République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOHAMED KHIDER, BISKRA

FACULTE des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DEPARTEMENT DE MATHEMATIQUES

Mémoire présenté par

Saad Chaima

En vue de l'obtention du Diplôme de

MASTER en Mathématiques

Option : **Statistique**

Titre

Efficacité d'un Estimateur

Membres du Comité d'Examen

Pr.	**YAHIA Djabrane**	UMKB	Président
Pr.	**MERAGHNI Djamel**	UMKB	Encadreur
Dr.	**CHINE Amel**	UMKB	Examinateur

Juin 2021

Dédicace

Avec mes sentiments de gratitude les plus profonds, Je dédie ce modeste travail :

A mes très chers parents, pour leur soutien inconditionnel tout au long de mes études et pour la confiance qu'ils m'ont toujours témoigné.

A mes grands-mères.

A mes chers frères et sœurs ; Sofiane, Walid, Ahmed Yazid, Somia, Marwa, Anfel et mon ange Ikhlasse

A mes tantes et mes oncles.

A mes cousines et cousins, en particulier Zeineb.

Que l'unité familiale reste pour nous tout l'objectif premier. Votre soutien et amour me donnent la volonté d'aller encore plus loin pour que vous soyez fiers de moi. Merci de m'avoir aidé et encouragé. Que Dieu préserve l'union dans notre famille et vous donne longue vie et santé. Je vous adore.

« Ma réussite est aussi la vôtre »

Chaima Saad

REMERCIEMENTS

En premier lieu, je remercie Dieu le Tout Puissant de m'avoir donné le courage et la santé pour achever ce travail.

Mes sincères remerciements et ma profonde reconnaissance vont à l'endroit de Monsieur ***MERAGHNI Djamel*** *Professeur à l'université de Biskra, pour avoir dirigé ce mémoire dans les règles de l'art sans jamais ménager ni son temps, ni ses compétences pour m'aider et me guider également pour le respect qu'il a eu vis-à-vis de mes recherches.*

Ma reconnaissance s'adresse également à Monsieur ***DJABRANE Yahia*** *et Madame* ***CHINE Amel****, qui m'ont fait l'honneur de constituer le jury de mémoire. Veuillez accepter chers professeurs mes sentiments de gratitude.*

Je remercie tous ceux qui m'ont enseignée durant toutes mes études au département de mathématiques.

J'adresse aussi, un merci particulier à mes amies pour leur amitié, leurs encagements, leur sympathie et leur bonne humeur. Je garde un excellent souvenir et j'espère vous en avoir laissé un aussi bon.

Enfin, que toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail, trouvent ici ma sincère reconnaissance et mes remerciements.

« Aucun travail ne s'accomplit dans la solitude »

Chaima Saad

Table des matières

Liste des tableaux

Introduction

La statistique est un ensemble de méthodes permettant de dégager les caractéristiques ou la répartition d'objets en fonction de critères d'étude déterminés. Son objectif est d'extraire des informations utiles des données et de mesurer les dépendances entre variables ainsi que leur variabilité. Ces données sont issues de domaines très vairées comme la médecine, l'économie, la sociologie, l'ingénierie, l'astrophysique, l'internet, ...

L'estimation consiste à donner une valeur approchée à un (ou plusieurs) paramètre inconnu d'une population sur la base d'un échantillon issu de cette dernière.

Lorsque l'on résume l'échantillon par une statistique, il ne faut pas trop réduire l'information apportée par l'échantillon initial. L'information peut être définie en suivant deux approches : la statistique exhaustive et l'information de Fisher.

L'un des problèmes de la statistique mathématique est la recherche d'estimateurs efficaces, c'est à dire des estimateurs à variance minimale. Notre objectif, à travers ce travail, est de rappeler les deux principales méthodes de construction d'un estimateur, puis de définir les critères permettant de mesurer son l'efficacité.

Ce mémoire se compose de deux chapitres :

- Chapitre 1 : Estimation ponctuelle

  Dans ce premier chapitre nous aborderons la définition de l'estimateur et ses propriétés (le biais, la consistance et la précision), aussi on va définir les propriétés des deux estimateurs intéressant qui sont la moyenne et la variance empirique. Enfin, pour estimer un paramètre on va déterminer deux méthodes d'estimation statistique, la méthode du maximum de vraisemblance et des moments.

- Chapitre 2 : Efficacité

  Le contenu du premier chapitre représente les concepts de base du deuxième chapitre. Ce dernier est consacré aux notions d'exhaustivité, famille exponentielle, information de Fisher et borne de Cramer-Rao. Ceci permettra d'arriver à la définition de l'efficacité d'un estimateur.

Chapitre 1

Estimation ponctuelle

En statistique paramétrique, l'estimation ponctuelle est une opération qui consiste à chercher une valeur approchée pour la vraie valeur inconnue d'un paramètre. Ceci se fait, sur la base d'une série d'observations (échantillon), par le biais de différentes méthodes. Les résultats obtenus font l'objet d'analyse dans le but de vérifier certaines qualités (consistance, biais, efficacité,...).

1.1 Statistique

X est une variable aléatoire (v.a) dont la fonction de répartition $F(x;\theta)$ et la densité $f(x;\theta)$ dépendent d'un (ou plusieurs) paramètre réel $\theta \in \Theta \subseteq \mathbb{R}$ (ou $\theta = (\theta_1, ..., \theta_k) \in \Theta \subseteq \mathbb{R}^k$). On considère un échantillon, de taille n (un n-échantillon), $(X_1, ..., X_n)$ de cette variable.

Définition 1.1.1 *Une statistique T est une fonction mesurable des v.a X_i :*

$$T := T(X_1, ..., X_n).$$

Remarque 1.1.1 *Elle permet d'avoir des informations sur la valeur théorique inconnue du paramètre θ.*

Exemple 1.1.1

1. *La moyenne empirique* $\overline{X}$ *et la variance empirique* S^2 *respectivement définies par* (1.3) *et* (1.4) *sont deux statistiques. Elles sont très utilisées en statistique mathématique.*

2. *Plus généralement, le moment empirique d'ordre* $k \geq 1$, *défini par*

$$M_k := \frac{1}{n} \sum_{i=1}^{n} X_i^k, \tag{1.1}$$

est une statistique. On l'utilise quand on s'intéresse au moment théorique

$$\mu_k := E\left(X^k\right), \ k \geq 1. \tag{1.2}$$

1.2 Estimateur ponctuel

L'estimation statistique consiste à attribuer une valeur approchée au paramètre inconnu (moyenne, variance,...) d'une population à l'aide d'observations $x_1, ..., x_n$ (réalisations de l'échantillon $(X_1, ..., X_n)$). Le but en général est de choisir, parmi toutes les statistiques, celle qui apporte le plus d'information possible sur le paramètre en question. De amples détails sur ce sujet peuvent être trouvés dans [6], [8], [9] et [11].

1.2.1 Définition

Définition 1.2.1 *Soit* $(X_1, ..., X_n)$ *un* n*-échantillon d'une v.a* X *dont la loi de probabilité dépend d'un paramètre* θ. *On appelle estimateur de* θ *toute statistique* T, *donnée dans la définition* 1.1.1, *ayant des propriétés bien définies. On le note par* $\widehat{\theta}_n$ *ou plus simplement par* $\widehat{\theta}$.

Exemple 1.2.1

- *La moyenne empirique* $\overline{X}$ *est un estimateur de la moyenne théorique* $\mu = E(X)$.

- *La variance empirique* S^2 *est un estimateur de la variance* $\sigma^2 = V(X)$.

1.2.2 Propriétés d'un estimateur

Pour juger si l'estimateur est bon or non, on vérifie certaines bonnes propriétés.

Biais d'un estimateur

Définition 1.2.2 *Soit* $\widehat{\theta}$ *un estimateur du paramètre* θ

1. *On appelle biais de* $\widehat{\theta}$ *pour* θ *la quantité*

$$b_\theta(\widehat{\theta}) = E(\widehat{\theta}) - \theta.$$

2. *L'estimateur* $\widehat{\theta}$ *est dit sans biais pour* θ *si son biais est nul, c-à-d si*

$$E(\widehat{\theta}) = \theta.$$

3. *L'estimateur* $\widehat{\theta}$ *est dit asymptotiquement sans biais si*

$$E(\widehat{\theta}) \longrightarrow \theta \text{ quand } n \longrightarrow \infty.$$

Remarque 1.2.1 *On dit que* $\widehat{\theta}$ *est biaisé si*

$$E(\widehat{\theta}) \neq \theta.$$

Consistance d'un estimateur

Il y a plusieurs types de convergence dont les plus importants sont la convergence en probabilité (convergence faible), la convergence presque sure (convergence forte) et la convergence en loi. La consistance d'un estimateur est définie par rapport au premier type.

Définition 1.2.3 *Un estimateur $\widehat{\theta}$ est dit consistant pour θ s'il converge en probabilité vers θ (on écrit $\widehat{\theta} \xrightarrow{P} \theta$), c-à-d pour tout $\varepsilon > 0$, on a*

$$P(\left|\widehat{\theta} - \theta\right| < \varepsilon) \longrightarrow 1 \text{ ou } P(\left|\widehat{\theta} - \theta\right| > \varepsilon) \longrightarrow 0, \; n \longrightarrow \infty.$$

Pour vérifier cette propriété de convergence, on a le résultat ci-dessous, qui est déduit directement de l'inégalité de Bienaymé-Tchebychev (voir [9], page 61).

Proposition 1.2.1 *Tout estimateur sans biais (ou asymptotiquement sans biais) dont la variance tend vers zéro est consistant.*

$$\left.\begin{array}{c} E(\widehat{\theta}) = \theta \text{ (ou } E(\widehat{\theta}) \longrightarrow \theta) \\ V(\widehat{\theta}) \longrightarrow 0 \end{array}\right\} \Longrightarrow \widehat{\theta} \xrightarrow{P} \theta, \; n \longrightarrow \infty$$

Précision d'un estimateur

Pour évaluer la précision d'un estimateur on utilise l'erreur quadratique moyenne définie par

$$EQM\left(\widehat{\theta}\right) := E\left((\widehat{\theta} - \theta)^2\right).$$

Proposition 1.2.2 *On a*

$$EQM\left(\widehat{\theta}\right) = \left(b_\theta(\widehat{\theta})\right)^2 + V(\widehat{\theta}).$$

Preuve. On a

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

Par la linéarité de l'espérance, on a

$$E[\widehat{\theta} - E(\widehat{\theta})] = 0,$$

et comme $E(\widehat{\theta}) - \theta$ est une constante, alors

$$E[(E(\widehat{\theta}) - \theta)^2] = \left(E(\widehat{\theta}) - \theta\right)^2 = \left(b_\theta(\widehat{\theta})\right)^2.$$

Par conséquent, on obtient le résultat voulu. ■

Corollaire 1.2.1 *Si $\widehat{\theta}$ est un estimateur sans biais, alors*

$$EQM\left(\widehat{\theta}\right) = V(\widehat{\theta}).$$

Définition 1.2.4 *Soient $\widehat{\theta}_1$ et $\widehat{\theta}_2$ deux estimateurs de θ. On dit que $\widehat{\theta}_1$ est plus précis (meilleur) que $\widehat{\theta}_2$ si*

$$EQM\left(\widehat{\theta}_1\right) < EQM\left(\widehat{\theta}_2\right).$$

On dit que un estimateur $\widehat{\theta}_1$ est relativement plus efficace que $\widehat{\theta}_2$ s'il est plus précis que le second.

Remarque 1.2.2

1. *On définit l'efficacité relative de $\widehat{\theta}_2$ par $\widehat{\theta}_1$ par*

$$eff\left(\widehat{\theta}_2, \widehat{\theta}_1\right) := \frac{EQM\left(\widehat{\theta}_1\right)}{EQM\left(\widehat{\theta}_2\right)}.$$

   Si cette quantité est inférieure à 1, alors $\widehat{\theta}_1$ est plus précis que $\widehat{\theta}_2$.

2. *Si les deux estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont sans biais, alors la comparaison se fait par rapport aux variances.*

1.3 Moyenne et variance empiriques

On considère un échantillon $(X_1, ..., X_n)$, de taille n, d'une v.a X d'espérance μ et de variance σ^2. Une description détaillée de ces deux statistiques se trouve, par exemple, dans [1] et [9].

1.3.1 Moyenne empirique

Définition 1.3.1 *On appelle moyenne empirique correspondant à cet échantillon, la statistique (notée $\overline{X}$) définie par*

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1.3}$$

Proposition 1.3.1 *On a*

$$E(\overline{X}) = \mu \text{ et } V(\overline{X}) = \frac{\sigma^2}{n}$$

Preuve. Pour l'espérance, on utilise la linéarité. On a

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

Pour la variance, on utilise l'indépendance des X_i. En effet,

$$V(\overline{X}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}V(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.$$

C'est ce qu'il fallait trouvé. ■

De la proposition 1.3.1, on déduit le résultat suivavnt.

Corollaire 1.3.1 *$\overline{X}$ est un estimateur sans biais et consistant de l'espérance μ.*

Proposition 1.3.2 (convergence en loi) *La moyenne empirique converge en loi vers la loi de normale :*

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow[n \longrightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

On dit que la distribution de $\overline{X}$ est approximée, pour n assez grand, par la loi normale. On note cela par

$$\overline{X} \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n).$$

Preuve. C'est l'application directe du théorème central-limite (TCL). Pour les détails, se référer à [9], page 66. ■

1.3.2 Variance empirique

Définition 1.3.2 *On appelle variance empirique la statistique (notée S^2) définie par*

$$S^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2. \tag{1.4}$$

Proposition 1.3.3 *On a*

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

Preuve. On a

$$\begin{aligned}
E(S^2) &= \frac{1}{n}\sum_{i=1}^{n}E(X_i^2) - E(\overline{X}^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}(V(X_i) + E(X_i)^2) - V(\overline{X}) - E(\overline{X})^2 \\
&= \frac{1}{n}(nV(X) + nE(X)^2) - \frac{1}{n}V(X) - E(X)^2 \\
&= V(X) + E(X)^2 - \frac{1}{n}V(X) - E(X)^2 \\
&= \frac{n-1}{n}V(X).
\end{aligned}$$

D'où le résultat. ■

Remarque 1.3.1 *Il est clair que S^2 est un estimateur biaisé de la variance σ^2. Mais, il est asymptotiquement sans biais.*

On définit alors la variance empirique corrigée, qui est sans biais, par

$$S^{*^2} := \frac{n}{n-1} S^2.$$

En effet, on a $E\left(S^{*^2}\right) = \dfrac{n}{n-1} E\left(S^2\right) = \dfrac{n}{n-1} \times \dfrac{n-1}{n} \sigma^2 = \sigma^2.$

Proposition 1.3.4 (corrélation entre $\overline{X}$ et S^2) *On a*

$$cov\left(\overline{X}, S^2\right) = \frac{n-1}{n^2} \mu_3^*,$$

où $\mu_3^* := E\left(X-\mu\right)^3$ *désigne le moment centré d'ordre* 3.

Preuve. La formule de la covariance est

$$cov\left(\overline{X}, S^2\right) = E\left(\overline{X} S^2\right) - E\left(\overline{X}\right) E\left(S^2\right).$$

En supposant que $\mu = 0$ (dans ce cas $\mu_3^* = \mu_3$) on obtient

$$cov\left(\overline{X}, S^2\right) = E\left(\overline{X} S^2\right).$$

En utilisant les définitions de $\overline{X}$ et S^2, on a

$$\begin{aligned} cov\left(\overline{X}, S^2\right) &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} X_j^2 - \overline{X}^2\right)\right] \\ &= \frac{1}{n^2} E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j^2\right) - E\left(\overline{X}^3\right) \\ &= \frac{1}{n^2} E\left(\sum_{i=1}^{n} X_i^3\right) - \frac{1}{n^3} E\left(\sum_{i=1}^{n} X_i\right)^3 \\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n} E\left(X_i^3\right)\right) - \frac{1}{n^3}\left(\sum_{i=1}^{n} E\left(X_i^3\right)\right), \end{aligned}$$

car $E\left(X_i X_j^2\right) = 0$ pour $i \neq j$ à cause de l'indépendance des éléments de l'échantillon.

$$cov\left(\overline{X}, S^2\right) = \frac{\mu_3}{n} - \frac{\mu_3}{n^2}.$$

D'où le résultat. Ceci reste valable lorsque $\mu \neq 0$, car la covariance est insensible aux transformations des variables. ■

Remarque 1.3.2 *La convariance est nulle (c-à-d non corrélation entre $\overline{X}$ et S^2) dans le cas où le troisième moment centré est nul. Ce qui est les cas des distributions symétriques. Si de plus, la population mère est normale alors les statistiques $\overline{X}$ et S^2 sont indépendantes.*

1.4 Méthodes d'estimation statistique

Il existe plusieurs méthodes pour déterminer un estimateur d'un paramètre sur la base d'un échantillon. Parmi lesquelles, les plus connues sont celles du maximum de vraisemblance, des moments et des moindres (en régression linéaire). Dans ce paragraphe, on s'intéressera aux deux premières méthodes.

1.4.1 Méthode du maximum de vraisemblance

C'est la plus universelle et attractive pour les raisons suivantes (voir [9], page 306) :

1. Elle est facile à mettre en oeuvre et plus simple que les méthodes alternatives. Elle se ramène à un problème classique de résolution d'équations.

2. Elle est optimale et asymptotiquement efficace (c-à-d quand la taille de l'échantillon tend vers l'infini théoriquement). D'un point de vue pratique, pour un échantillon suffisamment grand $(n > 30)$, elle fournit des estimateurs de très bonne qualité.

3. Elle donne des estimateurs dont la variance asymptotique est la plus faible parmi tous les estimateurs sans biais.

Définition 1.4.1 (fonction de vraisemblance) *On appelle fonction de vraismblance (ou plus simplement vraisemblance) de θ pour une réalisation $(x_1, ..., x_n)$ de l'échantillon $(X_1, ..., X_n)$, la fonction de θ*

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

où

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

Le principe de la méthode est de déterminer la valeur de θ qui rend la fonction de vraisemblance $L(\theta)$ maximale. Puisque la fonction logarithme est croissante, alors la maximisation de $L(\theta)$ revient à celle de $\ln L(\theta)$, appelée fonction log-vraisemblance (voir [1] et [6]).

Définition 1.4.2 (estimateur du maximum de vraisemblance (emv)) *On appelle emv de θ toute fonction $\widehat{\theta}$ de $(X_1, ..., X_n)$ vérifiant*

$$L(x_1, ..., x_n; \widehat{\theta}) = \max_{\theta \in \Theta} L(x_1, ..., x_n; \theta) \text{ ou } \ln L(x_1, ..., x_n; \widehat{\theta}) = \max_{\theta \in \Theta} \ln L(x_1, ..., x_n; \theta).$$

En d'autre termes,

$$\widehat{\theta} := \arg\max_{\theta \in \Theta} L(x_1, ..., x_n; \theta) = \arg\max_{\theta \in \Theta} \ln L(x_1, ..., x_n; \theta)$$

Remarque 1.4.1 (détermination de l'emv) *En général, l'emv doit vérifier les deux conditions suivantes :*

$$\begin{cases} \dfrac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = 0, \\ \dfrac{\partial^2 \ln L(x_1, ..., x_n; \theta)}{\partial^2 \theta} < 0. \end{cases}$$

Exemple 1.4.1 (loi continue) *Soit $(X_1, ..., X_n)$ un n-échantillon d'une v.a $X \sim \mathcal{N}(\mu, \sigma^2)$, où σ^2 est supposée connue. On souhaite trouver l'emv du paramètre $\mu \in \mathbb{R}$. La fonction de* *densité est*

$$f(x;\mu) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right),\ x \in \mathbb{R}.$$

Les fonctions de vraisemblance et log-vraisemblance sont respectivement

$$L(x_1,...,x_n;\mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right), \tag{1.5}$$

et

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

Alors on a

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

qui s'annule lorsque

$$\sum_{i=1}^n (x_i-\mu) = 0 \Longleftrightarrow \sum_{i=1}^n x_i - n\mu = 0.$$

La solution de cette équation est $\mu = \frac{1}{n}\sum_{i=1}^n x_i$. *On vérifie qu'il s'agit bien d'un maximum. En effet, on a*

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

Par conséquent, l'emv de μ *est*

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i,$$

c'est la moyenne empirique $\overline{X}$.

Exemple 1.4.2 (loi discrète) *Soit* $(X_1,...,X_n)$ *un* n*-échantillon d'une v.a* $X \sim \mathcal{P}(\lambda)$. *On souhaite estimer le paramètre* $\lambda > 0$. *On a*

$$P(X=x) = \frac{\lambda^x}{x!}e^{-\lambda},\ x \in \mathbb{N}.$$

Distribution	**Paramètre**	**Estimateur**
uniforme $\mathcal{U}([0,\theta])$	$\theta > 0$	$\max\limits_{1\leq i\leq n} X_i$
binomiale $\mathcal{B}(n,p)$	$0 < p < 1$	fréquence empirique
Poisson $\mathcal{P}(\lambda)$	$\lambda > 0$	$\overline{X}$
exponentielle $\mathcal{E}(\theta)$	$\theta > 0$	$1/\overline{X}$
normale $\mathcal{N}(\mu,\sigma^2)$	μ σ^2	$\overline{X}$ S^2

TAB. 1.1 – Estimateurs du maximum de vraismblance des paramètres de quelques lois usuelles

Les fonctions de vraisemblance et log-vraisemblance sont respectivement

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

Alors la dérivée par rapport à λ est

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

Elle s'annule pour $\lambda = \frac{1}{n}\sum_{i=1}^{n}x_i$, qui vérifie que

$$\frac{\partial^2 \ln L(x_1,...,x_n;\lambda)}{\partial\lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n}x_i < 0.$$

D'où l'esimateur de λ est

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}X_i,$$

c-à-d la moyenne empirique.

Dans le tableau 1.1, on résume les estimateurs obtenus par la méthode du maximum de vraisemblance pour les paramètres de qielques unes des distributions les plus usuelles (voir [11]).

1.4.2 Méthode des moments

Cette méthode est la plus naturelle, elle consiste à écrire le paramètre en termes des moments théoriques puis à remplacer ces derniers par leurs contre-parties empiriques. En d'autres termes, si $\theta = h(\mu_1, \mu_2, ...)$ alors l'estimateur des moments (em) de θ est

$$\widehat{\theta} := h(M_1, M_2, ...),$$

où M_k et μ_k sont définis par (1.2) et (1.1) respectivement. Pour plus de détails, voir, par exemple, [2] et [4].

Remarque 1.4.2

1. *Si le paramètre à estimer est* $\theta = \mu_1 = E(X)$ *(espérance de la v.a), alors l'em de* θ *est* $\widehat{\theta} = M_1 = \overline{X}$.
2. *Si* $\theta = V(X) = \mu_2 - \mu_1^2$ *(variance de la v.a), alors l'em de* θ *est* $\widehat{\theta} = M_2 - M_1^2 = S^2$.

Exemple 1.4.3 (loi continue) *Soit* $(X_1, ..., X_n)$ *un échantillon d'une v.a* $X \sim \mathcal{U}([0, \theta])$, *de densité de probabilité*

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}, \text{ avec } \theta > 0. \tag{1.6}$$

On a $E(X) = \mu_1 = \theta/2$, *d'où* $\theta = 2\mu_1$. *Donc, on déduit que l'em de* θ *est*

$$\widehat{\theta} = 2M_1 = 2\overline{X}.$$

Exemple 1.4.4 (loi discrète) *Soit* $(X_1, ..., X_n)$ *un échantillon de* $X \sim \mathcal{B}(p)$ *de loi de probabilité*

$$P(X = x) = p^x (1-p)^{1-x}, \ x = 0, 1.$$

On a $E(X) = p,$ *c-à-d* $p = E(X) = \mu_1$. *Donc l'em de* p *est*

$$\widehat{p} = M_1 = \overline{X}.$$

Remarque 1.4.3 *Les deux d'estimation ci-dessus ne donnent pas nécessairement les mêmes estimateurs. C'est le cas, par exemple, de l'estimation du paramètre* θ *de la loi uniforme où l'emv est* $\max\limits_{1\leq i\leq n} Xi$ *(voir le tableau 1.1) alors que l'em est* $2\overline{X}$. *La question qui se pose dans de tels cas est de savoir lequel, parmi les différents estimateurs, est le meilleur.*

Chapitre 2

Efficacité

En tant que statistique, un estimateur constitue souvent une réduction de l'information de manière à ne retenir que ce qui parait utile lors de l'estimation d'un certain paramètre. La qualité d'un estimateur est définie par rapport à plusieurs critères (voir le chapitre 1) dont l'efficacité qui se mesure en termes de l'information de Fisher.

2.1 Statistique exhaustive

Dans un problème statistique, on cherche à obtenir le maximum de l'information possible sur un (ou plusieurs) paramètre θ inconnu. Pour cela, on dispose de l'observation $(x_1, ..., x_n)$ d'un échantillon $(X_1, ..., X_n)$, où n est très grand. Il est alors intéressant de réduire les données en les résumant par une statistique $T(X_1, ..., X_n)$ de dimension inférieure à n. Généralement, il est logique de s'attendre à ce que ceci résulte en moins d'information sur θ que celle contenue dans l'ensemble des données initiales. Cependant, il existe des statistiques qui résument les observations tout en conservant l'intégralité de l'information sur θ. Ce sont les statistiques exhaustives.

2.1.1 Exhaustivité

Définition 2.1.1 *Une statistique* T *est dit exhaustive si la loi conditionnelle de* $(X_1, ..., X_n)$ *sachant que* $T(x_1, ..., x_n) = t$ *ne dépend pas de* θ*, c-à-d*

$$P\left((X_1, ..., X_n) / T(x_1, ..., x_n) = t\right) \text{ indépendante de } \theta.$$

Dire que T est exhaustive signifie qu'elle contient toute l'information, induite par l'observation $(x_1, ..., x_n)$, sur la valeur inconnue du paramètre θ.

Exemple 2.1.1 (contrôle de qualité [3]) *Dans un processus de production, on considère* $(X_1, ..., X_n)$*, où*

$$X_i = \begin{cases} 1 & \text{si la } i^{\text{ème}} \text{ pièce est défectueuse} \\ 0 & \text{sinon} \end{cases}$$

Les X_i *sont des v.a indépendantes et de même loi* $\mathcal{B}(p)$*, où* p *est la probabilité qu'une pièce soit défectueuse. On montre que le nombre total de pièces défectueuses* $T = \sum_{i=1}^{n} X_i$ *est une statistique exhaustive. Cela revient à montrer que* $P\left((X_1, ..., X_n) / T = t\right)$ *ne dépend pas de* p*.*

En effet, on a $T \sim \mathcal{B}(n, p)$ *(loi binomiale). Alors*

$$\begin{aligned} P\left((X_1, ..., X_n) / T(x_1, ..., x_n) = t\right) &= P\left(X_1 = x_1, ..., X_n = x_n / \sum_{i=1}^{n} x_i = t\right) \\ &= \frac{P\left(X_1 = x_1, ..., X_n = x_n, \sum_{i=1}^{n} x_i = t\right)}{P\left(\sum_{i=1}^{n} x_i = t\right)} \\ &= \begin{cases} 0 & si \sum_{i=1}^{n} x_i \neq t \\ \frac{P(X_1 = x_1, ..., X_n = x_n)}{P\left(\sum_{i=1}^{n} x_i = t\right)} & si \sum_{i=1}^{n} x_i = t \end{cases} \end{aligned}$$

Comme les X_i sont indépendants, on a

$$\frac{P\left(X_1=x_1,...,X_n=x_n\right)}{P\left(\sum\limits_{i=1}^{n}x_i=t\right)}=\frac{\prod\limits_{i=1}^{n}P\left(X_i=x_i\right)}{P\left(T=t\right)}=\frac{\prod\limits_{i=1}^{n}p^{x_i}\left(1-p\right)^{1-x_i}}{C_n^t p^t\left(1-t\right)^{n-t}}$$

$$=\frac{p^{\sum\limits_{i=1}^{n}x_i}\left(1-p\right)^{n-\sum\limits_{i=1}^{n}x_i}}{C_n^t p^t\left(1-t\right)^{n-t}}=\frac{1}{C_n^t}$$

Donc $P\left(\left(X_1,...,X_n\right)/T(x_1,...,x_n)=t\right)=\begin{cases}0 & si\ \sum\limits_{i=1}^{n}x_i\neq t\\ \frac{1}{C_n^t} & si\ \sum\limits_{i=1}^{n}x_i=t\end{cases}.$

Il est clair que la probabilité conditionnelle ci-dessus ne dépend pas de p.

2.1.2 Théorème de factorisation

En raison du calcul compliqué de probabilité conditionnelle, il est très difficile de montrer l'exhaustivité en utilisant la définition ci-dessus. Il existe un théorème, du à Fisher-Neyman et connu sous le nom de théorème de factorisation, facile d'utilisation permettant de trouver des statistiques exhaustives.

Théorème 2.1.1 (factorisation) *Soit un n-échantillon $\left(X_1,...,X_n\right)$ d'une v.a X et $T=T\left(X_1,...,X_n\right)$ une statistique de loi $g\left(t;\theta\right)$ (densité dans le cas continu, $P\left(T=t\right)$ dans le cas discret). La statistique T est exhaustive pour θ si et seulement si, il existe une fonction mesurable h ne dépendant que de $\left(x_1,...,x_n\right)$ et une fonction mesurable g ne dépendant que de t et θ telles que la fonction de vraisemblance puisse s'écrire sous la forme*

$$L\left(x_1,...,x_n;\theta\right)=g\left(t;\theta\right)h\left(x_1,...,x_n\right).$$

Preuve. Voir [3], page 16. ■

Exemple 2.1.2 (cas discret) *On a vu dans l'exemple 2.1.1 que*

$$L(x_1, ..., x_n; p) = p^t (1-p)^{n-t} = g(t; p)\, h(x_1, ..., x_n),$$

où $g(t; p) = p^t (1-p)^{n-t}$ *et* $h(x_1, ..., x_n) = 1$, *avec* $t = \sum_{i=1}^n x_i$. *Donc,* T *est une statistique exhaustive pour* p.

Exemple 2.1.3 (cas continu) *Soit* $(X_1, ..., X_n)$ *un* n*-échantillon d'une v.a* $X \sim \mathcal{N}(\mu, \sigma^2)$, *où le paramètre* $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$, *est inconnu. On a d'après 1.5*

$$L(x_1, ..., x_n; \mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right], \tag{2.1}$$

qui est de la forme $g\left(\left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right); \mu, \sigma^2\right) h(x_1, ..., x_n)$, *avec* h *égale à la constante* 1. *Donc le couple* $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ *est une statistique exhaustive pour le paramètre* (μ, σ^2).

Remarque 2.1.1 T *est une statistique exhaustive pour* θ *et* Ψ *une fonction strictement monotone de* T. *Alors, la statistique* $S = \Psi(T)$ *est une statistique exhaustive pour* θ.

Exemple 2.1.4 *La relation (2.1) peut être réécrite comme*

$$\begin{aligned} L(x_1, ..., x_n; \mu, \sigma^2) &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left(ns^2 - 2n\mu\overline{x} + n\mu^2\right)\right] \\ &= g\left(\left(\overline{x}, s^2\right); \mu, \sigma^2\right). \end{aligned}$$

Donc le couple $\left(\overline{X}, S^2\right)$ *est aussi une statistique exhaustive pour le paramètre* (μ, σ^2).

Proposition 2.1.1 (maximum de vraisemblance et exhaustivité) *Si* T *est une statistique exhaustive pour* θ *et si* $\widehat{\theta}$ *est un emv de* θ, *alors*

$$\widehat{\theta} := \arg\max_{\theta\in\Theta} g(T(x_1, ..., x_n); \theta) = \arg\max_{\theta\in\Theta} \ln g(T(x_1, ..., x_n); \theta).$$

Dans ce cas, l'emv $\widehat{\theta}$ doit vérifier les deux conditions suivantes :

$$\begin{cases} \dfrac{\partial \ln g(T(x_1,...,x_n);\theta)}{\partial \theta} = 0, \\ \dfrac{\partial^2 \ln g(T(x_1,...,x_n);\theta)}{\partial^2 \theta} < 0. \end{cases}$$

Preuve. D'après la proposition 2.1.1, on a $L(x_1,...,x_n;\theta) = g(t;\theta)\,h(x_1,...,x_n)$, avec h ne dépendant pas de θ. Par conséquent, maximiser la vraisemblance L revient à maximiser la fonction g. ■

Remarque 2.1.2

1. *Tout emv est fonction d'une statistique exhaustive. En d'autres termes, si T est une statistique exhaustive et si $\widehat{\theta}$ est un emv, alors il existe une fonction Φ telle que $\widehat{\theta} = \Phi \circ T$.*
2. *Un emv n'est pas nécessairement exhaustif.*

2.2 Famille exponentielle

C'est un modèle paramétrique très important en statistique, car elle possède un certain nombre de propriétés intéressantes. Pour une description détaillée de cette famille, on réfère le lecteur aux livres [7], [8] et [9].

Définition 2.2.1 *Soit X une v.a dont la loi de probabilité dépend d'un paramètre $\theta = (\theta_1,...,\theta_k) \in \Theta \subseteq \mathbb{R}^k$. On dit que la loi de X (ou que X) appartient à la famille exponentielle si et seulement si sa densité de probabilité $f(x;\theta)$ dans le cas continu (ou sa fonction de probabilité $P(X = x;\theta)$ dans le (cas discret) s'écrit de la forme*

$$f(x;\theta) = \exp\left\{\sum_{i=1}^{k} a_i(x)\,\alpha_i(\theta) + b(x) + \beta(\theta)\right\},\ \ x \in \mathbb{R}, \tag{2.2}$$

où a_i, b_i, α et β sont des fonctions.

Remarque 2.2.1

1. *Dans la formule* (2.2) , *le nombre* k *de termes de la somme est égal à la dimension du paramètre* θ.

2. *Dans le cas d'un seul paramètre* $(k = 1)$, *l'écriture* (2.2) *devient*

$$f(x;\theta) = \exp\{a(x)\alpha(\theta) + b(x) + \beta(\theta)\}. \tag{2.3}$$

3. *Une forme équivalente à* (2.2) *est*

$$f(x;\theta) = c(x)\, d(\theta) \exp\left\{\sum_{i=1}^{k} a_i(x)\, \alpha_i(\theta)\right\},$$

avec $c(x) = \exp(b(x))$ *et* $d(\theta) = \exp(\beta(\theta))$.

Exemple 2.2.1 (loi discrète) $X \sim \mathcal{B}(n,p)$. *On a pour* $x \in \{0, ..., n\}$

$$\begin{aligned} P(X = x; p) &= C_n^x p^x (1-p)^{n-x} = C_n^x \left(\frac{p}{1-p}\right)^x (1-p)^n \\ &= \exp\left\{\ln C_n^x + x \ln\left(\frac{p}{1-p}\right) + n\ln(1-p)\right\} \\ &= \exp\left\{\ln\left(\frac{n!}{x!\,(n-x)!}\right) + x\ln\left(\frac{p}{1-p}\right) + n\ln(1-p)\right\}. \end{aligned}$$

C'est de la forme (2.3) , *avec* $a(x) = x,\ \alpha(p) = \ln\left(\frac{p}{1-p}\right)$, $b(x) = \ln\left(\frac{n!}{x!\,(n-x)!}\right)$ *et* $\beta(p) = n\ln(1-p)$.

Exemple 2.2.2 (loi continue) 1. $X \sim \mathcal{E}(\lambda)$. *Pour* $x \geq 0$, *on a*

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

qui est de la forme (2.3) , *avec* $a(x) = x,\ \alpha(\lambda) = -\lambda,\ b(x) = 0$ *et* $\beta(\lambda) = \ln\lambda$.

2. $X \sim \mathcal{N}(\mu, \sigma^2)$. *On a*

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

qui appartient de la forme (2.2), *avec* $a_1(x) = x$, $\alpha_1(\mu, \sigma^2) = \mu/\sigma^2$, $a_2(x) = x^2$, $\alpha_2(\mu, \sigma^2) = -1/2\sigma^2$, $b(x) = 0$ *et* $\beta(\mu, \sigma^2) = -\mu^2/2\sigma^2 - \ln\left(\sigma\sqrt{2\pi}\right)$.

Remarque 2.2.2 *La loi de Weibull de paramètre* $\theta = (\eta, \lambda)$ *n'appartient pas à la famille exponentielle. En effet, on a*

$$\begin{aligned} f(x;\eta,\lambda) &= \frac{\eta}{\lambda}\left(\frac{x}{\lambda}\right)^{\eta-1}e^{-\left(\frac{x}{\lambda}\right)^{\eta}} \\ &= \exp\left\{-\left(\frac{x}{\lambda}\right)^{\eta}+(\eta-1)\ln\left(\frac{x}{\lambda}\right)+\ln\left(\frac{\eta}{\lambda}\right)\right\} \end{aligned}$$

Le terme $(x/\lambda)^\eta$ *ne peut pas être mis sous la forme* $a(x)\,\alpha(\eta, \lambda)$.

Le théorème suivant, connu sous le nom de théorème de Darmois, donne le lien entre famille exponentielle et exhaustivité.

Théorème 2.2.1 (Darmois) *Soit* X *une v.a de densité (ou loi de probabilité)* f *et dont le support (ensemble où* f *est positive strictement) ne dépend pas du paramètre* θ*. Une condition nécessaire et suffisante pour qu'un échantillon admette une statistique exhaustive pour* $\theta = (\theta_1, ..., \theta_k)$ *est que* f *appartienne à la famille exponentielle définie par* (2.2). *Une statistique exhaustive particulière est* $T := \left(\sum_{i=1}^n a_1(X_i), ..., \sum_{i=1}^n a_k(X_i)\right)$.

Preuve. Dans la cas d'un paramètre, la démonstration est faite dans [9], page 293. ■

Exemple 2.2.3 *De l'exemple 2.2.2, on déduit que :*

1. *$\sum_{i=1}^{n} X_i$ est une statistique exhaustive pour le paramètre λ de $\mathcal{E}(\lambda)$.*

2. *$\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ est une statistique exhaustive pour le couple (μ, σ^2) de $\mathcal{N}(\mu, \sigma^2)$.*

2.3 Information de Fisher

L'information de Fisher est une notion centrale en statistique paramétrique pour l'étude de l'efficacité d'un estimateur. Elle a éte définie par R.A. Fisher pour mesurer l'information relative à un paramètre, apportée par un échantillon. Pour plus de détails se référer à [2], [5], [7] et [9].

2.3.1 Définitions

Soit X une v.a dont la loi de probabilité dépend d'un paramètre $\theta = (\theta_1, ..., \theta_k) \in \Theta \subseteq \mathbb{R}^k$. On se concentrera surtout sur le cas $k = 1$, c-à-d le cas d'un seul paramètre. Les quatre hypothèses **H1-H4** suivantes seront utiles pour la suite.

H1. Le support de X ne dépend pas de θ.
H2. La fonction $\frac{\partial}{\partial\theta} L(x_1, ..., x_n; \theta)$ est intégrable.
H3. La fonction $\frac{\partial}{\partial\theta} L(x_1, ..., x_n; \theta)$ est de carré intégrable.
H4. La fonction $\frac{\partial^2}{\partial\theta^2} L(x_1, ..., x_n; \theta)$ existe et est intégrable.

Remarque 2.3.1

1. *Les distributions appartenant à la famille exponentielle vérifient les quatre hypothèses* **H1-H4**.

2. *Sous les hypothèses* **H1**, **H2** *et* **H4** *on peut échanger les signes d'intégration et de dérivation, c-à-d*

$$\int \frac{\partial}{\partial\theta} L(x_1, ..., x_n; \theta)\, dx = \frac{\partial}{\partial\theta} \int L(x_1, ..., x_n; \theta)\, dx,$$
$$\int \frac{\partial^2}{\partial\theta^2} L(x_1, ..., x_n; \theta)\, dx = \frac{\partial^2}{\partial\theta^2} \int L(x_1, ..., x_n; \theta)\, dx.$$

Définition 2.3.1 *La quantité d'information de Fisher apportée, sur le paramètre θ, par un n-échantillon $(X_1, ..., X_n)$ vérifiant l'hypothèse* **H3**, *est définie par*

$$I_n(\theta) := E\left[\left(\frac{\partial \ln L(X_1, ..., X_n; \theta)}{\partial \theta}\right)^2\right] = E\left[(S(X_1, ..., X_n; \theta))^2\right],$$

où

$$S(x_1, ..., x_n; \theta) := \frac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = \frac{\partial L(x_1, ..., x_n; \theta)/\partial \theta}{L(x_1, ..., x_n; \theta)}, \tag{2.4}$$

est appelée fonction score.

Remarque 2.3.2 *L'information de Fisher d'une seule observation x_i de la v.a X est donnée par*

$$I(\theta) = E\left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta}\right)^2\right] = E\left[(S(X; \theta))^2\right].$$

Proposition 2.3.1 *Sous les hypothèses* **H1** *et* **H2**, *le score est centré.*

Preuve. On montre que $E[S(X_1, ..., X_n; \theta)] = 0$. On a

$$\begin{aligned}
E[S(X_1, ..., X_n; \theta)] &= E\left[\frac{\partial \ln L(X_1, ..., X_n; \theta)}{\partial \theta}\right] \\
&= \int \frac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} L(x_1, ..., x_n; \theta) dx \\
&= \int \frac{\partial L(x_1, ..., x_n; \theta)/\partial \theta}{L(x_1, ..., x_n; \theta)} L(x_1, ..., x_n; \theta) dx \\
&= \int \frac{\partial}{\partial \theta} L(x_1, ..., x_n; \theta) dx \\
&= \frac{\partial}{\partial \theta} \int L(x_1, ..., x_n; \theta) dx.
\end{aligned}$$

Finalement, on utilise le fait que $L(x_1, ..., x_n; \theta)$ est une densité pour obtenir le résultat. ∎

Corollaire 2.3.1 *Sous les hypothèses* **H1**, **H2** *et* **H3**, *l'information de Fisher est tout simplement égale à la variance du score :* $I_n(\theta) = V[S(X_1, ..., X_n; \theta)]$.

Exemple 2.3.1

1. *Soit* $(X_1, ..., X_n)$ *un échantillon, de taille* $n \geq 1,$ *d'une population* $X \sim \mathcal{B}(p)$

$$P(X = x) = p^x (1 - p)^{1-x}, \ x = 0, 1.$$

Les fonctions de vraisemblance et log-vraisemblance sont respectivement

$$L(x_1, ..., x_n; p) = p^t (1 - p)^{n-t} \ \text{ et } \ \ln L(x_1, ..., x_n; p) = t \ln(p) + (n - t) \ln(1 - p),$$

où $T = \sum_{i=1}^n X_i, \sim \mathcal{B}(n, p)$ *avec* $E(T) = np$ *et* $V(T) = np(1 - p)$.

On calcule le score :

$$S(x_1, ..., x_n; p) = \frac{\partial \ln L(x_1, ..., x_n; p)}{\partial p} = \frac{t}{p} - \frac{n - t}{1 - p}.$$

Donc l'information de Fisher est

$$I_n(p) = V\left(\frac{T}{p} - \frac{n - T}{1 - p}\right) = \frac{n}{p(1 - p)}.$$

2. *Soit* $(X_1, ..., X_n)$ *un* n*-échantillon d'une v.a* $X \sim \mathcal{N}(\mu, \sigma^2)$, *avec* σ^2 *connue.*

Le score a déjà été calculé dans l'exemple 1.4.1. On a donc

$$I_n(\mu) = V\left(\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)\right) = \frac{1}{\sigma^4}\sum_{i=1}^n V(X_i) = \frac{1}{\sigma^4}\sum_{i=1}^n \sigma^2 = \frac{n}{\sigma^2}.$$

Proposition 2.3.2 *Sous les trois hypothèses* **H1**, **H3** *et* **H4**, *l'information de Fisher est égale à*

$$I_n(\theta) := -E\left(\frac{\partial^2 \ln L(X_1, ..., X_n; \theta)}{\partial \theta^2}\right) = -E\left(\frac{\partial S(X_1, ..., X_n; \theta)}{\partial \theta}\right). \tag{2.5}$$

Cette écriture est souvent la plus pratique pour les calculs.

Preuve. D'après la proposition 2.3.1, on a

$$E\left[S\left(X_1,...,X_n;\theta\right)\right]=\int\frac{\partial\ln L(x_1,...,x_n;\theta)}{\partial\theta}L(x_1,...,x_n;\theta)dx=0.$$

En dérivant les deux membres, on obtient

$$\int\frac{\partial^2\ln L(x_1,...,x_n;\theta)}{\partial\theta^2}L(x_1,...,x_n;\theta)dx+\int\frac{\partial\ln L(x_1,...,x_n;\theta)}{\partial\theta}\frac{\partial L(x_1,...,x_n;\theta)}{\partial\theta}dx=0.$$

En utilisant la relation (2.4) , on écrit

$$\int\frac{\partial^2\ln L(x_1,...,x_n;\theta)}{\partial\theta^2}L(x_1,...,x_n;\theta)dx+\int\left(\frac{\partial\ln L(x_1,...,x_n;\theta)}{\partial\theta}\right)^2L(x_1,...,x_n;\theta)dx=0,$$

c-à-d

$$E\left(\frac{\partial^2\ln L\left(X_1,...,X_n;\theta\right)}{\partial\theta^2}\right)+I_n\left(\theta\right)=0.$$

D'où le résultat. ∎

Exemple 2.3.2 *On applique la formule 2.5 sur l'exemple 2.3.1*

1. *Soit $(X_1,...,X_n)$ un échantillon de v.a $X\sim\mathcal{B}(p)$. On a*

$$\frac{\partial^2\ln L\left(x_1,...,x_n;p\right)}{\partial p^2}=-\frac{t}{p^2}-\frac{n-t}{\left(1-p\right)^2}.$$

Donc, l'information de Fisher, sur la proportion p, est

$$I_n\left(p\right)=-E\left(-\frac{T}{p^2}-\frac{n-T}{\left(1-p\right)^2}\right)=\frac{n}{p\left(1-p\right)}.$$

2. *Soit $(X_1,...,X_n)$ un n-échantillon d'une v.a $X\sim\mathcal{N}\left(\mu,\sigma^2\right)$, avec σ^2 connue. On a*

$$\frac{\partial^2\ln L\left(x_1,...,x_n;\mu\right)}{\partial\mu^2}=-\frac{n}{\sigma^2}.$$

Donc, l'information de Fisher, sur la moyenne μ, est

$$I_n(\mu) = -E\left(-\frac{n}{\sigma^2}\right) = \frac{n}{\sigma^2}.$$

Remarque 2.3.3 *Si le paramètre est multidimensionnel $\theta = (\theta_1, ..., \theta_k)$, alors :*

- *le score est un vecteur de dimension k :*

$$S(x_1, ..., x_n; \theta) := \begin{pmatrix} S_1(x_1, ..., x_n; \theta) \\ \vdots \\ S_k(x_1, ..., x_n; \theta) \end{pmatrix},$$

où $S_i(x_1, ..., x_n; \theta) = \partial \ln L(x_1, ..., x_n; \theta) / \partial\theta_i, i = 1, ..., k$. C'est le gradient $(\nabla \ln L)$ de la log-vraisemblance :

- *La quantité d'information de Fisher est égale à la matrice de covariance du score.*
- *Sous les hypothèses* **H1**, **H3** *et* **H4** *on exprime l'information de Fisher en termes des dérivées des scores $S_i(x_1, ..., x_n; \theta)$:*

$$I_n(\theta) = -E\begin{pmatrix} \frac{\partial S_1(x_1,...,x_n;\theta)}{\partial\theta_1} & ... & \frac{\partial S_1(x_1,...,x_n;\theta)}{\partial\theta_k} \\ ... & ... & ... \\ \frac{\partial S_k(x_1,...,x_n;\theta)}{\partial\theta_1} & ... & \frac{\partial S_k(x_1,...,x_n;\theta)}{\partial\theta_k} \end{pmatrix}.$$

Exemple 2.3.3 *Soit une v.a $X \sim \mathcal{N}(\mu, \sigma^2)$ de paramètres μ et σ^2 inconnus (dans ce cas $\theta = (\mu, \sigma^2)$).*

On a, d'après l'exemple 1.4.1,

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

Distribution	**Paramètre**	**Information de Fisher**
Bernoulli $\mathcal{B}(p)$	$0 < p < 1$	$1/p(1-p)$
Binomiale $\mathcal{B}(n,p)$	$0 < p < 1$	$n/p(1-p)$
Binomiale négative $\mathcal{B}n(r,p)$	$0 < p < 1$	$r/p^2(1-p)$
Poisson $\mathcal{P}(\lambda)$	$\lambda > 0$	$1/\lambda$
Exponentielle $\mathcal{E}(\lambda)$	$\lambda > 0$	$1/\lambda^2$
Normale $\mathcal{N}(\mu,\sigma^2)$ Normale $\mathcal{N}(\mu,\sigma^2)$	$-\infty < \mu < +\infty$ $\sigma^2 > 0$	$1/\sigma^2$ $1/2\sigma^4$
Uniforme $\mathcal{U}([0,\theta])$	$\theta > 0$	$1/\theta^2$
Pareto de paramètre θ	$\theta > 0$	$1/\theta^2$

TAB. 2.1 – Quantités d'information de Fisher apportées par une observation sur les paramètres de distributions de probabilités

Les deux scores sont

$$S_1(x_1,...,x_n;\theta) = \frac{\partial \ln L(x_1,...,x_n;\theta)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu),$$

et

$$S_2(x_1,...,x_n;\theta) = \frac{\partial \ln L(x_1,...,x_n;\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Donc, le vecteur score est

$$S(x_1,...,x_n;\theta) = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \\ -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 \end{pmatrix}.$$

Donc, la matrice d'infomation de Fisher est

$$I_n(\theta) = -E\begin{pmatrix} -\frac{n}{\sigma^2} & \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu) \\ \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu) & \frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2 \end{pmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

A la fin de cette partie, on résume, dans la tableau 2.1, les quantités d'information de Fisher apportées par une seule observation sur les paramètres de quelques unes des distributions de probabilités usuelles (extrait du tableau 6.1 de [7]).

2.3.2 Propriétés

- Positivité : $I_n(\theta) \geq 0$ pour tout θ.
- Additivité : si X et Y sont deux v.a indépendantes dont les supports ne dépendent pas de θ, alors on a

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta),$$

où $I_X(\theta)$, $I_Y(\theta)$ et $I_{(X,Y)}(\theta)$ sont les quantités information de Fisher au sur θ liées respectivement à X, Y et (X, Y).

On en déduit le résultat important suivant :

Proposition 2.3.3 *Soit* $(X_1, ..., X_n)$ *un* n*-échantillon vérifiant hypothèses* **H1** *et* **H3**. *Alors*

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

Preuve. On a

$$\begin{aligned} I_n(\theta) &= V[S(X_1, ..., X_n; \theta)] = V\left(\frac{\partial \ln L(X_1, ..., X_n; \theta)}{\partial \theta}\right) \\ &= V\left(\sum_{i=1}^{n} \frac{\partial \ln f(X_i; \theta)}{\partial \theta}\right) = \sum_{i=1}^{n} V\left(\frac{\partial \ln f(X_i; \theta)}{\partial \theta}\right) \\ &= nI(\theta). \end{aligned}$$

Donc, l'information apportée par un n-échantillon est n fois celle apportée par une seule variable. ■

Remarque 2.3.4 *On dit que chaque observation apporte la même proportion de l'informantion globale apportée par tout l'échantillon. Autrement dit, les v.a éléments de l'échantillon ont la même importance.*

Exemple 2.3.4

1. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, avec σ^2 connue. On a $I_n(\mu) = n/\sigma^2$ (voir l'exemple 2.3.1) et $I(\mu) = 1/\sigma^2$ (voir le tableau 2.1). On constate que

$$I_n(\mu) = \frac{n}{\sigma^2} = n \times \frac{1}{\sigma^2} = nI(\mu).$$

2. Soit une v.a $X \sim \mathcal{U}([0, \theta])$, $\theta > 0$. D'après (1.6), les fonctions de vraisemblance et log-vraisemblance sont respectivement

$$L(x_1, ..., x_n; \theta) = \frac{1}{\theta^n} I_{\max\{x_i\} \leq \theta} \text{ et } \ln L(x_1, ..., x_n; \theta) = -n \ln(\theta) I_{\max\{x_i\} \leq \theta}.$$

Le score est alors

$$S(x_1, ..., x_n; \theta) = \frac{-n}{\theta} I_{\max\{x_i\} \leq \theta}.$$

Donc

$$I_n(\theta) = E\left[(S(X_1, ..., X_n; \theta))^2\right] = \frac{n^2}{\theta^2}.$$

D'autre part, on a $I(\theta) = 1/\theta^2$ (voir le tableau 2.1). On remarque que $I_n(\theta) \neq nI(\theta)$. Ceci est dû au fait que le support de X dépend du paramètre θ (l'hypothèse **H1** n'est pas vérifiée).

2.3.3 Information de Fisher et exhaustivité

On suppose que les hypothèses **H1-H4** sont satisfaites.

Proposition 2.3.4 (dégradation de l'information) *L'information $I_T(\theta)$ portée par une statistique T est inférieure ou égale à celle apportée par l'échantillon $(X_1, ..., X_n)$, c-à-d*

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

Preuve. Soit T une statistique de densité $g\left(t;\theta\right)$. Alors

$$L(x_1,...,x_n;\theta)=g\left(t,\theta\right)h\left(x_1,...,x_n;\theta/t\right),$$

où h est la densité de l'échantillon conditionnellement à $T=t$. En dérivant deux fois la log-vraisemblance, on a

$$\frac{\partial^2\ln L\left(x_1,...,x_n;\theta\right)}{\partial\theta^2}=\frac{\partial^2\ln g\left(t,\theta\right)}{\partial\theta^2}+\frac{\partial^2\ln h\left(x_1,...,x_n,\theta/t\right)}{\partial\theta^2}.$$

Alors

$$-I_n\left(\theta\right)=-I_T\left(\theta\right)-I_{n/T}\left(\theta\right)\text{ c-à-d }I_T\left(\theta\right)=I_n\left(\theta\right)-I_{n/T}\left(\theta\right),$$

où

$$I_{n/T}\left(\theta\right):=E\left(\frac{\partial^2\ln h\left(X_1,...,X_n,\theta/t\right)}{\partial\theta^2}\right),$$

est la quantité d'information conditionnelle (par rapport à une distribution conditionnele). Or cette dernière est positive ou nulle, d'où le résultat. ■

Corollaire 2.3.2 *Si T est exhaustive, alors il n'y a pas de dégradation de l'information. La réciproque est aussi vraie. En d'autres termes, on a*

$$T\textit{ est exhaustive}\Longleftrightarrow I_T\left(\theta\right)=I_n\left(\theta\right).$$

En effet, le terme $I_{n/T}\left(\theta\right)$ est nul car h ne dépend pas de θ.

Remarque 2.3.5 *Puisque les statistiques exhaustives permettent de ne pas perdre d'information, il vaut mieux les choisir pour l'estimation.*

2.4 Borne de Cramer-Rao

En vue de définir ce qu'est le meilleur estimateur, on commence par définir la borne de Cramer-Rao. Elle va indiquer la borne inférieure pour la variance d'un estimateur sans biais. La proposition suivante précise cette propriété.

Proposition 2.4.1 (inégalité de Cramer-Rao) *Sous les hypothèses* **H1-H4** *(avec* $0 < I_n(\theta) < \infty$*), on a pour tout estimateur* $\widehat{\theta}$ *de* θ

$$V\left(\widehat{\theta}\right) \geq \frac{1}{I_n(\theta)} \left(\frac{\partial E\left(\widehat{\theta}\right)}{\partial \theta}\right)^2.$$

En particulier, si $\widehat{\theta}$ *est un estimateur sans biais de* θ*, alors*

$$V\left(\widehat{\theta}\right) \geq \frac{1}{I_n(\theta)}. \tag{2.6}$$

Preuve. On montre le cas particulier. On rappelle que $\partial \ln L(\theta)/\partial\theta$ est la fonction score $S = S(X_1, ..., X_n; \theta)$ et que $I_n(\theta) = V(S)$. On applique l'inégalité de Cauchy-Schwarz :

$$\left[cov\left(\widehat{\theta}, S\right)\right]^2 \leq V\left(\widehat{\theta}\right) V(S) = V\left(\widehat{\theta}\right) I_n(\theta). \tag{2.7}$$

On a

$$cov\left(\widehat{\theta}, S\right) = E\left(\widehat{\theta} S\right) - E\left(\widehat{\theta}\right) E(S).$$

Or $E(S) = 0$, d'où

$$\begin{aligned} cov\left(\widehat{\theta}, S\right) = E\left(\widehat{\theta} S\right) &= \int \widehat{\theta} S L(x_1, ..., x_n; \theta)\, dx_1 ... dx_n \\ &= \int \widehat{\theta} \left(\frac{\partial L}{\partial \theta}/L\right) L dx = \frac{\partial}{\partial \theta} \int \widehat{\theta} L(x_1, ..., x_n; \theta)\, dx_1 ... dx_n \\ &= \frac{\partial}{\partial \theta} E\left(\widehat{\theta}\right) = \frac{\partial \theta}{\partial \theta} = 1. \end{aligned}$$

En substituant ce résultat dans (2.7) , on obtient le résultat. ■

Remarque 2.4.1

1. *Le deuxième membre dans les inégalités de la proposition 2.4.1 est appelé borne de Cramer-Rao.*
2. *Si $\widehat{\theta}$ est un estimateur sans biais de $\varphi(\theta)$, alors on a*

$$V\left(\widehat{\theta}\right) \geq \frac{1}{I_n(\theta)} \left(\frac{\partial \varphi(\theta)}{\partial \theta}\right)^2.$$

Exemple 2.4.1 (loi discrète) *Soit $X \sim \mathcal{P}(\lambda)$. D'après l'exemple 1.4.2, la dérivée du score est*

$$\frac{\partial S(x_1, ..., x_n; \lambda)}{\partial \lambda} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i = -\frac{1}{\lambda^2}t,$$

où $t = \sum_{i=1}^n x_i$. La quantité d'information de Fisher est

$$I_n(\lambda) = -E\left(-\frac{1}{\lambda^2}T\right) = \frac{1}{\lambda^2}E(T) = \frac{1}{\lambda^2} \times n\lambda = \frac{n}{\lambda}.$$

Comme la moyenne empirique $\overline{X}$ est un estimateur sans biais de λ, alors la Cramer-Rao est égale à

$$\frac{1}{I_n(\lambda)} = \frac{\lambda}{n}.$$

Exemple 2.4.2 (loi continue) *Soit $X \sim \mathcal{E}(\lambda), \lambda > 0$. Les fonctions de vraisemblance et log-vraisemblance sont respectivement*

$$L(x_1, ..., x_n; \lambda) = \lambda^n \exp(-\lambda t) \;\; et \;\; \ln L(x_1, ..., x_n; \lambda) = n \ln \lambda - \lambda t,$$

où $t = \sum_{i=1}^n x_i$. On déduit le score et sa dérivée

$$S(x_1, ..., x_n; \lambda) = \frac{n}{\lambda} - t \;\; et \;\; \frac{\partial S(x_1, ..., x_n; \lambda)}{\partial \lambda} = -\frac{n}{\lambda^2},$$

La quantité d'information de Fisher est donc

$$I_n(\lambda) = -E\left(-\frac{n}{\lambda^2}\right) = \frac{n}{\lambda^2}.$$

La moyenne empirique $\overline{X}$ *est un estimateur sans biais de* $\varphi(\lambda) = 1/\lambda$. *Donc, la borne de Cramer-Rao est égale à*

$$\frac{1}{I_n(\lambda)}\left(\frac{\partial\varphi(\lambda)}{\partial\lambda}\right)^2 = \frac{\lambda^2}{n} \times \left(-\frac{1}{\lambda^2}\right)^2 = \frac{1}{n\lambda^2}.$$

2.5 Estimateur efficace

L'inégalité (2.6) de Cramer-Rao fournit la borne inférieure pour l'ensemble des variances des estimateurs sans biais, qui peut être atteinte ou non. Si cette borne est atteinte par un estimateur, ce dernier sera donc le meilleur et sera qualifié d'estimateur optimal dans la classe des estimateurs sans biais. On parle alors d'estimateur sans biais de variance minimale (ESBVM).

D'aprés les hypothèses **H1**, **H2**, **H3** et **H4** on a

Définition 2.5.1 (efficacité) *Un estimateur sans biais* $\widehat{\theta}$ *de* θ *est dit efficace si sa variance est égale à la borne de Cramer-Rao, c-à-d*

$$V\left(\widehat{\theta}\right) = \frac{1}{I_n(\theta)}.$$

Remarque 2.5.1

1. *Si* $\widehat{\theta}$ *est un estimateur sans biais pour* $\varphi(\theta)$, *alors il est dit efficace si*

$$V\left(\widehat{\theta}\right) = \frac{1}{I_n(\theta)}\left(\frac{\partial\varphi(\theta)}{\partial\theta}\right)^2.$$

2. *On dit que* $\widehat{\theta}$ *est asymptotiquement efficace si sa variance est équivalente (lorsque*

$n \longrightarrow \infty$) à la borne de Cramer-Rao.

La proposition suivante donne la relation l'existence d'un estimateur efficace et la famille exponentielle.

Proposition 2.5.1 (famille exponentielle et efficacité) *La borne de Cramer-Rao ne peut être atteinte que si la loi de X appartient à la famille exponentielle (2.3).*
D'autre part, si la loi de X est bien de la forme précédente, il n'existe (à une transformation linéaire près) qu'une seule fonction $\varphi(\theta)$ du paramètre qui puisse être estimée efficacement, où la fonction φ et son estimateur sont respectiement

$$\varphi(\theta) = -\frac{\partial\beta(\theta)/\partial\theta}{\partial\alpha(\theta)/\partial\theta} \text{ et } \widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} a(X_i).$$

La variance minimale est alors

$$V\left(\widehat{\theta}\right) = \frac{\partial\varphi(\theta)/\partial\theta}{n\partial\alpha(\theta)/\partial\theta}.$$

Preuve. Voir [9], page 303. ■

Exemple 2.5.1

1. *Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, avec σ^2 connue. On a vu dans l'exemple 2.3.1 que $I_n(\mu) = n/\sigma^2$. La borne de Cramer-Rao, égale à $1/I_n(\mu) = \sigma^2/n$, est atteinte par la variance de la moyenne empirique $\overline{X}$. En effet, on a*

$$V\left(\overline{X}\right) = \frac{\sigma^2}{n} = \frac{1}{I_n(\mu)}.$$

Donc, $\overline{X}$ est l'estimateur (sans biais) efficace de μ.

2. *Soit $X \sim \mathcal{E}(\lambda), \lambda > 0$. De l'exemple 2.2.2, on a $a(x) = x$, $\alpha(\lambda) = -\lambda$, $b(x) = 0$ et*

$\beta(\lambda) = \ln \lambda$. *La fonction qu'on estime efficacement est donc*

$$\varphi(\lambda) = -\frac{\partial \beta(\lambda) / \partial \lambda}{\partial \alpha(\lambda) / \partial \lambda} = -\frac{1/\lambda}{-1} = \frac{1}{\lambda},$$

et son estimateur efficace est la moyenne empirique

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} a(X_i).$$

La variance (minimale) de cet estimateur est alors

$$V\left(\overline{X}\right) = \frac{\partial \varphi(\lambda) / \partial \lambda}{n \partial \alpha(\lambda) / \partial \lambda} = \frac{-1/\lambda^2}{-n} = \frac{1}{n\lambda^2}.$$

Conclusion

Dans ce mémoire, on a conclu que la notion d'efficacité d'un estimateur est fondée sur le célèbre résultat connu sous le nom d'inégalité de Cramer-Rao. En effet, ce résultat précise une borne inférieure pour la variance d'un estimateur quelconque. Autrement dit, cette dernière est forcément supérieure à la borne de Cramer-Rao. Dans le cas d'un estimateur sans biais, cette borne n'est rien d'autre que l'inverse de la quantité d'information de Fisher apportée, sur le paramètre d'intérêt, par un ensemble d'observations (échantillon).

Un estimateur est dit efficace si sa variance atteint la borne de Cramer-Rao. Dans la classe des estimateurs sans biais, le meilleur estimateur du point de l'efficacité est celui ayant une variance (minimale) égale à l'inverse de l'information de Fisher.

L'appartenance de la distribution de probabilité de la population à la famille exponentielle garantit l'existence d'une statistique exhaustive et, par conséquent, d'un estimateur efficace pour le paramètre en question ou pour une fonction de ce dernier.

Bibliographie

[1] Dusart, B. Cours de statistique inférentielles. Lisence 2-S4 SI-MASS. *http ://www.unilim.fr.*

[2] Dauxois, J.Y. (2012) . Statistique inférentielle. Université de Franche-Comté.

[3] Gaudoin,O. Statistique inférentielle avancée, notes de cours. Ensimag-2$^{\text{éme}}$ année. *https ://membres-ljk.imag.fr/Olivier.Gaudoin/courses.html.*

[4] Gaudoin, O. Principes et méthodes statistiques, notes de cours. Ensimag-2$^{\text{éme}}$ année. *https ://membres-ljk.imag.fr/Olivier.Gaudoin/courses.html.*

[5] Keribin,C. (2020) . Modélisation statistique. Université Paris-Saclay.

[6] Laoudj Chekraoui, F. Cours de statistique mathématique. Université de Jijel. *www.exoco-lmd.com.*

[7] Lejeune, M. (2010). Statistique, la théoie et ses applications. Springer.

[8] Lecoutre, J.P. (2012). Statistique et probabilités. Dunod, Paris.

[9] Saporta, G. (2011). Probabilités, analyse des données et statistique. Technip, Paris.

[10] Ammar Kessentini, S. (2010) . Sciences et technologies de l'information et de mathematiques. Universite de Nantes.

[11] Veysseyre, R. (2006). Aide-mémoire, statistique et probabilités pour l'ingénieur. Dunod, Paris .

Annexe : Abréviations et Notations

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

Notation	**Signification**
$b_\theta(T)$	Biais de l'estimateur T pour θ.
c-à-d	C'est-à-dire.
cov	Covariance.
em	Estimateur des moments.
emv	Estimateur du maximum de vraisemblance.
EQM	Erreur quadratique moyenne.
$E_\theta(T)$	Espérance ou moyenne d'un estimateur T.
$E(X)$ ou μ	Espérance mathématique ou moyenne de X.
$F(x;\theta)$	Fonction de répartition.
$f(x;\theta)$	Densité de probabilité.
$I_X(\theta)$	Quantité d'information de Fisher d'une variable X.
$I_n(\theta)$	Quantité d'information de Fisher sur θ.
i.i.d	Indépendantes identiquement distribuées.
$\xrightarrow{\mathcal{L}}$	Convergence en loi.
$L(x_1, ..., x_n;\theta)$ ou $L(\theta)$	Fonction de vraisemblance de θ.
$\xrightarrow{P}$	Convergence en probabilité.
S^{*^2}	Variance empirique corrigée.

$S(x_1, ..., x_n; \theta)$	Fonction score.
S^2	Variance empirique.
Θ	Ensemble des valeurs possibles de θ.
θ	Paramètre inconnu.
$\widehat{\theta}$	Estimateur de θ.
TCL	Théorème central-limite.
v.a X	Variable aléatoire X.
$V(X)$ ou σ^2	Variance de X.
$\overline{X}$	Moyenne empirique.
$(X_1, ..., X_n)$	Echantillon de taille n.

ملخص

يتم قياس كفاءة فعالية مقدر معلمة توزيع مجتمع إحصائي معين، بواسطة حد كرامر- راو. بالنسبة للمقدر الغير منحاز، فان هذا الأخير يساوي معكوس كمية معلومات فيشر المقدمة من عينة على القيمة غير المعروفة للمعلمة. نظرا لأن التباين دائما ما يكون محدودا من الأدنى فان المقدر الفعال هو الذي يصل تباينه الى حد كرامر- راو.

الكلمات المفتاحية: *إحصاء شامل؛ الاحتمالية العظمى؛ العائلة الأسية؛ الفعالية؛ حد كرامر- راو؛ معلومات فيشر؛ مقدر.*

Abstract

The efficiency of an estimator of a parameter, relative to the distribution of a certain population, is measured with respect to Cramer-Rao bound. For an unbiased estimator, the latter is equal to the inverse of the amount of Fisher information provided by a sample on the unknown value of the parameter. Since the variance is always bounded from below, an efficient estimator is that whose variance reaches Cramer-Rao bound.

Keywords: Cramer-Rao bound; Efficiency; Estimator; Exponential family; Fisher information; Maximum likelihood; Sufficient statistics.

Résumé

L'efficacité d'un estimateur d'un paramètre, relatif à la distribution d'une certaine population, est mesurée par rapport à la borne de Cramer-Rao. Pour un estimateur sans biais, cette dernière est égale à l'inverse de la quantité d'information de Fisher apportée, par un échantillon, sur la valeur inconnue du paramètre. La variance étant toujours minorée, un estimateur efficace est celui dont la variance atteint la borne de Cramer-Rao.

Mots clés : Borne de Cramer-Rao; Efficacité; Estimateur; Famille exponentielle; Information de Fisher; Maximum de vraisemblance; Statistique exhaustive.