

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A. MIRA - Béjaia
Faculté des Sciences Exactes
Département de Mathématiques

Polycopié de cours

Statistique Inférentielle

Mme BOURAINE née BERDJOU DJ Louiza

Année universitaire 2013/2014

Avant-Propos

Ce polycopié de cours est destiné principalement aux étudiants de la Licence Mathématiques mais peut-être utile à toute personne souhaitant connaître et surtout utiliser les principales méthodes de la statistique inférentielle.

Le niveau mathématique requis est celui de la première année et deuxième année Licence avec quelques notions (probabilités unidimensionnelles, séries, intégrales multiples) souvent enseignées en deuxième année Licence.

la première partie du polycopié est consacré à quelques rappels de la théorie des probabilités (couples de variables aléatoires, fonctions caractéristiques et la convergence des suites de variables aléatoires) constitue un passage obligé pour donner des bases rigoureuses à la méthode statistique. La deuxième partie statistique correspond aux chapitres d'échantillonnage, d'estimation et de tests d'hypothèses.

Chaque chapitre se conclut par des exercices permettant de contrôler l'acquisition des notions essentielles qui ont été introduites.

Table des matières

Introduction	4
1 Rappels	7
1.1 Couples de variables aléatoires	7
1.1.1 Définition et propriétés	7
1.1.2 Vecteurs aléatoires discrets et continus	8
1.1.3 Variables aléatoires indépendantes	15
1.1.4 Espérance mathématique	17
1.1.5 Covariance :	19
1.1.6 Changement de variables bidimensionnelles	24
1.1.7 Régression	27
1.1.8 Vecteurs aléatoires gaussiens	33
1.1.9 Les lois de probabilités usuelles	39
1.1.10 Lois continues	39
1.2 Fonctions caractéristiques et convergence de suites de variables aléatoires .	44
1.2.1 Fonctions caractéristiques	44
1.2.2 Fonctions caractéristiques des variables aléatoires marginales et conditionnelles	51
1.2.3 Convergence des suites de variables aléatoires	52
1.2.4 Exercices	60
2 Echantillonnage	65

2.1	Position du problème	65
2.1.1	Avantages de l'échantillonnage	66
2.1.2	Choix de l'échantillon	66
2.2	Notions fondamentales	67
2.2.1	Population de référence	67
2.2.2	Echantillon	67
2.2.3	Distribution d'un échantillon	68
2.2.4	Fonction de répartition empirique d'un échantillon	68
2.2.5	Statistique d'ordre d'un échantillon	70
2.2.6	Moments d'un échantillon	71
2.3	Modèle statistique	75
2.4	Exercices	76
3	L'estimation	77
3.1	Généralités	77
3.1.1	Estimateur	77
3.1.2	Exemples élémentaires	78
3.1.3	Qualités d'un estimateur	78
3.2	Statistique suffisante (exhaustive)	80
3.2.1	Définition d'une statistique exhaustive	80
3.2.2	Lois permettant une statistique exhaustive	81
3.2.3	L'information de Fisher	82
3.2.4	Information de Fisher et suffisance	85
3.2.5	Borne de Freshet-Damois-Cramer-Rao (FDCR)	86
3.2.6	Estimateur efficace	87
3.2.7	Estimateur sans biais de variance minimale (MVUE)	87
3.2.8	Généralisation (Cas multidimensionnel)	87
3.2.9	Statistique complète	89
3.3	Méthodes de calcul d'un estimateur	90

3.3.1	Méthode des moments	90
3.3.2	Méthode du maximum de vraisemblance	91
3.4	Estimation par intervalle de confiance	96
3.4.1	Principe de construction	96
3.4.2	Intervalles de confiance classiques	98
3.5	Exercices	102
4	Tests statistiques	104
4.1	Notions générales	104
4.2	Test entre deux hypothèses simples	108
4.2.1	La méthode de Neyman-Pearson	108
4.3	Tests d'hypothèses simple contre hypothèse composite	113
4.4	Tests paramétriques usuels	114
4.5	Application du test de N-P dans le cas de famille exponentielle	116
4.6	Tests et statistique exhaustive	118
4.7	Test entre deux hypothèses multiples (test de Lehman)	118
4.7.1	Famille à rapport de vraisemblance monotone	118
4.7.2	Tests bilatérales	119
4.7.3	Tests unilatérales	120
4.8	Test de rapport de vraisemblance maximale	121
4.9	Test de comparaison d'échantillons	123
4.9.1	Test de Student et F.S pour des échantillons indépendants	123
4.10	Exercices	128
	Bibliographie	129

Introduction

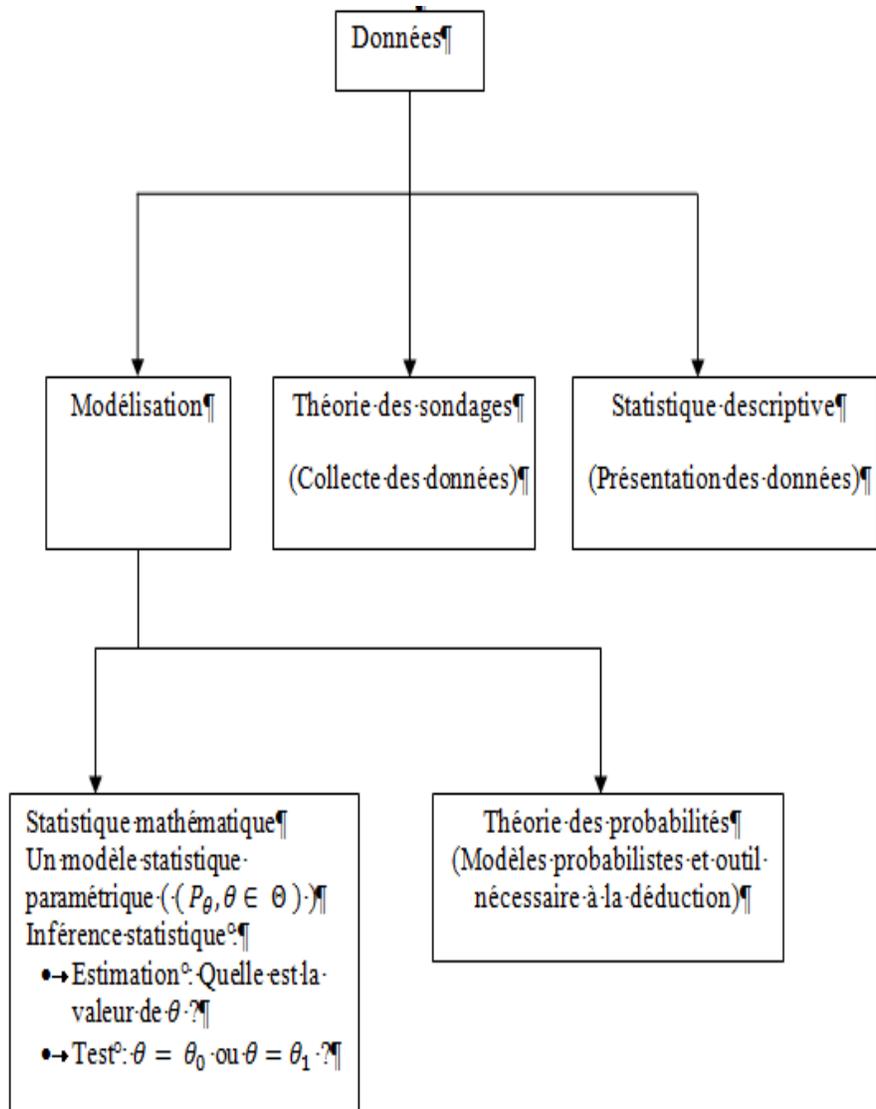
La statistique a une origine très ancienne, se réduisant initialement à une collecte de données (recensement).

Le terme "Statistique", vient du latin *statisticus*, relatif à l'état (*status*), il est employé alors dans un sens purement descriptif de recueil ou de collection de données chiffrées, "les statistiques". Le mot employé au singulier, avec l'article défini "la statistique" évoque la méthode utilisée pour étendre les résultats et dégager des lois (l'inférence) c'est donc une méthode scientifique d'analyse des données recueillies.

La théorie des probabilités joue un rôle important dans l'inférence statistique.

La théorie des probabilités a commencé avec Blaise Pascal, Pierre Fermat et Jacques Bernoulli par l'analyse des jeux dits de hasard. La théorie des probabilités servira ensuite d'outil de base à un ensemble de données pour fixer la précision avec laquelle l'on estime certains paramètres : la statistique mathématique (ou inférentielle).

Tout ceci peut se résumer au moyen de la hiérarchie suivante :



Chapitre 1

Rappels

1.1 Couples de variables aléatoires

1.1.1 Définition et propriétés

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Un couple aléatoire ou un vecteur aléatoire à deux dimensions est une application notée (X, Y) définie par :

$$\begin{aligned}(X, Y) : \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto (X(\omega), Y(\omega))\end{aligned}$$

telle que $\forall x, y \in \mathbb{R}$:

$A_x = \{\omega \in \Omega : X(\omega) < x\} \in \mathcal{A}$ et $A_y = \{\omega \in \Omega : Y(\omega) < y\} \in \mathcal{A}$ ou bien $A_{x,y} = \{\omega \in \Omega : X(\omega) < x \text{ et } Y(\omega) < y\} \in \mathcal{A}$ (Image réciproque du pavé $] - \infty, x[\times] - \infty, y[$). Ceci revient à supposer que la probabilité attachée à tout pavé de la forme $] - \infty, x[\times] - \infty, y[$ est définie grâce à la définition de \mathbb{P} (voir FIG.1.1).

Ou bien $\forall B = (B_1, B_2) \in \mathcal{B}_{\mathbb{R}} \times \mathcal{B}_{\mathbb{R}} = \mathcal{B}_{\mathbb{R}^2}$ (Tribu borélienne engendrée par les pavés : $\{\omega \in \Omega : X(\omega) \in B_1\} \in \mathcal{A}$, $\{\omega \in \Omega : Y(\omega) \in B_2\} \in \mathcal{A}$).

Définition 1.1. L'ensemble des valeurs possibles du couple est appelé *ensemble de définition* ou *support du couple*.

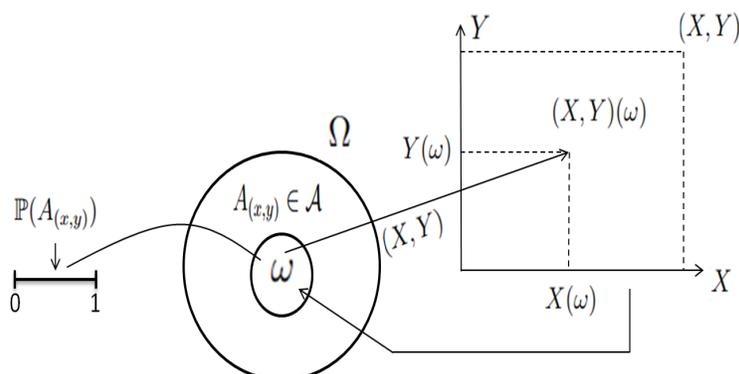


FIG. 1.1 – Couple de variables aléatoires

Propriétés de la fonction de répartition

Soit F la fonction de répartition du couple (X, Y) . On a

$$\begin{aligned} F : \mathbb{R}^2 &\longrightarrow [0, 1] \\ (x, y) &\longmapsto F(x, y) \end{aligned}$$

F vérifie les propriétés suivantes :

1. $0 \leq F(x, y) \leq 1, \quad \forall (x, y) \in \mathbb{R}^2$.
2. $\lim_{x \text{ ou } y \rightarrow -\infty} F(x, y) = 0$ et $\lim_{x \text{ et } y \rightarrow +\infty} F(x, y) = 1$
3. $F(x, y)$ est continue à gauche.
4. $F(x, y)$ est non décroissante i.e : $\forall (a, b) \in \mathbb{R}^2, \quad a > 0, \quad b > 0, \quad \forall (x, y) \in \mathbb{R}^2 :$

$$\begin{cases} F(x + a, y) - F(x, y) \geq 0, \\ F(x, y + b) - F(x, y) \geq 0, \\ F(x + a, y + b) - F(x + a, y) - F(x, y + b) + F(x, y) \geq 0. \end{cases}$$

1.1.2 Vecteurs aléatoires discrets et continus

- **Couple aléatoire discret** : L'ensemble de définition ou le support D du couple (X, Y) est constitué d'un nombre fini ou dénombrable de points de \mathbb{R}^2 . Il est caractérisé par une loi conjointe des variables aléatoires X et Y .
- **Vecteur aléatoire absolument continu** : L'ensemble de définition $D = X(\Omega) \times Y(\Omega)$ est une partie de \mathbb{R}^2 . La fonction de répartition conjointe est

continue et possède une dérivée partielle d'ordre deux $\frac{\partial^2 F}{\partial x \partial y}$. Ces vecteurs seront caractérisés par une fonction densité conjointe des variables aléatoires X et Y .

Etude du cas discret

Soit (X, Y) un couple de variables aléatoires discrètes. $X(\Omega) = \{x_i, i \in I\}$, $Y(\Omega) = \{y_j, j \in J\}$. La loi de probabilité du couple (X, Y) est définie par la donnée d'une suite de réels positifs ou nuls p_{ij} tels que :

$$\mathbb{P}(X = x_i \text{ et } Y = y_j) = p_{ij}$$

avec $\sum_{i \in I} \sum_{j \in J} p_{ij} = 1$.

- **Lois marginales**

Considérons la suite des réels $p_{i\bullet} \geq 0$ définie par :

$$p_{i\bullet} = \mathbb{P}(X = x_i), \quad i \in I.$$

On a : $\{X = x_i\} = \left\{ \bigcup_{j \in J} (X = x_i \text{ et } Y = y_j) \right\}$,

donc : $p_{i\bullet} = \sum_{j \in J} p_{ij}$

et

$$\sum_{i \in I} p_{i\bullet} = \sum_{i \in I} \sum_{j \in J} p_{ij} = 1.$$

Les $p_{i\bullet}$ constituent une loi de probabilité de la variable aléatoire X , cette loi est appelée "loi marginale de la variable aléatoire X ".

De même la loi marginale de la variable aléatoire Y est définie par les réels $p_{\bullet j} \geq 0$ tels que :

$$p_{\bullet j} = \mathbb{P}(Y = y_j) = \sum_{i \in I} p_{ij}$$

et on a : $\sum_{j \in J} p_{\bullet j} = \sum_{j \in J} \sum_{i \in I} p_{ij} = 1$.

Exemple 1.1. Une urne contient 4 boules blanches, 2 boules noires et 4 boules rouges. On extrait au hasard 3 boules de cette urne sans remise. On note X : "Le nombre de boules blanches" et Y : "le nombre de boules noires" figurant parmi les trois boules tirées. Déterminer la loi conjointe du couple (X, Y) et les lois marginales de X et Y .

L'espace fondamental Ω associé à l'expérience aléatoire qui consiste à tirer 3 boules sans remise est l'ensemble de dispositions non ordonnées, sans répétition de 3 boules parmi 10 (c'est donc des combinaisons sans répétition).

$$|\Omega| = C_{10}^3 = \frac{10!}{3!7!} = 120$$

L'ensemble de définition du couple est :

$$\begin{aligned} D &= \{(i, j) \text{ tels que } 0 \leq i \leq 3, 0 \leq j \leq 2\} \\ &= \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (3, 0)\} \text{ fini.} \end{aligned}$$

Les résultats sont donnés dans le tableau suivant :

$X \ Y$	0	1	2	$p_{i\bullet}$
0	1/30	3/30	1/30	$p_0 = 5/30$
1	6/30	8/30	1/30	$p_1 = 15/30$
2	6/30	3/30	0	$p_2 = 9/30$
3	1/30	0	0	$p_3 = 1/30$
$p_{\bullet j}$	14/30	14/30	2/30	1

Etude du cas continu

- **Fonction de densité conjointe**

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et (X, Y) un couple aléatoire absolument continu. $X(\Omega) \times Y(\Omega)$ est un pavé de \mathbb{R}^2 (ensemble produit de deux intervalles de \mathbb{R}) de fonction de répartition F possédant une dérivée seconde $\frac{\partial^2 F}{\partial x \partial y}$.

Définition 1.2. On appelle fonction de densité de probabilité toute application f de $\mathbb{R}^2 \rightarrow \mathbb{R}$ possédant les propriétés suivantes :

- $f(x, y) \geq 0, \quad \forall (x, y) \in \mathbb{R}^2$;
- f continue sur \mathbb{R}^2 , sauf peut être sur un ensemble de surface nulle ;
- $\int \int_{\mathbb{R}^2} f(x, y) dx dy = 1$;
- $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$;
- $\forall a, b, c, d : \mathbb{P}(a \leq X \leq b \text{ et } c \leq Y \leq d) = \int \int_{\Delta} f(x, y) dx dy$, où Δ est le pavé $[a, b] \times [c, d]$.

Remarque 1.1.

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv.$$

- **Fonction de densité marginale**

La loi marginale de X possède la fonction de répartition G et la densité g définies par :

$$G(x) = F(x, +\infty).$$

$$g(x) = G'(x) = \int_{-\infty}^{+\infty} f(x, y) dy = f_X(x).$$

De même la loi marginale de Y possède la fonction de répartition H et la densité h définies par :

$$H(y) = F(+\infty, y).$$

$$h(y) = H'(y) = \int_{-\infty}^{+\infty} f(x, y) dx = f_Y(y).$$

Exemple 1.2. Soit (X, Y) un couple de variables aléatoires de fonction de densité :

$$f(x, y) = \begin{cases} x + y, & \text{si } 0 \leq x \leq 1 \text{ et } 0 \leq y \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

Le support de (X, Y) est représenté comme suit :

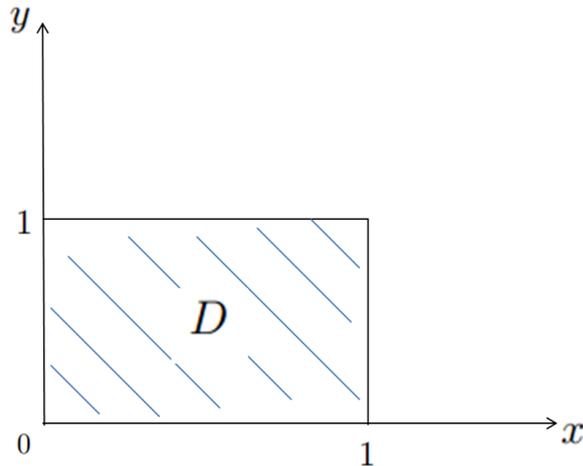


FIG. 1.2 – Support D

Les fonctions de densités marginales sont respectivement :

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ &= \int_0^1 (x + y) dy \\ &= x + 1/2. \end{aligned}$$

D'où :

$$f_X(x) = \begin{cases} x + 1/2, & \text{si } 0 \leq x \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

De même :

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx \\ &= \int_0^1 (x + y) dx \\ &= y + 1/2. \end{aligned}$$

D'où :

$$f_Y(y) = \begin{cases} y + 1/2, & \text{si } 0 \leq y \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

Exemple 1.3. Soit

$$f(x, y) = \begin{cases} 3(x + y), & \text{si } x > 0, y > 0 \text{ et } x + y < 1; \\ 0, & \text{sinon.} \end{cases}$$

Le support :

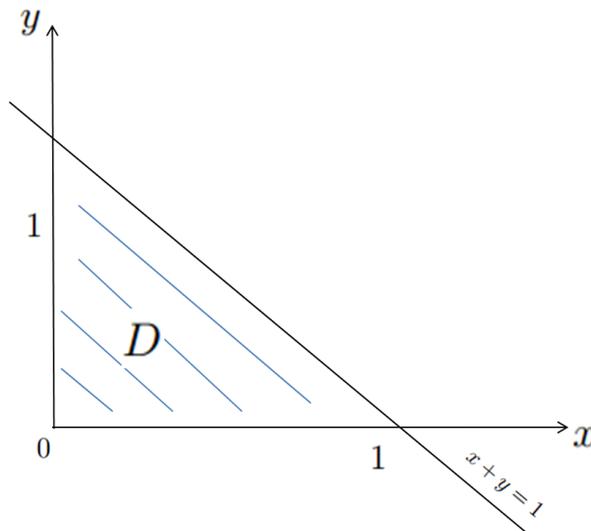


FIG. 1.3 – Support D

$$f_X(x) = \int_0^{1-x} 3(x+y)dy = \frac{3}{2}(1-x^2), \quad 0 \leq x \leq 1.$$

$$f_Y(y) = \int_0^{1-y} 3(x+y)dx = \frac{3}{2}(1-y^2), \quad 0 < y \leq 1.$$

Lois conditionnelles

Soit A un événement de $(\Omega, \mathcal{A}, \mathbb{P})$ tels que : $\mathbb{P}(A) \neq 0$.

$$\begin{aligned} \mathbb{P}(\cdot|A) : \mathcal{A} &\longrightarrow [0, 1] \\ B &\longmapsto \mathbb{P}(B/A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \end{aligned}$$

Appliquons cette définition aux variables aléatoires.

▷ **Cas discret** : Soit (X, Y) un couple aléatoire discret tels que : $X(\Omega) = \{x_i, i \in I\}$, $Y(\Omega) = \{y_j, j \in J\}$. Lorsque la variable aléatoire Y est fixée $Y = y_j$ avec $p_{\bullet j} = \mathbb{P}(Y = y_j) \neq 0$.

La variable aléatoire prend différentes valeurs x_i tel que $(x_i, y_j) \in D$, ($D = X(\Omega) \times Y(\Omega)$: support du couple).

On peut alors définir une variable aléatoire unidimensionnelle X sachant $Y = y_j$ dont la loi de probabilité est :

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{p_{ij}}{p_{\bullet j}}.$$

La variable aléatoire $X|Y = y_j$, $j \in J$ est appelée variable aléatoire conditionnelle de X sachant $Y = y_j$. On a bien

$$\sum_{i \in I} \mathbb{P}(X = x_i | Y = y_j) = \frac{\sum_{i \in I} p_{ij}}{p_{\bullet j}} = \frac{p_{\bullet j}}{p_{\bullet j}} = 1.$$

De même on définit la variable aléatoire Y sachant $X = x_i$ par :

$$\mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}(Y = y_j, X = x_i)}{\mathbb{P}(X = x_i)} = \frac{p_{ij}}{p_{i\bullet}}$$

et on a bien

$$\sum_{j \in J} \mathbb{P}(Y = y_j | X = x_i) = \frac{\sum_{j \in J} p_{ij}}{p_{i\bullet}} = \frac{p_{i\bullet}}{p_{i\bullet}} = 1, \quad (p_{i\bullet} \neq 0).$$

Exemple 1.4. Soit le couple aléatoire (X, Y) de support

$D = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ de loi conjointe :

$$p_{00} = 0.4, \quad p_{01} = 0.2, \quad p_{10} = 0.1, \quad p_{11} = 0.3.$$

Trouver la loi de X sachant que $Y = 1$, $X(\Omega) = \{0, 1\}$.

$$\begin{aligned} \mathbb{P}(X = 0 | Y = 1) &= \frac{p_{01}}{p_{\bullet 1}} = \frac{0.2}{\sum_{i=0}^1 p_{i1}} = \frac{0.2}{p_{01} + p_{11}} = \frac{0.2}{0.5} = 2/5 \\ \mathbb{P}(X = 1 | Y = 1) &= \frac{p_{11}}{p_{01} + p_{11}} = \frac{0.3}{0.5} = 3/5 \end{aligned}$$

et $\sum_{i \in I} \mathbb{P}(X = x_i | Y = y_j) = 1$.

▷ **Cas continu** : Soit (X, Y) un couple aléatoire de densité conjointe f . On définit la densité conditionnelle de Y sachant $X = x$ (respectivement de X sachant $Y = y$) par la relation $f_{Y|X}(y) = \frac{f(x,y)}{f_X(x)}$ avec $f_X(x) \neq 0$ (respectivement $f_{X|Y}(x) = \frac{f(x,y)}{f_Y(y)}$ avec $f_Y(y) \neq 0$).

Exemple 1.5. Soit (X, Y) un couple aléatoire de densité conjointe :

$$f(x, y) = \begin{cases} \frac{e^{-x/y} e^{-y}}{y}, & x \in]0, +\infty[, \quad y \in]0, +\infty[, \quad D = \mathbb{R}^* \times \mathbb{R}^* ; \\ 0, & \text{sinon.} \end{cases}$$

Cherchons $\mathbb{P}(X > 1 | Y = y)$.

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

$$\begin{aligned} f_Y(y) &= \int_0^{+\infty} f(x, y) dx \\ &= \int_0^{+\infty} \frac{e^{-x/y} e^{-y}}{y} dx \\ &= \frac{e^{-y}}{y} \int_0^{+\infty} e^{-x/y} dx \\ &= e^{-y}. \end{aligned}$$

D'où :

$$f_{X|Y}(x) = \frac{e^{-x/y} e^{-y}}{y e^{-y}} = \begin{cases} \frac{1}{y} e^{-x/y}, & x \in]0, +\infty[; \\ 0, & \text{sinon.} \end{cases}$$

$$\begin{aligned} \mathbb{P}(X > 1 | Y = y) &= \int_1^{+\infty} f_{X|Y}(x) dx \\ &= \frac{1}{y} \int_1^{+\infty} e^{-x/y} dx \\ &= e^{-1/y}. \end{aligned}$$

1.1.3 Variables aléatoires indépendantes

Définition 1.3. Soit (X, Y) un couple aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$. Les variables aléatoires X et Y sont indépendantes si :

a) Cas discret :

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j) \quad \text{où } x_i \in X(\Omega), i \in I, \quad y_j \in Y(\Omega), j \in J.$$

b) Cas absolument continu :

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{ou bien}$$

$$F(x, y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y) = F_X(x) \cdot F_Y(y).$$

Dans le cas contraire on dit que X et Y sont dépendantes.

Remarque 1.2. Lorsque X et Y sont indépendantes :

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X=x_i)\mathbb{P}(Y=y_j)}{\mathbb{P}(Y=y_j)} = \mathbb{P}(X = x_i) = P_{i\bullet} \quad (\text{Loi marginale de } X) \quad \text{et :}$$

$$f_{X|Y}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x) \quad (\text{Densité marginale de } X).$$

Somme de deux variables aléatoires indépendantes

Cas continu :

Soient X, Y deux variables aléatoires continues indépendantes,

$$\begin{aligned} F_{X+Y}(t) &= \mathbb{P}(X + Y \leq t) \\ &= \int \int_{x+y \leq t} f(x, y) dx dy \\ &= \int \int_{x+y \leq t} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{t-y} f_X(x) dx \right] f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} F_X(t-y) f_Y(y) dy. \end{aligned}$$

$$\begin{aligned} f_{X+Y}(t) &= \frac{d}{dt} \left[\int_{-\infty}^{+\infty} F_X(t-y) f_Y(y) dy \right] \\ &= \int_{-\infty}^{+\infty} \frac{d}{dt} (F_X(t-y) f_Y(y) dy) \\ &= \int_{-\infty}^{+\infty} f_X(t-y) f_Y(y) dy. \end{aligned}$$

Cas discret :

$$\begin{aligned}\mathbb{P}(X + Y = t) &= \sum_x \mathbb{P}(X = x, Y = t - x) \\ &= \sum_y \mathbb{P}(X = t - y, Y = y) \\ &= \sum_x \mathbb{P}(X = x)\mathbb{P}(Y = t - x).\end{aligned}$$

1.1.4 Espérance mathématique

Définition 1.4. L'espérance mathématique du vecteur aléatoire (X, Y) est le vecteur $(\mathbb{E}(X), \mathbb{E}(Y))$ s'il existe (i.e : si $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ existent).

Propriétés

Proposition 1.1. Soit (X, Y) un couple aléatoire et $a \in \mathbb{R}$, alors $\mathbb{E}(a(X, Y)) = a\mathbb{E}(X, Y)$.

Proposition 1.2. Soit

$$\begin{aligned}\varphi : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \varphi(x, y)\end{aligned}$$

et (X, Y) un couple aléatoire de support D .

$$\mathbb{E}(\varphi(x, y)) = \sum_{(x, y) \in D} \varphi(x, y)\mathbb{P}(X = x, Y = y) \quad \text{si la série converge (cas discret),}$$

$$\mathbb{E}(\varphi(x, y)) = \int \int_{\mathbb{R}^2} \varphi(x, y)f(x, y)dxdy \quad \text{si l'intégrale est convergente (cas continu).}$$

Exemple 1.6. $\mathbb{E}(XY) = \int \int_{\mathbb{R}^2} xyf(x, y)dxdy$.

Espérance d'une somme

Proposition 1.3. Soit (X, Y) un couple aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ tel que $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ existent.

Alors :

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Démonstration. Cas discret :

$$\begin{aligned}
 \mathbb{E}(X + Y) &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (x + y) \mathbb{P}(X = x, Y = y) \\
 &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} x \mathbb{P}(X = x, Y = y) + \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} y \mathbb{P}(X = x, Y = y) \\
 &= \sum_{x \in X(\Omega)} x \left(\sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y) \right) + \sum_{y \in Y(\Omega)} y \left(\sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y) \right) \\
 &= \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) + \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y) \\
 &= \mathbb{E}(X) + \mathbb{E}(Y).
 \end{aligned}$$

Cas continu :

$$\begin{aligned}
 \mathbb{E}(X + Y) &= \int \int_{\mathbb{R}^2} (x + y) f(x, y) dx dy \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy \\
 &= \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} f(x, y) dy \right) dx + \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f(x, y) dx \right) dy \\
 &= \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} y f_Y(y) dy \\
 &= \mathbb{E}(X) + \mathbb{E}(Y).
 \end{aligned}$$

□

Exercice : Soit X, Y deux variables aléatoires indépendantes de loi $\mathcal{U}_{[0,1]}$ et $Z = X + Y$.

1. Déterminer la loi de probabilité de Z ,
2. Calculer $\mathbb{E}(Z)$ et $V(Z)$.

Corollaire 1.1. Soient X, Y deux variables aléatoires indépendantes, si $\mathbb{E}(X)$, $\mathbb{E}(Y)$ et $\mathbb{E}(XY)$ existent alors :

$$\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y).$$

Démonstration. Dans le cas continu

$$\begin{aligned}
 \mathbb{E}(XY) &= \int \int xy f(x,y) dx dy \\
 &= \int \int xy f_X(x) \cdot f_Y(y) dx dy \quad \text{car } X \text{ et } Y \text{ sont indépendantes.} \\
 &= \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right) \left(\int_{-\infty}^{+\infty} y f_Y(y) dy \right) \\
 &= \mathbb{E}(X)\mathbb{E}(Y).
 \end{aligned}$$

□

1.1.5 Covariance :

Soient X, Y deux variables aléatoires. On appelle covariance de X et de Y la valeur suivante si elle existe :

$$\text{cov}(X, Y) = \mathbb{E}[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]. \quad (1.1)$$

Remarque 1.3.

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Il suffit de développer l'expression (1.1).

Propriétés de la covariance et de la variance d'une somme

Proposition 1.4. Soient X, X', Y, Y' des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ admettant chacune un moment d'ordre 2 et $\lambda \in \mathbb{R}$ alors :

1. $\text{cov}(X, X) = \text{Var}(X) \geq 0$;
2. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y)$;
3. $\text{cov}(X, Y) = \text{cov}(Y, X)$;
4. $\text{cov}(X + X', Y) = \text{cov}(X, Y) + \text{cov}(X', Y)$;
5. $\text{cov}(X, Y + Y') = \text{cov}(X, Y) + \text{cov}(X, Y')$;
6. $\text{cov}(\lambda X, Y) = \text{cov}(X, \lambda Y) = \lambda \text{cov}(X, Y)$;
7. $\text{cov}(aX + b, cY + d) = a c \text{cov}(X, Y) \quad \forall a, b, c, d \in \mathbb{R}$.

Démonstration. 1.

$$\begin{aligned} \text{cov}(X, X) &= \mathbb{E}(XX) - \mathbb{E}(X) \cdot \mathbb{E}(X) \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \text{Var}(X) \geq 0. \end{aligned}$$

2.

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}(X + Y))^2 \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}^2(X) - \mathbb{E}^2(Y) - 2\mathbb{E}(X) \cdot \mathbb{E}(Y). \\ &= \text{Var}(X) + \text{Var}(Y) + 2[\mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y) \end{aligned}$$

3.

$$\begin{aligned} \text{cov}(X + X', Y) &= \mathbb{E}[(X + X')Y] - \mathbb{E}(X + X')\mathbb{E}(Y) \\ &= \mathbb{E}[XY + X'Y] - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X')\mathbb{E}(Y) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X'Y) - \mathbb{E}(X')\mathbb{E}(Y) \\ &= \text{cov}(X, Y) + \text{cov}(X', Y). \end{aligned}$$

4.

$$\begin{aligned} \text{cov}(\lambda X, Y) &= \mathbb{E}(\lambda XY) - \mathbb{E}(\lambda X)\mathbb{E}(Y) \\ &= \lambda \mathbb{E}(XY) - \lambda \mathbb{E}(X)\mathbb{E}(Y) \\ &= \lambda \text{cov}(X, Y). \end{aligned}$$

□

La covariance est une forme bilinéaire symétrique positive sur l'ensemble des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ ayant un moment d'ordre 2 (la forme quadratique associée étant la variance).

Remarque 1.4. La covariance n'est pas définie positive, car $\text{Var}(X) = 0 \Leftrightarrow X = \text{cste}$ et non pas pour X nulle.

Cas où les variables aléatoires sont indépendantes

Proposition 1.5. Soit X, Y deux variables aléatoires indépendantes dont les variances existent alors :

1. $cov(X, Y) = 0$;
2. $cov(X + Y) = Var(X) + Var(Y)$.

Démonstration. $cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$

car $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

D'autre part : $Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y) = Var(X) + Var(Y)$. \square

Coefficient de corrélation

Définition 1.5. Soit X, Y deux variables aléatoires, leur coefficient de corrélation est le nombre : $\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$ s'il existe.

Où σ_X l'écart-type de X et σ_Y l'écart-type de Y .

Propriétés :

1. $|\rho| \leq 1$,
2. $\rho(X, X) = 1$ et $\rho(X, -X) = -1$,
3. $\rho(aX + b, cY + d) = \rho(X, Y)$ si $ac \neq 0$ où $a, b, c, d \in \mathbb{R}$
4. Si X, Y sont indépendantes alors $\rho(X, Y) = 0$, la réciproque est fautive.

Démonstration. 1. On propose deux méthodes pour la première propriété.

Première méthode On a :

$$Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \geq 0.$$

$$\begin{aligned} Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) &= \frac{Var(X)}{\sigma_X^2} + \frac{Var(Y)}{\sigma_Y^2} + \frac{2cov(X, Y)}{\sigma_X \sigma_Y} \\ &= 1 + 1 + 2\rho = 2(1 + \rho(X, Y)) \geq 0 \end{aligned}$$

$$\rho \geq -1.$$

D'autre part :

$$\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \geq 0$$

$$\begin{aligned} \text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \geq &= \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} - \frac{2\text{cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= 2 - 2\rho = 2(1 - \rho) \geq 0 \Rightarrow \rho \leq 1 \end{aligned}$$

D'où :

$$-1 \leq \rho \leq 1$$

Deuxième méthode Soit a un nombre réel quelconque et notons V la quantité

$$\mathbb{E}[a(X - \mathbb{E}(X)) + (Y - \mathbb{E}(Y))]^2.$$

$$\begin{aligned} V &= \mathbb{E}[a^2(X - \mathbb{E}(X))^2 + (Y - \mathbb{E}(Y))^2 + 2a(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= a^2\mathbb{E}[(X - \mathbb{E}(X))^2] + \mathbb{E}[Y - \mathbb{E}(Y)]^2 + 2a\mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= a^2\text{Var}(X) + \text{Var}(Y) + 2a\text{cov}(X, Y) \geq 0. \end{aligned}$$

Ce polynôme en a est positif ou nul donc son discriminant est négatif car

$$\text{Var}(X) \geq 0.$$

$$\begin{aligned} \Delta' &= \text{cov}^2(X, Y) - \text{Var}(X)\text{Var}(Y) \leq 0 \\ &\Rightarrow \text{cov}^2(X, Y) \leq \text{var}(X)\text{Var}(Y) \\ &\Rightarrow \frac{\text{cov}^2(X, Y)}{\text{var}(X)\text{Var}(Y)} \leq 1 \\ &\Rightarrow -1 \leq \rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{Var}(Y)}} < 1. \end{aligned}$$

$$2. \rho(X, X) = \frac{\text{cov}(X, X)}{\text{Var}(X)} = \frac{\text{Var}(X)}{\text{Var}(X)} = 1.$$

$$\rho(X, -X) = \frac{\text{cov}(X, -X)}{\sigma(X)\sigma(-X)} = \frac{-\text{cov}(X, X)}{\text{Var}(X)} = -1.$$

3.

$$\begin{aligned}
\rho(aX + b, cY + d) &= \frac{\text{cov}(aX + b, cY + d)}{\sqrt{\text{Var}(aX + b)\text{Var}(cY + d)}} \\
&= \frac{\text{accov}(X, Y)}{\sqrt{a^2\text{Var}(X)c^2\text{Var}(Y)}} \\
&= \pm \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \\
&= \pm \rho(X, Y).
\end{aligned}$$

4. Si X et Y sont indépendantes : $\text{cov}(X, Y) = 0 \Rightarrow \rho(X, Y) = 0$. La réciproque est fausse.

□

Exercice : Soit X une variable aléatoire continue, admettant une densité de probabilité f et ayant les moments des deux premiers ordres finis. On suppose que f est paire.

1. Montrer que $\mathbb{E}(X) = 0$,
2. Montrer que $\rho(X, |X|) = 0$, conclure.

Solution :

1. On a $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$,
on pose $u = -x$

$$\begin{aligned}
\mathbb{E}(X) &= \int_{+\infty}^{-\infty} -(u) f(-u) du \\
&= \int_{+\infty}^{-\infty} u f(u) du \\
&= - \int_{-\infty}^{+\infty} u f(u) du \\
&= -\mathbb{E}(X).
\end{aligned}$$

D'où : $\mathbb{E}(X) = 0$.

2. On a :

$$\begin{aligned}\rho &= \frac{\text{cov}(X, |X|)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \\ \text{cov}(X, |X|) &= \mathbb{E}(X, |X|) - \mathbb{E}(X)\mathbb{E}(|X|), \\ &= \mathbb{E}(X, |X|) \\ \mathbb{E}(X, |X|) &= \int_{-\infty}^{+\infty} x|x|f(x)dx \\ &= \int_0^{+\infty} x^2 f(x)dx - \int_{-\infty}^0 x^2 f(x)dx \\ &= \int_0^{+\infty} x^2 f(x)dx - \int_0^{-\infty} x^2 f(x)dx.\end{aligned}$$

Comme $\int_0^{-\infty} x^2 f(x)dx = -\int_0^{+\infty} x^2 f(-x)dx = -\int_0^{+\infty} x^2 f(x)dx$ donc $\mathbb{E}(X, |X|) = 0$
d'où : $\rho = 0$ mais X et $|X|$ ne sont pas indépendantes.

1.1.6 Changement de variables bidimensionnelles

Considérons deux variables aléatoires X, Y conjointement continues de densité $f(x, y)$.

On s'intéresse à la densité conjointe de U, V fonction de X et Y .

i.e : supposons que $U = g_1(X, Y), V = g_2(X, Y)$, g_1, g_2 vérifient les deux conditions suivantes :

a) On peut résoudre par rapport à x, y le système d'équations $u = g_1(x, y), v = g_2(x, y)$.

Les solutions sont notées $x = h_1(u, v), y = h_2(u, v)$.

b) g_1, g_2 sont continûment différentiables de Jacobien non nul, i.e :

$$J(x, y) = \begin{vmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{vmatrix} = \frac{\partial g_1}{\partial x} \frac{\partial g_2}{\partial y} - \frac{\partial g_1}{\partial y} \frac{\partial g_2}{\partial x} \neq 0$$

Sous ces conditions on peut montrer que les variables U et V sont continues de densité :

$$f_{(U,V)}(u, v) = f_{(X,Y)}(x, y) |J(x, y)|^{-1} \quad (1.2)$$

où il faut remplacer x par $h_1(u, v)$ et y par $h_2(u, v)$.

En effet :

$$\mathbb{P}(U \leq u, V \leq v) = \int \int_{\Delta} f_{(X,Y)}(x, y) dx dy$$

où $\Delta = \{(x, y) : g_1(x, y) \leq u, g_2(x, y) \leq v\}$.

En dérivant par rapport à u et v on obtient la formule (1.2).

Exemple 1.7. Soit (X, Y) un couple de variables aléatoires absolument continues de densité f . Posons $U = X + Y$ et $V = X - Y$. Exprisons la densité du couple (U, V) en fonction de celle de (X, Y) .

Posons

$$\begin{cases} g_1(x, y) = x + y = u \\ g_2(x, y) = x - y = v \end{cases} \Rightarrow \begin{cases} x = \frac{u+v}{2} \\ y = \frac{u-v}{2} \end{cases}$$

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \neq 0$$

$$f_{(U,V)}(u, v) = \frac{1}{2} f_{(X,Y)}\left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

Application numérique : Si X, Y sont indépendantes et uniforme sur $[0, 1]$, on a :

$$f_{(X,Y)}(x, y) = \begin{cases} 1, & \text{si } (x, y) \in [0, 1]^2; \\ 0, & \text{sinon.} \end{cases}$$

$$f_{(U,V)}(u, v) = \begin{cases} 1/2, & 0 \leq u+v \leq 2 \text{ et } 0 \leq u-v \leq 2; \\ 0, & \text{sinon.} \end{cases}$$

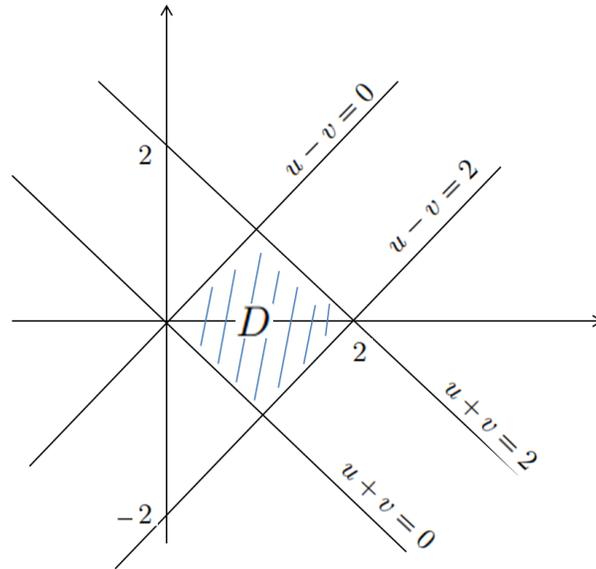
On a $D = \{(u, v) \in \mathbb{R}^2 \mid 0 \leq u+v \leq 2, 0 \leq u-v \leq 2\}$.

Vérifions que $f_{U,V}$ est bien une densité de probabilité :

$$\begin{aligned} I &= \int \int_D f_{(U,V)}(u, v) du dv \\ &= 1/2 \int \int du dv \\ &= 1/2 \int_0^1 du \int_{-u}^u dv + 1/2 \int_1^2 du \int_{u-2}^{2-u} dv \\ &= 2 \cdot 1/2 \int_0^1 \left(\int_{-u}^u dv \right) du \\ &= \int_0^1 2udu = 1. \end{aligned}$$

Cas général :

Soit X_1, X_2, \dots, X_n n variables aléatoires de densité conjointe connue. Si l'on s'intéresse à

FIG. 1.4 – Support du couple (U, V)

la densité conjointe et Y_1, Y_2, \dots, Y_n où $Y_1 = g_1(X_1, X_2, \dots, X_n)$,

$Y_2 = g_2(X_1, X_2, \dots, X_n), \dots, Y_n = g_n(X_1, X_2, \dots, X_n)$ on admettra que les g_i ont des dérivées partielles continues et que :

$$J(x_1, x_2, \dots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0$$

et on admet que le système d'équation :

$$\begin{cases} Y_1 = g_1(X_1, X_2, \dots, X_n), \\ Y_2 = g_2(X_1, X_2, \dots, X_n), \\ \vdots, \\ Y_n = g_n(X_1, X_2, \dots, X_n), \end{cases}$$

admet une solution unique notée :

$$\begin{cases} x_1 = h_1(y_1, y_2, \dots, y_n), \\ x_2 = h_2(y_1, y_2, \dots, y_n), \\ \vdots, \\ x_n = h_n(y_1, y_2, \dots, y_n), \end{cases}$$

Si ces conditions sont réalisées, la densité conjointe des variables Y_i est :

$$f_{(Y_1, Y_2, \dots, Y_n)}(y_1, y_2, \dots, y_n) = f_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) |J(x_1, x_2, \dots, x_n)|^{-1}$$

où : $x_i = h_i(y_1, y_2, \dots, y_n)$ pour $x = \overline{1, n}$.

1.1.7 Régression

Les densités conditionnelles permettent de calculer les moments conditionnels, comme par exemple les espérances.

On peut définir notamment la fonction :

$$x \longmapsto \mathbb{E}(Y|X = x)$$

qui s'appelle fonction de régression (non linéaire), son graphe étant la courbe de régression de Y en X .

a) Espérance conditionnelle :

- Pour une loi absolument continue, on obtient :

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y) dy = \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)} dy.$$

- Pour une loi discrète, on obtient :

$$\begin{aligned} \mathbb{E}(Y|X = x) &= \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y|X = x) \\ &= \sum_{y \in Y(\Omega)} y \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}. \end{aligned}$$

- Notons que $\mathbb{E}(Y|X)$ est une variable aléatoire, comme fonction de la variable aléatoire X et que ses réalisations sont les valeurs $\mathbb{E}(Y|X = x)$ avec les probabilités $\mathbb{P}(X = x)$. On peut donc calculer l'espérance mathématique de cette variable aléatoire.

Théorème de l'espérance totale

Propriété (linéarité) :

$$\mathbb{E}(Y_1 + Y_2|X) = \mathbb{E}(Y_1|X) + \mathbb{E}(Y_2|X).$$

Théorème 1.1 (Théorème de l'espérance totale).

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y).$$

Ce théorème est un outil très puissant pour calculer l'espérance mathématique d'une loi compliquée dont les lois conditionnelles sont simples.

Démonstration. Cas absolument continu :

$$\begin{aligned} \mathbb{E}[\mathbb{E}(Y|X)] &= \int_{-\infty}^{+\infty} \mathbb{E}(Y|X = x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)} f_X(x) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f(x, y) dx \right) dy = \int_{-\infty}^{+\infty} y f_Y(y) dy = \mathbb{E}(Y). \end{aligned}$$

Cas discret :

$$\begin{aligned} \mathbb{E}[\mathbb{E}(Y|X)] &= \sum_{x \in X(\Omega)} \mathbb{E}(Y|X = x) \mathbb{P}(X = x) \\ &= \sum_{x \in X(\Omega)} \left(\sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y|X = x) \right) \mathbb{P}(X = x) \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} y \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} \mathbb{P}(X = x) \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} y \mathbb{P}(X = x, Y = y) \\ &= \sum_{y \in Y(\Omega)} y \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y) \\ &= \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y) = \mathbb{E}(Y). \end{aligned}$$

□

b) Variance conditionnelle

On peut également calculer la variance conditionnelle :

$$\begin{aligned} \text{Var}(Y|X = x) &= \mathbb{E}[(Y - \mathbb{E}(Y|X = x))^2|X = x] \\ \text{Var}(Y|X = x) &= \mathbb{E}[Y^2|X = x] - \mathbb{E}^2[Y|X = x] \end{aligned} \quad (1.3)$$

De même $\text{Var}(Y|X = x)$ est une variable aléatoire, étant fonction de X , on peut calculer son espérance.

Théorème 1.2 (Théorème de la variance totale).

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}[\mathbb{E}(Y|X)]. \quad (1.4)$$

Démonstration.

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}[Y - \mathbb{E}(Y)]^2 \\ &= \mathbb{E}[Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \mathbb{E}(Y)]^2 \\ &= \mathbb{E}[(Y - \mathbb{E}(Y|X))^2] + 2\mathbb{E}[(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - \mathbb{E}(Y))] + \mathbb{E}[(\mathbb{E}(Y|X) - \mathbb{E}(Y))^2] \end{aligned}$$

Le premier terme n'est autre que $\mathbb{E}[\text{Var}(Y|X)]$, en effet en appliquant le théorème de l'espérance totale :

$$\mathbb{E}[(Y - \mathbb{E}(Y|X))^2] = \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}(Y|X))^2|X]].$$

D'où :

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}[\mathbb{E}(Y|X)]. \quad (1.5)$$

□

Interprétation géométrique : Soit L_X^2 le sous espace de L^2 constitué des variables aléatoires fonction seulement de X de type $\varphi(X) = \mathbb{E}(Y|X)$.

L_X^2 est convexe et contient la droite des constantes D . C'est donc un sous espace de Hilbert fermé.

Alors $\mathbb{E}(Y|X)$ s'interprète comme la projection orthogonale de Y sur L_X^2 .

En effet $\mathbb{E}(Y|X)$ est un opérateur linéaire, pour montrer que c'est un projecteur orthogonal il suffit de vérifier qu'il est idempotent et auto-adjoint.

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

Auto-adjoint $\langle Z, \mathbb{E}(Y|X) \rangle = \langle \mathbb{E}(Z|X), Y \rangle$.

Le minimum de $\mathbb{E}[(Y - \varphi(X))^2]$ est atteint pour $\varphi(X) = \mathbb{E}(Y|X)$.

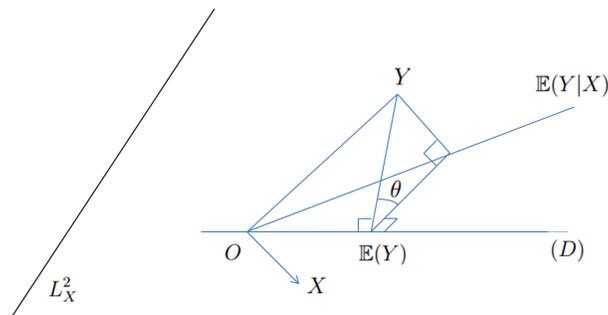


FIG. 1.5 – Représentation géométrique de $\mathbb{E}(Y|X)$

Le théorème de la variance totale s'interprète comme le théorème de pythagore appliqué au triangle rectangle $Y, \mathbb{E}(Y), \mathbb{E}(Y|X)$.

$$\begin{aligned} \|Y - \mathbb{E}(Y)\|^2 &= \text{Var}(Y) = \|\mathbb{E}(Y|X) - \mathbb{E}(Y)\|^2 + \|Y - \mathbb{E}(Y|X)\|^2 \\ &= \text{Var}(\mathbb{E}(Y|X)) + \mathbb{E}[(Y - \mathbb{E}(Y|X))^2] \\ &= \text{Var}(\mathbb{E}(Y|X)) + \mathbb{E}[\mathbb{E}(Y - \mathbb{E}(Y|X))^2] \\ &= \text{Var}(\mathbb{E}(Y|X)) + \mathbb{E}[\text{Var}(Y|X)]. \end{aligned}$$

Exemple 1.8. Soit (X, Y) un couple aléatoire défini par la densité :

$$f(x, y) = \begin{cases} e^{-y}, & 0 \leq x \leq y; \\ 0, & \text{sinon.} \end{cases}$$

1. Déterminer la fonction de répartition du couple (X, Y) .
2. Déterminer la loi marginale de X .
3. Déterminer la loi conditionnelle de $Y|X = x$.
4. Calculer $\mathbb{E}(Y|X = x)$ et $\mathbb{E}[\mathbb{E}(Y|X)]$.

Solution :

1.

$$\begin{aligned}
 F(x, y) &= \mathbb{P}(X \leq x, Y \leq y) \\
 &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \\
 &= \int_0^x \int_u^y f(u, v) du dv \text{ si } (x, y) \in D \\
 &= \int_0^x \left(\int_u^y e^{-y} dv \right) du \\
 &= \int_0^x [-e^{-v}]_u^y du \\
 &= \int_0^x (e^{-u} - e^{-y}) du \\
 &= -e^{-x} - xe^{-y} + 1
 \end{aligned}$$

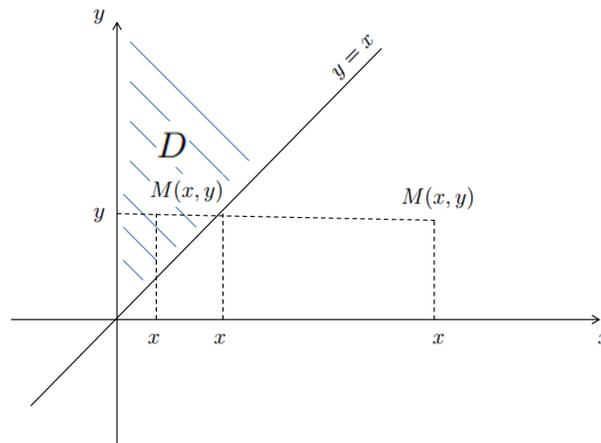


FIG. 1.6 – Support de (X, Y)

$$F(x, y) = 1 - e^{-x} - xe^{-y} \text{ si } (x, y) \in D$$

si $x \leq 0$ et $y \leq 0$: $F(x, y) = 0$,

si $x \geq y \geq 0$: $F(x, y) = F(y, y) = 1 - e^{-y} - ye^{-y}$

D'où :

$$F : \mathbb{R}^2 \longrightarrow [0, 1]$$

$$(x, y) \longmapsto F(x, y) = \begin{cases} 0, & \text{si } x \leq 0, \quad y \leq 0; \\ 1 - e^{-x} - xe^{-y}, & 0 \leq x \leq y; \\ 1 - e^{-y} - ye^{-x}, & 0 \leq y \leq x. \end{cases} \quad (1.6)$$

2. Loi marginale de X

$$F_X(x) = F(x, +\infty) = 1 - e^{-x} \quad x > 0$$

$$f_X(x) = F'_X(x) = e^{-x} \quad (X \rightsquigarrow \mathcal{Exp}(1)) \text{ ou bien } f_X(x) = \int_x^{+\infty} e^{-y} dy = e^{-x}$$

$$F_Y(y) = F(+\infty, y) = F(y, y) = 1 - e^{-y} - ye^{-y}$$

$$f_Y(y) = ye^{-y}, \quad y > 0 \text{ ou bien } f_Y(y) = \int_0^y e^{-y} dx = ye^{-y}.$$

3. $f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{x-y}$ pour x fixé

$$\begin{aligned} \mathbb{E}(Y|X = x) &= \int_x^{+\infty} y f_{Y|X}(y) dy \\ &= \int_x^{+\infty} (ye^{x-y}) dy \\ &= e^x \int_x^{+\infty} ye^{-y} dy \\ &= x + 1. \end{aligned}$$

D'où $\mathbb{E}(Y|X) = X + 1$.

$$\begin{aligned} \mathbb{E}[\mathbb{E}(Y|X)] &= \mathbb{E}(X) + 1 \\ &= \int_0^{+\infty} e^{-x} dx + 1 = 2. \end{aligned}$$

$$\begin{aligned} \mathbb{E}(Y) &= \int_0^{+\infty} y f_Y(y) dy \\ &= \int_0^{+\infty} y^2 e^{-y} dy \\ &= -y^2 e^{-y} \Big|_0^{+\infty} + \int_0^{+\infty} 2ye^{-y} dy = 2. \end{aligned}$$

1.1.8 Vecteurs aléatoires gaussiens

Rappels

Une variable aléatoire réelle X est dite gaussienne (normale) centrée réduite ($\mathbb{E}(X) = 0$, $Var(X) = 1$), si elle admet comme densité de probabilité :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}$$

et on note $X \rightsquigarrow \mathcal{N}(0, 1)$.

Une variable aléatoire réelle $Y \rightsquigarrow \mathcal{N}(m, \sigma^2)$ (normale de moyenne m et de variance σ^2) si $Y = \sigma X + m$ avec $X = \frac{Y-m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$ et :

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-m)^2}, \quad y \in \mathbb{R}.$$

Définition 1.6. Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ de \mathbb{R}^n est dit gaussien si toute combinaison linéaire de ses composantes est une variable aléatoire réelle gaussienne.

Remarque 1.5. Il est d'usage d'identifier un vecteur de \mathbb{R}^n à des matrices colonnes, ainsi on écrira le vecteur $X = (X_1, X_2, \dots, X_n)$ sous la forme $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$

et alors

$X^t = (X_1, X_2, \dots, X_n)$ (est une matrice ligne).

Définition 1.7. Si $X : \Omega \longrightarrow \mathbb{R}^n$ est un vecteur aléatoire de carré intégrable (i.e : $\mathbb{E}(X_i^2) < \infty$), on définit sa matrice des covariances, avec les notations matricielles par :

$$\Gamma = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^t].$$

avec $[X - \mathbb{E}(X)][X - \mathbb{E}(X)]^t = \begin{pmatrix} X_1 - \mathbb{E}(X_1) \\ \vdots \\ X_n - \mathbb{E}(X_n) \end{pmatrix} (X_1 - \mathbb{E}(X_1), \dots, X_n - \mathbb{E}(X_n))$ est une matrice carrés d'ordre n dont les termes sont : $(X_i - \mathbb{E}(X_i), \dots, X_j - \mathbb{E}(X_j)) \quad 1 \leq i, j \leq n$ et :

$$\Gamma = (cov(X_i, X_j)_{1 \leq i, j \leq n}).$$

Remarque 1.6. La matrice de covariance est symétrique et définie positive i.e : $\Gamma^t = \Gamma$ et $u^t \Gamma u > 0$, $\forall u \in \mathbb{R}^{n*}$, (i.e : la forme quadratique associée est une forme positive).

Propriétés des vecteurs gaussiens

La loi du vecteur aléatoire gaussien $X = (X_1, \dots, X_n)$ est entièrement déterminée par la donnée de $M = \mathbb{E}(X)$ et de la matrice Γ de covariance. On note :

$$X \rightsquigarrow \mathcal{N}_n(M, \Gamma).$$

D'après la définition, les composantes d'un vecteur gaussien sont des variables aléatoires réelles gaussiennes. La réciproque est fautive. Cependant si toutes les composantes du vecteur X sont des variables aléatoires gaussiennes indépendantes alors le vecteur X est gaussien.

Densité de probabilité : Le vecteur gaussien X admet une densité de probabilité f si et seulement si $\det(\Gamma) \neq 0$ et :

$$f(\underline{x}) = f(x_1, x_2, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \Gamma}} \exp \left[\frac{-1}{2} (\underline{x} - M)^t \Gamma^{-1} (\underline{x} - M) \right],$$

où $X^t = (X_1, X_2, \dots, X_n)$.

Vecteurs gaussiens à deux dimensions

Soit (X, Y) un vecteur aléatoire gaussien tel que :

$$X \rightsquigarrow \mathcal{N}(\mathbb{E}(X), \sigma_X^2), \quad Y \rightsquigarrow \mathcal{N}(\mathbb{E}(Y), \sigma_Y^2).$$

Soit $\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$ le coefficient de corrélation linéaire entre X et Y .

la matrice de covariance est :

$$\begin{aligned} \Gamma &= \begin{pmatrix} \sigma_X^2 & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \sigma_Y^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix} \end{aligned}$$

$$\det(\Gamma) = \sigma_X^2 \sigma_Y^2 (1 - \rho^2)$$

$$\Gamma^{-1} = \frac{1}{\det(\Gamma)} \begin{pmatrix} \sigma_Y^2 & -\rho \sigma_X \sigma_Y \\ -\rho \sigma_X \sigma_Y & \sigma_X^2 \end{pmatrix}.$$

La densité de probabilité du couple (X, Y) est donnée par :

$$f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho^2}} \exp \left[\frac{-1}{2} (x - \mathbb{E}(X), y - \mathbb{E}(Y)) \Gamma^{-1} \begin{pmatrix} x - \mathbb{E}(X) \\ y - \mathbb{E}(Y) \end{pmatrix} \right]$$

En notant A la quantité

$$A = (x - \mathbb{E}(X), y - \mathbb{E}(Y))\Gamma^{-1} \begin{pmatrix} x - \mathbb{E}(X) \\ y - \mathbb{E}(Y) \end{pmatrix}.$$

Alors

$$\begin{aligned} A &= \frac{1}{\det(\Gamma)}(x - \mathbb{E}(X), y - \mathbb{E}(Y)) \begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix} \begin{pmatrix} x - \mathbb{E}(X) \\ y - \mathbb{E}(Y) \end{pmatrix} \\ &= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)} [\sigma_Y^2(x - \mathbb{E}(X))^2 - 2\rho\sigma_X\sigma_Y(x - \mathbb{E}(X))(y - \mathbb{E}(Y)) + \sigma_X^2(y - \mathbb{E}(Y))^2] \\ &= \frac{1}{(1-\rho^2)} \left[\frac{(x - \mathbb{E}(X))^2}{\sigma_X^2} - 2\rho\frac{(x - \mathbb{E}(X))(y - \mathbb{E}(Y))}{\sigma_X\sigma_Y} + \frac{(y - \mathbb{E}(Y))^2}{\sigma_Y^2} \right] \end{aligned}$$

En remplaçant dans $f(x, y)$ on aura :

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ &\times \exp \left[\frac{-1}{2(1-\rho^2)} \left(\frac{(x - \mathbb{E}(X))^2}{\sigma_X^2} - 2\rho\frac{(x - \mathbb{E}(X))(y - \mathbb{E}(Y))}{\sigma_X\sigma_Y} + \frac{(y - \mathbb{E}(Y))^2}{\sigma_Y^2} \right) \right] \end{aligned}$$

Remarque 1.7. 1. Supposons que $\text{cov}(X, Y) = 0 \Rightarrow \rho = 0 \Rightarrow X, Y$ indépendantes,

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left[\frac{-1}{2} \left(\frac{(x - \mathbb{E}(X))^2}{\sigma_X^2} + \frac{(y - \mathbb{E}(Y))^2}{\sigma_Y^2} \right) \right].$$

2. Si $Z = (X, Y)$ vecteur aléatoire tel que X, Y sont deux variables aléatoires indépendantes avec $X \rightsquigarrow \mathcal{N}(0, 1)$, $Y \rightsquigarrow \mathcal{N}(0, 1)$ alors :

$$\begin{aligned} g(z) = g(x, y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}. \end{aligned}$$

Et si $X = (X_1, \dots, X_n)$ un n -uplet de variables aléatoires indépendantes tel que

$X_i \rightsquigarrow \mathcal{N}(0, 1)$ alors :

$$\begin{aligned} g(\underline{x}) = g(x_1, \dots, x_n) &= \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} \right) \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2} \quad x_i \in \mathbb{R} \quad i = \overline{1, n}. \end{aligned}$$

Exemple 1.9. Soit $X \rightsquigarrow \mathcal{N}(0, 1)$ et T une variable aléatoire définie par :

$\mathbb{P}(T = 1) = \mathbb{P}(T = -1) = \frac{1}{2}$, deux variables aléatoires indépendantes. On définit la variable aléatoire $Y = TX$, alors :

$$\begin{aligned} \mathbb{P}(Y < y) &= \mathbb{P}(TX < y) \\ &= \mathbb{P}((TX < y) \cap (T = 1)) + \mathbb{P}((TX < y) \cap (T = -1)) \\ &= \frac{1}{2}\mathbb{P}(X < y) + \frac{1}{2}\mathbb{P}(-X < y) \\ &= \frac{1}{2}[\phi(y) + 1 - \phi(-y)] \\ &= \phi(y). \end{aligned}$$

Donc : $Y \rightsquigarrow \mathcal{N}(0, 1)$.

Par ailleurs :

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(XY) \\ &= \mathbb{E}(X \cdot TX) \\ &= \mathbb{E}(T)\mathbb{E}(X^2) = 0. \end{aligned}$$

Si on conclut que X et Y étaient indépendantes, la variable aléatoire $X + Y \rightsquigarrow \mathcal{N}(0, 2)$

or :

$$\begin{aligned} \mathbb{P}(X + Y = 0) &= \mathbb{P}(X + TX = 0) = \mathbb{P}((1 + T)X = 0) \\ &= \mathbb{P}(1 + T = 0) = \mathbb{P}(T = -1) = \frac{1}{2}. \end{aligned}$$

contradiction avec la loi $X + Y$ est absolument continue.

Conclusion : X, Y sont deux variables aléatoires normales tel que $\text{cov}(X, Y) = 0$ mais X, Y sont indépendantes. $\begin{pmatrix} X \\ TX \end{pmatrix}$ n'est pas gaussien.

Propriété : Si X, Y sont des composantes d'un vecteur gaussien ; X, Y indépendant $\Leftrightarrow \text{cov}(X, Y) = 0$.

Exemple 1.10. Soit (X, Y) un vecteur gaussien de densité :

$$f(x, y) = K \exp \left[\frac{-1}{2} \left(x^2 - \frac{2x(y+5)}{4} + \frac{(y+5)^2}{4} \right) \right]$$

1. Quelle est la matrice de covariance, déterminer K .
2. Donner les lois marginales de X et Y .

Solution :

On a :

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{(X-\mathbb{E}(X))^2}{\sigma_X^2} - 2\rho\frac{(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))}{\sigma_X\sigma_Y} + \frac{(Y-\mathbb{E}(Y))^2}{\sigma_Y^2}\right)\right].$$

par identification on déduit : $\sigma_X^2 = \text{Var}(X) = \frac{1}{1-\rho^2}$ et $\sigma_Y^2 = \text{Var}(Y) = \frac{4}{1-\rho^2}$

$$\begin{aligned} \frac{\rho}{\sigma_X\sigma_Y(1-\rho^2)} &= \frac{1}{4} \Rightarrow \frac{\rho}{\frac{1}{\sqrt{1-\rho^2}}\frac{2}{\sqrt{1-\rho^2}}(1-\rho^2)} = \frac{1}{4} \\ &\Rightarrow \frac{\rho}{2} = \frac{1}{4} \Rightarrow \rho = \frac{1}{2} \end{aligned}$$

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X\sigma_Y} \Rightarrow \text{cov}(X, Y) = \rho\sigma_X\sigma_Y = 4/3$$

$\sigma_X^2 = 4/3$ et $\sigma_Y^2 = 16/3$, d'où :

$$\Gamma = \begin{pmatrix} 4/3 & 4/3 \\ 4/3 & 16/3 \end{pmatrix}$$

$$K = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} = \frac{1}{2\pi \cdot 2\sqrt{4/3} \cdot \sqrt{4/3}} = \frac{\sqrt{3}}{8\pi}.$$

$$\mathbb{E}(X) = 0, \mathbb{E}(Y) = -5$$

d'où : $X \rightsquigarrow \mathcal{N}(0, 4/3)$ et $Y \rightsquigarrow \mathcal{N}(-5, 16/3)$

$$f_X(x) = \frac{\sqrt{3}}{2\sqrt{2\pi}} e^{-\frac{3}{8}x^2}, \quad x \in \mathbb{R}.$$

$$f_Y(y) = \frac{\sqrt{3}}{4\sqrt{2\pi}} e^{-\frac{3}{32}(y+5)^2}, \quad y \in \mathbb{R}.$$

Transformation affine d'un vecteur gaussien

Soit X un vecteur de loi $\mathcal{N}_n(\mu, \Lambda)$, de composantes X_i , $i = \overline{1, n}$.

Soit $Y = AX + b$ avec $A \in \mathcal{M}_{\mathbb{R}}(m, n)$ et $b \in \mathbb{R}^m$, alors Y est aussi gaussien.

En effet

si $a \in \mathbb{R}^m$

$$\begin{aligned} t_a Y &= (t_a A)X + a^t b \\ &= \sum_{i=1}^n \alpha_i X_i + B \rightsquigarrow \mathcal{N}_1(\mu, \Lambda) \end{aligned}$$

On pose $t_\alpha = t_a A$: vecteur de composantes α_i , $1 \leq i \leq n$ et $B = t_a b \in \mathbb{R}$.

On a $t_\alpha X$ est une variable aléatoire gaussienne et donc $t_\alpha X + B$ est gaussienne, ainsi toute combinaison linéaire des composantes de Y suit une loi normale $\Rightarrow Y$ est gaussienne de \mathbb{R}^m .

et on a : $\mathbb{E}(Y) = A\mathbb{E}(X) + b = A\mu + b = \mu'$

et $Var(Y) = A\Lambda A^t = \Lambda'$ d'où $Y \rightsquigarrow \mathcal{N}_m(\mu', \Lambda')$.

Cas particulier d'une transformation linéaire i.e : $b = 0$.

Si $X \rightsquigarrow \mathcal{N}_n(0, I_n)$ alors $Y = AX \rightsquigarrow \mathcal{N}_m(0, AA^t)$.

Exemple 1.11. $X, Y, Z \rightsquigarrow \mathcal{N}(0, 1)$ indépendantes.

$L = (X, Y, Z) \rightsquigarrow \mathcal{N}_3(0, I_3)$.

$S = \frac{1}{6}(5X^2 + 2Y^2 + 5Z^2 + 4XY - 2XZ + 4YZ)$.

On remarque que : $S = L^t A L$ avec $A = \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$

A étant symétrique, idempotente $A^2 = A$, $\text{rang}(A) = \text{trace}(A) = 2$, $S \rightsquigarrow \chi_2^2$.

$T = L^t L - S = L^t(I - A)L \rightsquigarrow \chi_1^2$ car $I - A$ idempotente.

De plus $A(I - A) = A - A^2 = A - A = 0$ donc $S = L^t A L$ et $T = L^t(I - A)L$ sont indépendantes.

Si $Y = X^t A X$ avec $X \rightsquigarrow \mathcal{N}(0, I_n)$, $A = A^2$ symétrique et idempotente de $\text{rang}(A) = r$ alors $Y \rightsquigarrow \chi_r^2$.

1.1.9 Les lois de probabilités usuelles

Lois discrètes

Bernoulli $\mathcal{B}(p)$, binomiale $\mathcal{B}(n, p)$, Poisson $\mathcal{P}(\lambda)$, Géométrique $\mathcal{G}(p)$, hypergéométrique $\mathcal{G}(N, n, p)$ (voir cours calcul des probabilités).

1.1.10 Lois continues

Loi uniforme $\mathcal{U}_{[a,b]}$, loi exponentielle $\mathcal{Exp}(\lambda)$, loi normale $\mathcal{N}(\mu, \sigma^2)$, loi normale centrée réduite $\mathcal{N}(0, 1)$, loi de Cauchy : $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.

Loi Gamma

Une variable aléatoire X suit une loi de gamma de paramètres $a > 0$ et $\theta > 0$ si c'est une variable aléatoire positive dont la densité de probabilité est :

$$f(x) = \frac{\theta^a}{\Gamma(a)} e^{-\theta x} x^{a-1}, \quad x > 0$$

on note $X \rightsquigarrow \gamma(a, \theta)$.

La fonction gamma étant définie sur \mathbb{R}_+^* par :

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

Parmi les propriétés de la fonction Γ , on peut montrer en intégrant par parties que pour $x > 1$: $\Gamma(x) = (x-1)\Gamma(x-1)$, $\Gamma(x+1) = x\Gamma(x)$ et pour $n \in \mathbb{N}^*$: $\Gamma(n) = (n-1)!$, $\Gamma(1) = 1$, on montre aussi que : $\Gamma(1/2) = \sqrt{\pi}$

Remarque 1.8. $n! = n^n e^{-n} \sqrt{2\pi n} = \Gamma(n+1)$.

Pour calculer $\mathbb{E}(X)$, on effectue le changement de variable $y = \theta x$:

$$\begin{aligned} \mathbb{E}(X) &= \frac{\theta^a}{\Gamma(a)} \int_0^{+\infty} e^{-\theta x} x^a dx \\ &= \frac{1}{\Gamma(a)} \int_0^{+\infty} e^{-y} y^a \frac{dy}{\theta} \\ &= \frac{1}{\theta} \frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\theta}. \end{aligned}$$

$\mathbb{E}(X^2) = \frac{a(a+1)}{\theta^2}$. D'où :

$$\text{Var}(X) = \frac{a}{\theta^2}, \quad \mathbb{E}(X) = \frac{a}{\theta}.$$

loi de la variable aléatoire $Y = \theta X$:

$$G(y) = \mathbb{P}(Y < y) = \mathbb{P}(\theta X < y) = \mathbb{P}(X < \frac{y}{\theta}) = F(y/\theta)$$

où F est la fonction de répartition de X . La densité de Y est :

$$\begin{aligned} g(y) &= \frac{1}{\theta} f\left(\frac{y}{\theta}\right) \\ &= \frac{\theta^{a-1}}{\Gamma(a)} e^{-y} \left(\frac{y}{\theta}\right)^{a-1} \\ &= \frac{1}{\Gamma(a)} e^{-y} y^{a-1} \quad y > 0 \end{aligned}$$

d'où : $Y \rightsquigarrow \gamma(a, 1)$ qu'on note $\gamma(a) \equiv \gamma(a, 1)$ (dite loi d'Erlang).

$$\mathbb{E}(Y) = \text{Var}(Y) = a.$$

D'où : $X \rightsquigarrow \gamma(a, \theta) \Leftrightarrow Y = X\theta \rightsquigarrow \gamma(a) \Leftrightarrow X = \frac{Y}{\theta} \rightsquigarrow \gamma(a, \theta)$.

La somme de deux variables aléatoires indépendantes de loi gamma est encore une loi gamma : si $X \rightsquigarrow \gamma(a, \theta)$ et $Y \rightsquigarrow \gamma(b, \theta)$ sont deux variables aléatoires indépendantes alors la variable aléatoire $X + Y \rightsquigarrow \gamma(a + b, \theta)$.

Remarque 1.9. pour $a = 1$: $\gamma(1, \theta) \equiv \text{Exp}(\theta)$: $f(x) = \theta e^{-\theta x}$, $x > 0$, si $X_1, \dots, X_n \rightsquigarrow \text{Exp}(\theta)$ alors $S_n = \sum_{i=1}^n X_i \rightsquigarrow \gamma(n, \theta)$ et $\theta S_n \rightsquigarrow \gamma(n)$.

Loi de Khi-deux (très utilisée en statistique)

La loi de Khi-deux à n degré de liberté, noté χ_n^2 , est la loi $\gamma(n/2, 1/2)$ où n est un entier positif, donc la densité d'une variable aléatoire $X \rightsquigarrow \chi_n^2$ est :

$$f(x) = \frac{1}{2^{n/2} \Gamma(n/2)} e^{-x/2} x^{n/2-1} \quad x > 0$$

Ces moments se déduisent de ceux de la loi γ :

$$\mathbb{E}(\chi_n^2) = \frac{n/2}{1/2} = n \quad \text{et} \quad \text{Var}(\chi_n^2) = \frac{n/2}{1/4} = 2n.$$

Remarque 1.10. • On peut passer de la loi gamma non tabulée à la loi khi-deux qui est tabulée :

$$Y = \frac{1}{2}X \rightsquigarrow \gamma(n) \Leftrightarrow X = 2Y \rightsquigarrow \gamma(n, 1/2) \equiv \chi_{2n}^2$$

si $X \rightsquigarrow \mathcal{Exp}(\theta) \Rightarrow 2\theta X \rightsquigarrow \chi_2^2$ et si $S_n = \sum_{i=1}^n X_i$ avec $X_i \rightsquigarrow \mathcal{Exp}(\theta)$ indépendantes :

$$2\theta S_n \rightsquigarrow \chi_{2n}^2.$$

- Il existe encore un lien entre la loi γ et la loi normale qui explique son importance en statistique, si $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$ alors :

$$\left(\frac{X - \mu}{\sigma}\right)^2 \rightsquigarrow \chi_1^2.$$

En effet : la variable aléatoire $Y = \frac{X - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$ et si on pose $Z = Y^2$ sa fonction de répartition est définie pour $z > 0$ par :

$$\begin{aligned} G(z) = \mathbb{P}(Z < z) &= \mathbb{P}(-\sqrt{z} < Y < \sqrt{z}) \\ &= \phi(\sqrt{z}) - \phi(-\sqrt{z}) \end{aligned}$$

et sa densité est :

$$\begin{aligned} g(z) &= \frac{1}{2\sqrt{z}} [\phi'(\sqrt{z}) - \phi'(-\sqrt{z})] \\ &= \frac{e^{-z/2}}{\sqrt{2\pi z}} \quad (\Gamma(1/2) = \sqrt{\pi}) \end{aligned}$$

c'est la densité de $\gamma(1/2, 1/2)$ i.e : χ_1^2 .

Somme de lois Khi-deux : Si $X \rightsquigarrow \chi_n^2$ et $Y \rightsquigarrow \chi_m^2$ deux variables aléatoires indépendantes alors : $X + Y \rightsquigarrow \chi_{n+m}^2$.

Si $X_1, \dots, X_n \rightsquigarrow \mathcal{N}(0, 1)$ sont indépendantes alors : $\sum_{i=1}^n X_i^2 \rightsquigarrow \chi_n^2$.

Le degré de liberté correspond au nombre de variables aléatoires indépendantes qui la constituent. Si ces variables aléatoires étaient liées par "k" relations le degré de liberté serait $n - k$.

Loi Bêta

Il existe deux familles de lois Bêta qui se déduisent de la famille des lois gamma.

1. Loi Bêta de second espèce :

Si X et Y sont deux variables aléatoires indépendantes de lois : $\gamma(p)$, $\gamma(q)$, alors la variable aléatoire $Z = \frac{X}{Y} \rightsquigarrow \beta(p, q)$ de 2^{ème} espèce notée $\beta_{II}(p, q)$ et de densité, pour $z > 0$:

$$f(z) = \frac{1}{\beta(p, q)} \frac{z^{p-1}}{(1+z)^{p+q}}$$

où :

$$\beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

est la fonction Bêta définie sur $\mathbb{R}_+^* \times \mathbb{R}_+^*$.

Des propriétés de la fonction Γ , on déduit :

$$\mathbb{E}(Z) = \frac{p}{q-1}, \quad q > 1.$$

$$\mathbb{E}(Z^2) = \frac{p(p+1)}{(q-1)(q-2)}, \quad q > 2$$

et

$$Var(Z) = \frac{p(p+q-1)}{(q-1)^2(q-2)}, \quad q > 2.$$

$$\beta(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$$

2. Loi Bêta de première espèce :

La loi bêta de 1^{ière} est définie par rapport à la loi précédente :

$$T = \frac{X}{X+Y} = \frac{Z}{1+Z} \quad \text{avec } X \rightsquigarrow \gamma(p), Y \rightsquigarrow \gamma(q), Z \rightsquigarrow \beta_{II}(p, q)$$

T est à valeur dans $[0, 1]$, sa densité est :

$$f(t) = \frac{1}{\beta(p, q)} t^{p-1}(1-t)^{q-1}, \quad t \in [0, 1].$$

on écrit $T \rightsquigarrow \beta_I(p, q)$.

On obtient les moments :

$$\mathbb{E}(T) = \frac{p}{p+q}, \quad \mathbb{E}(T^2) = \frac{p(p+1)}{(p+q)(p+q+1)},$$

$$\text{Var}(T) = \frac{pq}{(p+q)^2(p+q+1)}$$

Remarque 1.11. Ces lois sont utilisées en statistique bayésienne pour représenter la distribution a priori de la probabilité d'un évènement.

Loi de Student :

Elle joue un rôle important dans l'estimation par intervalle de confiance.

Définition 1.8. Soit $X \rightsquigarrow \mathcal{N}(0, 1)$, et $Y \rightsquigarrow \chi_n^2$ deux variables aléatoires indépendantes. Alors la variable aléatoire $Z = \frac{X}{\sqrt{Y/n}} \rightsquigarrow$ Student à n degré de liberté.

Propriétés : si $X \rightsquigarrow \mathcal{T}(n)$:

$$\begin{cases} \mathbb{E}(X) = 0, & n > 1; \\ \text{Var}(X) = \frac{n}{n-2}, & n > 2. \end{cases}$$

Remarque 1.12. pour $n = 1$, T_1 est la loi de Cauchy de densité $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$, T_1 est le rapport de deux gaussiennes indépendantes.

Loi de Fisher-Snedecor

Définition 1.9. Soit $X_1 \rightsquigarrow \chi_{n_1}^2$ et $X_2 \rightsquigarrow \chi_{n_2}^2$ deux variables aléatoires indépendantes, alors :

$$F = \frac{X_1/n_1}{X_2/n_2} \rightsquigarrow \mathfrak{F}(n_1, n_2).$$

C'est la loi de Fisher à (n_1, n_2) degré de liberté.

Propriétés : si $X \rightsquigarrow \mathfrak{F}(n_1, n_2)$:

$$\begin{cases} \mathbb{E}(X) = \frac{n_1}{n_2-2}, & n_2 > 2; \\ \text{Var}(X) = \frac{2n_1^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}, & n_2 > 4. \end{cases}$$

Remarque 1.13. $T_n^2 = \mathfrak{F}(1, n)$.

1.2 Fonctions caractéristiques et convergence de suites de variables aléatoires

1.2.1 Fonctions caractéristiques

Fonction génératrice des moments

Définition 1.10. On appelle fonction génératrice des moments d'une variable aléatoire X discrète, à valeurs possibles entières non négatives, ($X(\Omega) \subset \mathbb{N}$), l'espérance mathématique de z^X :

$$g_X(z) = \mathbb{E}(z^X), \quad \text{avec } 0 \leq z \leq 1.$$

$$g_X(z) = \sum_{x \in X(\Omega)} z^x \mathbb{P}(X = x) \leq \sum_{x \in X(\Omega)} \mathbb{P}(X = x) = 1.$$

on a : $g_X(0) = 0$ et $g_X(1) = 1$.

Obtention des moments factoriels :

Supposons que le moment factoriel d'ordre k , de la variable aléatoire X existe i.e :

$\mu_{[k]}(X) = \mathbb{E}[X(X-1), \dots, (X-k+1)]$ existe ($k > 0$).

pour $z \leq 1$: la somme $\sum_{x \in X(\Omega)} x(x-1), \dots, (x-k+1)z^{x-k} \mathbb{P}(X = x) \leq \mu_{[k]}(X)$ est absolument et uniformément convergente en z .

Donc, on peut dériver sous signe somme :

$$\begin{aligned} g^{(k)}(z) &= \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \frac{d^{(k)} z^k}{dz^k} \\ &= \sum_{x \in X(\Omega)} \mathbb{P}(X = x) x(x-1), \dots, (x-k+1) z^{x-k}. \end{aligned}$$

et pour $z = 1$: $g_X^{(k)}(1) = \mu_{[k]}(X)$.

En particulier : $\mathbb{E}(X) = g'_X(1)$ et $\mathbb{E}(X^2) = g''(1) + g'(1)$.

Exemple 1.12. la fonction génératrice de $X \rightsquigarrow \mathcal{B}(n, p)$,

$$\mathbb{P}(X = x) = C_n^x p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n$$

$$\begin{aligned}
g(z) &= \sum_{x=0}^n C_n^x p^x q^{n-x} z^x \\
&= \sum_{x=0}^n C_n^x (pz)^x q^{n-x} \\
&= (q + pz)^n.
\end{aligned}$$

$$g'(z) = np(q + pz)^{n-1}, \text{ pour } z = 1 \Rightarrow \mathbb{E}(X) = np,$$

$$\mathbb{E}(X^2) = g''(1) + g'(1).$$

$$g''(z) = n(n-1)p^2(q + pz)^{n-2}, \text{ pour } z = 1 : \mathbb{E}(X^2) = n(n-1)p^2 + np$$

$$\begin{aligned}
\text{Var}(X) &= n(n-1)p^2 + np - n^2p^2 \\
&= np - np^2 = np(1-p) = npq.
\end{aligned}$$

Propriété :

Si deux variables aléatoires X et Y ont la même fonction génératrice, X et Y suivent la même loi de probabilité. En effet, soit g cette fonction génératrice et soit $g(z) = \sum_{i=0}^{\infty} a_i z^i$ son développement en série entière au $v(0)$, comme ce D.L est unique alors g est la fonction génératrice d'une variable aléatoire prenant les valeurs $i = 0, 1, \dots$ avec la probabilité a_i .

Remarque 1.14. On peut définir la fonction génératrice d'une variable aléatoire X par :

$$g_X(t) = \mathbb{E}(e^{tX}) = \begin{cases} \sum_{x \in X(\Omega)} e^{tx} \mathbb{P}(X = x), & \text{si } X \text{ est discrète;} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) dx, & \text{si } X \text{ est continue.} \end{cases}$$

Cette fonction peut ne pas exister car $\mathbb{E}(e^{tX})$ n'est pas toujours définie. Les moments d'ordre n sont calculés par :

$$g^{(n)}(0) = \mathbb{E}(X^n).$$

Fonctions caractéristiques

Définition 1.11. On appelle fonction caractéristique d'une variable aléatoire réelle X la fonction φ_X de la variable réelle t définie par :

$$\varphi_X(t) = \mathbb{E}(e^{itX}).$$

$$\varphi_X(t) = \mathbb{E}[\cos(tX) + i\sin(tX)] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]$$

$$\varphi_X(t) = \begin{cases} \int_{-\infty}^{+\infty} e^{itx} f(x) dx, & \text{si } X \text{ est absolument continue;} \\ \sum_{x \in X(\Omega)} e^{itx} \mathbb{P}(X = x), & \text{si } X \text{ est discrète.} \end{cases}$$

Remarque 1.15. φ_X existe toujours car $|e^{itx}| = 1 \Rightarrow |\varphi_X(t)| \leq 1$, du fait $|\varphi_X(t)| \leq \int_{-\infty}^{+\infty} |e^{itx}| f(x) dx = \int_{-\infty}^{+\infty} f(x) dx = 1$.

Pour une variable aléatoire X qui prend des valeurs entières positives : $\mathbb{P}(X = k) = p_k$ alors : $\varphi_X(t) = \sum_k p_k e^{itk} = g_X(e^{it})$.

$$\varphi_X(t) = g_X(e^{it}).$$

Propriétés :

- a) $\varphi(0) = 1$. En effet pour X absolument continue : $\varphi_X(0) = \int_{-\infty}^{+\infty} f(x) dx = 1$ et pour X discrète $\varphi_X(0) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) = 1$.
- b) Si $Y = aX + b$, $a, b \in \mathbb{R}$,

$$\varphi_Y(t) = \varphi_X(at) e^{ibt}.$$

En effet :

$$\begin{aligned} \varphi_Y(t) &= \mathbb{E}(e^{itY}) \\ &= \mathbb{E}[e^{it(aX+b)}] \\ &= \mathbb{E}[e^{itaX} e^{itb}] \end{aligned}$$

$$\varphi_Y(t) = e^{itb} \varphi_X(at).$$

Théorème 1.3. Soient X_1, \dots, X_n n variables aléatoires indépendantes alors :

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t).$$

Démonstration.

$$\begin{aligned}\varphi_{X_1+\dots+X_n}(t) &= \mathbb{E}(e^{it}) \sum_{i=1}^n X_i \\ &= \mathbb{E}\left[\prod_{j=1}^n e^{itX_j}\right] \\ &= \prod_{j=1}^n \mathbb{E}[e^{itX_j}] = \prod_{j=1}^n \varphi_{X_j}(t).\end{aligned}$$

Car si X_j sont indépendantes alors e^{itX_j} le sont aussi.

La réciproque est fautive, on peut avoir : $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)$ sans que X_1 et X_2 soient indépendantes. \square

Exemple 1.13. Soit $X_1 = X_2 = X$ une variable aléatoire de Cauchy, sa fonction caractéristique est :

$$\varphi_X(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{itx}}{1+x^2} dx = e^{-|t|} \quad (\text{intégrale calculée en utilisant les résidus.})$$

$\varphi_{2X}(t) = e^{-2|t|} = \varphi_X(t)\varphi_X(t)$, mais X n'est pas indépendante d'elle même.

Exemples

1. Si $X \rightsquigarrow \mathcal{B}(p)$: $\varphi_X(t) = \mathbb{E}(e^{itX}) = pe^{it \cdot 1} + (1-p)e^{it \cdot 0}$,

$$\varphi_X(t) = pe^{it} + q, \quad (q = 1-p).$$

2. $X \rightsquigarrow \mathcal{B}(n, p)$, $Y = \sum_{i=1}^n X_i$, $X_i \rightsquigarrow \mathcal{B}(p)$ indépendantes,

$$\varphi_Y(t) = \prod_{i=1}^n \varphi_{X_i}(t) = (pe^{it} + q)^n$$

3. $X \rightsquigarrow \mathcal{P}(\lambda)$:

$$\begin{aligned}\varphi_X(t) &= \mathbb{E}(e^{itX}) \\ &= \sum_{x=0}^{\infty} e^{itx} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{it})^x}{x!} = e^{-\lambda} e^{\lambda e^{it}}\end{aligned}$$

$$\varphi_X(t) = e^{\lambda(e^{it}-1)}.$$

4. $X \rightsquigarrow \mathcal{U}_{[-a,a]}$, on a :

$$f(x) = \begin{cases} \frac{1}{2a}, & x \in [-a, a]; \\ 0, & \text{Sinon.} \end{cases}$$

$$\begin{aligned} \varphi_X(t) &= \int_{-a}^a \frac{1}{2a} e^{itx} dx \\ &= \frac{1}{2a} \left[\frac{1}{it} e^{itx} \right]_{-a}^a \\ &= \frac{1}{2a} \left[\frac{e^{ita} - e^{-ita}}{it} \right] \\ &= \frac{1}{2a} \frac{1}{it} [i \sin ta + i \sin ta] \end{aligned}$$

$$\varphi_X(t) = \frac{\sin ta}{ta}, \quad t \in \mathbb{R}$$

5. $X \rightsquigarrow \mathcal{Exp}(\theta)$, $\theta > 0$.

$$f(x) = \theta e^{-\theta x}, \quad x > 0$$

$$\begin{aligned} \varphi_X(t) &= \theta \int_0^{+\infty} e^{-\theta x} e^{itx} dx \\ &= \theta \int_0^{+\infty} e^{-(\theta-it)x} dx \\ &= \theta \left[-\frac{1}{\theta-it} e^{-(\theta-it)x} \right]_0^{+\infty} \\ \varphi_X(t) &= \frac{\theta}{\theta-it}, \quad \forall t \in \mathbb{R}. \end{aligned}$$

6. $X \rightsquigarrow \mathcal{N}(0,1)$,

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

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{itx} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-it)^2 - \frac{t^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-it)^2} dx. \end{aligned}$$

D'où :

$$\varphi_X(t) = e^{-\frac{t^2}{2}}, \quad t \in \mathbb{R}.$$

Moments d'ordre n d'une variable aléatoire

Théorème 1.4. *Si les n premiers moments $\mathbb{E}(X^k)$, $k = 1, 2, \dots, n$ existent alors φ_X est n fois différentiable et on a :*

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k), \quad k = 1, 2, \dots, n.$$

Démonstration : Cas absolument continu.

Si $\int_{-\infty}^{+\infty} |x|f(x)dx < +\infty \Rightarrow \int_{-\infty}^{+\infty} xe^{itx}f(x)dx$ converge uniformément donc on peut dériver sous le signe \int :

$$\varphi_X'(t) = \int_{-\infty}^{+\infty} ixe^{itx}f(x)dx$$

$$\varphi_X'(0) = i \int_{-\infty}^{+\infty} xf(x)dx = i\mathbb{E}(X).$$

Donc $\varphi_X^{(k)}(t) = i^k \int_{-\infty}^{+\infty} x^k e^{itx}f(x)dx$, d'où

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k).$$

La démonstration est identique dans le cas discret.

Théorème 1.5. *Si les n premiers moments de X existent alors au $v(0)$:*

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)(it)^k}{k!} + o(t^n).$$

Démonstration : Ceci résulte du théorème précédent et de la formule de Taylor.

Exemple 1.14. $X \rightsquigarrow \mathcal{P}(\lambda)$, sa fonction caractéristique est : $\varphi_X(t) = e^{\lambda(e^{it}-1)}$.

$$\varphi_X'(t) = \lambda i e^{it} e^{\lambda(e^{it}-1)},$$

$$\varphi_X'(0) = \lambda i \Rightarrow \lambda i = i\mathbb{E}(X) \Rightarrow \mathbb{E}(X) = \lambda.$$

Inversion de la fonction caractéristique :

A chaque fonction de répartition F correspond une et une seule fonction caractéristique φ et inversement, à toute fonction caractéristique φ correspond une et une seule fonction de répartition F .

Théorème 1.6. Si φ_X est la fonction caractéristique de la variable aléatoire X , la loi de probabilité de X est déterminée par :

$$F_X(b) - F_X(a) + \frac{1}{2} [\mathbb{P}(X = a) + \mathbb{P}(X = b)] = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt, \quad \forall a, b \in \mathbb{R}.$$

En particulier, si a, b sont deux points de continuité de F_X , alors :

$$F_X(b) - F_X(a) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt, \quad \forall a, b \in \mathbb{R}.$$

Si la variable aléatoire X est discrète, la formule d'inversion est :

$$\mathbb{P}(X = x) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \varphi_X(t) dt.$$

Corollaire 1.2. Si deux variables aléatoires réelles ont la même fonction caractéristique, alors elles ont la même loi de probabilité.

Exemple 1.15. (a) La somme de deux variables aléatoires indépendantes normales est normale. Soit $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$:

$$\begin{aligned} \varphi_{X+Y}(t) &= \varphi_X(t) \cdot \varphi_Y(t) = e^{it\mu_1} e^{-\sigma_1^2 \frac{t^2}{2}} e^{it\mu_2} e^{-\sigma_2^2 \frac{t^2}{2}} \\ &= e^{it(\mu_1 + \mu_2)} e^{-(\sigma_1^2 + \sigma_2^2) \frac{t^2}{2}}. \end{aligned}$$

Donc

$$X + Y \rightsquigarrow \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

(b) La somme de deux variables aléatoires binomiales indépendantes de paramètres (n, p) et (n', p) est une binomiale de paramètre $(n + n', p)$. Soit $X \rightsquigarrow \mathcal{B}(n, p)$, $Y \rightsquigarrow \mathcal{B}(n', p)$,

$$X = \sum_{i=1}^n X_i, \quad Y = \sum_{i=1}^{n'} Y_i \quad \text{avec } X_i \rightsquigarrow \mathcal{B}(p), \quad Y_i \rightsquigarrow \mathcal{B}(p).$$

$X + Y$ est une somme de $n + n'$ variables aléatoires indépendantes de Bernoulli de paramètre p donc c'est une $\mathcal{B}(n + n', p)$.

$$\text{On a : } \varphi_X(t) = (q + pe^{it})^n, \quad \varphi_Y(t) = (q + pe^{it})^{n'},$$

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) = (q + pe^{it})^{n+n'}, \quad (\text{fonction caractéristique d'une } \mathcal{B}(n+n', p)).$$

(c) La somme de deux variables aléatoires indépendantes de Poisson de paramètres λ_1, λ_2 est une variable de Poisson de paramètre $\lambda_1 + \lambda_2$:

$$\begin{aligned}\varphi_{X+Y}(t) &= \varphi_X(t) \cdot \varphi_Y(t) = e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{it}-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^{it}-1)} \quad (\text{fonction caractéristique d'une v.a. } \mathcal{P}(\lambda_1 + \lambda_2).)\end{aligned}$$

Corollaire 1.3. Si φ_X est intégrable sur \mathbb{R} , alors la variable aléatoire X admet une densité de probabilité f continue, donnée par :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t) dt.$$

Fonction caractéristique d'une variable aléatoire multidimensionnelle

Définition 1.12. Soit $Y = (X_1, \dots, X_n)^t \in \mathbb{R}^n$ un vecteur aléatoire et $t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \in \mathbb{R}^n$.

On appelle fonction caractéristique de la variable aléatoire Y :

$$\varphi_Y(t) = \mathbb{E}[e^{it'Y}] = \mathbb{E}[e^{i(t_1 X_1 + \dots + t_n X_n)}],$$

cette fonction existe toujours avec $t' = (t_1, \dots, t_n)$, (transposée de t).

1.2.2 Fonctions caractéristiques des variables aléatoires marginales et conditionnelles

La fonction caractéristique de la variable aléatoire marginale X_1 , s'obtient à partir de la fonction caractéristique de Y en faisant : $t_2 = t_3 = \dots = t_n = 0$:

$$\varphi_{X_1}(t_1) = \mathbb{E}[e^{it_1 X_1}] = \varphi_Y(t_1, 0, \dots, 0).$$

Cas particulier : Soit (X, Y) un couple absolument continu, la fonction caractéristique du couple est :

$$\varphi_{(X,Y)}(t_1, t_2) = \mathbb{E}[e^{i(t_1 X + t_2 Y)}] = \mathbb{E}_X [e^{it_1 X} (\mathbb{E}_{Y|X}(e^{it_2 Y}))]$$

Soit : $\varphi_{(X,Y)}(t_1, t_2) = \int_{-\infty}^{+\infty} e^{it_1 x} \varphi_{Y|X=x}(t_2) f(x) dx$, où : $\varphi_{Y|X=x}$ est la fonction caractéristique de $Y|X = x$.

1.2.3 Convergence des suites de variables aléatoires

Une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles est une suite de fonctions mesurables de Ω dans \mathbb{R} . Il existe diverses façons de définir la convergence de (X_n) . On considère dans la suite, une suite (X_n) de variables aléatoires réelles et une variable aléatoire X définie sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Différents types de convergence

La convergence en probabilité (Convergence faible)

Définition 1.13. (X_n) converge en probabilité vers X si :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0,$$

On note : $X_n \xrightarrow{\mathbb{P}} X$.

Exemple 1.16. Soit $(X_n)_{n \geq 1}$, la suite de variables aléatoires réelles définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ par : $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$, $\mathbb{P}(X_n = 1) = \frac{1}{n}$.

Cette suite converge en probabilité vers 0, en effet :

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = 1) = \frac{1}{n} \longrightarrow 0.$$

Théorème 1.7. (Loi faible des grands nombres)

Soit $(X_n)_{n \in \mathbb{N}^*}$, une suite de variables aléatoires indépendantes 2 à 2, identiquement distribuées, de moyenne m et d'écart-type σ . Alors la variable aléatoire :

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} m,$$

i.e. : $\forall \varepsilon > 0, \quad \mathbb{P}(|Y_n - m| > \varepsilon) \longrightarrow 0$ quand $n \rightarrow \infty$.

Inégalité de Bienaymé-Chebychev

Inégalité de Markov :

Lemme 1.1. Soit X une variable aléatoire positive de moyenne $\mu = \mathbb{E}(X)$ finie. Pour toute constante $c > 0$, on a :

$$\mathbb{P}(X \geq c) \leq \frac{\mu}{c}.$$

Démonstration. Supposons que X est absolument continue :

$$\begin{aligned}
 \mu = \mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\
 &= \int_{x \geq c} x f(x) dx + \int_{x < c} x f(x) dx \\
 &\geq \int_{x \geq c} x f(x) dx \\
 &\geq c \int_{x \geq c} f(x) dx \\
 &\geq c \mathbb{P}(X \geq c).
 \end{aligned}$$

D'où

$$\mathbb{P}(X \geq c) \leq \frac{\mu}{c}.$$

□

Inégalité de Chebychev : Appliquons l'inégalité de Markov à la variable aléatoire $(X - \mu)^2$, en posant $c = \varepsilon^2$.

$$\mathbb{P}((X - \mu)^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}((X - \mu)^2)}{\varepsilon^2},$$

on déduit alors l'inégalité de B-C :

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \quad \text{I.B Chebychev.}$$

$$\text{ou bien } \mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad \text{en posant } k\sigma = \varepsilon.$$

Démonstration. du théorème 1.7 On a :

$$\mathbb{E}(Y_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = m \text{ et } \text{Var}(Y_n) = \frac{\sigma^2}{n}.$$

D'après l'inégalité de B.Chebychev :

$$\begin{aligned}
 \mathbb{P}(|Y_n - m| > \varepsilon) &\leq \frac{\sigma^2}{n\varepsilon^2} \\
 \lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - m| > \varepsilon) &\leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0
 \end{aligned}$$

Donc $\forall \varepsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - m| > \varepsilon) = 0$, d'où

$$Y_n \xrightarrow{\mathbb{P}} m.$$

□

Convergence presque sûre (Convergence forte)

Définition 1.14. (X_n) converge presque sûrement (P.S) vers X (ou avec une probabilité égale à 1) si :

$$\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1 \quad \text{ou bien} \quad \mathbb{P}(\lim_{n \rightarrow \infty} X_n \neq X) = 0$$

et on note $X_n \xrightarrow{p.s.} X$.

En d'autres termes, l'ensemble des points de divergence est de probabilité nulle.

$$\mathbb{P} \left[\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right] = 1.$$

$$\begin{aligned} X_n \xrightarrow{p.s.} X &\Leftrightarrow \forall \varepsilon > 0, \lim_n \mathbb{P} \left[\sup_{m \geq n} |X_m - X| > \varepsilon \right] = 0 \\ &\Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P} \left[\bigcup_{m \geq n} (|X_m - X| > \varepsilon) \right] = 0 \\ &\Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P} \left[\bigcap_{m \geq n} (|X_m - X| < \varepsilon) \right] = 1. \end{aligned}$$

On a les évènements $\bigcup_{m \geq n} (|X_m - X| > \varepsilon)$ et $(\sup_{m \geq n} |X_n - X| > \varepsilon)$ sont équivalents et ont comme complémentaire :

$$\bigcap_{m \geq n} (|X_n - X| < \varepsilon),$$

d'où l'équivalence des 3 conditions :

la première condition signifie que :

$$\sup_{m \geq n} |X_n - X| \xrightarrow{\mathbb{P}} 0 \Rightarrow |X_n - X| \xrightarrow{\mathbb{P}} 0,$$

donc la convergence P.S est plus forte que la convergence en probabilité :

$$X_n \xrightarrow{p.s.} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$$

la dernière condition exprime que, à partir d'un certain rang $N = N(\varepsilon)$, tous les évènements $(|X_n - X| < \varepsilon)$ pour $n > N$ sont réalisés avec une probabilité qui tend vers 1.

Théorème 1.8. (*Loi forte des grands nombres*)

Si (X_n) est une suite de variables aléatoires indépendantes de même loi et d'espérance m , alors :

$$X_n \xrightarrow{p.s.} m.$$

Convergence en moyenne d'ordre p

Si $\mathbb{E}(|X_n - X|^p)$ existe on a :

Définition 1.15. (X_n) converge en moyenne d'ordre p vers X si :

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0,$$

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

Remarque 1.16. 1. La convergence la plus utilisée est la convergence dans L^2 dite encore "convergence en moyenne quadratique", i.e :

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^2) = 0,$$

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} : \forall n > N : \mathbb{E}(|X_n - X|^2) < \varepsilon.$$

2. Si $X_n \xrightarrow{L^p} X$, $1 \leq p \leq \infty$ alors $X_n \xrightarrow{L^q} X$, $1 \leq q \leq p$.

3. La convergence en moyenne quadratique entraîne la convergence en probabilité.

La démonstration est une conséquence directe de l'inégalité de Chebychev.

En effet :

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|^2)}{\varepsilon^2}$$

$$\text{Si } \mathbb{E}(|X_n - X|^2) \longrightarrow 0 \Rightarrow \mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$$

convergence en moyenne quadratique \Rightarrow *convergence en \mathbb{P} .*

Convergence en loi (convergence faible)

Bien qu'elle soit la plus faible, elle est la plus utilisée en pratique car elle permet d'approximer la loi de X_n par celle de X .

Définition 1.16. La suite (X_n) converge en loi vers X de fonction de répartition F si la suite (F_n) de fonction de répartition des X_n converge vers F en tout points de continuité de F . $X_n \xrightarrow{\mathcal{L}} X$.

Remarque 1.17. 1. La convergence en loi n'est réalisée qu'aux points de continuité de F . En un point de discontinuité x de F , $F_n(x)$ peut ne pas tendre vers aucune limite, ou tendre vers une limite différente de $F(x)$.

2. La convergence en loi ne suppose pas que les variables aléatoires X_n et X soient de même loi. On montre par exemple qu'une suite de variables aléatoires binomiale ou de Poisson converge vers une variable gaussienne.

3. La convergence de $F_n(x)$ vers une valeur limite pour tout x ne suffit pas pour avoir la convergence en loi de (X_n) . Il faut que la fonction limite soit une fonction de répartition. Ainsi la suite de variables aléatoires $(X_n)_{n \geq 1}$ uniformément réparties sur $[-n, n]$:

$$F_n(x) = \begin{cases} \frac{x+n}{2n}, & \text{si } x \in [-n, n]; \\ 1, & x \geq n; \\ 0, & x \leq -n. \end{cases}$$

$$\lim_{n \rightarrow \infty} F_n(x) = \frac{1}{2}, \quad \forall x \in \mathbb{R}.$$

Cette fonction limite n'est pas une fonction de répartition par conséquent, la suite (X_n) ne converge pas en loi.

Convergence en loi et fonction caractéristique

Théorème 1.9 (de P.Lévy). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles. Si $X_n \xrightarrow{\mathcal{L}} X$ alors : quand $n \rightarrow \infty$ $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$. Réciproquement si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires réelles tel que : $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ et si φ est continue en 0 alors (X_n) converge en loi vers une variable aléatoire réelle X dont la fonction caractéristique est φ .*

Exemple 1.17. *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles de loi $\mathcal{B}(n, p)$,*

$$0 \leq p \leq 1.$$

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{avec } X \rightsquigarrow \mathcal{P}(np).$$

On a : $\varphi_{X_n}(t) = (1 - p + pe^{it})^n = \varphi_n(t)$.

posant $\lambda = np \Rightarrow p = \lambda/n \Rightarrow \varphi_n(t) = (1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^{it})^n$.

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi_n(t) &= \lim_{n \rightarrow \infty} \left[1 - \frac{\lambda}{n}(1 - e^{it}) \right]^n \\ &= \lim_{n \rightarrow \infty} \left[1 + \frac{\lambda}{n}(e^{it} - 1) \right]^n \\ &= e^{\lambda(e^{it} - 1)} = \varphi(t) \text{ fonction caractéristique de } X \rightsquigarrow \mathcal{P}(\lambda = np). \end{aligned}$$

Comme φ est continue en 0, alors : $X_n \xrightarrow{\mathcal{L}} X$, i.e :

$$(X_n \mathcal{B}(n, p)) \xrightarrow{\mathcal{L}} X \quad (X \rightsquigarrow \mathcal{P}(np)).$$

Remarque 1.18. La continuité de φ en 0 est indispensable.

Exemple 1.18. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles continues uniformes

sur $[-n, n]$:

$$f(x) = \begin{cases} \frac{1}{2n}, & x \in [-n, n]; \\ 0, & \text{sinon.} \end{cases}$$

$$\begin{aligned} \varphi_{X_n}(t) = \varphi_n(t) &= \mathbb{E}(e^{itX_n}) \\ &= \int_{-n}^n \frac{1}{2n} e^{itx} dx \\ &= \frac{1}{2nit} [e^{itn} - e^{-itn}] \\ &= \frac{\sin(nt)}{nt}. \end{aligned}$$

Si $t \neq 0$ $|\varphi_n(t)| \leq \frac{1}{nt} \xrightarrow{n \rightarrow \infty} 0 = \varphi(t) \quad \forall t \neq 0$ et $\varphi(0) = 1$.

La fonction φ n'étant pas continue en 0 : $X_n \not\xrightarrow{\mathcal{L}}$

$((X_n)$ ne converge pas en loi).

Théorème central limite

Le théorème central limite concerne le comportement asymptotique de la somme de n variables aléatoires indépendantes. Il comporte de nombreux énoncés qui indiquent des conditions suffisantes de convergence vers la loi normale. La forme la plus utilisée est le théorème Lindeberg et Lévy.

Théorème 1.10. Soit une suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ indépendantes obéissant à la même loi de probabilité, de moyenne $\mu < \infty$ et de variance $\sigma^2 < \infty$. Soit

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i, \text{ alors :}$$

$$Y_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} X \text{ de loi } \mathcal{N}(0, 1)$$

Démonstration. On a : $Y_n = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma\sqrt{n}} \right)$.

On a le développement en série de Taylor de e^{itY} s'écrit :

$$e^{itY} = 1 + itY - \frac{t^2}{2}Y^2 + \dots + \frac{(it)^n}{n!}Y^n + o(t^n)$$

$$\mathbb{E}(e^{itY}) = 1 + it\mathbb{E}(Y) - \frac{t^2}{2}\mathbb{E}(Y^2) + \dots + \frac{(it)^n}{n!}\mathbb{E}(Y^n) + \dots$$

Soit $\varphi_{X_i}(t)$ la fonction caractéristique de la variable aléatoire X_i , $i = \overline{1, n}$.

On a : $\mathbb{E} \left(\frac{X_i - \mu}{\sigma\sqrt{n}} \right) = 0$ et $\text{Var} \left(\frac{X_i - \mu}{\sigma\sqrt{n}} \right) = \frac{1}{n}$

$$\varphi_{\frac{X_i - \mu}{\sigma\sqrt{n}}}(t) = 1 - \frac{t^2}{2} \cdot \frac{1}{n} + o\left(\frac{1}{n}\right).$$

Donc :

$$\begin{aligned} \varphi_{Y_n}(t) &= \prod_{i=1}^n \varphi_{\frac{X_i - \mu}{\sigma\sqrt{n}}}(t) \\ &= \left[\varphi_{\frac{X_i - \mu}{\sigma\sqrt{n}}}(t) \right]^n \\ &= \left[1 - \frac{t^2}{2} \cdot \frac{1}{n} + o\left(\frac{1}{n}\right) \right]^n. \end{aligned}$$

Pour t fixé : $\ln(\varphi_{Y_n}(t)) = n \ln \left[1 - \frac{t^2}{2} \cdot \frac{1}{n} + o\left(\frac{1}{n}\right) \right] \longrightarrow \frac{-t^2}{2}$,

d'où : $\varphi_{Y_n}(t) \xrightarrow{n \rightarrow \infty} e^{\frac{-t^2}{2}}$ qui est la fonction caractéristique d'une variable aléatoire qui suit une $\mathcal{N}(0, 1)$.

D'où

$$Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X \quad (\mathcal{N}(0, 1)).$$

□

Application du théorème central limite :

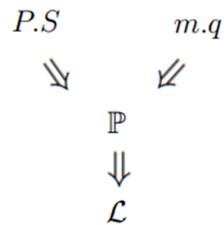
Soit X_1, X_2, \dots, X_n n variables aléatoires indépendantes de même loi de moyenne μ et de

variance σ^2 finis alors en posant : $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$

$$\frac{Y_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Relation entre les différents types de convergence :

Théorème 1.11. *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires alors :*



$X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$, la réciproque est fausse sauf si $X = a$ (p.s), $a \in \mathbb{R}$ (constante).

Propriétés : Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} a$, $a \in \mathbb{R}$ alors :

- $X_n + Y_n \xrightarrow{\mathcal{L}} X + a$.
- $X_n Y_n \xrightarrow{\mathcal{L}} aX$.
- $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{a}$ (si $a \neq 0$).

Théorème 1.12. (*Théorème de Slutsky*)

Si g est une application réelle continue alors :

$$X_n \xrightarrow{\mathcal{L}} X \Rightarrow g(X_n) \xrightarrow{\mathcal{L}} g(X).$$

Propriété :

Si la suite (X_n) est telle que : $a_n(X_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ avec $a_n \rightarrow \infty$ et $\sigma > 0$, alors, si g est une application réelle dérivable, $g(X_n) \xrightarrow{\mathcal{L}}$ avec :

$$a_n[g(X_n) - g(m)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 |g'(m)|).$$

1.2.4 Exercices

Exercice 1.1. *Etant donné deux v.a.r. indépendantes X et Y telle que*

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k) = pq^{k-1}, \quad q = 1 - p, \quad 0 < p < 1, \quad k = 1, 2, \dots$$

Calculer

$$\mathbb{P}(X = Y), \quad \mathbb{P}(X > Y), \quad \mathbb{P}(X + Y = n), \quad \mathbb{P}(X = k / X + Y = n).$$

Exercice 1.2. *Soit (X, Y) une v.a.r. discrètes à valeurs dans \mathbb{N}^2 de loi conjointe :*

$$\mathbb{P}(X = i, Y = j) = \frac{a}{2^{i+1}j!}.$$

1. Déterminer toutes les valeurs possibles de a .
2. Déterminer les lois marginales.
3. X et Y sont-elles indépendantes ?

Exercice 1.3. Soit (X, Y) un couple de v.a. de densité

$$f_{(X,Y)}(x, y) = \begin{cases} c \frac{e^{-x}}{y^2}, & \text{si } x > 0 \text{ et } y > 1; \\ 0, & \text{sinon.} \end{cases}$$

1. Calculer la constante c .
2. Déterminer les lois marginales de X et Y .
3. Les v.a. X et Y sont-elles indépendantes ?
4. Calculer $\mathbb{P}(Y > 2, X < 1)$.
5. Calculer $\mathbb{P}(XY > 1)$.

Exercice 1.4. Soit (X, Y) un couple de v.a. dont la loi est déterminée par la densité :

$$f(x, y) = \begin{cases} \frac{xy}{2}, & \text{si } 0 \leq x \leq 2 \text{ et } 0 \leq y \leq x; \\ 0, & \text{sinon.} \end{cases}$$

1. Déterminer la fonction de répartition de ce couple.
2. Déterminer les lois marginales de X et Y . Ces variables sont-elles indépendantes ?
3. Déterminer la loi conditionnelle de Y sachant $X = x$.

Exercice 1.5. Soient X et Y deux v.a. indépendantes de même loi uniforme sur $] - 1, 1[$.

Déterminer la loi de probabilité de la v.a. $Z = Y - X$.

Exercice 1.6. Soient X et Y deux variables aléatoires à valeurs dans \mathbb{N}

telles que $X \hookrightarrow \mathcal{P}(\lambda)$ et $Y/X = n \hookrightarrow \mathcal{B}(n, p)$ où $p \in]0, 1[$.

1. Montrer que $\mathbb{E}(Y) = \lambda p$.
2. On définit la variance conditionnelle de Y sachant X par

$$\mathbb{V}(Y/X = x) = \mathbb{E}(Y^2/X = x) - (\mathbb{E}(Y/X = x))^2$$

Montrer que $\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y/X)) + \mathbb{V}(\mathbb{E}(Y/X))$. En déduire la variance de Y

3. Déterminer la loi de Y .

Exercice 1.7. Soit X, Y deux variables aléatoires indépendantes de même loi $\exp(1)$, de densité e^{-x} sur \mathbb{R}^+ .

1. Soit les variables aléatoires

$$U = X - Y \quad \text{et} \quad V = X + Y$$

Donner la loi jointe du couple (U, V) .

2. En déduire la loi marginale de V .

3. Calculer $\text{Cov}(U, V)$. Pouvez-vous en déduire que U et V sont indépendantes ?

4. U et V sont-elles indépendantes.

5. Donner la densité de la loi de $T = \frac{U}{V}$.

6. Calculer $\mathbb{E}(T)$ et $\text{Cov}(T, U)$.

Exercice 1.8. Soit $X = (X_1, X_2, X_3)$ un vecteur aléatoire gaussien centré de matrice de covariance

$$A = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{pmatrix}$$

1. Le vecteur aléatoire X admet-il une densité ?

2. Déterminer a tel que les variables aléatoires X_1 et $Y = X_2 - aX_1$ soient indépendantes. Quelle est la loi de (X_1, Y) ?

Exercice 1.9. Soit (X, Y) un vecteur gaussien centré de matrice de covariance

$$\Lambda = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \text{avec} \quad |\rho| < 1.$$

On pose $U = \frac{X+Y}{\sqrt{2}}$ et $V = \frac{X-Y}{\sqrt{2}}$.

1. Déterminer la matrice de covariance de (U, V) . Déduire uniquement de cette question la loi de (U, V) . Les v.a. U et V sont-elles indépendantes ?

2. Calculer $\mathbb{E}(X^4)$.

Exercice 1.10. Soit (X, Y, Z) un vecteur Gaussien d'espérance $(1, 0, 1)$ et de matrice de covariances

$$B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 6 & 2 \\ 0 & 2 & 1 \end{pmatrix}$$

Donner la loi du vecteur $U = (X + Y, X + Z)$.

Exercice 1.11. Soit (X, Y) un couple de variables aléatoires. On suppose que X suit la loi uniforme $\mathcal{U}_{]0,1[}$. $Y|X = x$ suit la loi :

$$\mathbb{P}(Y = -2|X = x) = \frac{x}{2}, \quad \mathbb{P}(Y = 0|X = x) = 1 - x, \quad \mathbb{P}(Y = 2|X = x) = \frac{x}{2}.$$

1. Calculer $\mathbb{E}(Y)$ et $\mathbb{V}(Y)$
2. Déterminer la fonction caractéristique de la loi de Y .

Exercice 1.12. Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes de loi $\mathcal{P}(\lambda)$.

On pose $Y_n = \sum_{i=1}^n X_i$.

Montrer que

$$\frac{Y_n - \mathbb{E}(Y_n)}{\sqrt{\text{var}(Y_n)}}$$

converge en loi vers une v.a. gaussienne centrée réduite.

Exercice 1.13. Soit (X_n) une suite de v.a. indépendantes de même loi définie par

$$\mathbb{P}(X_n = 1) = p \text{ et } \mathbb{P}(X_n = -1) = 1 - p = q \text{ avec } 0 < p < 1. \text{ On pose}$$

$$Y_n = \prod_{i=1}^n X_i.$$

1. Déterminer la loi de probabilité de Y_n .
2. Déterminer la loi limite de la suite (Y_n) .

Exercice 1.14. Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes et de même loi, de densité f définie par :

$$f(x) = \begin{cases} e^{-(x-\theta)}, & \text{si } x > \theta; \\ 0, & \text{si } x \leq \theta. \end{cases}$$

où θ est un nombre positif fixé.

1. Montrer que $I_n = \inf(X_1, X_2, \dots, X_n)$ converge en moyenne quadratique vers θ
2. Etudier la convergence en loi de $Y_n = n(I_n - \theta)$.

Exercice 1.15. Soit $(X_n)_{n \geq 1}$ une s.v.a. indépendantes identiquement distribuées de loi $\mathcal{U}_{[\theta, \theta+1]}$. On définit la v.a. $Z_n = \min(X_1, X_2, \dots, X_n)$.

1. Donner la loi de Z_n .

2. Montrer que Z_n converge en probabilité vers θ .
3. Montrer que $n(Z_n - \theta)$ converge en loi vers une v.a. Z dont on donnera la loi.

Exercice 1.16. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de loi $\mathcal{U}_{[0, \theta]}$.

1. Déterminer la loi de la v.a.

$$S_n = \sup_{1 \leq i \leq n} X_i$$

2. Montrer que

$$\mathbb{E}(S_n) = \frac{n}{n+1}\theta; \quad \mathbb{V}(S_n) = \frac{n}{(n+2)(n+2)^2}\theta^2$$

3. Dédurre que (S_n) converge en probabilité vers θ .
4. On cherche à trouver la loi limite de $Y_n = n(\theta - S_n)$
 - a) Déterminer la fonction de répartition de Y_n .
 - b) En utilisant le résultat d'analyse

$$\lim_{n \rightarrow \infty} \left(1 - \frac{y}{n}\right)^n = e^{-y},$$

montrer que Y_n converge en loi vers une v.a. de loi $\mathcal{E}(\frac{1}{\theta})$

Chapitre 2

Echantillonnage

Introduction

Le mot statistique désigne à la fois un ensemble de données et l'activité qui consiste à les recueillir, les traiter et les interpréter.

Faire de la statistique c'est étudier un ensemble d'objets équivalents appelés individus ou unités statistique, sur lesquels on observe des caractéristiques appelées variables.

Recensement : observer toutes les unités statistiques d'une population finie.

Echantillon : Partie de la population.

Sondage : Etudier les unités de l'échantillon.

La démarche statistique

Statistique descriptive : Son but est de synthétiser, résumer l'information contenue dans les données. Elle utilise les tableaux statistiques et les représentations graphiques.

Statistique inférentielle : Son but est d'étendre les propriétés constatées sur l'échantillon à toute la population et valider des hypothèses a priori. Le calcul des probabilités joue un rôle important.

2.1 Position du problème

On veut à partir d'un échantillon, déduire des informations sur la population. Le problème qui se pose alors est comment choisir une partie de la population qui repro-

duit le plus fidèlement possible ses caractéristiques . C'est le problème d'échantillonnage (sondage).

2.1.1 Avantages de l'échantillonnage

- Impossibilité d'étudier toute la population lorsqu'elle est infinie.
- Le coût : Le choix d'un échantillon est de moindre coût qu'un recensement.
- Le temps : la rapidité nécessaire de certaines prises de décisions empêche le recours à un recensement.

2.1.2 Choix de l'échantillon

Il ya 3 grands modes de tirage d'un échantillon :

- Tirage empirique.
- Tirage aléatoire.
- Tirage artificiel (simulation).

a) **Echantillonnage empirique** : Utilise des connaissances préalables qu'on a sur la population. La méthode la plus utilisée est la méthode des quotas. Cette méthode se base sur la construction d'un échantillon de taille n dans lequel les proportions des individus sont égales à celles de la population. Une fois ces quotas déterminés, il faut les respecter dans le choix de l'échantillon.

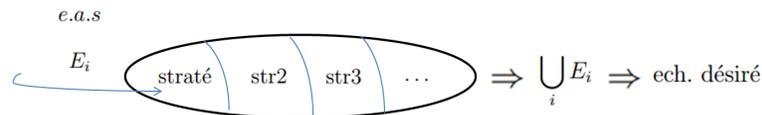
b) **Echantillonnage aléatoire** : L'échantillon est dit aléatoire si chaque individu de la population a une probabilité connue d'appartenir à l'échantillon.

* Echantillon aléatoire simple : Chaque individu a la même probabilité d'être choisi (la loi uniforme discrète).

* Echantillon aléatoire non simple : Tient compte de certain facteurs de pondération. On distingue deux types de tirage :

1. Tirage exhaustif : Sans remise (les observations ne sont pas indépendantes).
2. Tirage non exhaustif : Avec remise bernoullien (les observations sont toutes indépendantes).

- * Echantillonnage stratifié : On subdivise la population en sous ensembles (strates) relativement homogènes. On extrait de chaque strate un échantillon aléatoire simple (e.a.s). La réunion de tous ces échantillons constitue l'échantillon désiré.



Cette méthode a le but d'améliorer la précision de l'estimation.

- * L'échantillonnage par grappes : On subdivise la population en sous grappes, on tire un échantillon aléatoire de grappes. L'échantillon désiré est constitué de tous les individus de chaque grappe.

2.2 Notions fondamentales

2.2.1 Population de référence

C'est la totalité des éléments pris en considération, et sur lesquels on désire obtenir des informations.

2.2.2 Echantillon

Soit $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ une population de taille N . Soit X le caractère que l'on voudrait étudier sur cette population. Avec l'échantillon aléatoire simple : soit X_k le résultat aléatoire du $k^{\text{ième}}$ tirage, c'est une variable aléatoire qui suit la même loi que X . On note x_k le résultat du $k^{\text{ième}}$ tirage et on note (X_1, X_2, \dots, X_n) le résultat aléatoire de ces n tirages.

Définition 2.1. (X_1, X_2, \dots, X_n) est un n -uplet de variables aléatoires indépendantes de même loi (celle de X). Il est appelé n -échantillon aléatoire simple.

Définition 2.2. L'unique réalisation (x_1, x_2, \dots, x_n) est appelée ensemble des valeurs observées.

Définition 2.3. Une statistique Y sur un échantillon (X_1, X_2, \dots, X_n) est une variable aléatoire fonction mesurable des X_i :

$$Y = f(X_1, X_2, \dots, X_n) \text{ prend la valeur } f(x_1, x_2, \dots, x_n)$$

2.2.3 Distribution d'un échantillon

Soit (X_1, X_2, \dots, X_n) un échantillon aléatoire simple issu d'une variable aléatoire X de densité de probabilité f . La distribution (loi) de probabilité du n -échantillon aléatoire simple est :

$$\begin{aligned} g(x_1, x_2, \dots, x_n) = g(\underline{x}) &= f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_n}(x_n) \\ &= \prod_{i=1}^n f_{X_i}(x_i) = (f_X(x))^n \end{aligned}$$

g est appelée *vraisemblance du n -échantillon*.

Exemple 2.1. 1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une variable aléatoire

$$X \rightsquigarrow \mathcal{N}(0, 1)$$

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

$$g(\underline{x}) = \prod_{i=1}^n f_{X_i}(x_i) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}, \quad x_i \in \mathbb{R} \quad \forall i = \overline{1, n}.$$

2. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une variable aléatoire $X \rightsquigarrow \mathcal{B}(\theta)$,

$$f(x) = \theta^x (1 - \theta)^{1-x} \text{ avec } x \in \{0, 1\}.$$

La vraisemblance de cet échantillon est :

$$g(x) = \prod_{i=1}^n f_{X_i}(x_i) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

2.2.4 Fonction de répartition empirique d'un échantillon

Définition 2.4. On appelle fonction de répartition empirique d'un échantillon (e.a.s avec remise) $F_n(x)$ la proportion des n variables aléatoires X_1, X_2, \dots, X_n qui sont inférieures à x . C'est donc une fonction aléatoire (v.a pour tout x) dont les réalisations sont des fonctions

en escalier de sauts égaux à $1/n$.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(X_i < x)} = \begin{cases} 0, & \text{si } x < x_1; \\ \frac{1}{n}, & \text{si } x_1 \leq x < x_2; \\ \vdots & \\ \frac{(i-1)}{n}, & \text{si } x_{i-1} \leq x < x_i; \\ \vdots & \\ 1, & \text{si } x \geq x_n. \end{cases}$$

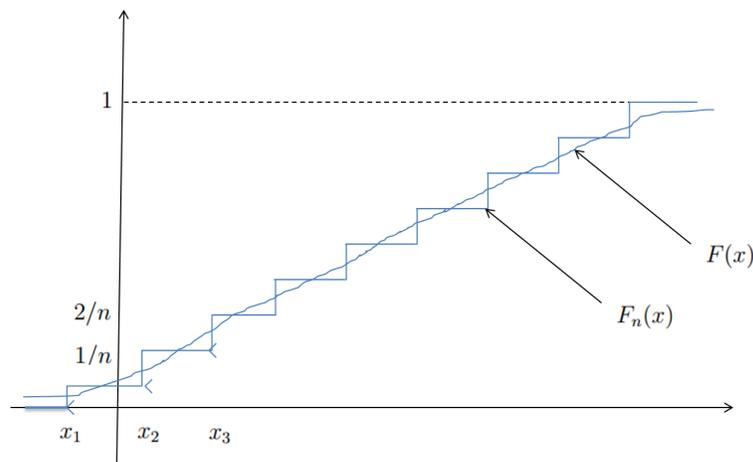


FIG. 2.1 – Graphe de la fonction de répartition empirique

Convergence de $F_n(x)$ vers $F(x)$

Théorème 2.1. Soit (X_1, X_2, \dots, X_n) un n -échantillon de fonction de répartition empirique $F_n(x)$ et $F(x)$ fonction de répartition de X (variable aléatoire parente).

Alors : $F_n(x) \xrightarrow{p.s.} F(x)$. i.e : $\mathbb{P}\left(x : \lim_{x \rightarrow \infty} F_n(x) = F(x)\right) = 1$

Démonstration. $F_n(x)$ étant une moyenne empirique de variable aléatoire réelle indépendante (puisque les X_i le sont), d'après la loi forte des grands nombres :

$$F_n(x) \xrightarrow{p.s.} \mathbb{E}[\mathbb{I}_{(X_i < x)}] = \mathbb{E}[\mathbb{I}_{(X < x)}] = F(x).$$

□

Convergence en loi de $F_n(x)$

On a : $\mathbb{1}_{(X_i < x)}$ est une variable aléatoire de Bernoulli $\mathcal{B}(F(x))$,

$$\mathbb{P}(F_n(x) = k/n) = \mathbb{P}(nF_n(x) = k) = C_n^k (F(x))^k (1 - F(x))^{n-k},$$

donc $nF_n(x) \rightsquigarrow \mathcal{B}(n, F(x))$. Alors d'après le théorème central limite :

$$\frac{nF_n(x) - nF(x)}{\sqrt{nF(x)(1 - F(x))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

i.e

$$\frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Théorème 2.2. (*Glivenko-Contelli*)

La convergence de F_n vers F est presque sûrement uniforme, i.e :

$$D_n = \sup_x |F_n(x) - F(x)| \xrightarrow{p.s.} 0.$$

Théorème 2.3. (*Kolmogorov*)

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n}D_n < y) = K(y) = \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2 y^2}.$$

Ce théorème signifie que la distribution asymptotique de la variable aléatoire D_n est connue et ne dépend pas de la variable de départ X , et permet de calculer des limites pour les valeur de D_n . La loi exacte de la variable D_n a été tabulée.

2.2.5 Statistique d'ordre d'un échantillon

Soit (X_1, X_2, \dots, X_n) un n-échantillon aléatoire simple issu d'une variable aléatoire X , les réalisations x_1, x_2, \dots, x_n peuvent être réordonnées en y_1, y_2, \dots, y_n où $y_1 < y_2 < \dots < y_n$. Les y_i constituent une permutation particulière des x_i .

Les y_i sont des réalisations du n-uplet de variables aléatoires (Y_1, Y_2, \dots, Y_n) qui constitue l'échantillon ordonné de X .

Soit F (resp. f) la fonction de répartition (resp. la densité) de la variable aléatoire X . Soit

H_k (resp. h_k) les fonctions de répartition (resp. les densités) de Y_k .

On cherche la loi de $Y_1 = \inf_{1 \leq i \leq n} X_i$ et $Y_n = \sup_{1 \leq i \leq n} X_i$.

a) Loi de $Y_1 = \inf_{1 \leq i \leq n} X_i$:

On a :

$$\begin{aligned} H_1(y) = \mathbb{P}(Y_1 < y) &= \mathbb{P}(\inf X_i < y) \\ &= 1 - \mathbb{P}(\inf X_i \geq y) \\ &= 1 - \mathbb{P}\left(\prod_{i=1}^n (X_i < y)\right) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(X_i < y) \\ &= 1 - \prod_{i=1}^n [1 - F(y)] \end{aligned}$$

$$H_1(y) = 1 - [1 - F(y)]^n.$$

$$\text{D'où : } h_1(y) = n(1 - F(y))^{n-1} f(y).$$

b) Loi de $Y_n = \sup_{1 \leq i \leq n} X_i$:

$$H_n(y) = \mathbb{P}(\sup X_i < y) = \prod_{i=1}^n \mathbb{P}(X_i \leq y),$$

$$H_n(y) = (F(y))^n,$$

$$\text{et } h_n(y) = n(F(y))^{n-1} f(y).$$

Remarque 2.1. Ces deux lois servent en particulier à détecter les valeurs aberrantes de l'échantillon : valeurs trop grandes ou trop petites.

2.2.6 Moments d'un échantillon

Définition 2.5. Soit (X_1, X_2, \dots, X_n) un n-échantillon aléatoire issu d'une variable aléatoire X et r un nombre entier ($r \in \mathbb{N}$). On appelle *moment d'ordre r* de l'échantillon et on note m'_r , la quantité :

$$m'_r = \frac{1}{n} \sum_{i=1}^n X_i^r.$$

Théorème 2.4. Soit (X_1, X_2, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X , alors $\mathbb{E}(m'_r) = m_r$, où m_r est le moment d'ordre r de la variable aléatoire X : $\mathbb{E}(m'_r) = m_r = \mathbb{E}(X_i^r)$. En effet :

$$\mathbb{E}(m'_r) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i^r\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^r] = m_r.$$

Moyenne empirique d'un échantillon

Définition 2.6. Soit $m'_r = \frac{1}{n} \sum_{i=1}^n X_i^r$ le moment d'ordre r de l'échantillon. Pour $r = 1$: $m'_1 = \frac{1}{n} \sum_{i=1}^n X_i$ est la moyenne du n -échantillon aléatoire simple (X_1, X_2, \dots, X_n) on la note \bar{X} ,

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

propriétés :

1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . Alors $\mathbb{E}(\bar{X}) = \mu$ (on dit que \bar{X} est une statistique sans biais).
En effet : $\mathbb{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mu$.
2. $Var(\bar{X}) = \frac{\sigma^2}{n}$.
En effet : $Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$.
3. $\bar{X} \xrightarrow{p.s} \mu$. (Loi forte des grands nombres).

Variance empirique d'un échantillon

Soit (X_1, X_2, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 .

On appelle variance empirique de l'échantillon la quantité :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Propriétés :

1. $\mathbb{E}(S^2) = \frac{\sigma^2(n-1)}{n}$,
2. $S^2 \xrightarrow{p.s} \sigma^2$,
3. $Var(S^2) = \frac{n-1}{n^3}[(n-1)\mu_4 - (n-3)\sigma^4]$, où $\mu_4 = \mathbb{E}(X - \bar{X})^4$.

Démonstration de (1) : $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

Décomposons S^2 :

$$\begin{aligned} X_i - \mu &= X_i - \bar{X} + \bar{X} - \mu \\ \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2 \sum_{i=1}^n (X_i - \bar{X})(\bar{X} - \mu) + \sum_{i=1}^n (\bar{X} - \mu)^2 \\ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 + (\bar{X} - \mu)^2. \end{aligned}$$

D'où :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - (\bar{X} - \mu)^2.$$

$$\begin{aligned} \mathbb{E}(S^2) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i - \bar{X})^2 - \mathbb{E}(\bar{X} - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n Var(X_i) - Var(\bar{X}) \\ &= \frac{1}{n} n \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} = \frac{\sigma^2}{n} (n-1). \end{aligned}$$

D'où :

$$\mathbb{E}(S^2) = \frac{\sigma^2}{n} (n-1).$$

Remarque 2.2. $\mathbb{E}(S^2) \neq \sigma^2$, on dit que S^2 est une statistique biaisée, son biais vaut : $\frac{\sigma^2}{n}$.

Autre démonstration :

$$\begin{aligned}
 \mathbb{E}(S^2) &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] \\
 &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right] \\
 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}^2) \\
 &= \frac{1}{n} n \mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}^2) \\
 &= \mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}^2) = \sigma^2 \frac{n-1}{n}.
 \end{aligned}$$

Cas de grands échantillons

Théorème 2.5. *Pour n assez grand :*

$$\bar{X} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ i.e. } \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

(Ceci résulte du théorème central limite)

Cas de petits échantillons

Le théorème central limite ne s'applique pas, alors on fait l'hypothèse : (X_1, \dots, X_n) est issu d'une variable aléatoire $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

Alors : $\bar{X} \longrightarrow \mathcal{N}(\mu, \sigma^2/2)$.

D'après la décomposition de S^2 :

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

En divisant par σ^2 :

$$\begin{aligned}
 \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 &= \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \frac{n}{\sigma^2} (\bar{X} - \mu)^2 \\
 &= \frac{nS^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2.
 \end{aligned}$$

On a : $\frac{X^i - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1) \Rightarrow \left(\frac{X^i - \mu}{\sigma}\right)^2 \rightsquigarrow \chi_1^2$.

D'où : $\sum_{i=1}^n \left(\frac{X^i - \mu}{\sigma}\right)^2 \rightsquigarrow \chi_n^2$, (Comme somme de n carrés de variables aléatoires indépendantes normales centrées réduites).

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1) \Rightarrow \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \rightsquigarrow \chi_1^2$$

D'où, on en déduit :

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$$

$$\text{i.e. } S^2 \rightsquigarrow \frac{\sigma^2}{n} \chi_{n-1}^2$$

$$\mathbb{E}(S^2) = \frac{\sigma^2}{n}(n-1) \quad \text{et} \quad \text{Var}(S^2) = \frac{\sigma^4}{n^2} 2(n-1).$$

2.3 Modèle statistique

Définition 2.7. On appelle *modèle statistique paramétrique* tout couple $(\chi, (\mathbb{P}_\theta)_{\theta \in \Theta})$ où χ est un ensemble dit espace des observations et $(\mathbb{P}_\theta)_{\theta \in \Theta}$ est une famille de lois de probabilités définies sur une tribu \mathcal{A} fixée de χ et Θ espace des paramètres.

Exemple 2.2. Dans une production de 10000 pièces mécaniques, une certaine proportion θ de ces pièces est défectueuses. On prélève au hasard 20 pièces.

Un modèle courant de cette situation est le modèle binômiale $(\chi, (\mathbb{P}_\theta)_{\theta \in \Theta})$, où $\Theta = [0, 1]$, $\chi = \{0, 1, \dots, 20\}$ et \mathbb{P}_θ est la loi $\mathcal{B}(20, \theta)$, θ est inconnu, donc on a affaire à un problème de statistique inférentielle.

Remarque 2.3. Si χ est fini ou dénombrable : $\mathcal{A} = \mathcal{P}(\chi)$, si $\chi = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$.

Définition 2.8. Soit $(\chi, (\mathbb{P}_\theta)_{\theta \in \Theta})$ un modèle statistique. On appelle statistique sur ce modèle toute application φ indépendante de θ , mesurable relativement à la tribu \mathcal{A} considérée sur χ à valeurs dans un certain espace Υ muni d'une tribu \mathcal{B} .

Le modèle $(\Upsilon, (q_\theta)_{\theta \in \Theta}) = (\Upsilon, (\varphi(\mathbb{P}_\theta))_{\theta \in \Theta})$, où $\forall \theta \in \Theta : q_\theta = \varphi(\mathbb{P}_\theta)$ est la loi image par φ de la loi \mathbb{P}_θ est dit modèle image par φ du modèle $(\chi, (\mathbb{P}_\theta)_{\theta \in \Theta})$.

Reprenons l'exemple précédent :

On a chacun des prélèvements est régi par une loi de Bernoulli π_θ de paramètre θ sur l'ensemble $\{0, 1\}$ (1 : pièce défectueuse, 0 : pièce non défectueuse). En supposant l'indépendance entre les prélèvements (justifié par le grand effectif de la production).

Le résultat de l'expérience est alors un 20-uplet, soit (x_1, x_2, \dots, x_n) régi par la loi puissance $\pi_\theta^{\otimes 20}$ de la loi π_θ , le modèle sous-jacent est ainsi $\left(0, 1^{\otimes 20}, (\pi_\theta)_{\theta \in [0,1]}^{\otimes 20}\right)$ est définit un modèle de Bernoulli.

Le modèle binomial considéré s'en déduit en prenant $\forall \theta \in [0, 1]$ par

$$\varphi : (x_1, \dots, x_{20}) \mapsto \sum_{i=1}^{20} x_i \text{ l'image } \mathbb{P}_\theta \text{ de } \pi_\theta^{\otimes 20}.$$

Remarque 2.4. φ joue un rôle de passage d'un modèle statistique vers un autre.

2.4 Exercices

Exercice 2.1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une v.a. X de loi $\exp(1/\lambda)$ représentant la durée de vie d'un appareil. $\lambda > 0$ inconnu.

1. A l'instant 0, on met en fonctionnement un appareil, on le remplace immédiatement par un autre, dès qu'il tombe en panne, et ainsi de suite on poursuit le processus. On observe la v.a. Z représentant le nombre de défaillances pendant un intervalle de temps $[0, a]$, ($a > 0$ fixé). Donner le modèle statistique associé.

(Indication : $\mathbb{P}(Z = n) = \mathbb{P}\left(\sum_{i=1}^n X_i \leq a, \sum_{i=1}^{n+1} X_i > a\right)$.)

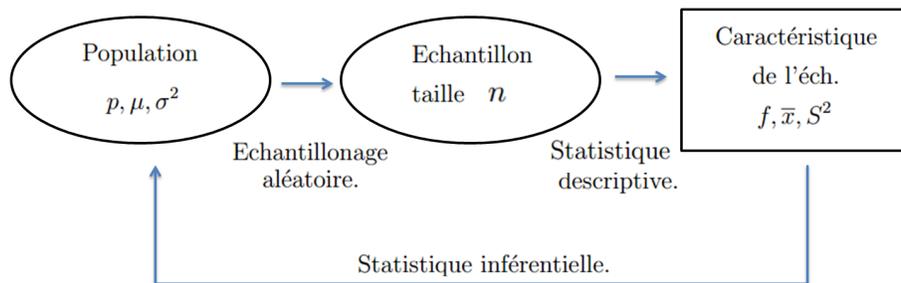
2. A l'instant 0 on met en fonctionnement k appareils. Dès qu'un appareil tombe en panne, on le remplace. On observe toujours le nombre de défaillance dans l'intervalle $[0, a]$. Donner le modèle statistique associé.

Chapitre 3

L'estimation

Introduction

L'estimation consiste à donner des valeurs approchées aux paramètres d'une population (p, μ, σ^2) ou (proportion, moyenne, variance) à partir des données de l'échantillon (f, \bar{x}, S^2) . On supposera vérifiée l'hypothèse d'échantillonnage aléatoire simple. La statistique inférentielle peut se résumer par le schéma suivant :



3.1 Généralités

3.1.1 Estimateur

Soient (X_1, X_2, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X (discrète ou continue) et θ un paramètre associé à la loi probabilité \mathbb{P}_θ de θ . Un estimateur de θ est une variable aléatoire T fonction des X_i

$$T = f(X_1, X_2, \dots, X_n).$$

Si on considère n observations x_1, x_2, \dots, x_n , l'estimateur T fournira une estimation de θ notée :

$$\hat{\theta} = f(x_1, x_2, \dots, x_n).$$

3.1.2 Exemples élémentaires

\bar{x} , S^2 sont des estimations de μ et de σ^2 (resp.).

Les variables aléatoires \bar{X} et S^2 , sont les estimateurs de μ et σ^2 (resp.).

Remarque 3.1. Le même paramètre peut-être estimé à l'aide d'estimateurs différents.

Exemple 3.1. Le paramètre λ d'une loi de Poisson peut-être estimé par \bar{X} et S^2 .

3.1.3 Qualités d'un estimateur

Estimateur convergent

Soit θ le paramètre à estimer et $T = f(X_1, X_2, \dots, X_n)$ un estimateur.

La première qualité d'un estimateur est d'être convergent (consistant : $\lim_{n \rightarrow \infty} T = \theta$). Deux estimateurs convergents ne convergent cependant pas nécessairement à la même vitesse, ceci est lié à la notion de précision d'un estimateur. On mesure généralement la précision d'un estimateur par l'erreur quadratique moyenne.

Perte quadratique : C'est l'écart au carré entre le paramètre et son estimateur :

$$l(\theta, T) = (\theta - T)^2.$$

Risque d'un estimateur : C'est la moyenne des pertes :

$$R(T, \theta) = \mathbb{E}[l(\theta, T)].$$

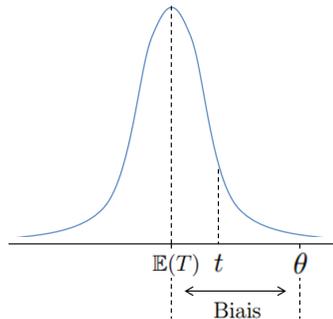
$R(T, \theta) = \mathbb{E}[(\theta - T)^2]$ est le risque quadratique moyen.

Un estimateur T est dit convergent si $\lim_{n \rightarrow \infty} R(T, \theta) = 0$.

Estimateur sans biais

$$T - \theta = T - \mathbb{E}(T) + \mathbb{E}(T) - \theta$$

La quantité $\mathbb{E}(T) - \theta$ est appelée biais de l'estimateur. T est dit sans biais si $\mathbb{E}(T) = \theta$.



Exemple 3.2. \bar{X} est sans biais pour μ , mais S^2 est biaisé pour σ^2 . On utilise souvent $S'^2 = \frac{n}{n-1}S^2$ pour estimer σ^2 .

Remarque 3.2.

$$\begin{aligned} \mathbb{E}((t - \theta)^2) &= \mathbb{E}[(T - \mathbb{E}(T) + \mathbb{E}(T) - \theta)^2] \\ &= \mathbb{E}[(T - \mathbb{E}(T))^2] + 2\mathbb{E}[(t - \mathbb{E}(T))(\mathbb{E}(T) - \theta)] + \mathbb{E}[(\mathbb{E}(T) - \theta)^2] \\ R(T, \theta) &= \mathbb{E}((t - \theta)^2) = \text{Var}(T) + (\mathbb{E}(T) - \theta)^2. \end{aligned}$$

Remarque 3.3. Si T est sans biais alors $R(T, \theta) = \text{Var}(T)$ et T convergent si $\lim_{n \rightarrow \infty} \text{Var}(T) = 0$. Entre deux estimateurs sans biais, le plus précis est donc celui de variance minimale.

Meilleur estimateur

Soient T_1, T_2 deux estimateur de θ . On dit que T_1 est meilleur que T_2 si :

$$R(T_1, \theta) < R(T_2, \theta), \quad \forall \theta.$$

Estimateur asymptotiquement sans biais

T est dit asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} \mathbb{E}(T) = \theta.$$

Pour un échantillon (X_1, X_2, \dots, X_n) issu de X de loi $f(x, \theta)$, on se contentera de rechercher un estimateur sans biais de variance minimale, ce problème est lié à l'existence de statistiques exhaustives.

3.2 Statistique suffisante (exhaustive)

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon nous apporte certaine information sur ce paramètre. Lorsque l'on résume cet échantillon par une statistique, il s'agit de ne pas perdre cette information. Une statistique qui conserve l'information sera dite suffisante (exhaustive).

3.2.1 Définition d'une statistique exhaustive

Soit (X_1, X_2, \dots, X_n) un n-échantillon et $L(\underline{x}, \theta)$ sa vraisemblance. Soit T une statistique fonction de X_1, X_2, \dots, X_n de loi $g(t, \theta)$ dans le cas continue, $\mathbb{P}(T = t)$ dans le cas discret. T est dite exhaustive si $L(\underline{x}, \theta) = g(t, \theta) \cdot h(\underline{x})$ (principe de factorisation, de Neyman-Fisher), en d'autres termes si la loi conditionnelle de l'échantillon est indépendante de θ . $\mathbb{P}_\theta(\underline{X} = \underline{x} | T(\underline{x}) = t) = k(\underline{x}, t)$ indépendante de θ .

Exemple 3.3. Soit X_1, X_2, \dots, X_n issu de $X \rightsquigarrow \mathcal{P}(\theta)$, θ inconnu. La statistique

$T = \sum_{i=1}^n X_i$ est-elle exhaustive pour θ ?

On a :

$$f(x, \theta) = e^{-\theta} \frac{\theta^x}{x!},$$

la loi de Poisson est stable donc $T = \sum_{i=1}^n X_i \rightsquigarrow \mathcal{P}(n\theta)$.

$$g(t, \theta) = e^{-n\theta} \frac{(n\theta)^t}{t!}, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{N}^n, \quad \forall t \in \mathbb{N}.$$

Notons P_θ la quantité $\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T = t)$.

$$\begin{aligned}
 P_\theta &= \frac{\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, T = t)}{\mathbb{P}_\theta(T = t)} \\
 &= \frac{\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, X_n = t - \sum_{i=1}^{n-1} x_i)}{\mathbb{P}_\theta(T = t)} \\
 &= \frac{\mathbb{P}_\theta(X_1 = x_1) \dots \mathbb{P}_\theta(X_n = x_n)}{\mathbb{P}_\theta(T = t)} \\
 &= \frac{\prod_{i=1}^n f(x_i, \theta)}{f(t, \theta)} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \cdot \frac{t!}{e^{-n\theta} (n\theta)^t} \\
 &= \frac{t!}{\prod_{i=1}^n x_i! n^t} = \frac{(\sum_{i=1}^n x_i)!}{\prod_{i=1}^n x_i!} \left(\frac{1}{n}\right)^{\sum_{i=1}^n x_i} \quad \text{indépendante de } \theta.
 \end{aligned}$$

Donc T est exhaustive pour θ .

Ou bien :

$$\begin{aligned}
 L(\underline{x}, \theta) &= \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \left(\frac{\theta^{x_i} e^{-\theta}}{x_i!} \right) = \frac{\theta^t e^{-n\theta}}{\prod_{i=1}^n x_i!} \\
 &= \frac{(n\theta)^t e^{-n\theta}}{t!} \cdot \frac{t!}{n^t (\prod_{i=1}^n x_i!)} \\
 &= g(t, \theta) \cdot h(\underline{x}).
 \end{aligned}$$

D'après le principe de factorisation, T est exhaustive pour θ .

3.2.2 Lois permettant une statistique exhaustive

Théorème 3.1. (Théorème de Darrois) [9]

Soit X une variable aléatoire dont le domaine de définition ne dépend pas de θ . Une condition nécessaire et suffisante pour que l'échantillon (X_1, X_2, \dots, X_n) admette une statistique exhaustive et que la forme de la densité soit :

$$f(x, \theta) = \exp[a(x)\alpha(\theta) + b(x) + \beta(\theta)] \quad (\text{famille exponentielle}).$$

Si la densité est de cette forme, alors $T = \sum_{i=1}^n a(X_i)$ est une statistique exhaustive.

Remarque 3.4. Ce théorème est un outil très puissant pour la recherche d'une statistique exhaustive.

Exemple 3.4. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \gamma(\theta)$.

$$f(x, \theta) = \frac{1}{\Gamma(\theta)} e^{-x} x^{\theta-1}, \quad x > 0.$$

$$\ln f(x, \theta) = -x + (\theta - 1) \ln x - \ln \Gamma(\theta).$$

$$T = \sum_{i=1}^n \ln(X_i) = \ln\left(\prod_{i=1}^n X_i\right) \text{ est exhaustive pour } \theta.$$

Remarque 3.5. Lorsque le domaine de X dépend de θ , le théorème de Darmais ne s'applique pas, ce qui n'empêche pas de trouver une statistique exhaustive.

Exemple 3.5. $X \rightsquigarrow \mathcal{U}_{[0, \theta]} \Rightarrow f(x, \theta) = \frac{1}{\theta}, \quad x \in [0, \theta].$ $T = \sup X_i$ est exhaustive pour θ .

$$\begin{aligned} L(\underline{x}, \theta) &= \frac{1}{\theta^n} \mathbb{I}(x_i) \quad \forall x_i \in [0, \theta] \\ &= \frac{1}{\theta^n} \mathbb{I}_{[\sup x_i, +\infty[}(\theta). \end{aligned}$$

$$\text{Comme } G(t) = \mathbb{P}(\sup X_i \leq t) = (F(t))^n = \left(\frac{t}{\theta}\right)^n$$

$$\begin{aligned} g(t, \theta) &= n \left(\frac{t}{\theta}\right)^{n-1} \frac{1}{\theta} = n \frac{t^{n-1}}{\theta^n} \\ L(\underline{x}, \theta) &= \frac{1}{\theta^n} \Rightarrow \frac{L}{g} = \frac{1}{nt^{n-1}} \text{ indépendant de } \theta. \end{aligned}$$

donc $T = \sup X_i$ est exhaustive pour θ .

3.2.3 L'information de Fisher

Définition 3.1. [9] On appelle quantité d'information de Fisher $I_n(\theta)$ apportée par un échantillon sur le paramètre θ , la quantité suivante positive ou nulle (si elle existe) :

$$I_n(\theta) = \mathbb{E} \left[\left(\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta} \right)^2 \right]$$

$$\text{avec } L(\underline{X}, \theta) = \prod_{i=1}^n f(X_i, \theta).$$

Théorème 3.2. [9] Si le domaine de définition de X ne dépend pas de θ alors :

$$I_n(\theta) = -\mathbb{E} \left[\frac{\partial^2 \ln L(\underline{X}, \theta)}{\partial \theta^2} \right].$$

Démonstration. On a $L(\underline{x}, \theta)$ est une densité de probabilité :

$$\int_{\mathbb{R}^n} L(\underline{x}, \theta) d\underline{x} = 1.$$

$$\begin{aligned} \text{on a } \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} &= \frac{\partial L(\underline{x}, \theta)}{\partial \theta L(\underline{x}, \theta)} \\ \text{donc } \frac{\partial L(\underline{x}, \theta)}{\partial \theta} &= L(\underline{x}, \theta) \cdot \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} \\ \int_{\mathbb{R}^n} \frac{\partial L(\underline{x}, \theta)}{\partial \theta} d\underline{x} = 0 &\Rightarrow \int_{\mathbb{R}^n} L(\underline{x}, \theta) \cdot \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} d\underline{x} = 0. \end{aligned}$$

Ce qui prouve que la variable aléatoire $\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta}$ est centrée i.e :

$$\mathbb{E} \left[\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta} \right] = 0.$$

Donc

$$I_n(\theta) = \mathbb{E} \left[\left(\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta} \right)^2 \right] = \text{Var} \left(\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta} \right).$$

En dérivant la deuxième fois :

$$\begin{aligned} \int_{\mathbb{R}^n} L(\underline{x}, \theta) \cdot \frac{\partial^2 \ln L(\underline{x}, \theta)}{\partial \theta^2} d\underline{x} + \int_{\mathbb{R}^n} \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} \frac{\partial L(\underline{x}, \theta)}{\partial \theta} d\underline{x} &= 0 \\ \int_{\mathbb{R}^n} L(\underline{x}, \theta) \cdot \frac{\partial^2 \ln L(\underline{x}, \theta)}{\partial \theta^2} d\underline{x} + \int_{\mathbb{R}^n} \left(\frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} \right)^2 L(\underline{x}, \theta) d\underline{x} &= 0 \\ I_n(\theta) &= - \int_{\mathbb{R}} \frac{\partial^2 \ln L(\underline{x}, \theta)}{\partial \theta^2} L(\underline{x}, \theta) d\theta \\ I_n(\theta) &= -\mathbb{E} \left[\frac{\partial^2 \ln L(\underline{X}, \theta)}{\partial \theta^2} \right]. \end{aligned}$$

□

Propriété d'additivité de l'information de Fisher :

Soit Y une variable aléatoire indépendante de X dont la loi dépend du même paramètre θ .

Soit $f(x, \theta)$, $g(y, \theta)$, $h(x, y, \theta)$ les densités de X , Y et du couple (X, Y) (resp.), d'information de Fisher $I_X(\theta)$, $I_Y(\theta)$, $I(\theta)$ (resp.)

Si les domaines de définition de X , Y ne dépendent pas de θ alors :

$$I(\theta) = I_X(\theta) + I_Y(\theta).$$

Démonstration. Puisque X , Y sont indépendantes :

$$\begin{aligned} h(x, y, \theta) &= f(x, \theta) \cdot g(y, \theta) \\ \ln h(x, y, \theta) &= \ln f(x, \theta) + \ln g(y, \theta) \\ \frac{\partial^2 \ln h(x, y, \theta)}{\partial \theta^2} &= \frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} + \frac{\partial^2 \ln g(y, \theta)}{\partial \theta^2} \\ \mathbb{E} \left(\frac{\partial^2 \ln h(X, Y, \theta)}{\partial \theta^2} \right) &= \mathbb{E} \left(\frac{\partial^2 \ln f(X, \theta)}{\partial \theta^2} \right) + \mathbb{E} \left(\frac{\partial^2 \ln g(Y, \theta)}{\partial \theta^2} \right) \\ -I(\theta) &= -I_X(\theta) - I_Y(\theta) \Rightarrow I(\theta) = I_X(\theta) + I_Y(\theta) \end{aligned}$$

□

Conséquence : Si le domaine de définition ne dépend pas de θ alors : $I_n(\theta) = nI_1(\theta)$

avec

$$I_1(\theta) = -\mathbb{E} \left(\frac{\partial^2 \ln f(X, \theta)}{\partial \theta^2} \right).$$

i.e. : que chaque observation a la même importance. Ce qui n'est pas le cas pour la loi uniforme sur $[0, \theta]$ où la plus grande observation est la plus intéressante.

Exemple 3.6. Soit $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Calculer l'information de Fisher apportée par un

n -échantillon issu de X sur le paramètre μ .

$$\begin{aligned}
 I_n(\mu) &= nI_1(\mu) \\
 I_1(\mu) &= -\mathbb{E} \left[\frac{\partial^2 \ln f(X, \mu)}{\partial \mu^2} \right] \\
 f(x, \mu) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\
 \ln f(x, \mu) &= -\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}(x-\mu)^2 \\
 \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2}(x-\mu) \\
 \frac{\partial^2 \ln L}{\partial \mu^2} &= \frac{-1}{\sigma^2} \Rightarrow I_n(0) = \frac{n}{\sigma^2}.
 \end{aligned}$$

3.2.4 Information de Fisher et suffisance

On montre que l'information portée par une statistique est inférieure ou égale à celle apportée par un échantillon. En effet : soit T la statistique de densité $g(t, \theta)$ que l'on substitue à l'échantillon.

On a $L(\underline{x}, \theta) = g(t, \theta) \cdot h(\underline{x}, \theta|t)$ où $h(\underline{x}, \theta|t)$ est la densité conditionnelle de l'échantillon.

$$\begin{aligned}
 \ln L(\underline{x}, \theta) &= \ln g(t, \theta) + \ln h(\underline{x}, \theta|t) \\
 \frac{\partial^2 \ln L(\underline{x}, \theta)}{\partial \theta^2} &= \frac{\partial^2 \ln g(t, \theta)}{\partial \theta^2} + \frac{\partial^2 \ln h(\underline{x}, \theta|t)}{\partial \theta^2}.
 \end{aligned}$$

En prenant l'espérance mathématique :

$$\begin{aligned}
 I_n(\theta) &= I_T(\theta) - \mathbb{E} \left[\frac{\partial^2 \ln h}{\partial \theta^2} \right] = I_T(\theta) + I_{n/T}(\theta). \\
 I_{n/T}(\theta) &= -\mathbb{E} \left[\frac{\partial^2 \ln h}{\partial \theta^2} \right] \geq 0
 \end{aligned}$$

D'où :

$$I_n(\theta) \geq I_T(\theta).$$

Si T est suffisante $I_n(\theta) = I_T(\theta)$.

La réciproque est vraie si le domaine de X est indépendant de θ .

3.2.5 Borne de Freshet-Damois-Cramer-Rao (FDCR)

Le résultat suivant nous indique que la variance d'un estimateur ne peut être inférieure à une certaine borne, qui dépend de la quantité d'information de Fisher apportée par l'échantillon sur le paramètre θ

Théorème 3.3. (*Inégalité de FDCR*)

Si le domaine de définition de X ne dépend pas de θ alors pour tout estimateur sans biais :

$$\text{Var}(T) \geq \frac{1}{I_n(\theta)}$$

Si T est un estimateur sans biais de $h(\theta)$:

$$\text{Var}(T) \geq \frac{[h'(\theta)]^2}{I_n(\theta)}.$$

Démonstration.

$$\begin{aligned} \text{cov}\left(T, \frac{\partial \ln L}{\partial \theta}\right) &= \mathbb{E}\left(T \frac{\partial \ln L}{\partial \theta}\right) \text{ puisque } \frac{\partial \ln L}{\partial \theta} \text{ est centrée} \\ &= \int_{\mathbb{R}^n} t \frac{\partial \ln L}{\partial \theta} L(\underline{x}; \theta) d\underline{x} \\ &= \int_{\mathbb{R}^n} t \frac{\partial L(\underline{x}; \theta)}{\partial \theta} d\underline{x} \\ &= \frac{d}{d\theta} \int_{\mathbb{R}^n} t L(\underline{x}, \theta) d\underline{x} \\ &= \frac{d}{d\theta} \mathbb{E}(T) \\ &= h'(\theta) \end{aligned}$$

D'autre part l'inégalité de Schwartz donne :

$$\left[\text{cov}\left(T, \frac{\partial \ln L}{\partial \theta}\right)\right]^2 \leq \text{Var}(T) \text{Var}\left(\frac{\partial \ln L}{\partial \theta}\right),$$

c'est à dire :

$$[h'(\theta)]^2 \leq \text{Var}(T) I_n(\theta).$$

□

3.2.6 Estimateur efficace

Définition 3.2. Un estimateur T est dit efficace si sa variance est égale à la borne de FDCR.

Propriété : Un estimateur sans biais efficace est convergent.

En effet : T efficace $\Rightarrow \text{Var}(T) = \frac{[h'(\theta)]^2}{I_n(\theta)}$ or $I_n(\theta) = nI_1(\theta)$ et T est sans biais

$$\begin{aligned} \lim_{n \rightarrow \infty} R(T, \theta) &= \lim_{n \rightarrow \infty} \text{Var}(T) \\ &= \lim_{n \rightarrow \infty} \frac{[h'(\theta)]^2}{I_n(\theta)} \\ &= \lim_{n \rightarrow \infty} \frac{[h'(\theta)]^2}{nI_1(\theta)} = 0. \end{aligned}$$

Donc T convergent.

Remarque 3.6. Un estimateur efficace T est un estimateur sans biais de variance minimale.

Exemple 3.7. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ estimateur de μ .

$\mathbb{E}(\bar{X}) = \mu$ et $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ et on a $I_n(\mu) = \frac{n}{\sigma^2}$.

La borne de FDCR est $\frac{1}{I_n(\mu)} = \frac{\sigma^2}{n} = \text{Var}(\bar{X})$.

Donc \bar{X} est efficace pour μ et c'est un estimateur sans biais de variance minimale de μ .

3.2.7 Estimateur sans biais de variance minimale (MVUE)

Théorème 3.4. S'il existe un estimateur de θ sans biais, de variance minimale, il est unique presque sûrement.

Démonstration. (indication : démonstration par l'absurde)

Supposer qu'il existe T_1, T_2 deux MVUE de θ et soit $T_3 = \frac{T_1 + T_2}{2}$. □

3.2.8 Généralisation (Cas multidimensionnel)

Statistique exhaustive

Soit le modèle statistique paramétrique : $(\mathcal{X}, (\mathbb{P}_\theta)_{\theta \in \Theta})$, avec $\theta \in \Theta \subset \mathbb{R}^p$.

Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow f(x, \theta)$, on suppose que $f(x, \theta)$ appar-

tienne à la famille exponentielle :

$$f(x, \theta) = c(\theta) \cdot h(x) \cdot \exp \sum_{j=1}^p \alpha_j(\theta) T_j(x)$$

Si $X(\Omega)$ ne dépend pas de θ .

La statistique $(\sum_{i=1}^n T_1(X_i), \sum_{i=1}^n T_2(X_i), \dots, \sum_{i=1}^n T_p(X_i))$ est exhaustive pour θ .

Information de Fisher

Soit $X \rightsquigarrow f(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}^p$. On note $\theta = (\theta_1, \theta_2, \dots, \theta_p)$.

On fait les hypothèses de régularité suivante :

$H_1 \forall x, \forall \theta, f(x, \theta) > 0$.

$H_2 \text{grad}_\theta f$ existe $\forall \theta$, i.e : on peut dériver f par rapport à θ_i , $\forall i = \overline{1, p}$, \mathbb{P}_θ p.s.

H_3 On peut dériver au moins deux fois $f(x, \theta)$ par rapport à θ_i , $\forall i = \overline{1, p}$, dériver $\int f(x, \theta) dx$ et permuter entre dérivation et intégration.

$\text{grad}_\theta f = \left(\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \dots, \frac{\partial f}{\partial \theta_p} \right)$ vecteur ligne $(1, p)$. On appelle score $S(x, \theta)$ le vecteur $(1, p)$ défini par $\text{grad}_\theta \text{Log} f$.

Définition 3.3. On appelle information en θ la matrice (p, p) de variance-covariance de $\text{grad}_\theta \text{Log} f(x, \theta)$.

$$I(\theta) = \mathbb{E}(S'S).$$

Comme dans le cas réel :

- $\mathbb{E}(\text{grad}_\theta \text{Log} f) = 0$
- $\mathbb{E} \left[\frac{1}{f} \frac{\partial^2 f(X, \theta)}{\partial \theta_i \partial \theta_j} \right] = 0, \quad \forall i, j = \overline{1, p}$.

Sous les hypothèses H_1, H_2, H_3 on obtient, $I(\theta)$: la matrice dont le terme général $I_{ij}(\theta)$,

$$I_{ij}(\theta) = -\mathbb{E} \left(\frac{\partial^2 f(X, \theta)}{\partial \theta_i \partial \theta_j} \right)$$

$I(\theta)$ est une matrice définie positive.

Exemple 3.8. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $\mathcal{N}(\mu, \sigma^2)$, avec μ, σ^2 inconnus.

$$L(\underline{x}, \mu, \sigma^2) = \frac{1}{(2\pi\sigma)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}, \quad x_i \in \mathbb{R}.$$

Pour μ :

$$I_1(\mu) = \frac{1}{\sigma^2} \Rightarrow I_n(\mu) = \frac{n}{\sigma^2}.$$

Pour σ^2 :

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

$$\begin{aligned} \ln f(x, \sigma^2) &= -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (x - \mu)^2 \\ \frac{\partial \ln f(x, \sigma^2)}{\partial \sigma^2} &= \frac{-1}{2\sigma^2} + \frac{1}{2\sigma^4} (x - \mu)^2 \\ \frac{\partial^2 \ln f(x, \sigma^2)}{(\partial \sigma^2)^2} &= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6} (x - \mu)^2 \\ -\mathbb{E} \left(\frac{\partial^2 \ln f(X, \sigma^2)}{(\partial \sigma^2)^2} \right) &= \frac{-1}{2\sigma^4} + \frac{1}{\sigma^4} \mathbb{E} \left(\frac{(x - \mu)^2}{\sigma^2} \right) = \frac{-1}{2\sigma^4} + \frac{1}{\sigma^4} = \frac{1}{2\sigma^4}. \end{aligned}$$

D'où $I_1(\sigma^2) = \frac{1}{2\sigma^4} \Rightarrow I_n(\sigma^2) = \frac{n}{2\sigma^4}$. Pour $\theta = (\mu, \sigma^2)$, $I_1(\theta)$ est une matrice carrée d'ordre 2 :

$$I_{11}(\theta) = I_1(\mu) = \frac{1}{\sigma^2} \text{ et } I_{22}(\theta) = I_1(\sigma^2) = \frac{1}{2\sigma^4}$$

$$I_{12} = -\mathbb{E} \left[\frac{\partial^2 \ln f(X, \mu, \sigma^2)}{\partial \mu \partial \sigma^2} \right]$$

$$\begin{aligned} \frac{\partial \ln f(x, \mu, \sigma^2)}{\partial \mu} &= \frac{1}{\sigma^2} (x - \mu) \\ \frac{\partial^2 \ln f(x, \mu, \sigma^2)}{\partial \mu \partial \sigma^2} &= \frac{-1}{\sigma^4} (x - \mu) \Rightarrow I_{12} = 0 \\ I_1(\theta) &= \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} \Rightarrow I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}. \end{aligned}$$

3.2.9 Statistique complète

Définition 3.4. On dit qu'une statistique T est complète pour une famille de loi de probabilité $(\mathbb{P}_\theta)_{\theta \in \Theta}$ si pour toute fonction borellienne h on ait :

$$\forall \theta \in \Theta : \mathbb{E}(h(T)) = 0 \Rightarrow h = 0, \quad \mathbb{P}_\theta \text{ p.s}$$

En particulier la statistique exhaustive de famille exponentielle est complète.

Exemple 3.9. Pour $X \rightsquigarrow \mathcal{P}(\lambda)$, λ inconnu.

$T = \sum_{i=1}^n X_i$ est complète.

$$\mathbb{E}(h(T)) = \sum_{t=0}^{\infty} h(t) e^{-n\lambda} \frac{(n\lambda)^t}{t!} \quad \text{car } T \rightsquigarrow \mathcal{P}(n\lambda).$$

$$\mathbb{E}(h(T)) = 0, \forall \lambda \Rightarrow h(t) = 0, \quad \forall t \in \mathbb{N}.$$

Exercice : Soit X une variable aléatoire à valeur dans $\{-1, 0, 1, 2, \dots\} = X(\Omega)$

$$\mathbb{P}_\theta(X = -1) = \theta$$

$$\mathbb{P}_\theta(X = x) = (1 - \theta)^2 \theta^x, \quad x \in \mathbb{N}.$$

Montrer que la famille $(\mathbb{P}_\theta, \theta \in]0, 1[)$ n'est pas complète.

La famille $(\mathbb{P}_\theta)_{\theta \in \Theta}$ est complète si :

$$\forall \theta \in \Theta : \mathbb{E}(h(X)) = 0 \Rightarrow h \equiv 0 \quad \mathbb{P}_\theta \text{ p.s.}$$

Théorème 3.5. Si T^* est un estimateur sans biais de θ dépendant d'une statistique exhaustive complète U alors T^* est l'unique estimateur MVUE de θ . En particulier si l'on dispose déjà d'un estimateur T sans biais de $\theta : T^* = \mathbb{E}[T/U]$.

3.3 Méthodes de calcul d'un estimateur

3.3.1 Méthode des moments

Soit (X_1, X_2, \dots, X_n) un n-échantillon issu d'une variable aléatoire X de densité $f(x, \theta_1, \theta_2, \dots, \theta_k)$ où : $\theta_1, \theta_2, \dots, \theta_k$ sont des paramètres inconnus.

La méthode des moments consiste à estimer les paramètres $\theta_1, \theta_2, \dots, \theta_k$, en égalisant les moments empiriques calculés à partir de l'échantillon avec les moments théoriques de même ordre.

Soit $\mu'_r = \mathbb{E}(X^r)$, $r = 1, 2, \dots, k$ moments d'ordre r de la population (théorique) et on note $M_r = \frac{1}{n} \sum_{i=1}^n X_i^r$ moment empirique d'ordre r de l'échantillon.

La solution du système $M_r = \mu'_r$, $r = \overline{1, k}$ nous donne les estimateurs de $\theta_1, \theta_2, \dots, \theta_k$

$$\begin{cases} M_1 = \mu'_1, \\ M_2 = \mu'_2, \\ \vdots, \\ M_k = \mu'_k, \end{cases} \quad k \text{ équations à } k \text{ inconnus ;}$$

Remarque 3.7. Dans la plupart des cas, les estimateurs obtenus par la méthode des moments sont consistants, convergents, asymptotiquement normaux mais en général ne sont pas efficaces.

Exemple 3.10. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, où μ, σ^2 sont inconnus. Estimer μ et σ^2 par la méthode des moments.

$$\begin{aligned} \mu'_1 &= \mathbb{E}(X) = \mu \\ \mu'_2 &= \mathbb{E}(X^2) = \text{Var}(X) + (\mathbb{E}(X))^2 = \sigma^2 + \mu^2 \\ M_1 &= \frac{1}{n} \sum_{i=1}^n X_i \\ M_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2. \end{aligned}$$

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n X_i, \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2, \end{cases} \quad L'estimateur de μ est $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$.$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 \end{aligned}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

3.3.2 Méthode du maximum de vraisemblance

Il s'agit de trouver un estimateur de θ , qui maximise la fonction de vraisemblance de l'échantillon. La valeur $\hat{\theta}$ qui maximise $L(\underline{x}, \theta)$ serait un bon estimateur car elle donne la

plus grande probabilité pour cet échantillon.

Cette méthode consiste à résoudre :

$$\begin{cases} \frac{\partial L(\underline{x}, \theta)}{\partial \theta} = 0, & \text{pour trouver } \hat{\theta}; \\ \frac{\partial^2 L(\underline{x}, \theta)}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} < 0, & \text{pour assurer l'existence du } \max_{\theta \in \Theta} L(\underline{x}, \theta). \end{cases}$$

Remarque 3.8. Maximiser $L(\underline{x}, \theta)$ revient à maximiser $\ln L(\underline{x}, \theta)$. Il est plus commode de maximiser $\ln L(\underline{x}, \theta)$.

Exemple 3.11. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, trouver l'estimateur EMV de μ et σ^2 .

La densité de X est :

$$\begin{aligned} f(x, \mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x > 0. \\ L(\underline{x}, \mu, \sigma^2) &= \prod_{i=1}^n f(x_i, \mu, \sigma^2) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2}. \\ \ln L(\underline{x}, \mu, \sigma^2) &= -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \\ \frac{\partial \ln L(\underline{x}, \mu, \sigma^2)}{\partial \mu} &= \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma^2} \right) = 0 \\ &\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ &\Rightarrow \sum_{i=1}^n (x_i - \mu) = 0 \Rightarrow n\mu = \sum_{i=1}^n x_i. \\ &\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i. \\ \frac{\partial^2 \ln L(\underline{x}, \mu, \sigma^2)}{\partial \mu^2} &= \frac{-1}{\sigma^2} < 0 \end{aligned}$$

d'où : $T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ est l'EMV de μ .

$$\begin{aligned}
L(\underline{x}, \mu, \sigma^2) &= \frac{1}{(\sigma^2 2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}. \\
\ln L(\underline{x}, \mu, \sigma^2) &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} + \frac{2 \sum_{i=1}^n (x_i - \mu)^2}{4(\sigma^2)^2} \\
&= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2} = 0. \\
\Rightarrow \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2} = \frac{n}{2\sigma^2} &\Rightarrow \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} = n \\
&\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.
\end{aligned}$$

On a :

$$\begin{aligned}
\frac{\partial^2 \ln L(\underline{x}, \mu, \sigma^2)}{\partial (\sigma^2)^2} &= \frac{2n}{4(\sigma^2)^2} \frac{-4\sigma^2 \sum_{i=1}^n (x_i - \mu)^2}{4(\sigma^2)^4} \\
&= \frac{n}{2(\sigma^2)^2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\sigma^2)^3} \Big|_{\sigma^2 = \hat{\sigma}^2} \\
&= \frac{1}{(\hat{\sigma}^2)^2} \left[\frac{n}{2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \right]_{\sigma^2 = \hat{\sigma}^2} \\
&= \frac{1}{(\hat{\sigma}^2)^2} \left[\frac{n}{2} - n \right] = \frac{-n}{2\hat{\sigma}^2} < 0.
\end{aligned}$$

$$D'où : \begin{cases} S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2, & \text{si } \mu \text{ est connue;} \\ S_2^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, & \text{si } \mu \text{ est inconnu.} \end{cases}$$

Propriété : S'il existe une statistique exhaustive U alors l'EMV en dépend.

En effet : $L(\underline{x}, \theta) = g(u, \theta) \cdot h(\underline{x})$ et résoudre $\frac{\partial \ln L}{\partial \theta} = 0$ revient à résoudre $\frac{\partial \ln g(u, \theta)}{\partial \theta} = 0$, donc $\hat{\theta} = f(u)$.

Caractéristiques de l'EMV

1. $L(\underline{x}, \theta)$ n'a aucune raison d'être différentiable en θ .

Exemple 3.12. Soit (X_1, \dots, X_n) issu de $X \rightsquigarrow \mathcal{U}_{[0, \theta]}$.

On a :

$$\begin{aligned} L(x, \theta) &= \frac{1}{\theta^n} \mathbb{I}_{[0, \theta]}(\sup x_i) \\ &= \frac{1}{\theta^n} \mathbb{I}_{[\sup x_i, +\infty]}(\theta). \end{aligned}$$

L est décroissante sur $[\sup x_i, +\infty[$. L est maximum pour θ minimum.

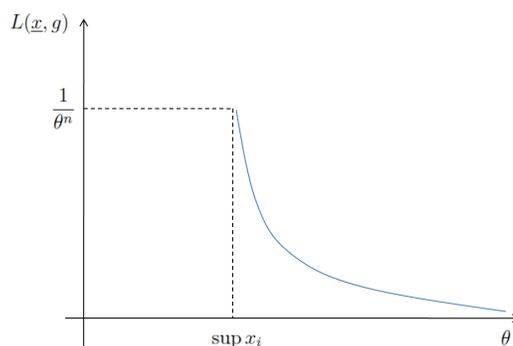


FIG. 3.1 – Graphe de L en fonction de θ

i.e. : $\hat{\theta} = \sup_{1 \leq i \leq n} x_i$ et donc $Y_n = \sup_{1 \leq i \leq n} X_i$ est l'EMV de θ .

2. L'EMV n'est pas forcément sans biais, donc pas forcément efficace.

Exemple 3.13. Reprenons l'exemple précédent :

$Y_n = \sup X_i$ où : $F_{Y_n}(y) = \mathbb{P}(Y_n < y) = \frac{y^n}{\theta^n}$, $y \in [0, \theta]$.

$$f(y) = \frac{ny^{n-1}}{\theta^n}, \quad y \in [0, \theta].$$

$$\mathbb{E}(Y_n) = \int_0^\theta \frac{ny^n}{\theta^n} dy = \frac{n}{n+1} \theta \neq \theta.$$

Y_n est biaisé.

On en déduit un estimateur sans biais :

$$\hat{Y}_n = \frac{n+1}{n} Y_n = \frac{n+1}{n} \sup X_i.$$

$$\begin{aligned}
 \text{Var}(\hat{Y}_n) &= \text{Var}\left(\frac{n+1}{n} \sup X_i\right) = \frac{(n+1)^2}{n^2} \text{Var}(\sup X_i). \\
 \text{Var}(\sup X_i) &= \text{Var}(Y_n) = \mathbb{E}(Y_n^2) - \mathbb{E}^2(Y_n). \\
 \mathbb{E}(Y_n^2) &= \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{n}{n+1} \theta^2. \\
 \text{Var}(Y_n) &= \frac{n}{n+2} \theta^2 - \frac{(n+1)^2}{n^2} \theta^2 \\
 &= \frac{n}{(n+2)(n+1)^2} \theta^2.
 \end{aligned}$$

$$\text{Var}\left(\frac{n+1}{n} \sup X_i\right) = \frac{\theta^2}{n(n+2)} \xrightarrow{n \rightarrow \infty} 0.$$

3. L'EMV n'est pas unique.

Exemple 3.14. $X \rightsquigarrow \mathcal{U}_{[\theta, \theta+1]}$, $\theta > 0$. On considère un n -échantillon issu de X :

$$\begin{aligned}
 L(\underline{x}, \theta) &= \mathbf{1}_{[\theta \leq \inf x_i \leq \sup x_i \leq \theta+1]} \\
 &= \mathbb{I}_{[\theta \leq \inf x_i]} \mathbb{I}_{[\theta \geq \sup x_i - 1]}.
 \end{aligned}$$

θ			
$\mathbf{1}_{[\theta \leq \inf x_i]}$	1	1	0
$\mathbf{1}_{[\theta \geq \sup x_i - 1]}$	0	1	1
L	0	1	0

Tout estimateur $\hat{\theta}$ compris entre $\hat{\theta}_1 = \sup X_i - 1$ et $\hat{\theta}_2 = \inf X_i$ est un MV.

$\hat{\theta}$ est unique si $\sup X_i = \inf X_i + 1$.

Propriété : Si $\hat{\theta}$ est l'estimateur du MV de θ alors $f(\hat{\theta})$ est l'EMV de $f(\theta)$.

Exemple 3.15. $X \rightsquigarrow \mathcal{P}(\lambda) : \hat{\lambda} = \bar{X}_n$, l'EMV de $e^{-\lambda}$ est $e^{-\bar{X}_n}$.

Propriété (admise) : Il existe une suite de valeurs $\hat{\theta}_n$ racine de l'équation de vraisemblance qui converge p.s vers θ quand $n \rightarrow \infty$. De plus $\exists N t.q \forall n > N \Rightarrow \hat{\theta}_n$ réalise effectivement un maximum pour L .

Propriété (admise) :

$$\frac{\hat{\theta} - \theta}{\frac{1}{\sqrt{I_n(\theta)}}} \underset{n \rightarrow \infty}{\overset{\mathcal{L}}{\rightarrow}} \mathcal{N}(0, 1).$$

Remarque 3.9. $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = \frac{1}{I_n(\theta)} \Rightarrow \hat{\theta}_n$ est asymptotiquement efficace.

3.4 Estimation par intervalle de confiance

Soit X une variable aléatoire dont la loi dépend d'un paramètre θ inconnu. Soit (X_1, X_2, \dots, X_n) un n-échantillon issu de X et $\alpha \in]0, 1[$.

Définition 3.5. Un intervalle de confiance pour le paramètre θ , de niveau de confiance $1 - \alpha \in]0, 1[$, est l'intervalle qui a la probabilité $1 - \alpha$ de contenir la vraie valeur du paramètre θ .

Remarque 3.10. Un intervalle de confiance indique la précision d'une estimation car pour un risque α donné, l'intervalle est d'autant plus grand que la précision est faible.

3.4.1 Principe de construction

Le principe de la méthode d'estimation par intervalle de confiance est le suivant :
Soit T un estimateur de θ (meilleur estimateur possible) dont on connaît la loi en fonction de θ .

On détermine par la suite un intervalle de probabilité de niveau $1 - \alpha$ pour T i.e :

$$\mathbb{P}(t_1(\theta) \leq T \leq t_2(\theta)) = 1 - \alpha.$$

Il faut ensuite inverser cet intervalle pour T , dont les bornes dépendent de θ pour obtenir un intervalle pour θ ,

$$\text{i.e } \mathbb{P}_\theta(a(T) \leq \theta \leq b(T)) = 1 - \alpha,$$

$$\text{ou bien } \mathbb{P}_\theta(t_2^{-1}(t) \leq \theta \leq t_1^{-1}(t)) = 1 - \alpha.$$

Graphiquement :

Verticalement, on lit l'intervalle de probabilité et horizontalement l'intervalle de confiance (FIG.3.2.)

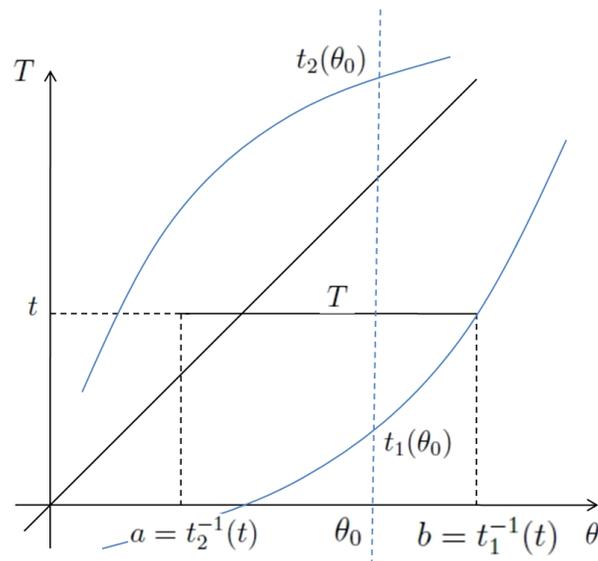


FIG. 3.2 – Intervalle de probabilité et intervalle de confiance

Remarque 3.11. Si $1 - \alpha$ augmente, on augmente la longueur de l'intervalle de probabilité donc les courbes s'écartent.

Si n augmente, comme T est supposé convergent, alors $Var(T)$ diminue donc $[t_1, t_2]$ diminue est les courbes se rapproche de la première bissectrice, donc l'intervalle de confiance diminue.

3.4.2 Intervalles de confiance classiques

Intervalle de l'espérance μ d'une loi $\mathcal{N}(\mu, \sigma^2)$

a) Cas où σ^2 est connue :

On sait que \bar{X} est le meilleur estimateur de μ .

$$\mathbb{E}(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

$$\bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1)$$

$$\mathbb{P}\left(-u_{\alpha/2} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{\alpha/2}\right) = 1 - \alpha$$

$$\mathbb{P}\left(\mu - \frac{\sigma}{\sqrt{n}}u_{\alpha/2} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right) = 1 - \alpha$$

d'où :

$$IC_{(1-\alpha)}(\mu) = \left[\bar{x} - \frac{\sigma}{\sqrt{n}}u, \bar{x} + \frac{\sigma}{\sqrt{n}}u \right]$$

où : $u_{\alpha/2}$ est le fractile d'ordre $(1 - \frac{\alpha}{2})$ de $\mathcal{N}(0, 1)$, i.e : $\mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{\alpha/2}\right) = 1 - \frac{\alpha}{2}$.

D'où :

$$u_{\alpha/2} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Exemple 3.16. Pour $1 - \alpha = 0.95$,

$$\mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{\alpha/2}\right) = 1 - \frac{\alpha}{2} = 0.975.$$

Sur la table de $\mathcal{N}(0, 1)$ on aura $u_{\alpha/2} = 1.96$.

$$IC(\mu) = \left[\bar{x} - \frac{\sigma}{\sqrt{n}}1.96, \bar{x} + \frac{\sigma}{\sqrt{n}}1.96 \right].$$

b) Cas où σ^2 est inconnue :

On a $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ est un estimateur sans biais de σ^2 et :

$$\frac{X_i - \bar{X}}{\sigma} \rightsquigarrow \mathcal{N}(0, 1) \Rightarrow \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

D'après la décomposition de S^2 :

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

en divisant par σ^2 :

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{nS^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2.$$

On a $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \rightsquigarrow \chi_n^2$: comme somme de n carrés de variables aléatoires centrées réduites.

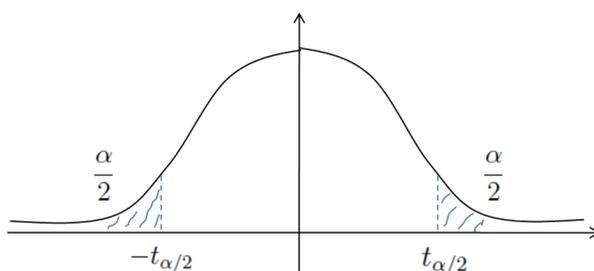
$\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \rightsquigarrow \chi_1^2$, on en déduit que $\frac{nS^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$.

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

Alors la variable aléatoire :

$$T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S'^2}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu}{S'} \sqrt{n} \rightsquigarrow T_{(n-1)} \text{ Student.}$$

L'intervalle de probabilité est :



$$\mathbb{P}\left(-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{S'} \leq t_{\alpha/2}\right) = 1 - \alpha$$

$$\mathbb{P}(|T_{n-1}| \leq t_{\alpha}) = 1 - \alpha$$

ou bien $\mathbb{P}(|T_{n-1}| > t_{\alpha}) = \alpha$ (t_{α} est tiré de la table de Student).

On trouve l'intervalle de confiance pour μ :

$$\bar{x} - t_{\alpha/2} \frac{S'}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{S'}{\sqrt{n}}$$

si $\alpha = 0.05$, $n = 10$ on trouve $t = 2.262$

$$IC(\mu) = \left[\bar{x} - 2.262 \frac{S'}{\sqrt{10}} \leq \mu \leq \bar{x} + 2.262 \frac{S'}{\sqrt{10}} \right].$$

Intervalle de confiance pour σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$

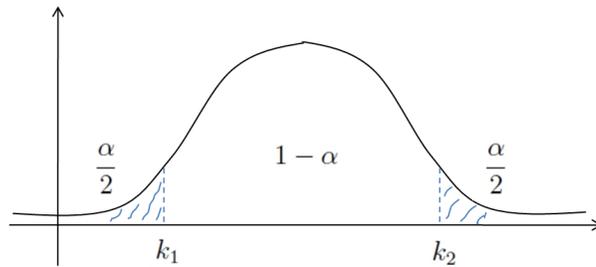
a) μ connu :

On a $T = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ est le meilleur estimateur de σ^2 .

$\frac{nT}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \rightsquigarrow \chi_n^2$ comme somme de n carrés de variable aléatoire $\mathcal{N}(0, 1)$.

$\mathbb{P}(k_1 \leq \frac{nT}{\sigma^2} \leq k_2) = 1 - \alpha$. L'intervalle de confiance pour σ^2 est :

$$\frac{nt}{k_2} \leq \sigma^2 \leq \frac{nt}{k_1},$$



avec $k_1 = \phi_{\chi_n^2}^{-1}(\alpha/2)$, $k_2 = \phi_{\chi_n^2}^{-1}(1 - \alpha/2)$.

b) μ inconnu :

On l'estime par $S'^2 + \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ et on a $\frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$.

$$\mathbb{P} \left(k_1 \leq \frac{(n-1)S'^2}{\sigma^2} \leq k_2 \right) = 1 - \alpha.$$

D'où :

$$IC(\sigma^2) = \left[\frac{(n-1)S'^2}{k_2}; \frac{(n-1)S'^2}{k_1} \right].$$

Exemple : Si $1 - \alpha = 0.95$, on cherche :

$$k_1 \text{ tel que } \mathbb{P}\left(\frac{(n-1)S'^2}{\sigma^2} < k_1\right) = \alpha/2 = 0.025.$$

$$k_2 \text{ tel que } \mathbb{P}\left(\frac{(n-1)S'^2}{\sigma^2} < k_2\right) = 1 - \alpha/2 = 0.975.$$

Estimation et intervalle de confiance d'une proportion

Soit une population comportant deux modalités A et B . Soit p la proportion d'individus possédant la modalité A . $1 - p$ est donc la proportion des individus possédant la modalité B .

On extrait de la population un échantillon de taille n . Soit la variable aléatoire K_n : nombre d'individus dans l'échantillon ayant la modalité A .

Définition 3.6. La variable aléatoire $F = \frac{K_n}{n}$ s'appelle fréquence empirique, sa réalisation f est la proportion d'individus dans l'échantillon ayant la modalité A .

Propriété : $K_n \rightsquigarrow \mathcal{B}(n, p)$, $\mathbb{E}(K_n) = np$, $Var(K_n) = npq$.

$$\mathbb{E}(F) = p \text{ et } Var(F) = \frac{p(1-p)}{n}.$$

Loi de probabilité de F :

Dés que $np > 5$ et $n(1-p) > 5$.

$F \rightsquigarrow \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$ i.e : $\frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0, 1)$ approximativement.

Théorème 3.6. La fréquence empirique $F = \frac{K_n}{n}$ est l'estimateur efficace de p .

F est l'estimateur sans biais et convergent de p , on a : $F \rightsquigarrow \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

$$\frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0, 1).$$

$$\mathbb{P}\left(-u \leq \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \leq u\right) = 1 - \alpha$$

$$\mathbb{P}\left(\frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \leq u\right) = 1 - \alpha/2, \quad u = \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha/2)$$

Problème : $p(1-p)$ inconnu.

On remplace $\sqrt{\frac{p(1-p)}{n}}$ par $\sqrt{\frac{f(1-f)}{n}}$, f étant l'estimation de p . D'où :

$$IC(p) = \left[f - u_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + u_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right].$$

3.5 Exercices

Exercice 3.1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une v.a. X . Dans chacun des cas suivants déterminer des statistiques exhaustives.

1) $X \rightsquigarrow \mathcal{G}(p)$, $0 < p < 1$.

2) X admet comme densité de probabilité $f(x; \lambda, \mu) = \frac{1}{\lambda} e^{-\left(\frac{x-\mu}{\lambda}\right)}$ avec $\lambda > 0$; $\mu > 0$ et $x > \mu$.

3) $X \hookrightarrow \mathcal{U}[\theta_1, \theta_2]$ avec $\theta_2 > \theta_1$.

Exercice 3.2. Déterminer les estimateurs du maximum de vraisemblance du couple de paramètres $(\theta; \mu)$, en se basant sur un échantillon issu d'une loi $\mathcal{U}_{[\theta; \theta+\mu]}$

Exercice 3.3. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une v.a. X dont la fonction de répartition est définie par :

$$F(x; \theta) = \begin{cases} 1 - \frac{x}{\theta}, & \text{si } x \geq \theta > 0; \\ 0, & \text{sinon.} \end{cases}$$

1. Déterminer la densité de probabilité de la v.a. X . Cette loi appartient-elle à la famille exponentielle ?
2. Trouver une statistique suffisante T pour θ . Déterminer sa loi de probabilité. Calculer son espérance et sa variance.
3. Trouver le MLE de θ . Est-il convergent ?
4. On propose un autre estimateur T^* pour θ tel que : $T^* = \frac{n-1}{n}T$. Calculer le risque quadratique associé à cet estimateur.
5. Quel est le meilleur des deux estimateurs ? Pourquoi ?

Exercice 3.4. On admet que la durée de vie d'un matériel est représentée par une v.a. X absolument continue de loi $\mathcal{E}(1/\theta)$, θ étant un paramètre inconnu strictement positif. Dans toute la suite, on dispose d'un échantillon (X_1, X_2, \dots, X_n) issu de X .

1. Montrer que la v.a. $Y = \frac{2X}{\theta}$ suit une loi de χ_2^2 . En déduire que $\frac{2}{\theta} \sum_{i=1}^n X_i$ suit la loi de χ_{2n}^2 .
2. Déterminer l'E.M.L $\hat{\theta}_n$ de θ . Montrer qu'il est sans biais et convergent. Est-il efficace ? Quelle estimation de θ proposeriez-vous après avoir observé $\sum_{i=1}^{10} x_i = 11.5$?
3. Construire pour $n = 10$ un intervalle de confiance pour θ au niveau 0.95. En notant L la v.a. longueur de cet intervalle, calculer, en fonction de θ , $\mathbb{E}(L)$ et $\mathbb{V}(L)$. Quelle estimation par intervalle de θ proposeriez-vous après avoir observé $\sum_{i=1}^{10} x_i = 11.5$?

Exercice 3.5. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une v.a X de densité de probabilité

$$f(x; \theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0$$

1. Déterminer une statistique exhaustive pour θ .
2. Déterminer la loi de probabilité de la v.a. $Y = -2\theta \ln X$.
3. Déterminer l'estimateur de maximum de vraisemblance de θ , noté T .
4. Déterminer la loi de $\frac{n}{T}$.
5. Déduire de la question 2 que $\frac{1}{T}$ est un estimateur sans biais de $\frac{1}{\theta}$.
6. Déduire de la question 2 la variance de $\frac{1}{T}$ et montrer qu'elle atteint la borne de Cramer-Rao. Conclure.
7. Déterminer un intervalle de confiance de niveau $1 - \alpha$, pour θ ($0 < \alpha < 1$ donné).

Chapitre 4

Tests statistiques

Introduction

Un test statistique est une procédure de décision qui permet de choisir entre deux hypothèses antagonistes (contraires) faites sur une population ou sur un ou plusieurs paramètres au vu d'un échantillon aléatoire. Un test statistique est basé sur les données d'une expérience aléatoire et l'essence d'un test est une règle qui dit au statisticien si les données qu'il a recueillies le conduisent à accepter ou rejeter l'hypothèse.

On distingue deux catégories de tests :

- a) Tests paramétriques : tester certaine hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire de densité de probabilité connue.
- b) Tests non paramétriques : utilisés lorsqu'on a à faire à un échantillon aléatoire dont la loi est complètement inconnue.

4.1 Notions générales

Un test statistique permet de trancher entre deux hypothèses :

Hypothèse nulle notée H_0 : c'est l'hypothèse à tester.

Hypothèse alternative notée H_1 : hypothèse contraire.

Définition 4.1. Une hypothèse statistique est une assertion (affirmation) sur la distribution d'une ou plusieurs variables ou sur les paramètres des distributions.

Il existe deux sorte d'hypothèses : hypothèse simple et composite.

Si l'hypothèse est réduite à un seul paramètre de l'espace des paramètres, on parlera d'hypothèse simple sinon on parlera d'hypothèse composite (multiple). $H_0 : \theta = \theta_0$ hypothèse simple. $H_1 : \theta > \theta_0$, $H_1 : \theta < \theta_0$, $H_1 : \theta \neq \theta_0$ sont composites (multiples).

Remarque 4.1. La nature de H_0 détermine la façon de formuler H_1 par conséquent la nature unilatérale ou bilatérale du test.

- Test bilatéral : $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ ($\theta > \theta_0$ ou $\theta < \theta_0$).
- Test unilatéral : $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$
ou bien $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$.

On ne cherche pas à déterminer si H_0 est vraie mais seulement si on ne possède pas de raisons majeurs au vu d'un échantillon de la rejeter.

H_0 est donc privilégiée par rapport à H_1 .

Exemple 4.1. *Test sur le paramètre p de la loi de Bernoulli.*

Soit $X \rightsquigarrow \mathcal{B}(p)$:

$H_0 : p = 1/2$ contre $H_1 : p > 1/2$. On peut aussi tester $H_0 : p \geq 1/2$ contre $H_1 : p < 1/2$.

Exemple 4.2. *Test de comparaison de deux échantillons.*

Soient (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$ et (Y_1, Y_2, \dots, Y_n) un n -échantillon issu de $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$

Comparer ces deux échantillons revient à comparer leur moyennes et leur variances, i.e tester :

$H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$

$H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Types d'erreurs

Lorsqu'on effectue un test statistique 4 situations peuvent se présenter :

1. Rejeter H_0 alors qu'elle est vraie : on dit qu'on commet une erreur de première espèce (erreur de type I).
2. Accepter H_0 alors qu'elle est fautive (erreur de deuxième espèce ou bien de type II).

3. Accepter H_0 alors qu'elle est vraie.
4. Rejeter H_0 alors qu'elle est fausse.

Vérité	H_0	H_1
Décision		
H_0	$1 - \alpha$	β erreur de deuxième espèce
H_1	α erreur de première espèce	$1 - \beta$

$$\alpha = \mathbb{P}(H_1|H_0 \text{ vraie}) = \mathbb{P}(\overline{H_0}|H_0).$$

$$\beta = \mathbb{P}(H_0|H_1 \text{ vraie}) = \mathbb{P}(H_0|\overline{H_0}).$$

α et β sont appelées risque de première espèce et risque de deuxième espèce (resp).

Exemple 4.3. *Après un diagnostic médical, un médecin doit décider si un individu est malade ou non. H_0 : présence de la maladie.*

Le médecin commet une erreur de première espèce, en déclarant qu'un individu n'est pas malade alors qu'il l'est. Il commet une erreur de deuxième espèce en signalant la présence de la maladie chez un individu qui n'est pas malade.

Remarque 4.2. En général les deux types d'erreurs sont liés mathématiquement : la croissance de l'une résulte de la décroissance de l'autre. Dans la pratique des tests statistiques on fixe α (niveau du test) : $\alpha \in \{0.05, 0.01, 0.1\}$.

Variable de décision

C'est une variable (statistique) qui doit apporté le maximum d'informations sur le problème posé et dont la loi sera différente sous H_0 et sous H_1 . Il faut que sa loi soit entièrement connue au moins sous H_0 (i.e lorsque H_0 est vraie).

Région critique (région de rejet)

La région critique W est l'ensemble des valeurs de la variable de décision qui conduisent à rejeter H_0 alors qu'elle est vraie.

La détermination de la région critique se fait en écrivant : $\mathbb{P}(W|H_0) = \alpha$.

La région d'acceptation est \overline{W} tel que :

$$\mathbb{P}(\overline{W}|H_0) = 1 - \alpha.$$

Remarque 4.3. Déterminer un test statistique, revient à déterminer sa région critique. Cette région se détermine à priori (sans connaître les résultats de l'expérience).

Puissance d'un test

On appelle puissance d'un test la probabilité $1 - \beta$ de rejeter H_0 alors qu'elle est fautive :

$$1 - \beta = \mathbb{P}(W|H_1).$$

Les étapes de construction d'un test statistique

1. Choix de H_0 et H_1 .
2. Détermination de la variable de décision.
3. Calcul de la région critique (W) en fonction de α .
4. Calcul de la puissance $1 - \beta$.
5. Calcul de la valeur expérimentale de la variable de décision.
6. Conclusion : rejet ou non rejet de H_0 .

Variable de décision et région critique optimale (test de Neyman-Pearson)

L'idée de N-P est de déterminer une région critique rendant β minimum ($1 - \beta$ maximale), pour un niveau α fixé. Une telle région, si elle existe est appelée "région critique optimale". Il s'agit donc de maximiser la puissance $1 - \beta$ pour un risque α fixé.

4.2 Test entre deux hypothèses simples

4.2.1 La méthode de Neyman-Pearson

Soit X une variable aléatoire de densité $f(x, \theta)$ où θ est un paramètre réel inconnu.

Il s'agit de tester :

$$\begin{cases} H_0 : \theta = \theta_0, \\ \text{contre,} \\ H_1 : \theta = \theta_1. \end{cases}$$

Supposons α connu et soit W une région critique de \mathbb{R}^n telle que :

$$\int_W L(\underline{x}, \theta_0) d\underline{x} = \alpha = \mathbb{P}(W|H_0).$$

Il s'agit de maximiser : $1 - \beta = \int_W L(\underline{x}, \theta_1) d\underline{x} = \alpha = \mathbb{P}(W|H_1)$, on peut écrire :

$$1 - \beta = \int_W \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x}.$$

Théorème 4.1. (Neyman-Pearson) [9]

La région critique optimale est définie par l'ensemble des points de \mathbb{R}^n tels que :

$$\frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > k_\alpha.$$

Démonstration. (a) S'il existe une constante k_α , telle que l'ensemble W des points de \mathbb{R}^n où : $\frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > k_\alpha$ soit de probabilité α sous H_0 : $\mathbb{P}(W|H_0) = \alpha$, alors cette région critique réalise le maximum de $1 - \beta$.

En effet : Soit W' une autre région de \mathbb{R}^n | $\mathbb{P}(W'|H_0) = \alpha$. W' diffère alors de W par des points où : $\frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} \leq k_\alpha$.

$\int_W \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x}$ diffère de $\int_{W'} \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x}$ pour les parties non communes à W et W'

$$\mathbb{P}(W|H_0) = \mathbb{P}(W'|H_0) = \alpha$$

$$\mathbb{P}(W - W'|H_0) = \mathbb{P}(W' - W|H_0)$$

$$\int_{W-W'} \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x} > \int_{W'-W} \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x}$$

En effet :

$$\int_{W'-W} \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x} = \frac{L(\xi', \theta_1)}{L(\xi', \theta_0)} \mathbb{P}(W' - W | H_0) \text{ avec } \xi' \in W' - W.$$

$$\int_{W-W'} \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} L(\underline{x}, \theta_0) d\underline{x} = \frac{L(\xi, \theta_1)}{L(\xi, \theta_0)} \mathbb{P}(W - W' | H_0) \text{ avec } \xi \in W - W'.$$

Alors :

$$\frac{L(\xi', \theta_1)}{L(\xi', \theta_0)} \leq k_\alpha < \frac{L(\xi, \theta_1)}{L(\xi, \theta_0)}.$$

D'où a) est démontré.

(b) Montrons que cette constante k_α existe.

Soit $A(k)$ la région de \mathbb{R}^n où, $L(\underline{x}, \theta_1) > kL(\underline{x}, \theta_0)$ et considérons $\mathbb{P}(A(k)|H_0)$ qui est une fonction monotone continue de k , si X est à densité continue.

Comme $L(\underline{x}, \theta_1) > 0$, car c'est une densité, on a $\mathbb{P}(A(0)|H_0) = 1$. D'autre part si $k \rightarrow \infty$ avec une densité bornée on a $\mathbb{P}(A(k)|H_0) \rightarrow 0$. Il existe alors une valeur intermédiaire k_α t.q $\mathbb{P}(A(k_\alpha)) = \alpha$.

Etude de $1 - \beta$: Puissance du test.

Montrons que $1 - \beta > \alpha$, un tel test est dit sans biais.

$$\mathbb{P}(W|H_1) > \mathbb{P}(W|H_0)?$$

Montrons que le test de N-P est sans biais :

Puisque $L(\underline{x}, \theta_1) > k_\alpha L(\underline{x}, \theta_0)$. D'où :

$$\int_W L(\underline{x}, \theta_1) d\underline{x} > k_\alpha \int_W L(\underline{x}, \theta_0) d\underline{x}.$$

Si $k_\alpha > 1$, la proposition est triviale.

Si $k_\alpha < 1$: ceci revient à montrer que $\beta < 1 - \alpha$.

$\beta = \mathbb{P}(\overline{W}|H_1)$ et $1 - \alpha = \mathbb{P}(\overline{W}|H_0)$.

$$\overline{W} = \{\underline{x} \in \mathbb{R}^n : \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} \leq k_\alpha\}.$$

Donc $\int_{\overline{W}} L(\underline{x}, \theta_1) d\underline{x} < k_\alpha \int_{\overline{W}} L(\underline{x}, \theta_0) d\underline{x}$. D'où :

$$\int_{\overline{W}} L(\underline{x}, \theta_1) d\underline{x} < \int_{\overline{W}} L(\underline{x}, \theta_0) d\underline{x}$$

i.e : $\beta < 1 - \alpha$, d'où le résultat.

Conclusion : Les test de N-P sont sans biais.

□

Remarque 4.4. On détermine les valeurs de k_α à partir de $\mathbb{P}(W|H_0) = \alpha$, α donné où :

$$W = \{\underline{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n, U(\underline{x}) = \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > k_\alpha\}.$$

Si $U(\underline{x}) > k_\alpha$: on rejette H_0 ;

Si $U(\underline{x}) \leq k_\alpha$: on ne rejette pas H_0 (on accepte H_0).

Exemple 4.4. Soit (X_1, X_2, \dots, X_n) un échantillon issu de $X \rightsquigarrow \mathcal{B}(\theta)$,

θ inconnu, $0 \leq \theta < 1$.

Tester au niveau α : $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ avec $\theta_0 < \theta_1$.

La loi de X s'écrit : $f(x, \theta) = \theta^x(1 - \theta)^{1-x}$, $x \in \{0, 1\}$.

$$L(\underline{x}, \theta) = \prod_{i=1}^n f(x_i, \theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

D'après N-P la région critique optimale est :

$$\begin{aligned} W &= \{\underline{x} \in \mathbb{R}^n \mid \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > k_\alpha\} \\ \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} &= \frac{\theta_1^{\sum_{i=1}^n x_i} (1 - \theta_1)^{n - \sum_{i=1}^n x_i}}{\theta_0^{\sum_{i=1}^n x_i} (1 - \theta_0)^{n - \sum_{i=1}^n x_i}} \\ &= \left(\frac{1 - \theta_1}{1 - \theta_0} \right)^n \left[\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right]^{\sum_{i=1}^n x_i} \\ \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > k_\alpha &\Rightarrow \ln \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > \ln k_\alpha \\ \ln \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} &= n \ln \frac{1 - \theta_1}{1 - \theta_0} + \sum_{i=1}^n x_i \ln \frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} > \ln k_\alpha \\ \Rightarrow \sum_{i=1}^n x_i &> \frac{\ln k_\alpha - n \ln \frac{1 - \theta_1}{1 - \theta_0}}{\ln \frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}} = k_\alpha. \end{aligned}$$

Donc $W = \{\underline{x} \in \mathbb{R}^n \mid \sum_{i=1}^n x_i > k_\alpha\}$.

k_α est telle que : $\mathbb{P}_{H_0}(W) = \mathbb{P}_{H_0}(\sum_{i=1}^n x_i > k_\alpha) = \alpha$, mais sous

$H_0 : \sum_{i=1}^n X_i \rightsquigarrow \mathcal{B}(n, \theta_0)$. La puissance du test est $1 - \beta = \mathbb{P}_{H_1}(W)$ i.e. : $\mathbb{P}_{H_1}(\sum_{i=1}^n x_i > k_\alpha)$.

Sous $H_1 : \sum_{i=1}^n X_i \rightsquigarrow \mathcal{B}(n, \theta_1)$ (la loi binomiale est tabulée).

Exemple 4.5. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, 1)$.

1. Tester au niveau $\alpha = 5\%$: $H_0 : \mu = m_0$ contre $H_1 : \mu = m_1$, m_0, m_1 données
 $m_1 > m_0$.
2. Calculer la puissance du test.
3. A.N $m_0 = 0$, $m_1 = 2$, $\bar{x} = 0.6$, $n = 16$.
4. Que se passera-t-il si $H_1 : m_1 = 0.5$?

$$\begin{aligned}
 W &= \left\{ \underline{x} \in \mathbb{R}^n, \frac{L(\underline{x}, m_1)}{L(\underline{x}, m_0)} > k_\alpha \right\} \\
 \frac{L(\underline{x}, m_1)}{L(\underline{x}, m_0)} &= \frac{\frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - m_1)^2}}{\frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - m_0)^2}} \\
 &= e^{-\frac{1}{2} \sum_{i=1}^n (x_i - m_1)^2} \cdot e^{\frac{1}{2} \sum_{i=1}^n (x_i - m_0)^2} \\
 \ln \frac{L(\underline{x}, m_1)}{L(\underline{x}, m_0)} > \ln k_\alpha &\Rightarrow -\frac{1}{2} \sum_{i=1}^n (x_i - m_1)^2 + \frac{1}{2} \sum_{i=1}^n (x_i - m_0)^2 > \ln k_\alpha \\
 &\Rightarrow -\frac{1}{2} \sum_{i=1}^n x_i^2 + m_1 \sum_{i=1}^n x_i - \frac{n}{2} m_1^2 + \frac{1}{2} \sum_{i=1}^n x_i^2 \\
 &\quad + m_0 \sum_{i=1}^n x_i + \frac{n}{2} m_0^2 > \ln k_\alpha \\
 &\Rightarrow (m_1 - m_0) \sum_{i=1}^n x_i - \frac{n}{2} (m_1^2 - m_0^2) > \ln k_\alpha \\
 &\Rightarrow (m_1 - m_0) \sum_{i=1}^n x_i > \frac{n}{2} (m_1^2 - m_0^2) + \ln k_\alpha \\
 &\Rightarrow \sum_{i=1}^n x_i > k_\alpha.
 \end{aligned}$$

D'où :

$$W = \left\{ \underline{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i > k_\alpha \right\}.$$

$\mathbb{P}(W|H_0) = \alpha$ or sous $H_0 : \sum_{i=1}^n X_i \rightsquigarrow \mathcal{N}(nm_0, n)$.

$$\mathbb{P}\left(\sum_{i=1}^n X_i > k_\alpha\right) = \alpha \Rightarrow \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - nm_0}{\sqrt{n}} < \frac{k_\alpha - nm_0}{\sqrt{n}}\right) = 1 - \alpha.$$

D'où :

$$\frac{k_\alpha - nm_0}{\sqrt{n}} = \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha).$$

(ϕ fonction de répartition de $\mathcal{N}(0, 1)$).

Alors :

$$k_\alpha = nm_0 + \sqrt{n}\phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha).$$

A.N : $m_0 = 0$, $m_1 = 2$, $\sum_{i=1}^n x_i = n\bar{x} = 16 \cdot 0.6 = 9.6$, $n = 16$.

$$\phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) = \phi_{\mathcal{N}(0,1)}^{-1}(0.95) = 1.64.$$

$$k_\alpha = 16 \cdot 0 + \sqrt{16} \cdot 1.64 = 4 \cdot 1.64 = 6.56.$$

On a : $\sum_{i=1}^n x_i = 9.6 > k_\alpha = 6.56 \Rightarrow$ on rejette H_0 .

Puissance du test :

$$\begin{aligned} 1 - \beta = \mathbb{P}(W|H_1) &= \mathbb{P}_{H_1}\left(\sum_{i=1}^n X_i > k_\alpha\right) \\ &= \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - nm_1}{\sqrt{n}} > \frac{k_\alpha - nm_1}{\sqrt{n}}\right) \\ &= 1 - \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - nm_1}{\sqrt{n}} \leq \frac{k_\alpha - nm_1}{\sqrt{n}}\right) \\ &= 1 - \phi_{\mathcal{N}(0,1)}\left(\frac{k_\alpha - nm_1}{\sqrt{n}}\right) \\ &= 1 - \phi\left[\frac{\phi^{-1}(1 - \alpha) + n(m_0 - m_1)}{\sqrt{n}}\right] \end{aligned}$$

$$1 - \beta = 1 - \phi\left[\sqrt{n}\phi^{-1}(1 - \alpha) + \sqrt{n}(m_0 - m_1)\right].$$

A.N. :

$$\begin{aligned} 1 - \beta &= 1 - \phi[1.64 + 4(-2)] = 1 - \phi(-6.36) \\ &= 1 - [1 - \phi(6.36)] \\ &= \phi(6.36) = 1 \end{aligned}$$

$$\Rightarrow \beta = 0.$$

Si $m_1 = 0.5$, la région critique ne change pas la puissance du test.

$$1 - \beta = \phi(0.36) = 0.64 \Rightarrow \beta = 0.36 \text{ (grand).}$$

4.3 Tests d'hypothèses simple contre hypothèse composite

Ce sont des tests de la forme : $H_0 : \theta = \theta_0$ contre $H_1 : \theta \in \Theta_1 \subset \mathbb{R}$.

Exemple 4.6.

$$\begin{cases} H_0 : \theta = \theta_0, & \text{contre;} \\ H_1 : \theta > \theta_0. \end{cases} \quad (4.1)$$

$$\begin{cases} H_0 : \theta = \theta_0, & \text{contre;} \\ H_1 : \theta < \theta_0. \end{cases} \quad (4.2)$$

$$\begin{cases} H_0 : \theta = \theta_0, & \text{contre;} \\ H_1 : \theta \neq \theta_0. \end{cases} \quad (4.3)$$

La nature du problème et l'idée principale de la solution ne changent pas par rapport au cas d'une hypothèse simple contre hypothèse simple. Cependant une différence apparaît au niveau des caractéristiques du test et qui est en rapport direct avec la complexité de l'hypothèse alternative. L'objectif est toujours de trouver la région critique (région de rejet de H_0).

Tout test d'hypothèse simple contre composite est caractérisé par le niveau α et une fonction puissance :

$$\begin{aligned} \pi(\theta) = 1 - \beta(\theta) : \Theta_1 &\longrightarrow [0, 1] \\ \theta_1 &\longmapsto 1 - \beta(\theta) \end{aligned}$$

Test uniformément le plus puissant

Un test est dit uniformément le plus puissant si $\forall \theta \in \Theta_1$, sa puissance $\pi(\theta) = 1 - \beta(\theta)$ est supérieure à la puissance de tout autre test.

Pour chaque valeur θ dans Θ_1 , on sait déterminer une région critique optimale, le problème est de construire une qui soit optimale pour tout l'ensemble Θ_1 , i.e : construire un test uniformément le plus puissant.

Définition 4.2. On dit que le test de région critique W est uniformément le plus puissant de niveau α pour un problème de test d'une hypothèse simple contre une hypothèse composite si $\forall W'$ une autre région critique d'un autre test au niveau α , on a :

$$1 - \beta(\theta) = \mathbb{P}_{H_1}(W) > 1 - \beta'(\theta) = \mathbb{P}_{H_1}(W'), \quad \forall \theta \in \Theta_1.$$

Théorème 4.2. Une condition nécessaire et suffisante d'existence d'un test uniformément le plus puissant au niveau α pour tester une hypothèse simple $H_0 : \theta = \theta_0$ contre une hypothèse composite $H_1 : \theta \in \Theta_1$ est :

1. $\forall \theta_1 \in \Theta_1$ il existe un test le plus puissant de niveau α pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ (N-P).
2. La forme de la région critique du test précédent ne doit pas dépendre de $\theta \in \Theta_1$ (sous H_1).

Remarque 4.5. Les exemples 4.1 et 4.2 sont uniformément les plus puissants, mais 4.3 ne l'est pas.

4.4 Tests paramétriques usuels

(a) Test sur la moyenne d'une loi normale $\mathcal{N}(\mu, \sigma^2)$.

- σ^2 connue : $\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu = \mu_1. \end{cases}$ avec $\mu_1 > \mu_0$.

On utilise la variable de décision \bar{X} .

$$W = \{\underline{x} \in \mathbb{R}^n, \bar{x} > k\}, \bar{X} \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n).$$

$$\begin{aligned} \mathbb{P}_{H_0}(W) = \alpha &\Rightarrow \mathbb{P}_{H_0}(\bar{X} > k) = \alpha \\ &\Rightarrow \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = \alpha \\ &\Rightarrow \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha \end{aligned}$$

$$k = \frac{\sigma}{\sqrt{n}} \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) + \mu_0.$$

Décision : $\begin{cases} \bar{x} > k, & \text{on rejette } H_0; \\ \bar{x} \leq k, & \text{on ne rejette pas } H_0. \end{cases}$

- σ^2 inconnue : σ^2 est estimée par $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

On a $\frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$ et

$$T_{n-1} = \frac{\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S'^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{S'} \sqrt{n} \rightsquigarrow \text{Student } a \text{ } (n-1).$$

$H_0 : \mu = \mu_0, H_1 : \mu = \mu_1 :$

$$W = \{\underline{x} \in \mathbb{R}^n, |T_{n-1}| > k\}$$

k est déterminé par $\mathbb{P}_{H_0}(W) = \alpha = \mathbb{P}(|\frac{\bar{X} - \mu_0}{S'} \sqrt{n}| > k) = \alpha$

Remarque sur les tests de la moyenne : Si la variable aléatoire parente X ne suit pas une loi normale, les tests précédents s'appliquent encore dès que n est assez grand ($n > 30$), en raison du théorème centrale limite.

(b) Test sur les proportions :

Soit X une variable aléatoire observée sur une population et soit (X_1, X_2, \dots, X_n) un n-échantillon extrait de cette population. On veut savoir si cet échantillon de fréquence observée $\frac{K}{n} = F$ estimateur de p , appartient à une population de fréquence p_0 (sous H_0) ou à une autre population inconnu de fréquence p (sous H_1).

$$\begin{cases} H_0 : p = p_0, & \text{contre} \\ H_1 : p \neq p_0. \end{cases}$$

On sait que $\frac{K}{n} = F \rightsquigarrow \mathcal{N}(p_0, \sqrt{\frac{p_0 q_0}{n}})$ sous H_0 .

$$W = \{\underline{x} \in \mathbb{R}^n, |\frac{K}{n} - p_0| > k\}.$$

$$\mathbb{P}(|\frac{K}{n} - p_0| > k) = \alpha \Rightarrow \mathbb{P}(\frac{|\frac{K}{n} - p_0|}{\sqrt{\frac{p_0 q_0}{n}}} > \frac{k}{\sqrt{\frac{p_0 q_0}{n}}}) = \alpha$$

$$k = \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) \sqrt{\frac{p_0 q_0}{n}} = \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) \sqrt{p_0 \frac{1 - p_0}{n}}.$$

Exemple 4.7. Sur un échantillon de 200 étudiants 45% sont favorables qu'il y ait synthèse à la fin de l'année. Ceci contredit-il l'hypothèse qu'un étudiant sur deux y est favorable ?

$$\begin{cases} H_0 : p = \frac{1}{2}, \\ H_1 : p \neq \frac{1}{2}. \end{cases}$$

$\alpha = 0.05, u = 1.96.$

$$\begin{aligned} W &= \{\underline{x} \in \mathbb{R}^n, |F - \frac{1}{2}| > 1.96 \sqrt{\frac{(1/2)^2}{200}}\} \\ &= \{\underline{x} \in \mathbb{R}^n, |F - 0.5| > 0.07\}. \end{aligned}$$

Comme $|f - 0.5| = |0.45 - 0.50| = 0.05$ donc on ne rejette pas H_0 au seuil $\alpha = 0.05$.

4.5 Application du test de N-P dans le cas de famille exponentielle

Si la loi de X appartient à la famille exponentielle pour θ , sa densité s'écrit :

$$\begin{aligned} L(\underline{x}, \theta) &= \prod_{i=1}^n f(x_i, \theta) \\ \ln L(\underline{x}, \theta) &= \sum_{i=1}^n \ln f(x_i, \theta) \\ &= \sum_{i=1}^n a(x_i) \alpha(\theta) + \sum_{i=1}^n b(x_i) + n\beta(\theta) \end{aligned}$$

$$\ln \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} > \ln k_\alpha \Rightarrow [\ln L(\underline{x}, \theta_1) - \ln L(\underline{x}, \theta_0)] > \ln k_\alpha$$

alors :

$$[\alpha(\theta_1) - \alpha(\theta_0)] \sum_{i=1}^n a(x_i) > k'_\alpha.$$

Exemple 4.8. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{P}(\lambda)$. Pour α donné tester : $H_0 : \lambda = \lambda_0$ contre $H_1 : \lambda = \lambda_1$. On a :

$$\begin{aligned} f(x, \lambda) &= e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, \dots \\ \ln f(x, \lambda) &= -\lambda + x \ln \lambda - \ln x_i! \\ &= x \ln \lambda - \lambda - \ln x_i! \\ &= a(x)\alpha(\lambda) + \beta(\lambda) + b(x) \end{aligned}$$

avec $\alpha(\lambda) = \ln \lambda$, $a(x) = x$. D'où :

$$W = \{\underline{x} \in \mathbb{R}^n \mid \ln \frac{\lambda_1}{\lambda_0} \sum_{i=1}^n x_i > k_\alpha\}.$$

Si $\lambda_1 > \lambda_0$: $W = \{\underline{x} \in \mathbb{R}^n \mid \sum_{i=1}^n x_i > k_1\}$,

si $\lambda_0 > \lambda_1$: $W = \{\underline{x} \in \mathbb{R}^n \mid \sum_{i=1}^n x_i \leq k_2\}$.

A.N : $\lambda_0 = 1$, $\lambda_1 = 2$, $n = 2$, $\alpha = 5\%$

$$W = \{\underline{x} \in \mathbb{R}^n, \sum_{i=1}^n x_i > k_1\}$$

$\alpha = \mathbb{P}(W|H_0) = \mathbb{P}(\sum_{i=1}^n X_i > k_1)$, $\sum_{i=1}^n X_i \rightsquigarrow \mathcal{P}(n\lambda_0)$ sous H_0 .

$$1 - \mathbb{P}(\sum_{i=1}^n X_i \leq k_2) = 1 - \alpha$$

$$1 - \sum_{j=0}^{k_1} \mathbb{P}_{H_0}(\sum_{i=1}^n X_i = j) = 1 - \alpha$$

$$1 - \sum_{j=0}^{k_1} e^{-n\lambda_0} \frac{(n\lambda_0)^j}{j!} = 1 - \alpha = 0.95.$$

$$\sum_{j=0}^{k_1} e^{-2} \frac{2^j}{j!} = 0.95 \Rightarrow k_1 = 4 \quad (\text{la loi de Poisson est tabulée}).$$

Si $x_1 + x_2 > 4$ on rejette H_0 ,

si $x_1 + x_2 \leq 4$ on ne rejette pas H_0 .

Remarque 4.6. Si $n\lambda > 20$, on peut approcher la loi de Poisson par la loi normale.

Calcul de la puissance :

Sous $H_1 : \lambda = \lambda_1$, $\sum_{i=1}^n X_i \rightsquigarrow \mathcal{P}(n\lambda_1)$ $\lambda_1 > \lambda_0$.

$$\begin{aligned}
 1 - \beta = \mathbb{P}(W|H_1) &= \mathbb{P}\left(\sum_{i=1}^n X_i > k_\alpha\right) \\
 &= 1 - \mathbb{P}\left(\sum_{i=1}^n X_i \leq k_\alpha\right) \\
 &= 1 - \sum_{j=0}^{k_\alpha} \mathbb{P}\left(\sum_{i=1}^n X_i = j\right) \\
 &= 1 - \mathbb{P}(\mathcal{P}(4) \leq 4) \\
 &= 1 - \sum_{j=0}^{k_\alpha} e^{-n\lambda_1} \frac{(n\lambda_1)^j}{j!} \\
 &= 1 - \sum_{j=0}^4 e^{-4} \frac{4^j}{j!} \\
 &= 1 - \mathbb{P}(\mathcal{P}(4) \leq 4) = 1 - 0.6288 = 0.3712.
 \end{aligned}$$

4.6 Tests et statistique exhaustive

La considération d'une statistique exhaustive simplifie la recherche de la région critique du test. S'elle existe une statistique exhaustive T pour θ , de densité $g(t, \theta)$ on a :

$$L(\underline{x}, \theta) = g(t, \theta) \cdot h(\underline{x}).$$

Le test de N-P se réduit alors à :

$$\frac{g(t, \theta_1)}{g(t, \theta_0)} > k_\alpha.$$

4.7 Test entre deux hypothèses multiples (test de Lehman)

4.7.1 Famille à rapport de vraisemblance monotone

Si H_0 est composite, le risque de première espèce est une fonction de θ , et le niveau du test est :

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta), \quad \Theta_0 \subset \mathbb{R}, \quad \alpha(\theta) \leq \alpha \text{ donné.}$$

Définition 4.3. Soit X_1, X_2, \dots, X_n un n-échantillon issu de X . La loi de X est dite à rapport de vraisemblance monotone s'il existe une statistique T à valeur dans \mathbb{R} tel que :

$$\frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} = f_{\theta_0, \theta_1}(t)$$

est une fonction monotone croissante de T si $\theta_0 < \theta_1$

Cas des familles exponentielles

$$\ln f(x, \theta) = a(x)q(\theta) + A(\theta) + B(x)$$

$$\ln \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)} = [q(\theta_1) - q(\theta_0)] \sum_{i=1}^n a(x_i) + nA(\theta_1) - nA(\theta_0).$$

$T = \sum_{i=1}^n a(X_i)$ est une statistique exhaustive.

$\ln \frac{L(\underline{x}, \theta_1)}{L(\underline{x}, \theta_0)}$ est une fonction monotone croissante en T si et seulement si $q(\theta_1) - q(\theta_0) > 0$ lorsque $\theta_1 > \theta_0$ i.e q doit être fonction croissante de θ .

Remarque 4.7. Les lois de famille exponentielle vérifiant cette propriété sont donc à rapport de vraisemblance monotone.

4.7.2 Tests bilatérales

Test de la forme $H_0 : \theta \leq \theta_1$ ou $\theta \geq \theta_2$ contre $H_1 : \theta_1 < \theta < \theta_2$.

Théorème 4.3. (*Lehmann*)

Soit X une variable aléatoire de loi $f(x, \theta)$ et T une statistique. Si la loi de X est à rapport de vraisemblance monotone en T alors il existe un test uniformément le plus puissant de niveau α donné dont la région de rejet est de la forme :

$$W = \{\underline{x} \in \mathbb{R}^n, K_1 < t(\underline{x}) < K_2\}.$$

Remarque 4.8. Les constantes K_1 et K_2 sont déterminées par : $\mathbb{P}_{H_0}(W) = \alpha$.

Test de la forme : $H_0 : \theta_1 \leq \theta \leq \theta_2$ contre $H_1 : \theta > \theta_2$ ou $\theta < \theta_1$

$$W = \{\underline{x} \in \mathbb{R}, t(\underline{x}) < c_1 \text{ ou } t(\underline{x}) > c_2\}$$

avec $\mathbb{P}_{H_0}(W) = \alpha \Rightarrow \mathbb{P}_{\theta_1}(T < c_1) = \mathbb{P}_{\theta_2}(T > c_2) = \frac{\alpha}{2}$.

4.7.3 Tests unilatérales

Test de la forme $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$.

Théorème 4.4. *Soit X une variable aléatoire de densité $f(x, \theta)$, si la loi de X est à rapport de vraisemblance monotone en une statistique T , alors il existe un test uniformément le plus puissant dont la région critique est de la forme :*

$$W = \{\underline{x} \in \mathbb{R}^n, t(\underline{x}) > k\}$$

tel que : $\mathbb{P}_{H_0}(W) = \alpha$.

Exemple 4.9. *Soit $X \rightsquigarrow \mathcal{B}(p)$, X_1, X_2, \dots, X_n un n -échantillon issu de X . Tester : $H_0 : p \leq p_0$ contre $H_1 : p > p_0$, on a :*

$$\begin{aligned} f(x, p) &= p^x(1-p)^{1-x}, \quad x \in 0, 1 \\ \ln f(x, p) &= x \ln \frac{p}{1-p} + \ln(1-p) \\ q(p) &= \ln \frac{p}{1-p} = \ln p - \ln(1-p) \\ q'(p) &= \frac{1}{p} + \frac{1}{1-p} > 0 \Rightarrow q \text{ est croissante en } p \end{aligned}$$

D'où la loi de $\mathcal{B}(p)$ est à rapport de vraisemblance monotone en $T = \sum_{i=1}^n a(X_i) = \sum_{i=1}^n X_i$, d'où :

$$W = \{\underline{x} \in \mathbb{R}^n, \sum_{i=1}^n X_i > k_\alpha\}$$

avec sous $H_0 : \sum_{i=1}^n X_i \rightsquigarrow \mathcal{B}(n, p_0)$ tabulée, on trouve k_α par : $\mathbb{P}_{H_0}(W) = \alpha$.

L'autre test unilatéral se détermine de la même manière : $H_0 : \theta \geq \theta_0$ contre $H_1 : \theta < \theta_0$

$$W = \{\underline{x} \in \mathbb{R}^n, T(\underline{x}) < k_\alpha\}.$$

Exemple 4.10. *Soit (X_1, X_2, \dots, X_n) issu de $X \rightsquigarrow \text{Exp}(\frac{1}{\theta})$. Tester :*

$$\begin{cases} H_0 : \theta \geq \theta_0, & \text{contre;} \\ H_1 : \theta < \theta_0. \end{cases}$$

On a :

$$f(x, \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} \quad x > 0$$

$$\ln f(x, \theta) = \ln \frac{1}{\theta} - \frac{1}{\theta}x$$

On a : $q(\theta) = -\frac{1}{\theta}$ est croissante en θ d'où $T = \sum_{i=1}^n X_i$. $f(x, \theta)$ est à rapport de vraisemblance monotone en T , d'où :

$$W = \{\underline{x} \in \mathbb{R}^n, \sum_{i=1}^n x_i < k_\alpha\}$$

$$\text{on a : } \mathbb{P}_{H_0}(W) = \alpha \Rightarrow \mathbb{P}_{H_0}\left(\sum_{i=1}^n x_i < k_\alpha\right) = \alpha$$

$$\text{or on a sous } H_0 : \frac{2}{\theta_0} \sum_{i=1}^n X_i \rightsquigarrow \chi_{2n}^2 \Rightarrow k_\alpha = \frac{\theta_0}{2} \phi_{\chi_n^2}^{-1}(\alpha)$$

$$W = \{\underline{x} \in \mathbb{R}^n, \sum_{i=1}^n x_i < \frac{\theta_0}{2} \phi_{\chi_n^2}^{-1}(\alpha)\}$$

indépendante de θ donc c'est un test uniformément le plus puissant.

4.8 Test de rapport de vraisemblance maximale

Ce test est utilisé dans le cas où les méthodes précédentes ne marchent pas.

(a) Test $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ où θ peut-être un vecteur de dimension p . On pose :

$$\lambda = \frac{L(\underline{x}, \theta_0)}{\sup_{\theta \in \Theta} L(\underline{x}, \theta)}$$

Remarque 4.9. $0 \leq \lambda \leq 1$, plus λ est grand, plus H_0 est vraisemblable, cela revient à remplacer sous H_1 , θ par son estimation du maximum de vraisemblance. La région critique du test sera :

$$W = \{\underline{x} \in \mathbb{R}^n, \lambda < k_\alpha\}.$$

Théorème 4.5. [9] La distribution de $-2 \ln \lambda$ est asymptotiquement celle de χ_p^2 sous H_0 .

Remarque 4.10. $X(\Omega)$ doit être indépendant de θ .

Démonstration. Pour $p = 1$, $-2 \ln \lambda \xrightarrow[n \rightarrow \infty]{} \chi_1^2$?

$$\ln \lambda = \ln L(\underline{x}, \theta_0) - \ln L(\underline{x}, \hat{\theta})$$

le développement en série de Taylor de $\ln L(\underline{x}, \theta_0)$ au $v(\hat{\theta})$ nous donne :

$$\ln \lambda = \ln L(\underline{x}, \hat{\theta}) + (\theta_0 - \hat{\theta}) \frac{\partial \ln L(\underline{x}, \hat{\theta})}{\partial \theta} + \frac{(\theta_0 - \hat{\theta})^2}{2} \frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial \theta^2} - \ln L(\underline{x}, \hat{\theta}), \quad \theta^* \in [\theta_0, \hat{\theta}].$$

$$\lambda = \frac{1}{2} (\theta_0 - \hat{\theta})^2 \frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial \theta^2}, \quad \theta^* \in [\theta_0, \hat{\theta}].$$

Sous $H_0 : \theta = \theta_0$, l'EMV $\xrightarrow[n \rightarrow \infty]{p.s.} \theta_0$

donc $\theta^* \xrightarrow[n \rightarrow \infty]{p.s.} \theta$.

$$\frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial \theta^2} \sim \frac{\partial^2 \ln \prod_{i=1}^n f(x_i, \theta)}{\partial \theta^2} = n \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln f(x_i, \theta)}{\partial \theta^2}$$

mais :

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln f(X_i, \theta)}{\partial \theta^2} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathbb{E} \left[\frac{\partial^2 \ln f(X, \theta)}{\partial \theta^2} \right] = -I_1(\theta)$$

donc $\frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial \theta^2} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} -nI_1(\theta) = -I_n(\theta)$

d'où : $-2 \ln \lambda \longrightarrow (\theta_0 - \hat{\theta})^2 I_n(\theta)$

mais on a :

$$\frac{\theta_0 - \hat{\theta}}{\sqrt{\frac{1}{I_n(\theta)}}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \Rightarrow \frac{(\theta_0 - \hat{\theta})^2}{\frac{1}{I_n(\theta)}} = (\theta_0 - \hat{\theta})^2 I_n(\theta) \longrightarrow \chi_1^2$$

d'où :

$$-2 \ln \lambda \longrightarrow \chi_1^2.$$

(b) Test entre deux hypothèse multiple :

On forme $\lambda = \frac{\sup_{\theta \in H_0} L(\underline{x}, \theta)}{\sup_{\theta \in H_1} L(\underline{x}, \theta)}$ et on obtient les mêmes propriétés que nous avons obtenues

en (a). □

Exemple 4.11. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, 1)$, tester au niveau α : $\begin{cases} H_0 : \mu = \mu_0 = 3, & \text{contre,} \\ H_1 : \mu \neq \mu_0. \end{cases}$

$$\lambda = \frac{L(\underline{x}, \mu_0)}{L(\underline{x}, \hat{\mu})} \text{ et } W = \{\underline{x} \in \mathbb{R}^n, \lambda < k\}$$

$$\begin{aligned} \lambda &= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu_0)^2}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2}} \\ &= e^{\frac{-1}{2} [\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \bar{x})^2]} \\ -2 \ln \lambda &= \sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= -2\mu_0 n \bar{x} + n\mu_0^2 + 2n\bar{x}^2 - n\bar{x}^2 \\ &= n(\bar{x}^2 - 2\mu_0 \bar{x} + \mu_0^2) \\ &= n(\bar{x} - \mu_0)^2 \end{aligned}$$

$$\begin{aligned} \lambda < k &\Rightarrow -2 \ln \lambda > -2 \ln k \\ &\Rightarrow n(\bar{x} - \mu_0)^2 > -2 \ln k \end{aligned}$$

$$\begin{aligned} \mathbb{P}(n(\bar{x} - \mu_0)^2 > -2 \ln k) = \alpha &\Rightarrow \mathbb{P}(\chi_1^2 > -2 \ln k) = 1 - \alpha \\ -2 \ln k = \phi_{\chi_1^2}^{-1}(1 - \alpha) &\Rightarrow k = e^{\frac{-1}{2} \phi_{\chi_1^2}^{-1}(1 - \alpha)}. \end{aligned}$$

$$W = \{\underline{x} \in \mathbb{R}^n : \lambda < e^{\frac{-1}{2} \phi_{\chi_1^2}^{-1}(1 - \alpha)}\}.$$

4.9 Test de comparaison d'échantillons

4.9.1 Test de Student et F.S pour des échantillons indépendants

Etant donné deux échantillons de taille n_1 et n_2 . Peut-on admettre qu'ils ont été prélevé dans une même population relativement à la variable étudiée, ces échantillons ayant-été prélevé indépendamment l'un de l'autre ?

Exemple 4.12. *Les résultats scolaires des filles et des garçons sont-ils comparables ?*

Les ampoules de l'entreprise A ont elles une durée de vie plus longue que celle de l'entreprise B ?

Mathématiquement le problème se formalise de la manière suivante : on observe sur le premier échantillon les réalisations d'une variable aléatoire X_1 de fonction de répartition

$F_1(x)$ et sur le deuxième échantillon les réalisations d'une variable aléatoire X_2 de fonction de répartition $F_2(x)$.

On teste : $\begin{cases} H_0 : F_1(x) = F_2(x), \\ H_1 : F_1(x) \neq F_2(x). \end{cases}$

Remarque 4.11. En pratique on se contente de vérifier l'égalité des espérances et des variances de X_1 et X_2 .

(A) Cas des échantillons Gaussiens :

$$X_1 \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2), X_2 \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2).$$

Les hypothèses reviennent à tester :

$$\begin{cases} H_0 : m_1 = m_2 \text{ et } \sigma_1^2 = \sigma_2^2, & \text{contre;} \\ H_1 : m_1 \neq m_2 \text{ et } \sigma_1^2 \neq \sigma_2^2. \end{cases}$$

Le test va consister à tester d'abord les variances, si elles ne sont pas significativement différentes à tester ensuite les espérances mathématiques en admettant $\sigma_1 = \sigma_2$.

(a) Test des variances de Fisher-Snédecor :

$$S_1^2 = \frac{1}{n_1} \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 \text{ et } S_2^2 = \frac{1}{n_2} \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2$$

$$\frac{n_1 S_1^2}{\sigma_1^2} \rightsquigarrow \chi_{n_1-1}^2, \frac{n_2 S_2^2}{\sigma_2^2} \rightsquigarrow \chi_{n_2-1}^2.$$

Sous $H_0 : \sigma_1 = \sigma_2 :$

$$F = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}} \rightsquigarrow \mathcal{F}(n_1 - 1, n_2 - 1).$$

Si $\sigma_1 = \sigma_2$, F ne doit pas être significativement différente de 1, F sera la variable de décision.

$W = \{\underline{x} \in \mathbb{R}^n, f > k, k > 1\}$, f est une réalisation de F .

Remarque 4.12. En pratique, on met au numérateur la plus grande des deux quantités. Si $n_1 = n_2 = n$, $F_{n_1-1, n_1-1} = \frac{S_1^2}{S_2^2}$.

Si le test aboutit à $\sigma_1 = \sigma_2$, on passe au test des espérances.

(b) Test sur les espérances de Student :

Supposons que $\sigma_1 = \sigma_2$ on a :

$$\left\{ \begin{array}{l} \frac{n_1 S_1^2}{\sigma_1^2} \rightsquigarrow \chi_{n_1-1}^2, \\ \bar{X}_1 \rightsquigarrow \mathcal{N}(m_1, \frac{\sigma}{\sqrt{n_1}}). \end{array} \right. \text{ et } \left\{ \begin{array}{l} \frac{n_2 S_2^2}{\sigma_2^2} \rightsquigarrow \chi_{n_2-1}^2, \\ \bar{X}_2 \rightsquigarrow \mathcal{N}(m_2, \frac{\sigma}{\sqrt{n_2}}). \end{array} \right.$$

$$\frac{n_1 S_1^2 + n_2 S_2^2}{\sigma^2} \rightsquigarrow \chi_{n_1+n_2-2}^2$$

mais $\bar{X}_1 - \bar{X}_2 \rightsquigarrow \mathcal{N}(m_1 - m_2, \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$

σ étant connu, on utilise la loi de Student.

$$T_{n_1+n_2-2} = \frac{\bar{X}_1 - \bar{X}_2 - (m_1 - m_2)}{\sqrt{(n_1 S_1^2 + n_2 S_2^2)(\frac{1}{n_1} + \frac{1}{n_2})}} \sqrt{n_1 + n_2 - 2}.$$

Sous $H_0 : m_1 = m_2$, la région critique est :

$$W = \{ \underline{x} \in \mathbb{R}^n : |T| > k \}.$$

(B) Cas des échantillons non Gaussiens :

Le test des variances F n'est plus valable car $\frac{nS^2}{\sigma^2} \not\rightarrow \chi_n^2$, mais pour n_1 et n_2 assez grand, on applique la formule de Student que σ_1 soit différent ou non de σ_2 .

Remarque 4.13. On dit que le test de Student est "robuste" car il résiste au changement de loi de X_1 et X_2 .

Exemple 4.13. Soit X_1, X_2, \dots, X_n un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$ tel que :

$$n = 7, \bar{x} = 14.2, S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.0114.$$

Soit Y_1, Y_2, \dots, Y_m un m -échantillon issu de $Y \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$ tel que : $m = 5, \bar{y} =$

$$\frac{1}{m} \sum_{i=1}^m y_i = 14.5 \text{ et } S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2 = 0.05.$$

Peut-on accepter que X et Y suivent la même loi normale au niveau $\alpha = 0.05$?

1. On test d'abord l'égalité des variances :

$H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$. Ce test est basé sur la statistique de Fisher :

on a :

$$\frac{(n-1)S_X^2}{\sigma_1^2} \rightsquigarrow \chi_{n-1}^2, \quad \frac{(m-1)S_Y^2}{\sigma_2^2} \rightsquigarrow \chi_{m-1}^2$$

$$F = \frac{\frac{S_Y^2}{\sigma_2^2}}{\frac{S_X^2}{\sigma_1^2}}.$$

Sous H_0 $F = \frac{S_Y^2}{S_X^2} \rightsquigarrow \text{Fisher}(m-1, n-1)$.

$$W = \{(\underline{x}, \underline{y}) \in \mathbb{R}^n \times \mathbb{R}^m, f > k\}$$

$$f = \frac{s_Y^2}{s_X^2} \frac{0.05}{0.0114} = 4.38.$$

$$\mathbb{P}_{H_0}(W) = \alpha \Rightarrow \mathbb{P}_{H_0}(F > k) = \alpha \Rightarrow \mathbb{P}_{H_0}(F_{(m-1, n-1)} < k) = 1 - \alpha = 0.95.$$

Sur la table $F(4.6, 0.95)$ on trouve $k = 4.53$ ou $f = 4.38 < 4.53 \Rightarrow$ on ne rejette pas H_0 i.e : $\sigma_1^2 = \sigma_2^2$.

2. Test sur les moyennes :

$H_0 : m_1 = m_2$ contre $H_0 : m_1 \neq m_2$ (test bilatéral) on a : $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

L'estimateur de σ^2 est $S^2 = \frac{1}{n+m-2} [\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2]$

$$\frac{(n+m-2)S^2}{\sigma^2} = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \rightsquigarrow \chi_{n+m-2}^2.$$

On a : $\bar{X} \rightsquigarrow \mathcal{N}(m_1, \frac{\sigma^2}{n})$, $\bar{Y} \rightsquigarrow \mathcal{N}(m_2, \frac{\sigma^2}{m})$

$$\bar{X} - \bar{Y} \rightsquigarrow \mathcal{N}(m_1 - m_2, \sigma \sqrt{\frac{1}{n} + \frac{1}{m}})$$

Sous H_0 : $\bar{X} - \bar{Y} \rightsquigarrow \mathcal{N}(0, \sigma \sqrt{\frac{1}{n} + \frac{1}{m}})$

$$\frac{\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{(n+m-2)\sigma^2}}} \rightsquigarrow T_{n+m-2}.$$

$W = \{(\underline{x}, \underline{y}) \in \mathbb{R}^n \times \mathbb{R}^m, |T| > t\}$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \sqrt{\frac{n+m-2}{(n+m-2)S^2}} = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$S^2 = 0.026 \Rightarrow S = 0.16, T = \frac{0.30}{0.16 \cdot 0.47} = 3.98.$$

Sur la table de Student : $\mathbb{P}(|T_{n+m-2}| > t) = 0.05 \Rightarrow t = 2.22 \Rightarrow$ on rejette H_0 .

Comparaison des moyennes d'échantillons appariés

Un même échantillon d'individus est soumis à deux mesures successives d'une même variable.

Exemple 4.14. 15 copies sont soumises à une double correction A et B, on veut tester si les deux séries de valeurs sont semblables.

	1	2	3	4	5
A	19.6	12.7	13.7	16.4	20.5
B	17.8	16.6	14.6	16.3	18.2

Soit X_1 la variable correspondante à la première série et X_2 l'autre. On se contente de tester l'hypothèse $H_0 : \mathbb{E}(X_1) = \mathbb{E}(X_2)$.

$X_1 - X_2 \rightsquigarrow \mathcal{N}(m_1 - m_2, \sigma)$, (X_1, X_2 sont gaussiennes).

Le test de $H_0 : m_1 = m_2$ contre $H_1 : m_1 \neq m_2$ consiste à former les différences.

$x_{i1} - x_{i2} = d_i$ et à faire un test de Student sur la moyenne des d_i car σ est en général inconnu.

$$t_{n-1} = \frac{\bar{d}}{S_d} \sqrt{n-1} = \frac{\bar{X}_1 - \bar{X}_2}{S_d} \sqrt{n-1}$$

On rejette H_0 si $|t_{n-1}| > k$ i.e : $W = \{\underline{x} \in \mathbb{R}^n, |t| > k\}$.

Remarque 4.14. La différence par rapport au test de Student vient du fait que X_1, X_2 ne peuvent être supposées indépendantes.

Exemple 4.15. Considérons une étude sur l'efficacité de deux traitements PH A et B administrés à des patients d'un laboratoire médical au niveau $\alpha = 0.05$. 5 patients ont été choisis, les résultats sont donnés dans ce tableau :

Patient	1	2	3	4	5
D	1.8	-3.9	-0.9	0.1	2.3

$$D = X_1 - X_2$$

$$\bar{D} = \frac{1.8 - 3.9 - 0.9 + 0.1 + 2.3}{5} = -0.02.$$

$$\begin{aligned}
S_D^2 &= \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2 \\
&= \frac{1}{n-1} (\sum_{i=1}^n D_i^2 - n\bar{D}^2) \\
&= \frac{(1.8)^2 + (-3.9)^2 + (-0.9)^2 + (0.1)^2 + (2.3)^2 - 5(0.02)^2}{5-1}
\end{aligned}$$

$$S_D^2 = 6.12$$

$$\hat{\sigma}_X = \frac{S_X}{\sqrt{n}} = \sqrt{\frac{6.12}{5}} = 1.10$$

$$T = \frac{\bar{D}}{S_X} \sqrt{n} \rightsquigarrow \text{Student à } (n-1) \text{ degrés de liberté, } |t| = \frac{0.02}{1.10} = 0.11.$$

$$W = \{\underline{x} \in \mathbb{R}^n, |t| > k\}$$

$k = 2.77 \Rightarrow$ on accepte $H_0 \Rightarrow$ le traitement A est aussi efficace que le traitement B au niveau $\alpha = 0.05$.

4.10 Exercices

Exercice 4.1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une v.a. X de loi $\mathcal{N}(0, \sigma^2)$, σ^2 étant inconnue. Soit le test $H_0 : \sigma^2 = 1$ contre $H_1 : \sigma^2 = 4$

Déterminer la région critique du test de Neyman-Pearson et calculer sa puissance dans le cas où $n = 15$ et $\alpha = 0.05$

Exercice 4.2. Soit $(X_1, X_2, \dots, X_{15})$ un n -échantillon issu de X de loi $\mathcal{N}(0, 1/\theta)$.

Soit le test $H_0 : \theta = 1$ contre $H_1 : \theta > 1$

Déterminer la région critique d'un test UPP de risque de 1^{ère} espèce α et préciser sa fonction puissance. Calculer cette puissance pour $\theta = 3$ et $\alpha = 0.05$.

Exercice 4.3. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de X de densité

$$f(x; \theta) = \begin{cases} \frac{1}{2\theta\sqrt{x}} e^{-\frac{\sqrt{x}}{\theta}}, & \text{si } x > 0; \\ 0, & \text{sinon.} \end{cases}$$

On considère le test $H_0 : \theta \leq 1$ contre $H_1 : \theta > 1$

Déterminer la région critique du test le plus puissant de niveau $\alpha = 0.01$ dans le cas où $n = 15$ et préciser sa fonction puissance.

Exercice 4.4. On admet que la durée de vie d'un matériel est représentée par une v.a. X absolument continue de loi $\mathcal{E}(1/\theta)$, θ étant un paramètre inconnu strictement positif. Dans toute la suite, on dispose d'un échantillon (X_1, X_2, \dots, X_n) issu de X .

1. Montrer que la v.a. $Y = \frac{2X}{\theta}$ suit une loi de χ_2^2 . En déduire que $\frac{2}{\theta} \sum_{i=1}^n X_i$ suit la loi de χ_{2n}^2 .
2. Déterminer l'E.M.L $\hat{\theta}_n$ de θ . Montrer qu'il est sans biais et convergent. Est-il efficace ? Quelle estimation de θ proposeriez-vous après avoir observé $\sum_{i=1}^{10} x_i = 11.5$?
3. Construire pour $n = 10$ un intervalle de confiance pour θ au niveau 0.95. En notant L la v.a. longueur de cet intervalle, calculer, en fonction de θ , $\mathbb{E}(L)$ et $\mathbb{V}(L)$. Quelle estimation par intervalle de θ proposeriez-vous après avoir observé $\sum_{i=1}^{10} x_i = 11.5$?
4. On considère le test $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$, θ_0, θ_1 étant des valeurs fixées vérifiant $\theta_1 > \theta_0$.
 - a) Déterminer la région critique optimale de N-P.
 - b) Déterminer le risque d'erreur de 2^{ème} espèce et la puissance du test.
 - c) Préciser comment varient la région critique, le risque d'erreur de 2^{ème} espèce et la puissance du test lorsque le niveau α choisi diminue.
 - d) Quelle décision prendriez-vous, pour $\alpha = 0.05$ et $n = 10$, après avoir observé $\sum_{i=1}^{10} x_i = 11.5$?
5. On considère maintenant le test $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$
 - a) Déterminer un test U.P.P. pour un niveau de signification α fixé
 - b) Déterminer la fonction puissance de ce test. En déduire le tableau de variation et le graphe de cette fonction.
 - c) Quelle décision prendriez-vous, pour $\alpha = 0.05$ et $n = 10$, après avoir observé $\sum_{i=1}^{10} x_i = 11.5$?

Bibliographie

- [1] J. Bass, « *Eléments de calcul de probabilités.* » Masson, 1974.
- [2] G. Calot, « *Cours de calcul des probabilités.* » Dunod, 1967.
- [3] D. Foudrinier, « *Statistique inférentielle. Cours et exercices* ». Dunod, Paris 2002.
- [4] J. Guégand, J. P. Gavini, « *Probabilités.* » 1998.
- [5] A. Krief, S. Levy, « *Calcul des probabilités.* » Hermann, 1972.
- [6] J.P. Lecoutre, S. Legait, P. Tassi, « *Statistique. Exercices corrigés et rappels de cours.)* » Masson, 1987.
- [7] J.P. Lecoutre, « *Statistique et probabilité, manuel et exercices corrigés.* » quatrième édition. Masson, 2009.
- [8] M. Sheldon, M. Ross, « *Initiation aux probabilités.* » Presses polytechniques et universitaires normandes, 1994.
- [9] G. Saporta, « *Probabilité, analyse des données et statistique.* » Editions Technip, 1990.
- [10] P. Tassi, « *Méthodes statistiques.* » Edition Economica, 2004.