sum_{i=1}^n X_i = k$. Dans une telle séquence, il y a $n-1+k$ symbols. Le nombre de manières différentes de combiner ces symbols est le nombre de façons de choisir les positions des $k*$ parmi $n-1+k$. Il s'agit d'une combinaison car les $k*$ et les "|" sont indistinguables, et donc il y a :

$$\frac{(n+k-1)!}{k!(n-1)!} \quad \text{possibilités}$$

Exemple :

Si vous avez un ensemble de 3 éléments distincts $\{A, B, C\}$ et vous voulez choisir 2 éléments avec répétition autorisée, le nombre de combinaisons possibles est :

$$\binom{3+2-1}{2} = \binom{4}{2} = \frac{(4)!}{2!(4-2)!} = 6$$

En effet, Les combinaisons possibles sont : $\{A, A\}, \{A, B\}, \{A, C\}, \{B, B\}, \{B, C\}, \{C, C\}$

5 Espace probabilisé et mesure de probabilité

5.1 Définitions et Exemples

L'étude des espaces probabilisés se fonde sur la théorie de la mesure. On considèrera donc ce dernier champs comme acquis. Les notations seront cependant légèrement différentes, pour respecter les conventions établies en théorie des probabilités.

Définition 5|5.1|1

Soit (Ω, \mathcal{A}) un espace mesurable. Une **probabilité** sur (Ω, \mathcal{A}) est une fonction $P : \mathcal{A} \mapsto \mathbb{R}_+$ telle que :

(i) $P(\emptyset) = 0$

(ii) $P(\Omega) = 1$

(iii) Pour toute suite $(A_n)_{n \geq 1}$ d'éléments 2 à 2 disjoints de (\mathcal{A}) :

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

(Ω, \mathcal{A}, P) est alors appelé "espace de probabilité". Dans ce contexte, les éléments de la tribu \mathcal{A} sont appelés des "événements".

Remarque :

Pour tout $A \in \mathcal{A}$, $P(A^c) = 1 - P(A)$. Il suffit de considérer $A_1 = A$, $A_2 = A^c$ et $A_n = \emptyset$ pour tout $n \geq 3$.

Exemples :

1) $\Omega = \{1, \dots, 6\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$

$$\begin{aligned} P : \mathcal{P}(\Omega) &\mapsto \mathbb{R} \\ A &\mapsto \frac{\#A}{6} \end{aligned}$$

Alors P est une probabilité sur (Ω, \mathcal{A}) . $P(\emptyset) = \frac{0}{6} = 0$ $P(\Omega) = \frac{6}{6} = 1$ et si $(A_n)_{n \geq 1}$ suite d'événements deux à deux disjoints de \mathcal{A} , on a $\#(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} \#A_n$, donc

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

2) Généralisation $\Omega = \{a_1, \dots, a_k\}$, $k \geq 1$. Soient $p_1, p_2, \dots, p_k \geq 0$ tels que $p_1 + \dots + p_k = 1$. Soit

$$\begin{aligned} P : \mathcal{P}(\Omega) &\mapsto \mathbb{R} \\ A &\mapsto \sum_{i \in \llbracket 1, k \rrbracket, a_i \in A} p_i = \sum_{i=1}^k p_i \mathbb{1}_A(a_i) \end{aligned}$$

Alors P est une probabilité sur (Ω, \mathcal{A}) .

En effet, $P(\emptyset) = 0$ $P(\Omega) = \sum_{i=1}^k p_i = 1$ et si $(A_n)_{n \geq 1}$ suite d'événements deux à deux disjoints de \mathcal{A} , on a $\mathbb{1}_{\left\{\bigcup_{n \geq 1} A_n\right\}} = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$ et donc $P\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

3) Soit $f : \mathbb{R} \mapsto \mathbb{R}$ une fonction mesurable positive telle que $\int_{\mathbb{R}} f(x) dx = 1$ et soit

$$\begin{aligned} P : \mathcal{B}(\mathbb{R}) &\mapsto \mathbb{R} \\ A &\mapsto \int_A f(x) dx \end{aligned}$$

Alors P est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

En effet, pour $(A_n)_{n \geq 1}$ suite d'événements deux à deux disjoints de $\mathcal{B}(\mathbb{R})$ on a

$$\mathbb{1}_{\left\{\bigcup_{n \geq 1} A_n\right\}} = \sum_{n \geq 1} \mathbb{1}_{A_n}$$

$$\begin{aligned} \int_{\bigcup_{n \geq 1} A_n} f(x) dx &= \int_{\mathbb{R}} f(x) \mathbb{1}_{\left\{\bigcup_{n \geq 1} A_n\right\}}(x) dx \\ &= \int_{\mathbb{R}} f(x) \sum_{n \geq 1} \mathbb{1}_{A_n}(x) dx \\ &= \sum_{n \geq 1} \int_{\mathbb{R}} f(x) \mathbb{1}_{A_n}(x) dx \\ &= \sum_{n \geq 1} P(A_n) \end{aligned}$$

Proposition 5[5.1]1

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soient $A, B \in \mathcal{A}$.

1. Si $A \cap B = \emptyset$ alors $P(A \cup B) = P(A) + P(B)$

2. $P(A \setminus B) = P(A) - P(A \cap B)$
3. $P(A^c) = 1 - P(A)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. "Croissance de P ". Si $A \subset B$ alors $P(A) \leq P(B)$
6. "Borne d'union".
 - (a) $P(A \cup B) \leq P(A) + P(B)$
 - (b) Si $(A_n)_{n \geq 1}$ est une suite d'événements alors $P\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$, où cette somme peut être éventuellement infinie.
7. Si $(A_n)_{n \geq 1}$ est une suite croissante d'événements alors $P\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$.
8. Si $(A_n)_{n \geq 1}$ est une suite décroissante d'événements alors $P\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$.

Preuve :

- ① On considère la suite $(A_n)_{n \geq 1}$ telle que $A_1 = A, A_2 = B, A_k = \emptyset, \quad \forall k \geq 3$ et on a

$$P(A \cup B) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = P(A) + P(B)$$

- ② On a $A \setminus B \cap (A \cap B) = \emptyset$, et $A \setminus B \cup (A \cap B) = A$ donc

$$P(A) = P(A \setminus B \cup (A \cap B)) = P(A \setminus B) + P(A \cap B)$$

- ③ On a $\Omega = A \cup A^c$ avec bien évidemment $A \cap A^c = \emptyset$ donc

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

- ④ On a $A \setminus B \cap (A \cap B) = \emptyset$ donc

$$P(A \setminus B \cup (A \cap B)) = P(A \setminus B) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$$

- ⑤

$$P(B) = P(B \setminus A \cup A) = P(B \setminus A) + P(A \cap B) \geq P(A)$$

- ⑥ On a pour le premier point

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$$

Pour le second point, on a

$$P\left(\bigcup_{n \geq 1} A_n\right) = P(A_1) + P\left(\bigcup_{n \geq 2} A_n^{A_1 \cap \dots \cap A_{n-1}}\right) = P(A_1) + \sum_{n \geq 2} P(A_n^{A_1 \cap \dots \cap A_{n-1}}) \leq \sum_{n \geq 1} P(A_n)$$

car pour tout $n \geq 1$, $A_n^{A_1 \cap \dots \cap A_{n-1}} \subset A_n$.

⑦ On a

$$\begin{aligned} P\left(\bigcup_{n \geq 1} A_n\right) &= P(A_1) + P\left(\bigcup_{n \geq 2} A_n^{A_{n-1}}\right) \\ &= P(A_1) + \sum_{n \geq 2} P(A_n^{A_{n-1}}) \\ &= \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

⑧ On a de même que si $A_1 \supset A_2 \dots \supset A_n$, alors $A_1^c \subset \dots \subset A_n^c$ donc $(A_n^c)_{n \geq 1}$ est une suite croissante d'événements. Ainsi

$$P\left(\bigcup_{n \geq 1} A_n^c\right) = \lim_{n \rightarrow \infty} P(A_n^c) \quad \xleftrightarrow{\text{passage au complémentaire}} \quad P\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

5.2 Expériences Aléatoires et Modélisation

Une expérience aléatoire est une expérience dans laquelle on modélise le résultat de manière aléatoire.

Une manière naturelle consiste à définir Ω comme l'ensemble des résultats possibles de l'expérience et \mathcal{A} comme la tribu engendrée par les événements observables de l'expérience, puis en définissant une probabilité P sur (Ω, \mathcal{A}) permettant de quantifier l'aléa.

Ici, un "événement observable" est une partie A de Ω telle qu'on peut déterminer, à l'issue de l'expérience, si le résultat est bien dans A .

Exemples.

1. Lancé d'un dé équilibré à 6 faces. Ici on connaît le résultat du lancer.

On prend $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$ et

$$\begin{aligned} P : \mathcal{P}(\Omega) &\longrightarrow \mathbb{R} \\ A &\longmapsto \frac{\#A}{6} \end{aligned}$$

2. Lancé d'un dé équilibré mais on n'observe que la parité du résultat. Ici on a uniquement l'information sur la parité du résultat.

On prend $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathcal{A} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ et la même proba qu'avant.

3. Une urne contient 6 boules : 3 jaunes, 1 rouge et 2 vertes. On en tire une au hasard. Deux boules de même couleur sont indiscernables.

On prend $\Omega = \{J_1, J_2, J_3, R, V_1, V_2\}$ et :

$$\mathcal{A} = \left\{ \emptyset, \{J_1, J_2, J_3\}, \{R\}, \{V_1, V_2\}, \{J_1, J_2, J_3, R\}, \{J_1, J_2, J_3, V_1, V_2\}, \{R, V_1, V_2\}, \Omega \right\}$$

4. Lancé d'une craie dans un bac à sable de longueur $L \in \mathbb{R}_+^*$.

On prend $\Omega = [0, L]$ et $\mathcal{A} = \mathcal{B}([0, L])$, ainsi que

$$\begin{aligned} P : \mathcal{B}([0, L]) &\longrightarrow \mathbb{R} \\ A &\longmapsto \frac{\lambda(A)}{L}, \quad \lambda \text{ la mesure de Lebesgues dans } \mathbb{R} \end{aligned}$$

5.3 Probabilités Conditionnelles

Soit (Ω, \mathcal{A}, P) un espace de probabilité.

Définition 5|5.3|1

Soit $A \in \mathcal{A}$ tel que $P(A) \neq 0$. Soit $P_A : \mathcal{A} \rightarrow \mathbb{R}$.

$$B \mapsto \frac{P(A \cap B)}{P(A)}$$

On appelle cette application **probabilité conditionnelle** sachant A et on la note généralement pour mesurer un événement $B \in \mathcal{A} : P(B|A)$. P_A est une probabilité sur (Ω, \mathcal{A}) (c'est trivial à montrer).

Proposition 5|5.3|1

Soient $A, B \in \mathcal{A}$ tels que $P(A) \neq 0$ et $P(B) \neq 0$.

— $P(A \cap B) = P(A)P(B | A) = P(B)P(A | B)$

— **Formule de Bayes.** $P(A | B) = \frac{P(B|A)P(A)}{P(B)}$.

— Généralisation. Soient $n \geq 2$ et $A_1, \dots, A_n \in \mathcal{A}$ tels que $P(A_1 \cap \dots \cap A_{n-1}) \neq 0$.

Pour des raisons d'inclusion cela implique que pour tout $k \in \{1, \dots, n-1\}$, $P(A_1 \cap \dots \cap A_k) \neq 0$.

Alors $P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1) \dots P(A_n | A_1 \cap \dots \cap A_{n-1})$.

— **Formule des probabilités totales.** Soient $A_1, \dots, A_n \in \mathcal{A}$ formant une partition de Ω tels que pour tout $i \in \{1, \dots, n\}$, $P(A_i) \neq 0$ (partition voulant dire $\bigcup_{i=1}^n A_i = \Omega$ et $\forall i \neq j, A_i \cap A_j = \emptyset$).

Alors

$$\forall B \in \mathcal{A}, \quad P(B) = \sum_{i=1}^n P(B \mid A_i)P(A_i)$$

6 Variables Aléatoires

Définition 6|6.0|1

Une **variable aléatoire** est une application allant d'un espace probabilisé (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) mesurable.

6.1 Indépendances

On considère dans cette partie un espace probabilisé (Ω, \mathcal{A}, P) .

Ce peut être par exemple $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$, avec $\mathcal{B}(\mathbb{R})$ le borélien de \mathbb{R} et P une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

On peut aussi avoir par exemple $(\Omega, \mathcal{A}, P) = (\{1, 2, \dots, 6\}, \mathcal{P}(\{1, 2, \dots, 6\}), P)$, et P mesure de probabilité sur $(\{1, 2, \dots, 6\}, \mathcal{P}(\{1, 2, \dots, 6\}))$

Définition 6|6.1|1

indépendance mutuelle

1) Soit $n \geq 2$ et $A_1 \dots A_n \in \mathcal{A}$. On dit que $A_1 \dots A_n$ sont mutuellement indépendants si et seulement si

$$\forall I \subset \{1, \dots, n\}, P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i) \quad (1)$$

Avec les conventions $\bigcap_{i \in \emptyset} A_i = \Omega$ et $\prod_{i \in \emptyset} u_i = 1$

2) Soit $(A_n)_{n \geq 1}$ une suite d'événements.

On dit que les $A_n, n \geq 1$, sont mutuellement indépendants si et seulement si pour toute partie **finie** I de \mathbb{N}^* , les $A_i, i \in I$ sont indépendants au sens de (1)

Preuve :

① Démontrons la première implication (\implies). Si pour tout $I \subset \mathbb{N}^*$ fini, les $A_i, i \in I$ sont indépendants, soit $n \in \mathbb{N}^*$, en prenant $I = \{1, \dots, n\}$, les A_1, \dots, A_n sont indépendants.

② (\impliedby) Supposons que pour tout $n \in \mathbb{N}^*$, A_1, \dots, A_n sont indépendants. Et soit $I \subset \mathbb{N}^*$ fini, si $I = \emptyset$, alors $P\left(\bigcup_{i \in \emptyset} A_i\right) = \prod_{i \in \emptyset} P(A_i) = 1$. Si $I \neq \emptyset$, on pose $n = \max(I)$. Par hypothèse, les A_1, \dots, A_n sont indépendants, donc

$$\forall J \subset \{1, \dots, n\}, \quad P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

Et comme $I \subset \{1, \dots, n\}$, c'est encore vrai pour tout $J \subset I$. Donc les $(A_i)_{i \in I}$ sont indépendants.

Définition 6|6.1|2

Indépendances de tribus :

1) Soient $\mathcal{A}_1, \dots, \mathcal{A}_n, n \geq 2$ des sous tribus de \mathcal{A} . On dit que $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont mutuellement indépendantes si et seulement si pour tout $A_1 \dots A_n \in \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ sont indépendants.

2) Soit $(\mathcal{A})_{n \geq 1}$ une suite de sous-tribus de \mathcal{A} . On dit que les $\mathcal{A}_n, n \geq 1$, sont indépendantes si et seulement si pour tout $I \subset \mathbb{N}^*$ fini, les $\mathcal{A}_i, i \in I$ sont indépendantes au sens de la première partie de la définition.

Proposition 6|6.1|1

Soient $(F_1, \mathcal{F}_1) \dots (F_n, \mathcal{F}_n)$ des espaces mesurables.

Et soient $f_1 : (E_1, \mathcal{E}_1) \rightarrow (F_1, \mathcal{F}_1) \dots f_n : (E_n, \mathcal{E}_n) \rightarrow (F_n, \mathcal{F}_n)$ des applications mesurables.

Si X_1, \dots, X_n sont indépendantes, alors $f_1(X_1) \dots f_n(X_n)$ le sont aussi.

Preuve :

On sait que les tribus $\sigma(X_i), \forall i \in \llbracket 1, n \rrbracket$ sont indépendantes, or $\forall B_i \in \mathcal{E}_i$ on a

$$(f_i \circ X_i(B_i))^{-1} = X_i^{-1} \circ f_i^{-1}(B_i) \subset \sigma(X_i)$$

Car $f_i^{-1}(B_i) \in \mathcal{E}_i$ puisque f_i est \mathcal{E}_i -mesurable.

Donc on a bien pour tout $i \in \llbracket 1, n \rrbracket$, $\sigma(f(X_i)) \subset \sigma(X_i)$. Donc $f_1(X_1) \dots f_n(X_n)$ sont indépendantes.

Définition 6|6.1|3

Indépendances de lois de probabilités :

X_1, \dots, X_n , des variables aléatoires à valeur dans $(E_1, \mathcal{E}_1) \dots (E_n, \mathcal{E}_n)$ sont indépendantes si et seulement si les tribus $\sigma(X_1) \dots \sigma(X_n)$ sont indépendantes ($\sigma(X) = \sigma(\{X^{-1}(B) : B \in \mathcal{E}\})$), ce qui est le cas si et seulement si pour tout $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$, on a

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$$

De manière équivalente, X_1, \dots, X_n sont indépendantes si et seulement si leur loi jointe est le produit de leurs lois marginales, *i.e.*

$$P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}$$

Preuve :

" \Leftarrow " : Soient $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$ on a alors

$$P_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) = (P_{X_1} \otimes \dots \otimes P_{X_n})(B_1, \dots, B_n)$$

C'est-à-dire

$$P((X_1, \dots, X_n) \in (B_1 \times \dots \times B_n)) = \prod_{i=1}^n P_{X_i}(B_i)$$

Donc X_1, \dots, X_n sont indépendants.

" \Rightarrow " : Soient $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$, alors :

$$P_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) = (P_{X_1} \otimes \dots \otimes P_{X_n})(B_1, \dots, B_n)$$

donc les probabilités $P_{(X_1, \dots, X_n)}$ et $P_{X_1} \otimes \dots \otimes P_{X_n}$ coïncident sur le π -système

$$\mathcal{C} = \{B_1 \times \dots \times B_n \mid B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n\}$$

Donc coïncident sur $\sigma(\mathcal{C}) = \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ tout entier.

Proposition 6|6.1|2

Soient X_1, \dots, X_n des v.a **indépendantes** à valeurs dans des espaces mesurables $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ respectivement. Supposons que pour tout $i = 1, \dots, n$, X_i admet une densité f_i par rapport à une mesure de référence ν_i (voir plus loin dans le chapitre pour la définition des densités). Alors (X_1, \dots, X_n) admet une densité par rapport à la mesure $\nu_1 \otimes \dots \otimes \nu_n$, donnée par

$$\begin{aligned} f : E_1 \times \dots \times E_n &\mapsto \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto \prod_{i=1}^n f_i(x_i) \end{aligned}$$

Preuve :

Soient $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$ on a alors

$$\begin{aligned}
P(X_1 \in B_1, \dots, X_n \in B_n) &= \int_{B_1 \times \dots \times B_n} f(x_1, \dots, x_n) d\nu_1 \otimes \dots \otimes \nu_n(x_1, \dots, x_n) \\
&= \int_{B_1 \times \dots \times B_n} \prod_{i=1}^n f_i(x_i) d\nu_1(x_1) \dots d\nu_n(x_n) \\
\text{(Fubini-Tonelli)} \quad &= \prod_{i=1}^n \int_{B_i} f_i(x_i) d\nu_i(x_i) \\
&= \prod_{i=1}^n P(X_i \in B_i)
\end{aligned}$$

Donc X_1, \dots, X_n sont indépendantes.

Proposition 6|6.1|3

Soient $(E_1, \mathcal{E}_1) \dots (E_n, \mathcal{E}_n)$ des espaces mesurables ($n \geq 2$). Soient X_1, \dots, X_n des variables aléatoires dans $E_1 \dots E_n$ respectivement.

Les assertions suivantes sont équivalentes :

- (i) X_1, \dots, X_n sont indépendantes.
- (ii) Pour toutes fonctions mesurables positives $g_1 : E_1 \rightarrow \mathbb{R}_+, \dots, g_n : E_n \rightarrow \mathbb{R}_+$ on a

$$\mathbb{E}[g_1(X_1) \dots g_n(X_n)] = \mathbb{E}[g_1(X_1)] \dots \mathbb{E}[g_n(X_n)]$$

- (iii) Pour toutes fonctions $g_1 : E_1 \rightarrow \mathbb{R}, \dots, g_n : E_n \rightarrow \mathbb{R}$ mesurables, telles que $g_i \in L^1(P_{X_i})$, $\forall i \in \llbracket 1, n \rrbracket$ on a :

- $g_1 \otimes \dots \otimes g_n \in L^1(P_{(X_1, \dots, X_n)})$
- $\mathbb{E}[g_1(X_1) \dots g_n(X_n)] = \mathbb{E}[g_1(X_1)] \dots \mathbb{E}[g_n(X_n)]$

(voir plus loin dans le chapitre pour la définition des espérances)

Preuve :

Procédons en 3 étapes.

① " $(i) \implies (ii)$ " : Supposons que X_1, \dots, X_n soient indépendantes dans $(E_i, \mathcal{E}_i), i \in \llbracket 1, n \rrbracket$. Posons pour tout $i \in \llbracket 1, n \rrbracket, g_i : E_i \mapsto \mathbb{R}_+$ des fonctions mesurables positives. Alors :

$$\begin{aligned}
\mathbb{E}[g_1(X_1) \dots g_n(X_n)] &= \int_{E_1 \times \dots \times E_n} g_1(x_1) \dots g_n(x_n) dP_{X_1, \dots, X_n}(x_1, \dots, x_n) \\
&= \int_{E_1 \times \dots \times E_n} g_1(x_1) \dots g_n(x_n) dP_{X_1} \otimes \dots \otimes P_{X_n}(x_1, \dots, x_n) \text{ car } \perp\!\!\!\perp \\
&= \prod_{i=1}^n \int_{E_i} g_i(x_i) dP_{X_i}(x_i) \text{ (Par Fubini-Tonelli)} \\
&= \prod_{i=1}^n \mathbb{E}[g_i(X_i)]
\end{aligned}$$

② "(ii) \implies (iii)" Soient pour tout $i \in \llbracket 1, n \rrbracket$, $g_i : E_i \mapsto \mathbb{R}$ des fonctions mesurables telles que $g_i \in L^1(P_{X_i})$. Pour commencer, comme on a que $|g_i|$ est positives (mesurables positives) on a :

$$\mathbb{E}\left[\prod_{i=1}^n |g_i(X_i)|\right] = \prod_{i=1}^n \mathbb{E}[|g_i(X_i)|] < \infty$$

Ainsi, $|g_1 \otimes \dots \otimes g_n|$ est intégrable, donc, par Fubini, on a $g_1 \otimes \dots \otimes g_n \in L^1(P_{X_1, \dots, X_n})$

Montrons maintenant l'égalité des espérances. Pour chaque $i \in \llbracket 1, n \rrbracket$, décomposons les $g_i = g_i^+ - g_i^-$ (comme en théorie de la mesure), avec $g_i^+ = \max(0, g_i)$, $g_i^- = \max(0, -g_i)$ mesurables positives. Alors :

$$g_1(X_1) \dots g_n(X_n) = \prod_{i=1}^n (g_i^+(X_i) - g_i^-(X_i)) = \sum_{\epsilon \in \{0,1\}^n} (-1)^{\sigma(\epsilon)} \prod_{i=1}^n g_i^{\epsilon_i}(X_i)$$

où $\epsilon = (\epsilon_1, \dots, \epsilon_n)$, $\epsilon_i \in \{+, -\}$, et $\sigma(\epsilon)$ est le nombre de ϵ_i égaux à $-$ dans ϵ . Par linéarité de l'espérance, on a

$$\begin{aligned}
\mathbb{E}[g_1(X_1) \dots g_n(X_n)] &= \sum_{\epsilon \in \{0,1\}^n} (-1)^{\sigma(\epsilon)} \mathbb{E}\left[\prod_{i=1}^n g_i^{\epsilon_i}(X_i)\right] \\
&= \sum_{\epsilon \in \{0,1\}^n} (-1)^{\sigma(\epsilon)} \prod_{i=1}^n \mathbb{E}[g_i^{\epsilon_i}(X_i)] \quad (\text{Fubini-Tonelli})
\end{aligned}$$

Or, on remarque que

$$\prod_{i=1}^n \mathbb{E}[g_i(X_i)] = \prod_{i=1}^n (\mathbb{E}[g_i^+(X_i)] - \mathbb{E}[g_i^-(X_i)]) = \sum_{\epsilon \in \{0,1\}^n} (-1)^{\sigma(\epsilon)} \prod_{i=1}^n \mathbb{E}[g_i^{\epsilon_i}(X_i)]$$

Ainsi, $\mathbb{E}[g_1(X_1) \dots g_n(X_n)]$ et $\prod_{i=1}^n \mathbb{E}[g_i(X_i)]$ coïncident ce qui est le résultat recherché.

③ "(iii) \implies (i)" Posons, pour tout $i \in \llbracket 1, n \rrbracket$, $A_i \in \mathcal{E}_i$ et considérons les fonctions $g_i = \mathbb{1}_{A_i}$ des fonctions mesurables. Ces dernières sont trivialement dans $L^1(P_{X_i})$, puisque $g_i \leq 1$, donc

$$\mathbb{E}[g(X_i)] = \int_{E_i} g_i(x) dP_{X_i}(x) \leq 1 \times \int_{E_i} dP_{X_i}(x) = P_{X_i}(E_i) = 1$$

également, on a

$$\mathbb{E}[g_1(X_1) \dots g_n(X_n)] = \mathbb{E}[\mathbb{1}_{A_1}(X_1) \dots \mathbb{1}_{A_n}(X_n)] = P(X_1 \in A_1, \dots, X_n \in A_n)$$

et

$$\mathbb{E}[g_1(X_1)] \dots \mathbb{E}[g_n(X_n)] = \mathbb{E}[\mathbb{1}_{A_1}(X_1)] \dots \mathbb{E}[\mathbb{1}_{A_n}(X_n)] = P(X_1 \in A_1) \dots P(X_n \in A_n)$$

Ce qui démontre que X_1, \dots, X_n sont indépendants.

6.2 Fonctions de Répartitions et Densités

Définition 6|6.2|1

Soit X une variable aléatoire réelle, on définit sa **fonction de répartition** comme

$$\begin{aligned} F : \mathbb{R} &\mapsto [0, 1] \\ t &\mapsto P_X((-\infty, t]) = P(X \leq t) \end{aligned}$$

Proposition 6|6.2|1

La fonction de répartition de X suffit à déterminer sa loi.

Proposition 6|6.2|2

Soit F la fonction de répartition d'une variable aléatoire réelle X , alors :

- F est croissante
- F est continue à droite et limitée à gauche *i.e.*

$$\forall t \in \mathbb{R} \quad \lim_{s \rightarrow t, s > t} F(s) = F(t) \quad \text{et} \quad \lim_{s \rightarrow t, s < t} F(s) \text{ existe, et vaut } P(X < t)$$

$$\bullet F(t) \xrightarrow[t \rightarrow +\infty]{} 1 \quad \text{et} \quad F(t) \xrightarrow[t \rightarrow -\infty]{} 0$$

Preuve :

① Montrons que F est croissante. Pour $s, t \in \mathbb{R}, s \leq t$, alors

$$\{\omega \in \Omega \mid X(\omega) \leq s\} \subset \{\omega \in \Omega \mid X(\omega) \leq t\} \implies F(s) \leq F(t)$$

② Montrons que $F(n) \xrightarrow{n \rightarrow +\infty} 1$. Puisque F est croissante, il suffit de montrer que $\lim_{t \rightarrow +\infty} F(t) = 1$.

Pour $n \geq 1$ notons $A_n = \{\omega \in \Omega \mid X(\omega) \leq n\}$ on a $\lim_{n \rightarrow +\infty} P(A_n) = P(\bigcup_{n \in \mathbb{N}^*} A_n)$ par croissance des A_n .

Pour tout $\omega \in \Omega$, on peut trouver $n \in \mathbb{N}^* \mid \omega \in A_n$ (en prenant par exemple $n = \lfloor X(\omega) \rfloor + 1$) donc $\Omega \subset \bigcup_{n \in \mathbb{N}^*} A_n$ Donc $\Omega = \bigcup_{n \in \mathbb{N}^*} A_n$ et donc $F(n) \xrightarrow{n \rightarrow +\infty} 1$

③ Soit $t \in \mathbb{R}$, montrons que F est continue à droite en t , *i.e.* : $F(x) \xrightarrow{x \rightarrow t, x > t} F(t)$

Par croissance de F , il est suffisant de montrer que $F(t + \frac{1}{n}) \xrightarrow{n \rightarrow +\infty} F(t)$, pour $n \geq 1$. Posons

$B_n = \{\omega \in \Omega \mid X(\omega) \leq t + \frac{1}{n}\}$, on a donc $F(t + \frac{1}{n}) = P(B_n), \forall n \geq 1$. Comme $(B_n)_{n \in \mathbb{N}^*}$ est une suite décroissante d'événements, on a :

$$\lim_{x \rightarrow t, x > t} F(x) = \lim_{n \rightarrow \infty} F(t + \frac{1}{n}) = \lim_{n \rightarrow \infty} P(B_n) = P(\bigcap_{n=1}^{+\infty} B_n)$$

Or, $\bigcap_{n \in \mathbb{N}^*} B_n = \{\omega \in \Omega : \forall n \geq 1, X(\omega) \leq t + \frac{1}{n}\} = \{\omega \in \Omega : X(\omega) \leq t\}$ Donc $F(x) \xrightarrow{x \rightarrow t, x > t} F(t)$

④ Soit $t \in \mathbb{R}$, montrons que F est limitée à gauche *i.e.* : $F(x) \xrightarrow{x \rightarrow t, x < t} P(X < t)$. Il suffit de déterminer $\lim_{n \rightarrow \infty} F(t - \frac{1}{n})$

Pour $n \geq 1$, $C_n = \{\omega \in \Omega : X(\omega) \leq t - \frac{1}{n}\}$, $(C_n)_{n \in \mathbb{N}^*}$ est une suite croissante d'événements, donc $\lim_{n \rightarrow +\infty} P(C_n) = P(\bigcup_{n \in \mathbb{N}^*} C_n)$, or $\bigcup_{n \in \mathbb{N}^*} C_n = \{\omega \in \Omega \mid \exists n \geq 1 : X(\omega) \leq t - \frac{1}{n}\} = \{\omega \in \Omega \mid X(\omega) < t\}$

Donc $\lim_{n \rightarrow \infty} F(t - \frac{1}{n}) = P(X < t)$

Corollaire

Les points de continuité de F sont le $t \in \mathbb{R}$ qui ne sont pas des atomes de X de sorte que

$$\lim_{x \rightarrow t, x < t} F(x) = P(X < t) = P(X \leq t) = F(t)$$

Remarque :

Soit X un vecteur aléatoire dans $\mathbb{R}^d, d \geq 1$, et $x \in \mathbb{R}^d$, on dit que x est un **atome** de X si

$$P(X = x) > 0$$

Définition 6|6.2|2

Soit X une variable aléatoire sur un espace probabilisé (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) , On dit que X **admet une densité** par rapport à η si, et seulement si il existe $f : E \rightarrow \mathbb{R}_+$ mesurable tel que $\forall B \in \mathcal{E}$,

$$P(X \in B) = \int_B f(x) d\eta(x)$$

f est alors appelée **densité** de X par rapport à η

Proposition 6|6.2|3

Si X admet deux densités f et g par rapport à une mesure η , alors $f = g, \eta$ p.p. *i.e.* :

$$\eta(\{x \in E : f(x) \neq g(x)\}) = 0$$

Preuve :

Soit X une variable aléatoire admettant deux densités f et g par rapport à une mesure η . Nous voulons montrer que $f = g$ η -presque partout, c'est-à-dire que :

$$\eta(\{x \in E : f(x) \neq g(x)\}) = 0.$$

Avec η est une mesure sur l'espace mesurable (E, \mathcal{E}) .

① Ensemble de Désaccord :

— Considérons l'ensemble $D = \{x \in E : f(x) \neq g(x)\}$. Nous devons montrer que $\eta(D) = 0$.

② Décomposition de D :

— Décomposons D en deux ensembles disjoints D_1, D_2 tels que $D_1 \cup D_2 = D$

$$D_1 = \{x \in E : f(x) > g(x)\} \quad \text{et} \quad D_2 = \{x \in E : f(x) < g(x)\}.$$

③ Intégrales sur D_1 et D_2 :

— Considérons l'intégrale de f et g sur D_1 :

$$\int_{D_1} f d\eta \quad \text{et} \quad \int_{D_1} g d\eta.$$

— Puisque $f(x) > g(x)$ pour tout $x \in D_1$, nous avons :

$$\int_{D_1} f d\eta > \int_{D_1} g d\eta.$$

— Cependant, comme f et g sont des densités de X par rapport à η , nous avons :

$$\int_{D_1} f d\eta = P(X \in D_1) = \int_{D_1} g d\eta.$$

— Cela implique que :

$$\int_{D_1} (f - g) d\eta = 0.$$

— Puisque $f - g > 0$ sur D_1 , cela implique que $\eta(D_1) = 0$.

— Avec le même raisonnement pour D_2 , on obtient que $\eta(D_2) = 0$.

On conclut :

— Puisque $\eta(D_1) = 0$ et $\eta(D_2) = 0$, nous avons :

$$\eta(D) = \eta(D_1 \cup D_2) = \eta(D_1) + \eta(D_2) = 0.$$

— Par conséquent, $f = g$ presque partout par rapport à η .

Et donc :

$$\eta(\{x \in E : f(x) \neq g(x)\}) = 0.$$

Définition 6|6.2|3

Soit (E, \mathcal{E}) un espace mesurable tel que \mathcal{E} contient tout les singletons. Une variable aléatoire X dans E est dite **discrète** si, t seulement si $P(X \in A) = 1$, où A est l'ensemble des atomes de X .

Remarque :

Cette dernière définition est équivalente à dire que X une variable aléatoire est dite **discrète** si elle admet une densité par rapport à la mesure de comptage sur A (l'ensemble des atomes de X) donnée par

$$\begin{aligned} f : E &\mapsto \mathbb{R} \\ x &\mapsto P(X = x) \end{aligned}$$

On rappelle que la mesure de comptage sur un sous-ensemble au plus dénombrable F de E est la mesure ∂ sur (E, \mathcal{E}) donnée par

$$\begin{aligned} \partial : (E, \mathcal{E}) &\mapsto \mathbb{R}_+ \cup \{\infty\} \\ B &\mapsto \#(B \cap F) \end{aligned}$$

Soit $f : E \rightarrow \mathbb{R}_+$ mesurable, alors

$$\int_E f(x) d\vartheta(x) = \sum_{x \in F} f(x)$$

Définition 6|6.2|4

Une variable aléatoire dans $\mathbb{R}^d, d \geq 1$ est dite "continue" si, et seulement si elle admet une densité par rapport à la mesure de Lebesgue.

Proposition 6|6.2|4

Une variable aléatoire dans $\mathbb{R}^d, d \geq 1$ continue n'admet pas d'atome. (La réciproque n'est pas vraie)

Théorème 6|6.2|1

Formule du changement de variable

Soit X une variable aléatoire dans \mathbb{R}^d admettant une densité f par rapport à la mesure de Lebesgue. Soit U un ouvert de \mathbb{R}^d tel que $X \in U$ p.s.. Soit φ un \mathcal{C}^1 -difféomorphisme de U dans un ouvert V de \mathbb{R}^d .

Alors $\varphi(X)$ admet une densité par rapport à la mesure de Lebesgue, donnée par

$$g : \mathbb{R}^d \mapsto \mathbb{R}$$

$$y \mapsto \begin{cases} f(\varphi^{-1}(y)) |\det J_{\varphi^{-1}}(y)| & \text{Si } y \in V \\ 0 & \text{Sinon} \end{cases}$$

preuve :

$\forall A \in \mathcal{B}(\mathbb{R}^d)$, on a

$$\begin{aligned} P(\varphi(X) \in A) &= P(\varphi(X) \in A, X \in U) + P(\varphi(X) \in A, X \notin U) \\ &= P(\varphi(X) \in A, X \in U) \\ &= P(X \in \varphi^{-1}(A) \cap U) \\ &= \int_{\varphi^{-1}(A) \cap U} f(x) dx \\ &= \int_U f(x) \mathbb{1}_{x \in \varphi^{-1}(A)} dx \end{aligned}$$

On effectue le changement de variable $y = \varphi(x), y \in V, x = \varphi^{-1}(y)$ et donc $dx = |\det J_{\varphi^{-1}(y)}|$ et on obtient

$$\begin{aligned} P(\varphi(X) \in A) &= \int_U f(x) \mathbb{1}_{x \in \varphi^{-1}(A)} dx \\ &= \int_V f(\varphi^{-1}(y)) |\det J_{\varphi^{-1}(y)}| \mathbb{1}_{y \in A} dy \\ &= \int_A f(\varphi^{-1}(y)) |\det J_{\varphi^{-1}(y)}| \mathbb{1}_{y \in V} dy \\ &= \int_A g(y) dy \end{aligned}$$

Remarque :

On dit que φ est un \mathcal{C}^1 -difféomorphisme dans V si :

- φ est bijective de U dans V
- φ est $\mathcal{C}^1(U)$
- $\varphi^{-1} : V \mapsto U$ est $\mathcal{C}^1(V)$

6.3 Espérance

Définition 6|6.3|1

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) . On définit l'espérance $\mathbb{E}[X]$ tel que

$$\mathbb{E}[X] = \begin{cases} \int_{\Omega} X(\omega) dP(\omega) & \text{si } X \in L^1(P) \\ +\infty & \text{sinon} \end{cases}$$

Remarque :

Soit X variable aléatoire dans (Ω, \mathcal{A}, P) dans un espace mesurable (E, \mathcal{E}) on a alors :

$$\mathbb{E}[X] = \int_E x dP_X(x)$$

(C'est direct en remarquant que P_X est la mesure image de P par X).

Si X admet une densité f par rapport à une mesure η , alors on a

$$\mathbb{E}[X] = \int_E x f(x) d\eta(x)$$

Proposition 6|6.3|1

Soit $(X_n)_{n \in \mathbb{N}^*}$ une variable aléatoire réelle positive, alors

$$\mathbb{E}[X] = \int_0^\infty P(X \geq t) dt$$

preuve :

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x dP_X(x) dx \text{ (théorème de transfert)} \\ &= \int_{\mathbb{R}_-^*} x dP_X(x) + \int_{\mathbb{R}_+} x dP_X(x) \\ &= \int_{\mathbb{R}_+} x dP_X(x) \text{ car } P_X(\mathbb{R}_-^*) = \mathbb{P}(X < 0) = 0 \\ &= \int_{\mathbb{R}_+} \left(\int_0^x dt \right) dP_X(x) \\ &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{t \leq x} dt \right) dP_X(x) \\ &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{t \leq x} dP_X(x) \right) dt \text{ (Fubini Toneli)} \\ &= \int_0^\infty P(X \geq t) dt \end{aligned}$$

Proposition 6|6.3|2

Soit X une variable aléatoire discrète, alors

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} P(X \geq k)$$

preuve :

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} kP(X = k) \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^k P(X = k) \text{ car } k = \sum_{j=1}^k 1 \\
&= \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(X = k) \text{ car } \forall n \in \overline{\mathbb{N}}^*, \sum_{k=1}^n \sum_{j=1}^k u_k = \sum_{j=1}^n \sum_{k=j}^n u_k \\
&= \sum_{j=1}^{\infty} P(X \geq j)
\end{aligned}$$

Proposition 6|6.3|3

Fonction test

Soit X une variable aléatoire à valeurs dans un espace mesurable (E, \mathcal{E}) et soit ∂ une mesure de référence sur (E, \mathcal{E}) .

Alors X admet une densité f par rapport à ∂ si, et seulement si pour toute fonction $g : E \mapsto \mathbb{R}$ mesurable et tel que $g \in L^1(P_X)$,

$$\mathbb{E}[g(X)] = \int_E g(x) f(x) d\partial(x)$$

6.4 Moments, Variance, Covariance

Définition 6|6.4|1

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) et $n \geq 1$, alors si $X \in L^n(P)$, on appelle $n^{\text{ième}}$ moment de X le réel $\mathbb{E}[X^n]$.

Définition 6|6.4|2

Soit $p \geq 1$ et X un vecteur aléatoire, on dit que $X \in L^p(P)$ si $\mathbb{E}[\|X\|^p] < +\infty$

Définition 6|6.4|3

Soit X un vecteur aléatoire sur un espace probablisé (Ω, \mathcal{A}, P) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, on dit que $X \in L^\infty(P)$

si

$$\exists M > 0, \quad \|X\| \stackrel{p.s.}{\leq} M \iff P(\omega \in \Omega, |X(\omega)| > M) = 0$$

Par l'équivalence des norme en dimension finie, ça ne dépend pas de la norme.

Proposition 6|6.4|1

Cauchy-Schwarz pour les espérances

Soient X et Y deux variables aléatoires réelles admettant un moment d'ordre deux. Alors :

i) $XY \in L^1(P)$

ii) $\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2] \mathbb{E}[Y^2]}$

Définition 6|6.4|4

Variance

Si $X \in L^2(P)$, on appelle variance de X le réel

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

Cette variable minimise la fonction

$$\begin{aligned} f : \mathbb{R} &\mapsto \mathbb{R} \\ x &\mapsto \mathbb{E}[(X - x)^2] \end{aligned}$$

Définition 6|6.4|5

Covariance

Soit X, Y deux variables aléatoires réelles dans $L^2(P)$. On définit la covariance de X et Y par

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

On a

- i) $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
- ii) $|\text{Cov}(X, Y)| \leq \sqrt{\mathbb{V}(X) \mathbb{V}(Y)}$
- iii) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- iv) $\text{Cov}(X, X) = \mathbb{V}(X)$

Proposition 6|6.4|2

Soit X un vecteur aléatoire dans $L^q(P)$, et $p \leq q$, $(q, p \in \mathbb{N})$, Alors

$$X \in L^p(P)$$

preuve :

On a

$$\|X\|^p \leq 1 \times \mathbb{1}_{\|X\|^p < 1} + \|X\|^q \mathbb{1}_{\|X\|^p \geq 1}$$

Donc

$$\begin{aligned} \mathbb{E}[\|X\|^p] &\leq \mathbb{E}[\mathbb{1}_{\|X\|^p < 1} + \|X\|^q \mathbb{1}_{\|X\|^p \geq 1}] \\ &\leq 1 + \mathbb{E}[\|X\|^q] < +\infty \end{aligned}$$

car $X \in L^q(P)$. Donc $X \in L^p(P)$.

7 Notions de Convergence

7.1 Tout d'abord, les Égalités

Pour toute cette sous partie, on prendra X et Y des variables aléatoires de (Ω, \mathcal{A}, P) dans (E, \mathcal{E})

Définition 7|7.1|1

On dit que X égal presque sûrement Y *i.e.*

$$X \stackrel{p.s}{=} Y$$

si

$$P(\omega \in \Omega \mid X(\omega) = Y(\omega)) = 1$$

Définition 7|7.1|2

On dit que X est égal en distribution (en loi) à Y *i.e.*

$$X \stackrel{(d)}{=} Y$$

si $\forall A \in \mathcal{E}$,

$$P(X^{-1}(A)) = P(X \in A) = \mathbb{P}(Y \in A) = P(Y^{-1}(A))$$

Remarque :

Si $X \stackrel{(d)}{=} Y$, X et Y ne sont pas nécessairement définis sur le même espace probabilisé, on peut avoir X dans (Ω, \mathcal{A}, P) et Y dans $(\Omega', \mathcal{A}', P)$. En revanche, ils sont forcément vers les même espace (E, \mathcal{E})

On a également que

$$X \stackrel{p.s}{=} Y \implies X \stackrel{(d)}{=} Y$$

.

7.2 Convergence de Variable et Vecteur Aléatoire

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires et X une variable aléatoire, sans mention contraire, elles seront toutes d'un espace probabilisé (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) . Pour rappel, si on écrit $(\mathbf{X}_n)_{n \geq 1}$ et \mathbf{X} alors ce sont des vecteurs aléatoires de dimension $d \geq 1$, de (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) .

7.2.1 Convergence Presque Sure

Définition 7|7.2|1

On dit que $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$ i.e. " X_n converge presque sûrement vers X " si

$$P(\{\omega \in \Omega \mid X_n(\omega) \xrightarrow[n \rightarrow +\infty]{} X(\omega)\}) = 1$$

Proposition 7|7.2|1

Soit $(\mathbf{X}_n)_{n \geq 1}$ et \mathbf{X} des vecteur aléatoires de dimension $d \geq 1$, alors

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{X} \iff \forall i \in \llbracket 1, d \rrbracket, \quad \mathbf{X}_n^{(i)} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{X}^{(i)}$$

C'est d'ailleurs équivalent à dire que si $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y$, alors

$$(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} (X, Y)$$

Preuve :

Proposition 7|7.2|2

Soient X_n et X deux variables aléatoires réelles, on a

$$\sup_{k \geq n} |X_k - X| \xrightarrow[n \rightarrow +\infty]{P} 0 \implies X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$$

Preuve :

Soit $V_n = \sup_{k \geq n} |X_k - X|$

$\forall n \in \mathbb{N}^*, V_n$ est positive (donc minorée) et décroissante, donc $(V_n)_{n \geq 1}$ est une suite décroissante minorée donc converge presque sûrement.

Ainsi, comme $V_n \xrightarrow[n \rightarrow +\infty]{P} 0$ alors $V_n \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ (on verra par la suite que convergence presque sûre implique convergence en probabilité).

Donc $\exists \Omega_0 \mid P(\Omega_0) = 1$ et $\forall \omega \in \Omega_0, \quad X_n(\omega)$ converge uniformément vers $X(\omega)$ *i.e.* : $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{U}} X$

Or, convergence uniforme implique convergence simple, donc $\forall \omega \in \Omega_0, \quad X_n(\omega) \xrightarrow[n \rightarrow +\infty]{} X(\omega)$, ainsi on a bien $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$

Proposition 7|7.2|3

Soit $(\mathbf{X}_n)_{n \geq 1}$ un vecteur aléatoire réel de taille $d \geq 1$, dans \mathbb{R}^p et \mathbf{X} un vecteur aléatoire réel de taille $d \geq 1$, dans \mathbb{R}^p

Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbf{X}$, et

$$\begin{aligned} h : \mathbb{R}^d &\mapsto \mathbb{R}^p \\ x &\mapsto h(x) \end{aligned}, p \geq 1$$

une fonction continue, alors

$$h(\mathbf{X}_n) \xrightarrow[n \rightarrow +\infty]{p.s.} h(\mathbf{X})$$

Preuve :

On a $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbf{X}$ *i.e.* : $P(\{\omega \in \Omega \mid \mathbf{X}_n(\omega) \xrightarrow[n \rightarrow +\infty]{} \mathbf{X}(\omega)\}) = 1$. Donc $\exists \Omega_0 \subset \Omega, \quad P(\Omega_0) = 1$ et $\forall \omega \in \Omega_0, \quad \mathbf{X}_n(\omega) \xrightarrow[n \rightarrow +\infty]{} \mathbf{X}(\omega)$ Or, comme h est \mathcal{C}^0 , $\forall \omega \in \Omega_0, \quad h(\mathbf{X}_n(\omega)) \xrightarrow[n \rightarrow +\infty]{} h(\mathbf{X}(\omega))$ D'où

$$h(\mathbf{X}_n) \xrightarrow[n \rightarrow +\infty]{p.s.} h(\mathbf{X})$$

7.2.2 Convergence dans L^p

Définition 7|7.2|2

Soit $(\mathbf{X}_n)_{n \geq 1}$ une suite de vecteurs aléatoires dans $\mathbb{R}^d, d \geq 1$. Si, pour $p \geq 1$, on a pour tout $n \geq 1$,

$$\mathbf{X}_n \in L^p(P) \iff \mathbb{E}[\|\mathbf{X}_n\|^p] < +\infty$$

On dit que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$ si

$$\mathbb{E}[\|\mathbf{X}_n - \mathbf{X}\|^p] \xrightarrow[n \rightarrow +\infty]{} 0$$

(Ne dépend pas de la norme)

Théorème 7|7.2|1

Théorème de convergence dominée, version probabiliste

Soit $(\mathbf{X}_n)_{n \geq 1}$ une suite de vecteurs aléatoires réels dans \mathbb{R}^d , et \mathbf{X} un vecteur aléatoire dans \mathbb{R}^d .

Supposons que :

1) $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{X}$

2) $\mathbf{X} \in L^p(P), \quad p \geq 1$

3) Il existe une variable aléatoire réelle $\mathbf{Y} \in L^p(P)$ tel que $\forall n \geq 1$

$$\|\mathbf{X}_n\| \leq \mathbf{Y} \text{ p.s.}$$

Alors

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$$

preuve :

Par (3), on a que $\mathbf{X}_n \in L^p(P), \quad \forall n \geq 1$. Pour montrer que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$, i.e.

$$\mathbb{E}[\|\mathbf{X}_n - \mathbf{X}\|^p] \xrightarrow[n \rightarrow +\infty]{} 0 \iff \int_{\Omega} \|\mathbf{X}_n(\omega) - \mathbf{X}(\omega)\|^p dP(\omega)$$

On sait que $\mathbf{X}_n(\omega) \xrightarrow[n \rightarrow +\infty]{} \mathbf{X}(\omega)$ P - presque partout donc

$$\|\mathbf{X}_n(\omega) - \mathbf{X}(\omega)\|^p \xrightarrow[n \rightarrow +\infty]{} 0$$

également, on a

$$\begin{aligned} \|\mathbf{X}_n - \mathbf{X}\|^p &\leq (\|\mathbf{X}_n\| + \|\mathbf{X}\|)^p \\ &\leq 2^p(\|\mathbf{X}_n\|^p + \|\mathbf{X}\|^p) \\ &\leq 2^p(\|\mathbf{Y}\|^p + \|\mathbf{X}\|^p) \in L^1(P) \text{ car } \mathbf{X}, \mathbf{Y} \in L^p(P) \end{aligned}$$

Donc $\|\mathbf{X}_n - \mathbf{X}\|$ est dominée par une fonction dans $L^p(P)$, ainsi on a

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{E}[\|\mathbf{X}_n - \mathbf{X}\|^p] &= \lim_{n \rightarrow +\infty} \int_{\Omega} \|\mathbf{X}_n(\omega) - \mathbf{X}(\omega)\|^p dP(\omega) \\ &= \int_{\Omega} \lim_{n \rightarrow +\infty} \|\mathbf{X}_n(\omega) - \mathbf{X}(\omega)\|^p dP(\omega) \text{ (TCD)} \\ &= 0 \end{aligned}$$

Donc $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$

Proposition 7|7.2|4

Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$ et $\mathbf{Y}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{Y}$ (tous des vecteurs aléatoires dans $\mathbb{R}^d, d \geq 1$), alors

$$\mathbf{X}_n + \mathbf{Y}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X} + \mathbf{Y}$$

Proposition 7|7.2|5

Pour X est une variable aléatoire réelle et pour $p, q > 1$ | $\frac{1}{p} + \frac{1}{q} = 1$. Si $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{L^q} Y$ alors :

$$X_n Y_n \xrightarrow[n \rightarrow +\infty]{L^1} XY$$

7.2.3 Convergence en Probabilité

Définition 7|7.2|3

Soit $(\mathbf{X}_n)_{n \geq 1}$ et \mathbf{X} des vecteurs aléatoires de dimension $d \geq 1$, on dit que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$ (\mathbf{X}_n converge en probabilité vers \mathbf{X}) si

$$\forall \epsilon > 0, \quad P(\|\mathbf{X}_n - \mathbf{X}\| > \epsilon) \xrightarrow[n \rightarrow +\infty]{} 0$$

Proposition 7|7.2|6

Soit $(\mathbf{X}_n)_{n \geq 1}$ et \mathbf{X} des vecteurs aléatoires dans $\mathbb{R}^d, d \geq 1$. Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$, et

$$\begin{aligned} h : \mathbb{R}^d &\mapsto \mathbb{R}^p \\ x &\mapsto h(x) \end{aligned}, p \geq 1$$

continue, alors

$$h(\mathbf{X}_n) \xrightarrow[n \rightarrow +\infty]{P} h(\mathbf{X})$$

preuve :

On a $h \in \mathcal{C}^0(\mathbb{R}^d)$ et $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$. Soit $\alpha > 0, M > 0$ tel que $\mathbb{P}(\|\mathbf{X}\| > M) \leq \alpha$

On a que $P(\|\mathbf{X}_n - \mathbf{X}\| > 1) \xrightarrow[n \rightarrow +\infty]{} 0$ car $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$

Soit $K = B(0, M + 1)$, d'après le théorème de Heine, comme h continue, h est absolument continue sur le compact K i.e. : $\forall \epsilon > 0, \exists \eta > 0 \mid \forall x, y \in K, \|x - y\| \leq \eta \implies \|h(x) - h(y)\| \leq \epsilon$

Soit $\epsilon > 0$,

$$\begin{aligned} P(\|h(\mathbf{X}_n) - h(\mathbf{X})\| > \epsilon) &= P(\{\|h(\mathbf{X}_n) - h(\mathbf{X})\| > \epsilon\} \cap (\{\|\mathbf{X}\| \leq M\} \cap \{\|\mathbf{X}_n - \mathbf{X}\| \leq 1\}) + \\ &\quad P(\{\|h(\mathbf{X}_n) - h(\mathbf{X})\| > \epsilon\} \cap (\{\|\mathbf{X}\| > M\} \cup \{\|\mathbf{X}_n - \mathbf{X}\| > 1\})) \\ &\leq P(\{\|h(\mathbf{X}_n) - h(\mathbf{X})\| > \epsilon\} \cap \{\mathbf{X}, \mathbf{X}_n \in K\}) + P(\{\|\mathbf{X}\| > M\} \cup \{\|\mathbf{X}_n - \mathbf{X}\| > 1\}) \\ &\leq P(\|\mathbf{X}_n - \mathbf{X}\| > \eta) + P(\|\mathbf{X}\| > M \cup \|\mathbf{X}_n - \mathbf{X}\| > 1) \\ &\leq \alpha + P(\|\mathbf{X}\| > M) + P(\|\mathbf{X}_n - \mathbf{X}\| > 1) \text{ (Pour } n \text{ assez grand)} \\ &\leq 3\alpha \end{aligned}$$

Donc, $\forall \epsilon > 0, P(\|h(\mathbf{X}_n) - h(\mathbf{X})\| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0$

Ainsi

$$h(\mathbf{X}_n) \xrightarrow[n \rightarrow +\infty]{P} h(\mathbf{X})$$

Proposition 7|7.2|7

Soit $(\mathbf{X}_n)_{n \geq 1}, \mathbf{X}$ et $(\mathbf{Y}_n)_{n \geq 1}, \mathbf{Y}$ des vecteurs aléatoires dans \mathbb{R}^d et \mathbb{R}^n , $d, n \geq 1$. Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$ et $\mathbf{Y}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{Y}$, alors

$$(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow[n \rightarrow +\infty]{P} (\mathbf{X}, \mathbf{Y})$$

Proposition 7|7.2|8

Soient \mathbf{X}_n et \mathbf{Y}_n deux vecteurs aléatoires. Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$ et $\mathbf{Y}_n - \mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} 0$ alors :

$$\mathbf{Y}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$$

Preuve :

Soit $\epsilon > 0$, alors :

$$\begin{aligned} P(\|\mathbf{Y}_n - \mathbf{X}\| > \epsilon) &= P(\|\mathbf{Y}_n - \mathbf{X}_n + \mathbf{X}_n - \mathbf{X}\| > \epsilon) \\ &\text{(I.T.)} \leq P(\|\mathbf{Y}_n - \mathbf{X}_n\| + \|\mathbf{X}_n - \mathbf{X}\| > \epsilon) \\ &\leq P(\|\mathbf{Y}_n - \mathbf{X}_n\| > \frac{\epsilon}{2} \cup \|\mathbf{X}_n - \mathbf{X}\| > \frac{\epsilon}{2}) \\ &\leq \underbrace{P(\|\mathbf{Y}_n - \mathbf{X}_n\| > \frac{\epsilon}{2})}_{\xrightarrow[n \rightarrow +\infty]{P} 0} + \underbrace{P(\|\mathbf{X}_n - \mathbf{X}\| > \frac{\epsilon}{2})}_{\xrightarrow[n \rightarrow +\infty]{P} 0} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

Proposition 7|7.2|9

Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{X}$ ou bien $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$, $p \geq 1$, Alors

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$$

Preuve :

Supposons que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{X}$, alors il existe Ω_0 tel que $P(\Omega_0) = 1$ et pour tout $\omega \in \Omega_0$ on a que pour tout $\epsilon > 0$, il existe $N_\omega \in \mathbb{N}$, $\forall n \geq N_\omega$, $\|\mathbf{X}_n(\omega) - \mathbf{X}(\omega)\| \leq \epsilon$

Donc prenons $\epsilon > 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon) &= \lim_{n \rightarrow \infty} P(\omega \in \Omega \mid (\|\mathbf{X}_n(\omega) - \mathbf{X}(\omega)\| \geq \epsilon)) \\ &= P(\Omega_0) = 0 \end{aligned}$$

Donc $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$

Pour la convergence dans L^1 , supposons que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^1} \mathbf{X}$, alors soit $\epsilon > 0$

$$P(\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon) \underbrace{\leq}_{(\text{Markov})} \frac{\mathbb{E}[\|\mathbf{X}_n - \mathbf{X}\|]}{\epsilon} \xrightarrow[n \rightarrow +\infty]{} 0$$

Donc $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$.

7.2.4 Convergence en Loi

Définition 7|7.2|4

Soit $(\mathbf{X}_n)_{n \geq 1}$ une suite de vecteurs aléatoires dans (E, \mathcal{E}) de taille $d \geq 1$ et \mathbf{X} vecteur aléatoire dans (F, \mathcal{F}) .

On dit que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$ (\mathbf{X}_n **converge en loi** vers \mathbf{X}) si pour toute fonction f **continue bornée**, $f : \mathbb{R}^d \mapsto \mathbb{R}$, on a :

$$\mathbb{E}[f(\mathbf{X}_n)] \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[f(\mathbf{X})]$$

Cette définition est **équivalente à**, si X_n et X sont dans \mathbb{R} à dire que pour tout $t \in \mathbb{R}^{\setminus A}$ où A représente l'ensemble des atomes de X

$$F_{X_n}(t) \xrightarrow[n \rightarrow +\infty]{} F_X(t)$$

Également équivalente à (Théorème de Levy) pour $(\mathbf{X}_n)_{n \geq 1}$ une suite de vecteurs aléatoires dans (E, \mathcal{E}) de taille $d \geq 1$ et \mathbf{X} vecteur aléatoire dans (F, \mathcal{F}) .

$$\forall t \in \mathbb{R}^d, \quad \varphi_{\mathbf{X}_n}(t) \xrightarrow{n \rightarrow +\infty} \varphi_{\mathbf{X}}(t)$$

Avec

$$\begin{aligned} \varphi : \mathbb{R}^d &\mapsto \mathbb{C} \\ x &\mapsto \mathbb{E}\left[e^{ix^T X}\right] \end{aligned}$$

(voir section suivante pour plus de précision sur la fonction caractéristique).

Proposition 7|7.2|10

Soit $\mathbf{X}_1, \dots, \mathbf{X}_n$ et \mathbf{X} des vecteurs aléatoires de dimension $d \geq 1$. Supposons que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$. Alors pour tout continue,

$$\begin{aligned} h : \mathbb{R}^d &\mapsto \mathbb{R}^p \\ x &\mapsto h(x) \end{aligned}$$

$p \geq 1$ On a alors

$$h(\mathbf{X}_n) \xrightarrow[n \rightarrow +\infty]{(d)} h(\mathbf{X})$$

Proposition 7|7.2|11

Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$ et $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{Y}$ alors \mathbf{X} et \mathbf{Y} ont la même loi.

Proposition 7|7.2|12

Soit $\mathbf{X}_1, \dots, \mathbf{X}_n$ et \mathbf{X} des vecteurs aléatoires de dimension $d \geq 1$. Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$, Alors $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$

preuve :

Supposons que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$, alors

$$\forall \epsilon > 0, \mathbb{P}(\|\mathbf{X}_n - \mathbf{X}\| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0$$

Soit $t \in \mathbb{R}$,

$$\begin{aligned}
|\varphi_{\mathbf{X}_n}(t) - \varphi_{\mathbf{X}}(t)| &= |\mathbb{E}[e^{it^T \mathbf{X}_n} - e^{it^T \mathbf{X}}]| \\
&= |\mathbb{E}[e^{it^T \mathbf{X}_n}(1 - e^{it^T \mathbf{X} - it^T \mathbf{X}_n})]| \\
&\leq \mathbb{E}[|e^{it^T \mathbf{X}_n}| |1 - e^{it^T \mathbf{X} - it^T \mathbf{X}_n}|] \quad (\text{I.T.}) \\
&\leq \mathbb{E}[|1 - e^{it^T (\mathbf{X} - \mathbf{X}_n)}|] \\
&\leq \mathbb{E}[|1 - e^{it^T (\mathbf{X} - \mathbf{X}_n)}| \mathbb{1}_{\|\mathbf{X}_n - \mathbf{X}\| \leq \epsilon}] + \mathbb{E}[|1 - e^{it^T (\mathbf{X} - \mathbf{X}_n)}| \mathbb{1}_{\|\mathbf{X}_n - \mathbf{X}\| > \epsilon}]
\end{aligned}$$

Or pour le deuxième terme, on a

$$\begin{aligned}
\mathbb{E}[|1 - e^{it^T (\mathbf{X} - \mathbf{X}_n)}| \mathbb{1}_{\|\mathbf{X}_n - \mathbf{X}\| > \epsilon}] &\leq 2\mathbb{E}[\mathbb{1}_{\|\mathbf{X}_n - \mathbf{X}\| > \epsilon}] \quad (\text{I.T.}) \\
&\leq 2\mathbb{P}(\|\mathbf{X}_n - \mathbf{X}\| > \epsilon) \xrightarrow[n \rightarrow +\infty]{} 0
\end{aligned}$$

Et pour le premier terme, on a

$$\begin{aligned}
\mathbb{E}[|1 - e^{it^T (\mathbf{X} - \mathbf{X}_n)}| \mathbb{1}_{\|\mathbf{X}_n - \mathbf{X}\| \leq \epsilon}] &\leq \mathbb{E}[|1 - (1 + it^T (\mathbf{X} - \mathbf{X}_n))| \mathbb{1}_{\|\mathbf{X}_n - \mathbf{X}\| \leq \epsilon}] \\
&\leq \mathbb{E}[|t| \epsilon] \\
&\leq \epsilon \mathbb{E}[|t|] \xrightarrow[n \rightarrow +\infty]{} 0
\end{aligned}$$

Donc on a bien que

$$|\varphi_{\mathbf{X}_n}(t) - \varphi_{\mathbf{X}}(t)| \xrightarrow[n \rightarrow +\infty]{} 0$$

Ainsi,

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$$

Proposition 7|7.2|13

Soit $(\mathbf{X}_n)_{n \geq 1}$ une suite de variables vecteurs réelles dans $\mathbb{R}^d, d \geq 1$, et $\mathbf{c} \in \mathbb{R}^d$ Alors

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{c} \iff \mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{c} \iff \mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \delta_{\mathbf{c}}$$

preuve :

Prouvons-le en dimension 1. $X_n \xrightarrow[n \rightarrow +\infty]{P} c \implies X_n \xrightarrow[n \rightarrow +\infty]{(d)} c$ est toujours vrai.

Montrons que $X_n \xrightarrow[n \rightarrow +\infty]{(d)} c \implies X_n \xrightarrow[n \rightarrow +\infty]{P} c$.

Soit $\epsilon > 0$, alors

$$\begin{aligned}\mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(\{X_n - c > \epsilon\} \cup \{X_n - c < -\epsilon\}) \text{ car ces deux événements sont disjoints} \\ &= \mathbb{P}(X_n - c > \epsilon) + \mathbb{P}(X_n - c < -\epsilon) \\ &= 1 - \mathbb{P}(X_n \leq c + \epsilon) + \mathbb{P}(X_n < c - \epsilon) \\ &\xrightarrow[n \rightarrow +\infty]{} 1 - P(c \leq c + \epsilon) + \mathbb{P}(c < c + \epsilon) \\ &\xrightarrow[n \rightarrow +\infty]{} 1 - 1 + 0 \\ &= 0\end{aligned}$$

Donc on a bien que

$$\forall \epsilon > 0, \mathbb{P}(|X_n - c| > \epsilon) \xrightarrow[n \rightarrow +\infty]{} 0 \iff X_n \xrightarrow[n \rightarrow +\infty]{P} c$$

Donc

$$X_n \xrightarrow[n \rightarrow +\infty]{(d)} c \iff X_n \xrightarrow[n \rightarrow +\infty]{P} c$$

Théorème 7|7.2|2

Théorème de Scheffé :

Soit $(\mathbf{X}_n)_{n \geq 1}$ une suite de vecteurs aléatoires réels de taille $d \geq 1$ et \mathbf{X} un vecteur aléatoire réel de taille d .

Soit μ une mesure sur $(\mathbb{R}^d; \mathcal{B}(\mathbb{R}^d))$ et supposons que \mathbf{X} et tout les \mathbf{X}_n admettent une densité par rapport à μ .

On note f la densité de \mathbf{X} par rapport à μ et f_n celle des \mathbf{X}_n par rapport à μ , pour tout $n \geq 1$.

Si pour μ presque tout $x \in \mathbb{R}^d$, $f_n(x) \xrightarrow[n \rightarrow +\infty]{} f(x)$ alors

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$$

preuve :

Soit g une fonction continue bornée, on a

$$\mathbb{E}[g(X_n)] = \int_{\mathbb{R}^d} g(x) f_n(x) d\mu(x)$$

et

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g(x) f(x) d\mu(x)$$

On a

$$\begin{aligned} & |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \\ &= \left| \int_{\mathbb{R}^d} g(x) f_n(x) d\mu(x) - \int_{\mathbb{R}^d} g(x) f(x) d\mu(x) \right| \\ &= \left| \int_{\mathbb{R}^d} g(x) (f_n(x) - f(x)) d\mu(x) \right| \\ &\leq \int_{\mathbb{R}^d} |g(x)| |f_n(x) - f(x)| d\mu(x) \\ &\leq M \int_{\mathbb{R}^d} |f_n(x) - f(x)| d\mu(x) \text{ avec } M = \sup_{x \in \mathbb{R}^d} |g(x)| \\ &\leq M \int_{\mathbb{R}^d} [2(f(x) - f_n(x))^+ - (f(x) - f_n(x))] d\mu(x) \\ &\text{car } |u| = 2u^+ - u \text{ avec } u^+ = \sup(0, u) \text{ avec ici } u = f(x) - f_n(x) \\ &\leq M \left(\int_{\mathbb{R}^d} 2(f(x) - f_n(x))^+ d\mu(x) - \int_{\mathbb{R}^d} (f(x) - f_n(x)) d\mu(x) \right) \\ &\leq 2M \left(\int_{\mathbb{R}^d} (f(x) - f_n(x))^+ d\mu(x) - (1 - 1) \right) \\ &\leq 2M \int_{\mathbb{R}^d} (f(x) - f_n(x))^+ d\mu(x) \end{aligned}$$

or on a :

$$\bullet f_n(x) \xrightarrow{n \rightarrow +\infty} f(x)$$

$$\bullet (f(x) - f_n(x))^+ \leq f(x)$$

$$\bullet f \in L^1(\mu)$$

donc par Théorème de convergence dominée, on a

$$\begin{aligned}
& |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \\
& \leq 2M \int_{\mathbb{R}^d} (f(x) - f_n(x))^+ d\mu(x) \\
& \xrightarrow{n \rightarrow +\infty} 0 \\
& \implies \lim_{n \rightarrow \infty} |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| = 0
\end{aligned}$$

Donc $X_n \xrightarrow[n \rightarrow +\infty]{(d)} X$

corollaire :

Soit $E \subset \mathbb{R}^d, d \geq 1$ au plus dénombrable, $\mathbf{X} \in E$ presque sûrement et $\forall n \geq 1, \mathbf{X}_n \in E$ presque sûrement, alors si $\forall \mathbf{x} \in E, P(\mathbf{X}_n = \mathbf{x}) \xrightarrow{n \rightarrow +\infty} P(\mathbf{X} = \mathbf{x})$ on a donc

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$$

En effet, il suffit juste de remarquer que $f_n(x) = P(\mathbf{X}_n = \mathbf{x})$ est une densité de \mathbf{X}_n par rapport à la mesure de comptage, et $f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$ celle de \mathbf{X} pour la même mesure. On sait que \mathbf{X} admet une densité par rapport à cette mesure car $\mathbf{X} \overset{p.s.}{\in} E$ dénombrable.

Proposition 7|7.2|14

Si $\mathbf{X}_n \perp\!\!\!\perp \mathbf{Y}_n$, deux vecteurs aléatoires dans \mathbb{R}^d et $\mathbb{R}^p, d, p \geq 1$ et $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}, \mathbf{Y}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{Y}$, alors on a

$$(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow[n \rightarrow +\infty]{(d)} (\mathbf{X}, \mathbf{Y})$$

preuve :

Soit $\mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{pmatrix} \in \mathbb{R}^{d+p}, \mathbf{t}_1 \in \mathbb{R}^d, \mathbf{t}_2 \in \mathbb{R}^p$, on a :

$$\begin{aligned}
\varphi_{(\mathbf{X}_n, \mathbf{Y}_n)}(\mathbf{t}) &= \mathbb{E}[e^{i\mathbf{t}_1 \mathbf{X}_n + i\mathbf{t}_2 \mathbf{Y}_n}] \\
&= \mathbb{E}[e^{i\mathbf{t}_1 \mathbf{X}_n}] \mathbb{E}[e^{i\mathbf{t}_2 \mathbf{Y}_n}] \quad (\text{par indépendance}) \\
&= \varphi_{\mathbf{X}_n}(\mathbf{t}_1) \varphi_{\mathbf{Y}_n}(\mathbf{t}_2) \xrightarrow{n \rightarrow +\infty} \varphi_{\mathbf{X}}(\mathbf{t}_1) \varphi_{\mathbf{Y}}(\mathbf{t}_2) \quad (\text{car } |\varphi| \leq 1) \\
&= \varphi_{(\mathbf{X}, \mathbf{Y})}(\mathbf{t})
\end{aligned}$$

Même si \mathbf{X} et \mathbf{Y} ne sont pas indépendants, on pourra toujours construire $\mathbf{X}' \stackrel{(d)}{=} \mathbf{X}$ et $\mathbf{Y}' \stackrel{(d)}{=} \mathbf{Y}$ tel que $\mathbf{X}' \perp \mathbf{Y}'$, de sorte que l'on ait une convergence vers $\varphi_{(\mathbf{X}', \mathbf{Y}')}(\mathbf{t}) = \varphi_{(\mathbf{X}, \mathbf{Y})}(\mathbf{t})$.

Pour prouver l'existence de telles variables aléatoires, on va les construire. soient \mathbf{X}', \mathbf{Y}' deux variables aléatoires sur l'espace probabilisé $(\Omega', \mathcal{A}', P') = (\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, P \times P)$. On note aussi pour tout $\omega_1, \omega_2 \in \Omega'$, $\mathbf{X}'(\omega_1, \omega_2) = \mathbf{X}(\omega_1)$ et $\mathbf{Y}'(\omega_1, \omega_2) = \mathbf{Y}(\omega_2)$. on a ainsi clairement $\mathbf{X}' \stackrel{(d)}{=} \mathbf{X}$ car $P_{\mathbf{X}'} = P_{\mathbf{X}}$ et $\mathbf{Y}' \stackrel{(d)}{=} \mathbf{Y}$. De plus, on a $\mathbf{X}' \perp \mathbf{Y}'$ car dépendent de coordonnées différentes de l'espace produit Ω' , ce qui conclut notre preuve.

Remarque :

i) Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$, \mathbf{X}_n et \mathbf{X} ne sont pas nécessairement définis sur le même espace probabilisé, on peut avoir \mathbf{X}_n dans (Ω, \mathcal{A}, P) et \mathbf{X} dans $(\Omega', \mathcal{A}', P)$. En revanche, ils sont forcément vers les même espace (E, \mathcal{E}) .

ii) Si $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbf{X}$ ou $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{X}$ ou $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{L^p} \mathbf{X}$ ($n, p \geq 1$) alors \mathbf{X}_n et \mathbf{X} sont forcément définis sur le même espace probabilisé (Ω, \mathcal{A}, P) (et sont à valeur dans le même espace).

iii) Pour la convergence en probabilité et dans L^p , on ne s'intéresse pas à la norme. En effet, par l'équivalence des normes en dimension finie, cela n'a pas d'importance, on peut donc prendre n'importe quelle norme.

8 Théorèmes de Convergences dans le Cas iid

Théorème 8|8.0|1

① Soit $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alors

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

② On note $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ la variance empirique, alors :

$$n \frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-1)$$

et

$$\hat{\sigma}_n^2 \perp \overline{X}_n$$

Preuve :

① Soit $t \in \mathbb{R}$, on a :

$$\begin{aligned} \varphi_{\overline{X}_n}(t) &= \mathbb{E} \left[e^{i \frac{t}{n} \sum_{i=1}^n X_i} \right] \\ &= (\varphi_{X_1}(\frac{t}{n}))^n \\ &= (e^{i \frac{t}{n} \mu - \frac{1}{2n^2} t^2 \sigma^2})^n \\ &= e^{it\mu - \frac{1}{2n} \sigma^2 t^2} \\ &= \varphi_{\psi}(t) \quad \text{Pour } \psi \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \end{aligned}$$

② Notons $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, $\mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$, avec $Z_i = \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ et donc $\mathbf{Z} \sim \mathcal{N}_n(0, \mathbf{I}_n)$ avec \mathbf{Z} un

vecteur

gaussien, puisque les Z_i sont iid

$$\begin{aligned}
n \frac{\widehat{\sigma}_n^2}{\sigma^2} &= \sum_{i=1}^n \left(\frac{X_i - \overline{X}_n}{\sigma} \right)^2 \\
&= \sum_{i=1}^n (Z_i - \overline{Z}_n)^2 \\
&= \left\| \mathbf{Z} - \frac{1}{n} \mathbb{1}_n^T \mathbf{Z} \mathbb{1}_n \right\|^2 \\
&= \left\| \left(\mathbf{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \right) \mathbf{Z} \right\|^2 \text{ Notons } \Pi_1 = \mathbf{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \text{ et } \Pi_2 = 1 - \Pi_1 \\
&= \left\| \Pi_1 \mathbf{Z} \right\|^2 \sim \chi^2(\text{rg}(\Pi_1)) \text{ Par Cochran (voir chapitre sur les vecteurs gaussiens)}
\end{aligned}$$

Or $\text{rg}(\Pi_1) = n - 1$ donc on a bien $n \frac{\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n - 1)$. Pour montrer que $\overline{X}_n \perp \widehat{\sigma}_n^2$, on sait que

$$\begin{pmatrix} \Pi_1 \mathbf{Z} \\ (\mathbf{I}_n - \Pi_1) \mathbf{Z} \end{pmatrix} = \begin{pmatrix} \Pi_1 \\ (\mathbf{I}_n - \Pi_1) \end{pmatrix} \mathbf{Z}$$

est un vecteur gaussien,

$$\begin{aligned}
\text{Cov}(\Pi_1 \mathbf{Z}, (1 - \Pi_1) \mathbf{Z}) &= \Pi_1 \mathbb{E}[\mathbf{Z} \mathbf{Z}^T] (\mathbf{I} - \Pi_1) \text{ (car } (\mathbf{I}_n - \Pi_1)^T = \mathbf{I}_n - \Pi_1) \\
&= \Pi_1 \mathbf{I}_n (\mathbf{I}_n - \Pi_1) \text{ (} \mathbb{E}[\mathbf{Z} \mathbf{Z}^T] = \mathbb{V}(Z) = \mathbf{I}_n) \\
&= \mathbf{I}_n
\end{aligned}$$

Ainsi , comme c'est un vecteur gaussien, (voir section sur les vecteurs gaussien), on a $\Pi_1 \mathbf{Z} \perp (1 - \Pi_1) \mathbf{Z}$. On a :

$$\begin{aligned}
&\Pi_1 \mathbf{Z} \perp (1 - \Pi_1) \mathbf{Z} \\
&\iff \widehat{\sigma}_n^2 \perp \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \mathbf{Z} \\
&\iff \widehat{\sigma}_n^2 \perp \frac{1}{n} \mathbb{1}_n^T \mathbf{Z} = \overline{X}_n
\end{aligned}$$

8.1 Convergences en Loi

Théorème 8|8.1|1

Théorème Central limite

Soit X_1, \dots, X_n iid $\in L^2(P)$ tel que $\mathbb{E}[X_1] = \mu$ et $\mathbb{V}(X_1) = \sigma^2$, alors

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \sigma^2)$$

Preuve :

Soit $t \in \mathbb{R}$, on a :

$$\begin{aligned} \varphi_{\sqrt{n}(\bar{X}_n - \mu)} &= \mathbb{E} \left[e^{it\sqrt{n}(\bar{X}_n - \mu)} \right] \\ &= \mathbb{E} \left[e^{i \frac{t}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)} \right] \\ &= \prod_{i=1}^n \mathbb{E} \left[e^{i \frac{t}{\sqrt{n}} (X_1 - \mu)} \right] \\ &= (\varphi_{X_1 - \mu}(\frac{t}{\sqrt{n}}))^n \\ &= (\varphi_{X_1 - \mu}(0) + \frac{t}{\sqrt{n}} \varphi'_{X_1 - \mu}(0) + \frac{t^2}{2n} \varphi''_{X_1 - \mu}(0) + o(\frac{1}{n^2}))^n \end{aligned}$$

Or on sait que $\varphi_{X_1 - \mu}(0) = 1$, $\varphi'_{X_1 - \mu}(0) = i\mathbb{E}[X_1 - \mu] = 0$ et que $\varphi''_{X_1 - \mu}(0) = i^2\mathbb{E}[X_1^2] = -\sigma^2$ (car $\mathbb{E}[X_1] = 0$). Donc :

$$\varphi_{\sqrt{n}(\bar{X}_n - \mu)} = (1 - \frac{t^2\sigma^2}{2n} + o(\frac{1}{n^2}))^n$$

Pour éviter l'introduction d'un logarithme complexe, que, pour tout nombre complexes u, z de module inférieur ou égal à 1 on peut observer que

$$\forall n \in \mathbb{N}^*, \quad |z^n - u^n| = |(z - u) \sum_{k=0}^{n-1} z^k u^{n-1-k}| \leq n|z - u|$$

On sait que $e^{-\frac{\sigma^2 t^2}{2n}} = 1 - \frac{\sigma^2 t^2}{2n} + o(\frac{1}{n})$ et $\varphi_{X_1 - \mu}(\frac{t}{\sqrt{n}}) = 1 - \frac{t^2\sigma^2}{2n} + o(\frac{1}{n^2})$ donc

$$\begin{aligned}
|\varphi_{\sqrt{n}(\bar{X}_n - \mu)} - e^{-\frac{\sigma^2 t^2}{2}}| &= |(\varphi_{X_1 - \mu}(\frac{t}{\sqrt{n}}))^n - (e^{-\frac{\sigma^2 t^2}{2n}})^n| \\
&\leq n |\varphi_{X_1 - \mu}(\frac{t}{\sqrt{n}}) - e^{-\frac{\sigma^2 t^2}{2n}}| \\
&\leq n |o(\frac{1}{n})| = o(1) \xrightarrow{n \rightarrow +\infty} 0
\end{aligned}$$

Donc $\varphi_{\sqrt{n}(\bar{X}_n - \mu)} \xrightarrow{n \rightarrow +\infty} e^{-\frac{\sigma^2 t^2}{2}} = \varphi_\psi(t)$ pour $\psi \sim \mathcal{N}(0, \sigma^2)$

Donc on a par théorème de Levy, que

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \sigma^2)$$

Théorème 8|8.1|2

Théorème de Slutsky :

Soit \mathbf{X}_n et \mathbf{X} des vecteurs aléatoires de taille $d \geq 1$ tel que $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{X}$, et \mathbf{Y}_n et $\mathbf{c} \in \mathbb{R}^p$ un vecteur fixé, tous dans \mathbb{R}^p , $p \geq 1$ tel que $\mathbf{Y}_n \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{c}$. Alors

$$\begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{(d)} \begin{pmatrix} \mathbf{X} \\ \mathbf{c} \end{pmatrix}$$

Preuve :

Soit $\mathbf{s} \in \mathbb{R}^p, \mathbf{t} \in \mathbb{R}^p$, montrons que $\varphi_{\mathbf{X}_n, \mathbf{Y}_n}(\mathbf{s}, \mathbf{t}) \xrightarrow{n \rightarrow +\infty} \varphi_{\mathbf{X}, \mathbf{c}}(\mathbf{s}, \mathbf{t})$

$$i.e. : \mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X}_n + \mathbf{t}^T \mathbf{Y}_n)} \right] \xrightarrow{n \rightarrow +\infty} \mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X} + \mathbf{t}^T \mathbf{c})} \right]$$

$$\begin{aligned}
|\mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X}_n + \mathbf{t}^T \mathbf{Y}_n)} \right] - \mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X} + \mathbf{t}^T \mathbf{c})} \right]| &= |\mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X}_n + \mathbf{t}^T \mathbf{Y}_n)} \right] - \mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X}_n + \mathbf{t}^T \mathbf{c})} \right] + \mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X}_n + \mathbf{t}^T \mathbf{c})} \right] - \mathbb{E} \left[e^{i(\mathbf{s}^T \mathbf{X} + \mathbf{t}^T \mathbf{c})} \right]| \\
&= |\mathbb{E} \left[e^{i\mathbf{s}^T \mathbf{X}_n} (e^{i\mathbf{t}^T \mathbf{Y}_n} - e^{i\mathbf{t}^T \mathbf{c}}) \right] + \mathbb{E} \left[e^{i\mathbf{t}^T \mathbf{c}} (e^{i\mathbf{s}^T \mathbf{X}_n} - e^{i\mathbf{s}^T \mathbf{X}}) \right]| \\
\text{(IT)} &\leq \mathbb{E} \left[|e^{i\mathbf{s}^T \mathbf{X}_n}| |e^{i\mathbf{t}^T \mathbf{Y}_n} - e^{i\mathbf{t}^T \mathbf{c}}| \right] + \mathbb{E} \left[|e^{i\mathbf{t}^T \mathbf{c}}| |e^{i\mathbf{s}^T \mathbf{X}_n} - e^{i\mathbf{s}^T \mathbf{X}}| \right] \\
\text{car } |e^{i\mathbf{s}^T \mathbf{X}_n}| &= 1 \leq \mathbb{E} \left[|e^{i\mathbf{t}^T \mathbf{Y}_n} - e^{i\mathbf{t}^T \mathbf{c}}| \right] + \mathbb{E} \left[|e^{i\mathbf{s}^T \mathbf{X}_n} - e^{i\mathbf{s}^T \mathbf{X}}| \right]
\end{aligned}$$

Or, soit $\epsilon > 0$, on a :

$$\begin{aligned}
\mathbb{E}\left[|e^{it^T \mathbf{Y}_n} - e^{it^T \mathbf{c}}|\right] &= \mathbb{E}\left[|e^{it^T \mathbf{Y}_n} - e^{it^T \mathbf{c}}| \mathbb{1}_{\|\mathbf{Y}_n - \mathbf{c}\| > \epsilon}\right] + \mathbb{E}\left[|e^{it^T \mathbf{Y}_n} - e^{it^T \mathbf{c}}| \mathbb{1}_{\|\mathbf{Y}_n - \mathbf{c}\| \leq \epsilon}\right] \\
&\leq 2P(\|\mathbf{Y}_n - \mathbf{c}\| > \epsilon) + \mathbb{E}\left[|\mathbf{t}^T \mathbf{Y}_n - \mathbf{t}^T \mathbf{c}| \mathbb{1}_{\|\mathbf{Y}_n - \mathbf{c}\| \leq \epsilon}\right] \\
&\leq 2P(\|\mathbf{Y}_n - \mathbf{c}\| > \epsilon) + \mathbb{E}\left[|\mathbf{t}^T (\mathbf{Y}_n - \mathbf{c})| \mathbb{1}_{\|\mathbf{Y}_n - \mathbf{c}\| \leq \epsilon}\right] \\
&\leq 2P(\|\mathbf{Y}_n - \mathbf{c}\| > \epsilon) + \mathbb{E}[\|\mathbf{t}\| \|\epsilon\|] \xrightarrow{n \rightarrow +\infty} 0
\end{aligned}$$

Avec le même raisonnement pour le second terme on a bien que $|\mathbb{E}\left[e^{i(\mathbf{s}^T \mathbf{X}_n + \mathbf{t}^T \mathbf{Y}_n)}\right] - \mathbb{E}\left[e^{i(\mathbf{s}^T \mathbf{X} + \mathbf{t}^T \mathbf{c})}\right]| \xrightarrow{n \rightarrow +\infty} 0$
Et ainsi que

$$\begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{(d)} \begin{pmatrix} \mathbf{X} \\ \mathbf{c} \end{pmatrix}$$

Théorème 8|8.1|3

Théorème de Kolmogorov

$X_1, \dots, X_n \sim P^*$ iid à valeur dans \mathbb{R} , avec $F^*(t) = P(x_i \leq t) = P^*((-\infty, t])$ et $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}}$

En posant $D_n = \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F^*(t)|$ on a :

- 1) La loi de D_n ne dépend pas de F^*
- 2) $\sqrt{n} D_n \xrightarrow[n \rightarrow +\infty]{(d)} K$ où $\forall t > 0, F_K(t) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 t^2}$, avec $F_K(t) = 0$ si $t \leq 0$

Théorème 8|8.1|4

Théorème de Donsker

Soient X_1, \dots, X_n des variables aléatoires iid de fonction de répartition F^* . On définit le processus aléatoire empirique $G_n(t, \omega) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}} - F^*(t) \right)$

Alors $G_n \xrightarrow[n \rightarrow +\infty]{(d)} G$ où G est un processus Gaussien dans l'espace des fonctions bornées continues sur $[0, 1]$ avec $\mathbb{E}[G] = 0$

Théorème 8|8.1|5

Delta methode

Soit $(\mathbf{X}_n)_{n \in \mathbb{N}^*}$ une suite de vecteurs aléatoires dans $\mathbb{R}^d, d \geq 1$ de carré intégrable, et

$$\begin{aligned} g : \mathbb{R}^d &\mapsto \mathbb{R}^p \\ x &\mapsto x \end{aligned}$$

($p \geq 1$) une fonction $\mathcal{C}^1(\mathbb{R}^d)$. Soit $\mathbf{a} \in \mathbb{R}^d$ et $\Sigma \in \mathcal{M}_{d \times d}(\mathbb{R})$ telle que :

$$\sqrt{n}(\mathbf{X}_n - \mathbf{a}) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, \Sigma)$$

Alors, si de plus on a $J_g(\mathbf{a})$ de rang plein, *i.e.* : $\text{rang}(J_g(\mathbf{a})) = \min(d, p)$ (équivalent à $g'(a) \neq 0$ en dimension 1, ou a $\det(J_g(\mathbf{a})) \neq 0$ si $d = p$), on a pour $k = \text{rang}(J_g(\mathbf{a}))$

$$\sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_k(0, \nabla g(\mathbf{a})^T \Sigma \nabla g(\mathbf{a}))$$

Preuve : On a :

$$\sqrt{n}(\mathbf{X}_n - \mathbf{a}) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, \Sigma)$$

Soit $g_1 \dots g_p$ des fonctions de \mathbb{R}^d dans \mathbb{R} tel que

$$\forall u \in \mathbb{R}^d, g(u) = \begin{pmatrix} g_1(u) \\ \vdots \\ g_p(u) \end{pmatrix}$$

notons $f_i, \forall i \in \llbracket 1, p \rrbracket$ des fonctions de \mathbb{R} dans \mathbb{R} tel que

$$f_i(t) = g_i(t\mathbf{X}_n + (1-t)\mathbf{a}), \forall t \in \mathbb{R}$$

On a par théorème fondamentale de l'analyse, $\forall i \in \llbracket 1, p \rrbracket$

$$\begin{aligned} f_i(1) - f_i(0) &= \int_0^1 f'_i(t) dt \\ \iff g_i(\mathbf{X}_n) - g_i(\mathbf{a}) &= \int_0^1 (\nabla g_i(t\mathbf{X}_n + (1-t)\mathbf{a}))^T dt (\mathbf{X}_n - \mathbf{a}) \end{aligned}$$

D'où on obtient

$$g(\mathbf{X}_n) - g(\mathbf{a}) = \int_0^1 J_g(t\mathbf{X}_n + (1-t)\mathbf{a}) dt (\mathbf{X}_n - \mathbf{a})$$

équivalent à écrire que

$$g(\mathbf{X}_n) - g(\mathbf{a}) = A_n(\mathbf{X}_n - \mathbf{a}) \quad A_n = \int_0^1 J_g(t\mathbf{X}_n + (1-t)\mathbf{a})dt$$

on a donc

$$\forall n \in \mathbb{N}^*, \sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) = A_n \sqrt{n}(\mathbf{X}_n - \mathbf{a})$$

Or $\sqrt{n}(\mathbf{X}_n - \mathbf{a}) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, \Sigma)$ et $\frac{1}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{} 0$

Donc, par Slutsky on a $\frac{1}{\sqrt{n}} \sqrt{n}(\mathbf{X}_n - \mathbf{a}) \xrightarrow[n \rightarrow +\infty]{(d)} 0 \times Z = 0$ pour $Z \sim \mathcal{N}_d(0, \Sigma)$

Équivalent à $\mathbf{X}_n - \mathbf{a} \xrightarrow[n \rightarrow +\infty]{(d)} 0$ et ainsi $\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{a}$

Comme $g \in \mathcal{C}^1(\mathbb{R}^d)$, on a donc $J_g \in \mathcal{C}^0(\mathbb{R}^p \times \mathbb{R}^d)$ ainsi :

$$J_g(\mathbf{X}_n) \xrightarrow[n \rightarrow +\infty]{P} J_g(\mathbf{a})$$

Montrons que $A_n \xrightarrow[n \rightarrow +\infty]{P} J_g(\mathbf{a})$. Soit $\epsilon > 0$ et $\mathbf{Y}_n(t) := t\mathbf{X}_n + (1-t)\mathbf{a}$, on a :

$$\begin{aligned} P(\|A_n - J_g(\mathbf{a})\| > \epsilon) &= P(\|\int_0^1 J_g(t\mathbf{X}_n + (1-t)\mathbf{a})dt - J_g(\mathbf{a})\| > \epsilon) \\ &= P(\|\int_0^1 [J_g(\mathbf{Y}_n(t)) - J_g(\mathbf{a})]dt\| > \epsilon) \\ &\leq P(\int_0^1 \|J_g(\mathbf{Y}_n(t)) - J_g(\mathbf{a})\|dt > \epsilon) \\ &\leq P(\int_0^1 \max_{x \in [0,1]} \|\{J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\}\|dt > \epsilon) \\ &\leq P(\max_{x \in [0,1]} \|\{J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\}\| > \epsilon) \end{aligned}$$

Soit $x \in [0, 1]$, soit $M > 0 \mid P(\|a\| > M) = 0$ et soit $\chi = B(0, M + 1)$, comme g est $\mathcal{C}^1(\mathbb{R}^d)$, J_g est continue, donc par le théorème de Heine, J_g est absolument continue sur un compact, en particulier, J_g est absolument continue sur le compact χ *i.e.*

$$\forall \epsilon > 0, \exists \eta > 0 \mid \forall \mathbf{x}_1, \mathbf{x}_2 \in \chi, \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \eta \implies \|J_g(\mathbf{x}_1) - J_g(\mathbf{x}_2)\| \leq \epsilon$$

Donc pour $\mathbf{x}_1 = \mathbf{Y}_n(\mathbf{x})$ et $\mathbf{x}_2 = \mathbf{a}$, on a

$$\|\mathbf{Y}_n(x) - \mathbf{a}\| \leq \eta \implies \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| \leq \epsilon$$

D'où par contraposée

$$\|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| > \epsilon \implies \|\mathbf{Y}_n(x) - \mathbf{a}\| > \eta$$

Pour un ω fixé, on pose $t \in \arg \max_{x \in [0,1]} \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\|$

$$\begin{aligned} P(\max_{x \in [0,1]} \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| > \epsilon) &= \mathbb{P}(\{\max_{x \in [0,1]} \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| > \epsilon\} \cap [\{\|\mathbf{a}\| \leq M\} \cap \{\|\mathbf{Y}_n(t) - \mathbf{a}\| \leq 1\}]) \\ &+ P(\{\max_{x \in [0,1]} \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| > \epsilon\} \cap [\{\|\mathbf{a}\| > M\} \cup \{\|\mathbf{Y}_n(t) - \mathbf{a}\| > 1\}]) \\ &\leq P(\{\max_{x \in [0,1]} \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| > \epsilon\} \cap \{\mathbf{Y}_n(t), \mathbf{a} \in \chi\}) \\ &+ P(\{\|\mathbf{a}\| > M\} \cup \{\|\mathbf{Y}_n(t) - \mathbf{a}\| > 1\}) \text{ on rappelle que } P(\|\mathbf{a}\| > M) = 0 \\ &\leq P(\|\mathbf{Y}_n(t) - \mathbf{a}\| > \eta) + 0 + \mathbb{P}(t\|\mathbf{X}_n - \mathbf{a}\| > 1) \quad (\mathbf{Y}_n(t) - \mathbf{a} = t(\mathbf{X}_n - \mathbf{a})) \\ &\leq P(\max_{t \in [0,1]} t\|\mathbf{X}_n - \mathbf{a}\| > \eta) + \mathbb{P}(\|\mathbf{X}_n - \mathbf{a}\| > 1) \quad t \text{ majoré par } 1 \\ &\leq P(\|\mathbf{X}_n - \mathbf{a}\| > \eta) + \mathbb{P}(\|\mathbf{X}_n - \mathbf{a}\| > 1) \xrightarrow{n \rightarrow +\infty} 0 \text{ car } \mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{P} \mathbf{a} \end{aligned}$$

Ainsi on a $\forall \epsilon > 0$,

$$P(\|A_n - J_g(\mathbf{a})\| > \epsilon) \leq P(\max_{x \in [0,1]} \|J_g(\mathbf{Y}_n(x)) - J_g(\mathbf{a})\| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0$$

Donc

$$A_n \xrightarrow[n \rightarrow +\infty]{P} J_g(\mathbf{a})$$

Ainsi, par Slutsky, on obtient

$$\begin{aligned} A_n \sqrt{n}(\mathbf{X}_n - \mathbf{a}) &= \sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_k(0, J_g(\mathbf{a})\Sigma J_g(\mathbf{a})^T) \\ &\iff \sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_k(0, \nabla g(\mathbf{a})^T \Sigma \nabla g(\mathbf{a})) \end{aligned}$$

Remarque :

① Si g est de \mathbb{R}^d dans \mathbb{R}^d il suffit donc simplement que $\nabla g(\mathbf{a})$ ne soit pas de déterminant nul (invertible) et on a donc

$$\sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, \nabla g(\mathbf{a})^T \Sigma \nabla g(\mathbf{a}))$$

② En réalité, on a pas réellement besoin que $\text{rang}(J_g(\mathbf{a}))$ soit de rang plein (soit dit en passant

équivalent $\det(\nabla g(\mathbf{a})\nabla g(\mathbf{a})^T) \neq 0$ si $d \leq p$ ou bien $\det(\nabla g(\mathbf{a})^T \nabla g(\mathbf{a})) \neq 0$ si $p \leq d$). C'est juste que si elle n'est pas de rang plein, on obtiendra une loi normale dégénérée, (une constante) et que cela perd en intérêt, et devient assez inexploitable.

③ Par exemple, on lit régulièrement dans les manuels de statistiques que $\nabla g(\mathbf{a})$ doit être de rang plein équivalent quand $g : \mathbb{R} \mapsto \mathbb{R}$ à dire $g'(a) \neq 0$. Mais en réalité, ça fonctionne tout autant si $g'(a) = 0$, simplement notre loi normale sera une constante. On voit bien que la preuve ne requiert pas le rang plein. Nous pouvons donner l'exemple très simple avec $Z \sim \mathcal{N}(0, 1)$, $X_n = a + \frac{Z}{\sqrt{n}}$ et $g(x) = (x - a)^2$, on a bien $g'(a) = 0$ et $\sqrt{n}(X_n - a) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, 1)$ et on a aussi $\sqrt{n}(g(X_n) - g(a)) = \frac{Z}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{(d)} 0 \stackrel{(d)}{=} \mathcal{N}(0, g'(a) \times 1) = \mathcal{N}(0, 0)$.

Proposition 8|8.1|1

X_1, \dots, X_n iid de loi P^* , alors :

$$(i) \ n\widehat{F}_n(t) \sim \mathcal{B}(n, F^*(t))$$

$$(ii) \ \sqrt{n}(\widehat{F}_n(t) - F^*(t)) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, F^*(t)(1 - F^*(t)))$$

8.2 Convergence Presque Sûre

Théorème 8|8.2|1

Lemme de Borel-Cantelli probabilistique

Soit X_1, \dots, X_n une suite de variables aléatoires réelles $\in L^1(P)$, et X une variable aléatoire réelle, toutes sur (Ω, \mathcal{A}, P) . Si,

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \sum_{i=1}^n P(|X_i - X| > \epsilon) < +\infty$$

Alors, $X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$

Si de plus, les X_1, \dots, X_n sont indépendantes, alors on a

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \sum_{i=1}^n P(|X_i - X| > \epsilon) < +\infty \iff X_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X$$

Théorème 8|8.2|2

Loi forte des grands nombres :

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires dans \mathbb{R}^d iid et intégrable, alors

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[X_1]$$

Preuve :

On va faire le cas où $(X_n)_{n \geq 1} \in L^2(P)$.

Soit $Y_n = X_n - \mathbb{E}[X_1]$, alors montrer que $\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} X \iff \overline{Y}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$ car $\overline{Y}_n = \overline{X}_n - \mathbb{E}[X_1]$

Par Chebychev, on a $\forall \epsilon > 0$,

$$\begin{aligned} \mathbb{P}(|\overline{Y}_{n^2}| > \epsilon) &\leq \frac{\mathbb{V}(\overline{Y}_{n^2})}{\epsilon^2} \\ &\leq \frac{\mathbb{V}\left(\frac{1}{n^2} \sum_{i=1}^{n^2} (X_i - \mathbb{E}[X_1])\right)}{\epsilon^2} \\ &\leq \frac{\mathbb{V}(X_1)}{(n\epsilon)^2} \xrightarrow[n \rightarrow +\infty]{} 0 \end{aligned}$$

Donc

$$\overline{Y}_n^2 \xrightarrow[n \rightarrow +\infty]{P} 0$$

$$\overline{Y}_{n^2} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0 \iff \mathbb{P}(\{\omega \in \Omega \mid \overline{Y}_{n^2}(\omega) \xrightarrow[n \rightarrow +\infty]{} 0\}) = 1$$

Or, Soit $\omega \in \Omega, \overline{Y}_{n^2}(\omega) \xrightarrow[n \rightarrow +\infty]{} 0 \iff \forall l \in \mathbb{Q}, l > 0, \exists n_0 \in \mathbb{N} \mid \forall n \geq n_0, |\overline{Y}_{n^2}| \leq \frac{1}{l}$

Donc $\omega \in \{\omega \in \Omega \mid \overline{Y}_{n^2}(\omega) \xrightarrow[n \rightarrow +\infty]{} 0\} \iff \omega \in \bigcap_{l \in \mathbb{Q}_+^*} \bigcup_{n_0 \in \mathbb{N}} \bigcap_{n=n_0}^{\infty} \{|\overline{Y}_{n^2}| \leq \frac{1}{l}\}$

Soit $l \in \mathbb{Q}_+^*$,

$$\begin{aligned} P(\{\omega \in \Omega \mid \overline{Y}_{n^2}(\omega) \xrightarrow[n \rightarrow +\infty]{} 0\}) &= P(\omega \in \bigcup_{n_0 \in \mathbb{N}} \bigcap_{n=n_0}^{\infty} \{|\overline{Y}_{n^2}| \leq \frac{1}{l}\}) \\ &\iff P(\{\omega \in \Omega \mid \overline{Y}_{n^2}(\omega) \text{ ne converge pas vers } 0\}) = P(\omega \in \bigcap_{n_0 \in \mathbb{N}} \bigcup_{n=n_0}^{\infty} \{|\overline{Y}_{n^2}| > \frac{1}{l}\}) \end{aligned}$$

Or, par Borel Cantelli, si pour une suite d'évènement $(A_n)_{n \in \mathbb{N}}, \sum_{n=1}^{\infty} P(A_n) < \infty$,

alors $P(\limsup A_n) = 0$, en notant $A_n = \{\omega \in \Omega \mid |Y_{n^2}(\omega)| > \frac{1}{l}\}$ (on a montré que

$$\sum_{n=1}^{\infty} \mathbb{P}(|\bar{Y}_{n^2}| > \frac{1}{l}) \leq \sum_{n=1}^{\infty} \frac{\mathbb{V}(X_1)}{(n\epsilon)^2} < \infty \text{ par Riemann})$$

On a donc par Borel Cantelli que

$$\begin{aligned} P(\limsup A_n) &= \mathbb{P}(\omega \in \bigcap_{n_0 \in \mathbb{N}} \bigcup_{n=n_0}^{\infty} \{|\bar{Y}_{n^2}(\omega)| > \frac{1}{l}\}) \\ &= 0 \\ \implies P((\limsup A_n)^c) &= \mathbb{P}(\bigcup_{n_0 \in \mathbb{N}} \bigcap_{n=n_0}^{\infty} \{|\bar{Y}_{n^2}| \leq \frac{1}{l}\}) = 1 \end{aligned}$$

Donc $\bar{Y}_{n^2} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$

Soit $k_n \in \mathbb{N}$ l'unique entier tel que $\forall n \in \mathbb{N}, k_n^2 \leq n \leq (k_n + 1)^2$, alors, soit $n \in \mathbb{N}$,

$$\begin{aligned} |\bar{Y}_n| &= \frac{1}{n} \left| \sum_{i=1}^{k_n^2} Y_i + \frac{1}{n} \sum_{j=k_n^2+1}^n Y_j \right| \\ &= \frac{k_n^2}{n} \left| \frac{1}{k_n^2} \sum_{i=1}^{k_n^2} Y_i + \frac{1}{k_n^2} \sum_{j=k_n^2+1}^n Y_j \right| \\ &\leq |\bar{Y}_{k_n^2}| + \frac{k_n^2}{n} \left| \frac{1}{k_n^2} \sum_{j=k_n^2+1}^n Y_j \right| \\ &\leq |\bar{Y}_{k_n^2}| + \frac{1}{k_n^2} \left| \sum_{j=k_n^2+1}^n Y_j \right| \text{ (on pose } S_n = \sum_{i=1}^n Y_i) \\ &\leq |\bar{Y}_{k_n^2}| + \frac{1}{k_n^2} |S_n - S_{k_n^2}| \\ &\leq |\bar{Y}_{k_n^2}| + \frac{1}{k_n^2} \max_{l \in \{k_n^2, \dots, (k_n+1)^2-1\}} |S_l - S_{k_n^2}| \end{aligned}$$

Ainsi, si $\frac{1}{k_n^2} \max_{l \in \{k_n^2, \dots, (k_n+1)^2-1\}} |S_l - S_{k_n^2}| \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$ on aura donc automatiquement que $\bar{Y}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$ puisqu'on a déjà montré que $\bar{Y}_{k_n^2} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$

Soit $\epsilon > 0$,

$$\begin{aligned}
& \sum_{n \geq 1} \mathbb{P}\left(\frac{1}{k_n^2} \max_{l \in \{k_n^2, \dots, (k_n+1)^2-1\}} |S_l - S_{k_n^2}| > \epsilon\right) \\
&= \sum_{n \geq 1} \mathbb{P}\left(\bigcup_{l=k_n^2}^{(k_n+1)^2-1} \frac{|S_l - S_{k_n^2}|}{k_n^2} > \epsilon\right) \\
&\leq \sum_{n \geq 1} \sum_{l=k_n^2}^{(k_n+1)^2-1} \mathbb{P}\left(\frac{|S_l - S_{k_n^2}|}{k_n^2} > \epsilon\right) \text{ or } \mathbb{E}[S_l - S_{k_n^2}] = 0 \text{ donc} \\
&\leq \sum_{n \geq 1} \sum_{l=k_n^2}^{(k_n+1)^2-1} \frac{\mathbb{V}(S_l - S_{k_n^2})}{\epsilon^2 n k_n^4} \\
&\leq \sum_{n \geq 1} \sum_{l=k_n^2}^{(k_n+1)^2-1} \sum_{i=k_n^2+1}^l \frac{\mathbb{V}(Y_i)}{\epsilon^2 n k_n^4} \\
&\leq \sum_{n \geq 1} \sum_{l=k_n^2}^{(k_n+1)^2-1} \frac{(l - k_n^2) \mathbb{V}(Y_1)}{\epsilon^2 n k_n^4} \\
&\leq \sum_{n \geq 1} \frac{((k_n+1)^2 - 1 - k_n^2) \mathbb{V}(Y_1)}{\epsilon^2 n k_n^4} \\
&\leq \sum_{n \geq 1} \frac{2}{\epsilon^2 k_n^3} \mathbb{V}(Y_1)
\end{aligned}$$

Or,

$$\sum_{n \geq 1} \frac{((k_n+1)^2 - 1 - k_n^2) \mathbb{V}(Y_1)}{\epsilon^2 n k_n^4} \leq \sum_{n \geq 1} \frac{2}{\epsilon^2 n^{\frac{3}{2}}} \mathbb{V}(Y_1) < \infty \quad (\text{Selon Riemman})$$

On a donc bien

$$\frac{1}{k_n^2} \max_{l \in \{k_n^2, \dots, (k_n+1)^2-1\}} |S_l - S_{k_n^2}| \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$$

Comme

$$|\bar{Y}_n| \leq |\bar{Y}_{k_n^2}| + \frac{1}{k_n^2} \max_{l \in \{k_n^2, \dots, (k_n+1)^2-1\}} |S_l - S_{k_n^2}|$$

On a bien que

$$\bar{Y}_n = \bar{X}_n - \mathbb{E}[X_1] \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$$

équivalent à

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[X_1]$$

Théorème 8|8.2|3
Glivenko-Cantelli

Soient X_1, \dots, X_n des variables aléatoires iid dans \mathbb{R}^d , alors

$$\sup_{t \in \mathbb{R}^d} |\widehat{F}_n(t) - F^*(t)| \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0$$

Avec $\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq t}$ la fonction de répartition empirique et F^* la fonction de répartition de X_1, \dots, X_n

Remarque :

Par loi forte des grands nombres, on a : $\widehat{F}_n(t) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} F^*(t), \quad \forall t \in \mathbb{R}^d.$

9 Fonction Caractéristique et Fonction Gamma

9.1 Rappel sur la Loi Gamma

On définit la fonction gamma :

$$\begin{aligned}\Gamma : \mathbb{C} \setminus -\mathbb{N} &\mapsto \mathbb{R} \\ x &\mapsto \int_0^\infty t^{x-1} e^{-t} dt\end{aligned}$$

Avec $\forall x \in \mathbb{R} \Gamma(x+1) = x\Gamma(x)$ et donc naturellement, si $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$

Valeurs à connaître

• $\Gamma(1) = 1$ (à noter que $\Gamma(0)$ n'est pas définie donc on ne peut pas dans ce cas aller plus loin dans la récurrence)

$$\bullet \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

On obtient par récurrence :

$$\bullet \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

9.2 Fonction Caractéristique

Soit \mathbf{X} un vecteur aléatoire dans \mathbb{R}^d , $d \geq 1$, alors pour tout $x \in \mathbb{R}^d$, on a la **fonction caractéristique** de \mathbf{X}

$$\begin{aligned}\varphi : \mathbb{R}^d &\mapsto \mathbb{C} \\ x &\mapsto \mathbb{E}\left[e^{ix^T \mathbf{X}}\right]\end{aligned}$$

Théorème de Levy :

Soit $(\mathbf{X}_n)_{n \geq 1}$ une suite de vecteurs aléatoires dans \mathbb{R}^d et \mathbf{X} un vecteur aléatoire dans \mathbb{R}^d , on a alors que

$$\mathbf{X}_n \xrightarrow[n \rightarrow +\infty]{(d)} X \iff \varphi_{\mathbf{X}_n}(t) \xrightarrow[n \rightarrow +\infty]{} \varphi_{\mathbf{X}}(t)$$

avec pour $n \geq 1$, $\varphi_{\mathbf{X}_n}$ la fonction caractéristique de \mathbf{X}_n et $\varphi_{\mathbf{X}}$ la fonction caractéristique de \mathbf{X} .

Proposition 9|9.2|1

Soit \mathbf{X} un vecteur aléatoire dans $\mathbb{R}^d, d \geq 1$ tel que $\mathbf{X} \in L^p(P), p \geq 1$ alors pour tout $k \leq p$ et pour tout $\mathbf{t} \in \mathbb{R}^d$, on a

$$\frac{\partial^k \varphi}{\partial \mathbf{t}^k}(\mathbf{t}) = \frac{1}{i^k} \mathbb{E} \left[\mathbf{X}^k e^{i \mathbf{t}^T \mathbf{X}} \right]$$

Proposition 9|9.2|2

caractérisation de l'indépendance par la fonction caractéristique

Soient \mathbf{X} et \mathbf{Y} deux vecteurs aléatoires sur l'espace probabilisé (Ω, \mathcal{A}, P) dans \mathbb{R}^d et \mathbb{R}^m , avec $d, m \geq 1$. Alors $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ si, et seulement si

$$\forall (\mathbf{s}, \mathbf{t}) \in \mathbb{R}^d \times \mathbb{R}^m, \quad \varphi_{(\mathbf{X}, \mathbf{Y})}(\mathbf{t}, \mathbf{s}) = \varphi_{\mathbf{X}}(\mathbf{t}) \varphi_{\mathbf{Y}}(\mathbf{s})$$

Proposition 9|9.2|3

Soit \mathbf{X} un vecteur aléatoire réel de fonction caractéristique φ . Si, pour tout $t \in \mathbb{R}$, $|\varphi(t)| = 1$, alors \mathbf{X} est une constante, *i.e.* : $\exists c \in \mathbb{R}^d \mid \mathbf{X} \stackrel{p.s.}{=} c$

10 Vecteurs Aléatoires

On définit un vecteur aléatoire \mathbf{X} de dimension $d \geq 1$ sur un ensemble probabilité (Ω, \mathcal{A}, P) dans $(E^d, \mathcal{E}^{\otimes d})$ un vecteur composé de variables aléatoires X_1, \dots, X_d sur (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) . On peut ainsi écrire :

$$\forall \omega \in \Omega, \quad \mathbf{X}(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_d(\omega) \end{pmatrix}$$

10.1 Règles de Calcul à Connaître

Pour un vecteur aléatoire réel $\mathbf{X} \in \mathbb{R}^m, m \geq 1, \mathbf{u} \in \mathbb{R}^m$ on a :

$$1) \mathbb{V}(\mathbf{u}^T \mathbf{X}) = \mathbf{u}^T \mathbb{V}(\mathbf{X}) \mathbf{u}$$

$$2) \text{ Soit } A \in \mathcal{M}_{m \times m}(\mathbb{R}), \mathbb{V}(A\mathbf{X}) = A\mathbb{V}(\mathbf{X})A^T$$

Pour \mathbf{X} un vecteur aléatoire dans $\mathbb{R}^d, d \geq 2$, et \mathbf{Y} un vecteur aléatoire dans $\mathbb{R}^m, m \geq 2$, on définit la matrice de covariance par :

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

et on a donc

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \text{Cov}(\mathbf{Y}, \mathbf{X})^T$$

On définit ainsi $\mathbb{V}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X})$

Proposition 10|10.1|1

inégalité de Cauchy Schwarz

Soient \mathbf{X}, \mathbf{Y} deux vecteurs aléatoires réels de taille $d \geq 1$, de carré intégrable tel que $\mathbb{E}[\mathbf{Y}\mathbf{Y}^T]$ est inversible. Alors

$$\mathbb{E}[\mathbf{X}\mathbf{Y}^T] \mathbb{E}[\mathbf{Y}\mathbf{Y}^T]^{-1} \mathbb{E}[\mathbf{Y}\mathbf{X}^T] \leq \mathbb{E}[\mathbf{X}\mathbf{X}^T]$$

10.2 Vecteur Gaussiens

Définition 10|10.2|1

Soit $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ un vecteur aléatoire de taille $d \geq 1$.

On dit que \mathbf{X} est un **vecteur gaussien** si et seulement si toute combinaison linéaire des coordonnées de \mathbf{X} suit une loi normale dans \mathbb{R} i.e. : $\forall t_1 \dots t_d \in \mathbb{R}, \quad \sum_{i=1}^d t_i X_i \sim \mathcal{N}(\mu, \sigma^2), \quad \sigma^2 \geq 0$

Remarque :

• On rappelle que $\forall \mu \in \mathbb{R}^d, \mathcal{N}(\mu, 0) \stackrel{(d)}{=} \delta_\mu$ et donc que $\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ est un vecteur gaussien.

• si $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ est un vecteur gaussien, alors X_1, \dots, X_n suivent une loi normale.

Définition 10|10.2|2

On appelle $\mathcal{N}_d(\mu, \Sigma)$ la loi de n'importe quel vecteur gaussien \mathbf{X} tel que $\mathbb{E}[\mathbf{X}] = \mu$ et $\mathbb{V}(\mathbf{X}) = \Sigma$

Remarque :

Si $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, alors $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim \mathcal{N}(0, \mathbf{I}_n)$

Proposition 10|10.2|1

Soit $\mu \in \mathbb{R}^d$ et $\Sigma \in \mathcal{S}_d^+$ où \mathcal{S}_d^+ représente l'ensemble des matrices symétriques semi-définies positives, i.e. : $\Sigma \geq 0$

Soit $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$ Alors \mathbf{X} admet une densité par rapport à la mesure de Lebesgue si, et seulement

si Σ est inversible. Dans ce cas, sa densité est donnée par

$$\begin{aligned} f : \mathbb{R}^d &\mapsto \mathbb{R} \\ x &\mapsto \frac{1}{\sqrt{2\pi}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \end{aligned}$$

Preuve :

Si Σ n'est pas inversible, alors il existe $u \in \mathbb{R}^{d \setminus 0_d}$ tel que $\Sigma u = 0_d$ alors $\mathbb{V}(u^T \mathbf{X}) = u^T \Sigma u = 0$ donc $u^T \mathbf{X}$ est une constante presque surement égale à $u^T \mu$.

On a alors pour $H = \{x \in \mathbb{R}^d \mid u^T x = u^T \mu\}$, que $X \in H$ presque surement, or H est un hyperplan affine, donc de mesure de Lebesgue nulle.

Supposons désormais Σ inversible, on a donc que $\Sigma^{-1/2}(\mathbf{X} - \mu) \sim \mathcal{N}_d(0, \mathbf{I}_d)$

Soit $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ tel que $\varphi(y) = \mu + \Sigma^{1/2}y$, on a que $\varphi \in \mathcal{C}^1$, bijective : $\varphi^{-1}(x) = \Sigma^{-1/2}(x - \mu)$ qui est bien \mathcal{C}^1

Donc, d'après la formule du changement de variable, on a que \mathbf{X} admet une densité donnée par

$$\begin{aligned} \mathbb{R}^d &\mapsto \mathbb{R} \\ x &\mapsto f(\varphi^{-1}(x)) |\det J_{\varphi^{-1}}(x)| \end{aligned}$$

où

$$\begin{aligned} f : \mathbb{R}^d &\mapsto \mathbb{R} \\ y &\mapsto \frac{e^{-\frac{\|y\|^2}{2}}}{(2\pi)^{d/2}} \end{aligned}$$

Avec $\det(J_{\varphi^{-1}}(x)) = \Sigma^{-1/2}$ (φ^{-1} est affine)

D'où l'on tire la densité de \mathbf{X} :

$$\begin{aligned} g : \mathbb{R}^d &\mapsto \mathbb{R} \\ x &\mapsto \frac{1}{\sqrt{2\pi}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \end{aligned}$$

Proposition 10|10.2|2

Soit \mathbf{X}, \mathbf{Y} deux vecteurs aléatoires de tailles respectives p et q tel que $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$ est un vecteur gaussien.

Alors $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \iff \text{Cov}(\mathbf{X}, \mathbf{Y}) = 0$

Preuve :

Si $\text{Cov}(\mathbf{X}, \mathbf{Y}) = 0$, alors :

$$\mathbb{V}\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\right) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$$

Alors $\forall (s, t) \in \mathbb{R}^p \times \mathbb{R}^q$ notons $\mathbb{E}[\mathbf{X}] = \mu_1, \mathbb{E}[\mathbf{Y}] = \mu_2, \Sigma_1 = \mathbb{V}(\mathbf{X}), \Sigma_2 = \mathbb{V}(\mathbf{Y})$

$$\begin{aligned} \varphi_{(X,Y)}(s, t) &= e^{i(s^T \mu_1 + t^T \mu_2) - \frac{1}{2} \begin{pmatrix} s \\ t \end{pmatrix}^T \Sigma \begin{pmatrix} s \\ t \end{pmatrix}} \\ &= e^{is^T \mu_1 - \frac{s^T \Sigma_1 s}{2}} e^{it^T \mu_2 - \frac{1}{2} t^T \Sigma_2 t} \\ &= \varphi_X(s) \varphi_Y(t) \end{aligned}$$

(L'autre sens est toujours vrai)

Théorème 10|10.2|1

On l'appelle parfois le théorème de **Cochran**.

Soit $\mathbf{X} \sim \mathcal{N}_d(0, \mathbf{I}_d)$, $d \geq 1$, un vecteur aléatoire. Alors

$$\|\mathbf{X}\|_2^2 \sim \chi^2(d)$$

Preuve :

On a :

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \stackrel{(d)}{=} \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = \mathbf{Y}, \quad Y_1, \dots, Y_d \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$$

Donc, prenons $t \in \mathbb{R}$, on a

$$\begin{aligned} \varphi_{\|\mathbf{X}\|_2^2}(t) &= \mathbb{E}\left[e^{it\|\mathbf{X}\|_2^2}\right] \\ &= \mathbb{E}\left[e^{it\|\mathbf{Y}\|_2^2}\right] \\ &= \varphi_{Y_1^2}(t)^d \end{aligned}$$

Regardons la loi de Y_1^2 , soit h une fonction mesurable positive, on a

$$\begin{aligned}
\mathbb{E}[h(Y_1^2)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h(x^2) e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{+\infty}^0 h(u) e^{-\frac{u}{2}} \frac{1}{\sqrt{2u}} du + \int_0^{+\infty} h(u) e^{-\frac{u}{2}} \frac{1}{\sqrt{2u}} du \right) \\
&= \int_0^{+\infty} h(u) \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} \frac{1}{\sqrt{u}} du
\end{aligned}$$

On reconnaît par la méthode de la fonction test la densité $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} \frac{1}{\sqrt{u}}$ de la loi $\Gamma^{-1}(\frac{1}{2}, \frac{1}{2}) = \chi^2(1)$.

Donc $\varphi_{Y_1^2}(t) = (\frac{\frac{1}{2}}{\frac{1}{2}-it})^{\frac{1}{2}} = (\frac{1}{1-2it})^{\frac{1}{2}}$ et donc

$$(\varphi_{Y_1^2}(t))^d = (\frac{1}{1-2it})^{\frac{d}{2}} = \varphi_{\Psi}(t), \quad \Psi \sim \Gamma^{-1}(\frac{d}{2}, \frac{1}{2}) = \chi^2(d)$$

Et donc on a bien

$$\|\mathbf{X}\|_2^2 \sim \chi^2(d)$$

10.3 Théorèmes de Convergences pour les Vecteurs Aléatoires

Théorème 10|10.3|1

Théorème central limite multi-varié

Soit une suite de vecteurs aléatoires réels $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ indépendants et identiquement distribués tel que $\mathbb{E}[\mathbf{X}_i] = \boldsymbol{\mu}$ et $\mathbb{V}(\mathbf{X}_1) = \boldsymbol{\Sigma}$, le **théorème central limite multivarié** stipule que :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma})$$

11 Espérances Conditionnelles

11.1 Espérances Conditionnelles pour les Variables Aléatoires de Carré Intégrable

Soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , supposons que X admette un moment d'ordre deux. *i.e.* : $X \in L^2(\Omega, \mathcal{A}, P)$.

Soit \mathcal{B} une sous tribu de \mathcal{A} *i.e.* : $\mathcal{B} \subset \mathcal{A}$, où $L^2(\Omega, \mathcal{B}, P)$ est l'ensemble des variables aléatoires \mathcal{B} -mesurable.

Lemme

$L^2(\Omega, \mathcal{A}, P)$ est un **espace Hilbertien** (*i.e.* un espace vectoriel complet muni d'un produit scalaire) avec comme produit scalaire

$$X, Y \in L^2(\Omega, \mathcal{A}, P) \mapsto \mathbb{E}[XY]$$

De plus, si \mathcal{B} est une sous tribu de \mathcal{A} , alors $L^2(\Omega, \mathcal{B}, P)$ est un **sous-espace vectoriel fermé** de $L^2(\Omega, \mathcal{A}, P)$.

Définition 11|11.1|1

Soit $X \in L^2(\Omega, \mathcal{A}, P)$, son espérance conditionnelle sachant \mathcal{B} est la projection orthogonale de X sur $L^2(\Omega, \mathcal{B}, P)$, *i.e.* : l'unique $\tilde{X} \in L^2(\Omega, \mathcal{B}, P) \mid \forall Y \in L^2(\Omega, \mathcal{B}, P)$

$$\mathbb{E}[(X - \tilde{X})Y] = 0 \iff \mathbb{E}[\tilde{X}Y] = \mathbb{E}[XY]$$

On note alors $\tilde{X} = \mathbb{E}[X|\mathcal{B}]$

Remarque :

Si $\mathcal{B} = \{\emptyset, \Omega\}$, alors $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$, en effet, soit $\omega \in \Omega$, et soit $\lambda = \mathbb{E}[X|\mathcal{B}](\omega)$, alors

$$\mathbb{E}[X|\mathcal{B}]^{-1}(\lambda) = \Omega \quad (\text{car } \omega \in \mathbb{E}[X|\mathcal{B}]^{-1}(\lambda)).$$

Donc $\forall \omega \in \Omega$, $\mathbb{E}[X|\mathcal{B}](\omega) = \lambda$. Or, pour toute variable aléatoire \mathcal{B} -mesurable, on a

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]Y]$$

Pour $Y(\omega) = 1$ (constante pour tout $\omega \in \Omega$, donc \mathcal{B} -mesurable) on a donc

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \lambda$$

d'où $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$

Proposition 11|11.1|1

Il est important de noter aussi que $\mathbb{E}[X|\mathcal{B}]$ est l'unique variable aléatoire $\tilde{X} \in L^2(\Omega, \mathcal{B}, P) \mid \forall B \in \mathcal{B}$,

$$\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[\tilde{X}\mathbb{1}_B]$$

Preuve :

Par définition de $\mathbb{E}[X|\mathcal{B}]$, comme $\mathbb{1}_B$ est \mathcal{B} -mesurable, on a bien que

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[\mathbb{E}[X|\mathcal{B}]\mathbb{1}_B] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_B]$$

Montrons désormais que $\mathbb{E}[X|\mathcal{B}]$ est la seule variable aléatoire vérifiant cette propriété.

Soit $\tilde{X} \in L^2(\Omega, \mathcal{B}, P)$ vérifiant

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[\tilde{X}\mathbb{1}_B] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_B]$$

et montrons que $\tilde{X} = \mathbb{E}[X|\mathcal{B}]$. Soit $Z \in L^2(\Omega, \mathcal{B}, P)$, montrons que $\mathbb{E}[XZ] = \mathbb{E}[\tilde{X}Z]$.

On sait qu'il existe une suite $(Z_n)_{n \geq 1}$ de variable aléatoires étagées \mathcal{B} -mesurables telles que

$$\mathbb{E}[(Z_n - Z)^2] \xrightarrow{n \rightarrow +\infty} 0$$

Pour tout $n \geq 1$, $\mathbb{E}[XZ_n] = \mathbb{E}[\tilde{X}Z_n]$ (car $(Z_n)_{n \geq 1}$ étagées), on a juste à montrer que

$$\mathbb{E}[XZ_n] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[XZ] \quad \text{et} \quad \mathbb{E}[\tilde{X}Z_n] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[\tilde{X}Z]$$

On a

$$|\mathbb{E}[XZ_n] - \mathbb{E}[XZ]| = |\mathbb{E}[X(Z_n - Z)]| \stackrel{C.S.}{\leq} \sqrt{\mathbb{E}[X^2] \mathbb{E}[(Z_n - Z)^2]} \xrightarrow{n \rightarrow +\infty} 0$$

De même,

$$|\mathbb{E}[\tilde{X}Z_n] - \mathbb{E}[\tilde{X}Z]| = |\mathbb{E}[\tilde{X}(Z_n - Z)]| \stackrel{C.S.}{\leq} \sqrt{\mathbb{E}[\tilde{X}^2] \mathbb{E}[(Z_n - Z)^2]} \xrightarrow{n \rightarrow +\infty} 0$$

Ce qui montre bien que $\mathbb{E}[XZ_n] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[XZ]$ et $\mathbb{E}[\tilde{X}Z_n] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[\tilde{X}Z]$

Et donc que $\tilde{X} = \mathbb{E}[X|\mathcal{B}]$

Remarque :

Si \mathcal{B} est une sous-tribu de \mathcal{A} engendrée par une variable aléatoire Z sur Ω i.e. $\mathcal{B} = \sigma(Z)$, on note

$$\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X|Z]$$

De plus, pour Y dans (Ω, \mathcal{F}, P) et Z dans (Ω, \mathcal{G}, P) , on a

$$\mathbb{E}[X|Y, Z] = \mathbb{E}[X|\sigma(Y, Z)]$$

Avec $\sigma(Y, Z) = \{(Y, Z)^{-1}(A), A \in \mathcal{F} \otimes \mathcal{G}\}$

Proposition 11|11.1|2

Lemme de Doob-Dynkin

Toute variable aléatoire $\sigma(Z)$ -mesurable peut s'écrire sous la forme $h(Z)$ où h est une fonction mesurable.

Preuve :

Soit X une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{A}, P) , à valeurs dans un espace (E, \mathcal{E}) quelconque. et soit Y sur (Ω, \mathcal{B}, P) dans (F, \mathcal{F}) avec $\mathcal{B} = \sigma(X) = \sigma(\{X^{-1}(A), A \in \mathcal{E}\})$

Montrons que $\exists h : E \mapsto F$ mesurable, tel que $Y = h(X)$. Pour cela, procédons en plusieurs étapes :

① Supposons que $Y = \mathbb{1}_A$, pour un certain $A \in \mathcal{A}$

Alors, nécessairement, il existe $B \in \mathcal{E}$ tel que $A = X^{-1}(B)$, car $A = Y^{-1}(\{1\}) \subset \mathcal{B} = \sigma(X) = \sigma(\{X^{-1}(A), A \in \mathcal{E}\})$ donc $\exists B \in \mathcal{E} \mid X^{-1}(B) = A$. Donc

$$Y^{-1}(\{1\}) = A = X^{-1}(B)$$

$$A^c = (X^{-1}(B))^c = X^{-1}(B^c)$$

$$\Omega = X^{-1}(B) \cup X^{-1}(B^c)$$

$$\phi = X^{-1}(\phi)$$

On prend alors $Y = \mathbb{1}_B(X)$, en effet, $\forall \omega \in \Omega, Y(\omega) = \mathbb{1}_B(X(\omega))$, en effet,

Si $\omega \in A, Y(\omega) = 1$ et $\mathbb{1}_B(X(\omega)) = 1$ car $X(\omega) \in X(A) = X(X^{-1}(B)) \subset B$

Si $\omega \in A^c, Y(\omega) = 0, \mathbb{1}_B(X(\omega)) = 0$ car $X(\omega) \in X(A^c) = X(X^{-1}(A)^c) = X(X^{-1}(A^c)) \subset B^c$

Par conséquent, $Y = \mathbb{1}_B(X)$ et donc $h = \mathbb{1}_B$

② Si Y est étagée, $Y = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} = \sum_{i=1}^n \alpha_i h_i(X) = h(X)$ avec h mesurable comme composée de fonctions mesurables.

③ Supposons enfin Y variable aléatoire quelconque. On peut, sans perte de généralité, supposer que Y est positive car pour tout Y , on pourra se ramener à $Y = h^+ - h^-$ avec $h^+ = \sup(h, 0), h^- = \sup(-h, 0)$

Supposons donc Y une variable aléatoire positive quelconque, et $(Y_n)_{n \geq 1}$ une suite croissante de variables aléatoires étagées positives. Alors $\forall n \in \mathbb{N}^*, Y_n = h_n(X)$, où h_n est une fonction mesurable, et donc

$$\lim_{n \rightarrow +\infty} Y_n = Y = \sup h_n(X) = h(X)$$

avec h mesurable comme sup de fonctions mesurables.

Remarque :

On déduit de la définition des espérances conditionnelles que pour toute fonction f mesurable, et pour toute variable aléatoire X, Y , on a

$$\mathbb{E}[(X - \mathbb{E}[X|Y])f(X)] = 0$$

En effet, on a :

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X|Y])f(X)] &= \mathbb{E}[Xf(X)] - \mathbb{E}[\mathbb{E}[X|Y]f(X)] \\ &= \mathbb{E}[Xf(X)] - \mathbb{E}[Xf(X)] = 0 \text{ (par définition)} \end{aligned}$$

Proposition 11|11.1|3

Soit $X, Y \in L^2(\Omega, \mathcal{A}, P)$, et \mathcal{B} une sous tribu de \mathcal{A} , alors :

1) $\forall \lambda, \mu \in \mathbb{R}$ on a :

$$\mathbb{E}[\lambda X + \mu Y | \mathcal{B}] = \lambda \mathbb{E}[X | \mathcal{B}] + \mu \mathbb{E}[Y | \mathcal{B}]$$

2) Si $X \geq 0$ p.s., Alors :

$$\mathbb{E}[X | \mathcal{B}] \geq 0 \text{ p.s.}$$

3)

$$\mathbb{E}[\mathbb{E}[X | \mathcal{B}]] = \mathbb{E}[X]$$

4) Si \mathcal{E} sous tribu de \mathcal{B} , alors :

$$\mathbb{E}[\mathbb{E}[X | \mathcal{B}] | \mathcal{E}] = \mathbb{E}[X | \mathcal{E}]$$

5) $\forall X \in L^2((\Omega, \mathcal{A}, P))$,

$$|\mathbb{E}[X | \mathcal{B}]| \leq \mathbb{E}[|X| | \mathcal{B}] \text{ p.s.}$$

6) Soit Y une variable aléatoire \mathcal{B} -mesurable bornée, alors :

$$\mathbb{E}[XY | \mathcal{B}] = Y \mathbb{E}[X | \mathcal{B}]$$

Preuve :

① C'est direct par linéarité de l'espérance.

② Supposons que $X \stackrel{p.s.}{\geq} 0$. On pose $\tilde{X} = \mathbb{E}[X | \mathcal{B}]$ et $B = \{\omega \in \Omega \mid \tilde{X}(\omega) < 0\} = \tilde{X}^{-1}(\mathbb{R}_-) \in \mathcal{B}$, car \tilde{X} est \mathcal{B} -mesurable. Ainsi, $\mathbb{1}_B \in L^2(\Omega, \mathcal{B}, P)$, d'où

$$\mathbb{E}[\tilde{X} \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_B]$$

Or, $\mathbb{E}[X\mathbb{1}_B] \geq 0$ car X est positive, donc $\mathbb{E}[\tilde{X}\mathbb{1}_B] \geq 0$. D'autres part, $\tilde{X}\mathbb{1}_B \leq 0$, donc $\tilde{X}\mathbb{1}_B \stackrel{p.s.}{=} 0$.
Donc

$$P(\underbrace{\tilde{X}\mathbb{1}_B \neq 0}_{\supset B}) = 0$$

Donc $P(B) = 0$. et ainsi $P(\tilde{X} \geq 0) = 1$ donc on a bien $\mathbb{E}[X|\mathcal{B}] \stackrel{p.s.}{\geq} 0$

③ Pour $Y \stackrel{p.s.}{=} 1$, \mathcal{B} -mesurable, on a

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]]$$

④ C'est direct par le fait que l'on projette X sur $L^2(\Omega, \mathcal{B}, P)$ puis sur $L^2(\Omega, \mathcal{C}, P)$

⑤ On a $|X| - X \geq 0$, donc $\mathbb{E}[|X| - X|\mathcal{B}] \stackrel{p.s.}{\geq} 0$ i.e.

$$\mathbb{E}[|X||\mathcal{B}] \geq \mathbb{E}[X|\mathcal{B}]$$

De même pour $|X| + X \geq 0$, on a

$$\mathbb{E}[|X||\mathcal{B}] \geq -\mathbb{E}[X|\mathcal{B}]$$

D'où

$$|\mathbb{E}[X|\mathcal{B}]| \stackrel{p.s.}{\leq} \mathbb{E}[|X||\mathcal{B}]$$

⑥ Soit $Z \in L^2(\Omega, \mathcal{B}, P)$, il suffit de montrer que

$$\mathbb{E}[XYZ] = \mathbb{E}[UZ]$$

où $U = Y\mathbb{E}[X|\mathcal{B}]$. On a

$$\mathbb{E}[XYZ] = \mathbb{E}[X(YZ)] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}](YZ)] = \mathbb{E}[(Y\mathbb{E}[X|\mathcal{B}])Z]$$

car $YZ \in L^2(\Omega, \mathcal{B}, P)$.

11.2 Généralisation aux Variable Intégrables

Proposition 11|11.2|1

Soit $X \in L^1(\Omega, \mathcal{A}, P)$, et \mathcal{B} une sous tribu de \mathcal{A} . Alors $\exists! \tilde{X} \in \mathcal{L}^1(\Omega, \mathcal{B}, P) \mid \forall B \in \mathcal{B}$,

$$\mathbb{E}[\tilde{X} \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_B]$$

On note donc $\tilde{X} = \mathbb{E}[X|\mathcal{B}]$.

Preuve :

① Existence : Supposons que $X \geq 0$. Pour $n \geq 1$, notons $X_n = X \mathbb{1}_{X \leq n}$. Alors $X_n \in L^2(\Omega, \mathcal{A}, P)$ car X_n est bornée pour tout n . De plus, $(X_n)_{n \in \mathbb{N}^*}$ est croissante et $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$.

Pour tout $n \geq 1$, soit $\tilde{X}_n = \mathbb{E}[X_n|\mathcal{B}]$. Alors $(\tilde{X}_n)_{n \in \mathbb{N}^*}$ est une suite positive et croissante.

En particulier, \tilde{X}_n converge P -p.p vers une variable aléatoire \tilde{X} à valeur dans $\mathbb{R}_+ \cup \{+\infty\}$ presque-sûrement.

De plus, par la propriété 5|5.1|1, on a

$$\forall n \geq 1, \quad \forall B \in \mathcal{B}, \quad \mathbb{E}[X_n \mathbb{1}_B] = \mathbb{E}[\tilde{X}_n \mathbb{1}_B]$$

En prenant $B = \Omega \in \mathcal{B}$, $\mathbb{E}[X] = \mathbb{E}[\tilde{X}]$ (par théorème de convergence monotone), on a $\mathbb{E}[\tilde{X}] < \infty$, donc $\tilde{X} \in L^1(\Omega, \mathcal{B}, P)$. Lorsque X est de signe quelconque, on peut considérer X^+ et X^- les parties positives et négatives de X

② unicité : Soit \tilde{X} et \tilde{X}' deux variables aléatoires dans $L^1(\Omega, \mathcal{B}, P)$ telles que

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X \mathbb{1}_B] = \mathbb{E}[\tilde{X} \mathbb{1}_B] = \mathbb{E}[\tilde{X}' \mathbb{1}_B]$$

Soit $B = \{\omega \in \Omega \mid \tilde{X}(\omega) < \tilde{X}'(\omega)\} \in \mathcal{B}$. Alors $\mathbb{E}[(\tilde{X} - \tilde{X}') \mathbb{1}_B] = 0$, or $(\tilde{X} - \tilde{X}') \mathbb{1}_B \xrightarrow{p.s.} 0$. On a donc

$$P((\tilde{X} - \tilde{X}') \mathbb{1}_B = 0) = 1 \iff P(\underbrace{(\tilde{X} - \tilde{X}') \mathbb{1}_B}_{\supset B} \neq 0) = 0$$

D'où

$$P(B) = 0$$

Et donc $\tilde{X} \xrightarrow{p.s.} \tilde{X}'$.

En faisant exactement le même raisonnement avec $B' = \{\omega \in \Omega \mid \tilde{X}'(\omega) < \tilde{X}(\omega)\}$ on obtient donc que

$$\tilde{X}' \stackrel{p.s}{=} \tilde{X}$$

Proposition 11|11.2|2

Dans le cas $L^1(\Omega, \mathcal{A}, P)$ aussi, on a les **même propriétés** que les espérances conditionnelles dans $L^2(\Omega, \mathcal{A}, P)$.

Théorème 11|11.2|1

Théorème de Transfert Conditionnelle :

Soient X, Y deux variable aléatoires quelconques à valeurs dans des espaces mesurables (E, \mathcal{E}) et (F, \mathcal{F}) , supposons que $X \perp\!\!\!\perp Y$ et soit h mesurable :

$$\begin{aligned} h : E \times F &\mapsto \mathbb{R} \\ (x; y) &\mapsto h(X, Y) \quad (x=X, y=Y) \end{aligned}$$

$$h \in \mathcal{L}^1(\Omega, \mathcal{A}, P), \text{ i.e. } h \in \mathcal{L}^1(\mathbb{P}_X \otimes \mathbb{P}_Y)$$

Alors,

$$\mathbb{E}[h(X, Y)|Y] = H(Y)$$

où

$$\begin{aligned} H : F &\mapsto \mathbb{R} \\ x &\mapsto \mathbb{E}[h(X, x)] = \int_E h(y, x) dP_X(y) \end{aligned}$$

$$\text{i.e. : } H(Y) = \mathbb{E}[h(X, y)]$$

Preuve :

posons $\mathcal{B} = \sigma(Y) = \sigma(\{Y-1(A), A \in \mathcal{F}\})$, soit $A \in \mathcal{B}$, on a :

$$\begin{aligned} \mathbb{E}[h(X, Y)\mathbb{1}_{Y \in A}] &= \int_{E \times F} h(x, y)\mathbb{1}_{y \in A} dP_X(x) dP_Y(y) \\ &= \int_F \left(\int_E h(x, y)\mathbb{1}_{y \in A} dP_X(x) \right) dP_Y(y) \\ &= \int_F H(y)\mathbb{1}_A dP_Y(y) \\ &= \mathbb{E}[H(Y)\mathbb{1}_A] \end{aligned}$$

Corollaire :

Soit X intégrable, Y une variable aléatoire quelconque tel que $X \perp\!\!\!\perp Y$. Alors

$$\mathbb{E}[X|Y] = \mathbb{E}[h(X, Y)|Y] \quad \text{où} \quad \begin{array}{ll} h : E \times F & \mapsto \mathbb{R} \\ (x; y) & \mapsto x \end{array}$$

$$\begin{aligned} &= H(Y) \\ &= \int_{\mathbb{R}} h(x, Y) dP_X(x) \\ &= \int_{\mathbb{R}} X dP_X(x) \\ &= \mathbb{E}[X] \end{aligned}$$

$$\implies \mathbb{E}[X|Y] = \mathbb{E}[X]$$

Proposition 11|11.2|3

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) telle que $X \in L^1(P)$ et A un évènement tel que $P(A) > 0$.
Alors

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X \mathbb{1}_{\{A\}}]}{\mathbb{E}[\mathbb{1}_{\{A\}}]} = \frac{\mathbb{E}[X \mathbb{1}_{\{A\}}]}{P(A)}$$

Preuve :

On a $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$. Soit $B \in \sigma(A)$, alors $\mathbb{E}[X \mathbb{1}_B] = \mathbb{E}[\mathbb{E}[X|A] \mathbb{1}_B]$. Notons $\tilde{X} = \mathbb{E}[X|\sigma(A)]$
Pour tout $\omega \in \Omega$, on a

- Soit $a \in A$ et $x_a = \tilde{X}(a)$ alors $\tilde{X}^{-1}(x_a) = A$ ou Ω (comme $a \in A$, on a $a \in \tilde{X}^{-1}(x_a)$, alors $\tilde{X}^{-1}(x_a) \neq A^c$, puisque de manière générale, $\tilde{X}^{-1}(\tilde{X}(A)) \supset A$ et $\tilde{X}(\tilde{X}^{-1}(A)) \subset A$)
- Soit $a_c \in A^c$ et $x_{a_c} = \tilde{X}(a_c)$ alors $\tilde{X}^{-1}(x_{a_c}) = A^c$ ou Ω avec le même raisonnement.

(Par ailleurs, si on a X dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ on a donc que $\exists \lambda, \mu \in \mathbb{R}, \forall \omega \in \Omega, \quad \tilde{X}(\omega) = \lambda \mathbb{1}_A + \mu \mathbb{1}_{A^c}$)

Dans tous les cas, \tilde{X} est constante sur A , d'où le fait que $\mathbb{E}[X|A] = \lambda$, ainsi :

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_A] &= \mathbb{E}[\mathbb{E}[X|\sigma(A)] \mathbb{1}_A] \\ &= \underbrace{\mathbb{E}[X|A]}_{=\lambda} \mathbb{E}[\mathbb{1}_A]\end{aligned}$$

D'où

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X\mathbb{1}_A]}{\mathbb{E}[\mathbb{1}_A]}$$

Proposition 11|11.2|4

Formule des espérances totales

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) , $X \in L^1(P)$, et E un ensemble tel que pour $(B_i)_{i \in E}$ une partition de Ω i.e. $\bigcup_{i \in E} B_i = \Omega$ et $\forall i \neq j, B_i \cap B_j = \emptyset$, et tel que $\forall i \in E, P(B_i) > 0$, alors

$$\mathbb{E}[X] = \sum_{i \in E} \mathbb{E}[X|B_i] P(B_i)$$

Preuve :

Comme $(B_i)_{i \in E}$ forme une partition de Ω , on a que $\sum_{i \in E} \mathbb{1}_{B_i} \stackrel{p.s.}{=} 1$.

De plus, comme pour tout $i \in E, P(B_i) > 0$, on a que pour tout $i \in E, \mathbb{E}[X|B_i] = \frac{\mathbb{E}[X\mathbb{1}_{B_i}]}{P(B_i)}$. Ainsi :

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X \sum_{i \in E} \mathbb{1}_{B_i}\right] \\ &= \sum_{i \in E} \mathbb{E}[X\mathbb{1}_{B_i}] \text{ (Fubini)} \\ &= \sum_{i \in E} \mathbb{E}[X|B_i] P(B_i)\end{aligned}$$

Définition 11|11.2|1

Soit \mathbf{X} un vecteur aléatoire réel dans $\mathbb{R}^d, d \geq 1$, intégrable. On définit l'espérance conditionnelle du vecteur \mathbf{X} sachant \mathcal{B} comme

$$\mathbb{E}[\mathbf{X}|\mathcal{B}] = \begin{pmatrix} \mathbb{E}[X_1|\mathcal{B}] \\ \vdots \\ \mathbb{E}[X_d|\mathcal{B}] \end{pmatrix}$$

12 Lois Conditionnelles

12.1 Définitions et Premières Propriétés

Soient $(E, \mathcal{E}), (F, \mathcal{F})$ deux espaces mesurables. Un **noyau de transition** de (E, \mathcal{E}) vers (F, \mathcal{F}) est une fonction

$$\partial : E \times \mathcal{F} \mapsto \mathbb{R}$$

telle que

- $\forall B \in \mathcal{F}, \quad \partial(., B)$ est mesurable
- $\forall x \in E, \quad \partial(x, .)$ est une mesure de probabilité sur (F, \mathcal{F})

Exemple :

$$\forall x \in \mathbb{R}, \mu_x = \mathcal{N}(x, 1)$$

$$\begin{aligned} \partial : \mathbb{R} \times \mathcal{B}(\mathbb{R}) &\mapsto \mathbb{R} \\ (x, B) &\mapsto \mu_x(B) = \frac{1}{2\pi} \int_B e^{-\frac{1}{2}(t-x)^2} dt \end{aligned}$$

est un noyau de transition de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vers $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Proposition 12|12.1|1

Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables, et soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) à valeurs dans E et F . Alors il existe un noyau de transition ∂ de (E, \mathcal{E}) vers (F, \mathcal{F}) tel que :

$$\forall A \in \mathcal{E}, \quad \forall B \in \mathcal{F}, \quad P(\{X \in A\} \cap \{Y \in B\}) = \int_A \underbrace{\partial(x, B)}_{P_{Y|X=x}(B)} dP_X(x)$$

Pour tout $x \in A$, $\partial(x, .)$ est appelée **loi conditionnelle de Y sachant $X = x$** et on la note $P_{Y|X=x}$, ∂ est appelée "loi conditionnelle de Y sachant X "

$$\begin{aligned} \partial : E \times F &\mapsto \mathbb{R} \\ (x, B) &\mapsto P_{Y|X=x}(B) \end{aligned}$$

Exemple :

Soit X une variable aléatoire dans \mathbb{N} , soit $\epsilon \sim \mathcal{N}(0, 1)$ tel que $\epsilon \perp\!\!\!\perp X$. Soit $Y = X + \epsilon$.
Alors, $\forall A \in \mathcal{P}(\mathbb{N}), \quad \forall B \in \mathcal{B}(\mathbb{R}),$

$$\begin{aligned}
P(\{X \in A\} \cap \{Y \in B\}) &= \int_A \mathbb{P}_{Y|X=x}(B) dP_X(x) \\
&= P\left(\bigcup_{x \in A} [X = x] \cap [Y \in B]\right) \\
&= \sum_{x \in A} P(\{X = x\} \cap \{Y \in B\}) \\
&= \sum_{x \in A} P(\{X = x\} \cap \{\epsilon + x \in B\}) \\
&= \sum_{x \in A} P(\{X = x\}) \mathbb{P}(\{\epsilon + x \in B\}) \quad (\text{par indépendance de } X \text{ et } \epsilon) \\
&= \int_A \mu_x(B) dP_X(x) \quad (\epsilon + x \sim \mathcal{N}(x, 1))
\end{aligned}$$

Ainsi, on a

$$\begin{aligned}
\partial : \mathbb{N} \times \mathcal{B}(\mathbb{R}) &\mapsto \mathbb{R} \\
(x, B) &\mapsto \mu_x(B)
\end{aligned}$$

et donc, $\forall x \in \mathbb{N}, \quad P_{Y|X=x} = \mathcal{N}(x, 1)$

Proposition 12|12.1|2

Si P est une mesure de probabilité sur (E, \mathcal{E}) et ∂ est un noyau de transition de (E, \mathcal{E}) vers (F, \mathcal{F}) , alors il existe une unique probabilité Q sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$, telle que :

$$\forall A \in \mathcal{E}, \quad \forall B \in \mathcal{F}, \quad Q(A \times B) = \int_A \partial(x, B) dP(x)$$

12.2 Densités Conditionnelles

Proposition 12|12.2|1

Soient μ_1 et μ_2 deux mesures σ -finies sur (E, \mathcal{E}) et (F, \mathcal{F}) respectivement.

Soient X, Y deux variables aléatoires dans E et F respectivement tel que

$$P_{(X,Y)} \ll \mu_1 \otimes \mu_2, \quad \forall x \in E, \quad \frac{dP_X}{d\mu_1}(x) \neq 0$$

Soit la densité sur F

$$\begin{aligned} f_{Y|X=x} : F &\mapsto \mathbb{R} \\ y &\mapsto \frac{\frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x,y)}{\frac{dP_X}{d\mu_1}(x)} \end{aligned}$$

La fonction associée est une loi conditionnelle de Y sachant $\{X = x\}$. $f_{Y|X=x}$ "**densité conditionnelle de Y sachant $\{X = x\}$** ".

Remarque :

À noter que, d'une manière générale, quand on écrit $\frac{dP_X}{d\mu_1}(x)$, on parlera toujours de la **densité de P par rapport à μ** , il s'agit donc d'un abus de notation en soit mais très intuitif. Par exemple, si on a \mathcal{P} qui admet une densité f par rapport à une mesure μ , on a alors que

$$f = \frac{d\mathcal{P}}{d\mu}$$

Preuve :

Soient $A \in \mathcal{E}$ et $B \in \mathcal{F}$

$$\begin{aligned} P(\{X \in A\} \cap \{Y \in B\}) &= \int_{A \times B} dP_{(X,Y)}(x, y) \\ &= \int_{A \times B} \frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) d\mu_1 \otimes \mu_2(x, y) \\ &= \int_A \left(\int_B \frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) d\mu_2(y) \right) d\mu_1(x) \end{aligned}$$

Soit $\tilde{A} = \{x \in A \mid \frac{dP_X}{d\mu_1}(x) \neq 0\} \in \mathcal{E}$, alors $A = \tilde{A} \cup \tilde{A}^c$, $\forall x \in \tilde{A}^c$, $0 = \frac{dP_X}{d\mu_1}(x) = \int_F \frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) dy$

Mais comme $\frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) \geq 0$, on a que $\forall x \in \tilde{A}^c$, $\frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) = 0$ pour μ_2 - presque tout $y \in F$

D'où :

$$\begin{aligned}
P(\{X \in A\} \cap \{Y \in B\}) &= \int_{\tilde{A}} \left(\int_B \frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) d\mu_2(y) \right) d\mu_1(x) \\
&= \int_{\tilde{A}} \left(\int_B \underbrace{\frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y)}_{=f_{X,Y}} / \underbrace{\frac{dP_X}{d\mu_1}(x)}_{=f_X} d\mu_2(y) \right) dP_X(x)
\end{aligned}$$

Soit

$$\begin{aligned}
\partial : E \times \mathcal{F} &\mapsto \mathbb{R} \\
(x, B) &\mapsto \begin{cases} \int_B \frac{f_{(X,Y)}(x,y)}{f_X(x)} d\mu_2(y), & \text{si } x \in \tilde{A} \\ P_Y(B), & \text{sinon} \end{cases}
\end{aligned}$$

∂ est alors un noyau de transition de (E, \mathcal{E}) dans (F, \mathcal{F}) puisque

- $\forall B \in \mathcal{F}, \quad x \mapsto \partial(x, B)$ est \mathcal{E} -mesurable
- $\forall x \in E, \quad B \mapsto \partial(x, B)$ est une mesure de probabilité

En **conclusion**, $\forall A \in \mathcal{E}, B \in \mathcal{F}$,

$$P(\{X \in A\} \cap \{Y \in B\}) = \int_A \partial(x, B) dP_X(x)$$

Donc $\partial(x, \cdot)$ est une loi conditionnelle de Y sachant $X = x$, et si $\frac{dP_X}{d\mu_1}(x) \neq 0$, $\partial(x, \cdot)$ admet une densité donnée par

$$P_{Y|X=x}(B) = \int_B \frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y) / \frac{dP_X}{d\mu_1}(x) d\mu_2(y)$$

Remarque :

Si $\frac{dP_X}{d\mu_1}(x) \neq 0$, alors,

$$\forall y \in F, f_{Y|X=x}(y) = \frac{\frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y)}{\frac{dP_X}{d\mu_1}(x)} = C \frac{dP_{(X,Y)}}{d\mu_1 \otimes \mu_2}(x, y)$$

exemple

Si X et Y sont deux variables aléatoires sur (Ω, \mathcal{A}, P) de loi jointe continue et de densité

$$\begin{aligned} f : \quad \mathbb{R}^2 &\mapsto \mathbb{R} \\ (x, y) &\mapsto 2e^{-(x+y)} \mathbb{1}_{0 \leq x \leq y} \end{aligned}$$

$$\forall x \geq 0, \quad \frac{dP_X}{d\lambda_1}(x) = \int_{\mathbb{R}} 2e^{-(x+y)} \mathbb{1}_{0 \leq x \leq y} dy > 0 \text{ et}$$

$$\begin{aligned} f_{Y|X=x}(y) &= C \frac{dP_{(X,Y)}}{d\lambda_1 \otimes \lambda_2}(x, y) \\ &= ce^{-(x+y)} \mathbb{1}_{x \leq y} \\ &= (ce^{-x})e^{-y} \mathbb{1}_{x \leq y} \\ &= c'e^{-y} \mathbb{1}_{x \leq y} \end{aligned}$$

Or $\int_{\mathbb{R}} f_{Y|X=x}(y) dy = 1$ si et seulement si

$$c' \int_x^{\infty} e^{-y} dy = 1$$

Si, et seulement si $c' = e^x$, si et seulement si $c = e^{2x}$. Ainsi, pour tout $x \geq 0$,

$$f_{Y|X=x}(y) = e^{-(y-x)} \mathbb{1}_{x \leq y}(y)$$

12.3 Cas Discret

Proposition 12|12.3|1

Soient E et F des ensembles au plus dénombrables, munis de leurs tribus discrète, et X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) et (F, \mathcal{F}) telle que pour tout $x \in E$, $P(X = x) \neq 0$.

Alors, $\forall x \in E$, $P_{Y|x=x}$ est une **loi discrète** sur F donnée par :

$$P_{Y|x=x}(\{y\}) = P_{[X=x]}(\{Y = y\})$$

Preuve :

La densité conditionnelle de Y sachant $X = x$ est donnée par

$$\begin{aligned} f_{Y|X=x} : F &\mapsto \mathbb{R} \\ y &\mapsto \frac{P(\{X=x\} \cap \{Y=y\})}{P(\{X=x\})} = P_{[X=x]}(Y = y) \end{aligned}$$

12.4 Caractérisation de l'Indépendance via les Lois Gaussiennes

Proposition 12|12.4|1

Soient X, Y deux variables aléatoires sur (Ω, \mathcal{A}, P) dans des espaces mesurables (E, \mathcal{E}) et (F, \mathcal{F}) .
Alors

$$X \perp\!\!\!\perp Y \iff \forall x \in E, P_Y \text{ est une loi conditionnelle de } Y \text{ sachant } X = x$$

Preuve :

" \implies " : Supposons que $X \perp\!\!\!\perp Y$, $\forall A \in \mathcal{E}, B \in \mathcal{F}$,

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) = \int_A P(\{Y \in B\})dP_X(x)$$

Donc

$$\begin{aligned} \partial : E \times F &\mapsto \mathbb{R} \\ (x, B) &\mapsto P(Y \in B) \end{aligned}$$

est un noyau de transition qui donne une loi conditionnelle de Y sachant X .

$\forall x \in E$, $\partial(x, B) = P_Y(\cdot)$ est une loi conditionnelle de Y sachant $\{X = x\}$.

" \impliedby " : $\forall A \in \mathcal{E}, B \in \mathcal{F}$,

$$\begin{aligned} P(\{X \in A\} \cap \{Y \in B\}) &= P((X, Y) \in A \times B) \\ &= \int_A P_Y(B)dP_X(x) \\ &= \left(\int_A dP_X(x) \right) P_Y(B) \\ &= P_X(A)P_Y(B) \end{aligned}$$

Ainsi, $X \perp\!\!\!\perp Y$

Proposition 12|12.4|2

S'il existe une loi Q sur (F, \mathcal{F}) telle que $\forall x \in E$, Q est une loi conditionnelle de Y sachant $X = x$, alors $Q = P_Y$, et $X \perp\!\!\!\perp Y$.

Preuve :

$\forall B \in \mathcal{F}$,

$$\begin{aligned} P_Y(B) &= P(\{X \in E\} \cap \{Y \in B\}) \\ &= \int_E Q(B) dP_X(x) \\ &= Q(B) \end{aligned}$$

12.5 Théorème de Fubini Généralisé

Théorème 12|12.5|1

Soient (E, \mathcal{E}) , (F, \mathcal{F}) deux espaces mesurables, et P_1 une probabilité sur (E, \mathcal{E}) . Soit ∂ un **noyau de transition de (E, \mathcal{E}) vers (F, \mathcal{F})** .

Alors notons $P_1.\partial$ la loi alors définie sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$:

i) $\forall f : E \times F \mapsto \mathbb{R}_+$ mesurable, alors

$$\int_{E \times F} f(x, y) d(P_1.\partial)(x, y) = \int_E \underbrace{\left(\int_F f(x, y) \partial(x, dy) \right)}_{\mathbb{P}_{Y|X=x}(y)} dP_1(x)$$

ii) Soit $f : E \times F \mapsto \mathbb{R}$ mesurable, alors $f \in L^1(E \times F, P_1.\partial)$ si, et seulement si

$$\begin{cases} \forall P_1, \quad \forall x \in E, \quad f(x, \cdot) \in L^1(F, \partial(x, \cdot)) \\ E \mapsto \mathbb{R} \\ x \mapsto \int_F f(x, y) \partial(x, dy) \in L^1(E, P_1) \end{cases}$$

et i) est alors vérifié

12.6 Liens avec l’Espérance Conditionnelle

12.6.1 Théorème de Transfert Conditionnel

Théorème 12|12.6|1

Théorème de transfert conditionnel

Soient X et Y deux variables aléatoires sur des espaces mesurables (E, \mathcal{E}) et (F, \mathcal{F}) , et soit $h : E \times F \mapsto \mathbb{R}$ une fonction mesurable telle que : $h \geq 0$ ou $h(X, Y) \in L^1(P)$ i.e. : $h \in L^1(P(X, Y))$

Soit

$$\begin{aligned}\varphi : E &\mapsto \mathbb{R} \\ x &\mapsto \int_F h(x, y) dP_{Y|X=x}(y)\end{aligned}$$

$$\mathbb{E}[h(X, Y)|X] = \varphi(X)$$

Preuve :

Cas où $h \geq 0$. Soit $B \in \sigma(X)$. Montrons que $\mathbb{E}[h(X, Y)\mathbb{1}_B] = \mathbb{E}[\varphi(X)\mathbb{1}_B]$

B s’écrit $B = X^{-1}(A)$, $A \in \mathcal{E}$

$$\begin{aligned}\mathbb{E}[h(X, Y)\mathbb{1}_B] &= \mathbb{E}[h(X, Y)\mathbb{1}_{\{X \in A\}}] \\ &= \int_{E \times F} h(x, y)\mathbb{1}_A(x) dP_{(X, Y)}(x, y) \text{ (théorème de transfert)} \\ &= \int_E \left(\int_F h(x, y)\mathbb{1}_A(x) dP_{Y|X=x}(y) \right) dP_X(x) \text{ (Fubini généralisé cas positif)} \\ &= \int_E \mathbb{1}_A(x) \left(\underbrace{\int_F h(x, y) dP_{Y|X=x}(y)}_{\varphi(x)} \right) dP_X(x) \\ &= \int_E \mathbb{1}_A(x) \varphi(x) dP_X(x) \\ &= \mathbb{E}[\varphi(X)\mathbb{1}_{\{X \in A\}}] \text{ (théorème de transfert)} \\ &= \mathbb{E}[\varphi(X)\mathbb{1}_B]\end{aligned}$$

$\varphi(X)$ étant bien $\sigma(X)$ – mesurable, on a bien $\mathbb{E}[h(X, Y)|X] = \varphi(X)$

Exemple :

1) Soit X une variable aléatoire réelle, et $\epsilon \sim \mathcal{N}(0, 1)$. $Y = X + \epsilon$ tel que $X \perp\!\!\!\perp \epsilon$ qu'en est-il de $\mathbb{E}[Y^2|\sigma(X)]$?

D'après le théorème de transfert conditionnel, $\mathbb{E}[Y^2|\sigma(X)] = \varphi(X)$ avec

$$\begin{aligned}\varphi : \mathbb{R} &\mapsto \mathbb{R} \\ x &\mapsto \int_{\mathbb{R}} (x + y)^2 dP_{\epsilon|X=x}(y)\end{aligned}$$

or, $P_{\epsilon|X=x} = P_{\epsilon}$ car $X \perp\!\!\!\perp \epsilon$

2) Si X est une variable aléatoire dans \mathbb{N} , intégrable, $\epsilon \sim \mathcal{N}(0, 1)$ tel que $X \perp\!\!\!\perp \epsilon$.

Soit $Y = X + \epsilon$, alors, $\forall n \in \mathbb{N}$, $P_{Y|X=x} = \mathcal{N}(x, 1)$, donc, d'après le théorème de transfert conditionnel, $\mathbb{E}[Y|X] = \varphi(X)$ où, $\forall x \in \mathbb{N}$,

$$\begin{aligned}\varphi(x) &= \int_{\mathbb{R}} y dP_{Y|X=x}(y) \\ &= \int_{\mathbb{R}} y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y-x)^2}{2}} dy \\ &= x\end{aligned}$$

Ainsi, $\mathbb{E}[Y|X] = X$

Autre manière : $\mathbb{E}[Y|X] = \mathbb{E}[X|X] + \mathbb{E}[\epsilon|X] = X + 0$ par indépendance de X et ϵ
 $\forall x \in \mathbb{R}$,

$$\begin{aligned}\varphi(x) &= \int_{\mathbb{R}} (x + y)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y^2} dy \\ &= x^2 + 2x\mathbb{E}[\epsilon] + \mathbb{E}[\epsilon^2] \\ &= x^2 + 1\end{aligned}$$

Ainsi, $\mathbb{E}[Y^2|\sigma(X)] = X^2 + 1$

12.6.2 Outils pour le Calcul des Loix Conditionnelles

Dans la suite, de cette partie, $(E, \mathcal{E}), (F, \mathcal{F}), (G, \mathcal{G})$ sont des espaces mesurables, X et Y sont des variables aléatoires sur (Ω, \mathcal{A}, P) dans (E, \mathcal{E}) et (F, \mathcal{F}) .

On notera, si Q est une mesure de probabilité sur (E, \mathcal{E}) , et $g : E \mapsto F$ est mesurable, $g\#Q$ la probabilité sur (F, \mathcal{F}) donnée par : $\forall B \in \mathcal{F}, (g\#Q)(B) = Q(g^{-1}(B))$ **la mesure image** de Q par g

donc $g\#P_X = P_{g(X)} \dots$

Proposition 12|12.6|1

Si $g : F \mapsto G$ est mesurable, alors : $\forall x \in E, P_{g(Y)|X=x} = g\#P_{Y|X=x}$ i.e. : si $Z \sim P_{Y|X=x}$, alors $g(Z) \sim P_{g(Y)|X=x}$

Preuve :

$\forall A \in \mathcal{E}, \forall C \in \mathcal{G},$

$$P(\{X \in A\} \cap \{g(Y) \in C\}) = P(\{X \in A\} \cap \{Y \in g^{-1}(C)\}) = \int_A \underbrace{P_{Y|X=x}(g^{-1}(C))}_{=g\#P_{Y|X=x}(C)} dP_X(x)$$

$\forall x \in E, g\#P_{Y|X=x}$ est la loi conditionnelle de $g(Y)$ sachant $X = x$, on vérifie en effet que

$$\begin{aligned} E \times G &\mapsto \mathbb{R} \\ (x, c) &\mapsto (g\#P_{Y|X=x})(C) \end{aligned}$$

est un noyau de transition

Proposition 12|12.6|2

Soit $h : E \times F \mapsto G$ mesurable. Alors

$$\forall x \in E, \quad P_{h(X,Y)|X=x} = P_{h(x,Y)|X=x}$$

Exemple :

Si on prend $Y = X + \epsilon$, alors, pour $x \in \mathbb{R}$, $P_{Y|X=x} = P_{x+\epsilon|X=x} = \mathcal{N}(x, 1)$

13 Inégalités

On considèrera dans cette section (sauf indications contraires) que toutes les variables aléatoires introduites sont réelles et sur un espace probabilisé (Ω, \mathcal{A}, P) .

Inégalité de Markov

Soit X une variable aléatoire positive admettant un moment d'ordre 1, alors :

$\forall \epsilon > 0$

$$P(X > \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}$$

Preuve :

Soit $t \in \mathbb{R}_+^*$, on a

$$\begin{aligned} X \geq t \mathbb{1}_{X \geq t} &\implies \mathbb{E}[X] \geq t \mathbb{E}[\mathbb{1}_{X \geq t}] \\ &\iff \mathbb{E}[X] \geq t P(X \geq t) \\ &\iff P(X \geq t) \leq \frac{\mathbb{E}[X]}{t} \end{aligned}$$

Et ainsi on a bien $P(X > \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}, \forall \epsilon > 0$

Inégalité de Paley-Zygmund

Soit X une variable aléatoire telle que $X \stackrel{p.s.}{\geq} 0$, et $X \in L^2(P)$. Alors, pour tout $\theta \in [0, 1]$ on a

$$P(X \geq \theta \mathbb{E}[X]) > (1 - \theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$$

Preuve :

$$\mathbb{E}[X] = \mathbb{E} \left[\underbrace{X \mathbb{1}_{X \leq \mathbb{E}[X]\theta}}_{\leq \mathbb{E}[X]\theta} \right] + \underbrace{\mathbb{E}[X^2]^{1/2} P(X > \mathbb{E}[X]\theta)^{1/2}}_{\stackrel{(C.S.)}{\leq} \mathbb{E}[X^2]^{1/2} P(X > \mathbb{E}[X]\theta)^{1/2}}$$

Donc

$$\begin{aligned} \mathbb{E}[X] &\leq \mathbb{E}[X] \theta + \mathbb{E}[X^2]^{1/2} P(X > \mathbb{E}[X] \theta)^{1/2} \\ &\iff P(X > \theta \mathbb{E}[X]) \geq (1 - \theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \end{aligned}$$

Inégalité de Hölder

Soit $X \in L^p(P)$ et $Y \in L^q(P)$, avec $p, q \geq 1, p, q \in \mathbb{R}$ deux variables aléatoires, tel que $\frac{1}{p} + \frac{1}{q} = 1$ Alors on a

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}$$

Preuve :

On a $p, q > 1$ car $\frac{1}{p} + \frac{1}{q} = 1$. (si on avait $p \leq 1 \implies \frac{1}{p} \geq 1, \forall q > 0, \frac{1}{p} + \frac{1}{q} > 1 \neq 1$). En rappelant que pour tout réels a, b positifs, et pour tout $\lambda > 0$, on a, par concavité du log :

$$\begin{aligned} \log\left(\frac{1}{p}(\lambda a)^p + \frac{1}{q}\left(\frac{b}{\lambda}\right)^q\right) &= \log\left(\frac{1}{p}(\lambda a)^p + \left(1 - \frac{1}{p}\right)\left(\frac{b}{\lambda}\right)^q\right) \\ &\geq \frac{1}{p} \log((\lambda a)^p) + \frac{1}{q} \log\left(\left(\frac{b}{\lambda}\right)^q\right) = \ln(ab) \end{aligned}$$

Par croissance de l'exponentielle, on a donc

$$ab \leq \frac{1}{p}(\lambda a)^p + \frac{1}{q}\left(\frac{b}{\lambda}\right)^q$$

Donc $|XY|$ positif, ainsi $|XY| \leq \frac{1}{p}(\lambda |X|)^p + \frac{1}{q}\left(\frac{|Y|}{\lambda}\right)^q, \forall \lambda > 0$ donc

$\mathbb{E}[|XY|] \leq \frac{\lambda^p}{p} \mathbb{E}[|X|^p] + \frac{1}{\lambda^q q} \mathbb{E}[|Y|^q]$ or, $X \in L^p(P)$ et $Y \in L^q(P)$ donc $XY \in L^1(P)$

Minimisons la fonction $f(\lambda) = \frac{\lambda^p \mathbb{E}[|X|^p]}{p} + \frac{\mathbb{E}[|Y|^q]}{\lambda^q q}$

$$f'(\lambda) = 0 \iff 0 = \lambda^{p-1} \mathbb{E}[|X|^p] - \lambda^{-q-1} \mathbb{E}[|Y|^q]$$

$$f'(\lambda) = 0 \iff \lambda = \left(\frac{\mathbb{E}[|Y|^q]}{\mathbb{E}[|X|^p]} \right)^{\frac{1}{p+q}}$$

Et donc

$$\begin{aligned}
\mathbb{E}[|XY|] &\leq \frac{1}{p} \left(\frac{\mathbb{E}[|Y|^q]^{\frac{1}{p+q}}}{\mathbb{E}[|X|^p]^{\frac{1}{p+q}}} \right)^p \mathbb{E}[|X|^p] + \frac{1}{q} \left(\frac{\mathbb{E}[|X|^p]^{\frac{1}{p+q}}}{\mathbb{E}[|Y|^q]^{\frac{1}{p+q}}} \right)^q \mathbb{E}[|Y|^q] \\
&\leq \frac{1}{p} \left(\frac{\mathbb{E}[|Y|^q]^{\frac{p}{p+q}}}{\mathbb{E}[|X|^p]^{\frac{p}{p+q}-1}} \right) + \frac{1}{q} \left(\frac{\mathbb{E}[|X|^p]^{\frac{q}{p+q}}}{\mathbb{E}[|Y|^q]^{\frac{q}{p+q}-1}} \right) \\
&\leq \frac{1}{p} \left(\frac{\mathbb{E}[|X|^p]^{1-\frac{p}{p+q}}}{\mathbb{E}[|Y|^q]^{\frac{-p}{p+q}}} \right) + \frac{1}{q} \left(\frac{\mathbb{E}[|X|^p]^{\frac{q}{p+q}}}{\mathbb{E}[|Y|^q]^{\frac{-p}{p+q}}} \right) \\
&\leq \mathbb{E}[|X|^p]^{1-\frac{p}{p+q}} \mathbb{E}[|Y|^q]^{\frac{p}{p+q}} \\
&\leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|Y|^q]^{\frac{1}{q}} \quad \text{car } \frac{1}{p} + \frac{1}{q} = 1 \iff p+q=pq
\end{aligned}$$

Inégalité de Minkovski

Pour $X, Y \in L^p(P)$, $p \geq 1$ quelconque, on a

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

Preuve :

Si $X, Y \in L^p(P)$, alors $X+Y \in L^p(P)$ car $\mathbb{E}[|X+Y|^p] \stackrel{\text{I.T.}}{\leq} \mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p] < \infty$

On a aussi que $|X||X+Y|^{p-1}$ et $|Y||X+Y|^{p-1} \in L^1(P)$ car on a par Hölder :

$$\mathbb{E}[|X||X+Y|^{p-1}] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|X+Y|^p]^{\frac{p-1}{p}} < \infty$$

et

$$\mathbb{E}[|Y||X+Y|^{p-1}] \leq \mathbb{E}[|Y|^p]^{\frac{1}{p}} \mathbb{E}[|X+Y|^p]^{\frac{p-1}{p}} < \infty$$

(car $\frac{1}{p} + \frac{p-1}{p} = 1$)

et on peut ainsi dire que

$$\mathbb{E}[|X+Y|^p] = \mathbb{E}[|X+Y||X+Y|^{p-1}] \stackrel{\text{I.T.}}{\leq} \mathbb{E}[|X||X+Y|^{p-1}] + \mathbb{E}[|Y||X+Y|^{p-1}]$$

Donc

$$\mathbb{E}[|X+Y|^p] \leq \mathbb{E}[|X||X+Y|^{p-1}] + \mathbb{E}[|Y||X+Y|^{p-1}]$$

Par Hölder, avec $\frac{1}{p} + \frac{p-1}{p} = 1$

on a

$$\begin{aligned}
\mathbb{E}[|X + Y|^p] &\leq \mathbb{E}[|X||X + Y|^{p-1}] + \mathbb{E}[|Y||X + Y|^{p-1}] \\
&\leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|X + Y|^p]^{\frac{p-1}{p}} + \mathbb{E}[|Y|^p]^{\frac{1}{p}} \mathbb{E}[|X + Y|^p]^{\frac{p-1}{p}} \\
&\leq \mathbb{E}[|X + Y|^p]^{\frac{p-1}{p}} (\mathbb{E}[|X|^p]^{\frac{1}{p}} + \mathbb{E}[|Y|^p]^{\frac{1}{p}})
\end{aligned}$$

Donc

$$\mathbb{E}[|X + Y|^p] \leq \mathbb{E}[|X + Y|^p]^{\frac{p-1}{p}} (\mathbb{E}[|X|^p]^{\frac{1}{p}} + \mathbb{E}[|Y|^p]^{\frac{1}{p}})$$

équivalent à

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

équivalent à

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

Inégalité de Jensen

Soit $X \in L^1(P)$ une variable aléatoire, et f une fonction convexe tel que $f \circ X \in L^1(P)$, alors

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

Preuve :

f est une fonction convexe, faisons la preuve dans le cas où f est différentiable. $\forall x, y$, on a

$$f(x) \geq f(y) + \nabla f(y)^T(x - y)$$

donc pour $y = \mathbb{E}[X]$, et $x = X$, on a

$$f(X) \geq f(\mathbb{E}[X]) + \nabla f(\mathbb{E}[X])(X - \mathbb{E}[X])$$

ainsi,

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(\mathbb{E}[X])] + \nabla f(\mathbb{E}[X])(\mathbb{E}[X] - \mathbb{E}[X]) = \mathbb{E}[f(\mathbb{E}[X])] = f(\mathbb{E}[X])$$

donc on a bien :

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(X)]$$

Proposition 13|13.0|1

De même, pour les espérances conditionnelles, on a que si $X \in L^1(P)$ une variable aléatoire, et f une fonction convexe tel que $f \circ X \in L^1(P)$, alors pour une tribu $\mathcal{B} \subset \mathcal{A}$ on a

$$\mathbb{E}[f(X)|\mathcal{B}] \leq f(\mathbb{E}[X|\mathcal{B}])$$

Preuve :

Identiquement, on fera la preuve dans le cas où f différentiable. On a :

$$f(X) \geq f(\mathbb{E}[X|\mathcal{B}]) + \nabla f(\mathbb{E}[X|\mathcal{B}]) (X - \mathbb{E}[X|\mathcal{B}])$$

$$\iff \mathbb{E}[f(X)|\mathcal{B}] \geq \mathbb{E}[f(\mathbb{E}[X|\mathcal{B}])|\mathcal{B}] + 0$$

$$\iff \mathbb{E}[f(X)|\mathcal{B}] \geq f(\mathbb{E}[X|\mathcal{B}])$$

Inégalité de Kolmogorov

Soit $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires indépendantes, dans $L^2(P)$.

Alors, $\forall x > 0, \forall n \geq 1$, (n peut valoir $+\infty$)

$$P\left(\sup_{N \in \llbracket 1, n \rrbracket} \left\{ \left| \sum_{j=1}^N X_j \right| \right\} > x\right) \leq \frac{\sum_{i=1}^n \mathbb{E}[X_i^2]}{x^2}$$

13.1 Inégalités de Concentration

Dans la théorie des probabilités, les inégalités de concentration fournissent des bornes sur la probabilité qu'une variable aléatoire dévie d'une certaine valeur (généralement l'espérance de cette variable aléatoire).

13.1.1 Variables Aléatoires bornées

Inégalité de Tchebychev

Soit X une variable aléatoire admettant un moment d'ordre deux.

Alors $\forall \epsilon > 0$

$$P(|X - \mathbb{E}[X]| > \epsilon) \leq \frac{\mathbb{V}(X)}{\epsilon^2}$$

On prend $Y = (X - \mathbb{E}[X])^2$ Alors selon l'inégalité de Markov, comme Y est positive, on a :

$$\forall \epsilon^2 (\epsilon > 0) P(Y > \epsilon^2) \leq \frac{\mathbb{E}[Y]}{\epsilon^2} \iff P((X - \mathbb{E}[X])^2 > \epsilon^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\epsilon^2}$$

$$\text{donc } P(|X - \mathbb{E}[X]| > \epsilon) \leq \frac{\mathbb{V}(X)}{\epsilon^2}$$

Pour démontrer les inégalités qui suivent, on utilisera la **Méthode de Chernoff**. L'idée est d'appliquer Markov à la variable aléatoire positive $e^{\lambda X}$, avec $\lambda \in \mathbb{R}$. Démontrons d'abord un lemme, avant d'utiliser la méthode.

Lemme d'Hoeffding

Soit Y une variable aléatoire réelle bornée, centrée, telle qu'il existe $a, b \in \mathbb{R}, a < b$ tel que $Y \in [a, b]$ p.s.. Alors, pour tout $t > 0$, on a

$$\mathbb{E}[e^{tY}] \leq e^{(a-b)^2 \frac{t^2}{8}}$$

Preuve :

On a $x \mapsto e^{tx}$ qui est une fonction convexe, donc, come $Y \in [a, b]$ p.s., on a, presque surement :

$$\begin{aligned} e^{tY} &\leq \frac{b-Y}{b-a} e^{ta} + \left(1 - \frac{b-Y}{b-a}\right) e^{tb} \\ &\leq \frac{b-Y}{b-a} e^{ta} + \frac{a-Y}{b-a} e^{tb} \end{aligned}$$

Et donc

$$\mathbb{E}[e^{tY}] \leq \frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb}$$

On pose pour $u = t(b-a)$

$$\begin{aligned}
\psi(u) &= \log\left(\frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb}\right) \\
&= \log\left(\frac{b}{b-a}e^{\frac{ua}{b-a}} - \frac{a}{b-a}e^{\frac{ub}{b-a}}\right) \\
&= \log\left(e^{\frac{ua}{b-a}}\left(\frac{b}{b-a} - \frac{a}{b-a}e^{\frac{u(b-a)}{b-a}}\right)\right) \\
&= \frac{a}{b-a}u + \log\left(1 + \frac{a}{b-a}(1 - e^u)\right)
\end{aligned}$$

on a $\psi(0) = \psi'(0) = 0$ et

$$\psi''(u) = \left(\frac{a}{b-a} - \frac{ae^u}{b-ae^u}\right)' = \frac{-bae^u}{(b-ae^u)^2} \quad \underbrace{\text{car de la forme } \frac{\alpha\beta}{(\alpha+\beta)^2}}_{\leq} \quad \frac{1}{4}$$

En effet, par inégalité arithmético-géométrique, on a, pour tout $x, y \in \mathbb{R}$ de même signe (si ils ne sont pas de même signes, alors on a directement $\frac{xy}{(x+y)^2} \leq 0 \leq \frac{1}{4}$) :

$$\frac{|x| + |y|}{2} \leq \sqrt{|xy|} \iff \frac{(|x| + |y|)^2}{4} \leq |xy| \iff \frac{(|x| + |y|)^2}{4} \leq xy \iff \frac{xy}{(x+y)^2} \leq \frac{1}{4}$$

Et donc, par Taylor-Lagrange, on a qu'il existe un réel $s \in [0, u]$ tel que

$$\psi(u) = \psi(0) + u\psi'(0) + \frac{1}{2}u^2\psi''(s) \leq \frac{(a-b)^2}{8}t^2$$

Donc on a

$$\mathbb{E}[e^{tY}] \leq e^{\frac{(a-b)^2}{8}t^2}$$

Inégalité de Hoeffding

Soit $X_1, \dots, X_n \in L^1(P)$, des variables aléatoires réelles indépendantes, et telles que $\forall i \in \llbracket 1, n \rrbracket, a_i \leq X_i \leq b_i$ p.s. Alors $\forall t > 0$,

$$P\left(\left|\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

et ainsi

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

Preuve :

On pose $Y_i = X_i - \mathbb{E}[X_i]$, les Y_i sont donc centrées, et sont entre $a_i - \mathbb{E}[X_i]$ et $b_i - \mathbb{E}[X_i]$ presque sûrement.

On pose $S_n = \sum_{i=1}^n Y_i$, on a pour tout $x \geq 0, u > 0$, par Markov :

$$\begin{aligned} P(S_n - \mathbb{E}[S_n] \geq x) &= P(e^{u(S_n - \mathbb{E}[S_n])} \geq e^{ux}) \\ &\leq \frac{\mathbb{E}[e^{u(S_n - \mathbb{E}[S_n])}]}{e^{ux}} \quad (\text{Markov}) \\ &\leq e^{-ux} \prod_{i=1}^n \mathbb{E}[e^{uY_i}] \end{aligned}$$

On obtient par le lemme précédent, car $b_i - a_i = b_i - \mathbb{E}[X_i] - (a_i - \mathbb{E}[X_i])$

$$P(S_n - \mathbb{E}[S_n] \geq x) \leq e^{-ux + \frac{u^2}{8} \sum_{i=1}^n (b_i - a_i)^2}$$

Sachant que c'est pour tout $u > 0$, on optimise et on trouve $u = \frac{4x}{\sum_{i=1}^n (b_i - a_i)^2}$

Ce qui montre que

$$\begin{aligned} P(S_n - \mathbb{E}[S_n] \geq x) &= P(e^{u(S_n - \mathbb{E}[S_n])}) \leq e^{-\frac{4x}{\sum_{i=1}^n (b_i - a_i)^2} x + \frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}} \\ &= e^{-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}} \end{aligned}$$

En remplaçant X_i par $-X_i$ on obtient aussi

$$P(S_n - \mathbb{E}[S_n] \leq -x) \leq e^{-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

et donc $\forall x > 0$, on a

$$P\left(\left|\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right| \geq x\right) \leq 2\exp\left(-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Inégalité de Bennett

Soit X_1, \dots, X_n des variables aléatoires réelles centrées indépendantes, tel que pour tout $i \in \llbracket 1, n \rrbracket$ il existe $c \in \mathbb{R}_+^*$ tel que $X_i \stackrel{p.s.}{\leq} c$.

En posant $S_n = \sum_{i=1}^n X_i$ et $V_n = \sum_{i=1}^n \mathbb{E}[X_i^2] = \sum_{i=1}^n \mathbb{V}(X_i)$ on a alors, pour tout $\lambda \geq 0$:

$$\log(\mathbb{E}[e^{\lambda S_n}]) \leq \frac{V_n}{c^2} \phi(c\lambda), \quad \phi(u) = e^u - u - 1$$

Et donc, pour tout $t \geq 0$ on a :

$$P(S_n \geq t) \leq e^{-\frac{V_n}{c^2} h(\frac{ct}{V_n})}$$

Avec $h(x) = (1+x)\log(1+x) - x$

Également, si on a en également, $|X_i| \stackrel{p.s.}{\leq} c$, alors on en déduit alors :

$$P(|S_n| \geq t) \leq 2e^{-\frac{V_n}{c^2}h(\frac{ct}{V_n})}$$

Preuve :

Supposons d'abord que $c = 1$.

Soit $\lambda \geq 0$, on peut déjà remarquer que, comme $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ croissante et que $X_i \overset{p.s.}{\leq} 1$

$$e^{\lambda X_i} = 1 + \lambda X_i + \phi(\lambda X_i) \leq 1 + \lambda X_i + X_i^2 \phi(\lambda) \quad (\phi(\lambda X_i) \leq \phi(\lambda) \stackrel{X_i \leq 1}{\leq} X_i^2 \phi(\lambda))$$

Donc

$$\begin{aligned} \log(\mathbb{E}[e^{\lambda S_n}]) &= \sum_{i=1}^n \log(\mathbb{E}[e^{\lambda X_i}]) \\ &= \sum_{i=1}^n \log(\mathbb{E}[e^{\lambda X_i}]) \\ &\leq \sum_{i=1}^n \log(\mathbb{E}[1 + \lambda X_i + X_i^2 \phi(\lambda)]) \\ (\text{car } \log(1+x) \leq x) &\leq \sum_{i=1}^n \mathbb{E}[\lambda X_i + X_i^2 \phi(\lambda)] \\ &\leq \sum_{i=1}^n \mathbb{E}[X_i^2] \phi(\lambda) = V_n \phi(\lambda) \end{aligned}$$

Pour $X_i \stackrel{p.s.}{\leq} c$ en général, il suffit d'observer que $\frac{X_i}{c} \stackrel{p.s.}{\leq} 1$, on a directement

$$e^{\lambda X_i} = 1 + \lambda X_i + \overbrace{\phi(\lambda X_i)}^{=\phi(\lambda c(\frac{X_i}{c}))} \leq 1 + \lambda X_i + \left(\frac{X_i}{c}\right)^2 \phi(\lambda c)$$

et donc que

$$\log(\mathbb{E}[e^{\lambda S_n}]) \leq \sum_{i=1}^n \mathbb{E}\left[\left(\frac{X_i}{c}\right)^2\right] \phi(\lambda) = \frac{V_n}{c^2} \phi(\lambda c)$$

Maintenant, montrons la deuxième inégalité. Soit $\lambda \geq 0, t \geq 0$ alors :

$$\begin{aligned} P(S_n \geq t) &= P(e^{\lambda S_n} \geq e^{\lambda t}) \\ (\text{Markov}) &\leq \frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda t}} \\ \log\left(\frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda t}}\right) &\leq \frac{V_n}{c^2} \phi(\lambda c) - \lambda t \end{aligned}$$

(Minimiser $\frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda t}}$ revient à minimiser $\log\left(\frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda t}}\right)$ par croissance du log).

Minimisons donc la fonction

$$\begin{aligned} \psi : \mathbb{R}_+ &\mapsto \mathbb{R} \\ \lambda &\mapsto \frac{V_n}{c^2} (e^{\lambda c} - \lambda c - 1) - \lambda t \end{aligned}$$

or $\psi''(\lambda) = V_n e^{\lambda c} > 0$ donc ψ est une fonction convexe. De plus,

$$\begin{aligned} \psi'(\lambda) &= 0 \\ \iff -t + \frac{V_n}{c^2} (ce^{\lambda c} - c) &= 0 \\ \iff \lambda &= \frac{1}{c} \log\left(1 + \frac{tc}{V_n}\right) \end{aligned}$$

Donc, on a, en évaluant en ce λ :

$$\begin{aligned}
\frac{V_n}{c^2} \phi(\lambda c) - \lambda t &= \frac{V_n}{c^2} (e^{\lambda c} - \lambda c - 1) - \lambda t \\
&= \frac{V_n}{c^2} \left(e^{\frac{1}{c} \log(1 + \frac{tc}{V_n}) c} - \frac{1}{c} \log(1 + \frac{tc}{V_n}) c - 1 \right) - \frac{1}{c} \log(1 + \frac{tc}{V_n}) t \\
&= \frac{V_n}{c^2} \left(1 + \frac{tc}{V_n} - \log(1 + \frac{tc}{V_n}) - 1 \right) + \frac{t}{c} \log(1 + \frac{tc}{V_n}) \\
&= \frac{1}{c^2} \left[\log(1 + \frac{tc}{V_n}) (-V_n + tc) + tc \right] \\
&= \frac{-V_n}{c^2} \left[\log(1 + \frac{tc}{V_n}) \left(1 - \frac{tc}{V_n} \right) - \frac{tc}{V_n} \right] \\
&= \frac{-V_n}{c^2} h\left(\frac{tc}{V_n}\right) \quad \text{avec } h(x) = \log(1+x)(1+x) - x
\end{aligned}$$

On a donc bien que

$$P(S_n \geq t) \leq \frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda t}} \leq e^{\frac{-V_n}{c^2} h(\frac{tc}{V_n})}$$

On peut l'appliquer aussi à $-X_i$ si $|X_i| \stackrel{p.s.}{\leq} c$, alors on a dans ce cas :

$$P(-S_n \geq t) \leq e^{\frac{-V_n}{c^2} h(\frac{tc}{V_n})}$$

et donc

$$P(|S_n| \geq t) = P(S_n \geq t \cup -S_n \geq t) = P(S_n \geq t) + P(-S_n \geq t) \leq 2e^{\frac{-V_n}{c^2} h(\frac{tc}{V_n})}$$

Inégalité de Bernstein

① Soit $X_1, \dots, X_n \in L^\infty(P)$, des variables aléatoires réelles indépendantes, et $S_n = \sum_{i=1}^n X_i$. Si il existe $c_n, v_n \in \mathbb{R}$ tel que $\sum_{i=1}^n \mathbb{E}[X_i^2] \leq v_n$ et, en posant $(X_i)_+ = \max(0, X_i)$, on vérifie :

$$\forall k \geq 3, \quad \sum_{i=1}^n \mathbb{E}[(X_i)_+^k] \leq \frac{v_n k! c_n^{k-2}}{2}$$

Alors, pour tout $\lambda \in [0, \frac{1}{c_n})$ on a

$$\log \left(\mathbb{E}[e^{\lambda(S_n - \mathbb{E}[S_n])}] \right) \leq \frac{v_n \lambda^2}{2(1 - c_n \lambda)}$$

Ce qui implique que pour tout $t \geq 0$ on a :

$$P(S_n - \mathbb{E}[S_n] \geq t) \leq e^{\frac{-t^2}{2(v_n + tc_n)}}$$

② Si X_1, \dots, X_n sont des variables aléatoires réelles indépendantes centrées, avec $V_n = \sum_{i=1}^n \mathbb{E}[X_i^2]$, et telles qu'il existe $M > 0$ tel que $\forall i \in \llbracket 1, n \rrbracket, |X_i| \stackrel{p.s.}{\leq} M$ alors, pour tout $t > 0$ on a :

$$P\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(\frac{-t^2}{2V_n + \frac{2Mt}{3}}\right)$$

Et donc

$$P\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2\exp\left(\frac{-t^2}{2V_n + \frac{2Mt}{3}}\right)$$

Preuve :

① Soit

$$\begin{aligned}\phi : \mathbb{R} &\mapsto \mathbb{R} \\ x &\mapsto e^x - x - 1\end{aligned}$$

On a alors $\forall x \leq 0, \phi(x) \leq \frac{x^2}{2}$ car $\phi(x) = \sum_{k=2}^{+\infty} \frac{x^k}{k!}$ et donc $\phi(x) - \frac{x^2}{2} = \sum_{k=3}^{+\infty} \frac{x^k}{k!}$ de terme dominant $x^3 \leq 0$, quand $x \leq 0$.

On rappelle que l'on note $X_+ = \max(0, X)$ et $X_- = \max(0, -X)$, $\forall X \in \mathbb{R}$ on a donc, pour tout $\lambda \geq 0$:

$$\begin{aligned}\phi(\lambda X_i) &= \phi(\lambda(X_i)_-) + \phi(\lambda(X_i)_+) \\ &\leq \frac{\lambda^2(X_i)_-^2}{2} + \phi(\lambda(X_i)_+) \\ &= \frac{\lambda^2(X_i)_-^2}{2} + \sum_{k=2}^{+\infty} \frac{(\lambda(X_i)_+)^k}{k!} \\ &= \frac{\lambda^2(X_i)_-^2}{2} + \sum_{k=3}^{+\infty} \frac{(\lambda(X_i)_+)^k}{k!} \quad (*)\end{aligned}$$

On a donc

$$\begin{aligned}
\log \left(\mathbb{E} \left[e^{\lambda(S_n - \mathbb{E}[S_n])} \right] \right) &= \sum_{i=1}^n \log \left(\mathbb{E} \left[e^{\lambda(X_i - \mathbb{E}[X_i])} \right] \right) \\
&= \sum_{i=1}^n \log \left(\mathbb{E} \left[e^{\lambda X_i} \right] \right) - \lambda \mathbb{E}[X_i] \\
&= \sum_{i=1}^n \log \left(1 + \mathbb{E}[\lambda X_i + \phi(\lambda X_i)] \right) - \lambda \mathbb{E}[X_i] \\
(\log(1+x) \leq x) &\leq \sum_{i=1}^n \mathbb{E}[\lambda X_i + \phi(\lambda X_i)] - \lambda \mathbb{E}[X_i] \\
\text{par (*)} &\leq \sum_{i=1}^n \mathbb{E} \left[\frac{\lambda^2 (X_i)^2}{2} \right] + \sum_{k=3}^{+\infty} \frac{\mathbb{E}[(\lambda(X_i)_+)^k]}{k!} \\
&\leq \frac{\lambda^2}{2} v_n + \sum_{k=3}^{+\infty} \frac{\lambda^k}{k!} \sum_{i=1}^n \mathbb{E}[(X_i)_+^k] \\
&\leq \frac{\lambda^2 v_n}{2} + \sum_{k \geq 3} \frac{\lambda^k v_n c_n^{k-2}}{2} \\
&\leq \frac{\lambda^2 v_n}{2} \sum_{k \geq 0} (\lambda c_n)^k \\
(\text{pour } \lambda \in [0, \frac{1}{c_n}]) &\leq \frac{v_n \lambda^2}{2(1 - c_n \lambda)}
\end{aligned}$$

Ainsi, $\forall \lambda \in [0, \frac{1}{c_n})$

$$\log \left(\mathbb{E} \left[e^{\lambda(S_n - \mathbb{E}[S_n])} \right] \right) \leq \frac{v_n \lambda^2}{2(1 - c_n \lambda)}$$

Ainsi, par cette inégalité, on a pour tout $t \geq 0$:

$$\begin{aligned}
P(S_n - \mathbb{E}[S_n] \geq t) &= P(e^{\lambda(S_n - \mathbb{E}[S_n])} \geq e^{\lambda t}) \\
&\leq \mathbb{E} \left[e^{\lambda(S_n - \mathbb{E}[S_n])} \right] e^{-\lambda t} \\
\log \left(P(S_n - \mathbb{E}[S_n] \geq t) \right) &\leq \frac{v_n \lambda^2}{2(1 - c_n \lambda)} - \lambda t
\end{aligned}$$

or $\lambda \mapsto \frac{v_n \lambda^2}{2(1 - c_n \lambda)} - \lambda t$ est convexe sur l'ensemble (convexe) $[0, \frac{1}{c})$.
on a en effet

$$\begin{aligned}
\left(\frac{v_n \lambda^2}{2(1 - c_n \lambda)} - \lambda t\right)'' &= \left(\frac{\lambda v_n (2 - \lambda c_n)}{2(1 - \lambda c_n)}\right)' \\
&= v_n \frac{-c_n^2 \lambda^2 + \lambda c_n + 1}{\underbrace{(1 - \lambda c_n)^3}_{\geq 0, \lambda \in [0, \frac{1}{c_n})}}
\end{aligned}$$

Or, le polynôme $\lambda \mapsto -c_n^2 \lambda^2 + \lambda c_n + 1$ est toujours positif sur $[0, \frac{1}{c_n})$. On résout donc

$$\left(\frac{v_n \lambda^2}{2(1 - c_n \lambda)} - \lambda t\right)' = 0$$

et on obtient le résultat $P(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-\frac{t^2}{2(v_n + t c_n)}}$

② Pour la deuxième assertion, pour prouver que $P(|\sum_{i=1}^n X_i| \geq t) \leq 2\exp(\frac{-t^2}{2V_n + \frac{2Mt}{3}})$, il nous suffit de remarquer qu'on peut appliquer l'inégalité de Bennett à X_1, \dots, X_n , en effet, on a bien qu'elles sont centrées et que $|X_i| \stackrel{p.s.}{\leq} M$, donc on a

$$P(S_n \geq t) \leq e^{\frac{-V_n}{M^2} h(\frac{tM}{V_n})}$$

On a donc toujours $\frac{tM}{V_n} \geq 0$, montrons que pour tout $x \geq 0$, on a $h(x) \geq \frac{x^2}{2 + \frac{2}{3}x}$.

D'abord, on remarque que $h(0) = 0 \geq \frac{0^2}{2 + \frac{2}{3}0} = 0$, et, pour $x \geq 0$ on a

$$\left(h(x) - \frac{x^2}{2 + \frac{2}{3}x}\right)' = \log(1+x) - \frac{x}{(2 + \frac{2}{3}x)^2} = \frac{\log(1+x)(2 + \frac{2}{3}x)^2 - x}{(2 + \frac{2}{3}x)^2} \geq 0$$

Donc on a bien que pour tout $x \geq 0$, $h(x) \geq \frac{x^2}{2 + \frac{2}{3}x}$, donc par décroissance de e^{-x} on a :

$$P(S_n \geq t) \leq e^{\frac{-V_n}{M^2} h(\frac{tM}{V_n})} \leq e^{\frac{-V_n}{M^2} \frac{(\frac{tM}{V_n})^2}{2 + \frac{2}{3} \frac{tM}{V_n}}} = \exp\left(\frac{-t^2}{2V_n + \frac{2Mt}{3}}\right)$$

Et ainsi, avec le même raisonnement que pour la preuve précédente (Bennett) on obtient :

$$P(|S_n| \geq t) = P(S_n \geq t \cup -S_n \geq t) = P(S_n \geq t) + P(-S_n \geq t) \leq 2\exp\left(\frac{-t^2}{2V_n + \frac{2Mt}{3}}\right)$$

Illustration :

Montrons une application concrète des inégalités de concentrations précédentes (*i.e.* : Chebychev, Hoeffding, Bennett et Bernstein) à une suite de variables aléatoires réelles $(X_n)_{n \geq 1} \stackrel{\text{iid}}{\sim} \mathcal{U}([0, 1])$. Précisons notre méthode pour Hoeffding, Bennett et Bernstein.

① Pour utiliser Hoeffding on applique l'inégalité à la variable aléatoire $Y_i = X_i/n$, avec donc presque sûrement $0 \leq Y_i \leq \frac{1}{n}$, $\forall i \in \llbracket 1, n \rrbracket$. Pour $t \geq 0$ on a

$$P(|\sum_{i=1}^n Y_i - \mathbb{E}[Y_i]| \geq t) = P(|\overline{X_n} - \mathbb{E}[X_1]| \geq t) \leq 2 \exp\left(\frac{-2t^2}{n \frac{1}{n^2}}\right) = 2e^{-2nt^2}$$

② Pour Bennett, on a pour tout $n \geq 1$, pour tout $i \in \llbracket 1, n \rrbracket$, avec $Y_i = \frac{X_i}{n} - \frac{1}{2n}$, $(Y_i)_{1 \leq i \leq n}$ est donc centrée et $|Y_i| = |\frac{X_i - 1/2}{n}| \stackrel{p.s.}{\leq} \frac{1}{2n} = c$. Donc :

$$V_n = \sum_{i=1}^n \mathbb{E}[Y_i^2] = \sum_{i=1}^n \mathbb{V}(Y_i^2) = \sum_{i=1}^n \frac{1}{n^2} n \mathbb{V}(X_i - 1/2) = \frac{1}{n^2} n \frac{1}{12} = \frac{1}{12n}$$

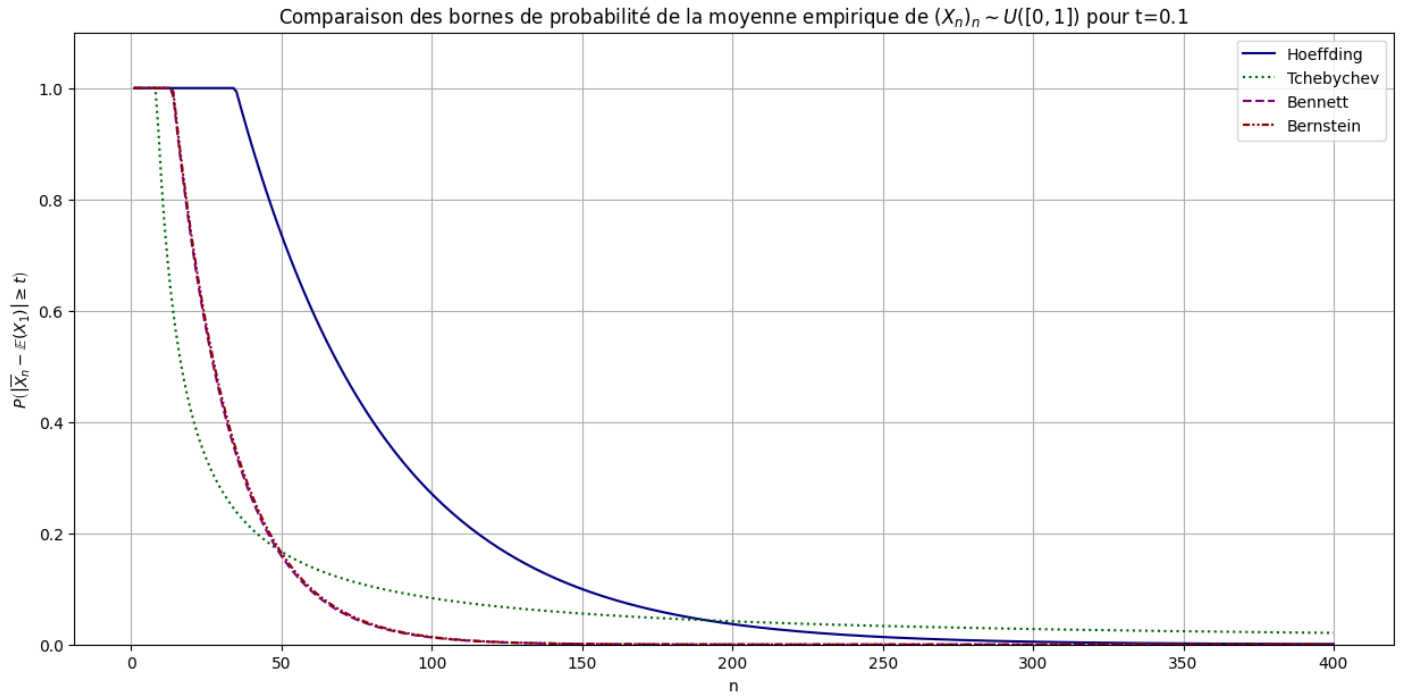
Ainsi, pour $t \geq 0$ on a :

$$P(|\sum_{i=1}^n Y_i - \mathbb{E}[Y_i]| \geq t) = P(|\overline{X_n} - \mathbb{E}[X_1]| \geq t) \leq 2e^{-\frac{V_n}{c^2} h(\frac{ct}{V_n})} = 2e^{-\frac{\frac{1}{12n}}{(2n)^2} h(t \frac{1/2n}{1/12n})} = 2e^{-\frac{n}{3} h(6t)}$$

③ Enfin, pour Bernstein, on applique l'inégalité à $Y_i = \frac{X_i}{n} - \frac{1}{2n}$ (centrée). Or, pour tout $n \geq 1$, pour tout $i \in \llbracket 1, n \rrbracket$, on a donc $|Y_i| = \underbrace{\frac{1}{n} |X_i - \frac{1}{2}|}_{\leq 1/2} \leq \frac{1}{2n} = M$.

Et donc, on a $V_n = \sum_{i=1}^n \mathbb{E}[Y_i^2] = \sum_{i=1}^n \mathbb{V}(Y_i) = \sum_{i=1}^n \frac{1}{n^2} \mathbb{V}(X_i - 1/2) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_1) = \frac{1}{12n}$. Ainsi, pour $t \geq 0$

$$P(|\sum_{i=1}^n Y_i| \geq t) = P(|\overline{X_n} - \mathbb{E}[X_1]| \geq t) \leq 2 \exp\left(\frac{-t^2}{2V_n + \frac{2Mt}{3}}\right) = 2 \exp\left(\frac{-t^2}{\frac{1}{6n} + \frac{t}{3n}}\right) = 2 \exp\left(\frac{-6nt^2}{1 + 2t}\right)$$

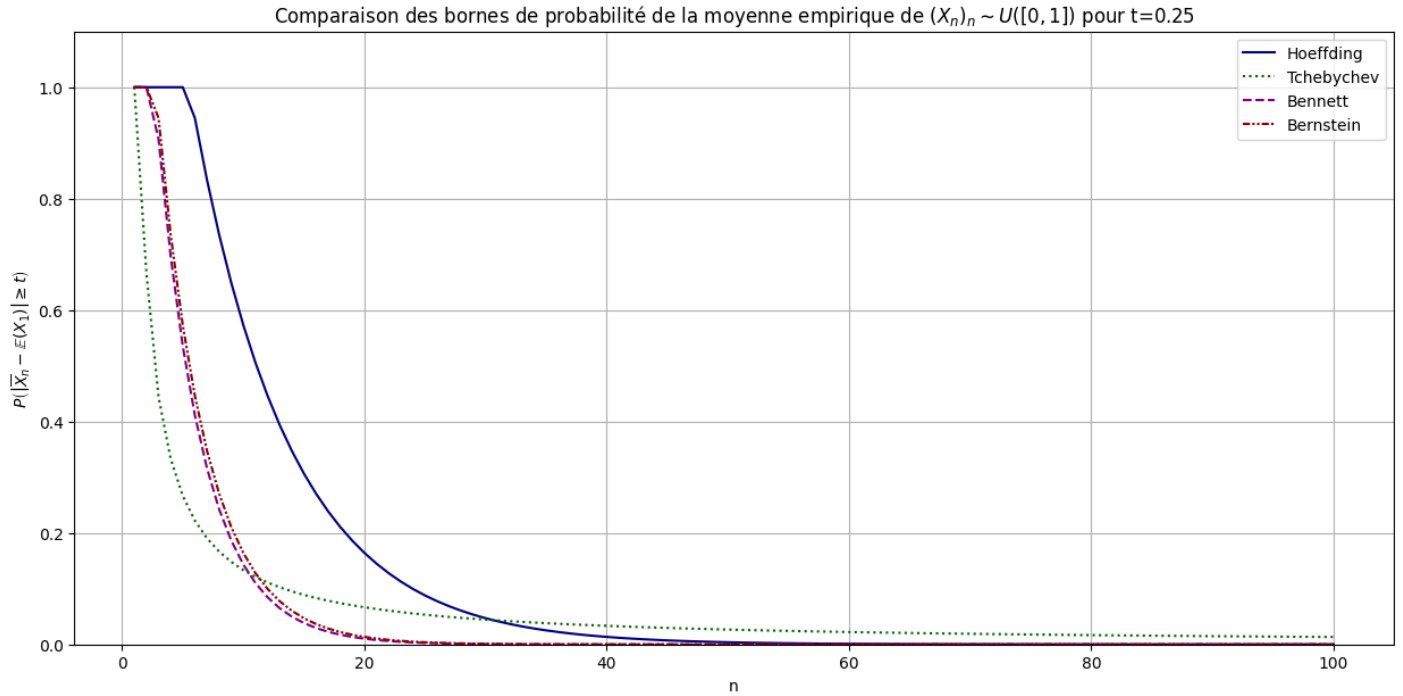


On observe que la précision de toutes les inégalités finit par surpasser celle de Tchebychev, ce qui est toujours vrai pour toutes les variables aléatoires respectant les conditions spécifiques de Hoeffding, Bennett et Bernstein. En revanche, on peut noter que l'inégalité de Tchebychev est plus précise que les autres pour des valeurs "petites" de n , ce qui est également généralement vrai pour d'autres types de variables aléatoires. On constate ainsi que la qualité de la borne de Tchebychev diminue (par rapport aux autres) à mesure que le nombre d'itérations augmente. En effet, grâce à la décroissance exponentielle des inégalités de Hoeffding, Bennett et Bernstein, on remarque qu'au début, la borne n'est pas très précise, mais qu'avec le temps, la décroissance de la borne s'accélère, tandis que celle de Tchebychev ralentit.

En résumé, il n'existe pas a priori de "meilleure" inégalité de concentration ; on sait simplement que l'inégalité de Tchebychev est préférable pour un faible nombre d'itérations, tandis que celles de Hoeffding, Bennett et Bernstein sont à privilégier lorsque n est élevé. Cependant, l'inégalité de Bennett est toujours plus précise que celle de Bernstein, car Bernstein est une majoration simple de Bennett, qui reste cependant plus pratique à manipuler. Comme on peut le voir, la différence entre Bennett et Bernstein est souvent infime, ce qui fait que les statisticiens ont tendance à privilégier cette dernière. L'inégalité de Bernstein est elle-même très souvent plus précise que celle de Hoeffding, en effet l'avantage des inégalités de Bennett et de Bernstein réside dans leur utilisation de l'information sur la variance de la loi, tandis que Hoeffding, plus grossière, se contente d'exploiter les bornes des variables aléatoires. La plus précise des inégalités est donc celle de Bennett à moyen terme, même si Hoeffding et Bernstein sont très souvent suffisantes, avec une très faible différence de borne entre Bennett et Bernstein. Par ailleurs, il est important de savoir qu'actuellement, sous les conditions données par les trois inégalités, on ne peut pas vraiment faire mieux que l'inégalité de **Bennett** en termes de bornes supérieures.

On peut aussi voir une autre image, pour un t plus important, on remarque que la différence entre

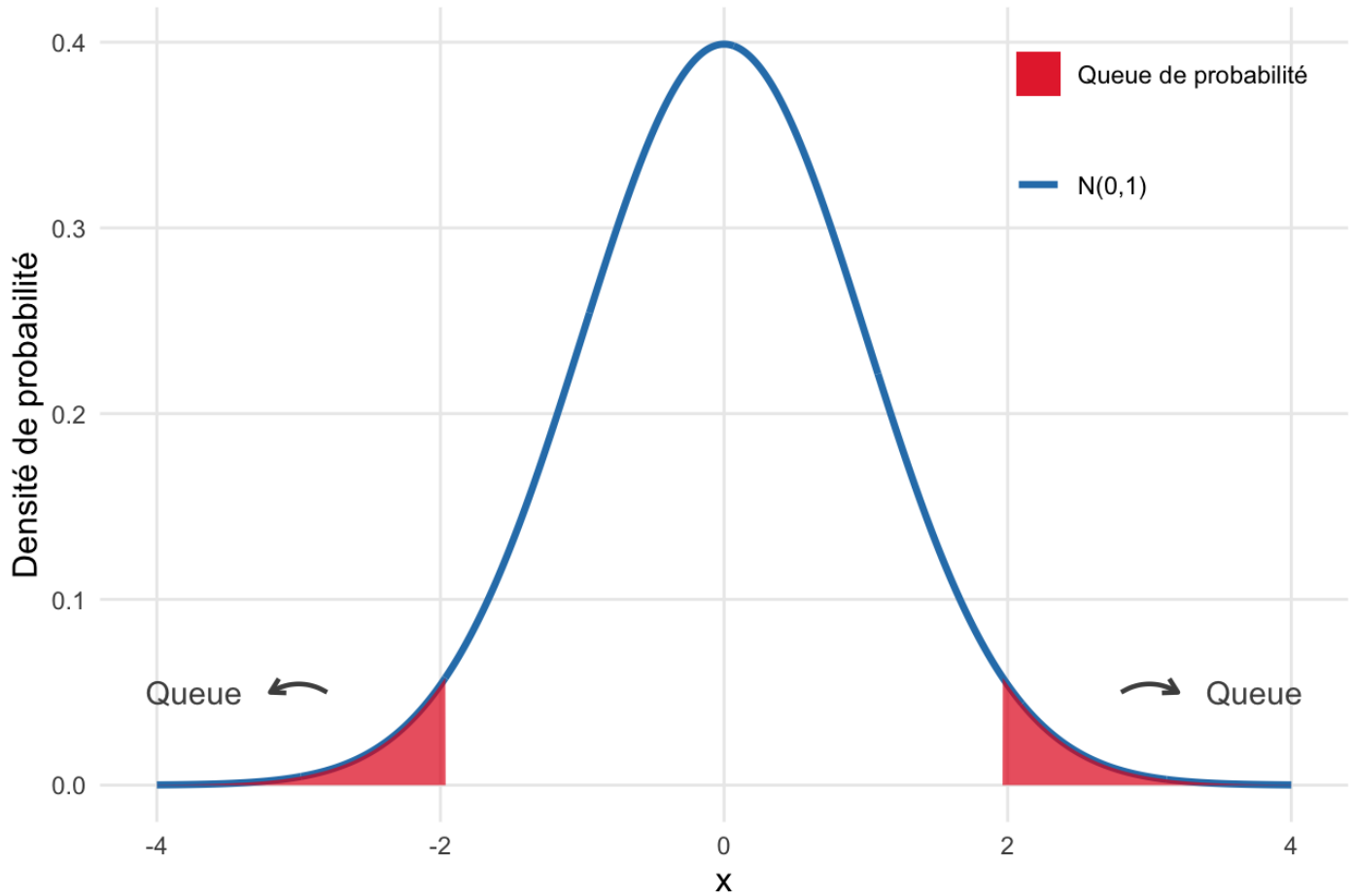
Bennett et Bernstein se fait plus ressentir.



13.1.2 Le Cas des Variables Aléatoires Non Bornées

Nous avons vu maintenant les meilleures inégalités de concentrations dans le cas où X_1, \dots, X_n sont bornées. Maintenant, nous allons étudier comment contrôler la concentrations des variables aléatoires réelles quand elles ne sont pas bornées. Pour cela, on va étudier la "queue" de distribution de nos variables aléatoires, c'est à dire, à quel point la densité de nos variables "s'écrasent" vers 0 sur leurs queues. En réalité, plus nos variables aléatoires ont un affaissement de queue rapide, plus on pourra contrôler fortement la concentration de nos variables autour de leur moyenne. Nous utiliserons toujours la méthode de **Chernoff** en commençant par démontrer un théorème que nous utiliserons abondamment par la suite et nous continuerons sur les variables aléatoires dites "sous-gaussiennes", qui ont un écrasement exponentiel (voir ci-dessous).

Distribution Normale Standard



Théorème 13|13.1|1

Soit X une variable aléatoire réelle, et $b \in [0, \infty]$ tel que pour tout $\lambda \in (0, b)$ on a $\mathbb{E}[e^{\lambda X}] < \infty$. Alors

$$\forall t \geq 0, \quad P(X - \mathbb{E}[X] > t) \leq e^{-\psi^*(t)}$$

$$\text{où } \psi^*(t) = \sup_{\lambda \in (0, b)} \{ \lambda t - \log(\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}]) \}$$

Preuve :

Soit $t \geq 0$ et $\lambda \in (0, b)$, on a $P(X - \mathbb{E}[X] > t) = P(e^{\lambda(X - \mathbb{E}[X])} > e^{\lambda t})$ et donc, par Markov,

$$P(e^{\lambda(X - \mathbb{E}[X])} > e^{\lambda t}) \leq \frac{\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}]}{e^{\lambda t}} = \exp[-(\lambda t - \log(\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}])]$$

Comme c'est vrai pour tout $\lambda \in (0, b)$ on peut juste optimiser en $\lambda \in (0, b)$ et cela donne le résultat escompté.

L'intérêt de la méthode de Chernoff est qu'elle permet d'obtenir des résultats non-asymptotiques, c'est à dire valables pour toute valeur de n , sous des conditions sur la loi des variables aléatoires qui

sont moins restrictives que de préciser un modèle paramétrique pour la loi de la variable aléatoire.

Définition 13|13.1|1

Soit X une variable aléatoire réelle intégrable et $\sigma \geq 0$. On dit que X est σ^2 -**sous-gaussienne** si

$$\forall \lambda \in \mathbb{R}, \quad \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leq e^{\sigma^2 \lambda^2 / 2}$$

Lemme

Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, alors pour tout $\lambda \in \mathbb{R}$:

$$\mathbb{E}[e^{\lambda X}] = e^{\lambda\mu + \frac{1}{2}\sigma^2\lambda^2}$$

Preuve :

$$\mathbb{E}[e^{\lambda X}] = \int_{\mathbb{R}} e^{\lambda x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx$$

$$\text{Or } \lambda x - \frac{(x - \mu)^2}{2\sigma^2} = \lambda\mu + \frac{\lambda^2\sigma^2}{2} - \frac{1}{2\sigma^2}(x - \mu - \lambda\sigma^2)^2$$

Ainsi,

$$\mathbb{E}[e^{\lambda X}] = e^{\lambda\mu + \frac{\lambda^2\sigma^2}{2}} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu - \lambda\sigma^2)^2\right) dx}_{=1}$$

Remarque :

① Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $X - \mathbb{E}[X] \sim \mathcal{N}(0, \sigma^2)$ et donc, par le lemme, toute variable aléatoire réelle gaussienne est σ^2 -sous-gaussienne.

② Toutes variables aléatoires Gaussienne de variance σ^2 est σ^2 -sous-gaussienne.

③ Si X_1, \dots, X_n sont des variables aléatoires indépendantes σ_i^2 -sous-gaussiennes, alors $\sum_{i=1}^n X_i$ est $\sum_{i=1}^n \sigma_i^2$ -sous-gaussienne.

④ Par le **lemme d'Hoeffding**, toutes les variables aléatoires bornées (entre a et b) sont $(\frac{b-a}{2})^2$ -

sous-gaussiennes.

Proposition 13|13.1|1

Soient X_1, \dots, X_n des variables aléatoires iid et σ^2 –sous-gaussienne. Alors, pour tout $t \geq 0$,

$$P\left(\sum_{i=1}^n X_i - \mu \geq t\right) \leq e^{\frac{-t^2}{2n\sigma^2}}$$

et donc

$$P\left(\left|\sum_{i=1}^n X_i - \mu\right| \geq t\right) \leq 2e^{\frac{-t^2}{2n\sigma^2}}$$

Preuve :

Soient X_1, \dots, X_n des variables aléatoires σ^2 –sous-gaussienne. On a pour tout $t \geq 0$, pour tout $\lambda > 0$,

$$\begin{aligned} P\left(\sum_{i=1}^n X_i - \mu \geq t\right) &= P\left(e^{\lambda \sum_{i=1}^n (X_i - \mu)} \geq e^{\lambda t}\right) \\ &\stackrel{\text{(Markov)}}{\leq} \frac{\mathbb{E}\left[e^{\lambda \sum_{i=1}^n (X_i - \mu)}\right]}{e^{\lambda t}} \\ &\stackrel{\text{(indépendance)}}{=} e^{-\lambda t} \prod_{i=1}^n \mathbb{E}\left[e^{\lambda(X_i - \mu)}\right] \\ &\stackrel{\text{(sous-gaussienne)}}{\leq} e^{-\lambda t} (e^{\sigma^2 \lambda^2 / 2})^n \\ &\stackrel{\text{(en prenant le } \lambda \text{ qui optimise)}}{=} e^{\frac{-t^2}{2n\sigma^2}} \end{aligned}$$

Pour le second résultat, on a bien que si X est σ^2 –sous-gaussienne, alors $-X$ l'est aussi.

corollaire

Si X_1, \dots, X_n sont iid et σ^2 sous-gaussiennes, alors pour tout $t \in \mathbb{R}_+$,

$$P(\overline{X_n} - \mathbb{E}[X_1] \geq t) \leq \exp\left(\frac{-nt^2}{2\sigma^2}\right) = \alpha$$

Donc, pour tout $\alpha \in (0, 1)$,

$$P(|\overline{X_n} - \mathbb{E}[X_1]| \geq \sigma \sqrt{\frac{2 \log(\frac{2}{\alpha})}{n}}) \leq \alpha$$

Lemme (Symétrisation)

Soit X une variable aléatoire réelle et $\lambda > 0$ tel que $\mathbb{E}[e^{\lambda X}] < \infty$. Alors, si X' est une variable aléatoire indépendante de X et de même loi que X et ϵ suivant une loi de Rademacher indépendante de X et X' , alors on a

$$\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leq \mathbb{E}[e^{\lambda\epsilon(X - X')}]$$

Preuve :

Par indépendance (et même loi) de X et X' on a $\mathbb{E}[X] = \mathbb{E}[X'|X]$. De plus, soit $A \in \mathcal{B}(\mathbb{R})$, on a

$$\begin{aligned} P(\epsilon(X - X') \in A) &= P(X - X' \in A \cap \epsilon = 1) + P(X' - X \in A \cap \epsilon = -1) \\ &= P(X - X' \in A)P(\epsilon = 1) + P(X' - X \in A)P(\epsilon = -1) \\ &= P(X - X' \in A)(P(\epsilon = 1) + P(\epsilon = -1)) = P(X - X' \in A) \end{aligned}$$

Donc $\epsilon(X - X') \stackrel{(d)}{=} X - X'$

et donc

$$\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] = \mathbb{E}[e^{\lambda(\mathbb{E}[X - X'|X])}] \stackrel{(\text{Jensen})}{\leq} \mathbb{E}[\mathbb{E}[e^{\lambda(X - X')}|X']] = \mathbb{E}[e^{\lambda(X - X')}] = \mathbb{E}[e^{\lambda\epsilon(X - X')}]$$

Proposition 13|13.1|2

Soit X une variable aléatoire réelle centrée et telle que ses moments sont "sous géométriques" au sens où il existe $b > 0$ tel que

$$\forall k \geq 2, \quad |\mathbb{E}[X^k]| \leq b^k$$

Alors X est $4b^2$ -sous-gaussienne.

Preuve :

Soit $\lambda > 0$, par le lemme de symétrisation on a pour $X' \stackrel{(d)}{=} X, X \perp\!\!\!\perp X'$:

$$\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}[e^{\lambda\epsilon(X - X')}]$$

Par symétrie de la variable $\epsilon(X - X')$, on a pour tout $k \in \mathbb{N}$, que $\mathbb{E}[(\epsilon(X - X'))^{2k+1}] = 0$. En effet,

$$\begin{aligned}
\mathbb{E}[\epsilon(X - X')^{2k+1}] &= \mathbb{E}[(\mathbb{1}_{\epsilon=1}(X - X')^{2k+1})] + \mathbb{E}[(\mathbb{1}_{\epsilon=-1}(X - X')^{2k+1})] \\
&= P(\epsilon = 1)\mathbb{E}[(X - X')^{2k+1}] + (1 - P(\epsilon = 1))\mathbb{E}[(X - X')^{2k+1}] \\
&= \mathbb{E}[(X - X')^{2k+1}] \\
&= \mathbb{E}[(-1)^{2k+1}(X' - X)^{2k+1}] = -\mathbb{E}[(X - X')^{2k+1}] \\
&= -\mathbb{E}[(X' - X)^{2k+1}] = -\underbrace{\mathbb{E}[\epsilon(X - X')^{2k+1}]}_{\text{car } X' - X \stackrel{(d)}{=} X - X'}
\end{aligned}$$

Donc $\mathbb{E}[\epsilon(X - X')^{2k+1}] = 0$. On en déduit par développement en séries entières que

$$\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}[e^{\lambda \epsilon(X - X')}] = \sum_{k=0}^{\infty} \frac{\lambda^k \mathbb{E}[(\epsilon(X - X'))^k]}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^{2k} \mathbb{E}[(X - X')^{2k}]}{(2k)!}$$

Or, en rappelant que $\epsilon(X - X') \stackrel{(d)}{=} X - X'$ on a :

$$\mathbb{E}[(X - X')^{2k}] = \sum_{l=0}^{2k} \binom{2k}{l} \mathbb{E}[X^{2k-l}] \mathbb{E}[X^l] \leq b^{2k} \sum_{l=0}^{2k} \binom{2k}{l} \stackrel{=(1+1)^{2k}}{=} b^{2k} 2^{2k}$$

En utilisant le fait que $(2k)! \geq 2^k k!$, en effet, c'est vraie pour $k = 0$, supposons désormais que $(2k)! \geq 2^k k!$ on a :

$$\begin{aligned}
(2(k+1))! &= (2k+2)(2k+1)2k! \\
&\geq (2k+2)(2k+1)2^k k! \\
&= 2^{k+1}(k+1)!(2k+1) \geq 2^{k+1}(k+1)!
\end{aligned}$$

Ce qui démontre le résultat par récurrence. Ainsi, nous obtenons :

$$\mathbb{E}[e^{\lambda k}] \leq \sum_{k=0}^{+\infty} \frac{\lambda^{2k} b^{2k} 2^k 2^k}{(2k)!} \leq \sum_{k=0}^{+\infty} \frac{(\sqrt{2}b\lambda)^{2k}}{k!} = e^{2\lambda^2 b^2}$$

Définition 13|13.1|2

Passons désormais aux variables aléatoires dites "**sous-poissonniennes**". Il s'agit d'une majoration moins forte de la queue de notre variable aléatoire. C'est donc moins restrictif.

Une variable aléatoire X est dite **sous Poissonnienne** s'il existe v^2 et $b \geq 0$ tels que

$$\forall \lambda > 0, \quad \log \left(\mathbb{E} \left[e^{\lambda(X - \mathbb{E}[X])} \right] \right) \leq \frac{v^2}{b^2} (e^{b\lambda} - 1 - b\lambda)$$

On note alors $X \in \text{sPoi}(v^2, b)$, et si $b = 0$ on doit comprendre qu'alors X est v^2 -sous-gaussienne.

Remarque :

Si $X \sim \mathcal{P}(\theta)$, $\theta > 0$ alors X est sous Poissonnienne.

Preuve :

Soit $\lambda > 0$, on rappelle que $\mathbb{E}[X] = \theta$.

$$\begin{aligned} \mathbb{E} \left[e^{\lambda(X - \mathbb{E}[X])} \right] &= \mathbb{E} \left[e^{-\lambda \mathbb{E}[X]} \right] \varphi_X \left(\frac{\lambda}{i} \right) \\ &= e^{-\lambda \theta} e^{\theta(e^\lambda - 1)} \end{aligned}$$

Donc $\log \left(\mathbb{E} \left[e^{\lambda(X - \mathbb{E}[X])} \right] \right) = -\lambda \theta + \theta(e^\lambda - 1) = \theta(e^\lambda - \lambda - 1)$, donc on a bien que $X \in \text{sPoi}(\theta, 1)$.

Proposition 13|13.1|3

Soit X une variable aléatoire telle que, pour $v^2, b \geq 0$ on a :

$$\forall k \geq 2, \quad \mathbb{E} \left[(X - \mathbb{E}[X])^2 (X - \mathbb{E}[X])_+^{k-2} \right] \leq v^2 b^{k-2}$$

Alors $X \in \text{sPoi}(v^2, b)$.

Preuve :

14 Estimation Statistique

14.1 Principes d'Échantillonnage

En pratique, on a des données $x_1, \dots, x_n \in \mathbb{R}^m, m \geq 1$ et on pose l'hypothèse E : On observe une réalisation de n variables aléatoires iid X_1, \dots, X_n de même loi P^* . On appelle X_1, \dots, X_n l'échantillon.

On suppose aussi que X_1, \dots, X_n sont sur un espace probabilisé (Ω, \mathcal{A}, P) avec

$$X_i : \Omega \rightarrow \mathbb{R}^m$$

Les données sont donc les réalisations de ces variables aléatoires, *i.e.*

$$(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$$

On a $\mathbf{P}^* = \{\mathcal{P}(X_i \in A), A \in \mathcal{A}\} = \text{Loi de } X_i$. On restera (sauf cas contraire mentionné) dans le cas $\Omega = \mathbb{R}^m, \mathcal{A} = \mathcal{B}(\mathbb{R}^m)$, donc \mathbf{P}^* est une loi sur $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$.

Le problème statistique est de trouver l'inconnue \mathbf{P}^* , à partir de X_1, \dots, X_n

- **D'abord, on se pose un problème d'estimation**

Soit \mathcal{P} l'ensemble des lois sur $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ et on pose

$$T : \mathcal{P} \rightarrow \mathbb{R}^p$$

On cherche la meilleure approximation de $T(\mathbf{P}^*)$

- **Puis, on se pose un problème de test d'hypothèse**

On prend $\mathcal{P}_0 \subset \mathcal{P}$ et l'on souhaite savoir si $\mathbf{P}^* \in \mathcal{P}_0$

On peut avoir, par exemple :

- $\mathcal{P}_0 = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\}$
- $\mathcal{P}_0 = \{\mathbf{P}^*, \mathbb{E}[X_1] = 0\}$

14.2 Principe d'un Estimateur Statistique

Comme évoqué ci-dessus, on cherche à estimer $T(\mathbf{P}^*)$, avec θ^* le paramètre à estimer, on a donc

$$T(\mathbf{P}^*) = \theta^*$$

On sait également que la fonction de répartition F^* suffit à déterminer la loi \mathbf{P}^* , ce qui signifie concrètement qu'il existe une bijection

$$\Psi : \mathcal{F} \rightarrow \mathcal{P}$$

Avec \mathcal{F} l'ensemble des fonctions de répartitions sur \mathbb{R}^m

Pour résumer, on estime donc

$$\theta^* = T(\mathbf{P}^*) = T \circ \Psi(F^*)$$

par

$$\hat{\theta}_n = T(\hat{P}_n) = T \circ \Psi(\hat{F}_n)$$

Avec $\hat{P}_n \stackrel{\text{def}}{=} \Psi(\hat{F}_n)$ et $\hat{\theta}_n$ notre estimateur.

Définition 14|14.2|1

On appelle **statistique** toute fonction mesurable des observations $S(X_1, \dots, X_n)$ dans \mathbb{R}^m . Une statistique S est donc une variable aléatoire ou un vecteur aléatoire qui ne dépend **que** de l'échantillon. Une statistique est aussi appelée estimateur si elle est utilisée pour estimer des paramètres (ou d'autres caractéristiques) d'une loi de probabilité.

Remarque :

1) On peut par ailleurs prouver que

$$\forall A \in \mathcal{B}(\mathbb{R}^m), \hat{P}_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_A(X_i) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(A)$$

2) Si $h : \mathbb{R}^m \rightarrow \mathbb{R}^m$ est une fonction mesurable, alors

$$\int h(x) d\hat{P}_n(x) = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

On définit donc pour cela la moyenne empirique

$$\bar{X}_n = \int_{\mathbb{R}^m} x d\hat{P}_n(x) = \frac{1}{n} \sum_{i=1}^n X_i$$

et la variance empirique

$$\hat{\Sigma}_n = \int_{\mathbb{R}^m} x x^T d\hat{P}_n(x) = \frac{1}{n} \sum_{i=1}^n X_i X_i^T - \bar{X}_n \bar{X}_n^T = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T$$

On utilise aussi régulièrement des **méthodes d'inversions** pour générer des variables aléatoires à partir d'une fonction de répartition donnée. C'est aussi une méthode très efficace pour simuler des distributions continues lorsque le pseudo inverse F^{-1} est explicitement calculable. On les utilise dans de nombreux domaines tel que par exemple les algorithmes fondés sur des méthodes de Monte-Carlo, ou bien en *Machine Learning*.

Proposition 14|14.2|1

Soit U une variable aléatoire uniforme sur $[0, 1]$, F une fonction de répartition et

$$\begin{aligned} F^{-1} : [0, 1] &\mapsto \mathbb{R} \\ u &\mapsto \inf_{x \in \mathbb{R}} \{F(x) \geq u\} \end{aligned}$$

son inverse généralisés. Alors

- (i) La variable aléatoire $X = F^{-1}(U)$ a pour fonction de répartition F .
- (ii) Si $X \sim F^*$, et que F^* est continue, alors $F^*(X) \sim \mathcal{U}([0, 1])$.

Preuve :

① Soit $X = F^{-1}(U)$, et $x \in \mathbb{R}$. Montrons que pour tout $u \in (0, 1]$, on a

$$F^{-1}(u) \leq x \iff u \leq F(x)$$

Par définition de $F^{-1}(u)$, si $u \leq F(x)$ alors $F^{-1}(u) \leq x$. On a donc $u \leq F(x) \implies F^{-1}(u) \leq x$.

Inversement, si $F^{-1}(u) \leq x$, alors, pour tout $\epsilon > 0$, on a $F^{-1}(u) \leq x + \epsilon$, donc, par définition de $F^{-1}(u)$, on a $u \leq F(x + \epsilon)$.

Et comme, par propriété des fonctions de répartitions, F est continue à droite, on a $u \leq F(x)$, par conséquent, on a bien que

$$F^{-1}(u) \leq x \iff u \leq F(x)$$

Ainsi on a, pour tout $x \in [0, 1]$:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

② Soit $X \sim F^*$, trouvons la loi de $F^*(X)$. Posons

$$\begin{aligned} F^{*-1} : [0, 1] &\mapsto \mathbb{R} \\ u &\mapsto \inf_{x \in \mathbb{R}} \{F^*(x) \geq u\} \end{aligned}$$

On a donc bien $\forall t \in [0, 1]$, $F^*(F^{*-1}(t)) = t$ car : soit $t \in [0, 1]$.

$$F^*(F^{*-1}(t)) \geq t \quad (\text{par définition})$$

Mais, pour tout $\epsilon > 0$, on a $F^*(F^{*-1}(t - \epsilon)) \leq t$ par définition. Par continuité de F , on a bien par passage à la limite que $F^*(F^{*-1}(t)) = t$.

De plus, Soit $t \in [0, 1]$, et considérons $A_t = \{x \in \mathbb{R} \mid F(x) \leq t\}$, on a

$$\begin{aligned} P(F^*(X) \leq t) &= P(X \in A_t) \\ &= P(X \in (-\infty, \sup A_t)) \end{aligned}$$

Montrons que $\sup A_t = \sup\{x \in \mathbb{R} : F^*(x) \leq t\} = \inf \overbrace{\{x \in \mathbb{R} : F(x) \geq t\}}^{B_t} = \inf B_t \stackrel{(\text{def})}{=} F^{*-1}(t)$. Directement, par croissance de F^* on a $\sup A_t \leq \inf B_t$. De plus, pour montrer que $\sup A_t \geq \inf B_t$, supposons par l'absurde que $\sup A_t < \inf B_t$. Alors, par définition du sup et de l'inf, il n'existe aucun $x \in \mathbb{R} \mid \sup A_t < x < \inf B_t$ et $F^*(x) = t$. Cependant, par croissance de F^* , pour tout $x \in (\sup A_t, \inf B_t)$, $F^*(x) > t$ car $x \notin A_t$ mais aussi $F^*(x) < t$ car $x \notin B_t$ ce qui est une contradiction. Par conséquent, on a bien $\sup A_t \geq \inf B_t$ et $\sup A_t \leq \inf B_t$, donc $\sup A_t = \inf B_t$ et ainsi :

$$\sup A_t = \inf B_t \iff \sup\{x \in \mathbb{R} : F^*(x) \leq t\} = \inf_{x \in \mathbb{R}} \{F(x) \geq t\} = F^{*-1}(t)$$

Cette dernière propriété est donc vraie peu importe la fonction de répartition. Et ainsi :

$$\begin{aligned} P(F^*(X) \leq t) &= P(X \in A_t) \\ &= P(X \in (-\infty, \sup A_t)) \\ &= P(X \leq \inf B_t) \\ &= F^*(F^{*-1}(t)) \\ &= t \end{aligned}$$

Avec aussi, $\forall t > 1$, $P(F^*(X) \leq t) = 1$ et $\forall t < 0$, $P(F^*(X) \leq t) = 0$ par propriété de la fonction de répartition. Ainsi, on a bien que $F^*(X) \sim \mathcal{U}([0, 1])$

14.3 Modèle Statistique

On appelle **modèle statistique** le triplet

$$\{\mathcal{X}, \mathcal{F}, \{\mathbb{P}_\theta, \theta \in \Theta\}\}$$

Avec \mathcal{X} appelé "espace d'état", \mathcal{F} la tribu sur \mathcal{X} et \mathbb{P}_θ une mesure probabilité sur \mathcal{X} .

Dans notre cas toujours égale à $\{\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \{\mathbb{P}_\theta, \theta \in \Theta\}\}$ qu'on simplifiera donc par

$$\{\mathbb{P}_\theta, \theta \in \Theta\}$$

On fera dans toute la section sur les estimateurs l'hypothèse (P) étant que $\mathbf{P}^* \in \{\mathbf{P}_\theta, \theta \in \Theta\}$ avec donc pour un certain $\theta^* \in \Theta$ tel que $\mathbf{P}^* = \mathbb{P}_{\theta^*}$.

14.3.1 Définitions

1) On dit qu'un modèle statistique est **paramétrique** si

$$\exists K \in \mathbb{N}, \Theta \subset \mathbb{R}^K$$

2) On dit qu'un modèle statistique est **identifiable** si

$$\forall \theta, \theta' \in \Theta, \quad \mathbb{P}_\theta = \mathbb{P}_{\theta'} \implies \theta = \theta'$$

i.e. : la fonction $\theta \rightarrow \mathbb{P}_\theta$ est injective

3) On dit qu'un modèle statistique est **dominé** si il existe une mesure μ σ -finie tel que

$$\forall \theta \in \Theta, \mathbb{P}_\theta \ll \mu \iff \exists f \text{ mesurable } | \forall A \in \mathcal{F}, \quad \mathbb{P}_\theta(A) = \int_A f(x) d\mu(x) \text{ (Radon-Nykodim)}$$

Remarque :

• On dit qu'un modèle est **à densité** (continue) s'il est dominé par la mesure de Lebesgue

• on dit qu'un modèle est **discret** si il existe A au plus dénombrable tel que

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta(A) = 1$$

un modèle est discret si et seulement si $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est dominé par la mesure de comptage sur Θ dénombrable

4) On dit que $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est **régulier**, si il vérifie les 4 hypothèses suivantes :

H1 : le support de \mathbb{P}_θ ne dépend pas de θ , avec le support défini par

$$\text{supp}(\mathbb{P}_\theta) = \{x \in \mathbb{R}^m, \forall \epsilon > 0, \mathbb{P}_\theta(B(x, \epsilon) > 0)\}$$

Cette hypothèse est équivalente à dire que

$$\forall \theta, \theta' \in \Theta, \quad f(x, \theta) > 0 \implies f(x, \theta') > 0$$

H2 : $f(x, \cdot)$ et $\log(f(x, \cdot))$ sont $\mathcal{C}^2(\Theta)$ pour μ presque tout x (μ la mesure dominante du modèle).

H3 : $\forall \theta \in \Theta, \exists U \subset \Theta$ tel que U est ouvert, et tel qu'il existe Λ une fonction mesurable tel que

$$\forall x \in \mathbb{R}, \sup_{\theta \in U} (|l'(x, \theta)| + l'(x, \theta)^2 + |l''(x, \theta)|) \leq \Lambda(x)$$

avec $l(x, \theta) = \log(f(x, \theta))$, $l'(x, \theta) = \frac{\partial l(x, \theta)}{\partial \theta}$ et

$$\int_{\mathbb{R}} \sup_{\theta \in \Theta} f(x, \theta) \Lambda(x) d\mu(x) < \infty$$

H4 : $\forall \theta \in \Theta, \quad \mathbb{E}_\theta[l'(X_1, \theta)^2] > 0$ (dans \mathbb{R}^m , on remplace par $\mathbb{E}_\theta[\nabla(l)(x, \theta) \nabla(l)(x, \theta)^T] > 0$)

Remarque :

Les modèles gaussiens, exponentielles, Bernouilli, Binomiale, Géométrique, Poisson sont réguliers. En revanche, le modèle Laplace ($f(x, \theta) \propto e^{-(x-\theta)}$) et uniforme ne sont pas réguliers.

14.4 Estimateurs

Toute statistique $\hat{\theta}_n, \bar{\theta}_n, \tilde{\theta}_n$ est appelé estimateur si $\exists g$ mesurable tel que $\hat{\theta}_n, \bar{\theta}_n, \tilde{\theta}_n = g(X_1, \dots, X_n)$

- Un estimateur $\hat{\theta}_n$ est dit **consistant** si $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} \theta^*$
- Un estimateur $\hat{\theta}_n$ est dit **fortement consistant** si $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta^*$

On définit le **biais** d'un estimateur $\hat{\theta}_n$ la fonction $B_\theta : \Theta \rightarrow \mathbb{R}$ tel que

$$B_\theta(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n - \theta]$$

14.4.1 Comparer les Estimateurs

Définition 14|14.4|1

Soit $\hat{\theta}_n$ un estimateur, on appelle **risque** de $\hat{\theta}_n$ la fonction $R_\theta : \Theta \rightarrow \mathbb{R}$ tel que

$$R_\theta(\hat{\theta}_n) = \mathbb{E}[\|\hat{\theta}_n - \theta\|_2^2]$$

Remarque :

On a aussi

$$R_\theta(\hat{\theta}_n) = \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\|_2^2] + \|\mathbb{E}[\hat{\theta}_n] - \theta\|_2^2 = \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\|_2^2] + \|B_\theta(\hat{\theta}_n)\|_2^2$$

Preuve :

$$\begin{aligned} R_\theta(\hat{\theta}_n) &= \mathbb{E}[\|\hat{\theta}_n - \theta\|_2^2] \\ &= \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n] + \mathbb{E}[\hat{\theta}_n] - \theta\|_2^2] \\ &= \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\|_2^2] + 2\mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])^T(\mathbb{E}[\hat{\theta}_n] - \theta)] + \mathbb{E}[\|\mathbb{E}[\hat{\theta}_n] - \theta\|_2^2] \\ &= \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\|_2^2] + 2(\mathbb{E}[\hat{\theta}_n] - \theta)(\mathbb{E}[\hat{\theta}_n] - \mathbb{E}[\hat{\theta}_n])^T + \|\mathbb{E}[\hat{\theta}_n] - \theta\|_2^2 \\ &= \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\|_2^2] + \|\mathbb{E}[\hat{\theta}_n] - \theta\|_2^2 \\ &= \mathbb{E}[\|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\|_2^2] + \|B_\theta(\hat{\theta}_n)\|_2^2 \end{aligned}$$

Pour des variables réelles univariée, on a donc

$$R_\theta(\hat{\theta}_n) = \mathbb{V}(\hat{\theta}_n) + B_\theta(\hat{\theta}_n)^2$$

On considère qu'un estimateur $\hat{\theta}_n$ est **asymptotiquement meilleur** qu'un autre estimateur $\bar{\theta}_n$ si, en définissant $V_\theta(\hat{\theta})$ et $V_\theta(\bar{\theta})$ la variance limite de ces estimateurs *i.e.* : $V_\theta(\hat{\theta}) = \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\sqrt{n}\hat{\theta}_n)$ sous $\theta \in \Theta$ on a :

$$\begin{cases} \forall \theta \in \Theta, V_\theta(\hat{\theta}) \leq V_\theta(\bar{\theta}) \\ \exists \theta_0 \in \Theta, V_{\theta_0}(\hat{\theta}) < V_{\theta_0}(\bar{\theta}) \end{cases}$$

Pour comparer des estimateurs de manière non asymptotique, plusieurs moyens différents existent :

- **1er moyen, l'approche min-max Worst case risk**

On veut minimiser $R_\theta^*(\hat{\theta}_n) = \sup_{\theta \in \Theta} R_\theta(\hat{\theta}_n)$

On dit que $\hat{\theta}_n$ est **préférable** à $\bar{\theta}_n$ si $R_\theta^*(\hat{\theta}_n) < R_\theta^*(\bar{\theta}_n)$

On dit donc que $\hat{\theta}_n$ est **min-max optimal** si

$$\hat{\theta}_n \in \arg \min_{\bar{\theta}_n} R_\theta^*(\bar{\theta}_n)$$

- **2e moyen, l'approche bayésienne**

On introduit une fonction de poids $w : \Theta \rightarrow [0, 1]$ tel que le risque intégré est

$$R_\theta^B(\hat{\theta}_n) = \int_{\Theta} R_\theta(\hat{\theta}_n) w(\theta) d\theta$$

On a alors que $\hat{\theta}_n$ est préférable à $\bar{\theta}_n$ si

$$R_\theta^B(\hat{\theta}_n) < R_\theta^B(\bar{\theta}_n)$$

Efficacité asymptotique

On ne considère que les estimateurs asymptotiquement normaux, *i.e.* : tel que

$$\forall \theta \in \Theta, \sqrt{n}(\bar{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, V(\bar{\theta}))$$

Définition 14|14.4|2

Soit deux estimateurs, $\hat{\theta}_n, \bar{\theta}_n$ asymptotiquement normaux de variance limite $V_\theta(\hat{\theta}), V_\theta(\bar{\theta})$

On dit que $\hat{\theta}_n$ est **asymptotiquement meilleur** que $\bar{\theta}_n$ si

$$\begin{cases} V_{\theta}(\hat{\theta}) \leq V_{\theta}(\bar{\theta}), & \forall \theta \in \Theta \\ \exists \theta_0 \in \Theta \mid V_{\theta_0}(\hat{\theta}) < V_{\theta_0}(\bar{\theta}) \end{cases}$$

On dit qu'un estimateur est **asymptotiquement efficace** si il n'existe pas d'estimateur asymptotiquement meilleur que lui.

Définition 14|14.4|3

Dans un **modèle régulier**, un estimateur est **asymptotiquement efficace** si, et seulement si

$$V(\bar{\theta}_n) = \frac{1}{I(\theta)}$$

Avec $I(\theta)$ l'information de Fischer.

14.4.2 Estimateur de la Méthode des Moments

$X_1, \dots, X_n \stackrel{iid}{\sim} \mathbf{P}_{\theta*}$, des variables aléatoires dans \mathbb{R}^m , $\theta \in \Theta \subset \mathbb{R}^k$, $k \geq 1$

Soit $\varphi_1 \dots \varphi_k : \mathbb{R}^m \rightarrow \mathbb{R}$ tel que $\mathbb{E}_{\theta}[\varphi_j(X_i)] < \infty$, $\forall \theta \in \Theta$, $\forall i, j$

On pose

$$m_j(\theta) = \mathbb{E}_{\theta}[\varphi_j(X_1)] \text{ et } \hat{m}_j(\theta) = \frac{1}{n} \sum_{i=1}^n \varphi_j(X_i)$$

Définition 14|14.4|4

On dit que $\hat{\theta}^{MM}$ est un estimateur de la **méthode des moments** si il vérifie

$$\forall j \in \llbracket 1, k \rrbracket, \quad m_j(\hat{\theta}_n^{MM}) = \hat{m}_j$$

Autrement dit, pour $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}^k$, on a

$$M(\theta) = \mathbb{E}_{\theta}[\Phi(X_1)]$$

et

$$\widehat{M}_n = \frac{1}{n} \sum_{i=1}^n \Phi(X_i)$$

On a alors :

$$M(\widehat{\theta}_n^{MM}) = \widehat{M}_n$$

Théorème 14|14.4|1

Si $M : \Theta \rightarrow \mathbb{R}^k$ est une fonction continue et injective, et que $\widehat{M}_n \in \text{Im}(M)$ alors $\widehat{\theta}^{MM}$ existe et

$$\widehat{\theta}^{MM} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta$$

Preuve :

Comme M est continue et injective, on a $M : \Theta \rightarrow \text{Im}(M)$ qui est bijective et continue, donc M^{-1} est continue par théorème de la fonction réciproque

$$\widehat{M}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} M(\theta)$$

Donc

$$\widehat{\theta}_n^{MM} = M^{-1}(\widehat{M}_n) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta$$

Corollaire :

Si M est injective et $\widehat{M}_n \in \text{Im}(M)$, alors $\widehat{\theta}_n^{MM}$ existe et est unique.

$$\widehat{\theta}_n^{MM} = M^{-1}(\widehat{\theta}_n^{MM})$$

Théorème 14|14.4|2

Si $\mathbb{E}[\|\Phi(X_1)\|^2] < \infty$, $M \in \mathcal{C}^1(\Theta)$ et est injective, et que $\widehat{\theta}_n^{MM}$ existe, alors $\widehat{\theta}_n^{MM}$ est asymptotiquement normal.

Preuve :

M est \mathcal{C}^1 donc comme M est injective, M^{-1} l'est aussi,

Par TCl, on a

$$\sqrt{n}(\widehat{M}_n - \mathbb{E}[\Phi(X_1)]) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_k(0, \mathbb{V}(\Phi(X_1)))$$

Donc avec δ -méthode, on a :

$$\sqrt{n}(M^{-1}(\widehat{M}_n) - M^{-1}(\mathbb{E}[\Phi(X_1)])) = \sqrt{n}(\widehat{\theta}_n^{MM} - \theta) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_k[0, \nabla M^{-1}(\mathbb{E}[\Phi(X_1)])^T \mathbb{V}(\Phi(X_1)) \nabla M^{-1}(\mathbb{E}[\Phi(X_1)])]$$

14.4.3 Estimateur du Maximum de Vraisemblance

$$X_1, \dots, X_n \stackrel{iid}{\sim} \mathbb{P}_{\theta^*}, \quad \theta^* \in \Theta \subset \mathbb{R}^k$$

On se place tout le long de ce chapitre, dans un modèle $\{\mathbb{P}_{\theta}, \theta \in \Theta\}$ **dominé** $\mathbb{P}_{\theta^*} \ll \mu, \quad \forall \theta \in \Theta$.

Définition 14|14.4|5

On appelle **Vraisemblance** du modèle $\{\mathbb{P}_{\theta}, \theta \in \Theta\}$ la fonction

$$\begin{aligned} L_n : \quad \mathbb{R}^n \times \Theta &\mapsto \mathbb{R}_+ \\ (x_1, \dots, x_n, \theta) &\mapsto \prod_{i=1}^n f(x_i, \theta) \end{aligned}$$

Remarque :

Si les X_i ne sont pas indépendants, on a alors $L_n(x_1, \dots, x_n, \theta) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n, \theta)$

Définition 14|14.4|6

On dit que $\widehat{\theta}_n^{MV}$ est un **estimateur du maximum de vraisemblance** si

$$L_n(X_1, \dots, X_n, \widehat{\theta}_n^{MV}) = \max_{\theta \in \Theta} L_n(X_1, \dots, X_n, \theta)$$

Donc si $\widehat{\theta}_n^{MV} \in \arg \max_{\theta \in \Theta} L_n(X_1, \dots, X_n, \theta)$

Remarque :

Dans certains modèles, l'estimateur du maximum de vraisemblance n'existe pas ou n'est pas unique.

Définition 14|14.4|7

On appelle la **log-vraisemblance** la fonction

$$l_n(\theta) = -\frac{1}{n} \log(L_n(X_1, \dots, X_n, \theta)) = -\frac{1}{n} \sum_{i=1}^n \log(f(X_i, \theta))$$

D'où $\hat{\theta}_n^{MV} \in \arg \min_{\theta \in \Theta} l_n(\theta)$

Conditions

$$\int \log(f(x, \theta)) |f(x, \theta^*) d\mu(x) < \infty \quad \forall \theta, \theta^* \in \Theta$$

On pose $Z_i = -\log(f(X_i, \theta))$, $\bar{Z}_n \xrightarrow[n \rightarrow +\infty]{p.s.} -\mathbb{E}_{\theta^*}(Z_1)$ et $J(\theta) = \mathbb{E}_{\theta^*}(Z_1)$

Théorème 14|14.4|3

$$\bullet J(\theta) \geq J(\theta^*) \quad \forall \theta \in \Theta$$

$$\bullet J(\theta) = J(\theta^*) \iff \mathbb{P}_\theta = \mathbb{P}_{\theta^*} \text{ (si } \{\mathbb{P}_\theta, \theta \in \Theta\} \text{ identifiable alors } J(\theta) = J(\theta^*) \iff \theta = \theta^*)$$

Preuve :

$$\begin{aligned} J(\theta) &= \mathbb{E}_{\theta^*}[-\log(f(X_1, \theta))] \\ &= \mathbb{E}_{\theta^*}[-\log(f(X_1, \theta))] + \mathbb{E}_{\theta^*}[-\log(\frac{f(X_1, \theta)}{f(X_1, \theta^*)})] \\ &\geq J(\theta^*) - \log(\mathbb{E}_{\theta^*}[\frac{f(X_1, \theta)}{f(X_1, \theta^*)}]) \text{ (Jensen)} \\ &\geq J(\theta^*) \text{ Car } \mathbb{E}_{\theta^*}[\frac{f(X_1, \theta)}{f(X_1, \theta^*)}] = 1 \end{aligned}$$

D'où, $J(\theta) \geq J(\theta^*)$, $\forall \theta, \theta^* \in \Theta$. D'autres part,

$$J(\theta) - J(\theta^*) = -\mathbb{E}_{\theta^*}[\log(\frac{f(X_1, \theta)}{f(X_1, \theta^*)})] = -\mathbb{E}_{\theta^*}[\log(\frac{f(X_1, \theta)}{f(X_1, \theta^*)}) - \frac{f(X_1, \theta)}{f(X_1, \theta^*)} + 1]$$

Si $J(\theta) = J(\theta^*)$, alors

$$\mathbb{E}_{\theta^*}[\log(\frac{f(X_1, \theta)}{f(X_1, \theta^*)}) - \frac{f(X_1, \theta)}{f(X_1, \theta^*)} + 1] = 0$$

Donc comme $\log(x) \leq x - 1$ et $\log(x) = x - 1 \iff x = 1$, on a

$$\log(\frac{f(X_1, \theta)}{f(X_1, \theta^*)}) - \frac{f(X_1, \theta)}{f(X_1, \theta^*)} + 1 \stackrel{p.s.}{=} 0$$

et

$$f(X_1, \theta) = f(X_1, \theta^*) \iff \mathbb{P}_\theta = \mathbb{P}_{\theta^*}$$

Remarque :

L'idée donc derrière l'utilisation du maximum de vraisemblance comme estimateur du paramètre θ , est la suivante. Nous cherchons les paramètres qui "expliquent le mieux" les données observées. Cela revient à trouver les paramètres qui rendent l'événement observé le plus probable.

Théorème 14|14.4|4

Soit $\Theta \in \mathbb{R}$ un ouvert, de plus on a :

- i) $f(x, \theta) \in \mathcal{C}^0(\Theta), \quad \forall x \in \mathbb{R}$
- ii) Le modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est identifiable
- iii) $\int |\log(f(x, \theta))| f(x, \theta^*) d\mu(x) < \infty \quad \forall \theta, \theta^* \in \Theta$
- iv) L'ensemble des minima locaux de $l_n(\theta)$ forment un intervalle fermé de \mathbb{R}

Alors

$$\hat{\theta}_n^{MV} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta^*$$

Définition 14|14.4|8

On appelle **information de Fisher** dans un modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$

$$I(\theta) = \mathbb{E}_\theta[l'(x, \theta)^2] = \mathbb{E}_\theta[\nabla[l](x, \theta) \nabla[l](x, \theta)^T]$$

avec $l(x, \theta) = \log(f(X_1, \theta)), l'(x, \theta) = \frac{\partial l(x, \theta)}{\partial \theta}$

Si le modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est régulier, alors on note le **score** du modèle

$$s(x, \theta) = l'(x, \theta)$$

et on a :

$$I(\theta) = \mathbb{V}_\theta(s(X_1, \theta))$$

preuve :

Il suffit de montrer que $\mathbb{E}_\theta[s(X_1, \theta)] = 0$. On a :

$$\begin{aligned} & \int f(x, \theta) d\mu(x) = 1 \\ \iff & \frac{\partial}{\partial \theta} \int f(x, \theta) d\mu(x) = 0 \\ \text{(H2+H3)} \iff & \int \frac{\partial}{\partial \theta} f(x, \theta) d\mu(x) = 0 \\ \iff & \int \frac{f'(x, \theta)}{f(x, \theta)} f(x, \theta) d\mu(x) = 0 \\ \iff & \int l'(x, \theta) f(x, \theta) d\mu(x) = 0 \end{aligned}$$

Théorème 14|14.4|5

Dans un modèle régulier, on a

$$I(\theta) = -\mathbb{E}_\theta[l''(X_1, \theta)] = -\mathbb{E}_\theta[\nabla^2[l](X_1, \theta)]$$

preuve :

Par ce qui précède on a :

$$\begin{aligned} & \int l'(x, \theta) f(x, \theta) d\mu(x) = 0 \\ \text{par H3, H2} \iff & \int [l''(x, \theta) f(x, \theta) + l'(x, \theta) f'(x, \theta)] d\mu(x) = 0 \\ \iff & \mathbb{E}_\theta[l''(X_1, \theta)] + \int (l'(x, \theta))^2 f(x, \theta) d\mu(x) = 0 \\ \iff & I(\theta) = -\mathbb{E}_\theta[l''(X_1, \theta)] \end{aligned}$$

Théorème 14|14.4|6

normalité asymptotique de l'estimateur du maximum de vraisemblance

Soit $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta^*}, \mathbb{P}_{\theta^*} \ll \mu, \theta^* \in \Theta$ ouvert, $\Theta \subset \mathbb{R}$

Supposons que

1) $\hat{\theta}_n^{MV}$ existe

2) $\hat{\theta}_n^{MV}$ est consistant

3) Le modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est régulier et identifiable,

Alors on a

$$\sqrt{n}(\hat{\theta}_n^{MV} - \theta^*) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \frac{1}{I(\theta^*)})$$

Preuve :

Comme Θ est un ouvert, et $l(x, \theta)$ est \mathcal{C}^1 en θ , on a

$$\hat{\theta}_n^{MV} \in \arg \min_{\theta \in \Theta} l_n(\theta) \implies l'_n(\hat{\theta}_n^{MV}) = 0$$

On sait par TAF, comme $l_n(x, \theta) \in \mathcal{C}^2(\Theta)$ pour μ presque tout x , qu'il existe $\tilde{\theta}_n \in [\theta^*, \hat{\theta}_n^{MV}]$ tel que

$$l'_n(\hat{\theta}_n^{MV}) - l'_n(\theta^*) = l''_n(\tilde{\theta}_n)(\hat{\theta}_n^{MV} - \theta^*)$$

Donc

$$\sqrt{n}(\hat{\theta}_n^{MV} - \theta^*) = \sqrt{n} \frac{-l'(\theta^*)}{l''(\tilde{\theta}_n)}$$

Or on a

$$-\sqrt{n}l'_n(\theta^*) = \sqrt{n}(\frac{1}{n} \sum_{i=1}^n \log(f(X_i, \theta^*))') = \sqrt{n}(\frac{1}{n} \sum_{i=1}^n s(X_i, \theta^*) - \mathbb{E}_\theta[s(X_1, \theta)]) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \mathbb{V}_\theta(s(X_1, \theta^*)))$$

Par TCL en utilisant le fait que $\mathbb{E}_\theta[s(X_1, \theta)] = 0$ par régularité de $\{\mathbb{P}_\theta, \theta \in \Theta\}$

Donc on a

$$-\sqrt{n}l'_n(\theta^*) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, I(\theta^*))$$

on a aussi que $l''_n(\tilde{\theta}_n) = l''_n(\theta^*) + l''_n(\tilde{\theta}_n) - l''_n(\theta^*)$

or $l''_n(\theta^*) = \frac{1}{n} \sum_{i=1}^n l''(X_i, \theta^*) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}_\theta[l''(\theta^*)] = -I(\theta^*)$ par la loi forte des grands nombres

On a aussi que $l''_n(\tilde{\theta}_n) - l''_n(\theta^*) \xrightarrow[n \rightarrow +\infty]{P} 0$ donc, par slusky, on a

$$\sqrt{n}(\hat{\theta}_n^{MV} - \theta^*) = \sqrt{n} \frac{l'(\theta^*)}{l''(\hat{\theta}_n)} \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \frac{1}{I(\theta^*)})$$

Remarque :

- 1) Le théorème est valable même si $\hat{\theta}_n^{MV}$ n'est pas unique
- 2) Le théorème peut être étendu au cas où $\Theta \in \mathbb{R}^d$, $d \geq 2$ et on a

$$\sqrt{n}(\hat{\theta}_n^{MV} - \theta^*) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, I^{-1}(\theta^*))$$

Définition 14|14.4|9

Divergence de Kulback Leibler

Soit P, Q deux mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$

On appelle **divergence de Kulback Leibler** la quantité

$$D_{K,L}(P||Q) = \begin{cases} \int_{\mathcal{X}} \frac{dP}{dQ}(x) \log\left(\frac{dP}{dQ}(x)\right) dQ(x) & \text{si } P \ll Q \\ +\infty & \text{sinon} \end{cases}$$

Remarque :

Si $\exists \mu \mid P \ll \mu, Q \ll \mu$ alors on a

$$D_{K,L}(P||Q) = \int_{\mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) d\mu(x)$$

où $p(x) = \frac{dP}{d\mu}(x)$ et $q(x) = \frac{dQ}{d\mu}(x)$ respectivement les densités de P et Q par rapport à μ

Proposition 14|14.4|1

- 1) $D_{K,L}(P||Q) \geq 0$
- 2) $D_{K,L}(P||Q) = 0 \iff P = Q$
- 3) $D_{K,L}(P||Q) \neq D_{K,L}(Q||P)$ en générale

4) $\left(\frac{P}{Q}\right) \rightarrow D_{K,L}(P||Q)$ est convexe, *i.e.*

$$D_{K,L}(\alpha P_1 + (1 - \alpha)P_2 || \alpha Q_1 + (1 - \alpha)Q_2) \leq \alpha D_{K,L}(P_1 || Q_1) + (1 - \alpha)D_{K,L}(P_2 || Q_2)$$

Proposition 14|14.4|2

Dans un modèle régulier, l'information de Fisher est égale à la courbure de la divergence de Kulbak Leibler :

$$I(\theta) = \frac{\partial^2}{\partial \theta^2} D_{K,L}(\mathbb{P}_{\theta*} || \mathbb{P}_{\theta})$$

Théorème 14|14.4|7

Inégalité de Cramer-Rao

Si $\{\mathbb{P}_{\theta}, \theta \in \Theta\}$ est régulier et dominé alors pour tout estimateur $\bar{\theta}_n$ de θ fondé sur $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta*}$, on a :

$$\mathbb{E}[(\bar{\theta}_n - \theta)^2] \geq \frac{(1 + (B_{\theta}(\bar{\theta}_n))')^2}{nI(\theta)} + B_{\theta}(\bar{\theta}_n)^2$$

équivalent à

$$\mathbb{V}(\bar{\theta}_n) \geq \frac{(1 + (B_{\theta}(\bar{\theta}_n))')^2}{nI(\theta)}$$

Avec $B_{\theta}(\bar{\theta}_n)'$ la dérivée du biais de $\bar{\theta}_n$ par rapport à θ

Preuve :

Notons $s_n(X, \theta) = \sum_{i=1}^n l'(X_i, \theta)$. Par Cauchy Schwarz on a

$$\text{Cov}(s_n(X, \theta)) \leq \mathbb{V}(s_n(X, \theta)) \mathbb{V}(\bar{\theta}_n)$$

Premièrement, on a par régularité du modèle :

$$\mathbb{V}(s_n(X, \theta)) \stackrel{\text{iid}}{=} \sum_{i=1}^n \mathbb{V}(l'(X_i, \theta)) = nI(\theta) \quad (1)$$

Deuxièmement,

$$\begin{aligned}
\text{Cov}(s_n(X, \theta), \bar{\theta}_n) &= \mathbb{E}_\theta[s_n(X, \theta)\bar{\theta}_n] - \mathbb{E}_\theta[s_n(X, \theta)]\mathbb{E}_\theta[\bar{\theta}_n] \\
&= \mathbb{E}_\theta\left[\sum_{i=1}^n l'(X_i, \theta)\bar{\theta}_n\right] \\
&= \mathbb{E}_\theta\left[\sum_{i=1}^n \frac{\partial \log(f(X_i, \theta))}{\partial \theta} \bar{\theta}_n\right] \\
&= \mathbb{E}_\theta\left[\frac{\partial \log(L_n(X_1, \dots, X_n, \theta))}{\partial \theta} \bar{\theta}_n\right] \\
&= \int_{\mathcal{X}^n} \frac{\partial \log(L_n(X_1, \dots, X_n, \theta))}{\partial \theta} L_n(X_1, \dots, X_n, \theta) \bar{\theta}_n d\mu(x_1, \dots, x_n) \\
\text{Par H2, H3} &= \frac{\partial}{\partial \theta} \int_{\mathcal{X}^n} L_n(X_1, \dots, X_n, \theta) \bar{\theta}_n d\mu(x_1, \dots, x_n) \\
&= \frac{\partial}{\partial \theta} \mathbb{E}_\theta[\bar{\theta}_n] \\
&= \frac{\partial}{\partial \theta} (B_\theta(\bar{\theta}_n) + \theta) \quad (2) \\
&= (1 + B_\theta(\bar{\theta}_n)')
\end{aligned}$$

D'où en combinant (1) et (2) on a

$$\mathbb{V}(\bar{\theta}_n) \geq \frac{(1 + (B_\theta(\bar{\theta}_n))')^2}{nI(\theta)}$$

14.4.4 Estimateurs de Bayes

On se place dans un modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ **dominé**, $\Theta \in \mathbb{R}^p, p \in \mathbb{N}$ avec $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta^*}$ on notera $\vec{X} = (X_1, \dots, X_n)$, $\vec{x} = x_1, \dots, x_n$

On suppose cette fois-ci que θ suit une loi dite **à priori** Π_0 . On supposera toujours que Π_0 admet une densité par rapport à la mesure de Lebesgue et on notera, par abus de langage, Π_0 sa densité.

Définition 14|14.4|10

On dit que $\bar{\theta}_n^B$ est un **estimateur de Bayes** s'il minimise le risque intégré par rapport à une loi à priori Π_0

$$\bar{\theta}_n^B \in \arg \min_{\bar{\theta}_n} \int_{\Theta} \mathbb{E}_\theta[\|\bar{\theta}_n - \theta\|_2^2] d\Pi_0(\theta)$$

$$\bar{\theta}_n^B \in \arg \min_{\bar{\theta}_n} \int_{\Theta} \int_{\mathcal{X}^n} \|\bar{\theta}_n(\vec{x}) - \theta\|_2^2 L_n(\vec{x}, \theta) d\mu_n(\vec{x}) d\Pi_0(\theta)$$

Avec $d\mu_n(\vec{x}) = d\mu_1(x_1) \times \dots \times d\mu_n(x_n)$

Remarque :

C'est la définition de l'estimateur Bayésien quand le risque quadratique est utilisé (ce que l'on considèrera tout le long de cette section) mais de manière plus générale, on a

$$\bar{\theta}_n^B(\vec{X}) \in \arg \min_{\bar{\theta}_n} \int_{\Theta} \mathbb{E}_{\theta}[\varphi(\bar{\theta}_n, \theta)] \Pi_0(\theta) d\theta$$

pour φ une fonction mesurable qu'on choisit.

Théorème 14|14.4|8

Soit $\Pi_n(\theta)$ la densité donnée par

$$\Pi_n(\theta) \propto L_n(\vec{X}, \theta) \Pi_0(\theta) \quad \forall \theta \in \Theta$$

Alors, si on a $\int_{\Theta} \|\theta\|_2^2 \Pi_0(\theta) d\theta < +\infty$ Alors, l'estimateur Bayésien existe est donné par

$$\bar{\theta}_n^B = \int_{\Theta} \theta \Pi_n(\theta) d\theta$$

Interprétation

Si on considère que $L_n(\vec{X}, \theta)$ est la densité de $\vec{X}|\theta$ où $\theta \sim \Pi_0$, alors $\Pi_n(\theta)$ est la densité de $\theta|\vec{X}$

Par Bayes, on obtient facilement donc que $\Pi_n(\theta) \propto L_n(\vec{X}, \theta) \Pi_0(\theta)$, comme Π_n est une densité de probabilité, on a en fait une autre formule de Π_n donnée par

$$\Pi_n(\theta) = \frac{\Pi_0(\theta) L_n(\vec{X}, \theta)}{\int_{\Theta} \Pi_0(t) L_n(\vec{X}, t) dt}$$

Pour le choix de la loi à priori

Il n'y a pas de choix universel de Π_0 , ce dernier est toujours subjectif, il vaut donc mieux le choisir tel que

- 1) Π_0 a pour support Θ , avec Θ un ouvert
- 2) Choisir Π_0 tel que le calcul de $\bar{\theta}_n^B$ soit faisable

Définition 14|14.4|11

Soit \mathcal{P}_0 une famille de loi sur Θ . On dit que \mathcal{P}_0 est une **famille conjuguée** pour le modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ si $\forall \Pi_0 \in \mathcal{P}_0$, la loi a priori $\Pi_n \in \mathcal{P}_0$. (Par exemple $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{P}(\theta)$ et \mathcal{P}_0 la famille des loi Gamma)

Théorème 14|14.4|9

Dans un modèle régulier, si Π_0 a pour support Θ et si $\bar{\theta}_n^B$ est consistant, alors

$$\sqrt{n}(\bar{\theta}_n^B - \theta^*) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, \frac{1}{I(\theta^*)})$$

15 Test Statistique

15.1 Notion de Test

Définition 15|15.1|1

Soit $\{\mathbb{P}_\theta, \theta \in \Theta\}$ un modèle statistique sur \mathcal{X}^n , soient $\Theta_0, \Theta_1 \subset \Theta \mid \Theta_0 \cap \Theta_1 = \emptyset$

Durant cette partie, on notera $\mathcal{Z} = \mathcal{X}^n$ et $Z = (X_1(\omega), \dots, X_n(\omega)) = (x_1, \dots, x_n)$ (pour un certain $\omega \in \Omega$)

On dit que $H_0 : \theta \in \Theta_0$ est l'**hypothèse nulle** et que $H_1 : \theta \in \Theta_1$ est l'**hypothèse alternative**.

Un **test** de H_0 contre H_1 est une fonction mesurable

$$\phi : \mathcal{Z} \mapsto \{0, 1\}$$

tel que si $\phi(Z) = 1$ on dit que l'on **rejette** l'hypothèse nulle et si $\phi(Z) = 0$, on l'accepte.

Deux types d'erreurs :

$\phi \backslash \theta$	H_0	H_1
$\phi(Z) = 0$	Cool	Erreur de type II
$\phi(Z) = 1$	Erreur de type I	Cool

On cherche généralement à minimiser l'erreur de type I.

Définition 15|15.1|2

Soit $\phi : \mathcal{Z} \mapsto \{0, 1\}$ un test. La fonction de **puissance** de ϕ est la fonction

$$\begin{aligned} \beta_\phi : \Theta &\mapsto [0, 1] \\ \theta &\mapsto \mathbb{P}_\theta(\phi(Z) = 1) \end{aligned}$$

Idéalement, on aimerait que $\beta_\phi(\theta) = 0$ si $\theta \in \Theta_0$ et $\beta_\phi(\theta) = 1$ si $\theta \in \Theta_1$. Mais c'est impossible si les supports de lois de Θ_0 et de Θ_1 se recouvrent.

Autre objectif : pour un ϵ petit,

$$\begin{cases} \beta_\phi(\theta) \leq \epsilon, \forall \theta \in \Theta_0 \\ \beta_\phi(\theta) \geq 1 - \epsilon, \forall \theta \in \Theta_1 \end{cases}$$

C'est possible, si Θ_0, Θ_1 sont "séparés" (il y a une distance non nulle entre les deux).

Définition 15|15.1|3

Le **niveau** du test $\phi : \mathcal{Z} \mapsto \{0, 1\}$ est défini par :

$$\alpha(\phi) = \sup_{\theta \in \Theta_0} \beta_\phi(\theta) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi(Z) = 1)$$

On dit que ϕ est de niveau (au plus) α , si $\alpha(\phi) \leq \alpha$ (probabilité de type I $\leq \alpha$).

Point de vue de Neymann-Pierson :

Chercher un test de niveau α (souvent $\alpha = 10\%/5\%/1\%$) aussi puissant que possible sous H_1 .

Définition 15|15.1|4

Soit $(\phi_\alpha)_{\alpha \in [0,1]}$ une famille de tests $\phi_\alpha : \mathcal{Z} \mapsto \{0, 1\}$ de H_0 contre H_1 tel que

- ① ϕ_α est de niveau α , $\forall \alpha \in [0, 1]$.
- ② Pour tout $\alpha, \alpha' \in [0, 1]$ tel que $\alpha \leq \alpha'$, on a $\phi_\alpha \leq \phi_{\alpha'}$

Définition 15|15.1|5

On définit alors la **p-valeur** par

$$\hat{\alpha} = \hat{\alpha}(Z) = \inf \{ \alpha \in [0, 1] : \phi_\alpha(Z) = 1 \}$$

Donc on a $\forall \alpha \in [0, 1]$:

- $\hat{\alpha} < \alpha \implies \phi_\alpha(Z) = 1$
- $\hat{\alpha} > \alpha \implies \phi_\alpha(Z) = 0$

Attention : on calcul la p-valeur **uniquement** sous H_0 !

Proposition 15|15.1|1

- ① Sous H_0 , la p-valeur $\hat{\alpha}(Z)$ est "sur-uniforme" *i.e.* : $\forall \theta_0 \in \Theta_0$ on a

$$\mathbb{P}_{\theta_0}(\hat{\alpha}(Z) \leq t) \leq t, \forall t \in [0, 1]$$

- ② Si de plus, $\Theta_0 = \{\theta_0\}$ et si $\alpha(\phi_\alpha) = \alpha, \forall \alpha \in [0, 1]$, alors sous \mathbb{P}_{θ_0} on a

$$\hat{\alpha} \sim \mathcal{U}([0, 1])$$

Preuve :

- ① Soit $\alpha > t$, si $\hat{\alpha} \leq t$ alors $\hat{\alpha} \leq \alpha$, donc $\phi_\alpha(Z) = 1$ donc

$$\mathbb{P}_{\theta_0}(\hat{\alpha} \leq t) \leq \mathbb{P}_{\theta_0}(\phi_\alpha(Z) = 1) \stackrel{\text{par def}}{\leq} \alpha$$

Car ϕ_α est de niveau α . Ceci est vrai $\forall \alpha > t$ donc $\mathbb{P}_{\theta_0}(\hat{\alpha} \leq t) \leq t$
(si $\hat{\alpha} > t$, on a directement $\mathbb{P}_{\theta_0}(\hat{\alpha} \leq t) = 0 \leq t$)

- ② Si $\hat{\alpha} > t$, alors $\phi_t(Z) = 0$ donc

$$\mathbb{P}_{\theta_0}(\hat{\alpha} > t) \leq \mathbb{P}_{\theta_0}(\phi_t(Z) = 0) = 1 - t$$

Par hypothèse. Donc $\mathbb{P}_{\theta_0}(\hat{\alpha} \leq t) \geq t$ d'où on a $\hat{\alpha} \sim \mathcal{U}([0, 1])$

Définition 15|15.1|6

Un **test randomisé** de H_0 contre H_1 est une fonction

$$\phi : \mathcal{Z} \mapsto [0, 1]$$

Le résultat du test est obtenu de la façon suivante :

Soit $U \sim \mathcal{U}([0, 1])$ indépendant de Z . On prend $\hat{\Psi} = \mathbb{1}_{\{U \leq \phi(Z)\}}$. Alors

$$P(\hat{\Psi} = 1 | \phi(Z)) = P(U \leq \phi(Z) | \phi(Z)) = \phi(Z)$$

On définit

$$\beta_\phi(\theta) = \mathbb{P}_\theta(\hat{\Psi} = 1) = \mathbb{E}_\theta[P(\hat{\Psi} = 1 | \phi(Z))] = \mathbb{E}_\theta[\phi(Z)]$$

15.1.1 Régions de Confiance

Définition 15|15.1|7

Soit $\alpha \in (0, 1)$. Une **région de confiance** de niveau $1 - \alpha$ est une fonction

$$\begin{aligned} C_\alpha : \mathcal{Z} &\mapsto \mathcal{P}(\Theta) \text{ (ensemble des partie de } \Theta) \\ Z &\mapsto C_\alpha(Z) \end{aligned}$$

telle que en notant $\hat{C}_\alpha = C_\alpha(Z) \subset \Theta$ on a pour tout $\theta \in \Theta$

$$\mathbb{P}_\theta(\theta \in C_\alpha(Z)) \geq 1 - \alpha$$

Remarque :

$\theta \in \Theta$ est déterministe, la région \hat{C}_α est aléatoire.

Il y a une correspondance exacte entre :

- régions de confiance de niveau $1 - \alpha$
- tests d'adéquations : pour tout $\theta_0 \in \Theta$, le test $\phi_\alpha^{(\theta_0)} : \mathcal{Z} \mapsto \{0, 1\}$ de niveau α de $H_0^{(\theta_0)} : \theta = \theta_0$ contre $H_1^{(\theta_0)} : \theta \neq \theta_0$

(Les tests d'adéquations sont une classe spécifique de test)

Proposition 15|15.1|2

Lien régions de confiance/test d'adéquation

- ① Soit \hat{C}_α une région de confiance de niveau $1 - \alpha$. Alors pour tout $\theta_0 \in \Theta$, le test $\phi_\alpha^{(\theta_0)}(Z) = \mathbb{1}_{\{\theta_0 \notin \hat{C}_\alpha\}}$ est un test d'adéquation de niveau α
- ② Soit $(\phi_\alpha^{(\theta_0)})_{\theta_0 \in \Theta}$ une famille de test d'adéquation de niveau α . Alors

$$\hat{C}_\alpha = \{\theta_0 \in \Theta \mid \phi_\alpha^{(\theta_0)}(Z) = 0\}$$

est une région de confiance de niveau $1 - \alpha$. De plus, ces deux transformations sont inverses l'une de l'autre.

Preuve :

① Soit $\theta_0 \in \Theta$. Le niveau de $\phi_\alpha^{(\theta_0)}$ est donné par

$$\mathbb{P}_{\theta_0}(\phi_\alpha^{(\theta_0)}(Z) = 1) = \mathbb{P}_{\theta_0}(\theta_0 \notin \hat{C}_\alpha) = 1 - \mathbb{P}_{\theta_0}(\theta_0 \in \hat{C}_\alpha) \leq \alpha$$

② Pour tout $\theta_0 \in \Theta$. Le niveau de $\phi_\alpha^{(\theta_0)}$ est donné par

$$\mathbb{P}_{\theta_0}(\theta_0 \notin \hat{C}_\alpha) = \mathbb{P}_{\theta_0}(\phi_\alpha^{(\theta_0)}(Z) = 1) \leq \alpha$$

Méthode du pivot

Définition 15|15.1|8

Un **pivot** est une fonction $G : \mathcal{Z} \times \Theta \mapsto E$ telle que la loi de $G(Z, \theta)$ lorsque $Z \sim P_\theta$ ne dépend pas de θ

Remarque :

Avec les notations ci-dessus, notons Q la loi de $G(Z, \theta)$. Soit $B_\alpha \subset E$ tel que $Q(B) \geq 1 - \alpha$. Alors $\hat{C}_\alpha = \{\theta \in \Theta \mid G(Z, \theta) \in B_\alpha\}$ est une région de confiance de niveau $1 - \alpha$. En effet, pour tout $\theta \in \Theta$,

$$\mathbb{P}_\theta(\theta \in \hat{C}_\alpha) = \mathbb{P}_\theta(G(Z, \theta) \in B) = Q(B) \geq 1 - \alpha$$

15.2 Tests Uniformément Plus Puissants

Définition 15|15.2|1

Soient $\alpha \in (0, 1)$ et $\phi_\alpha : \mathcal{Z} \mapsto [0, 1]$. On dit que ϕ_α est **uniformément plus puissant de niveau α** (UPP- α) si

i) ϕ_α est de niveau α

ii) pour tout autre test ϕ de niveau α , ϕ_α est "meilleur que ϕ ", au sens où :

$$\forall \theta \in \Theta_1, \quad \beta_{\phi_\alpha}(\theta) \geq \beta_\phi(\theta)$$

Remarque :

Il n'existe pas nécessairement de test UPP- α de H_0 contre H_1 , mais nous allons voir des conditions dans lesquelles il en existe.

15.2.1 Dans le Cas de Deux Hypothèses Simples

On considère dans cette partie le cas $Z \sim P$, sur \mathcal{Z} , et on a $\Theta_0 = \{\theta_0\} = \{0\}$ et $\Theta_1 = \{\theta_1\} = \{1\}$
Avec les hypothèses : $H_0 : P = P_0$, $H_1 : P = P_1$.

Si $\phi : \mathcal{Z} \mapsto [0, 1]$ est un test randomisé, on a :

- $\alpha(\phi) = \mathbb{E}_0[\phi(Z)] = \beta_\phi(0)$
- $\beta(\phi) = \beta_\phi(1) = \mathbb{E}_1[\phi(Z)]$

On suppose que P_0 et P_1 admettent des densités $p_0, p_1 : \mathcal{Z} \mapsto \mathbb{R}_+$ par rapport à une même mesure μ sur \mathcal{Z} (on peut prendre par exemple $\mu = P_1 + P_0$).

On considère aussi la condition :

$$p_0, p_1 > 0, \mu.p.p. \tag{C}$$

On note que ϕ est UPP- α si :

- $\alpha(\phi) \leq \alpha$
- Pour tout ϕ' , tel que $\alpha(\phi') \leq \alpha$, on a $\beta(\phi) \geq \beta(\phi')$

si ϕ est une solution du problème d'optimisation sous contrainte :

$$\arg \max_{\substack{\phi: \mathcal{Z} \mapsto [0,1] \\ \int_{\mathcal{Z}} \phi p_0 d\mu \leq \alpha}} \int_{\mathcal{Z}} \phi p_1 d\mu$$

Par théorème de changement de variable on a :

$$\mathbb{E}_i[\phi(Z)] = \int_{\mathcal{Z}} \phi(z) dP_i(z) = \int_{\mathcal{Z}} \phi(z) p_i(z) d\mu(z)$$

Il est pour cela bon de randomiser car $\phi : \mathcal{Z} \mapsto [0, 1]$ espace convexe, ce qui permet d'avoir un problème d'optimisation convexe et linéaire.

Le problème est équivalent à

$$\arg \min_{\phi: \int \phi p_0 d\mu \leq \alpha} \left(- \int \phi p_1 d\mu \right) \quad (*)$$

Afin de résoudre (*), on considère la version régularisée/Lagrangienne étant donnée $\lambda \in \mathbb{R}_+$, on définit

$$E_\lambda(\phi) = - \int \phi p_1 d\mu + \lambda \int \phi p_0 d\mu$$

Lemme

Pour tout $\lambda \in \mathbb{R}_+$, et $q \in [0, 1]$, on définit le test

$$\phi_{\lambda,q}(z) = \mathbb{1}_{p_1(z) > \lambda p_0(z)} + q \mathbb{1}_{p_1(z) = \lambda p_0(z)} = \begin{cases} 1 & \text{si } p_1(z) > \lambda p_0(z) \\ q & \text{si } p_1(z) = \lambda p_0(z) \\ 0 & \text{sinon} \end{cases}$$

Alors, pour tout $q \in [0, 1]$, $\phi_{\lambda,q} \in \arg \min E_\lambda$

Réciproquement, si $\phi \in \arg \min_{\phi} E_\lambda(\phi)$, et si $p_0, p_1 > 0$, alors

$$\phi(z) = \begin{cases} 1 & \text{si } \frac{p_1(z)}{p_0(z)} > \lambda \\ 0 & \text{si } \frac{p_1(z)}{p_0(z)} < \lambda \end{cases}$$

Preuve :

Pour tout $\phi : \mathcal{Z} \mapsto [0, 1]$, on a

$$E_\lambda(\phi) = \int_{\mathcal{Z}} \phi(z) (\lambda p_0(z) - p_1(z)) d\mu(z)$$

Or, pour tout $z \in \mathcal{Z}$:

$$\arg \min_{\phi(z) \in [0,1]} \phi(z) (\lambda p_0(z) - p_1(z)) = \begin{cases} 0 & \text{si } \lambda p_0(z) - p_1(z) > 0 \\ 1 & \text{si } \lambda p_0(z) - p_1(z) < 0 \\ [0, 1] & \text{si } \lambda p_0(z) - p_1(z) = 0 \end{cases}$$

En particulier,

$$\phi(z) (\lambda p_0(z) - p_1(z)) \geq \phi_{\lambda,q}(z) (\lambda p_0(z) - p_1(z))$$

On notera $F(z) = \phi(z)(\lambda p_0(z) - p_1(z))$ et $G(z) = \phi_{\lambda,q}(z)(\lambda p_0(z) - p_1(z))$

Donc $E_\lambda(\phi) \geq E_\lambda(\phi_{\lambda,q})$ et donc

$$\phi_{\lambda,q} \in \arg \min E_\lambda$$

Pour la réciproque, on utilise que si $F(z) \geq G(z), \forall z \in \mathcal{Z}$, et si $\int F d\mu = \int G d\mu$, alors $F \stackrel{\mu.p.p.}{=} G$

Théorème 15|15.2|1

Neymann-Pearson

① Pour tout $\alpha \in [0, 1]$, il existe $\lambda \in \mathbb{R}^*$ et $q \in [0, 1]$ |

$$\alpha(\phi_{\lambda,q}) = \alpha$$

② Dans ce cas, $\phi_{\lambda,q}$ est UPP- α . De plus, tout test ϕ UPP- α est un minimiseur de E_λ , en particulier,

$$\mu.p.s., \phi(z) = \begin{cases} 1 & \text{si } p_1(z) > \lambda p_0(z) \\ 0 & \text{si } p_1(z) < \lambda p_0(z) \end{cases}$$

Preuve :

① On pose $F : \mathbb{R}_+ \mapsto [0, 1]$ | $F(\lambda) = P_0(p_1(Z) \leq p_0(Z)\lambda) = P_0(\frac{p_1(Z)}{p_0(Z)} \leq \lambda)$

Alors F est une fonction de répartition, donc F est croissante, cadlåg, et telle que $\lim_{\lambda \rightarrow +\infty} F(\lambda) = 1$, $F(\lambda) = 0$, si $\lambda < 0$

Remarque :

$$\begin{aligned} \alpha(\phi_{\lambda,q}) &= P_0(p_1(z) > p_0(z)\lambda) + qP_0(p_1(z) = p_0(z)\lambda) \\ &= 1 - F(\lambda) + q(F(\lambda) - F(\lambda_-)) \\ &\in [1 - F(\lambda), 1 - F(\lambda_-)] \end{aligned}$$

On pose $\lambda^* = \inf\{\lambda \in \mathbb{R}_+ : F(\lambda) \geq 1 - \alpha\} < \infty$

$F(\lambda) \xrightarrow{\lambda \rightarrow \infty} 1 > 1 - \alpha$ Alors, par définition : $F(\lambda) < 1 - \alpha, \quad \forall \lambda < \lambda^*$

Donc $F(\lambda_-^*) \leq 1 - \alpha$ i.e. : $1 - F(\lambda_-^*) \geq \alpha$. Par croissance de F , on a $F(\lambda) \geq 1 - \alpha$ pour tout $\lambda > \lambda^*$

Donc $F(\lambda^*) = \lim_{\substack{\lambda \rightarrow \lambda^* \\ \lambda > \lambda^*}} F(\lambda) \geq 1 - \alpha$ i.e. : $1 - F(\lambda^*) \leq \alpha$. On cherche $q \in [0, 1]$ tel que

$$\alpha(\phi_{\lambda^*, q}) = \alpha = 1 - F(\lambda^*) + q(F(\lambda^*) - F(\lambda_-^*)) \iff q = \frac{\alpha - (1 - F(\lambda^*))}{F(\lambda^*) - F(\lambda_-^*)}$$

② Soient $\lambda \in \mathbb{R}^*$ et $q \in [0, 1] \mid \alpha(\phi_{\lambda, q}) = \alpha$ montrons que $\phi_{\lambda, q}$ est UPP- α . Soit ϕ un autre test tel que $\alpha(\phi) \leq \alpha$, alors :

$$\begin{aligned} \beta(\phi) &= \lambda\alpha(\phi) - E_\lambda(\phi) \\ &\leq \lambda\alpha - E_\lambda(\phi_{\lambda, q}) \text{ (Lemme)} \\ &\leq \lambda\alpha(\phi_{\lambda, q}) - E_\lambda(\phi_{\lambda, q}) \\ &\leq \beta(\phi_{\lambda, q}) \end{aligned}$$

Donc $\phi_{\lambda, q}$ est UPP- α . Réciproquement, si ϕ est UPP- α , alors

$$\begin{aligned} E_\lambda(\phi) &= \lambda\alpha(\phi) - \beta(\phi) \\ &\leq \lambda\alpha - \beta(\phi_{\lambda, q}) \text{ (\phi est UPP-}\alpha\text{)} \\ &\leq E_\lambda(\phi_{\lambda, q}) = \inf E_\lambda \end{aligned}$$

Remarque :

Si la loi de $\frac{p_1(z)}{p_0(z)}$ sous $Z \sim P_0$ est \mathcal{C}^0 , alors il existe un test de rapport de vraisemblance **pur** UPP- α

$$\phi_\lambda(z) = \mathbb{1}_{p_1(z) > \lambda p_0(z)}$$

Sinon, il peut être nécessaire d'avoir recours à un test randomisé.

15.2.2 Test UPP Cas Général

Rappel : $\phi : \mathcal{Z} \mapsto [0, 1]$ est UPP- α si ϕ est de niveau α et si $\forall \phi'$ de niveau α ,

$$\forall \theta_1 \in \Theta_1, \quad \beta_\phi(\theta_1) = \mathbb{E}_{\theta_1}[\phi(Z)] \geq \beta_{\phi'}(\theta_1)$$

Observation :

On sait que si $\Theta_0 = \{\theta_0\}$ et si ϕ est UPP- α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \in \Theta_1$ alors ϕ est UPP- α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ pour tout $\theta_1 \in \Theta_1$

En particulier, par le cours précédent, il existe $\lambda \in \mathbb{R}_+$ tel que

$$\phi(z) = \begin{cases} 1 & \text{si } \frac{p_{\theta_1}(z)}{p_{\theta_0}(z)} > \lambda \\ 0 & \text{si } \frac{p_{\theta_1}(z)}{p_{\theta_0}(z)} < \lambda \end{cases} \mu.p.p.$$

Questions :

Si ϕ est UPP- α de $H_0 : \theta = \Theta_0$ contre $H_1 : \theta = \Theta_1$, est-il vrai que ϕ est aussi UPP- α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1, \forall \theta_0 \in \Theta_0, \theta_1 \in \Theta_1$?

La réponse est non. En effet, prenons $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\theta, 1)$, $H_0 : \theta \leq 0$ contre $H_1 : \theta > 0$

Alors $\phi(Z_n) = \mathbb{1}_{\bar{X}_n > \frac{q_{1-\alpha}}{\sqrt{n}}}$ est UPP- α , en revanche, si $\theta_0 < 0$ et $\theta_1 > 0$, ϕ n'est pas UPP- α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ car le test UPP- α est donné par

$$\phi'(Z_n) = \mathbb{1}_{\bar{X}_n > \theta_0 + \frac{q_{1-\alpha}}{\sqrt{n}}}$$

Définition 15|15.2|2

Soient $\{\mathbb{P}_\theta, \theta \in \Theta\}$ un modèle statistique identifiable et dominé par μ sur \mathcal{Z} . On dit que le modèle admet des **rapports de vraisemblance monotones** si $\Theta \subset \mathbb{R}$ et s'il existe une statistique $T : \mathcal{Z} \mapsto \mathbb{R}$ telle que pour tout $\theta_0, \theta_1 \in \Theta$ avec $\theta_0 < \theta_1$, il existe une fonction croissante $\Psi_{\theta_0, \theta_1} : \mathbb{R} \mapsto \mathbb{R}_+$ telle que

$$\frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)} = \Psi_{\theta_0, \theta_1}(T(Z)), \quad \forall Z \in \mathcal{Z}$$

En particulier, si $\Psi_{\theta_0, \theta_1}$ est strictement croissante, on a

$$\mathbb{1}_{\frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)} > \lambda} = \mathbb{1}_{T(Z) > \Psi_{\theta_0, \theta_1}^{-1}(\lambda)}$$

Lemme

Soit Y une variable aléatoire, et $f, g : \mathbb{R} \mapsto \mathbb{R}$ croissantes et positives (ou intégrables). Alors

$$\mathbb{E}[f(Y)g(Y)] \geq \mathbb{E}[f(Y)] \mathbb{E}[g(Y)]$$

Preuve :

Soit Y' une variable aléatoire indépendante de Y et de même loi.

En distinguant les cas $Y \leq Y'$ et $Y > Y'$, grâce à la croissance de f et g on prouve facilement l'inégalité :

$$(f(Y) - f(Y'))(g(Y) - g(Y')) \geq 0$$

D'où

$$\begin{aligned} 0 &\leq \mathbb{E}[(f(Y) - f(Y'))(g(Y) - g(Y'))] \\ &= \mathbb{E}[f(Y)g(Y)] + \mathbb{E}[f(Y')g(Y')] - \mathbb{E}[f(Y)g(Y')] - \mathbb{E}[f(Y')g(Y)] \\ &= 2(\mathbb{E}[f(Y)g(Y)] - \mathbb{E}[f(Y)] \mathbb{E}[g(Y)]) \end{aligned}$$

Ce qui démontre le résultat. (équivalent à dire que $\text{Cov}(f(Y), g(Y)) \geq 0$)

Lemme

Sous les conditions du théorème (qui suit), pour toute fonction $g : \mathbb{R} \mapsto \mathbb{R}$ croissante, la fonction $\theta \mapsto \mathbb{E}_\theta[g(T(Z))]$ est croissante. En particulier, $\beta_{\phi_{\lambda,q}}$ est croissante.

Preuve :

Soient $\theta_0 < \theta_1$, on a

$$\begin{aligned} \mathbb{E}_{\theta_1}[g(T(Z))] &= \int_{\mathcal{Z}} g(T(z)) p_{\theta_1}(z) d\mu(z) \\ &= \int_{\mathcal{Z}} g(T(z)) \frac{p_{\theta_1}(z)}{p_{\theta_0}(z)} p_{\theta_0}(z) d\mu(z) \\ &= \mathbb{E}_{\theta_0}[g(T(Z)) \frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)}] \\ &= \mathbb{E}_{\theta_0}[g(T(Z)) \Psi_{\theta_0, \theta_1}(T(Z))] \\ (\text{Lemme}) &\geq \mathbb{E}_{\theta_0}[g(T(Z))] \mathbb{E}_{\theta_0}[\Psi_{\theta_0, \theta_1}(T(Z))] \end{aligned}$$

Or

$$\mathbb{E}_{\theta_0}[\Psi_{\theta_0, \theta_1}(T(Z))] = \int_{\mathcal{Z}} \frac{p_{\theta_1}(z)}{p_{\theta_0}(z)} p_{\theta_0}(z) d\mu(z) = 1$$

Donc $\mathbb{E}_{\theta_1}[g(T(Z))] \geq \mathbb{E}_{\theta_0}[g(T(Z))]$, d'où la croissance de $\theta \mapsto \mathbb{E}_\theta[g(T(Z))]$.

Pour la deuxième assertion, on note $\phi_{\lambda,q}(Z) = g_{\lambda,q}(T(Z))$, où $g_{\lambda,q}(t) = \mathbb{1}_{\{t > \lambda\}} + q \mathbb{1}_{\{t = \lambda\}}$ croissante

Théorème 15|15.2|2

Neymann-Pierson

Soit $\{\mathbb{P}_\theta, \theta \in \Theta\}$ un modèle admettant des rapports de vraisemblance monotones pour la statistique T . On cherche à tester $H_0 : \theta < \theta_0$ contre $H_1 : \theta > \theta_0$. Alors

1) Il existe $\lambda \in \mathbb{R}$ et $q \in [0, 1]$ tels qu'en notant $\phi_{\lambda,q}(z) = \mathbb{1}_{\{T(z) > \lambda\}} + q\mathbb{1}_{\{T(z) = \lambda\}}$, on a :

$$\beta_{\phi_{\lambda,q}}(\theta_0) = P_{\theta_0}(T(Z) > \lambda) + qP_{\theta_0}(T(Z) = \lambda) = \alpha$$

2) Dans ce cas, $\phi_{\lambda,q}$ est un test UPP- α .

Preuve :

① Existence de $\lambda \in \mathbb{R}, q \in [0, 1] \mid \beta_{\phi_{\lambda,q}}(\theta_0) = \alpha$ est admise, car la preuve est identique à celle faite précédemment.

② Montrons que $\phi_{\lambda,q} = \mathbb{1}_{\{T(Z) > \lambda\}} + q\mathbb{1}_{\{T(Z) = \lambda\}}$ est UPP- α

Ⓐ Montrons d'abord que $\alpha(\phi_{\lambda,q}) \leq \alpha$.

Pour tout $\theta \leq \theta_0$, par le lemme, on a

$$\beta_{\phi_{\lambda,q}}(\theta) \leq \beta_{\phi_{\lambda,q}}(\theta_0) = \alpha$$

Ⓑ Montrons que $\phi_{\lambda,q}$ est UPP- α

Soit $\theta_1 > \theta_0$, soit ϕ un autre test de niveau α pour Θ_0 . On va montrer que $\phi_{\lambda,q}$ est un test de rapport de vraisemblance. On pose $\lambda(\theta_1) = \Psi_{\theta_0, \theta_1}(\lambda)$.

- Si $\frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)} > \lambda(\theta_1)$, alors $T(Z) > \lambda$ donc $\phi_{\lambda,q}(Z) = 1$.
- Si $\frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)} < \lambda(\theta_1)$, alors $T(Z) < \lambda$ donc $\phi_{\lambda,q}(Z) = 0$.

De plus, $\alpha(\phi_{\lambda,q}) = \alpha$, donc, par le cours précédent, $\phi_{\lambda,q}$ est UPP- α de $\{\theta_0\}$ contre $\{\theta_1\}$.
Donc $\beta_{\phi,q}(\theta_1) \geq \beta_{\phi,q}(\theta_0)$

Remarque :

Si $\Theta \subset \mathbb{R}$, $(\mathbb{P}_\theta)_{\theta \in \Theta}$ modèle avec des rapports de vraisemblances monotones en $T(Z)$, alors le modèle

$(\mathbb{P}_\theta)_{\theta \in \Theta}$ admet des rapports des rapports de vraisemblances monotones en $-T(Z)$, et c'est un test UPP- α de $H_0 : \theta \geq \theta_0$ contre $H_1 : \theta < \theta_0$ sous la forme

$$\phi(Z) = \mathbb{1}_{\{T(Z) < c\}} + q\mathbb{1}_{\{T(Z) = c\}}$$

Remarque :

$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$

Il n'y a pas de test UPP- α . En effet, si ϕ est UPP- α , alors

• ϕ est UPP- α de $H_0 : \theta = 0$ contre $H'_1 : \theta > 0$. Par l'unicité dans Neymann-Pierson,

$$\phi(Z) = \mathbb{1}_{\{\overline{X_n} > \frac{q_{1-\alpha}}{\sqrt{n}}\}} \quad \text{p.s. sous } P_0 \quad (1)$$

• ϕ est UPP- α de $H_0 : \theta = 0$ contre $H''_1 : \theta < 0$. Par le même raisonnement, on a, presque sûrement :

$$\phi(Z) = \mathbb{1}_{\{\overline{X_n} < \frac{q_{1-\alpha}}{\sqrt{n}}\}} \quad (2)$$

(1) et (2) sont contradictoires (valeurs différentes) si $Z > \frac{q_{1-\alpha}}{\sqrt{n}}$, de probabilité > 0 .

15.2.3 Test de Rapport de Vraisemblance

Méthode générale pour construire un test de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$

\mathbb{P}_θ admet une densité p_θ par rapport à μ sur \mathcal{Z}

Définition 15|15.2|3

Un test de **rapport de vraisemblance** de Θ_0 contre Θ_1 est un test de la forme $\phi(Z) = \mathbb{1}_{\{T(Z) > c\}}$

$$T(Z) = \frac{\sup_{\theta \in \Theta_1} p_\theta(Z)}{\sup_{\theta \in \Theta_0} p_\theta(Z)}$$

Remarque :

On peut aussi considérer la statistique

$$\tilde{T}(Z) = \frac{\sup_{\theta \in \Theta} p_{\theta}(Z)}{\sup_{\theta \in \Theta_0} p_{\theta}(Z)} = \frac{\max(\sup_{\theta \in \Theta_0} p_{\theta}(Z), \sup_{\theta \in \Theta_1} p_{\theta}(Z))}{\sup_{\theta \in \Theta_0} p_{\theta}(Z)}$$

$$\begin{aligned}\tilde{T}(Z) &= \max(1, \frac{\sup_{\theta_1 \in \Theta} p_{\theta_1}(Z)}{\sup_{\theta \in \Theta_0} p_{\theta}(Z)}) \\ &= \max(1, T(Z))\end{aligned}$$

Si $c \geq 1$, $\tilde{T}(Z) > c$ si, et seulement si $T(Z) > c$.

En théorie, on détermine $c_{\alpha} > 0$ tel que le test soit de niveau α . (Si $c_{\alpha} > 1$, on peut remplacer $T(Z)$ par $\tilde{T}(Z)$)

15.2.4 Cas des Modèles à Rapports de Vraisemblances Monotones

On suppose que $\Theta \subset \mathbb{R}$ et que $\forall \theta_0, \theta_1 \in \Theta_1, \quad \theta_0 \leq \theta_1$,

$$\frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)} = \Psi_{\theta_0, \theta_1}(S(Z))$$

$H_0 : \theta \leq \bar{\theta}$ contre $H_1 : \theta > \bar{\theta}$.

$$\begin{aligned}T(Z) &= \frac{\sup_{\theta_1 > \bar{\theta}} p_{\theta_1}(Z)}{\sup_{\theta_0 \leq \bar{\theta}} p_{\theta_0}(Z)} \quad (\theta_0 \leq \bar{\theta} < \theta_1) \\ &= \sup_{\theta_1 > \bar{\theta}} \inf_{\theta_0 \leq \bar{\theta}} \frac{p_{\theta_1}(Z)}{p_{\theta_0}(Z)} \\ (\text{RVM}) &= \sup_{\theta_1 > \bar{\theta}} \inf_{\theta_0 \leq \bar{\theta}} \Psi_{\theta_0, \theta_1}(S(Z)) \\ &= \Psi(S(Z))\end{aligned}$$

où $\Psi(s) = \sup_{\theta_1 > \bar{\theta}} \inf_{\theta_0 \leq \bar{\theta}} \Psi_{\theta_0, \theta_1}(s)$ croissante, car $\Psi_{\theta_0, \theta_1}$ l'est.

Tests de rapports de vraisemblances : $\mathbb{1}_{\{\Psi(S(Z)) > t\}}, t = \Psi(c) \iff \mathbb{1}_{\{S(Z) > c\}}$ On retrouve le test UPP- α .

Test d'adéquation

$H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$,

$$\tilde{T}(Z) = \frac{\sup_{\theta \in \Theta} p_{\theta}(Z)}{\sup_{\theta \in \Theta_0} p_{\theta}(Z)} = \frac{p_{\hat{\theta}}(Z)}{p_{\theta_0}(Z)}$$

où $\hat{\theta} = \arg \max_{\theta \in \Theta} p_{\theta}(Z)$ (estimateur du maximum de vraisemblance)

15.3 Test Asymptotiques

Dans l'analyse statistique, la connaissance de la loi de probabilité exacte d'une statistique de test est souvent un luxe que la réalité des modèles complexes ne nous offre pas. C'est dans ce contexte que les tests asymptotiques deviennent des outils indispensables. Ils nous permettent de contourner l'absence d'une distribution explicite en s'appuyant sur le comportement de la statistique lorsque la taille de l'échantillon tend vers l'infini, offrant ainsi une méthode robuste pour l'inférence statistique même face à l'incertitude théorique.

En effet, illustrons le problème avec un exemple concret. Soit

$$\phi(Z) = \mathbb{1}_{\left\{ \frac{p_{\hat{\theta}_n^{MV}}(Z)}{p_{\theta_0}(Z)} > c_{\alpha}(\theta_0) \right\}}$$

Test de rapport de vraisemblance où $c_{\alpha}(\theta_0)$ quantile d'ordre $(1 - \alpha)$ de $\frac{p_{\hat{\theta}_n^{MV}}(Z)}{p_{\theta_0}(Z)} > c_{\alpha}(\theta_0)$ sous $Z \sim \mathbb{P}_{\theta_0}$

Problème :

Il nous faut connaître $c_{\alpha}(\theta_0)$... On a, comme région de confiance :

$$\hat{R}_{\alpha} = \left\{ \theta_0 \in \Theta : \frac{p_{\hat{\theta}(Z)}(Z)}{p_{\theta_0}(Z)} \leq c_{\alpha}(\theta_0) \right\}$$

Ainsi, afin de contourner ces difficultés, il peut être utile d'avoir recours à des approximations.

Déterminons maintenant le cadre asymptotique dans lequel nous nous placerons de cette sous-partie. On pose $\mathcal{Z} = \mathcal{Z}_n = \mathcal{X}^n$, $Z_n = (X_1, \dots, X_n)$, où X_1, \dots, X_n est iid de loi \mathbb{P}_{θ} sur \mathcal{X} .

$$\begin{aligned} p_{\theta} : \mathcal{X} &\mapsto \mathbb{R}_+ \\ p_{\theta}(x_1, \dots, x_n) &\mapsto \prod_{i=1}^n p_{\theta}(x_i) \end{aligned}$$

Avec $p_{\theta} = \frac{d\mathbb{P}_{\theta}}{d\mu}$. On a un modèle $\{\mathbb{P}_{\theta}, \theta \in \Theta\}$, avec Θ_0, Θ_1 et $\theta_0 \in \Theta_0$, fixés puis $n \rightarrow \infty$.

Définition 15|15.3|1

On dit qu'une famille $(\phi_n)_{n \geq 1}$ de tests où $\phi_n : \mathcal{X}^n \mapsto \{0, 1\}$ test de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ est **asymptotiquement de niveau α** si pour tout $\theta_0 \in \Theta_0$ on a :

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup \beta_{\phi_n}(\theta_0) \\ &= \lim_{n \rightarrow \infty} \sup \mathbb{P}_{\theta_0}(\phi_n(X_1, \dots, X_n) = 1) \leq \alpha \end{aligned}$$

On suppose que $\{\mathbb{P}_\theta : \theta \in \Theta\}$ est un modèle régulier donc $l(x, \theta) = \log(p_\theta(x))$ de classe \mathcal{C}^2 en θ .

Rappelons aussi que $I(\theta) = \mathbb{E}_\theta[\nabla l(X_1, \theta) \nabla l(X_1, \theta)^T] = -\mathbb{E}_\theta[\nabla^2 l(X, \theta)]$. On notera ici $l_n(\theta) = \log(p_\theta(X_1, \dots, X_n)) = \sum_{i=1}^n l(X_i, \theta)$, et on a donc $\hat{\theta}_n^{MV} \in \arg \max_{\theta \in \Theta} l_n(\theta)$.

On pose l'hypothèse (H) :

Il existe une fonction $M : \mathcal{X} \mapsto \mathbb{R}_+$ tel que

- pour tout $\theta, \theta' \in \Theta, x \in \mathcal{X}$, on a :

$$\|\nabla^2 l(x, \theta) - \nabla^2 l(x, \theta')\|_F \leq M(x) \|\theta - \theta'\|_2 \quad (\|\cdot\|_F \text{ est la norme de Frobenius})$$

- Pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[M(X)] < +\infty$

Théorème 15|15.3|1

Théorème de Wilks

Soit $\{\mathbb{P}_\theta, \theta \in \Theta\}$ un modèle régulier avec $\Theta \subset \mathbb{R}^d$ ouvert et convexe, on suppose l'hypothèse H.

Également, on suppose que $\hat{\theta}_n^{MV} \xrightarrow[n \rightarrow +\infty]{P} \theta_0$ et $\sqrt{n}(\hat{\theta}_n^{MV} - \theta_0) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, I^{-1}(\theta_0))$. Alors, pour tous $\theta_0 \in \Theta$ sous \mathbb{P}_{θ_0} , avec $\hat{\theta}_n^{MV} \in \arg \max_{\theta} l_n$, on a

$$\begin{aligned} & 2 \log\left(\frac{p_{\hat{\theta}_n^{MV}}(X_1, \dots, X_n)}{p_{\theta_0}(X_1, \dots, X_n)}\right) \\ &= 2(l_n(\hat{\theta}_n^{MV}) - l_n(\theta_0)) \xrightarrow[n \rightarrow +\infty]{(d)} \chi^2(d) \end{aligned}$$

En particulier, pour un test asymptotique de niveau α d'adéquation à θ_0 :

$$\phi_{n,\alpha}(X_1, \dots, X_n) = \mathbb{1}_{\{2(l_n(\hat{\theta}_n) - l_n(\theta_0)) > c_{1-\alpha}(\chi^2(d))\}}$$

asymptotique de niveau α pour $H_0 : \theta = \theta_0$.

Preuve :

Posons

$$\begin{aligned} \phi : [0, 1] &\mapsto \mathbb{R} \\ t &\mapsto l_n((1-t)\hat{\theta}_n^{MV} + t\theta_0) \end{aligned}$$

Comme le modèle est régulier, l_n est $\mathcal{C}^2(\Theta)$ donc ϕ l'est aussi. Ainsi, par théorème de Taylor Lagrange, il existe $\tilde{\theta}_n \in [\theta_0, \hat{\theta}_n^{MV}]$ (il existe donc $t \in [0, 1]$, $\tilde{\theta}_n = t\hat{\theta}_n^{MV} + (1-t)\theta_0$) tel que

$$\begin{aligned} \phi(1) &= \phi(0) + \phi'(0) + \frac{1}{2}\phi''(\tilde{\theta}_n) \\ \iff l_n(\theta_0) - l_n(\hat{\theta}_n^{MV}) &= \overbrace{< \nabla l_n(\hat{\theta}_n^{MV}), \theta_0 - \hat{\theta}_n^{MV} >}^{=0} + \frac{1}{2}(\theta - \hat{\theta}_n^{MV})^T \nabla^2 l_n(\tilde{\theta}_n)(\theta - \hat{\theta}_n^{MV}) \\ \iff l_n(\hat{\theta}_n^{MV}) - l_n(\theta_0) &= \frac{1}{2}(\hat{\theta}_n^{MV} - \theta_0)^T (-\nabla^2 l_n(\tilde{\theta}_n))(\hat{\theta}_n^{MV} - \theta_0) \end{aligned}$$

Mais on sait que l'on a $\sqrt{n}(\hat{\theta}_n^{MV} - \theta_0) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, I^{-1}(\theta_0))$, avec $I(\theta_0)$ l'information de Fisher, équivalent à

$$\sqrt{n}I^{\frac{1}{2}}(\theta_0)(\hat{\theta}_n^{MV} - \theta_0) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}_d(0, I_d)$$

donc, par Cochran on a

$$\|\sqrt{n}I^{\frac{1}{2}}(\theta_0)(\hat{\theta}_n^{MV} - \theta_0)\|_2^2 = n(\hat{\theta}_n^{MV} - \theta_0)^T I(\theta_0)(\hat{\theta}_n^{MV} - \theta_0) \xrightarrow[n \rightarrow +\infty]{(d)} \chi^2(d)$$

Mais, on peut aussi écrire

$$(\hat{\theta}_n^{MV} - \theta_0)^T [-\nabla^2 l_n(\tilde{\theta}_n)](\hat{\theta}_n^{MV} - \theta_0) = n(\hat{\theta}_n^{MV} - \theta_0)^T \frac{1}{n} [-\nabla^2 l_n(\tilde{\theta}_n)](\hat{\theta}_n^{MV} - \theta_0)$$

Or, en rappelant que l'estimateur du maximum de vraisemblance est consistant, on a

$$\begin{aligned}
\frac{1}{n} \|\nabla^2 l_n(\tilde{\theta}_n) - (-\nabla^2 l_n(\theta_0))\|_F &= \frac{1}{n} \left\| \sum_{i=1}^n \nabla^2 l(X_i, \tilde{\theta}_n) - \nabla^2 l(X_i, \theta_0) \right\|_F \\
&\leq \frac{1}{n} \sum_{i=1}^n \|\nabla^2 l(X_i, \tilde{\theta}_n) - \nabla^2 l(X_i, \theta_0)\|_F \\
&\stackrel{(H)}{\leq} \frac{1}{n} \sum_{i=1}^n M(X_i) \|\tilde{\theta}_n - \theta_0\|_2 \\
&\leq |t| \underbrace{\|\hat{\theta}_n^{MV} - \theta_0\|_2}_{\xrightarrow[n \rightarrow +\infty]{P} 0} \underbrace{\frac{1}{n} \sum_{i=1}^n M(X_i)}_{\xrightarrow[n \rightarrow +\infty]{P} \mathbb{E}[M(X_1)] < \infty} \xrightarrow[n \rightarrow +\infty]{P} 0
\end{aligned}$$

Comme $-\frac{1}{n} \nabla^2 l_n(\theta_0) \xrightarrow[n \rightarrow +\infty]{P} \mathbb{E}[-\nabla^2 l(\theta_0, X_1)] = I(\theta_0)$ on a donc bien que $-\frac{1}{n} \nabla^2 l_n(\tilde{\theta}_n) \xrightarrow[n \rightarrow +\infty]{P} I(\theta_0)$ aussi.

$$\begin{aligned}
2(l_n(\hat{\theta}_n^{MV}) - l_n(\theta_0)) &= (\hat{\theta}_n^{MV} - \theta_0)^T (-\nabla^2 l_n(\tilde{\theta}_n)) (\hat{\theta}_n^{MV} - \theta_0) \\
&= n(\hat{\theta}_n^{MV} - \theta_0)^T \underbrace{\left(-\frac{1}{n} \nabla^2 l_n(\tilde{\theta}_n)\right)}_{\xrightarrow[n \rightarrow +\infty]{P} I(\theta_0)} (\hat{\theta}_n^{MV} - \theta_0) \underbrace{\xrightarrow[n \rightarrow +\infty]{(d)} \chi^2(d)}_{\text{(Slutsky)}}
\end{aligned}$$

Remarque :

Test asymptotiques d'adéquation à θ_0 : $\phi_\alpha^{(\theta_0)}(X_1, \dots, X_n) = \mathbb{1}_{\{2(l_n(\hat{\theta}_n^{MV}) - l_n(\theta_0)) > q_{1-\alpha}(\chi^2(d))\}}$

Région de confiance : $1 - \alpha \quad \hat{C}_\alpha = \{\theta_0 \in \Theta : l_n(\hat{\theta}_n^{MV}) - l_n(\theta_0) \leq \frac{q_{1-\alpha}}{2}\}$

15.3.1 Tests de Wald

Définition 15|15.3|2

On définit le **test de Wald I**

$$\phi_\alpha(X_1, \dots, X_n) = \mathbb{1}_{\{T_1 > q_{1-\alpha}(\chi^2(d))\}}$$

Avec

$$T_1 = n \left\langle I(\theta_0)(\hat{\theta}_n^{MV} - \theta_0), \hat{\theta}_n^{MV} - \theta_0 \right\rangle \xrightarrow[n \rightarrow +\infty]{(d)} \chi^2(d) \quad (1)$$

Et le **test Wald II** par

$$\phi'_\alpha(X_1, \dots, X_n) = \mathbb{1}_{\{T_2 > q_{1-\alpha}(\chi^2(d))\}}$$

avec

$$T_2 = n \left\langle I(\hat{\theta}_n^{MV})(\hat{\theta}_n^{MV} - \theta_0), (\hat{\theta}_n^{MV} - \theta_0) \right\rangle \xrightarrow[n \rightarrow +\infty]{(d)} \chi^2(d) \quad (2)$$

Proposition 15|15.3|1

Ces deux tests sont asymptotiquement de niveau α , on a

- Pour (1), vu en preuve du théorème de Wilks.

- Pour (2) On utilise (1) et le fait que $\|I(\hat{\theta}_n^{MV}) - I(\theta_0)\|_F \xrightarrow[n \rightarrow +\infty]{P} 0$ ce qui démontre le résultat attendu.

Le test de Wald II est utile pour déterminer des régions de confiance

$$T_2 = \|I(\hat{\theta}_n^{MV})^{\frac{1}{2}} \sqrt{n}(\hat{\theta}_n^{MV} - \theta_0)\| \leq q_{1-\alpha}(\chi^2(d)) = q_{1-\alpha}$$

si, et seulement si $I(\hat{\theta}_n^{MV})^{\frac{1}{2}} \sqrt{n}(\hat{\theta}_n^{MV} - \theta_0) \in B(0, q_{1-\alpha}) = q_{1-\alpha} B_2$ avec B_2 la boule unité de norme 2

si, et seulement si $\theta_0 \in \underbrace{\hat{\theta}_n^{MV} + \frac{I(\hat{\theta}_n^{MV})^{-\frac{1}{2}}}{\sqrt{n}} q_{1-\alpha} B_2}_{\hat{C}_{1-\alpha} \text{ ellipsoïde centre en } \hat{\theta}_n^{MV}}$.

Remarque :

Sous \mathbb{P}_{θ_0} , en posant $T_0 = 2(l_n(\hat{\theta}_n^{MV}) - l_n(\theta_0))$ on a

$$|T_0 - T_1| \xrightarrow[n \rightarrow +\infty]{P} 0 \quad |T_1 - T_2| \xrightarrow[n \rightarrow +\infty]{P} 0 \quad \text{sous } H_0 : \theta = \theta_0$$

Ces trois tests (Wilks, Wald I, Wald II), sont asymptotiquement équivalents sous $H_0 : \theta = \theta_0$, mais pas forcément sous $H_1 : \theta \neq \theta_0$. C'est pareil pour leurs régions de confiance associées d'ailleurs.

Par exemple, si $\theta \neq \theta_0$, sous \mathbb{P}_θ ,

$$T_2 \approx n < I(\underbrace{\theta}_{\text{pas } \theta_0})(\hat{\theta}_n^{MV} - \theta_0), \hat{\theta}_n^{MV} - \theta_0 > \approx n < I(\theta)(\theta - \theta_0), \theta - \theta_0 >$$

15.3.2 Lois Discrètes : Test du χ^2

Soient $d \geq 1$ et $\mathcal{X} = \{0, 1, \dots, d\}$. On pose $\Theta = \{(\theta_1, \dots, \theta_d) \in \mathbb{R}_+^* \mid \sum_{i=1}^d \theta_i < 1\}$. Pour tout $\theta \in \Theta$, on note P_θ la loi de probabilité sur \mathcal{X} définie par :

$$\forall j = 1, \dots, d, \quad P_\theta(\{j\}) = \theta_j \quad \text{et} \quad P_\theta(\{0\}) = 1 - \sum_{j=1}^d \theta_j$$

Notations : Pour tout $\theta, \theta' \in \Theta$, on pose $\chi^2(\theta, \theta') := \sum_{j=0}^d \frac{\theta_j - \theta'_j}{\theta_j}$

Proposition 15|15.3|2

Pour tout $\theta, \theta' \in \Theta = \{(\theta_1, \dots, \theta_d) \in \mathbb{R}_+^* \mid \sum_{i=1}^d \theta_i < 1\}$ si on note P_θ la mesure de probabilité $\sum_{i=0}^d \theta_i \delta_i$ sur $\mathcal{X} = \llbracket 1, d \rrbracket$ alors on a :

$$\|I(\theta)^{\frac{1}{2}}(\theta' - \theta)\|_2^2 = \chi^2(\theta, \theta')$$

Preuve :

On a donc la densité de P_θ qui est $p_\theta(x) = \sum_{i=0}^d \theta_i \mathbb{1}_{x=i}, \forall x \in \llbracket 1, d \rrbracket$.

on sait que $l(x, \theta) = \log \left(\sum_{i=0}^d \theta_i \mathbb{1}_{x=i} \right) = \log \left((1 - \sum_{j=1}^d \theta_j) \mathbb{1}_{x=0} + \sum_{i=1}^d \theta_i \mathbb{1}_{x=i} \right)$

Donc on a, pour $k \in \llbracket 1, d \rrbracket$, $\frac{\partial l(x, \theta)}{\partial \theta_k} = \frac{-\mathbb{1}_{x=0} + \mathbb{1}_{x=k}}{\sum_{i=0}^d \theta_i \mathbb{1}_{x=i}}$

Donc, pour $j, k \in \llbracket 1, d \rrbracket$ et $X \sim P_\theta$, on a :

$$\begin{aligned}
I(\theta)_{j,k} &= \mathbb{E}_\theta \left[\left(\frac{\partial l(X, \theta)}{\partial \theta} \right)_j \left(\frac{\partial l(X, \theta)}{\partial \theta} \right)_k^T \right] \\
&= \mathbb{E}_\theta \left[\frac{1}{\left(\sum_{i=0}^d \theta_i \mathbb{1}_{X=i} \right)^2} (\mathbb{1}_{X=j} - \mathbb{1}_{X=0})(\mathbb{1}_{X=k} - \mathbb{1}_{X=0}) \right] \\
&= \frac{P_\theta(X=j) \mathbb{1}_{j=k}}{\theta_j^2} - \frac{\theta_0}{\theta_0^2} \mathbb{1}_{j=0} - \frac{\theta_0}{\theta_0^2} \mathbb{1}_{k=0} + \frac{\theta_0}{\theta_0^2} \\
&= \frac{\mathbb{1}_{j=k}}{\theta_j} - \frac{\mathbb{1}_{j=0}}{\theta_0} - \frac{\mathbb{1}_{k=0}}{\theta_0} + \frac{1}{\theta_0} \\
&= \frac{\mathbb{1}_{j=k}}{\theta_j} + \frac{1}{\theta_0} \quad \text{si } k \neq 0, j \neq 0
\end{aligned}$$

Donc, en remarquant que $I(\theta)_{0,0} = 0$, on a :

$$\begin{aligned}
\|I(\theta)^{\frac{1}{2}}(\theta' - \theta)\|_2^2 &= (\theta' - \theta)^T I(\theta) (\theta' - \theta) \\
&= \sum_{k,j=0}^d I(\theta)_{k,j} (\theta'_j - \theta_j) (\theta'_k - \theta_k) \\
&= \frac{1}{\theta_0} \sum_{j \neq k, j,k > 0} (\theta'_j - \theta_j) (\theta'_k - \theta_k) + \sum_{k=1}^d \left(\frac{1}{\theta_k} + \frac{1}{\theta_0} \right) (\theta'_k - \theta_k)^2 \\
&= \frac{1}{\theta_0} \sum_{j,k=1}^d (\theta'_j - \theta_j) (\theta'_k - \theta_k) + \sum_{k=1}^d \frac{(\theta'_k - \theta_k)^2}{\theta_k} \\
&= \frac{1}{\theta_0} ((1 - \theta'_0 - (1 - \theta_0))((1 - \theta'_0 - (1 - \theta_0)))) + \sum_{k=1}^d \frac{(\theta'_k - \theta_k)^2}{\theta_k} \\
&= \sum_{k=0}^d \frac{(\theta'_k - \theta_k)^2}{\theta_k}
\end{aligned}$$

Conséquences : Si $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} P_\theta$, le test de Wald I de $H_0 : \theta = \theta^*$ contre $H_1 : \theta \neq \theta^*$ s'écrit en posant $\theta^* = (\theta_1, \dots, \theta_d)$ et en supposant que les conditions de Wilks soient respectées :

$$\phi_{n,\alpha}(X_1, \dots, X_n) = \mathbb{1}_{\left\{ n \sum_{j=0}^d \frac{\hat{\theta}_{j,n}^{MV} - \theta_j^*}{\theta_j^*} > q_{1-\alpha}(X^2(d)) \right\}} \quad (1)$$

où $\hat{\theta}_n^{MV} = (\hat{\theta}_{1,n}^{MV}, \dots, \hat{\theta}_{d,n}^{MV})$. On vérifie que

$$\hat{\theta}_{j,n}^{MV} = \frac{1}{n} \sum_{i=0}^d \mathbb{1}_{\{X_i=j\}}$$

i.e. $\hat{\theta}_{j,n}^{MV}$ est la fréquence de la classe j dans l'échantillon (X_1, \dots, X_n) .

Le test (1) est **le test du χ^2** . Par la propriété précédente, il est asymptotiquement de niveau α .

15.4 Méthodes non Asymptotiques et Non Paramétriques

On se place, dans cette partie, dans le cadre X_1, \dots, X_n, X iid de loi P sur \mathbb{R} et $F(x) = P((-\infty, x]) = P(X \leq x)$

15.4.1 Bandes de Confiance Uniformes sur la Fonction de Répartition

Par Hoeffding, on a pour tout $x \in \mathbb{R}$ et $t \in \mathbb{R}_+$

$$P(|\hat{F}_n(x) - F(x)| \geq t) \leq 2e^{-2nt^2}$$

On a en fait un résultat plus fort (uniforme en x)

Théorème 15|15.4|1

Inégalité de Dvoretzky-Kiefer-Wolfowitz (DKW)

On a pour tout $t \in \mathbb{R}_+$,

$$P(\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \geq t) \leq 2e^{-2nt^2}$$

$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = \|\hat{F}_n - F\|_\infty$ On a une "bande de confiance" sur F . Avec probabilité $1 - \alpha$, on a pour tout $x \in \mathbb{R}$, $F(x) \in [\hat{F}_n(x) - \sqrt{\frac{\log(2/\alpha)}{2n}}, \hat{F}_n(x) + \sqrt{\frac{\log(2/\alpha)}{2n}}]$. La preuve est cependant hors de portée de ce cours.

Proposition 15|15.4|1

Supposons que P n'admette pas d'atome (*i.e.* : F est continue). Soient $U_1, \dots, U_n \stackrel{\text{iid}}{\sim} \mathcal{U}([0, 1])$. Alors,

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \stackrel{(d)}{=} \sup_{t \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{U_i \leq t} - t \right|$$

Preuve :

On pose $F : \mathbb{R} \mapsto (0, 1)$ strictement croissante et bijective

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = \sup_{t=F^{-1}(x), t \in (0,1)} |\hat{F}_n(F^{-1}(t)) - \underbrace{F(F^{-1}(t))}_t|$$

$$\hat{F}_n(F^{-1}(t)) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq F^{-1}(t)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{F(X_i) \leq t}. \text{ Donc}$$

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = \sup_{t \in (0,1)} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{F(X_i) \leq t} - t \right|$$

Or, on a déjà démontré que si F est continue (ce qui est le cas car P n'a pas d'atome), alors $F(X_1), \dots, F(X_n) \stackrel{\text{iid}}{\sim} \mathcal{U}([0, 1])$.

Remarque :

DKW pour tout $n \geq 1$ et $t \in \mathbb{R}_+$, nous donne :

$$P(\sqrt{n} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \geq t) \leq 2e^{-t^2}$$

Quand $n \rightarrow \infty$ on a aussi :

$$\sqrt{n} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{(d)} \sup_{u \in [0,1]} |B_n|$$

où $(B_u)_{u \in [0,1]}$ est un pont gaussien :

Un **processus gaussien** est un processus tel que :

- $\forall N, \forall u_1, \dots, u_N, (B_{u_1}, \dots, B_{u_N})$ vecteur gaussien
- $\mathbb{E}[B_n] = 0 \quad \forall u$ et $\mathbb{E}[B_u B_v] = \min(u, v) - uv$

15.4.2 Test de Permutation

On considère le vecteur $(X_1, \dots, X_n) \sim P_n$ sur \mathcal{X}^n , et on pose l'hypothèse nulle : les observations X_1, \dots, X_n sont iid. C'est-à-dire qu'il existe une loi P sur \mathcal{X} tel que $P_n = P^{\otimes n}$

Exemples :

Détection d'anomalie : X_1, \dots, X_n iid et on veut tester si X_n est indépendante de X_1, \dots, X_{n-1} et de même loi. Cela revient à dire que X_1, \dots, X_n sont iid. Dans le cas contraire, on dit que X_n est une anomalie.

Test d'égalité de loi de deux populations : $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} P$ sur \mathcal{X} et $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} Q$ sur \mathcal{X} et $X_1, \dots, X_n \perp\!\!\!\perp Y_1, \dots, Y_n$. On veut tester si $P = Q$ (l'hypothèse nulle). Cela revient à dire que $(X_1, \dots, X_n, Y_1, \dots, Y_n)$ est iid de loi $P^{\otimes n} \otimes Q^{\otimes n}$

Test de signal : On pose $x_1, \dots, x_n \in \mathcal{X}$ fixés (covariables) et Y_1, \dots, Y_n des variables réelles réponses. On suppose aussi qu'il existe $\sigma > 0$ et une fonction $f : \mathcal{X} \mapsto \mathbb{R}$ telle que $Y_i = f(x_i) + \epsilon_i$ où $\epsilon_1, \dots, \epsilon_n \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. On veut tester s'il y a du signal. Hypothèse nulle H_0 : "pas d'information/signal" *i.e.* : La fonction de régression $f : \mathcal{X} \mapsto \mathbb{R}$ est **constante**. Dans ce cas, les observations Y_1, \dots, Y_n sont iid.

Difficulté : la loi commune (éventuelle) P de X_1, \dots, X_n peut être arbitraire. Il peut être difficile d'estimer P .

On observe que si X_1, \dots, X_n sont iid, alors, la loi de (X_1, \dots, X_n) est **invariante par permutation** au sens où pour toute permutation (bijection) σ de $\{1, \dots, n\}$,

$$(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \stackrel{(d)}{=} (X_1, \dots, X_n) \sim P^{\otimes n}$$

Définition 15|15.4|1 tests de permutations

Soit $F : \mathcal{X}^n \mapsto \mathbb{R}$, $Z = (X_1, \dots, X_n)$ une fonction mesurable. Pour toute $\sigma \in S_n$, $Z^\sigma = (X_{\sigma(1)}, \dots, X_{\sigma(n)})$ avec $\mathfrak{S}_n = \{\text{permutations de } \llbracket 1, n \rrbracket\}$, et donc $\#\mathfrak{S}_n = n!$. On note, pour toute $\sigma \in \mathfrak{S}_n$, $R(\sigma) = \tilde{R}(\sigma, Z) = \sum_{\tau \in \mathfrak{S}_n} \mathbb{1}_{\{F(Z^\tau) \leq F(Z^\sigma)\}}$

Soit id la fonction permutation identité tel que $Z^{\text{id}} = Z$ et prenons $\alpha \in (0, 1)$. Le **test par permutation** de niveau α correspondant à la statistique de test F est le test de $H_0 : X_1, \dots, X_n$ iid est défini par :

$$\phi_\alpha(Z) = \mathbb{1}_{\{R(\text{id}) \leq \alpha n!\}}$$

Théorème 15|15.4|2

Sous $H_0 : X_1, \dots, X_n$ est iid, si on suppose que les valeurs $(F(Z^\sigma))_{\sigma \in \mathfrak{S}_n}$ sont des réels 2 à 2 distincts, alors le **test par permutation** est de niveau α *i.e.*

$$P(\phi_\alpha(Z) = 1) \leq \alpha$$

Preuve :

① On commence par montrer que pour toute permutation σ , $R(\sigma) \stackrel{(d)}{=} R(id)$.

Si $\sigma, \tau \in \sigma_n$, $[(Z^\sigma)^\tau]_i = (Z^\sigma)_{\tau(i)} = Z_{\sigma(\tau(i))} = Z_{\sigma \circ \tau(i)} \quad (*)$.

Donc $(Z^\sigma)^\tau = Z^{\sigma \circ \tau}$, $\tilde{R}(\sigma, Z) = \sum_{\tau \in \sigma_n} \mathbb{1}_{F(Z^\tau) \leq F(Z^\sigma)}$ or, sous H_0 , on a $Z \stackrel{(d)}{=} Z^\tau$, donc $\tilde{R}(\sigma, Z) \stackrel{(d)}{=} \tilde{R}(\sigma, Z^\tau)$. Ainsi, pour σ' une autre permutation, on a :

$$\begin{aligned} R(\sigma) &= \tilde{R}(\sigma, Z) \stackrel{(d)}{=} \tilde{R}(\sigma, Z^\tau) = \sum_{\sigma' \in \mathfrak{S}_n} \mathbb{1}_{F((Z^\tau)^{\sigma'}) \leq F((Z^\tau)^\sigma)} \\ &\stackrel{\text{par } *}{=} \sum_{\sigma' \in \mathfrak{S}_n} \mathbb{1}_{F(Z^{\tau \circ \sigma'}) \leq F(Z^{\tau \circ \sigma})} = \sum_{\sigma'' \in \mathfrak{S}_n} \mathbb{1}_{F(Z^{\sigma''}) \leq F(Z^{\tau \circ \sigma})} \\ &= \tilde{R}(\tau \circ \sigma, Z) = \tilde{R}(\tau \circ \sigma) \end{aligned}$$

Ainsi, $R(\sigma) \stackrel{(d)}{=} \tilde{R}(\tau \circ \sigma)$ pour tout τ . On prend $\tau = \sigma^{-1}$, on conclut que $R(\sigma) \stackrel{(d)}{=} R(id)$.

② Ensuite, on pose $\sigma \sim \mathcal{U}(\mathfrak{S}_n)$ **indépendante de Z** , et on montre que $R(\sigma) \stackrel{(d)}{=} R(id)$.

Pour tout $A \subset \llbracket 1, n! \rrbracket$, on a

$$\begin{aligned} P(R(\sigma) \in A) &= \sum_{\tau \in \mathfrak{S}_n} P(R(\sigma) \in A, \sigma = \tau) \\ &\stackrel{\text{par } *}{=} \sum_{\sigma \perp\!\!\!\perp Z} \sum_{\tau \in \mathfrak{S}_n} \underbrace{P(\tilde{R}(\tau, Z) \in A)}_{=P(R(id, Z) \in A)} \underbrace{P(\sigma = \tau)}_{1/n!} \\ &= P(R(id) \in A) \end{aligned}$$

Donc $R(\sigma) \stackrel{(d)}{=} R(id)$.

③ On rappelle que les valeurs $(F(Z^\sigma))_{\sigma \in \mathfrak{S}_n}$ sont des réels 2 à 2 distincts, alors, si $\sigma \sim \mathcal{U}(\mathfrak{S}_n)$ indépendante de Z , on a $R(\sigma) \sim \mathcal{U}(\llbracket 1, n! \rrbracket)$

L'application

$$\begin{aligned} \Phi_Z : \mathfrak{S}_n &\mapsto \llbracket 1, n! \rrbracket \\ \tau &\mapsto \tilde{R}(\tau, Z) \end{aligned}$$

est une bijection. En effet, les éléments $F(Z^\tau)$ sont 2 à 2 distincts, donc c'est aussi le cas pour les $R(\tau)$, on a donc l'injection. Mais également, comme $\#\mathfrak{S}_n = \#\llbracket 1, n! \rrbracket$, on a que ϕ_Z est injective d'un espace de cardinal $\#\mathfrak{S}_n = \#\llbracket 1, n! \rrbracket = n!$ dans $\llbracket 1, n! \rrbracket$ (de cardinal $n!$) et donc est surjective. Par conséquent, on a la bijection. Ainsi :

$$\begin{aligned}
P(R(\boldsymbol{\sigma}) = j) &= P(\tilde{R}(\boldsymbol{\sigma}, Z) = j) \\
&= P(\boldsymbol{\sigma} = \Phi_Z^{-1}(j)) \\
&= \sum_{\boldsymbol{\tau} \in \mathfrak{S}_n} \underbrace{P(\boldsymbol{\sigma} = \boldsymbol{\tau})}_{=\frac{1}{n!}} P(\Phi_Z^{-1}(j) = \boldsymbol{\tau}) \\
&\stackrel{Z \perp \!\!\! \perp \boldsymbol{\sigma}}{=} \frac{1}{n!}
\end{aligned}$$

Conclusion : $R(id, Z) \sim \mathcal{U}([1, n!])$ donc $P(R(id, Z) \leq \alpha n!) = \frac{[\alpha n!]}{n!}$

Ainsi, $R(\boldsymbol{\sigma}) \stackrel{(d)}{=} R(\boldsymbol{\tau} \circ \boldsymbol{\sigma})$ pour tout $\boldsymbol{\tau} \in \mathfrak{S}_n$. On prend $\boldsymbol{\tau} = \boldsymbol{\sigma}^{-1}$, on conclut que

$$R(\boldsymbol{\sigma}) \stackrel{(d)}{=} R(id)$$

Ainsi, sous H_0 on a bien

$$P(\phi_\alpha(Z) = 1) = P(R(id) \leq \alpha n!) = \frac{[\alpha n!]}{n!} = \alpha \frac{[\alpha n!]}{\alpha n!} \leq \alpha$$

Remarque :

Il n'existe a priori pas d'application F fonctionnant à tout les coups (c'est-à-dire respectant $(F(Z^\sigma))_{\sigma \in \mathfrak{S}_n}$ sont des réels 2 à 2 distincts) pour tout échantillon X_1, \dots, X_n . En revanche, si les X_1, \dots, X_n admettent une densité par rapport à la mesure de Lebesgue, on peut prendre

$$\begin{aligned}
F : \mathcal{Z} &\mapsto \mathbb{R} \\
\mathbf{X} &\mapsto \sum_{i=1}^n ix_i
\end{aligned}$$

Avec donc $(F(Z^\sigma))_{\sigma \in \mathfrak{S}_n}$ étant presque sûrement des valeurs distinctes. En effet, puisque $\ker(F) = \{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n ix_i = 0\}$ est un hyperplan de mesure de Lebesgue nulle. pour tout $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathfrak{S}_n$ différents, on a $\sum_{i=1}^n iX_{\boldsymbol{\sigma}(i)} \stackrel{\text{p.s.}}{\neq} \sum_{i=1}^n iX_{\boldsymbol{\tau}(i)}$.

15.5 Test d'Hypothèses Multiples

Dans certaines applications, on peut être intéressé à effectuer simultanément plusieurs tests sur la loi des données.

Exemple :

régression univariée : $i = 1, \dots, n$, $Y_i = \beta_1 X_{i,1} + \dots + \beta_d X_{i,d} + \epsilon_i$, $\epsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$

Pour chaque $j = 1, \dots, d$ on test $H_{0,j} : \beta_j = 0$ contre $H_{1,j} : \beta_j \neq 0$. Formellement, étant donné le modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ sur Z dans \mathcal{Z} . Pour $i = 1, \dots, m$, on considère un test $H_{0,i} : \theta \in \Theta_{0,i}$ contre $H_{1,i} : \theta \in \Theta_{1,i}$

Définition 15|15.5|1

Un **test multiple** est une fonction $R : \mathcal{Z} \mapsto \mathcal{P}(\{1, \dots, m\})$, où $R(Z)$ est l'ensemble des indices $i = 1, \dots, m$ pour lesquelles on **rejette** $H_{0,i}$.

Pour $\theta \in \Theta$, on note $T_\theta = \{1 \leq i \leq m : \theta \in \Theta_{0,i}\}$ l'ensemble des indices des tests pour lesquels l'hypothèse nulle est **vraie**.

On suppose que, pour chaque $i = 1, \dots, m$, on dispose d'une famille $(\phi_{\alpha,i})_{\alpha \in (0,1)}$ de tests simples de niveau α . On peut définir, pour chaque $i = 1, \dots, m$, la p-valeur associée $\hat{\alpha}_i = \hat{\alpha}_i(Z)$. La question est de savoir comment combiner ces p-valeurs individuelles en un test multiple.

15.5.1 FWER et Correction de Bonferroni

Définition 15|15.5|2

Le **family-wise error rate** (FWER) de R est défini par :

$$\text{FWER}(R) = \sup_{\theta \in \Theta} \mathbb{P}_\theta(R(Z) \cap T_\theta \neq \emptyset) = \sup_{\theta \in \Theta} \mathbb{P}_\theta(\exists i \in \llbracket 1, m \rrbracket, \quad \theta \in \Theta_{0,i} \quad \text{mais} \quad i \in R(Z))$$

Cela correspond à la probabilité de faire au moins une erreur de type I.

On peut observer la méthode naïve en prenant $R(Z) = \{1 \leq i \leq m : \hat{\alpha}_i \leq \alpha\}$ (pas de correction pour le test multiple). Cette méthode ne fonctionne pas : considérons 20 tests avec des hypothèses nulles vraies au niveau $\alpha = 5\%$. En moyenne, au moins 1 test va être rejeté (à tort).

Définition 15|15.5|3

Correction de Bonferroni

$$R_B(Z) = \{i \in \llbracket 1, m \rrbracket : \hat{\alpha}_i \leq \frac{\alpha}{m}\}$$

Proposition 15|15.5|1

$$\text{FWER}(R_B) \leq \alpha$$

Preuve : Borne d'union :

$$\begin{aligned}
\mathbb{P}_\theta(R(Z) \cap T_\theta \neq \emptyset) &= \mathbb{P}_\theta(\exists i \in T_\theta : \hat{\alpha}_i \leq \frac{\alpha}{m}) \\
&\leq \sum_{i \in T_\theta} \mathbb{P}_\theta(\hat{\alpha}_i \leq \frac{\alpha}{m}) \\
&\leq \#(T_\theta) \frac{\alpha}{m} \quad (\text{sur-uniformité de la p-valeur sous } H_{0,i}) \\
&\leq \alpha
\end{aligned}$$

Remarque :

Le FWER est un critère conservateur, en effet, on ne souhaite effectuer aucune erreur de type I. Dans les applications où l'on effectue un grand nombre de tests simultanément (m grand, $\frac{\alpha}{m} \ll 1$), on risque de ne rien rejeter.

15.5.2 La FDR et la Correction de Benjamini-Hochberg

L'objectif est d'adopter un critère plus souple sur les erreurs de type I afin de faire davantage de découvertes. On peut par exemple utiliser le critère des **faux positifs** (erreur de type I) :

$$\text{FP}(\theta, Z) = \#(R(Z) \cap T_\theta) = \#\{i : \theta \in \Theta_{0,i}, \quad i \in R(Z)\}$$

ou bien l'ensemble des **faux négatifs** (erreur de type II) :

$$\text{FN}(\theta, Z) = \#(R(Z)^c \cap T_\theta^c)$$

Définition 15|15.5|4

On définit la **False discovery proportion** (FDP) ainsi :

$$\text{FDP}(\theta, R(Z)) = \frac{\#(R(Z) \cap T_\theta)}{\max(\#R(Z), 1)} = \frac{\text{FP}}{\max(\#R(Z), 1)} = \begin{cases} \frac{\text{FP}}{\#R(Z)} & \text{Si } R(Z) \neq \emptyset \\ 0 & \text{Si } R(Z) = \emptyset \end{cases}$$

Et le **False Discovery Rate** (FDR) comme cela :

$$\text{FDR}(\theta, R(Z)) = \mathbb{E}_\theta[\text{FDP}(\theta, R(Z))] = \mathbb{E}_\theta\left[\frac{\#(R(Z) \cap T_\theta)}{\#R(Z)} \mathbb{1}_{\#R(Z) \geq 1}\right]$$

On peut interpréter que $\text{FDR}(\theta, Z) \leq \alpha$ signifie qu'en moyenne, parmi les "découvertes" (hypothèses nulles rejetées), la proportion de fausses découvertes est d'au plus α .

Procédure de Benjamini-Hochberg

Soit $C_m = \sum_{j=1}^m \frac{1}{j}$. On note

$$R_{BH} = \{1 \leq i \leq m : \hat{\alpha}_i \leq \hat{\alpha}_{(\hat{\xi})}\}$$

où $\hat{\xi} = \max(1 \leq \xi \leq m : \hat{\alpha}_{(\xi)} \leq \frac{\xi\alpha}{mC_m})$

Théorème 15|15.5|1

$$\text{FDR}(\theta, R_{BH}) \leq \alpha$$

Preuve :

$$\begin{aligned} \text{FDR}(\theta, R_{BH}) &= \mathbb{E}_\theta \left[\frac{\#(R_{BH}(Z) \cap T_\theta)}{\#R_{BH}(Z)} \mathbb{1}_{\#R_{BH}(Z) \geq 1} \right] \\ &= \mathbb{E}_\theta \left[\frac{\sum_{i \in T_\theta} \mathbb{1}_{\hat{\alpha}_i \leq \hat{\alpha}_{(\hat{\xi})}}}{\hat{\xi}} \mathbb{1}_{\hat{\xi} \geq 1} \right] \quad (*) \end{aligned}$$

On remarque que :

$$\frac{1}{\hat{\xi}} = \sum_{j \geq \hat{\xi}} \left(\frac{1}{j} - \frac{1}{j+1} \right) = \sum_{j \geq 1} \frac{\mathbb{1}_{j \geq \hat{\xi}}}{j(j+1)}$$

Donc

$$\text{FDR} \stackrel{(*)}{=} \mathbb{E}_\theta \left[\sum_{i \in T_\theta} \sum_{j \geq 1} \frac{\mathbb{1}_{\hat{\alpha}_i \leq \hat{\alpha}_{(\hat{\xi})}} \mathbb{1}_{1 \leq \hat{\xi} \leq j}}{j(j+1)} \right]$$

Or, si $\hat{\xi} \leq j$, on a $\hat{\alpha}_{(\hat{\xi})} \stackrel{(def)}{\leq} \frac{\hat{\xi}\alpha}{mC_m} \leq \min(\alpha', \frac{j\alpha'}{m})$ car $\hat{\xi} \leq \min(j, m)$, avec $\alpha' = \frac{\alpha}{C_m}$. Donc

$$\begin{aligned}
\text{FDR} &\leq \mathbb{E}_\theta \left[\sum_{i \in T_\theta} \sum_{j \geq 1} \frac{\mathbb{1}_{\hat{\alpha}_i \leq \min(\frac{j\alpha'}{m}, \alpha')}}{j(j+1)} \right] \\
&= \sum_{i \in T_\theta} \sum_{j \geq 1} \frac{\mathbb{P}_\theta(\hat{\alpha}_i \leq \min(\frac{j\alpha'}{m}, \alpha'))}{j(j+1)} \\
&\leq \sum_{i \in T_\theta} \sum_{j \geq 1} \frac{\min(\frac{j\alpha'}{m}, \alpha')}{j(j+1)} \quad (\text{sur uniformité sous } H_{0,i} \text{ de } \hat{\alpha}_i) \\
&\leq \#T_\theta \sum_{j \geq 1} \frac{\min(\frac{j\alpha'}{m}, \alpha')}{j(j+1)} \\
&\leq \#T_\theta \alpha' \left[\sum_{j=1}^{m-1} \frac{\frac{j}{m}}{j(j+1)} + \sum_{j=m}^{\infty} \frac{1}{j(j+1)} \right] \\
&\leq \#T_\theta \frac{\alpha'}{m} \left(\sum_{j=2}^m \frac{1}{j} + 1 \right) \quad (\text{car } \sum_{j=m}^{\infty} \frac{1}{j(j+1)} = \sum_{j=m}^{\infty} \frac{1}{j} - \frac{1}{j+1} = \frac{1}{m}) \\
&\leq \#T_\theta \frac{\alpha}{C_m m} \times C_m \leq \alpha
\end{aligned}$$