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 en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

LADAOURI Chafia

Titre :

Estimation ponctuelle

Membres du Comité d'Examen :

Dr.	**CHINE Amel**	UMKB	Président
Dr.	**OUANOUGHI Yasmina**	UMKB	Encadreur
Dr.	**BENAMEUR Sana**	UMKB	Examinatrice

Juin 2021

Dédicace

$\mathcal{A}$ mes très chers parents.

$\mathcal{A}$ mes frères et sœurs.

$\mathcal{A}$ toute ma famille.

$\mathcal{A}$ mes collègues.

$\mathcal{A}$ mes amis.

REMERCIEMENTS

L*e moment est venu pour remercier qui m'ont aidé depuis que j'ai écrit la première lettre jusqu'à maintenant.*

*Avant tout, je tiens remercier "**ALLAH**" qui m'a donné la force, la patience et le courage pour arriver ici.*

*Mes remerciements à **Dr Ouanoughi Yasmina**, pour m'avoir encadré durant cette année.*

*Également, je remercie les membres du jury : **Dr Chine Amel** et **Dr Benameur Sana**, pour examiner et juger mon travail.*

*Je profite l'occasion pour remercier tous mes enseignants commençait par mes études primaires terminant par mes études supérieures. En particulier **Mr F. Benatia** et **Mr Mohammed Moumni**.*

*Une grande dédicace à mon cheikh qui j'ai eu le coran sous ces mains **Isaid Bouzid**. Merci pour votre prière pour moi, et à tous mes camarades de la mosquée de **Ali Ben Abi Talib**.*

Merci pour être toujours là.

Table des matières

Liste des tableaux

Introduction

En statistique, comme dans la théorie des probabilités le hasard intervient fortement. Mais dans la théorie des probabilités, on suppose la loi connue précisément et on cherche à donner les caractéristiques de la variable qui suit cette loi. L'objectif de la statistique est le contraire ; à partir de la connaissance de la variable, que peut-on dire de la loi de cette variable ?

Soit X une variable aléatoire dont la densité de probabilité $f(x;\theta)$ dépend d'un paramètre θ appartenant à l'ensemble de définition D_θ. A l'aide d'un échantillon issu de X, il s'agit de déterminer une valeur approchée $\hat{\theta}$ de ce paramètre. On pourra utiliser deux méthodes :

- Estimation ponctuelle : on calcule une valeur vraisemblable $\hat{\theta}$ de θ.
- Estimation par intervalle : on cherche un intervalle dans lequel θ se trouve avec une probabilité élevée.

Dans ce travail, nous sommes intéressés à l'estimation ponctuelle, nous avons répartit le présent travail en trois chapitres.

La théorie de l'estimation fait intervenir des fonctions ou statistiques particulières, appelées estimateurs, dont nous allons donner dans le premier chapitre les propriétés essentielles puis, nous étudierons principalement les statistiques exhaustives, et la quantité d'information apportée par un échantillon de taille n.

La recherche du meilleur estimateur d'un paramètre difficile à résoudre, c'est l'objectif du deuxième chapitre. En effet

- La précision d'un estimateur $\hat{\theta}$ dépende de sa variance.
- 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 mielleur 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.

Dans le troisième chapitre nous présentons quelques méthodes d'estimation ponctuelle.

Chapitre 1

Définitions générales

On va voir dans ce chapitre quelque définition usuelle associé à l'estimateur.

1.1 Définition d'une statistique

Soit X une $v.a$ dont la fonction de répartition $F(x;\theta)$ et la densité $f(x;\theta)$ dépondent du paramètre θ tel que D_θ est l'ensemble de définition de θ. On pose $X_1,\ldots,X_n$ est un échantillon de taille n tel que $n \in \mathbb{N}$ de la $v.a$ X.

Définition 1.1.1 *(Statistique)*

On dit que T est une statistique si T est une fonction mesurable des $v.a$'s X_i tel que $i = 1,\ldots,n$ qui donne une information d'un paramètre θ de la population $(X_1,\ldots,X_n)$.

i) À chaque échantillon on trouve différentes statistiques. Une statistique peut être à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^d$ on dit que T est une statistique vectorielle.

ii) La théorie de l'estimation consiste à définie des statistiques particulières, appelée estimateurs.

iii) Une fois l'échantillon effectivement réalisé, l'estimateur prend une valeur numérique, appelée estimation du paramètre θ. On notera $\hat{\theta}$.

Exemple 1.1.1

Soit la statistique suivante :

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

est la fonction moyenne arithmétique des n observation d'un échantillon *iid*. Cette statistique peut être considérée comme un estimateur, a priori raisonnable de l'espérance mathématique

$$\mathbb{E}[X] = m.$$

- L'espérance mathématique de T :

$$\begin{aligned}
\mathbb{E}[T] = \mathbb{E}\left[\overline{X}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\mathbb{E}[X_1 + \cdots + X_n] \\
&= \frac{1}{n}\left[\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]\right] \\
&= \frac{1}{n}[m + \cdots + m] \\
&= \frac{1}{n}[nm] \\
&= m.
\end{aligned}$$

- La variance mathématique de T :

La variance de T est égale à la variance σ^2 de la population divisé par la taille n de l'échantillon. Elle est définie comme suit :

$$\begin{aligned}
Var(T) = Var(\overline{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} Var(X_1 + \cdots + X_n) \\
&= \frac{1}{n^2}\left[Var(X_1) + \cdots + Var(X_n)\right] \\
&= \frac{1}{n^2}(\sigma^2 + \cdots + \sigma^2) \\
&= \frac{1}{n^2}(n\sigma^2) \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

1.2 Statistique exhaustive

1.2.1 Fonction de vraisemblance

Définition 1.2.1

On a un échantillon aléatoire $(X_1, \ldots, X_n)$ et soit X est une $v.a$ à n démentions donc elle admet une distribution de probabilité jointe de $(X_1, \ldots, X_n)$. Cette distribution est appelée *vraisemblance et on définie comme suit :*

– Pour X est continue :

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

Tel que f est la densité de la $v.a$ X.

– Pour X est discrète :

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

Tel que P est la probabilité de la $v.a$ X.

1.2.2 Définition d'une statistique exhaustive

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

Définition 1.2.2

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 θ.

Les variables aléatoires $(X_i)_{i=1,\ldots,n}$ étant indépendantes, la densité de l'échantillon $(X_1, \ldots, X_n)$ est :

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \qquad \forall \theta \in D_\theta \;\text{ et }\; \forall x_1, \ldots, x_n \in \mathbb{R}$$

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

Cette densité $L(x_1, \ldots, x_n; \theta)$ est une fonction de θ appelée *vraisemblance de l'échantillon*. Elle peut se mettre sous la forme :

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

Tel que :

- $g(t;\theta)$ est la densité de la statistique T.
- $h(x_1,\ldots,x_n;\theta/T=t)$ 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 :

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

Soit X une $v.a$ suivant une loi uniforme sur $[0,\theta]$.Tel que :

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

La statistique:

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

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

En effet :

– La fonction de vraisemblance de l'échantillon :

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

– Loi de la statistique T est :

La fonction de répartition est :

$$\begin{aligned}
F_T(t;\theta) &= P(T \leqslant t;\theta) \\
&= P(\sup_{i\in[1,n]} X_i \leqslant t;\theta) \\
&= P(X_1 \leqslant t,\ldots,X_n \leqslant t;\theta) \\
&= P(X_1 \leqslant t;\theta)\ldots P(X_n \leqslant t;\theta) \\
&= [P(X_1 \leqslant t;\theta)]^n \\
&= \begin{cases} 0 & \text{si } t\ < 0 \\ (\frac{t}{\theta})^n & \text{si } 0 \leqslant t \leqslant \theta \\ 1 & \text{si } t \geqslant \theta. \end{cases}
\end{aligned}$$

La densité est :

$$\begin{aligned}
f_T(t;\theta) &= [F_T(t;\theta)]^{'} \\
&= \begin{cases} \frac{n}{\theta^n} t^{n-1} & \text{si } t \in [0,\theta] \\ 0 & \text{sinon.} \end{cases}
\end{aligned}$$

Donc la vraisemblance est :

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

alors la statistique T est une statistique exhaustive.

- La propriété d'exhaustivité pour une statistique est intéressante si elle ne dépend pas de la taille de l'échantillon.
- Si T une statistique exhaustive pour le paramètre θ et g une fonction strictement

monotone de T, donc la statistique

$$R = g(T)$$

est une statistique exhaustive pour le paramètre θ.

1.2.3 Famille exponentielle

Définition 1.2.3

Soit X une $v.a$ réelle dont la loi de probabilité dépend d'un paramètre $\theta \in D_\theta$. On dit que la loi de X appartient à la famille exponentielle si et seulement si $P_X(x;\theta)$ (cas discrét) ou $f(x;\theta)$ (cas continue) est de la forme :

$$f(x;\theta) = \exp\left[a(x)\alpha(\theta) + b(x) + \beta(\theta)\right],$$

ou de la forme équivalente :

$$\ln\left[f(x;\theta)\right] = a(x)\alpha(\theta) + b(x) + \beta(\theta).$$

Remarque 1.2.1

La plupart des lois usuelles appartiennent à la famille exponentielle.

Exemple 1.2.1

On considère le modèle suivant :

$$B(n,p) : p \in [0,1], n \in \mathbb{N}. \quad \text{(Modèle Binomial).}$$

La densité de ce modèle est :

$$P_X(x;p) = C_n^x p^x (1-p)^{n-x},$$

donc

$$\begin{aligned}
P_X(x;p) &= C_n^x p^x (1-p)^{n-x} \\
&= C_n^x p^x (1-p)^{-x} (1-p)^n \\
&= C_n^x \left(\frac{p}{1-p}\right)^x (1-p)^n \\
&= \exp\left[\ln\left(C_n^x \left(\frac{p}{1-p}\right)^x (1-p)^n\right)\right] \\
&= \exp\left[x \ln\left(\frac{p}{1-p}\right) + \ln(C_n^x) + n\ln(1-p)\right].
\end{aligned}$$

Donc le modèle $B(n,p)$ est une famille exponentielle, tel que :

$$a(x) = x, \quad \alpha(p) = \ln\left(\frac{p}{1-p}\right), \quad b(x) = \ln(C_n^x), \quad \beta(p) = n\ln(1-p).$$

1.2.4 Théorème de Darmois

Nous sauvegardons les symboles dans les paragraphes précédents et on suppose que E l'ensemble de définition de la $v.a$ $X = (X_1, \ldots, X_n)$ ne dépend pas de θ. Une condition nécessaire et suffisante pour que l'échantillon $(X_1, \ldots, X_n)$ admette une statistique exhaustive est que la densité est de la forme exponentielle.

Et si de plus l'application :

$$x_j \longrightarrow t = \sum_{i=n}^{n} a(X_i)$$

est bijective et continûment différentiable pour tout x_j, alors la statistique T :

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

est une statistique exhaustive pour le paramètre θ.

Exemple 1.2.2

Reprenons l'exemple de la loi uniforme de densité

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

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

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

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

Exemple 1.2.3

La variable aléatoire X suit une loi normale $N(m;\sigma^2)$ de densité

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

Cette variable vérifie toutes les hypothèses du théorème de Darmois. En effet, la densité sous la forme exponentielle est donné par suite :

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

la formule ci-dessus est équivalente à :

$$\begin{aligned}
\ln f(x;(m,\sigma^2)) &= -\ln\sigma\sqrt{2\pi} - \frac{(x-m)^2}{2\sigma^2} \\
&= -\frac{x^2}{2\sigma^2} + \frac{mx}{\sigma^2} - \frac{m^2}{2\sigma^2} - \ln\sigma\sqrt{2\pi}.
\end{aligned}$$

Avec le domaine de définition ne dépend pas de m et σ.

Dans la léturature on peut trouver deux situation :

- Cas 1 : (σ est connu)

Dans ce cas on peut rédiger la forme exponentielle de la densité de probabilité par :

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

tel que :

$$a(x) = x, \quad \alpha(m) = \frac{m}{\sigma^2}, \quad b(x) = \frac{-x^2}{2\sigma^2}, \quad \beta(m) = \frac{-m^2}{2\sigma^2} - \ln\left(\sigma\sqrt{2\pi}\right).$$

Donc la statistique exhaustive est donnée par :

$$T = \sum_{i=1}^{n} a_i(x) = \sum_{i=1}^{n} X_i.$$

On remarque que T est une statistique exhaustive pour la moyenne m.

- Cas 2 : (m est connu)

Dans ce cas la formule exponentielle on peut la récrire par :

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

Tel que

$$a(x) = (x-m)^2, \quad \alpha(\sigma) = -\frac{1}{2\sigma^2} \quad b(x) = 0, \quad \beta(\sigma) = -\ln\sigma\sqrt{2\sigma}.$$

Donc

$$T = \sum_{i=1}^{n} a_i(x) = \sum_{i=1}^{n} \left(X_i - m\right)^2.$$

La statistique $T = \sum_{i=1}^{n}(X_i - m)^2$ est une statistique exhaustive pour la variance σ^2.

1.3 Information de Fisher

Définition 1.3.1

On appelle quantité d'information de Fisher $I_n(\theta)$ apportée sur un paramètre θ

par une échantillon est l'expression suivante :

$$\begin{aligned}
I_n(\theta) &= I_{x_1,\ldots,x_n}(\theta) \\
&= \mathbb{E}\left[\left(\frac{d\ln L(x_1,\ldots,x_n;\theta)}{d\theta}\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{L'(x_1,\ldots,x_n;\theta)}{L(x_1,\ldots,x_n;\theta)}\right)^2\right] \\
&= \int_{E_\theta}\left[\frac{d\ln L(x_1,\ldots,x_n;\theta)}{d\theta}\right]^2 L(x_1,\ldots,x_n;\theta)d\left(x_1,\ldots,x_n\right).
\end{aligned}$$

et on garde les mêmes notations des paragraphes précédents.

1.3.1 Propriétés de la quantité d'information

Si la vraisemblance $L(x;\theta)$ est dérivable au moins jusqu'à l'ordre deux et si l'ensemble E_θ ne dépend pas de θ, la quantité d'information de Fisher possède les propriétés suivantes :

$$\begin{aligned}
&\frac{d\ln L(x_1,\ldots,x_n;\theta)}{d\theta} \text{ est une variable aléatoire centré.} \\
&I_n(\theta) = Var\left[\frac{d\ln L(x_1,\ldots,x_n;\theta)}{d\theta}\right] \\
&I_n(\theta) = -\mathbb{E}\left[\frac{d^2\ln L(x_1,\ldots,x_n;\theta)}{d\theta^2}\right] \\
&I_n(\theta) = nI_1(\theta).
\end{aligned}$$

1.3.2 Précision apportée par un échantillon

Supposons que le paramètre θ à estimer soit la moyenne d'une loi normale. En remplaçant $f(x;\theta)$ par la densité de probabilité d'une loi normale dans les expressions précédentes, on obtient :

$$I_n(\theta) = \frac{n}{\sigma^2} = nI_1(\theta).$$

L'information $I_n(\theta)$ est donc d'autant plus grande que l'écart-type σ est petit (justification du mot « précision »).

Exemple 1.3.1

Soit X une variable aléatoire suit la loi de Bernoulli $\mathcal{B}(\theta)$

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

d'après des calculs, on trouve

$$\ln\left(f\left(x;\theta\right)\right) = x\ln(\theta) + (1-x)\ln(1-\theta).$$

Et

$$\begin{aligned}\frac{\partial}{\partial\theta}\ln(f(x;\theta)) &= \frac{x}{\theta} - \frac{1-x}{1-\theta}\\ \frac{\partial^2}{\partial\theta^2}\ln(f(x;\theta)) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}.\end{aligned}$$

Comme $\mathbb{E}\left[X\right] = \theta$, donc

$$I(\theta) = \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta)\right] = \frac{1}{\theta(1-\theta)}.$$

1.3.3 Dégradation de l'information

Si l'ensemble E_θ ne dépend pas du paramètre θ, l'information apportée par un échantillon est supérieure ou égale à l'information apportée par une statistique.

Il y a égalité si la statistique est exhaustive (justification du qualificatif « exhaustive »).

Chapitre 2

Qualité d'un estimateur ponctuelle

Pour décider que l'estimateur est bon il faut étudier leurs caractères. Dans ce chapitre on va voir des caractères comme : le biais, la convergence,...

2.1 Estimation

Estimer un paramètre, c'est chercher une valeur approchée en se basant sur les résultats obtenues dans un échantillon, lorsqu'un paramètre est estimé par un seul nombre, déduit des résultats de l'échantillon, ce nombre est appelé estimation ponctuelle du paramètre.

Définition 2.1.1

Soit X une $v.a$ dont la loi de probabilité P dépend d'un paramètre inconnu θ. On appelle estimateur de θ une $v.a$ $T = t(X_1, \ldots, X_n)$, où t une fonction mesurable de $\mathbb{R}^n$ dans $\mathbb{R}$ qui aux valeurs observées $(x_1, \ldots, x_n)$ fait corresponde un nombre réel $\hat{\theta} = t(x_1, \ldots, x_n)$ appelé estimation ponctuelle de θ.

Définition 2.1.2

L'estimation ponctuelle se fait à l'aide d'un estimateur, qui est une variable aléatoire d'échantillon. L'estimation est la valeur que prend la variable aléatoire dans l'échantillon observé.

Exemple 2.1.1

La moyenne arithmétique $\overline{X}$ des n observations est un exemple de statistique :

$$\overline{X} = t(X_{1,}, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

2.2 Mode de convergence

Soit $(X_n)_{n\geq 1}$ une suite de $v.a$ réelle définie sur un espace probabilisé $(E, \mathcal{A}, P)$ et X est un $v.a$ réelle définie sur la même espace.

2.2.1 Convergence en probabilité (convergence faible)

Définition 2.2.1

On dit que X_n converge vers X en probabilité, on note $X_n \xrightarrow{p} X$, quand $n \longrightarrow +\infty$ si et seulement si

$$\forall \varepsilon > 0, P\left(\{\omega \in E : |X_n(\omega) - X(\omega)| > \varepsilon\}\right) \longrightarrow 0 \text{ quand } n \longrightarrow +\infty$$

$$\Leftrightarrow \forall \varepsilon > 0, P\left(\{\omega \in E : |X_n(\omega) - X(\omega)| \leq \varepsilon\}\right) \longrightarrow 1 \text{ quand } n \longrightarrow +\infty.$$

Propriété 2.2.1

Soit $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de $v.a$ réelle, et X et Y deux $v.a$ réelles. Si $X_n \xrightarrow{p} X$ et $Y_n \xrightarrow{p} Y$ et si $f : \mathbb{R} \longrightarrow \mathbb{R}$ et $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ deux fonctions continues, alors

1) $f(X_n) \xrightarrow{p} f(X)$.

2) $g(X_n, Y_n) \xrightarrow{p} g(X, Y)$.

Théorème 2.2.1 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a réelle. Qui converge presque sûrement vers X. Alors $X_n \xrightarrow{p} X$.*

La loi faible des grandes nombres

Définition 2.2.2

Soit (X_n) une suite de $v.a$ et *iid* telle que $\mathbb{E}\,\|X_n\| < +\infty$ et $\mathbb{E}\,[X_n] = m$, alors

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} m.$$

2.2.2 Convergence presque-sûre

Définition 2.2.3

On dit qu'une suite $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X, quand $n \longrightarrow +\infty$, si l'ensemble sur lequel $\lim\limits_{n\longrightarrow+\infty} X_n \neq X$ est négligeable. On note $X_n \xrightarrow{p.s} X$ où $n \longrightarrow +\infty$ si et seulement si

$$P\left(\left\{\omega \in E : \lim_{n\longrightarrow+\infty} X_n(\omega) \neq X(\omega)\right\}\right) = 0$$

ou bien

$$P\left(\left\{\omega \in E : \lim_{n\longrightarrow+\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Propriété 2.2.2

Si $X_n \xrightarrow{p.s} X$ et $f : \mathbb{R} \longrightarrow \mathbb{R}$ et une fonction continue si et seulement si $f(X_n) \xrightarrow{p.s} f(X)$.

Théorème 2.2.2 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a réelle. Si $\sum_{n\in\mathbb{N}} P\left(|X_n - X| > \varepsilon\right) < +\infty$ pour tout $\varepsilon > 0$, alors $X_n \xrightarrow{p.s} X$.*

La loi forte des grands nombres

Définition 2.2.4

Soit (X_n) une suite de $v.a$ et iid telle que $\mathbb{E}\left\|X_n\right\| < +\infty$ et $\mathbb{E}\left[X_n\right] = m$, alors

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s} m.$$

2.2.3 Convergence en moyenne quadratique

Définition 2.2.5

Soit $(X_n)_{n\in\mathbb{N}}$ et X de $v.a$ réelle admettant un moment d'ordre 2. On dit que $(X_n)_{n\in\mathbb{N}}$ converge en moyenne quadratique vers X si et seulement si

$$\lim_{n\longrightarrow+\infty} \mathbb{E}\left[|X_n - X|^2\right] = 0$$

on note alors $X_n \xrightarrow{m.q} X$.

Propriétés 2.2.1

1) Soit $(X_n)_{n\in\mathbb{N}}$ une suite des $v.a$ réelle admettant une variance. Si $\mathbb{E}\left[X_n\right] \underset{n\longrightarrow+\infty}{\longrightarrow} m$ et $Var(X_n) \underset{n\longrightarrow+\infty}{\longrightarrow} 0$, alors $X_n \xrightarrow{m.q} m$.

2) Soit $(X_n)_{n\in\mathbb{N}}$ et X de $v.a$ réelle. Si $X_n \overset{m.q}{\longrightarrow} X$, alors $\lim\limits_{n\longrightarrow+\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$ et $\lim\limits_{n\longrightarrow+\infty} Var(X_n) = Var(X)$.

Théorème 2.2.3 *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de* $v.a$ *réelle admettant un moment d'ordre* 2 *et* X *une* $v.a$ *réelle. admettant un moment d'ordre* 2. *Si* $X_n \overset{m.q}{\longrightarrow} X$ *alors* $X_n \overset{p}{\longrightarrow} X$.

2.2.4 Convergence en loi

On dit que la suite des $v.a$ $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si pour tout x en lequel F est continue.

$$\lim_{n\longrightarrow+\infty} F_n(X) = F(X),$$

en écrit $X_n \overset{l}{\longrightarrow} X$.

Remarque 2.2.1

1) X_n et X n'est pas forcement de même loi.

2) La convergence de F_{X_n}vers une fonction F ne suffit pas dire que c'est une convergence en loi, il faut aussi que F_{X_n} soit aussi une fonction de distribution d'une $v.a$ X.

2.3 Estimateur convergent

Définition 2.3.1

Un estimateur $\hat{\theta}$ du paramètre θ est dit convergent si $\hat{\theta}$ converge en probabilité vers θ, *c'est-à-dire*

$$\forall\varepsilon > 0, P\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) \longrightarrow 0 \quad \text{lorsque } n \longrightarrow +\infty$$

on note $\hat{\theta} \overset{p}{\longrightarrow} \theta$ quand $n \longrightarrow +\infty$.

Remarque 2.3.1

On parle d'estimateur fortement convergent lorsqu'on a convergence presque sûr.

Propriété 2.3.1

Soit $\hat{\theta}$ un estimateur convergent du paramètre θ, et g une fonction de $\mathbb{R}$ dans $\mathbb{R}$, continue au point θ. Alors $g\left(\hat{\theta}\right)$ et un estimateur convergent de $g\left(\theta\right)$.

Exemple 2.3.1

Soit $X = (X_1, \ldots, X_n)$ est une $v.a$ suit la loi Normale $N(m, \sigma^2)$ et on a $T = \overline{X}$ est un estimateur de m. D'après la théorème central limite on a :

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

Donc on calcule :

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

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

$$\lim_{n \longrightarrow +\infty} P\left(\left|\overline{X} - m\right| > \varepsilon\right) = 2 - 2\phi(\infty) = 0.$$

Et d'après la définition la statistique $T = \overline{X}$ est un estimateur consiste de θ.

2.4 Estimateur sans biais

Définition 2.4.1

Le biais de l'estimateur $\hat{\theta}$ de θ est la fonction définition sur D_θ par :

$$B(\hat{\theta}) = \mathbb{E}\left[\hat{\theta}\right] - \theta$$

si bien sur $\mathbb{E}\left[\hat{\theta}\right]$ existe.

Définition 2.4.2

On dit qu'un estimateur $\hat{\theta}$ de θ est un estimateur sans biais si :

$$B(\hat{\theta}) = 0 \;\text{ donc }\; \mathbb{E}\left[\hat{\theta}\right] = \theta.$$

Exemple 2.4.1 *(Estimateur de l'espérance)*

La statistique $\overline{X}$ est un estimateur sans biais pour l'espérance mathématique $\mathbb{E}\left[X\right] = m$. En effet, on a $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ donc

$$\begin{aligned}
\mathbb{E}\left[\overline{X}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} m \\
&= m.
\end{aligned}$$

Alors

$$\mathbb{E}\left[\overline{X}\right] = m.$$

Exemple 2.4.2 *(Estimateur de la variance)*

La statistique $S_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ est un estimateur biaisé pour la variance. En effet,

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

Alors, on a

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

Et donc

$$\mathbb{E}\left[S_n^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

donc S_n^2 est un estimateur biaisé de σ^2, et son biais est

$$\begin{aligned} B(S_n^2) &= \mathbb{E}\left[S_n^2\right] - \sigma^2 \\ &= \frac{n-1}{n}\sigma^2 - \sigma^2 \\ &= -\frac{1}{n}\sigma^2. \end{aligned}$$

Définition 2.4.3

On dit qu'un estimateur $\hat{\theta}$ de θ est asymptotique sans biais si et seulement si

$$\lim_{n \longrightarrow +\infty} B(\hat{\theta}) = 0 \Leftrightarrow \lim_{n \longrightarrow +\infty} \mathbb{E}\left[\hat{\theta}\right] = \theta.$$

Exemple 2.4.3

1. Pour S_n^2 (l'estimateur biaisé de σ^2) on a

$$\mathbb{E}\left[S_n^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \quad B(S_n^2) = -\frac{1}{n}\sigma^2,$$

et pour $n \longrightarrow +\infty$ on trouve

$$\begin{aligned} \lim_{n \longrightarrow +\infty} B(S_n^2) &= \lim_{n \longrightarrow +\infty} -\frac{1}{n}\sigma^2 = 0 \\ &\Leftrightarrow \lim_{n \longrightarrow +\infty} \mathbb{E}\left[S_n^2\right] = \sigma^2. \end{aligned}$$

Alors S_n^2 est un estimateur asymptotique sans biais de σ^2.

2. Si on note par $S_n^{*2} = \frac{n}{n-1} S_n^2$ où

$$\begin{aligned} S_n^{*2} &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 \right) \\ &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 . \end{aligned}$$

Et on a

$$\begin{aligned} \mathbb{E}\left[S_n^{*2} \right] &= \mathbb{E}\left[\frac{1}{n} S_n^2 \right] \\ &= \frac{n}{n-1} \mathbb{E}\left[S_n^2 \right] \\ &= \frac{n}{n-1} \frac{n-1}{n} \sigma^2 \\ &= \sigma^2 . \end{aligned}$$

donc S_n^{*2} est un estimateur sans biais de σ^2 est dit estimateur de la variance modifie.

2.5 Risque d'un estimateur (Erreur quadratique moyenne)

Définition 2.5.1

Soit $\hat{\theta}$ un estimateur de θ. on suppose que pour tout $\theta \in D_\theta$, $\hat{\theta}$ admet un moment d'ordre 2 pour la probabilité P.

On appelle erreur quadratique de $\hat{\theta}$ et on note $R(\hat{\theta})$ le réel :

$$R(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta} - \theta \right)^2 \right] \quad \text{tel que } \; \theta \in D_\theta .$$

Propriété 2.5.1

On peut écrire le risque d'un estimateur d'autre façon par :

$$R(\hat{\theta}) = Var(\hat{\theta}) + \left[B(\hat{\theta})\right]^2.$$

Preuve. Notons que l'on

$$\begin{aligned}
R(\hat{\theta}) &= \mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= \mathbb{E}\left[\hat{\theta}^2+\theta^2-2\hat{\theta}\theta\right] \\
&= \mathbb{E}\left[\hat{\theta}^2\right]+\mathbb{E}\left[\theta^2\right]-2\mathbb{E}\left[\hat{\theta}\theta\right] \\
&= \mathbb{E}\left[\hat{\theta}^2\right]+\mathbb{E}\left[\theta^2\right]-2\theta\mathbb{E}\left[\hat{\theta}\right]+\left(\mathbb{E}\left[\hat{\theta}\right]^2-\mathbb{E}\left[\hat{\theta}\right]^2\right) \\
&= \left[\mathbb{E}\left[\hat{\theta}^2\right]-\mathbb{E}\left[\hat{\theta}\right]^2\right]+\theta^2-2\theta\mathbb{E}\left[\hat{\theta}\right]+\mathbb{E}\left[\hat{\theta}^2\right] \\
&= \left[\mathbb{E}\left[\hat{\theta}^2\right]-\mathbb{E}\left[\hat{\theta}\right]^2\right]+\left[\mathbb{E}\left[\hat{\theta}\right]-\theta\right]^2 \\
&= Var(\hat{\theta})+\left[B(\hat{\theta})\right]^2.
\end{aligned}$$

Donc

$$R(\hat{\theta}) = Var(\hat{\theta}) + \left[B(\hat{\theta})\right]^2.$$

■

Remarque 2.5.1

Si l'estimateur $\hat{\theta}$ est sans biais. Alors,

$$R(\hat{\theta}) = Var(\hat{\theta}).$$

2.6 Comparaison de deux estimateurs

Définition 2.6.1

Si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs du même paramètre θ, on dit que $\hat{\theta}_1$ est préférable à $\hat{\theta}_2$ si

$$R(\hat{\theta}_1) < R(\hat{\theta}_2).$$

Définition 2.6.2

Si les estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ sont sans biais, alors on dit que $\hat{\theta}_1$ est de variance minimale si :

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2) \Leftrightarrow R(\hat{\theta}_1) \leq R(\hat{\theta}_2).$$

Exemple 2.6.1

On a deux estimateurs $T = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - m)^2$ et $S_n^{*2} = \frac{n}{n-1}S_n^2$, en effet,

$$\begin{aligned} Var(T) &= \frac{1}{n^2}Var\left(\sum_{i=1}^{n}(X_i - m)^2\right) \\ &= \frac{1}{n}Var\left((X_i - m)^2\right) \\ &= \frac{1}{n}\left[\mathbb{E}\left[X - m\right]^4 - \left[\mathbb{E}\left[X - m\right]^2\right]^2\right] \\ &= \frac{1}{n}\left[m^4 - \sigma^4\right], \end{aligned}$$

et

$$\begin{aligned} Var\left(S_n^{*2}\right) &= \left(\frac{n}{n-1}\right)^2 Var\left(S_n^2\right) \\ &= \left(\frac{n}{n-1}\right)^2 \frac{n-1}{n^3}\left[(n-1)\,m^4 - (n-3)\,\sigma^4\right] \\ &= \frac{1}{n}\left[m^4 - \frac{n-3}{n-1}\sigma^4\right]. \end{aligned}$$

Donc

$$Var(T) < Var\left(S_n^{*2}\right).$$

Alors, si l'espérance mathématique m est connue, l'estimateur T est un estimateur de la variance meilleur que S_n^{*2}.

2.7 Comportement asymptotique d'un estimateur

Définition 2.7.1

Soit un modèle paramètre par θ ($\theta \in \mathbb{R}$) et $(X_1, \ldots, X_n)$ un échantillon de ce modèle *iid* et soit $\hat{\theta}$ est un estimateur de θ dans $\mathbb{R}$ donc

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(\theta, \sigma^2) \text{ d'après le théorème central limite.}$$

Tel que σ^2est la variance asymptotique de $\hat{\theta}$.

Théorème 2.7.1 *D'après la loi des grands nombre et le théorème centrale limite on :*

$$\overline{X} \xrightarrow{p.s} \mu \quad \textit{tel que} \quad n \longrightarrow +\infty.$$

2.8 Estimateur efficace

Définition 2.8.1

Un estimateur sans biais est efficace s'il n'est pas possible de trouver un autre estimateur sans biais qui a une variance plus petit.

2.8.1 Inégalité de Cramer-Rao et efficacité

Définition 2.8.2 *(Inégalité de Cramer-Rao)*

Si $\hat{\theta}$ est un estimateur sans biais de θ, $\mathbb{E}\left[\hat{\theta}\right] = \theta$, on obtient

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

- Si $Var(\hat{\theta}) = \frac{1}{I_n(\theta)}$, l'estimateur est efficace car il a la plus petite variance.

Chapitre 3

Méthodes d'estimation ponctuelle

Il existe plusieurs méthodes pour obtenir des estimation ponctuelles, et dans ce chapitre, on va étudier les deux méthodes le plus utilisée en statistique : La méthode des moments et la méthode de vraisemblance.

3.1 Méthode des moments

La méthode du moment est historiquement l'une des plus anciennes méthodes d'estimation, qui a été présenté par le statisticien *Karl Pearson* en 1894. L'essence de cette méthode est l'égalité entre les moments de la population et ceux de l'échantillon correspondant. Ainsi, nous obtenons un ensemble d'équations en résolvant les paramètres de la population, et nous obtenons les estimations requises, celles-ci sont appelées "*estimations des moments*".

3.1.1 Définition des moments

On appelle moment d'ordre k ($k \in \mathbb{N}$) d'une v.a $X = (X_1, \ldots, X_n)$ est l'espérance mathématique de X^k est définie par :

$$\begin{aligned} m_k &= \mathbb{E}\left[X^k\right] \\ &= \begin{cases} \sum\limits_{x \in E} x^k P_X(x) & \text{si } x \text{ discrèt} \\ \int\limits_{\mathbb{R}} x^k f_X(x)dx & \text{si } x \text{ continue.} \end{cases} \end{aligned}$$

Remarque 3.1.1

i) L'espérance mathématique de la v.a X est le moment d'ordre 1 de X

$$m = \mathbb{E}\left[X\right].$$

ii) La variance (moment centré d'ordre 2) de X

$$\begin{aligned} Var(X) &= \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right] \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2. \end{aligned}$$

3.1.2 Principe de la méthode

Supposons que l'application h définie de $D_\theta \subset \mathbb{R}^d$ dans $h(D_\theta) \subset \mathbb{R}^d$ est bijective et continue, et soit $\varphi : E \to \mathbb{R}^d$ tell que $\mathbb{E}\left[\varphi(X)\right]$ existe

$$\begin{aligned} &h : D_\theta \longrightarrow \mathbb{R}^d \\ &\theta \longrightarrow h(\theta) = \mathbb{E}\left[\varphi(X)\right] \quad \text{pour tout } \theta \in D_\theta. \end{aligned}$$

La méthode des moments consiste a estimer θ par :

$$
\begin{aligned}
h(\theta) &= \mathbb{E}\left[\varphi(X)\right] \\
&\Longrightarrow \widehat{h}(\theta) = h(\widehat{\theta}) = \mathbb{E}\left[\varphi(X)\right] = \frac{1}{n}\sum_{i=1}^{n}\varphi(X_i) \\
&\Longrightarrow \widehat{\theta} = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i)\right).
\end{aligned}
$$

Exemple 3.1.1

Soit X est une v.a suit la loi exponentielle de paramètre λ de densité

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

On suppose que l'application h de $D_\lambda \subset \mathbb{R}$ dans $h(D_\lambda) \subset \mathbb{R}$ est une fonction bijective et continue, et soit $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ tell que $\varphi(X) = X$. Alors

$$
\begin{aligned}
h(\lambda) &= \mathbb{E}\left[\varphi(X)\right] = \mathbb{E}\left[X\right] \\
&= \int_0^\infty x f_\lambda(x)dx \\
&= \int_0^\infty x(\lambda\exp(-\lambda x))dx.
\end{aligned}
$$

On pose $u = x \Longrightarrow u' = dx$ et $v' = \lambda \exp(-\lambda x) \Longrightarrow v = -\exp(-\lambda x)$, alors

$$\begin{aligned}
\mathbb{E}[X] &= -x\exp(-\lambda x)|_0^\infty - \int_0^\infty -\exp(-\lambda x)dx \\
&= \int_0^\infty \exp(-\lambda x)dx \\
&= -\frac{\exp(-\lambda x)}{\lambda}\Big|_0^\infty \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Donc

$$\mathbb{E}[X] = \frac{1}{\lambda} \Longrightarrow \hat{\lambda} = \frac{1}{\mathbb{E}[X]}.$$

Et on a

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Alors

$$\hat{\lambda} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{1}{\overline{\overline{X}}}.$$

Exemple 3.1.2

Soit X suit une loi normale $\{N(m,\sigma^2), m \in \mathbb{R}\}$ .On suppose que l'application h de $D_\theta \subset \mathbb{R}$ dans $h(D_\theta) \subset \mathbb{R}^d$ est une fonction bijective et continue et soit $\varphi : X \to \varphi(X) = (X, X^2)$, on pose $\theta = (m, \sigma^2)$. Donc,

$$\begin{aligned}
h(\theta) &= \mathbb{E}[\varphi(X)] \\
&= \mathbb{E}\left[(X, X^2)\right] \\
&= (\mathbb{E}[X], \mathbb{E}\left[X^2\right]).
\end{aligned}$$

Tel que

$$\mathbb{E}\left[X\right] = m \ \text{ et } \ \mathbb{E}\left[X^2\right] = \sigma^2 - m^2.$$

Alors

$$\begin{aligned} h(\theta) &= h(m, \sigma^2) \\ &= (\mathbb{E}\left[X\right], \mathbb{E}\left[X^2\right]) \\ &= (m, \sigma^2 - m^2). \end{aligned}$$

$$\Rightarrow (\hat{m}, \hat{\sigma}^2) = h^{-1}(\hat{E}(X), \hat{E}(X^2)) = h^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} X_i^2).$$

Et on pose

$$\begin{cases} u = m \\ v = \sigma^2 - m^2 \end{cases} \Rightarrow \begin{cases} m = u \\ \sigma^2 = v - u^2. \end{cases}$$

Alors

$$h^{-1}(u, v) = (u, v - u^2).$$

Donc

$$(\hat{m}, \hat{\sigma}^2) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2\right).$$

3.1.3 Quelque estimateur par la méthode de MM

On dit en général que les capacités de la méthode du moment ne sont pas efficaces. C'est pourquoi il est souvent utilisé en première approximation pour trouver des capacités plus efficaces que l'on peut obtenir par d'autres moyens, par exemple la méthode du maximum de vraisemblance.

Loi	Estimateur
$X \sim U([a,b])$	$\hat{a} = \overline{X} - \sqrt{3\frac{n-1}{n}S^2}; \hat{b} = \overline{X} + \sqrt{3\frac{n-1}{n}S^2}$
$X \sim U([0,b])$	$\hat{b} = 2\overline{X}$
$X \sim B(n,p)$	$\hat{p} = \frac{\overline{X}}{n}$
$X \sim P(\lambda)$	$\hat{\lambda} = \overline{X}$
$X \sim N(m,\sigma^2)$	$\hat{m} = \overline{X}; \hat{\sigma}^2 = \frac{n}{n-1}S^2$
$X \sim N(0,\sigma^2)$	$\hat{\sigma}^2 = \hat{E}(X^2)$
$X \sim Exp(\lambda)$	$\hat{\lambda} = \frac{1}{\overline{X}}$
$X \sim G(p)$	$\hat{p} = \frac{1}{\overline{X}}$

TAB. 3.1 – Estimateur obtenu par la méthode du moment

3.2 Méthode du maximum de vraisemblance

Dans deux articles publiés en 1920, En eux, le scientifique *Fischer*, l'un des pionniers des statisticiens de notre temps, a présenté une méthode générale d'estimation, qu'il a appelée la "*méthode du maximum de vraisemblance*" ou "*la plus grande probabilité*", et il a également expliqué les avantages de cette méthode. La méthode maximum de vraisemblance est considérée comme l'une des méthodes statistiques les plus importantes et les plus répandues pour estimer les paramètres de la distribution statistique de probabilité, le "*modèle statistique*" proposé.

Soit X une v.a réelle de loi paramétrique (discrète ou continue), dont on veut estimer le paramètre θ par la méthode de MV.

3.2.1 Principe de la méthode

La méthode de maximum de vraisemblance consistant à estimer θ par la valeur qui maximise L (fonction de vraisemblance)

$$\hat{\theta} = \left\{ \theta \in D_\theta \middle/ L(\hat{\theta}) = \sup_\theta L(\theta) \right\}.$$

Ceci est le problème d'optimisation. On utilise généralement le fait que si L est dérivable et si L admet un maximum global en une valeur, alors la dérivée première s'annule en et que la dérivée seconde est négative.

Réciproquement, si la dérivée première s'annule en $\theta = \hat{\theta}$, et que la dérivée seconde est négative en $\theta = \hat{\theta}$, alors $\hat{\theta}$ est un maximum local (et non global) de $L(x_1, \ldots, x_n; \theta)$. Il est alors nécessaire de vérifier qu'il s'agit bien d'un maximum global. La vraisemblance étant positive et le logarithme népérien une fonction croissante, il est équivalent et souvent plus simple de maximiser le logarithme népérien de la vraisemblance (le produit se transforme en somme, ce qui est plus simple à dériver).

Ainsi en pratique :

1) La condition nécessaire

$$\frac{\partial L(x_1, \ldots, x_n; \theta)}{\partial \theta} = 0 \quad \text{ou} \quad \frac{\partial \ln L(x_1, \ldots, x_n; \theta)}{\partial \theta} = 0.$$

Permet de trouver la valeur $\hat{\theta}$

2) $\theta = \hat{\theta}$ est un maximum local si la condition suffisante est remplie au point critique :

$$\frac{\partial^2 L(x_1, \ldots, x_n; \theta)}{\partial \theta^2}(\hat{\theta}) \leq 0 \quad \text{ou} \quad \frac{\partial^2 \ln L(x_1, \ldots, x_n; \theta)}{\partial \theta^2}(\hat{\theta}) \leq 0.$$

Exemple 3.2.1

Soit X suit la loi de Bernoulli de paramètre p tell que $P_X(x; p) = p^x(1-p)^{1-x}$, donc ce vraisemblance est :

$$L(x_1, \ldots, x_n; p) = \prod_{i=1}^{n} \left(p^{x_i}(1-p)^{1-x_i}\right).$$

Alors

$$\begin{aligned}\ln L(x_1,\ldots,x_n;p) &= \ln\left(\prod_{i=1}^{n}\left(p^{x_i}(1-p)^{1-x_i}\right)\right)\\ &= \sum_{i=1}^{n}\left(x_i\ln p+(1-p)\ln(1-p)\right).\end{aligned}$$

On pose

$$\begin{aligned}\Psi(p) &= \ln L(x_1,\ldots,x_n;p)\\ &= \ln(p)\left(\sum_{i=1}^{n}x_i\right)+\ln(1-p)\sum_{i=1}^{n}(1-x_i).\end{aligned}$$

La première dérivé est

$$\Psi'(p)=\frac{\sum\limits_{i=1}^{n}x_i}{p}-\frac{\sum\limits_{i=1}^{n}(1-x_i)}{(1-p)}.$$

Le deuxième dérivé est

$$\Psi''(p)=-\frac{\sum\limits_{i=1}^{n}x_i}{p^2}-\frac{\sum\limits_{i=1}^{n}(1-x_i)}{(1-p)^2}<0.$$

Tell que Ψ'' négative donc elle est admet une maximum, en effet

$$\begin{aligned}\Psi'(p)&=0\\ &\Rightarrow\frac{\sum\limits_{i=1}^{n}x_i}{\hat{p}}-\frac{\sum\limits_{i=1}^{n}(1-x_i)}{(1-\hat{p})}=0\\ &\Rightarrow(1-\hat{p})\sum_{i=1}^{n}x_i-\hat{p}\sum_{i=1}^{n}(1-x_i)=0\\ &\Rightarrow\sum_{i=1}^{n}x_i-\hat{p}n=0\\ &\Rightarrow\hat{p}=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{X_n.}\end{aligned}$$

Exemple 3.2.2

On souhaite estimer le paramètre λ d'une loi de Poisson à partir d'un n-échantillon.

On a $f(x;\lambda) = P_X(x;\lambda) = \exp(-\lambda)\frac{\lambda^x}{x!}$. La fonction de vraisemblance s'écrit

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

donc

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

La dérivée première du logarithme de la vraisemblance est

$$\begin{aligned}\frac{\partial \ln L(x_1,\ldots,x_n;\lambda)}{\partial\lambda} &= -n + \frac{\sum_{i=1}^{n}x_i}{\lambda}\\ &\Rightarrow -n + \frac{\sum_{i=1}^{n}x_i}{\lambda} = 0.\end{aligned}$$

La dérivée seconde

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

Alors elle est admet un maximum et

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

Donc l'estimateur du paramètre λ est

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Théorème 3.2.1 *S'il existe un estimateur efficace du paramètre θ, donc il est l'estimateur de maximum de vraisemblance de ce paramètre.*

3.2.2 Quelque estimateur par la méthode de MV

Loi	Estimateur
$X \sim U([0,b])$	$\hat{b} = \sup\limits_{i\in[1,n]} (x_i)$
$X \sim B(n,p)$	$\hat{p} = \frac{k}{n}$
$X \sim P(\lambda)$	$\hat{\lambda} = \overline{X}$
$X \sim N(m,\sigma^2)$	$\hat{m} = \overline{X}, \hat{\sigma} = \frac{n}{n-1}S^2$
$X \sim Exp(\lambda)$	$\hat{\lambda} = \frac{1}{\overline{X}}$

TAB. 3.2 – Estimateur obtenu par la méthode du MV

3.2.3 Propriétés sur la méthode de MV

1) L'EMV n'existe pas toujours et s'elle existe n'a aucune raison d'être unique.

2) S'il existe un estimateur efficace de θ, alors il est égal à l'unique EMV de θ. Mais la réciproque est fausse.

3) Si $\hat{\theta}$ est un estimateur de maximum de vraisemblance de θ, $f(\hat{\theta})$ est l'estimateur de maximum de vraisemblance de $f(\theta)$.

4) L'EMV est un estimateur convergent si la taille de l'échantillon est grande, l'EMV devient unique, et tend vers la vraie valeur du paramètre θ

$$\hat{\theta}_n \xrightarrow{p} \theta.$$

5) L'EMV est asymptotiquement efficace et asymptotiquement normale si $n \geq 30$.

6) Il n'y aucune raison pour que l'EMV soit sans biais.

Exemple 3.2.3

Soit $X \sim U\left([0,\theta]\right)$ tell que $\theta > 0$ alors

la fonction de vraisemblance est

$$\begin{aligned} L(x_1,\ldots,x_n;\lambda) &= \prod_{i=1}^{n} \frac{1}{\theta} 1_{[0,\theta]}\left(X_i\right) \\ &= \frac{1}{\theta^n} 1_{0 \leq X_{(1)} \leq X_{(n)} \leq \theta} \\ &= \frac{1}{\theta} 1_{\left[\sup\limits_{1\leq i \leq \theta} X_i, \infty\right[} (\theta) \end{aligned}$$

et donc l'EMV est $T = \sup\limits_{1\leq i\leq n} X_i$ car $1_{0\leq \inf\limits_{1\leq i\leq n} X_i} = 1$ $p.s.$ On peut montrer que $T\,/\theta \rightsquigarrow Beta(n,1)$, c'est-à-dire la loi de la $v.a$ T admet pour densité

$$f(x,\theta) = \frac{nx^{n-1}}{\theta} 1_{0\leq x\leq \theta}.$$

Il s'en suit que $\mathbb{E}\left[T/\theta\right] = \frac{n}{n+1}$ et $B(T) = \mathbb{E}\left[T\right] - \theta = \frac{-\theta}{n+1} \neq 0$ donc l'EMV est ici biaisé.

Conclusion

Le but de l'estimation ponctuelle est de choisir, parmi toutes les statistiques possibles, le mieilleur estimateur , c'est-à-dire celui qui donnera une estimation ponctuelle la plus proche possible du paramètre et ceci, quel que soit l'échantillon.

Bibliographie

[1] Bigot, J. (2014). Notes de cours de probabilités.

[2] Boulay, J-P. (2010). Statistique Mathématique, Applications commentées, Ellipses, Paris.

[3] Boziane, N. (2005). Mémoire Magister, Efficacité-Robustesse des estimateurs des modèles paramètrique. Université de Biskra, Algérie.

[4] Concordet, D. Introduction à la statistique inférentielle.Ecole Vétérinaire, Toulous.

[5] Delyon, B. (2021). Cours de Master 2, Estimation paramétrique. Université Rennes I, France.

[6] Druilhet, P. (2004-2005).Support du cours de statistique inférentielle. ENSAI.

[7] Dusart, P. (2018). cours de Licence 2, Statistique inférentielles.

[8] Gannaz, I. Cours de quatrième année, Introduction à la statistique. INSA, Lyon.

[9] Goldford, B. Pardoux, C. (2011). Introduction à la méthode statistique, Mannel et exercice corrigés. DUNOD, Paris.

[10] Lejeune, M. (2010). Statistique, La théorie et ses application (2^e édition). Springer, Paris.

[11] Lethielleux, M. (2006).Probabilités, Estimation statistique (3^e édition). DUNOD, Paris.

[12] Mustafa, A.H. (1999). Inférence statistique -1-, Théorie de l'estimation (en arabe). Université Nasser, Misurata, Libye.

[13] Polisano, K. (2018). Cours de statistique, niveau L1-L2 licence. HAL, france.

[14] Rancourt, F. Sur l'estimation ponctuelle pour des méthodes de mélange. Université de Sheerbrooke, Canada.

[15] Rosseau, J. (2004-2005). statistique inférentielle.

[16] Ruch, J-J. (2012-2013). statistique : Estimation.

[17] Saporta, G. (2006). Probabilités analyse des données et statistique. Editions TECHNIP, Paris.

[18] Veysseyre, R. (2006). Aide-mémoire, Statistique et probabilités pour l'ingénieur (2^e edition). DUNOD, Paris.

[19] Ycart, B. (2002). Estimation pararitrique tests statistique. Centre de publication universitaire, tunis.

Annexe A : Abréviations et Notations

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

F	Fonction de répartition.
f	Fonction de densité.
E	L'ensemble d'observation.
D	L'ensemble des paramètres.
$\overline{X}$	Moyenne empirique.
$\mathbb{E}(.)$	Espérence mathématique.
$Var(.)$	Variance mathématique.
P	Distribution de X.
$v.a$	Variable aléatoire.
iid	Indépendantes et identiquement distribuées.
$I(.)$	Quantité d'information de Fisher.
$\xrightarrow{p}$	Convergence en probabilité.
$\xrightarrow{p.s}$	Convergence presque-sûre.
$\xrightarrow{m.p}$	Convergence en moyenne quadratique.
$\xrightarrow{l}$	Convergence en loi.

$\mathcal{A}$	Tribu.
$B(.)$	Le biais.
S^2	Variance empirique.
S^{*2}	Variance empirique modifie.
$R(.)$	Risque d'un estimateur.

Résumé

Ce travail, porte sur l'estimation ponctuelle. Estime un paramètre, c'est donner une valeur approchée de ce paramètre, à partir des résultats obtenus sur un échantillon aléatoire extrait de la population.

Mots clés: Estimateur, statistique exhaustive, information de Ficher, Maximum de vraisemblance.

Abstract

This work focuses on point estimation. Estimate a parameter, that is, give an approximate value of this parameter, from the results obtained on a random sample taken from the population.

Keywords: Estimator, exhaustive statistics, Fisher information, maximum likelihood.

ملخص

يركز هذا العمل على التقدير النقطي. تقدير المعامل، أي إعطاء قيمة تقريبية لهذه المعلمة، انطلاقا من نتائج مأخوذة من عينة عشوائية مستخرجة من المجتمع.

كلمات مفتاحيه: مقدر، إحصاء شامل، معلومة فيشر، المعقولية العظمى.