

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 Appliquées**”

Option : **Statistique**

Par : **AMEUR RAHMOUN**

Titre :

**Qualités d'un estimateur paramétrique et
application**

Devant le Jury :

Dr. Benameur Sana	U. Biskra	Président
Dr. Ouanoughi Yasmina	U. Biskra	Encadreur
Dr. Khereddine Souraya	U. Biskra	Examinatrice

Soutenu Publiquement le **28/06/2022**

Dédicace

Je dédie ce modeste travail :

A Mon très cher Mère Rebiha

A mes chères soeurs : **Zeyneb, Somia, Karima, Khawla**

A mes très chère amis :Mohamed, Aymen, Adel, Addelwahab, Abdelmajid.

Je ne manque pas de mentionner les vertueux professeurs **Mohamed Kouadria**

et **Menacer Chihab Eddine**, qui m'ont aidé à accomplir cette note.

A tous mes enseignants depuis mes premières années d'études.

A tous ceux qui me sens chers et que j'ai omis de citer

Enfin, je ne pourrais manquer de souligner l'apport et le soutien de mes

collègues étudiants à

UNIVERSITÉ MOHAMED KHIDER BISKRA de qui ont rendu cette période qui

de ma vie si

intéressante et si enrichissante.

RAHMOUN AMEUR

Remerciements

Ce travail de mémoire de master est la première expérience dans l'activité de recherche. Il n'aurait pas été aussi fructueux sans l'aide de plusieurs personnes.

Je vais donc m'essayer à trouver les mots justes pour exprimer spécifiquement
ma reconnaissance

à tous ceux qui ont contribué de près ou de loin à ce travail.

Avant tout, je souhaite rendre grâce à **ALLAH**, le Clément et Miséricordieux de
m'avoir

donné la force, le courage et la patience de mener à bien ce modeste travail

Je tiens à remercier sincèrement mon encadreur :

OUANOUGHY YASMINA, de m'avoir suivi et dirigé tout au long de la
réalisation de ce mémoire.

Je souhaite également remercier les membres du jury :

BENAMEUR SANA ⊗ **KHEIREDDINE SOURAYA**

Maitres de conférences à UNIVERSITÉ MOHAMED KHIDER, BISKRA, pour
l'honneur qu'ils

m'ont fait en portant leur attention sur ce travail. Merci à toutes et à tous.

RAHMOUN AMEUR

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Table des figures	vi
Introduction	1
1 Quelques définitions utiles	3
1.1 Modèles statistiques	3
1.2 Modèle d'échantillonnage	5
1.3 Vraisemblance	6
1.4 Familles Exponentielles	7
1.5 Définition d'une statistique	9
1.6 Statistique exhaustive	11
1.6.1 Définition d'une statistique exhaustive	12
1.6.2 Propriétés d'une statistique exhaustive	13

1.6.3	Lois permettant une statistique exhaustive	14
1.7	Information de Fisher	16
1.7.1	Propriétés de la quantité d'information	17
1.7.2	Dégradation de l'information	18
2	Quelques méthodes de calcul d'un estimateur	20
2.1	Définition d'un estimateur	20
2.2	Construction d'estimateurs	21
2.2.1	La méthode des moments	21
2.2.2	La méthode du maximum de vraisemblance	24
2.2.3	La méthode des moindres carrés	29
3	Qualités d'un estimateur	33
3.1	Estimateur convergent	34
3.1.1	Convergence en probabilité	34
3.1.2	Convergence presque sûre	35
3.2	Estimateur sans biais	36
3.3	Estimateur asymptotiquement sans biais	37
3.4	Estimateur avec biais(biaisé)	38
3.5	Risque d'un estimateur (Erreur quadratique moyenne)	39
3.6	Estimateur robuste	40
3.7	Estimateur efficace	41
3.7.1	Inégalité de Fréchet-Darmois-Cramer-Rao(FDCR)	43
3.8	Estimateur absolument correct	43

3.9 Recherche du meilleur estimateur	45
Conclusion	46
Notations et symbols	47
Bibliographie	48

Table des figures

3.1	Comparaison d'estimateur avec $E(T_1) = E(T_2)$ et $Var(T_1) >$ $Var(T_2)$	36
-----	---	----

Introduction

En statistique, un estimateur est une fonction permettant d'évaluer un paramètre inconnu relatif à une loi de probabilité (comme son espérance ou sa variance). Il peut par exemple servir à estimer certaines caractéristiques d'une population totale à partir de données obtenues sur un échantillon comme lors d'un sondage. La définition et l'utilisation de tels estimateurs constitue la statistique inférentielle.

La qualité des estimateurs s'exprime par leur convergence, leur biais, leur efficacité et leur robustesse. Diverses méthodes permettent d'obtenir des estimateurs de qualités différentes

Le premier chapitre du mémoire est consacré à quelques définitions utiles (Modèle statistique, vraisemblance, familles exponentielles), puis, nous étudierons principalement les statistiques exhaustives et la qualité d'information apporté par un échantillon de taille n .

Dans le deuxième chapitre nous présentons quelques méthodes de calcul d'un estimateur comme la méthode des moments, la méthode du maximum de vraisemblance et la méthode des moindres carrés.

La première condition imposée à un estimateur est d'être convergent. Deux estimateurs convergents ne convergent cependant pas nécessairement à la notion de précision d'un estimateur. On mesure généralement la précision d'un estimateur

par l'erreur quadratique. pour rendre l'erreur quadratique moyenne la plus petite possible, il faut choisir un estimateur sans biais de variance minimale, cette propriété traduit l'efficacité de l'estimateur.

Il arrive que lors d'un sondage, une valeur extrême et rare apparaisse. On cherche à ce genre de valeur ne change que de manière très faible la valeur de l'estimateur. On dit alors que l'estimateur est robuste. Nous présenterons tout cela dans le troisième chapitre.

Chapitre 1

Quelques définitions utiles

L'objet de ce chapitre est de présenter le socle sur lequel vont s'appuyer toutes les techniques statistiques présentées dans les chapitres suivants. Ainsi nous présenterons la notion fondamentale de modèle statistique et en donnerons quelques cas particuliers importants que nous retrouverons dans les développements ultérieurs. Nous présenterons aussi une notion très liée à la notion de modèle statistiques : la vraisemblance. Elle est également très importante en statistiques.

1.1 Modèles statistiques

En statistique la loi de probabilité est inconnue. Le but étant justement de la connaître, totalement ou partiellement. La modélisation se fait à l'aide d'un modèle statistique.

Définition 1.1.1 (Modèle statistique) *On appelle modèle statistique un triplet $(E, \mathcal{E}, \mathcal{P})$ où :*

- E : ensemble des observation possibles de l'expérience.

- \mathcal{E} : ensemble des évènements (tribu de parties de \mathcal{E}).
- \mathcal{P} : famille de mesures de probabilités.

Exemple 1.1.1 *Modèle associée à l'observation d'un lancer d'une pièce de monnaie :*

$$(\{\pi, F\}, \mathcal{P}(\{\pi, F\}), \{B(p) / p \in [0, 1]\})$$

Modèle associée à l'observation de la note obtenue en mathématiques par un étudiant de ST choisi au hasard :

$$(R, B_R, \{N(m, \sigma^2) / m \in R, \sigma^2 \in R_+^*\})$$

En statistique on fait rarement une seule observation mais plusieurs, et on notera $(E, \mathcal{E}, \mathcal{P})^{(n)}$ le modèle associé à n observations d'une v.a X dans des conditions d'indépendance.

Exemple 1.1.2

$$(\{\pi, F\}, \mathcal{P}(\{\pi, F\}), \{B(p) / p \in [0, 1]\})^{(10)}$$

Modèle associée à l'observation de 10 lancers d'une pièce de monnaie

$$(R, B_R, \{N(m, \sigma^2) / m \in R, \sigma^2 \in R_+^*\})^{(5)}$$

Modèle associée à l'observation des notes obtenues par 5 étudiants de ST choisis au hasard.

On distingue deux types de modèles statistiques :

Définition 1.1.2 – *Un modèle paramétrique est un modèle où on fait l'hypothèse*

- que la loi de X appartient à une famille bien déterminée de lois, famille indexée par un nombre fini de paramètres.
- Un modèle non paramétrique est un modèle où on ne fait aucune hypothèse sur la loi de X ou une hypothèse très large, par exemple que la loi est absolument continue.
 - modèles statistiques paramétriques :

$$(E, \mathcal{E}, (P_\theta)_{\theta \in \Theta}), (E, \mathcal{E}, (f_\theta)_{\theta \in \Theta}).$$

- modèles statistiques non paramétriques :

$(E, \mathcal{E}, \{P/P \in \mathcal{P}\}) \mathcal{P} \rightarrow$ ensemble de toutes les lois de probabilités

$(E, \mathcal{E}, \{F/F \in \mathcal{F}\}) \mathcal{F} \rightarrow$ ensemble de toutes les fonctions de répartition

1.2 Modèle d'échantillonnage

Pour étudier un phénomène aléatoire, on a souvent intérêt à observer plusieurs réalisations indépendantes de celui-ci. C'est ce que l'on a fait dans l'exemple du premier chapitre. On parle alors d'échantillon ou d'échantillonnage. On appelle n -échantillon de la loi P_θ , la donnée d'un vecteur $X = (X_1, \dots, X_n)$ constitué de n v.a indépendantes et identiquement distribuées (*i.i.d*) de loi P_θ . On appelle modèle d'échantillonnage, le modèle

$$(E^n, \mathcal{E}^{\otimes n}, \mathcal{P}^n = \{P_\theta^{\otimes n} : \theta \in \Theta\}).$$

où $\mathcal{E}^{\otimes n}$ est la tribu produit (engendrée par les pavés) sur E^n et $P_\theta^{\otimes n} = P_\theta \otimes \dots \otimes P_\theta$ est la probabilité produit sur $(E^n, \mathcal{E}^{\otimes n})$ qui est la loi du vecteur $X = (X_1, \dots, X_n)$

Toutes les *v.a* ont la même loi, donc la même valeur de θ . Un échantillon est un vecteur aléatoire. Sa réalisation, fruit de n observations indépendantes du même phénomène, est notée $x = (x_1, \dots, x_n)$. On fera toujours cette distinction entre *v.a* et sa réalisation en utilisant majuscules ou minuscule. Grace à l'indépendance et l'identique distribution, la densité de l'échantillon sous la loi P_θ est alors :

$$x = (x_1, \dots, x_n) \rightarrow \prod_{i=1}^n f_\theta(x_i)$$

pour tous x de E . Si on considère le produit de droite non plus comme une fonction de x mais comme une fonction du paramètre θ , pour un $x = (x_1, \dots, x_n)$ fixé, on parle de vraisemblance.

1.3 Vraisemblance

Dans un modèle statistique paramétrique $(E, \mathcal{E}, \mathcal{P})$, on appelle vraisemblance de l'observation x la fonction

$$\begin{aligned} L(x; \cdot) &: \Theta \rightarrow \mathbb{R}^+ \\ \theta &\rightarrow L(x; \theta) = f_\theta(x) \end{aligned}$$

Bien sur, dans le cas d'un modèle d'échantillonnage, la vraisemblance de l'échantillon observé $x = (x_1, \dots, x_n)$ s'écrit sous la forme

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

C'est donc la loi conjointe du *n-échantillon* évaluée aux valeurs observées et considérée comme fonction du paramètre θ .

1.4 Familles Exponentielles

Un modèle paramétrique important en statistique est celui des familles exponentielles. Il recouvre de nombreux modèles paramétriques classiques : normal, binomial, poisson, gamma etc. Un modèle statistique $(E, \mathcal{E}, \mathcal{P})$ sur un espace des observations E est dit **famille exponentielle générale** s'il existe un entier p , des fonctions η, T, C et h tels que les densités puissent s'écrire, pour tout θ de Θ , sous la forme :

$$f_{\theta}(x) = e^{\langle \eta(\theta), T(x) \rangle} C(\theta) h(x)$$

avec les contraintes que

- T soit une fonction mesurable à valeur dans \mathbb{R}^p
- η soit une fonction à valeur dans \mathbb{R}^p .
- C soit une fonction réelle positive qui ne dépend pas x .
- h soit une fonction borélienne positive qui ne dépend pas de θ .

Le vecteur aléatoire $T(x)$ est appelé **statistique canonique** du modèle. Si la fonction T est l'identité, la famille exponentielle est dite **naturelle**. On parle de **forme canonique** d'une famille exponentielle générale quand les densités de probabilités ont la forme

$$f_{\theta}(x) = e^{\langle \theta, T(x) \rangle} C(\theta) h(x),$$

pour tout θ de Θ , ce qu'il est toujours possible d'obtenir quitte à reparamétriser la famille par $\theta' = \eta(\theta)$. Dans ce cas le paramètre θ de la famille exponentielle est appelé **paramètre canonique**. Revenons sur le modèle de Bernoulli. La densité

s'écrit :

$$\begin{aligned} f_p(x) &= p^x (1-p)^{1-x} = \left(\frac{p}{1-p}\right)^x (1-p) = \exp\left(x \ln\left(\frac{p}{1-p}\right)\right) (1-p), \\ &= \exp(\langle \eta(p), T(x) \rangle) C(p) h(x), \end{aligned}$$

avec

$$\eta(p) = \ln\left(\frac{p}{1-p}\right), \quad T(x) = x, \quad C(p) = (1-p), \quad h(x) = 1.$$

Le modèle de Bernoulli est donc une famille exponentielle naturelle puisque $T = Id$.

De plus, le modèle Bernoulli paramétré en fonction de η

$$(E, \mathcal{E}, \mathcal{P}) = (E, \mathcal{E}, \{B(1, e^\eta / (1 + e^\eta)) : \eta \in \mathbb{R}\}).$$

est sous forme canonique. Modèle échantillonnage construit à partir d'une famille exponentielle générale canonique reste une famille exponentielle générale canonique. En effet si $X = (X_1, \dots, X_n)$ est un échantillon de loi de densité

$$f_\theta(x) = e^{\langle \theta, T(x) \rangle} C(\theta) h(x).$$

alors le vecteur aléatoire X a pour densité

$$f_\theta(x_1, \dots, x_n) = e^{\langle \theta, \sum_{i=1}^n T(x_i) \rangle} C^n(\theta) \prod_{i=1}^n h(x_i).$$

et $\sum_{i=1}^n T(x_i)$ est la statistique canonique du nouveau modèle.

On en déduit l'expression de la vraisemblance pour un échantillon $x = (x_1, \dots, x_n)$ d'une famille exponentielle générale. La vraisemblance pour un échantillon $x =$

(x_1, \dots, x_n) d'une famille exponentielle générale canonique est la fonction :

$$\theta \rightarrow L(x_1, \dots, x_n, \theta) = e^{\langle \theta, \sum_{i=1}^n T(x_i) \rangle} C^n(\theta) \prod_{i=1}^n h(x_i)$$

1.5 Définition d'une statistique

Soit X est une variable aléatoire dont la fonction de répartition $F(x; \theta)$ dépendent du paramètre θ , Une statistique est une fonction mesurable T des variables aléatoires X_i :

$$T(X_1, \dots, X_n)$$

A un échantillon, on peut associer différentes statistiques. La théorie de l'estimation consiste à définir des statistiques particulières, appelées estimateur. Une fois l'échantillon effectivement réalisé, l'estimateur prend une valeur numérique, appelée estimation du paramètre θ . On notera $\hat{\theta}$ l'estimateur du paramètre θ .

Exemple 1.5.1 Soit \bar{X} la statistique :

$$\bar{X} = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

c'est-à-dire la fonction moyenne arithmétique des n observations d'un échantillon. Cette statistique peut être considérée comme un estimateur, a priori raisonnable, de l'espérance mathématique.

$$E(X) = m$$

Caractéristique de la statistique \bar{X} :

1. *L'espérance mathématique de \bar{X} est égale à la moyenne m de la popula-*

tion, et est définie comme suit :

$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n} E(X_1 + \dots + X_n) \\
 &= \frac{1}{n} [E(X_1) + \dots + E(X_n)] \\
 &= \frac{1}{n} [m + \dots + m] \\
 &= \frac{n \cdot m}{n} \\
 &= m
 \end{aligned}$$

2. **La variance de \bar{X}** est égale à la variance σ^2 de la population divisé par taille n l'échantillon, et est définie comme suit :

$$\begin{aligned}
 V(\bar{X}) &= \left[\frac{1}{n} (X_1 + \dots + X_n) \right]^2 \\
 &= \frac{1}{n^2} V[X_1 + \dots + X_n] \\
 &= \frac{1}{n^2} [V(X_1) + \dots + V(X_n)] \\
 &= \frac{1}{n^2} [\sigma^2 + \dots + \sigma^2] \\
 &= \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}
 \end{aligned}$$

La variance empirique S^2 de l'échantillon aléatoire est une statistique, qui

est défini comme

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

est un estimateur de la variance théorique σ^2 .

Caractéristique de la statistique S^2 :

L'espérance mathématique de S^2 est égale à la variance σ^2 de la population. En effet

$$\begin{aligned} (n-1) E(S^2) &= E\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) \\ &= \sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \\ &= nE(X_1^2) - nE(\bar{X}^2) \\ &= n(\sigma^2 + m^2) - n\left(\frac{\sigma^2}{n} + m^2\right) \\ &= (n-1)\sigma^2 \end{aligned}$$

alors

$$E(S^2) = \sigma^2$$

1.6 Statistique exhaustive

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon apporte une certaine information sur ce paramètre (information qui serait diffé-

rente pour un autre paramètre avec le même échantillon). 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 qualifiée d'exhaustive.

1.6.1 Définition d'une statistique exhaustive

La statistique T a donc apporté toute l'information possible sur le paramètre. Une telle statistique est appelée statistique exhaustive ou résumé exhaustif pour le paramètre. Les variables aléatoires $(X_i)_{i=1, \dots, n}$ étant indépendantes, la densité de l'échantillon (X_1, \dots, X_n) est :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{\theta}(x_i) \quad \forall \theta \in \Theta \quad \text{et} \quad \forall (x_1, \dots, x_n) \in \mathbb{R}$$

où x_1, \dots, x_n est la réalisation de l'échantillon (X_1, \dots, X_n)

Elle peut se mettre sous la forme :

$$L(x_1, \dots, x_n; \theta) = g(t; \theta) h(x_1, \dots, x_n; \theta/T = t)$$

Tel que :

- $g(t; \theta)$ est la densité de la statistique T .
- $h(x_1, \dots, x_n; \theta/T = t)$ est la densité conditionnelle de l'échantillon sachant $T = t$.

La statistique T sera dite exhaustive si la densité conditionnelle de X sachant $T(x) = t$ et indépendant du paramètre θ C'est-à-dire :

$$h(x_1, \dots, x_n; \theta/T = t) \text{ ne dépende pas de } \theta.$$

1.6.2 Propriétés d'une statistique exhaustive

La propriété d'exhaustivité pour une statistique est intéressante si elle ne dépend pas de la taille de l'échantillon. Soient T une statistique exhaustive pour le paramètre θ et φ une fonction strictement monotone de T . Alors : La statistique $S = \varphi(T)$ est une statistique exhaustive pour le paramètre θ . Soit X est une variable aléatoire suivant une loi uniforme sur $[0, \theta]$. Elle a pour densité :

$$f(x_1, \dots, x_n; \theta) = \begin{cases} \frac{1}{\theta} & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$$

La statistique :

$$T = \sup_i X_i \quad \text{tel que } i \in [1, n]$$

est un résumé exhaustif de l'échantillon $X = (X_1, \dots, X_n)$ pour le paramètre θ , tel que cette échantillon est *iid*. En effet :

– La fonction de vraisemblance de l'échantillon :

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n f(x_i; \theta) \\ &= \begin{cases} \frac{1}{\theta^n} & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

– Loi de la statistique T est :

La fonction de répartition est :

$$G(t, \theta) = \begin{cases} 0 & \text{si } t < 0 \\ P(T < t) = \left(\frac{t}{\theta}\right)^n & \text{si } 0 \leq t \leq \theta \\ 1 & \text{si } t \geq \theta \end{cases}$$

La densité est :

$$g(t, \theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{si } 0 \leq t \leq \theta \\ 0 & \text{sinon} \end{cases}$$

Donc la vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} = \frac{nt^{n-1}}{\theta^n} \frac{1}{nt^{n-1}}$$

On retrouve la factorisation d'une statistique exhaustive. Le principe de factorisation nous donne donc un moyen de reconnaître si une statistique est exhaustive, mais ne permet pas de la construire ou même de savoir s'il en existe une.

1.6.3 Lois permettant une statistique exhaustive

Le théorème suivant répond à ces deux préoccupations précédentes :

Théorème de Darmois

Soit une variable aléatoire X 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 est que la densité est de la forme exponentielle. Et si de plus l'application :

$$x \longrightarrow \sum_{i=1}^n t(X_i)$$

est bijective et continûment différentiable pour le paramètre θ , alors la statistique T :

$$T = \sum_{i=1}^n t(X_i)$$

est une statistique exhaustive particulière.

Ce théorème est un outil très puissant dans la recherche des statistiques exhaus-

tives et l'on remarque que la plupart des lois usuelles, lois de Poisson, de Gauss, lois gamma sont de la forme exponentielle. Reprenons l'exemple de la loi uniforme de densité

$$f(x; \theta) = \frac{1}{\theta} \quad \text{sur } [0, \theta]$$

La densité de la v.a X sous la forme exponentielle donnée par :

$$f(x; \theta) = \exp(-\ln \theta) \quad \text{ou} \quad \ln f(x; \theta) = -\ln \theta$$

On remarque que $t(x) = 1$. Cependant, on vérifie que la statistique T :

$$T = \sum_{i=1}^n a(X_i) = 1$$

n'est pas une statistique exhaustive pour le paramètre θ . En effet, le domaine de définition de la v.a X dépend de θ .

Autres exemples de statistiques exhaustives

1. loi de Bernoulli de paramètre p inconnu : $T = \sum_{i=1}^n X_i$ est exhaustif pour p
2. loi de Gauss : $N(m; \sigma^2)$
 - (a) si σ est connu : $T = \sum_{i=1}^n X_i$ est exhaustive pour la moyenne m
 - (b) si m est connu : $T = \sum_{i=1}^n (X_i - m)^2$ est exhaustive pour la variance σ .
 - (c) si m et σ sont tous les deux inconnus le couple $(\sum_{i=1}^n X_i, \sum_{i=1}^n (x_i - m)^2)$ ou (\bar{X}, S^2) est exhaustif pour le couple (m, σ) .

1.7 Information de Fisher

On appelle quantité d'infomation de Fisher $I_n(\theta)$ apportée par un n-échantillon sur le paramètre θ la quantité suivante positive ou nulle il est de la forme :

$$I_n(\theta) = E \left[\left(\frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \right)^2 \right] \quad (1.1)$$

Si le domaine de définition de X ne dépend pas de θ et si la vraisemblance est deux fois dérivable, alors :

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

si cette quantité existe. On a $L(x; \theta)$ est une densité de probabilité :

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

En dérivant les deux fois par rapport à θ et en remarquant que :

$$\frac{\partial (L(x, \theta))}{\partial \theta} = L(x, \theta) \frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \quad (1.2)$$

il vient :

$$\int_{\mathbb{R}^n} L(x, \theta) \frac{\partial (\ln(L(x, \theta)))}{\partial \theta} dx = 0$$

ce qui prouve que la variable aléatoire

$$\frac{\partial (\ln(L(x, \theta)))}{\partial \theta}$$

est centrée i.e :

$$E \left[\frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \right] = 0$$

Donc

$$I_n(\theta) = E \left[\left(\frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \right)^2 \right] = \text{Var} \left(\frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \right)$$

Dérivons une deuxième fois :

$$\int_{\mathbb{R}^n} L(x, \theta) \frac{\partial^2 (\ln(L(x, \theta)))}{\partial \theta^2} dx + \int_{\mathbb{R}^n} \frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \frac{\partial (L(x, \theta))}{\partial \theta} dx = 0$$

d'après 1.2

$$\int_{\mathbb{R}^n} L(x, \theta) \frac{\partial^2 (\ln(L(x, \theta)))}{\partial \theta^2} dx + \int_{\mathbb{R}^n} \left[\frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \right]^2 dx = 0$$

$$E \left[\left(\frac{\partial (\ln(L(x, \theta)))}{\partial \theta} \right)^2 \right] + E \left[\left(\frac{\partial^2 (\ln(L(x, \theta)))}{\partial \theta^2} \right) \right] = 0$$

1.7.1 Propriétés de la quantité d'information

Si le domaine de définition ne dépend pas du paramètre θ , on a les propriétés suivantes :

Additivité : Si le domaine de définition ne dépend pas de θ on a :

$$I_n(\theta) = n \cdot I_1(\theta)$$

En effet les opérateurs espérance et dérivée seconde sont linéaires. Ceci veut dire 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.

Précision : Soit X une variable aléatoire de Laplace-Gauss $N(\theta, \sigma^2)$ où σ^2 est connu. On a $I_1(\theta) = \frac{1}{\sigma^2}$ l'information apportée par une observation sur la moyenne est d'autant plus grande que la dispersion est petite.

1.7.2 Dégradation de l'information

Montrons que l'information portée par une statistique est inférieure ou égale à celle apportée par l'échantillon. Soit T de densité $g(t, \theta)$ la statistique que l'on substitue à l'échantillon, on a :

$$L(x; \theta) = g(t; \theta) h(x; \theta/T = t)$$

où $h(x, \theta/t)$ est la densité conditionnelle de l'échantillon. On a donc, en prenant l'espérance des dérivées secondes :

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

le dernier terme est la quantité d'information conditionnelle $I_{n/T}(\theta)$ (ou information supplémentaire), elle est positive ou nulle, donc :

$$I_T(\theta) \leq I_n(\theta)$$

on voit donc que si T est exhaustive $I_n(\theta) = I_T(\theta)$ et que la réciproque est vraie si le domaine de X est indépendant de θ . Soit X suit de v.a loi exponentielle $P(\lambda)$, et dispose d'un échantillon indépendant (x_1, x_2, \dots, x_n) . Calculer $I_n(\lambda)$

$$f(x) = \lambda \exp(-\lambda x)$$

$$L(x, \lambda) = \prod_{i=1}^n \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right)$$

$$\ln(L(x, \lambda)) = n \ln \lambda - \lambda \sum_{i=1}^n x_i$$

$$\frac{\partial^2 (\ln(L(x, \lambda)))}{\partial \lambda^2} = n$$

D'où

$$I_n(\lambda) = n \quad , \quad I(\lambda) = 1$$

Chapitre 2

Quelques méthodes de calcul d'un estimateur

Il existe plusieurs méthodes pour donner attribuer des estimations ponctuelles qui possèdent toutes ou certaines de ces caractéristiques. Dans ce chapitre, nous étudierons certaines des méthodes couramment utilisées en statistique pour obtenir des estimations.

2.1 Définition d'un estimateur

Soit X une variable aléatoire dont la loi de probabilité $f(x; \theta)$ dépend d'un seul paramètre θ .

$$X = (X_1, \dots, X_n)$$

est un échantillon de taille n de cette variable (variable parente). Une statistique est une fonction mesurable $T(X)$ des variables aléatoires X_i . On note, en général, $\hat{\theta}(X)$ ou $\hat{\theta}_n$, ou plus simplement, $\hat{\theta}$ l'estimateur du paramètre θ .

Exemple 2.1.1 *La moyenne arithmétique \bar{X} des n observations est un exemple de statistique :*

$$\bar{X} = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

Un estimateur est une statistique qui a des propriétés bien définies. Une suite T_n de statistiques, $T_n = \varphi(X_n)$, est appelé estimateur du paramètre θ si T_n tend θ quand n tend vers l'infini, la convergence étant une convergence en probabilité presque sûre ou en moyenne quadratique. Une fois l'échantillon effectivement réalisé, l'estimateur prend une valeur numérique, appelée estimation ponctuelle du paramètre θ par la statistique T .

Exemple 2.1.2 *Si l'on cherche à évaluer la taille moyenne des enfants de 10 ans, on peut effectuer un sondage sur un échantillon de la population des enfants de 10 ans (par exemple en s'adressant à des écoles réparties dans plusieurs milieux différents). La taille moyenne calculée sur cet échantillon, appelée moyenne empirique, sera un estimateur de la taille moyenne des enfants de 10 ans*

2.2 Construction d'estimateurs

Trois méthodes sont présentes; la méthode du maximum de vraisemblance, la méthode des moments, et la méthode des moindres carrés ordinaires.

2.2.1 La méthode des moments

La méthode des moments est un outil d'estimation intuitif qui date du début des statistiques. Elle consiste à estimer les paramètres recherchés en égalisant certains moments théoriques (qui dépendent de ces paramètres) avec leurs contreparties empiriques. L'égalisation se justifie par la loi des grands nombres qui implique que

l'on peut "approcher" une espérance mathématique par une moyenne empirique. On est donc amené à résoudre un système d'équations.

Soit X une variable aléatoire continue ayant la fonction de densité $f(x; \theta_1, \dots, \theta_k)$ ou une variable aléatoire discrète ayant la fonction de masse $P(x; \theta_1, \dots, \theta_k)$, cette variable se caractérisant par k paramètres inconnus. Si X_1, \dots, X_n forment un échantillon aléatoire de taille n des valeurs prises par X , on peut définir comme suit les k premiers moments de cet échantillon par rapport à l'origine :

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad k = 1, 2, \dots, n$$

Les k premiers moments de la population par rapport à l'origine se traduisent, quant à eux, par

$$\mu_k = E(X^k) = \begin{cases} \int_{-\infty}^{+\infty} x^k f(x; \theta_1, \dots, \theta_k) dx, k = 1, \dots, n \text{ si } X \text{ est continue} \\ \sum_{i=1}^n x_i^k P(x; \theta_1, \dots, \theta_k), k = 1, \dots, n \text{ si } X \text{ est discrète} \end{cases}$$

Les moments $\{\mu_k\}$ de la population sont en général des fonction des k paramètres inconnus $\{\theta_i\}$. En faisant correspondre les moments de l'échantillon à ceux de la population, on obtient k équation simultanées à k inconnus (les paramètres θ_i). En d'autres termes,

$$\mu_k = m_k \quad k = 1, \dots, n \quad (2.1)$$

La solution de l'équation 2.1 notée $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$, fournit les estimateurs de moment de $\theta_1, \theta_2, \dots, \theta_k$

Soit une variable X de loi $N(\mu, \sigma^2)$, les paramètres μ et σ^2 étant ici inconnus. pour trouver à chacun un estimateur par la méthode des moment, il faut se rappeler

que dans le cas de loi normale.

$$\begin{cases} \mu_1 = E(X) = \mu \\ \mu_2 = E(X^2) = \sigma^2 + \mu^2 \end{cases}$$

Les moments de l'échantillon se traduisent ici par

$$m_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

En recourant à l'équation (2.1), on obtient donc :

$$\begin{aligned} \mu &= \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma^2 + \mu^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 \end{aligned}$$

d'où on arrive aux solution

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

Loi de Gamma : X suit de la loi gamma de paramètre a et λ , son espérance et sa variance valent :

$$\begin{cases} E(X) = a\lambda \\ V(X) = a\lambda^2 \end{cases}$$

On peut donc exprimer a et λ en fonction de $E(X)$ et $V(X)$

$$\begin{cases} E(X) = a\lambda \\ V(X) = a\lambda^2 \end{cases} \implies \begin{cases} a = \frac{E(X)}{\lambda} \\ V(X) = \frac{E(X)}{\lambda} \lambda^2 \end{cases} \implies \begin{cases} a = \frac{E(X)^2}{V(X)} \\ \lambda = \frac{V(X)}{E(X)} \end{cases}$$

Si on dispose d'un échantillon (X_1, X_2, \dots, X_n) de la loi gamma de paramètre a et λ on a \bar{X} et S^2 son d'estimateurs convergents de $E(X)$ et $V(X)$ respectivement. D'après l'expimes. On en déduit deux estimateurs convergents a et λ

$$\begin{cases} \hat{a} = \frac{\bar{X}^2}{S^2} \\ \hat{\lambda} = \frac{\bar{X}}{S^2} \end{cases}$$

Dans certains cas, la méthode des moments n'est pas capable d'atteindre la borne de Cramér-Rao : l'estimation est donc dépassée par l'estimation par maximum de vraisemblance.

2.2.2 La méthode du maximum de vraisemblance

Soit une population X muni d'une distribution de probabilité quelecon que $P(X, \theta)$, qui dépend d'un paramètre θ . Le vrai paramètre θ est inconnu. On tire un échantillon (X_1, \dots, X_n) *iid* de X . Une observation de cet échantillon est notée x_1, \dots, x_n . La question qui se pose est comment estimer le vrai paramètre θ ? L'idée générale de l'estimateur du maximum de vraisemblance consiste à choisir le paramètre qui maximise de vraisemblance de l'échantillon observé. Qu'appelle-t-on une vraisemblance? La fonction vraisemblance notée $L(x_1, \dots, x_n; \theta)$ est la distribution de probabilité d'une population de paramètre θ que génère l'échantillon observé.

$$L(x_1, \dots, x_n; \theta) = P(x_1, \dots, x_n / \theta)$$

Puisque les X_1, \dots, X_n sont *iid* $P(x_1, \dots, x_n / \theta)$ est égale au produit de toutes les probabilité $P(X_i = x_i) = P(x_i = \theta)$ que X_i prend la valeur x_i .

$$L(x_1, \dots, x_n; \theta) = P(x_1, \dots, x_n, \theta) = P(x_1, \theta) \cdot P(x_2, \theta) \dots P(x_n, \theta)$$

L'estimateur du maximum de vraisemblance noté EMV est la valeur du paramètre qui maximise cette fonction de vraisemblance.

$$EMV \left[\hat{\theta} \right] = \theta_{MV} = \hat{\theta}(x_1, \dots, x_n) \text{ maximise } L(x_1, \dots, x_n; \theta)$$

$$L(x_1, \dots, x_n; \hat{\theta}_{MV}) = \max_{\theta} L(x_1, \dots, x_n; \theta)$$

$$L(x_1, \dots, x_n; \theta) = \max_{\theta} P(x_1, \theta) \dots P(x_n, \theta)$$

L'EMV est donc la valeur du paramètre inconnu, dont il est le plus vraisemblable qu'il génère l'échantillon effectivement observé. Si X est une variable aléatoire continue, la fonction de vraisemblance est déterminée à partir de la densité de probabilité $f(x, \theta) = P(x, \theta)$

Soit :

$$L(x_1, \dots, x_n; \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \dots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

On notera que la valeur $\hat{\theta}$ obtenue dépend de l'échantillon. Elle en constitue une estimation particulière.

Exemple 2.2.1 (variable aléatoire de Bernoulli) On suppose que $X \rightsquigarrow B(p)$ (p est le vrai paramètre inconnu) On tire un échantillon X_1, \dots, X_n iid de taille n . Les valeurs observées sont telles que l'on a s succès ($x_i = 1$) et $n - s$ échecs ($x_j = 0$), La vraisemblance de cet échantillon est donnée par :

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

soit

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

p étant la probabilité de succès $p [X_i = 1]$, et $(1 - p)$ est la probabilité d'échec

$$p [X_i = 0]$$

$0 \leq p \leq 1$ d'où

$$L(x_1, \dots, x_n; p) = p^s (1 - p)^{n-s}$$

L'EMV \hat{p} maximise $L(x_1, \dots, x_n; p)$ par rapport à p .

Une condition nécessaire pour obtenir le maximum de $L(x_1, \dots, x_n; p)$ est que la dérivée par rapport à p soit nulle.

puisque

$$\frac{\partial L(x_1, \dots, x_n; p)}{\partial p} = s.p^{s-1} \cdot (1 - p)^{n-s} - p^s \cdot (n - s) \cdot (1 - p)^{n-s-1}$$

Cette condition implique

$$s.\hat{p}^{s-1} \cdot (1 - \hat{p})^{n-s} - \hat{p}^s \cdot (n - s) \cdot (1 - \hat{p})^{n-s-1} = 0$$

En divisant par $p^{s-1} \cdot (1 - p)^{n-s-1}$, on obtient

$$s \cdot (1 - \hat{p}) - \hat{p} \cdot (n - s) = 0$$

et alor

$$s - \hat{p}n = 0 \Rightarrow \hat{p} = \frac{s}{n}$$

ou encore

$$\sum_{i=1}^n x_i - \hat{p}n = 0 \Rightarrow p = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$$

On peut facilement vérifier (dérivée seconde de $L(x_1, \dots, x_n; p)$ par rapport à p

est négative que $p = \bar{X}$ est le maximum unique de $L(x_1, \dots, x_n; p)$. En pratique, le calcul de l'estimateur du maximum de vraisemblance se fait de la manière suivante : partant de la fonction de vraisemblance

$$L(x_1, \dots, x_n; \theta) = p(x_1, \theta) \dots p(x_n, \theta)$$

θ étant toujours le paramètre inconnu. On introduit la fonction logarithme comme suit :

$$\log L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \log p(x_i, \theta)$$

Maximiser $L(\theta)$ est équivalent à maximiser $\log L(\theta)$:

On a $L(\theta_1) \geq L(\theta_2)$ si et seulement si $\log L(\theta_1) \geq \log L(\theta_2) \quad \forall \theta_1, \theta_2$

En conséquence l'EMV $\hat{\theta}_{MV} = \hat{\theta}_{MV}$ est la valeur de θ qui maximise

$$\log L(\theta) = \sum_{i=1}^n \log p(x_i, \theta)$$

La procédure de détermination de ce maximum consiste à :

- Déterminer la dérivée du $\log L(\theta)$ par rapport à θ

$$\frac{\partial \log L(\theta)}{\partial \theta}$$

- Calculer $\hat{\theta}_{MV}$ solution

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0$$

- vérifier que la solution $\hat{\theta}$ représente réellement le maximum de $\log L(\theta)$: La dérivée seconde de $\log L(\theta)$ par rapport à θ doit être négative.

Exemple 2.2.2 (variable aléatoire de normale) On suppose que $X \rightsquigarrow N(m, \sigma^2)$, où la vraie moyenne m et la vraie variance σ^2 sont inconnues on tire un échan-

tillon X_1, \dots, X_n iid de taille n . Les valeurs observées sont x_1, \dots, x_n . On essaye de trouver les estimateurs du maximum de vraisemblance de la moyenne \hat{m}_{MV} et de la variance $\hat{\sigma}_{MV}$. Considérons d'abord l'estimation de m . Puisque X est une v.a continue, la fonction de vraisemblance doit être définie à partir de la densité de probabilité.

Soit

$$L(x_1, \dots, x_n; \sigma^2) = P(x_1, m, \sigma^2) \dots P(x_n, m, \sigma^2)$$

Où

$$P(x_i, m, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - m)^2}{2\sigma^2}\right), \quad i = 1, \dots, n$$

par conséquent, la vraisemblance qu'une population de moyenne m et de variance σ^2 génère l'échantillon observé s'avère être :

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

Ainsi, \hat{m}_{MV} est déterminé par l'équation suivante :

$$\begin{aligned} \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^2} = 0 &\implies \sum_{i=1}^n (x_i - m)^2 = 0 \\ &\implies \hat{m}_{MV} = \sum_{i=1}^n \frac{x_i}{n} = \bar{X} \end{aligned}$$

Considérons maintenant l'EMV de σ^2 .

Soit toujours

$$\begin{aligned} \log L(x_1, \dots, x_n; m, \sigma^2) &= -n \log(\sigma \sqrt{2\pi}) - \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^2} \\ &= -\frac{n}{2} \log(\sigma^2) - \log(\sqrt{2\pi}) - \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^2} \\ \frac{\partial \log L(X; m, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^4} \end{aligned}$$

L'EMV $\hat{\sigma}^2$ est déterminé, ainsi, par l'équation suivante :

$$-\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^4} = 0 \iff \frac{-n\sigma^2 + \sum_{i=1}^n (x_i - m)^2}{2\sigma^2} = \hat{\sigma}_{MV}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$$

Si on remplace \hat{m}_{MV} par sa valeur devient

$$\hat{\sigma}_{MV}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$$

La méthode des moments donne les mêmes estimateurs de μ et de σ^2 que la méthode du maximum de vraisemblance.

2.2.3 La méthode des moindres carrés

Un grand nombre d'études empirique ont pour but d'analyser la façon dont une variable est reliée à d'autres variables. L'examen de ce type de problème est s'appeler l'analyse de la régression. Supposons que l'on dispose d'un échantillon de taille n d'observations de X et Y échantillon :

$$(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$$

Il faut déterminer la relation entre X et Y à partir de cet échantillon. Souvent on se retrouve dans la situation suivante : Y varie régulièrement avec X . La tendance globale est bien décrite par une droite (relation linéaire). En d'autres termes :

$$Y_i = a + bX_i + \epsilon_i \quad (2.2)$$

Les valeurs de a, b et ϵ_i sont inconnues dans l'équation 2.2, mais elles sont fixes, alors que ϵ varie d'une observation à l'autre. Seules les valeurs de a et b sont à estimer. Soient \hat{a} et \hat{b} les estimateurs de a et b respectivement, nous écrivons :

$$\hat{Y} = \hat{a} + \hat{b}X$$

où \hat{Y} est la valeur prédite pour un X donné lorsque \hat{a} et \hat{b} sont déterminés. Le problème revient à trouver les paramètres a et b de la droite $\hat{Y} = \hat{a} + \hat{b}X$ qui "approche le mieux" la dépendance des Y sur les X , c'est-à-dire qui "s'écarte le moins" du nuage de points (Y_i, X_i) . Nous nous basons sur la méthode des moindres carrés pour l'estimation des paramètres, en choisissant les valeurs \hat{a} et \hat{b} telles que la distance entre Y_i et $a + bX_i$ soit minimale. Il faut donc que :

$$\epsilon_i = Y_i - a - bX_i$$

Soit petit pour tout i , Pour ce faire, nous pouvons choisir parmi plusieurs critères comme :

1. $\min_{a,b} (\max(|\epsilon_i|))$
2. $\min_{a,b} (\sum_{i=1}^n |\epsilon_i|)$
3. $\min_{a,b} (\sum_{i=1}^n \epsilon_i^2)$

Mais pour des raisons de commodité, nous allons employer le troisième critère : celui privilégié par la méthode des moindres carrés. Supposons que nous disposons de n observations :

$$(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$$

L'équation liant les Y_i et les X_i est définie par :

$$Y_i = a + bX_i + \epsilon_i \quad i = 1, \dots, n$$

et la somme des carrés des écarts à la droite est :

$$S(a, b) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - a - bX_i)^2 \quad (2.3)$$

Nous devons estimer a et b de telle façon que lorsqu'on substitue les estimations de a et b dans l'équation 2.3, on obtienne la plus petite valeur possible de S . Mathématiquement, on peut déterminer les estimateurs de a et b , notés respectivement \hat{a} et \hat{b} , en prenant les dérivées partielles de l'équation 2.3 d'abord par rapport à a , en suite par rapport à b et ce en les posant égales à zéro. Dérivée partielle par rapport à a :

$$\frac{\partial S(a, b)}{\partial a} = -2 \sum_{i=1}^n (Y_i - a - bX_i) = 0$$

et dérivée partielle par rapport à b :

$$\frac{\partial S(a, b)}{\partial b} = -2 \sum_{i=1}^n (Y_i - a - bX_i) X_i = 0$$

Ainsi les valeurs estimées de \hat{a} et \hat{b} sont données par :

$$\sum_{i=1}^n (Y_i - \hat{a} - \hat{b}X_i) = 0 \quad (2.4)$$

$$\sum_{i=1}^n (Y_i - \hat{a} - \hat{b}X_i) X_i = 0 \quad (2.5)$$

En développant les équations 2.4 et 2.5 nous obtenons :

$$\sum_{i=1}^n Y_i - n\hat{a} - \hat{b} \sum_{i=1}^n X_i = 0 \quad (2.6)$$

$$\sum_{i=1}^n X_i Y_i - \hat{a} \sum_{i=1}^n X_i - \hat{b} \sum_{i=1}^n X_i^2 = 0 \quad (2.7)$$

Ces équations 2.6 et 2.7 sont appelées équation normales. Sachant que :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

nous trouvons pour les équations 2.6 et 2.7 les solutions suivantes :

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

$$\hat{b} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

En substituant l'équation \hat{a} dans l'équation \hat{b} , nous obtenons l'équation de régression estimée :

$$\hat{Y}_i = \bar{Y} + \hat{b}(X_i - \bar{X})$$

Chapitre 3

Qualités d'un estimateur

On a vu plusieurs techniques pour construire des estimateurs. Même si la présentation n'est pas exhaustive, abordons maintenant le problème de l'évaluation de la qualité d'un estimateur et la comparaison d'estimateurs entre-eux. Le but étant bien sûr de prendre le meilleur (s'il en existe un meilleur). On l'a vu, un estimateur $\hat{\theta}$ de θ est une *v.a.* Pour chaque échantillon observé, l'estimateur prendra de nouvelles valeurs. Il faut donc, pour parler de la qualité d'un estimateur, tenir compte de son comportement aléatoire. A priori donc, l'estimateur ne donnera pas toujours (en fait même rarement) la bonne valeur θ . Dans le cas où $\hat{\theta}$ est absolument continu, il sera même *p.s.* toujours différent de la valeur fixe θ . Il est à noter que la présence d'erreur n'est pas toujours la conséquence des variations aléatoires de l'estimateur. Naturellement, on voudra qu'un estimateur possède quelques unes (à défaut de toutes) des qualités suivantes.

- Quand la taille d'échantillon augmente, l'estimateur a tendance à se rapprocher (dans un sens à définir) de la valeur θ qu'il estime. On parlera dans ce cas d'estimateur convergent ou consistant.
- Même si l'estimateur commet une erreur d'estimation à chaque fois, "en moyenne"

- (en fait en espérance) il ne se trompe pas. On dira dans un tel cas que l'estimateur est sans biais.
- L'estimateur doit être le plus précis possible : les variations de l'estimateur autour de θ doivent être réduites, voire les plus petites possible. On mesurera cette précision au moyen de la notion de fonction de risque.

Il y aurait d'autres critères, mais nous n'aurons pas le temps de les étudier.

3.1 Estimateur convergent

3.1.1 Convergence en probabilité

On dit que la suite $(X_n)_{n \geq 1}$ converge en probabilité vers X , et on note $X_n \xrightarrow{p} X$ quand $n \rightarrow \infty$ si :

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

ou

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| \leq \epsilon) = 1$$

On dit que $\hat{\theta}_n$ est un estimateur convergent θ de si $\hat{\theta}_n$ tend vers en probabilité quand n tend vers ∞ c-à-d :

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} P\left[|\hat{\theta}_n - \theta| > \epsilon\right] = 0 \text{ c-à-d } \hat{\theta}_n \xrightarrow{p} \theta \text{ quand } n \rightarrow \infty$$

3.1.2 Convergence presque sûre

On dit que la suite $(X_n)_{n \geq 1}$ converge presque sûrement vers X , et on note $X_n \xrightarrow{p.s.} X$ quand $n \rightarrow \infty$ si :

$$P\left(\forall w : X_n(w) \xrightarrow{n \rightarrow \infty} X(w)\right) = 1 \iff P(\forall w : |\lim X_n(w) - X(w)|) = 1$$

ou

$$\forall \epsilon > 0; \sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty \quad \text{c-à-d la série converge.}$$

On parle d'estimateur fortement convergent lorsqu'on a convergence presque sûre. Soit $\hat{\theta}_n$ un estimateur convergent du paramètre θ , et soit une fonction de \mathbb{R} dans \mathbb{R} , continue au point θ . Alors $(\phi(\hat{\theta}_n))$ est un estimateur convergent de $\phi(\theta)$.

Exemple 3.1.1 Soit $X = (X_1, \dots, X_n)$ est une v.a suit la loi Normale $N(m, \sigma^2)$ et on a $\hat{\theta} = \bar{X}$ est un estimateur de m . D'après la théorème central limite on a :

$$\sqrt{n} \frac{\bar{X} - m}{\sigma} \rightsquigarrow N(0, 1)$$

Donc on calcule :

$$\begin{aligned} P(|\bar{X} - m| > \epsilon) &= 1 - P(|\bar{X} - m| \leq \epsilon) \\ &= 1 - P(-\epsilon \leq \bar{X} - m \leq \epsilon) \\ &= 1 - P\left(\frac{-\epsilon\sqrt{n}}{\sigma} \leq \frac{\bar{X} - m}{\sigma} \leq \frac{\epsilon\sqrt{n}}{\sigma}\right) \\ &= 2 - 2\phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) \end{aligned}$$

où ϕ est la fonction de répartition de la loi normale $N(0, 1)$. Alors, si $n \rightarrow \infty$

on trouve

$$\lim_{n \rightarrow \infty} P(|\bar{X} - m| > \epsilon) = 2 - 2\phi(\infty) = 0$$

Et d'après la définition la statistique $\hat{\theta} = \bar{X}$ est un estimateur consistant de θ .

3.2 Estimateur sans biais

L'erreur d'estimation est mesurée par la quantité $\hat{\theta} - \theta$ qui peut s'écrire :

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

1. $\hat{\theta} - E(\hat{\theta})$ représente les fluctuations de l'estimateur $\hat{\theta}$ autour de sa valeur moyenne $E(\hat{\theta})$ (espérance mathématique).

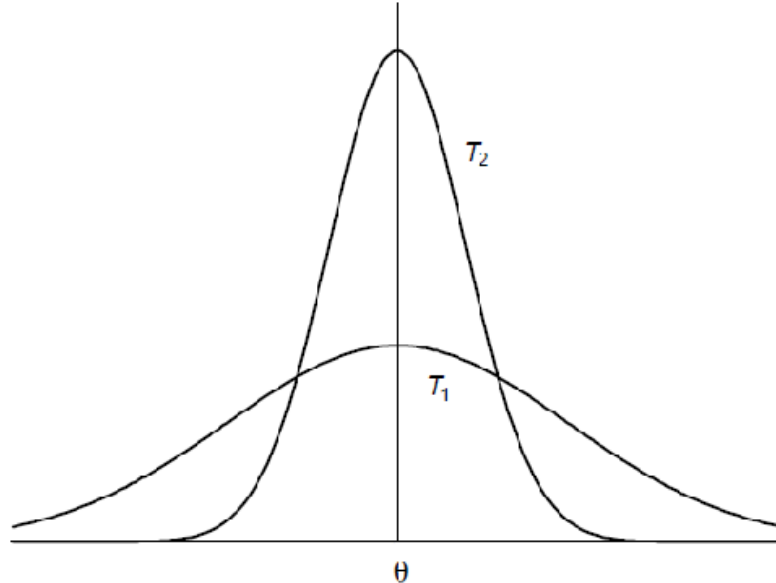


FIG. 3.1 – Comparaison d'estimateur avec $E(T_1) = E(T_2)$ et $Var(T_1) > Var(T_2)$

2. $E(\hat{\theta}) - \theta$ est une erreur systématique car l'estimateur $\hat{\theta}$ varie autour de son espérance mathématique $E(\hat{\theta})$ et non autour de la valeur θ du paramètre

sauf si $E(\hat{\theta}) = \theta$.

3. La quantité $E(\hat{\theta}) - \theta$ est le biais de l'estimateur.

4. Un estimateur est sans biais si $E(\hat{\theta}) = \theta$.

Exemple 3.2.1 (Estimateur de l'espérance mathématique) Soit (X_1, X_2, \dots, X_n)

un n échantillon telles que $E(X_i) = m, \forall i = 1, \dots, n$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ (la moyenne empirique) est un estimateur sans biais de m , en effet :

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &= \frac{1}{n} n E(X_1) \\ &= \frac{1}{n} nm \\ &= m \end{aligned}$$

Alors

$$E(\bar{X}) = m$$

3.3 Estimateur asymptotiquement sans biais

On dit que $\hat{\theta}_n$ est un estimateur asymptotiquement sans biais de θ si :

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

Exemple 3.3.1 Soit (X_1, X_2, \dots, X_n) un n échantillon

telles que $V(X_i) = \sigma^2 < \infty, \forall i = 1, \dots, n$

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

la variance empirique est un estimateur asymptotiquement sans biais de σ^2 ,
en effet :

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{1}{n} \sum_{i=1}^n EX_i^2 - E\bar{X}^2 \\ &= EX^2 - E\bar{X}^2 \\ &= (V(X) + E^2(X)) - (V(\bar{X}) + E^2(\bar{X})) \\ &= \sigma^2 + m^2 - \frac{\sigma^2}{n} - m^2 \\ &= \sigma^2 - \frac{\sigma^2}{n} \end{aligned}$$

Donc

$$\lim_{n \rightarrow \infty} E(S^2) = \lim_{n \rightarrow \infty} \left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2$$

3.4 Estimateur avec biais(biaisé)

On dit que $\hat{\theta}_n$ est un estimateur avec biais de θ si :

$$E(\hat{\theta}_n) \neq \theta \text{ c-à-d } E(\hat{\theta}_n - \theta) = E(\hat{\theta}_n) - \theta \neq 0$$

Exemple 3.4.1 Soit (X_1, \dots, X_n) un n échantillon.

Et S^2 est un estimateur biaisé de σ^2

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

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

Cependant, l'absence de biais n'est pas une garantie absolue de « bon estimateur ».

Il faut aussi tenir compte de sa variance (3.1).

3.5 Risque d'un estimateur (Erreur quadratique moyenne)

Comme l'indique son nom ce critère mesure la distance au carré à laquelle $\hat{\theta}$ se situe en moyenne par rapport à θ . Donc on appelle erreur quadratique moyenne de $\hat{\theta}$ par rapport à θ , la valeur notée EQM définie par :

$$EQM(\hat{\theta}) = E \left[(\hat{\theta} - \theta)^2 \right]$$

On peut écrire :

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

Comme $E(\hat{\theta}) - \theta$ est une constante et que $E[\hat{\theta}_n - E(\hat{\theta}_n)] = 0$ il vient :

$$E\left[(\hat{\theta} - \theta)^2\right] = V(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$$

Donc

$$EQM(\hat{\theta}) = V(\hat{\theta}) + \text{Biais}^2(\hat{\theta})$$

Le meilleur de deux estimateur $\hat{\theta}_1$ et $\hat{\theta}_2$, c'est-à-dire le plus efficace, est celui qui a la plus petite EQM : $\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$ si :

$$EQM(\hat{\theta}_1) < EQM(\hat{\theta}_2) \iff \frac{EQM(\hat{\theta}_1)}{EQM(\hat{\theta}_2)} < 1$$

De deux estimateurs sans biais, ceci revient à dire que le plus efficace est celui la variance minimale

3.6 Estimateur robuste

Il arrive que lors d'un sondage, une valeur extrême et rare apparaisse (par exemple un enfant de 10 ans mesurant 1,80 m). On cherche à ce que ce genre de valeur ne change que de manière très faible la valeur de l'estimateur. On dit alors que l'estimateur est robuste.

Exemple 3.6.1 *En reprenant l'exemple de l'enfant (Exemple2.1.2), la moyenne n'est pas un estimateur robuste car ajouter l'enfant très grand modifiera beaucoup la valeur de l'estimateur. La médiane par contre n'est pas modifiée dans un tel cas.*

3.7 Estimateur efficace

Un estimateur sans biais est efficace s'il n'est pas possible de trouver un autre estimateur sans biais qui a une variance plus petite. Soient $\hat{\theta}_1$ et $\hat{\theta}_2$ deux estimateurs sans biais de θ , $\hat{\theta}_1$ est dit efficace que $\hat{\theta}_2$ si $V(\hat{\theta}_1) \leq V(\hat{\theta}_2)$ et $\hat{\theta}$ est dit un estimateur efficace si

$$V(\hat{\theta}) = \frac{1}{I_n(\theta)}$$

ou $I_n(\theta)$ est la quantité d'information apportée par l'échantillon $X = (X_1, X_2, \dots, X_n)$ sur θ , et par rapport 1.1

Exemple 3.7.1 $X_i \sim N(m, \sigma^2)$ et $V(\bar{X}) = \frac{\sigma^2}{n}$:

$$\begin{aligned} L(X, m) &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_i (X_i - m)^2\right) \\ \ln(L(X, m)) &= -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_i (X_i - m)^2 \\ \frac{\partial \ln(L(X, m))}{\partial m} &= \frac{1}{\sigma^2} \sum_{i=1} (X_i - m) \\ \frac{\partial^2 \ln(L(X, m))}{\partial m^2} &= \frac{-n}{\sigma^2} \end{aligned}$$

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

d'où

$$V(\bar{X}) = \frac{1}{I_n(\theta)}$$

Il en résulte que \bar{X} est un estimateur efficace.

Exemple 3.7.2 Soit X_1, X_2, \dots, X_n un n échantillon telles que

$\hat{\theta}_1 = \frac{X_1+X_2}{2}$, $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n X_i + \frac{X_1-X_2}{2}$ deux estimateurs sans biais de $m = E(X_1)$

$$E(\hat{\theta}_1) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}(E(X_1) + E(X_2)) = \frac{1}{2}(E(X_1) + E(X_2)) = \frac{1}{2}2E(X_1) = m$$

et

$$\begin{aligned} V(\hat{\theta}_1) &= V\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}[V(X_1) + V(X_2) + 2cov(X_1, X_2)] \\ &= \frac{1}{4}V(X_1) + V(X_2) \\ &= \frac{1}{4}2V(X_1) \\ &= \frac{1}{2}\sigma^2 \end{aligned}$$

$$V(\hat{\theta}_1) = \frac{1}{2}\sigma^2$$

on a $cov(X_1, X_2) = 0$ car X_1, X_2 sont indépendante.

$$E(\hat{\theta}_2) = E\left(\frac{1}{n} \sum_{i=1}^n X_i + \frac{X_1 - X_2}{2}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) + \frac{1}{2}[E(X_1) - E(X_2)] = \frac{1}{n}nE(X_1) = m$$

et

$$\begin{aligned} V(\hat{\theta}_2) &= V\left(\frac{1}{n} \sum_{i=1}^n X_i + \frac{X_1 - X_2}{2}\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n V(X_i) + \frac{1}{4}[V(X_1) + V(X_2) - 2cov(X_1, X_2)] \\ &= \frac{1}{n^2}nV(X_1) + \frac{1}{4}2V(X_1) \\ &= \frac{1}{n}\sigma^2 + \frac{1}{2}\sigma^2 \end{aligned}$$

$$V(\hat{\theta}_2) = \frac{2+n}{2n}\sigma^2$$

Alors $V(\hat{\theta}_1) < V(\hat{\theta}_2)$, donc $\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$

3.7.1 Inégalité de Fréchet-Darmois-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 : Si le domaine de définition de X ne dépend pas de θ , on a pour tout estimateur $\hat{\theta}$ sans biais de θ :

$$V(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$$

et si $\hat{\theta}$ est un estimateur sans biais de $k(\theta)$:

$$V(\hat{\theta}) \geq \frac{[k'(\theta)]^2}{I_n(\theta)}$$

où k est une fonction dérivable.

3.8 Estimateur absolument correct

Un estimateur est convergent si sa distribution tend à se concentrer autour de la valeur inconnue du paramètre. Un estimateur sans biais, dont la variance tend vers 0 quand n tend vers l'infini, est donc convergent. Il est absolument correct. Mais il n'est pas nécessairement unique comme le montrent les exemples ci-dessous.

Exemple 3.8.1 *Comparaison d'estimateurs de l'espérance mathématique* Comme estimateur de l'espérance mathématique m , on peut choisir la statistique T_1 :

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad E(\hat{\theta}_1) = m \quad \text{Var}(\hat{\theta}_1) = \frac{\sigma^2}{n}$$

T_1 est un estimateur sans biais dont la variance tend vers 0 quand n tend vers l'infini, il est donc convergent. On peut aussi choisir la statistique $\hat{\theta}_2$, moyenne arithmétique des observations de rang impair. On obtient si on suppose n pair :

$$\hat{\theta}_2 = \frac{2}{n} \sum_{i=0}^{p-1} X_{2i+1} \quad n = 2p \quad E(\hat{\theta}_2) = m \quad \text{Var}(\hat{\theta}_2) = \frac{2\sigma^2}{n}$$

L'estimateur $\hat{\theta}_2$ est sans biais et sa variance tend vers 0 quand n tend vers l'infini, il est donc convergent. Les deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ ont les mêmes propriétés. Il est évident cependant que $\hat{\theta}_1$ est « meilleur » que $\hat{\theta}_2$. En effet :

- il tient compte de toute l'information apportée par l'échantillon.
- sa variance est la plus petite : $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$

Exemple 3.8.2 Estimation du paramètre p d'une loi binomiale. On veut estimer la proportion p d'électeurs qui voteront pour le candidat A lors des élections municipales. On interroge un échantillon représentatif de taille n de l'ensemble des électeurs et soit K_n le nombre de réponses favorables ou f_n la fréquence des réponses.

$$\left\{ \begin{array}{l} E(K_n) = np \\ \text{Var}(K_n) = np(1-p) \end{array} \right. \quad \left\{ \begin{array}{l} E(f_n) = p \\ \text{Var}(f_n) = \frac{p(1-p)}{n} \end{array} \right.$$

La fréquence f_n est un estimateur sans biais du paramètre p . De plus, sa variance tend vers 0 quand n tend vers l'infini, il est donc convergent.

Résumé : estimateurs convergents et sans biais

- d'une moyenne $E(X)$: la statistique $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

- d'une variance S^2 : la statistique $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Attention : S n'est pas un estimateur sans biais de l'écart-type σ , en effet :

$$E(\sqrt{S^2}) \neq \sqrt{E(S^2)}$$

- d'une proportion p : la fréquence $f_n = \frac{K_n}{n}$ (K_n nombre de réalisations de l'événement étudié au cours de n épreuves).

3.9 Recherche du meilleur estimateur

La recherche du meilleur estimateur d'un paramètre est un problème difficile à résoudre. En effet :

- La précision d'un estimateur $\hat{\theta}$ dépend de sa variance, c'est-à-dire de la loi de $\hat{\theta}$ qui dépend elle-même de la loi de la variable aléatoire X . Il faut donc connaître la forme de cette loi .
- Une statistique est un résumé apporté par un échantillon, il est donc très important de ne pas perdre d'information.

En tenant compte de ces deux impératifs, on peut aborder la recherche du meilleur estimateur suivant deux méthodes :

- Soit en recherchant des statistiques exhaustives qui conduisent à des estimateurs sans biais, de variance minimale.
- Soit en étudiant la quantité d'information de Fisher qui apporte des indications sur la précision d'un estimateur.

Conclusion

L'un des problèmes de la statistique mathématique est la recherche des estimateurs efficaces, c'est-à-dire des estimateurs à variance minimale.

L'estimateur du maximum de vraisemblance possède cette propriété d'efficacité, mais malheureusement cet estimateur est très sensible contre les valeurs extrêmes, en d'autre terme est un estimateur non robuste.

C'est pour cette raison que nous avons présenté ce travail qui s'intéresse aux qualités d'un estimateur.

Notations et symbols

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

$v.a$	→	Variable aléatoire.
$i.i.d$	→	Indépendantes identiquement distribuées.
$c.à.d$ ou $i.e$	→	C'est à dire.
$E(X)$ ou μ	→	Espérance mathématique ou moyenne di v.a X.
$V(X)$	→	Variance du v.a X.
\bar{X}	→	Moyenne empirique.
σ	→	Ecart type.
\xrightarrow{p}	→	Converge en probabilité.
S^2	→	Variance empirique.
$L(x; \theta)$ ou $L_\theta(x)$	→	Fonction de la vraisemblance.
$f(x; \theta)$	→	Densité de probabilité.
E	→	L'ensemble d'observation.
EMV	→	Estimateur du maximum de vraisemblance.
$EQM(\hat{\theta}_n, \theta)$	→	L'erreur quadratique moyenne de $\hat{\theta}_n$ par rapport à θ :
Θ	→	ensemble de définition θ

Bibliographie

- [1] Bouraine, (2014). Polycopié de cours Statistique inférentielles. Université A. MIRA-Béjaia, Algérie.
- [2] Dusart, P. (2018). cours de Licence 2, Statistique inférentielles.
- [3] Hubler, J. (2007). Statistique descriptive appliquée à la gestion et à l'économie. Editions Bréal.
- [4] JACQUES, J. Statistiques inférentielles.
- [5] Jean-Yves Dauxois (2011-2012) CTU, master enseignement des mathématiques Statistique inférentielle.
- [6] Lejeune, M. (2004). Statistique : La théorie et ses applications. Springer Science & Business Media.
- [7] Rabah, M. (2014). Cours de statistique inférentielles, USTHB-Faculte des mathématique, Algérie.
- [8] Ruch, J-J. (2012-2013). statistique : Estimation.
- [9] Saporta, G. (2006). Probabilités, analyse des données et statistique. Editions Technip

- [10] Saporta, G.(2011). Probabilités, analyse des données et statistique. Technip, Paris.
- [11] Velenik, Y. (2011). Probabilités et statistique. Université de Geneve.
- [12] Veysseyre, R. (2014). Aide-mémoire-Statistique et probabilités pour les ingénieurs. Dunod.