

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider, Biskra
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département de Mathématiques



Mémoire présenté pour obtenir le diplôme de
Master en “**Mathématiques**”

Option : Probabilités et Statistiques

Par Mr. Ara Selsabil

Titre :

Estimation par la méthode du maximum de vraisemblance

Devant le Jury :

Dr. Abdelli Jihane	UMKB	Présidente
Dr. Tour Madiha	UMKB	Encadrante
Dr. Berkane Hassiba	UMKB	Examinatrice

Soutenu Publiquement le 02/06/2025

Dédicace

Je dédie ce travail

A moi-même, grâce à Allah d'abord, puis grâce à mes efforts, j'ai terminé ce
mémoire
à celle qui a planté dans mon coeur l'amour du savoir et la détermination, à ma
mère "Diya"
à celui qui m'a appris à vivre pour un noble objectif, à mon cher père "m'hamed"
à ma colonne vertébrale et à la sécurité de mes jours, mes soeurs "Leila, Sana,
Asma"
à mon trésor et ceux qui me donnent le soutien dans la vie, mes frères "Samir,
Koiydar, Mohamed cherif"
à ma chère grand-mère, je te souhaite une bonne santé, à toute ma belle famille,
je vous aime
à mes amies des moments et des jours "Khouloud, Sabrine, Hadjira, Zeineb,
Sara, Chahinez, Selsabil"
à mes enseignants et à tous ceux qui ont supervisé mon éducation depuis mon
enfance jusqu'à aujourd'hui, qu'Allah vous accorde le paradis.

Remerciements

Je tiens à remercier ma encadrante Melle **Tour Madiha**
pour tous ces efforts tout au long de la réalisation de mon travail,
et un grand remerciement pour professeur **Abdelli Jihane** et **Berkane Hassiba**
d'avoir acceptée de juger mon sujet de fin d'étude.

Notations et symbols

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

Notation	Signification
θ	paramètre inconnue
Θ	ensemble des valeurs possible du paramètre θ
$(X_1, X_2, \dots, X_n)$	un n – échantillon
$\mathbb{R}$	ensemble des valeurs réelles
$n!$	$n \times (n - 1) \times (n - 2) \times \dots \times 3 \times 2 \times 1$
EMV	estimateur du maximum de vraisemblance
A	évènement
$\Gamma(\alpha)$	fonction gamma $\int_0^\infty y^{\alpha-1} e^{-y} dy$
$P(A)$	Probabilité de l'évènement A
$F_n(\cdot)$	Focction de répartition empirique
v.a.r	variable aléatoire réelle
t_n	réalisation d'un estimateur T_n
iid	indépendant et identiquement distribuées
P_θ	loi de l'échantillon, quand la valeur du paramètre est θ
T_n	un estimateur
F	fonction de répartition de loi

Notation	Signification
$(\Omega, \mathcal{F})$	espace probabilisable
$(\Omega, \mathcal{F}, P)$	espace de probabilité
X	variable aléatoire
f	densité de probabilité
$L_\theta(x_1, \dots, x_n)$	fonction de vraisemblance de l'observation dépend de paramètre θ
$\ln L_\theta(x_1, \dots, x_n)$	fonction log-vraisemblance du $L_\theta(x_1, \dots, x_n)$
T	statistique
$E(X)$ ou μ	espérance mathématique ou moyenne du v.a X
$Var(X)$ ou σ^2	variance du v.a X
$\bar{X}$	estimateur de la moyenne
S^2	estimateur de la variance
S^*	estimateur sans biais de la variance
$\hat{\theta}_{MV}$	estimateur au sens de maximum de vraisemblance du paramètre θ
$\mathcal{P}(\lambda)$	loi de Poisson à un paramètre $\lambda > 0$
$\mathcal{N}(\mu, \sigma^2)$	loi normale(ou de Gauss) à deux paramètre $\mu \in \mathbb{R}$ et $\sigma^2 > 0$
$\mathcal{N}(0, 1)$	loi normale standard (centrée réduite)
$\Gamma(\alpha, \lambda)$	loi gamma
$\mathcal{B}(p)$	loi de Bernoulli
$\mathcal{B}(n, p)$	loi Binomiale
$\mathcal{P}(\lambda)$	loi de Poisson
$\mathcal{U}[0, n]$	loi Uniforme
$\exp(\lambda)$	loi exponentielle
$\xrightarrow{P}$	convergence en probabilité
$\xrightarrow{L^2}$	convergence en moyenne quadratique
$\xrightarrow{p.s}$	convergence presque sûre

Table des matières

Dédicace	i
Remerciements	ii
Notations et symbols	iii
Table des matières	v
Table des figures	viii
Liste des tableaux	ix
Introduction	1
1 Rappel sur les notions de base en probabilités et statistiques :	4
1.1 Espace de probabilité :	4
1.2 Variables aléatoires	6
1.2.1 Variables aléatoires discrètes :	7
1.2.2 Variables aléatoires continues :	8
1.3 Quelques lois de probabilité usuelles en statistique	9

TABLE DES MATIÈRES

1.3.1	Loi discrètes	9
1.3.2	Lois continues	11
1.4	Type de convergence et théorème limites :	14
1.5	Echantillon :	17
1.6	Distribution d'un échantillon :	18
1.6.1	Fonction de répartition empirique d'un échantillon :	19
1.6.2	Moyenne, variance et moment empiriques :	20
2	Estimation paramétrique ponctuelle	22
2.1	Estimateur et estimation	22
2.2	Qualités d'un estimateur	23
2.3	Information de Fisher	26
2.4	Borne de Fréchet-Darmois-Cramer-Rao(FDCR)	27
2.5	Estimateur efficace	28
2.6	Quelques estimateurs classiques	30
3	Méthode du maximum de vraisemblance pour construction d'un	
	estimateur	32
3.1	Fonction de vraisemblance	33
3.2	Estimation par la méthode du maximum de vraisemblance	33
3.3	Propriétés asymptotiques	34
3.4	Exemples de EMV	35
3.4.1	Loi discrète	35
3.4.2	Loi continue	36

TABLE DES MATIÈRES

3.5 Simulation	39
Conclusion	46
Bibliographie	47

Table des figures

3.1	Estimation du maximum de vraisemblance de paramètre λ .	40
3.2	Estimation du maximum de vraisemblance de paramètre μ .	42
3.3	Estimation du maximum de vraisemblance de paramètre sigma σ^2 .	43
3.4	Comparaison entre la densité théorique $\mathcal{E}(2)$ et celle estimée par	
	EMV pour $n = 50$.	44
3.5	Comparaison entre la densité théorique $\mathcal{E}(2)$ et celle estimée par	
	EMV pour $n = 500$.	45

Liste des tableaux

3.1	Estimation de paramètre lambda par la méthode de EMV et son	
	biais avec taille $n=50,100,500$.	41
3.2	Estimation de paramètre mu et sigma 2 par la méthode de EMV	
	et son biais avec taille $n=50,100,500$.	43

Introduction

L'**estimation paramétrique** en statistique désigne le processus visant à déterminer la valeur d'un ou plusieurs paramètres inconnus dans un modèle statistique à partir des données observées. En termes simples, elle consiste à approximer les paramètres du modèle en exploitant les informations fournies par les échantillons de données. On distingue principalement deux types d'estimation paramétrique :

- **Estimation ponctuelle** : Cette méthode fournit une seule valeur comme meilleure approximation du paramètre inconnu. Par exemple, pour estimer la moyenne d'une population, on peut utiliser la moyenne calculée à partir de l'échantillon comme estimation ponctuelle.
- **Estimation par intervalle** : Contrairement à l'estimation ponctuelle, cette approche exprime l'estimation sous forme d'un intervalle accompagné d'un niveau de confiance. Par exemple, on peut affirmer que la moyenne réelle de la population se trouve entre deux valeurs avec une probabilité de 95%.

Dans le cadre d'estimation ponctuelle, on a plusieurs méthodes pour estimer un paramètre, comme la méthode des moindres carrés, la méthode des moments et la méthode du maximum de vraisemblance. Cette méthode consiste à rechercher les valeurs des paramètres qui maximisent la fonction de vraisemblance, c'est-à-dire celles qui rendent les données observées les plus probables en fonction du modèle.

Cette technique est d'une grande portée dans de nombreux secteurs comme ceux de l'économie, de la médecine, de l'ingénierie ou des sciences sociales dans la mesure où elle permet d'analyser des données et d'extraire les paramètres réels prévus par des modèles complexes qui décrivent des phénomènes réels. Dans le domaine médical, par exemple, l'EMV peut permettre d'estimer les probabilités de maladies ou l'efficacité des différents traitements expérimentés par le biais d'essais cliniques. En économie, il permet également d'estimer des modèles de prévision du comportement des consommateurs ou de fonctionnement marchés financiers. Dans la vie réelle, l'un de ses principaux atouts réside dans l'amélioration des décisions informées à partir des données disponibles, la possibilité de rendre des estimations fidèles qui sont utiles pour prendre des décisions informées dans différents secteurs comme l'amélioration de la qualité dans le domaine industriel ou le traitement des risques dans le domaine des entreprises.

L'objectif de notre travail est d'étudier **l'estimation par la méthode du maximum de vraisemblance** qui constitue un outil précieux pour déterminer les paramètres d'un modèle statistique, en particulier lorsque l'on dispose d'un volume important de données observées.

Ce mémoire est composé de trois chapitres :

- **Premier chapitre** : Dans ce chapitre, **les notions de probabilité et de statistique** sont introduites à travers différents concepts tels que les variables discrètes et continues, les lois de probabilité, les types de convergence, entre autres.
- **Deuxième chapitre** : Ce chapitre aborde le concept d'**estimation paramétrique ponctuelle**, en mettant l'accent sur les notions d'estimateur, les critères évaluant la qualité d'un estimateur.

- **Troisième chapitre** : Dans le dernier chapitre, nous aborderons la fonction de vraisemblance ainsi que l'estimateur du maximum de vraisemblance, accompagnés de quelques exemples illustratifs et une partie de simulation.

Chapitre 1

Rappel sur les notions de base en probabilités et statistiques :

Les statistiques et les probabilités sont des outils essentiels pour comprendre le monde qui nous entoure. En mathématiques, elles permettent de modéliser et d'analyser des données afin de prendre des décisions rationnelles. vous apprendra à collecter, organiser, analyser des données et calculer des probabilités.

Dans ce chapitre, nous mentionnerons les notions de base de probabilités et de statistiques.

1.1 Espace de probabilité :

Une **expérience** est appelée "aléatoire" s'il est impossible de prévoir à l'avance son résultat.

On appelle univers, noté ensemble fondamental Ω , l'ensemble associé à une expé-

CHAPITRE 1. RAPPEL SUR LES NOTIONS DE BASE EN PROBABILITÉS ET STATISTIQUES :

since aléatoire, telle que :

$$\Omega = \{\text{tous les résultats possibles de cette expérience}\}.$$

Définition 1.1.1 (σ -Algèbre) [8] Soit $\mathcal{F}$ un sous-ensemble de parties de l'ensemble Ω ($\mathcal{F} \subset P(\Omega)$), on dit que $\mathcal{F}$ est un σ -**algèbre** (ou **tribu**) si vérifier les conditions suivantes :

1. $\Omega \in \mathcal{F}$.
2. Si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$.
3. Si $(A_i)_{i \geq 0} \subset \mathcal{F}$ alors $\bigcup_{i \geq 0} A_i \in \mathcal{F}$.

Remarque 1.1.1 Si Ω un ensemble des possibilités d'une expérience aléatoire. Alors $(\Omega, \mathcal{F})$ dit espace probabilisable.

Définition 1.1.2 (Espace de probabilité) Soit P une application de $\mathcal{F}$ dans $[0, 1]$. P est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{F})$ si :

1. $P(A) \in [0, 1], \forall A \in \mathcal{F}$.
2. $P(\Omega) = 1$.
3. Si $A_1, A_2, \dots, A_n \in \mathcal{F}$ tels que $A_i \cap A_j = \emptyset$, si $i \neq j$ (i.e : les évènements A_i et A_j deux à deux disjoints) alors :

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j)$$

Le triplet $(\Omega, \mathcal{F}, P)$ est appelé **espace de probabilité** ou **espace probabilisé**.

Pour plus de détails sur espace de probabilité voir [16].

1.2 Variables aléatoires

Variable aléatoires réelle :

Définition 1.2.1 *Etant donné un univers Ω , une **variable aléatoire réelle**, est une **application** de Ω dans $\mathbb{R}$ [12] :*

$$X : \omega \in \Omega \rightarrow X(\omega) \in \mathbb{R} \quad (1.1)$$

Loi de probabilité :

Définition 1.2.2 *Soit Ω un univers muni d'une probabilité P , et soit X une v.a.r. On appelle **loi de probabilité** de X , notée P_X , l'application qui à toute partie A de $\mathbb{R}$ associe :*

$$P_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}). \quad (1.2)$$

Fonction de probabilité :

Définition 1.2.3 *La fonction de répartition de la v.a.r. X est définie par :*

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}. \quad (1.3)$$

propriétés de la fonction de répartition :

1. $F(x)$ est une fonction en escalier.
2. $0 \leq F(x) \leq 1$.
3. Elle est croissante au sens large si $a \leq b$ alors $F(a) \leq F(b)$.

4. L'importance pratique de la fonction de répartition est qu'elle permet de calculer la probabilité de tout intervalle dans $\mathbb{R}$.
5. $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$

1.2.1 Variables aléatoires discrètes :

Définition 1.2.4 Une variable aléatoire réelle discrète X prend ses valeurs sur un ensemble fini $x_1, x_2, \dots, x_n$ ou sur un ensemble infini dénombrable.

Lois de probabilités

Définition 1.2.5 Soit X une v.a.r. discrète prenant ses valeurs dans un ensemble discret (fini ou dénombrable) $\{x_1, x_2, \dots, x_n\}$, éventuellement infini. Alors la loi de X est caractérisée par l'ensemble des probabilités $P(X = x_i)$, c'est-à-dire les nombres réels positifs p_i tels que ;

$$P(X = x_i) = p_i \quad \text{avec : } 0 \leq p_i \leq 1 \quad \text{et} \quad \sum_{i=1}^n p_i = 1$$

Remarque 1.2.1

1. L'espérance mathématique (moment d'ordre 1) de la v.a.r. discrète X , notée $E(X)$ est définie par :

$$E(X) = \sum_{i=1}^n x_i P(X = x_i).$$

2. Le nombre :

$$Var(X) = E[(X - E(X))^2],$$

lorsqu'il existe, est appelé variance (moment d'ordre 2) de X et le nombre :

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

est l'écart type de X .

1.2.2 Variables aléatoires continues :

Définition 1.2.6 Une *variable aléatoire continue* X est une variable qui prend ses valeurs dans un intervalle de $\mathbb{R}$, [2].

Fonction densité de probabilité :

Définition 1.2.7 On appelle densité de probabilité toute application continue :

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\rightarrow f(x) \end{aligned}$$

telle que ;

$$\begin{aligned} \forall x \in \mathbb{R} : f(x) &\geq 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} f(x) dx = 1. \\ \text{et} \quad P(X \leq x) &= \int_{-\infty}^x f(t) dt \end{aligned}$$

Fonction de répartition :

Définition 1.2.8 On définit la fonction de répartition de v a X par :

$$\forall t \in \mathbb{R} : F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Proposition 1.2.1 On a l'identité :

$$P(a \leq X \leq b) = F_X(b) - F_X(a), \quad \forall a < b.$$

Remarque 1.2.2

1. L'espérance mathématique (moment d'ordre 1) de la v.a.r continue X , définie sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$, est donnée par l'intégrale, si elle converge :

$$E(X) = \int_{\Omega} X dP = \int_{\Omega} x dP_x dx$$

que l'on peut écrire, si f est la densité de probabilité de X :

$$E(X) = \int_{\mathbb{R}} x f(x) dx.$$

2. La variance (moment d'ordre 2), ou carré de l'écart-type σ , est donnée par l'intégrale si cette intégrale et la densité f existent :

$$Var(X) = E[(X - E(X))^2] = \int_{\mathbb{R}} [x - E(X)]^2 f(x) dx.$$

1.3 Quelques lois de probabilité usuelles en statistique

[17]

1.3.1 Loi discrètes

Loi de Bernoulli :

Une v.a.r. X est dite suite une loi de Bernoulli de paramètre $p \in]0, 1[$, notée $X \sim \mathcal{B}(p)$, l'orsque l'ensemble de résultat se réduit à deux résultats possibles, et

CHAPITRE 1. RAPPEL SUR LES NOTIONS DE BASE EN PROBABILITÉS ET STATISTIQUES :

on écrit $\Omega = \{\text{Succès}, \text{Echec}\}$, $X(\Omega) = \{0, 1\}$. Loi de probabilité de X :

$$P(X = x) = \begin{cases} p & \text{si } x = 1 \\ q = 1 - p & \text{si } x = 0 \end{cases}, \quad (1.4)$$

son espérance et sa variance sont données par :

$$E(X) = p \text{ et } Var(X) = p(1 - p)$$

Loi Binomiale :

Définition 1.3.1 : Soient n variables aléatoires X_i , ($i = 1 \dots n$) de Bernoulli indépendantes et de même paramètre p alors $X = \sum_{i=1}^n X_i$ est une variable aléatoire qui suit une loi binomiale de paramètre n et p . On note $X \sim \mathcal{B}(n, p)$, la loi de X s'écrit comme suit :

$$P(X = x) = \begin{cases} C_n^x p^x (1 - p)^{n-x} & \text{si } x = 0, 1, \dots, n \\ 0 & \text{sinon} \end{cases}, \quad (1.5)$$

son espérance et sa variance sont données par :

$$E(X) = np \text{ et } Var(X) = np(1 - p)$$

Loi de Poisson :

La loi de Poisson caractérise des événements rares, est très utilisée pour modéliser le nombre qu'un événement se produit dans une période de temps.

Définition 1.3.2 On dit qu'une variable aléatoire X suit une loi de Poisson de

paramètre réel et positif λ , notée $X \sim \mathcal{P}(\lambda)$ si :

$$P(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{si } x = 0, 1, 2, \dots \\ 0 & \text{sinon} \end{cases} \quad (1.6)$$

son espérance et sa variance sont données par :

$$E(X) = \text{Var}(X) = \lambda.$$

Loi Uniforme :

Définition 1.3.3 On dit que une variable aléatoire X suit une loi uniforme sur l'ensemble $\{1, \dots, n\}$, et on note $X \sim \mathcal{U}(n)$ si on a :

$$P(X = x) = \begin{cases} \frac{1}{n} & \text{si } x = 0, 1, 2, \dots, n \\ 0 & \text{si non} \end{cases}, \quad (1.7)$$

son espérance et sa variance sont données par :

$$E(X) = \frac{n+1}{2} \quad \text{et} \quad V(X) = \frac{n^2-1}{12}$$

1.3.2 Lois continues

Loi Uniforme :

Définition 1.3.4 Une variable aléatoire réelle continue X , suit une loi uniforme

CHAPITRE 1. RAPPEL SUR LES NOTIONS DE BASE EN PROBABILITÉS ET STATISTIQUES :

sur l'intervalle $[a, b]$, si sa loi de probabilité admet une densité f égale à :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}, \quad (1.8)$$

et sa fonction de répartition est :

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x \geq b \end{cases}. \quad (1.9)$$

Sa espérance et sa variance est :

$$E(x) = \frac{b+a}{2} \quad \text{et} \quad \text{Var}(x) = \frac{(b-a)^2}{12}$$

Loi exponentielle :

Définition 1.3.5 Une variable aléatoire réelle positive X suit une loi exponentielle, de paramètre λ positif, si sa densité de probabilité est donnée par :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}, \quad (1.10)$$

et sa fonction de répartition est :

$$F(X) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}.$$

CHAPITRE 1. RAPPEL SUR LES NOTIONS DE BASE EN PROBABILITÉS ET STATISTIQUES :

Sa espérance et sa variance est :

$$E(X) = \frac{1}{\lambda} \quad \text{et} \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

La loi gamma :

Définition 1.3.6 *La loi exponentielle est un cas particulier de la famille des lois gamma. Une variable aléatoire réelle positive X suit une loi gamma $\Gamma(\alpha; \lambda)$, de paramètres positifs α et λ , si sa densité de probabilité est donnée par :*

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}, \quad (1.11)$$

Γ est la fonction gamma définie par l'intégrale :

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx.$$

Sa espérance et sa variance est :

$$E(X) = \frac{\alpha}{\lambda} \quad \text{et} \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}$$

Loi normale :

Définition 1.3.7 *:Une variable aléatoire réelle X suit une loi normale (ou loi gaussienne, loi de Laplace-Gauss) d'espérance μ et d'écart type σ (nombre strictement positif, car il s'agit de la racine carrée de la variance σ^2) si cette variable aléatoire réelle X admet pour densité de probabilité la fonction $f(x)$ définie, pour*

tout nombre réel x , par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1.12)$$

et sa fonction de répartition est :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Sa espérance et sa variance est :

$$E(X) = \mu \quad \text{et} \quad \text{Var}(X) = \sigma^2$$

Remarque 1.3.1 Dans les cas particulier $\mu = 0$ et $\sigma^2 = 1$ la loi normale est la loi normale centrée réduite ou loi normal standard et sa densité est :

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

Pour plus de détails voir [17].

1.4 Type de convergence et théorème limites :

Type de convergence : [1]

1. convergence en loi :

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 1.4.1 *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 .*

$$\lim_{n \rightarrow \infty} F_n = F$$

et on note :

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X \quad (1.13)$$

2. convergence en probabilité :

Définition 1.4.2 *On considère une suite (X_n) de variables aléatoires définie sur Ω , X une autre variable, définie sur Ω . On dit que la suite (X_n) converge en probabilité vers X si :*

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0. \quad (1.14)$$

et on note :

$$X_n \xrightarrow[n \rightarrow \infty]{P} X \quad (1.15)$$

3. convergence en moyenne quadratique :

Définition 1.4.3 *Soit (X_n) Une suite de variables aléatoires de carré intégrable. On dit que (X_n) converge en moyenne quadratique vers une variable aléatoire X si :*

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0.$$

et on note :

$$X_n \xrightarrow[n \rightarrow \infty]{L^2} X \quad (1.16)$$

Propriété 1.4.1

1. La convergence en moyenne quadratique entraîne la convergence en probabilité.
2. Pour les (X_n) sont des variables aléatoires d'espérance et de variance finies, si $E(X_n) \rightarrow \mu$ et $Var(X_n) \rightarrow 0$ alors X_n converge en moyenne quadratique vers μ .
4. **convergence presque sure :**

Définition 1.4.4 Soit (X_n) une suite de variables aléatoires. On dit que cette suite converge presque sûrement vers X si :

$$P[\lim_{n \rightarrow \infty} X_n = X] = P[\{\omega \in \Omega; \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}] = 1$$

et on note :

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} X \quad (1.17)$$

Théorèmes limites :

– **Loi forte des grands nombres :**

Théorème 1.4.1 : Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées admettant une moyenne $m = E[X_1]$, et $S_n = \sum_{i=1}^n X_i$ alors :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} m. \quad (1.18)$$

Pour la démonstration voir [II, Statistique inf].

– **Loi faibles des grands nombres :**

Théorème 1.4.2 : Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées admettant une moyenne $m = E[X_1]$, et $S_n = \sum_{i=1}^n X_i$ alors :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{P} m. \quad (1.19)$$

Preuve. Voir [1, Statistique inf p 54]. ■

– **Théorème centrale limite** :

Théorème 1.4.3 (Théorème centrale limite) Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées admettant une moyenne $\mu = E[X_1]$ et de variance $\text{Var}[X_1] = \sigma^2$. Considérons la moyenne $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, quand $n \rightarrow \infty$ la loi de la variable aléatoire :

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$$

tend vers la loi normale centrée réduite $\mathcal{N}(0, 1)$, ce que l'on peut écrire comme :

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (1.20)$$

Preuve. voir [1, Statistique inf p 58.] ■

1.5 Echantillon :

[3]

Définition 1.5.1 (Population) C'est la totalité des éléments pris en considération, et sur lesquels on désire obtenir des informations.

Définition 1.5.2 (Echantillon) Soit X une variable aléatoire sur un référentiel Ω . Un **échantillon** de X de taille $n \geq 1$ dans population est un n -uplet $(X_1, \dots, X_n)$ de variables aléatoires indépendantes de même loi que X . La loi de X sera appelée loi mère. Une réalisation de cet échantillon est un n -uplet de réels $(x_1, \dots, x_n)$ où $X_i(\omega) = x_i$.

1.6 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 :

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

g est appelée vraisemblance du n -échantillon.

Exemple 1.6.1 1. Soit $(X_1, X_2, \dots, X_n)$ un n -échantillon issu d'une variable aléatoire $X \sim \mathcal{N}(0, 1)$

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

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

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

$$f(x) = \theta^x (1 - \theta)^{1-x}$$

avec $x \in \{0, 1\}$. La vraisemblance de cet échantillon est :

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

1.6.1 Fonction de répartition empirique d'un échantillon :

Définition 1.6.1 On appelle fonction de répartition empirique d'un échantillon $F_n(x)$ 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 $\frac{1}{n}$.

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

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

Théorème 1.6.1 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 : $F_n(x) \xrightarrow[n \rightarrow \infty]{p.s} F(x)$. i.e : $P(x : \lim_{n \rightarrow \infty} F_n(x) = F(x)) = 1$

Preuve. Voir [1, Statistique inf p 70]. ■

Théorème 1.6.2 (Glivenko Contelli) La convergence de F_n vers F est presque

surement uniforme, i.e :

$$D_n = \sup_x |F_n - F| \xrightarrow{p.s} 0. \quad (1.21)$$

1.6.2 Moyenne, variance et moment empiriques :

6

Dans la suite on considère $(X_1, X_2, \dots, X_n)$ un n -échantillon issu d'une variable aléatoire de moyenne μ et de variance σ^2 .

Définition 1.6.2 (Statistique) On appelle *statistique* T toute fonction mesurable des variables aléatoires $X_1, X_2, \dots, X_n$, telle que :

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

Remarque 1.6.1 Une *statistique* T est une variable aléatoire.

– **Moyenne empirique :**

Définition 1.6.3 On appelle *moyenne empirique* ou *moyenne de l'échantillon*, la statistique notée $\bar{X}$ définie par :

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

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 $E(\bar{X}) = \mu$ (on dit que $\bar{X}$ est une statis-

tique sans biais). En effet :

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

$$2. \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\text{En effet : } \text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

$$3. \bar{X} \xrightarrow{p.s} \mu. \text{ (Loi forte des grands nombres).}$$

– **Variance empirique :**

Définition 1.6.4 On appelle *variance empirique* ou *variance de l'échantillon*, la statistique notée $S^2(X)$ définie par :

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

Propriétés :

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

$$2. S^2 \xrightarrow{p.s} \sigma^2.$$

Preuve. Voir [1, Statistique inf chapitre 2 page 73] ■

– **Moment empirique :**

Définition 1.6.5 On appelle *moment d'ordre r* de l'échantillon et on note M_r , r un nombre entier ($r \in \mathbb{N}^*$), la quantité :

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

Chapitre 2

Estimation paramétrique ponctuelle

Soit X une variable aléatoire dont la loi, de forme connue, dépend d'un paramètre θ inconnu, $\theta \in \Theta \subset \mathbb{R}$. On suppose que l'on dispose de n observations $x_1, x_2, \dots, x_n$ qui sont les réalisations de variables aléatoires $X_1, X_2, \dots, X_n$ indépendantes et identiquement distribuées (i.i.d.), de même loi que X . On dit que $(X_1, X_2, \dots, X_n)$ est un n -échantillon de la loi de X .

On cherche à estimer θ à partir des n -observations $(x_1, x_2, \dots, x_n)$. Pour tout $\theta \in \Theta$, θ représente en général une valeur caractéristique de la loi de X telle que son espérance, sa variance, son étendue...

2.1 Estimateur et estimation

Définition 2.1.1 (Estimateur) Soit $(X_1, X_2, \dots, X_n)$ un n -échantillon d'une loi P_θ dépendant d'un paramètre inconnu $\theta \in \Theta \subset \mathbb{R}$. On appelle **estimateur** de θ une variable aléatoire T_n obtenue comme fonction du n -échantillon i.i.d

$(X_1, X_2, \dots, X_n)$; autrement dit : [9]

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

Définition 2.1.2 (Estimation) Soit T_n un estimateur de θ . On appelle **estimation** de θ , la réalisation t_n de la v.a. T_n , obtenue à partir de la réalisation $(x_1, x_2, \dots, x_n)$ du n -échantillon $(X_1, X_2, \dots, X_n)$:

$$t_n = f(x_1, x_2, \dots, x_n)$$

2.2 Qualités d'un estimateur

Soit θ le paramètre à estimer et on note $\hat{\theta}_n = T_n = f(X_1, X_2, \dots, X_n)$ un estimateur. [5].

1. Estimateur sans biais :

Définition 2.2.1 Le biais d'un estimateur $\hat{\theta}_n$ de θ est la quantité :

$$b_\theta(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta. \quad (2.1)$$

Un estimateur $\hat{\theta}_n$ est **sans biais** pour θ si et seulement si :

$$b_\theta(\hat{\theta}_n) = 0, \forall \theta \in \Theta.$$

2. Estimateur asymptotiquement sans biais :

$\hat{\theta}_n$ est dit asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta.$$

3. Estimateur convergente :

Définition 2.2.2 Un estimateur $\hat{\theta}_n$ est dit **convergent** si $\hat{\theta}_n$ tend vers θ lorsque n tend vers l'infini.

Définition 2.2.3 (Consistant) On dit que $\hat{\theta}_n$ est un **estimateur consistant** de θ si pour tout $\varepsilon > 0$ on a :

$$P_\theta(|\hat{\theta}_n - \theta| \geq \varepsilon) \rightarrow 0, n \rightarrow \infty. \quad (2.2)$$

Théorème 2.2.1 Si $\hat{\theta}_n$ est convergent et de variance tendant vers 0 lorsque n tend vers l'infini ($\text{Var}(\hat{\theta}_n) = 0, n \rightarrow \infty$), alors $\hat{\theta}_n$ est consistant.

Preuve. On a, pour tous réels θ et $\alpha > 0$.

$$|\hat{\theta}_n - \theta| > \alpha \Rightarrow |\hat{\theta}_n - E(\hat{\theta}_n)| > \alpha - |\theta - E(\hat{\theta}_n)|.$$

Si $\lim E(\hat{\theta}_n) = \theta$, alors à partir de certain rang N , on a $|\theta - E(\hat{\theta}_n)| < \frac{\alpha}{2}$. Ainsi

$$\begin{aligned} P(|\hat{\theta}_n - \theta| > \alpha) &\leq P(|\hat{\theta}_n - E(\hat{\theta}_n)| > \alpha - |\theta - E(\hat{\theta}_n)|) \\ &\leq P(|\hat{\theta}_n - E(\hat{\theta}_n)| > \frac{\alpha}{2}) \\ &\leq \frac{4}{\alpha^2} \text{Var}(\hat{\theta}_n). (\text{par Bienaymé-Chebichev}) \end{aligned}$$

borne supérieure qui tend vers 0 lorsque n tend vers l'infini. De façon générale, on peut écrire :

$$\hat{\theta}_n - \theta = (\hat{\theta}_n - E(\hat{\theta}_n)) + (E(\hat{\theta}_n) - \theta).$$

la grandeur $\hat{\theta}_n - E(\hat{\theta}_n)$ représente les fluctuations de $\hat{\theta}_n$ autour de sa moyenne et $E(\hat{\theta}_n) - \theta$ représente l'erreur systématique (biais). ■

4. Erreur quadratique moyenne :

La qualité d'un estimateur se mesure également par l'erreur quadratique moyenne.

Définition 2.2.4 Soit $\hat{\theta}_n$ un estimateur du paramètre θ . On appelle **risque quadratique**, ou **erreur quadratique moyenne** de $\hat{\theta}_n$ (comme estimateur de θ) la quantité :

$$EQM(\hat{\theta}_n) = E_\theta[(\hat{\theta}_n - \theta)^2]. \quad (2.3)$$

Théorème 2.2.2 Soit $\hat{\theta}_n$ un estimateur du paramètre θ à étudier on a :

$$E[(\hat{\theta}_n - \theta)^2] = Var(\hat{\theta}_n) + [E(\hat{\theta}_n) - \theta]^2. \quad (2.4)$$

où $Var(\hat{\theta}_n)$ est la variance de $\hat{\theta}_n$ et $[E(\hat{\theta}_n) - \theta]$ est le biais de $\hat{\theta}_n$ pour estimer θ .

$$\begin{aligned} E[(\hat{\theta}_n - \theta)^2] &= E[(\hat{\theta}_n - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \theta)^2] \\ &= E[(\hat{\theta}_n - E(\hat{\theta}_n))^2] + E[(E(\hat{\theta}_n) - \theta)^2] + 2E[(\hat{\theta}_n - E(\hat{\theta}_n)][E(\hat{\theta}_n) - \theta]] \\ &= Var(\hat{\theta}_n) + [E(\hat{\theta}_n) - \theta]^2. \end{aligned}$$

car $E(\hat{\theta}_n - E(\hat{\theta}_n)) = 0$.

5. Meilleur estimateur

Soient $\hat{\theta}_1, \hat{\theta}_2$ deux estimateurs de θ . On dit que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ si :

$$E[(\hat{\theta}_1 - \theta)^2] < E[(\hat{\theta}_2 - \theta)^2], \forall \theta.$$

Remarque 2.2.1

1. Entre deux estimateurs sans biais, le “meilleur” sera celui dont la variance est minimale (on parle d’**efficacité**).
2. Pour un estimateur sans biais, le risque quadratique est égal à la variance de l’estimateur.
3. Lorsque le **risque quadratique** $EQM(\hat{\theta}_n)$ tend vers 0 quand n tend vers l’infini, on dit que $\hat{\theta}_n$ converge en **moyenne quadratique**.
4. Si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs de θ (avec ou sans biais), on choisira celui qui a le plus petit risque quadratique. Si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont sans biais, choisir l’estimateur ayant le plus petit risque quadratique revient bien sûr à choisir celui de plus petite variance.
5. La convergence en moyenne quadratique implique la consistance.

2.3 Information de Fisher



Définition 2.3.1 On appelle **quantité d’information de Fisher** I_n apportée par un échantillon sur le paramètre θ , la quantité suivante positive ou nulle (si elle existe) :

$$I_n(\theta) = E \left[\left(\frac{\partial \ln L(X, \theta)}{\partial \theta} \right)^2 \right], \quad (2.5)$$

avec

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

Théorème 2.3.1 *Si le domaine de définition de X ne dépend pas de θ alors :*

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

Preuve. Voir [1, Statistique inf p 83]. ■

Remarque 2.3.1 *Si le domaine de définition de X ne dépend pas de θ alors :*

$$I_n(\theta) = nI_1(\theta),$$

avec

$$I_1(\theta) = -E \left[\left(\frac{\partial^2 \ln f(X, \theta)}{\partial \theta^2} \right) \right].$$

2.4 Borne de Fréchet-Darmois-Cramer-Rao(FDCR)

Théorème 2.4.1 (Inégalité de FDCR) *Si le domaine de définition de X ne dépend pas de θ , alors pour tout estimateur sans biais :*

$$\text{Var}(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)}. \quad (2.6)$$

Si $\hat{\theta}$ est un estimateur sans biais de $g(\theta)$:

$$\text{Var}(\hat{\theta}_n) \geq \frac{[g'(\theta)]^2}{I_n(\theta)}.$$

La borne $\frac{[g'(\theta)]^2}{I_n(\theta)}$ est appelée borne de **Cramer-Rao (FDCR)**.

Preuve. Voir [1, Statistique inf p 86]. ■

2.5 Estimateur efficace

Définition 2.5.1 Un estimateur $\hat{\theta}_n$ est dit **efficace** si sa variance est égale à la borne **FDCR**.

Propriété 2.5.1 Un estimateur sans biais efficace est convergente. En effet :

$\hat{\theta}_n$ efficace $\Rightarrow \text{Var}(\hat{\theta}_n) = \frac{[g'(\theta)]^2}{I_n(\theta)}$, or $I_n(\theta) = nI_1(\theta)$ et $\hat{\theta}_n$ est sans biais.

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left[\left(\hat{\theta}_n - \theta \right)^2 \right] &= \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) \\ &= \lim_{n \rightarrow \infty} \frac{[g'(\theta)]^2}{I_n(\theta)} \\ &= \lim_{n \rightarrow \infty} \frac{[g'(\theta)]^2}{nI_1(\theta)} = 0, \end{aligned}$$

donc $\hat{\theta}$ convergent.

Remarque 2.5.1 Un estimateur efficace $\hat{\theta}_n$ est un estimateur sans biais de variance minimale.

Exemple 2.5.1 Soit $(X_1, X_2, \dots, X_n)$ un n -échantillon issu de $X \sim \mathcal{N}(\mu, \sigma^2)$,

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

estimateur de μ . Calculer l'information de Fisher et la borne de FDCR. On a :

$$I_1(\mu) = -E \left[\left(\frac{\partial^2 \ln f(X, \mu)}{\partial \mu^2} \right) \right]$$

et

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

alors

$$\ln f(x, \mu) = -\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}(x - \mu)^2.$$

$$\frac{\partial \ln f}{\partial \mu} = \frac{-1}{\sigma^2} \Rightarrow I_n(\mu) = \frac{n}{\sigma^2}.$$

$$E(\bar{X}) = \mu \text{ et } \text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \text{ et on a : } I_n(\mu) = \frac{n}{\sigma^2}.$$

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

Définition 2.5.2 Un estimateur sans biais est **efficace** si sa variance est la plus faible parmi les variances des autres estimateurs sans biais. Ainsi, si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs sans biais du paramètre θ , l'estimateur $\hat{\theta}_1$ est efficace si :

$$V(\hat{\theta}_1) < V(\hat{\theta}_2) \quad \text{et} \quad E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta.$$

2.6 Quelques estimateurs classiques

Etant donné un échantillon $X_1, \dots, X_n$ d'une variable aléatoire X inconnue, on a :

1. Estimation d'une moyenne :

$\bar{X}$ est un **estimateur sans biais** de la moyenne $\mu = E[X]$ est :

$$\bar{X} = \frac{X_1 + \dots + X_n}{n},$$

qui représente la moyenne empirique dans l'échantillon, son estimation $\bar{x}$ est la moyenne observée dans une réalisation de l'échantillon.

2. Estimation d'une variance :

– S^2 est un **estimateur consistant** de σ^2 (mais biaisé), est :

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

– $\tilde{S}^2 = \frac{n}{n-1} S^2$ est un estimateur sans biais et consistant de la variance $\sigma^2 = V(X)$, est appelé la variance empirique :

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

Remarque 2.6.1 Si la moyenne μ de X est connue, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ est un meilleur estimateur de σ^2 que S^2 .

3. Estimation d'une proportion :

Dans le cas particulier, si $(X_1, \dots, X_n)$ un n -échantillon de la loi Bernoulli

$\mathcal{B}(p)$, $p \in [0, 1]$. On estime p par l'estimateur intuitif :

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

qui représente la proportion aléatoire dans l'échantillon. C'est un estimateur sans biais de p .

Chapitre 3

Méthode du maximum de vraisemblance pour construction d'un estimateur

La méthode du maximum de vraisemblance est une méthode d'estimation paramétrique. Le principe de la construction des estimateurs par maximisation de la vraisemblance est intuitivement évident, il s'agit de choisir comme estimateur le paramètre pour lequel l'observation est la plus probable, ou la plus vraisemblance.

3.1 Fonction de vraisemblance

Soit X une variable aléatoire réelle de loi paramétrique (discrète ou continue), on définit une fonction f telle que :

$$f(x; \theta) = \begin{cases} f_{\theta}(x) & \text{si } X \text{ est une v.a continue de densité } f \\ P_{\theta}(X = x) & \text{si } X \text{ est une v.a discrète de probabilité ponctuelle } P \end{cases}.$$

Définition 3.1.1 On appelle **fonction de vraisemblance** de θ de Θ pour une réalisation $(x_1, \dots, x_n)$ d'un $(X_1, \dots, X_n)$ n -échantillon indépendantes et de même loi, la densité noté $L(x_1, \dots, x_n; \theta)$, s'écrit :[\[5\]](#).

$$L(x_1, \dots, x_n; \theta) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta). \quad (3.1)$$

3.2 Estimation par la 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 $\hat{\theta}_{MV}$ qui maximise $L(x_1, \dots, x_n; \theta)$ serait un bon estimateur car elle donne la plus grande probabilité pour cet échantillon.

Définition 3.2.1 On appelle **estimation de maximum de vraisemblance (EMV)** de paramètre θ la statistique $\hat{\theta}_{MV}$ tel que :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \ln L(x_1, \dots, x_n; \theta). \quad (3.2)$$

Cette méthode consiste à résoudre :

$$\left\{ \begin{array}{l} \frac{\partial L(x; \theta)}{\partial \theta} = 0, \\ \left. \frac{\partial^2 L(x; \theta)}{\partial \theta^2} \right|_{\theta = \hat{\theta}_{MV}} < 0, \end{array} \right. \quad \begin{array}{l} \text{pour trouvé } \hat{\theta}_{MV} \\ \text{pour assurer l'existence du } \max_{\theta \in \Theta} L(x, \theta) \end{array} . \quad (3.3)$$

Remarque 3.2.1

1. Quand $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ et que toutes les dérivées partielles ci-dessous existes, $\hat{\theta}_{MV}$ est solution du système d'équations appelées équation de vraisemblance :

$$\forall j \in \{1, \dots, d\}, \frac{\partial}{\partial \theta_j} \ln L(x_1, \dots, x_n; \theta_j) = 0$$

2. Maximiser $L(x; \theta)$ revient à maximiser $\ln L(x; \theta)$. Il est plus commode de maximiser $\ln L(x; \theta)$.

3.3 Propriétés asymptotiques

Proposition 3.3.1 [11] Soit l'échantillon $(X_1, X_2, \dots, X_n)$ issu de la densité $f(x; \theta)$ avec $\theta \in \Theta \subseteq \mathbb{R}$.

On considère la suite $\hat{\theta}_n$ quand $n \rightarrow \infty$. Alors cette suite est telle que :

$$\sqrt{n}(\hat{\theta}_{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{I(\theta)}).$$

Le résultat énoncé dans cette proposition implique les propriétés suivantes :

1. $\hat{\theta}_{MV}$ est asymptotiquement sans biais, $E(\hat{\theta}_{MV}) \xrightarrow{n \rightarrow \infty} \theta$.
2. Pour n tend vers l'infini, la variance de $\hat{\theta}_{MV}$ se rapproche de $\frac{1}{nI(\theta)}$.

On dit que $\hat{\theta}_{MV}$ est asymptotiquement efficace.

3. Des propriétés 1 et 2 on déduit que $\hat{\theta}_{MV}$ converge vers θ en moyenne quadratique.
4. $\hat{\theta}_{MV}$ tend à devenir gaussien quand n s'accroît.

On résume ces propriétés en disant que l'EMV est un estimateur BAN (Best Asymptotically Normal).

3.4 Exemples de EMV

3.4.1 Loi discrète

Soit $(X_1, X_2, \dots, X_n)$ un n -échantillon issu de $X \sim \mathcal{P}(\lambda)$, on souhaite estimer le paramètre λ d'une loi de Poisson.

On a :

$$f(x; \lambda) = P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

La fonction de vraisemblance s'écrit

$$\begin{aligned} L(x_1, \dots, x_n; \lambda) &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-\lambda n} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}. \end{aligned}$$

Il est plus simple d'utiliser le logarithme, la vraisemblance étant positive :

$$\begin{aligned}\ln L(x_1, \dots, x_n; \lambda) &= \ln e^{-\lambda n} + \ln \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \\ &= -\lambda n + \sum_{i=1}^n \ln \frac{\lambda^{x_i}}{x_i!} \\ &= -\lambda n + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!).\end{aligned}$$

La dérivée première est :

$$\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^n x_i}{\lambda}.$$

S'annule pour $\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$.

La dérivée seconde :

$$\frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial^2 \theta} = -\frac{\sum_{i=1}^n x_i}{\lambda^2}$$

est toujours négative ou nulle. Ainsi l'estimation donnée par $\Lambda = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$ conduit à un estimateur du maximum de vraisemblance égal à $\hat{\lambda} = \bar{x}$. Il est normal de retrouver la moyenne empirique qui est le meilleur estimateur possible pour le paramètre λ (qui représente aussi l'espérance d'une loi de Poisson).

3.4.2 Loi continue

Soit $(X_1, X_2, \dots, X_n)$ un n -échantillon issu de $X \sim \mathcal{N}(\mu, \sigma^2)$, trouver l'estimateur EMV de μ et σ^2 .

La loi normale $\mathcal{N}(\mu, \sigma)$ a pour fonction densité :

$$f(x; \mu, \sigma^2) = f_{(\mu, \sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad x > 0.$$

Ecrivons la fonction de vraisemblance pour une réalisation d'un échantillon de n variables indépendantes :

$$\begin{aligned} L(x_1, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n f(x_i; \mu, \sigma^2) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right). \end{aligned}$$

$$\ln L(x_1, \dots, x_n; \mu, \sigma^2) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2.$$

$$\begin{aligned} \frac{\partial \ln L(x_1, \dots, x_n; \mu, \sigma^2)}{\partial \mu} &= \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma^2}\right) = 0 \\ &\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ &\Rightarrow \sum_{i=1}^n (x_i - \mu) = 0 \Rightarrow n\mu = \sum_{i=1}^n x_i \\ &\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i. \end{aligned}$$

et

$$\frac{\partial^2 \ln L(x_1, \dots, x_n; \mu, \sigma^2)}{\partial \mu^2} = \frac{-1}{\sigma^2} < 0$$

d'où : l'EMV de l'espérance μ :

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

Pour le second paramètre, on calcule :

$$\begin{aligned} \frac{\partial \ln L(x_1, \dots, x_n; \mu, \sigma^2)}{\partial \sigma^2} L &= \frac{\partial}{\partial \sigma^2} \left(\frac{-n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} + \frac{2 \sum_{i=1}^n (x_i - \mu)^2}{4(\sigma^2)^2} = 0 \\ &\Rightarrow \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2} = \frac{n}{2\sigma^2} \Rightarrow \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} = n \\ &\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

et

$$\begin{aligned} \frac{\partial^2 \ln L(x_1, \dots, x_n; \mu, \sigma^2)}{\partial (\sigma^2)^2} &= \frac{2n}{4(\sigma^2)^2} - \frac{-4\sigma^2 \sum_{i=1}^n (x_i - \mu)^2}{4(\sigma^2)^4} \\ &= \frac{n}{2(\sigma^2)^2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\sigma^2)^3} \\ &= \frac{1}{(\hat{\sigma}^2)^2} \left[\frac{n}{2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \right]_{\sigma^2 = \hat{\sigma}^2} \\ &= \frac{1}{(\hat{\sigma}^2)^2} \left[\frac{n}{2} - n \right] = \frac{-n}{2(\hat{\sigma}^2)^2} < 0. \end{aligned}$$

$$\text{D'où : } \begin{cases} \hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2, & \text{si } \mu \text{ est connue;} \\ \hat{\sigma}_2^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2, & \text{si } \mu \text{ est inconnu.} \end{cases}.$$

La méthode fournit un estimateur non biaisé de la moyenne ($E(\hat{\mu}) = \mu$) mais par contre, l'estimateur de la variance est biaisé ($E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$). Néanmoins l'estimateur est asymptotiquement sans biais.

3.5 Simulation

Dans cette section, on s'est intéressé à l'estimation de la paramètre inconnu de loi de probabilité spécifié, par la méthode du maximum de vraisemblance et on montrer graphiquement la performance de l'estimateur EMV par convergence de l'estimateur paramétrique de la fonction de densités $f_n(x)$ par la méthode du maximum de vraisemblance vers la vraie fonction de densités $f(x)$, utilisant le logiciel d'analyse statistique R. [15].

1. Cas où loi discrète

Simulation de paramètre λ (lambda) de la loi de Poisson $\mathcal{P}(\lambda)$

On a un ensemble des réalisations $(x_1, x_2, \dots, x_n)$ d'une distribution de Poisson avec un paramètre λ . On suppose que f représente la fonction de densité de probabilité avec la forme :

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

la fonction de vraisemblance s'écrit

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}.$$

La méthode d'estimation du maximum de vraisemblance consiste à choisir pour estimation la valeur du paramètre λ pour laquelle la probabilité est la plus élevée.

Algorithme de simulation

- Générer des données à partir de la distribution de poisson $\mathcal{P}(3)$. Dans le langage de R, nous pouvons générer facilement des nombres aléatoires à partir d'une distribution de probabilité spécifique. Cela est réalisé en utilisant la fonction : `rpois( $n$ , 3)` et on donne le nombre d'observation $n = 100$ de cette simulation.
- Donner l'intervalle de paramètre simulé λ .
- Calculer la log-vraisemblance de tous l'ensemble de données.
- Tracer le graphe des estimations par maximum du vraisemblance de la valeur de λ qui est inconnu.

Résultats de simulation

Estimation de λ par MLE pour la distribution de poisson est : $\hat{\lambda} = 3.039844$.

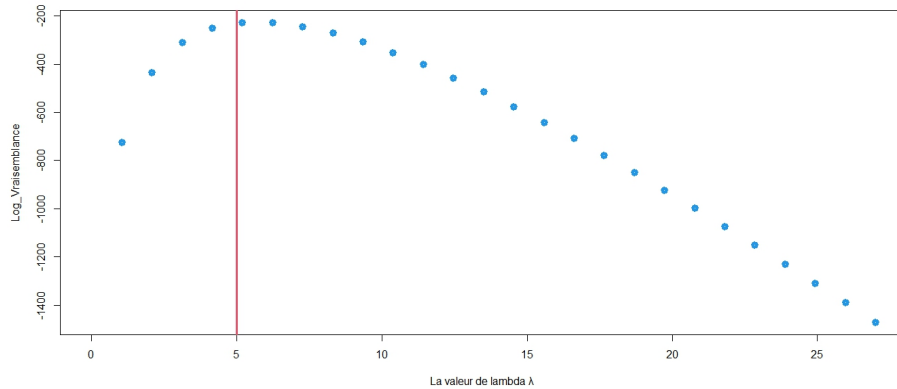


FIG. 3.1 – Estimation du maximum de vraisemblance de paramètre λ .

n	les valeurs de $\hat{\lambda}$ par méthode EMV	biais($\hat{\lambda}$)
50	2.820313	-0.1796875
100	2.929688	-0.0703125
500	3.052344	0.05234375

TAB. 3.1 – Estimation de paramètre lambda par la méthode de EMV et son biais avec taille n=50,100,500.

2. Cas où loi continue

Simulation de paramètre (μ, σ) (mu,sigma) de la loi Normale

On a un ensemble des réalisations $(x_1, x_2, \dots, x_n)$ d'une distribution normale $\mathcal{N}(\mu, \sigma)$.

On suppose que f représente la fonction de densité de probabilité avec la forme :

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad x > 0$$

$L(x; \mu, \sigma)$ est appelé une fonction de vraisemblance pour une réalisation d'un échantillon de *i.i.d* est :

$$L(x_1, \dots, x_n; \mu, \sigma) = \prod_{i=1}^n f(x_i; \mu, \sigma),$$

on applique la méthode d'estimation de maximum du vraisemblance pour estimation la valeur des paramètres (μ, σ^2) pour laquelle la vraisemblance maximale.

Algorithme de simulation

– Génération de données aléatoires suivant une distribution normale $\mathcal{N}((5, 3))$.

Dans le langage de R, nous pouvons utilisant la fonction : `rnorm(n, 5, 3)` et on donner le nombre d'observation $n = 100$ de cette simulation .

– Donner les intervalles des paramètres simulées (μ, σ^2) .

– Calculer la log-vraisemblance de tous l'ensemble de données

– Tracer le graphe des estimations par maximum du vraisemblance de la valeur

de μ et σ qui sont inconnus.

Résultats de simulation

1. Estimation de la moyenne μ par EMV pour la distribution normale est :

$$\hat{\mu} = 5.034556.$$

pour la valeur estimé de paramètre σ^2 (variance) est :

$$\hat{\sigma}^2 = 8.985305.$$

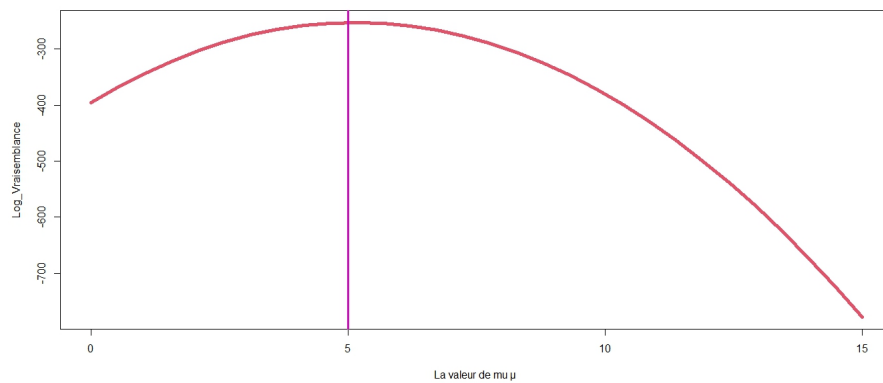


FIG. 3.2 – Estimation du maximum de vraisemblance de paramètre μ .

2. Estimateur de variance σ^2 par MLE pour la distribution normale est :

$$\hat{\sigma}^2 = 8.905478.$$

pour la valeur estimé de paramètre μ est :

$$\hat{\mu} = 4.961819.$$

CHAPITRE 3. MÉTHODE DU MAXIMUM DE VRAISEMBLANCE POUR CONSTRUCTION D'UN ESTIMATEUR

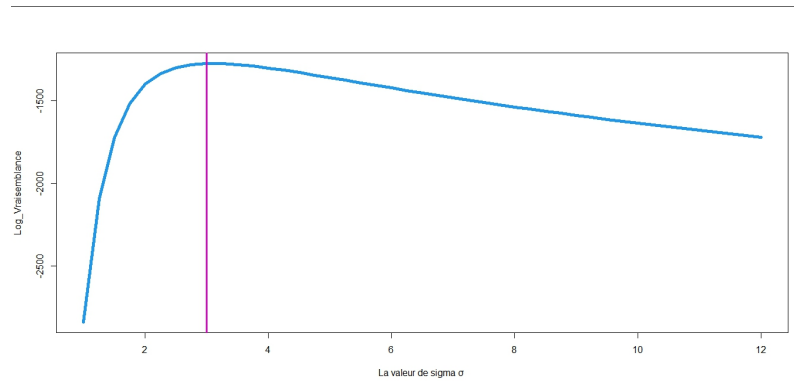


FIG. 3.3 – Estimation du maximum de vraisemblance de paramètre sigma σ^2 .

n	les valeurs de $\hat{\mu}$ (EMV)	biais($\hat{\mu}$)	les valeurs de $\hat{\sigma}^2$ (EMV)	biais($\hat{\sigma}^2$)
50	4.303069	-0.6969305	8.169505	1.3035745
100	4.782310	-0.2176904	8.549007	1.7824158
500	5.041300	0.04130015	9.206625	2.04079925

TAB. 3.2 – Estimation de paramètre mu et sigma 2 par la méthode de EMV et son biais avec taille n=50,100,500.

3. Utilisation du maximum de vraisemblance pour l'estimation paramétrique de la densité de la loi exponentielle de paramètre $\lambda > 0$.

Pour l'estimation de la densité paramétrique nous avons choisi de simuler la densité de probabilité de la loi exponentielle (de paramètre λ) est :

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Algorithme de simulation

L'algorithme de simulation que nous avons utilisé :

- Simuler un échantillon de taille $n = 50$ et $n = 500$, cela est réalisé en utilisant la fonction : `rexp(n, 2)`.
- Construire l'estimateur paramétrique par la méthode du **maximum de vraisemblance** de λ qui est inconnu à partir des observations.
- Tracer les deux graphes : densité test (théorique) et la densité estimée.

Résultats de simulation

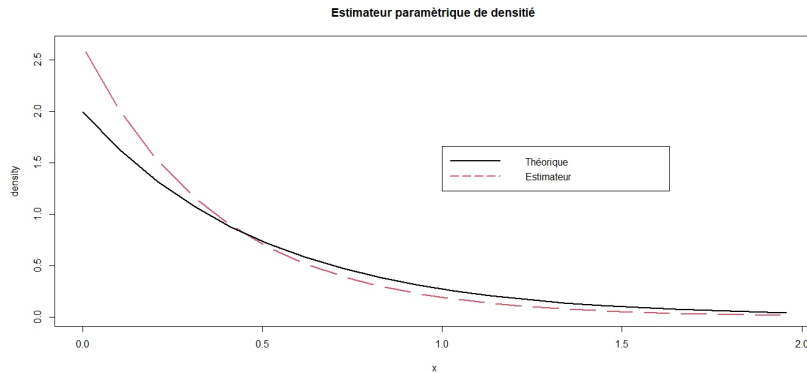


FIG. 3.4 – Comparaison entre la densité théorique $\mathcal{E}(2)$ et celle estimée par EMV pour $n = 50$.

Telle que l'estimation de λ par MLE pour la distribution de exponentielle avec $n = 50$ est : $\hat{\lambda} = 1.712305$.

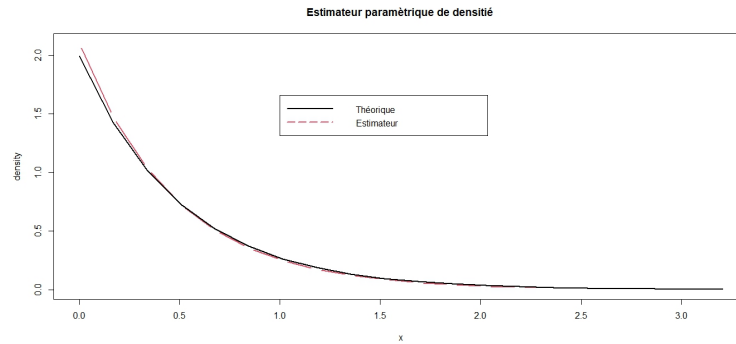


FIG. 3.5 – Comparaison entre la densité théorique $\mathcal{E}(2)$ et celle estimée par EMV pour $n = 500$.

Telle que l'estimation de λ par EMV pour la distribution de exponentielle avec $n = 500$ est : $\hat{\lambda} = 2.165625$.

Remarque 3.5.1 *Nous remarquons sur le graphe ci-dessus que quand n est grand l'estimateur de densité est proche de la densité théorique c'est à dire l'estimation de paramètre de densité par la méthode du maximum de vraisemblance donne de bons résultats.*

Conclusion de simulation : De cette simulation on peut conclure que la méthode du maximum de vraisemblance a donné de bons résultats pour estimé n'importe quel paramètre inconnu relatif à une loi de probabilité et on peut dire que son estimation par EMV est asymptotiquement sans biais.

Conclusion

Dans ce travail, nous nous sommes intéressés à une méthode d'estimation paramétrique, basée sur maximum de vraisemblance. Cette méthode nous a permis de construire des estimateurs performants des paramètres du la lois des probabilités, c'est les plus utilisées, parce que le vraisemblance est une fonction qui contient toute l'information des données sur le paramètre inconnu. L'estimation par maximum de vraisemblance constitue une méthode populaire, puissante, fondamentale, d'une grande utilité en statistique, surtout dans les situations les plus complexes où l'estimation des paramètres est cruciale aux fins de l'analyse et de la prédiction des données sous considération. Mais elle n'est pas sans difficulté surtout du point de vue du calcul de la vraisemblance avec des modèles complexes, tout en s'avérant incontournable pour les statisticiens et les analystes des données.

Bibliographie

- [1] Bouraine, et Berdjoudj, L, (2013/2014). Statistique inférentielle. Université A.MIRA Béjaia.
- [2] Breton,J.C.(2008,Octobre-Novembre). Stat-IUt. Université de la Rochelle.
- [3] Christophe C,Caen. (2018). Cour sur l'estimation du maximum de vraisemblance.
- [4] Diccio, T. Hall, P, & Romano, J. (1991). Empirical likelihood is Bartlett-correctable. The Annals of statistics,19(2), 1053 – 1061.
- [5] DUSART,P.(2018). Cours de statistique inférentielle. Licence 2-S4-SI-MASS.
- [6] Gaudoin, O, Note de cour de deuxième. Principes et méthodes statistiques.
- [7] Hall, P,& La Scala, B. (1990). Methodology and algorithms of empirical likelihood. International Statistical Review/Revue Internationale de statistique,109 – 127.
- [8] Jeanblanc,Monique,Master. (2006). Cours de calcul stochastique.2IF EVRY.
- [9] Jean J.R, (2012/2013). Statistique :estimation, préparation à l'Agrégation Bordeaux 1.
- [10] Lecoutre, P.Statistiques et Probabilités. DUNOD.
- [11] Lejeune, M. (2010). Statistique, la théorie et ses applications. Springer, Paris.

- [12] Magalie, F. Introduction à la théorie des probabilités. Universités de Rennes 1 et Rennes 2 France.
- [13] Mark J, Schervich.(1995). Theory of statistics.Springer,USA.
- [14] Owen,A. (2001). Empirical likelihood. CRC press. International Statistical Review/Revue Internationale de statistique,109 – 127.
- [15] Sébastien , D. Laurent, T. (2016). Encore besoin d' R?. Institut de Mathématiques de Toulouse.
- [16] Veyssaye, R. (2001,2006). Aide mémoire :Statistique et probabilités pour l'ingénieur. L'Usine nouvelle & DUNOD.
- [17] Yadolah, D, Giuseppe M.Springer (2008). Premiers pas en simulation.

Résumé

La méthode de l'estimation du maximum de vraisemblance (EMV) est une méthode statistique dont l'objet est de calculer des estimations des paramètres d'un modèle probabiliste à partir de données observées. Cette méthode consiste à trouver les valeurs de ses paramètres qui maximisent la fonction de vraisemblance, c'est-à-dire la fonction qui rend les données observées les plus vraisemblables. En statistique, et dans le cadre de l'apprentissage statistique, cette méthode est employée pour réaliser des estimations à partir de données mais aussi pour des distributions et des modèles paramétriques connus.

المخلص

تقدير الاحتمال الأقصى هو طريقة إحصائية الغرض منها هو حساب تقديرات بارامترات نموذج احتمالي من البيانات المرصودة, و تتكون هذه الطريقة من إيجاد قيم متغيراتها التي تزيد من دالة الاحتمال إلى أقصى حد أي الدالة التي تجعل البيانات المرصودة هي الأكثر احتمالاً. في علم الإحصاء, و كجزء من التعلم الإحصائي, تستخدم هذه الطريقة لإجراء تقديرات من البيانات, و كذلك للتوزيعات المعروفة و النماذج البارامترية.

Abstract

Maximum likelihood estimation (MLE) is a statistical method whose aim is to calculate estimates of the parameters of a probabilistic model from observed data. This method consists in finding the values of its parameters that maximize the likelihood function, i.e. the function that makes the observed data most likely. In statistics, and as part of statistical learning, this method is used to make estimates from data, but also for known distributions and parametric models.