

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



Mémoire présenté pour obtenir le diplôme de

Master en “**Mathématiques Appliquées**”

Option : Statistique

Par

HAMMOUDI Lamia

Titre :

Estimation par maximum de vraisemblance

Devant le Jury :

Mr.	Brahimi Brahim	Pr.	U. Biskra	Président
Mme.	Touba Sonia	Dr.	U. Biskra	Encadreur
Mme.	Roubi Afaf	Dr.	U. Biskra	Examinatrice

Soutenu Publiquement le 27/06/2022

Dédicace

Je dèdie ce humble travail à :

A ma très chère maman "Nadjia"

*Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se
doit. Tes prières me protègent,
et ta présence à cotés de moi a toujours été ma source de force pour affronter les
différents obstacles.*

A mon papa "Ramdanne"

A mes chères sœurs "Maroua", "Rima"

A mon frère "Housseem"

A mes amis.

Remerciements

*Tout d'abord je tiens à remercier "**Allah Le Tout Puissant**" de m'avoir donné le courage, la volonté et la santé pour mener à bien ce travail.*

Je tiens à remercier le Dr. **Touba Sonia** pour ses remarques tout au long de la réalisation de mon mémoire.

*Je remercie aussi les membres du jury Pr. **Brahimi Brahim** et Dr. **Roubi Afaf** d'avoir accepté d'évaluer et d'examiner ce travail.*

Je tiens aussi à remercier Dr **Yahia Djabrane** et Dr **Berkane Hassiba**, pour leurs aides et conseils.

Je remercie chaleureusement toute ma famille "**Mamam,Papa,Maroua,Rima et houssem**" et mes amis "**Sabrina,Mofida**" pour leur soutien, leur encouragement et leur amour illimité.

Notations et symbols

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

Notation	Signification
$E(x)$: espérance de la v.a X
$Var(x)$: variance de la v.a X
cov	: covariance.
f	: Fonction de densité.
$\hat{\theta}$: estimateur de θ .
$I(\theta)$: L'information.
$F_n(\cdot)$: Fonction de répartition empirique.
EMV	: Estimateur du maximum de vraisemblance.
$FDCR$: Frechet-Darמוש-Cramer-Rao.
$MRLM$: Model régression linéaire multiple.
GEV	: distribution des valeurs extrêmes généralisés.
\sum	: somme
B_{T_n}	: biais d'un estimateur T_n .
P_θ	: loi de probabilité de X si le paramètre θ .

$N(\mu, \sigma^2)$: loi normale
$B(p)$: loi de Bernoulli
$\exp(\lambda)$: loi exponentielle
(X_1, \dots, X_n)	: échantillon de taille n .
(x_1, \dots, x_n)	: l'échantillon X
T_n	: estimateur.
$G(p)$: loi géométrique.
$\mathcal{U}(0, \theta)$: loi uniforme.
$\mathcal{E}(\lambda)$: loi de poisson.
\bar{X}	: Moyenne empirique.
<i>iid</i>	: indépendantes identiquement distribuées.
<i>eff</i>	: efficacité.
$\Phi_X(\cdot)$: fonction caractéristique.
$\xrightarrow{p.s}$: convergence presque sûre
\xrightarrow{loi}	: Convergence en loi.
<i>va</i>	: variable aléatoire.
Ψ	: diagama.
Γ	: gama.
<i>c – à – d</i>	: C'est-à-dire.
\sim	: suit la loi.
\rightarrow	: converge vers.

Table des matières

Dédicace	i
Remerciements	ii
Notations et symboles	iii
Table des matières	v
Liste des tableaux	viii
Introduction	1
1 Estimation paramétrique	4
1.1 Modèle paramétrique	4
1.2 Estimateur	5
1.3 Propriétés des estimateurs	6
1.3.1 Estimateur avec biais	6
1.3.2 Estimateur sans biais	6
1.3.3 Estimateur asymptotiquement sans biais	7
1.3.4 Convergence estimateur	7

1.4	Estimateur efficace	8
1.5	Estimateur optimal	8
1.5.1	Ecart quadratique moyenne	8
1.5.2	Estimateur à variance minimal	9
1.6	Exhaustivité d'un estimateur	11
1.6.1	Théorème de Darmois	12
1.6.2	Famille exponentielle	13
1.7	Amélioration d'un estimateur	13
1.7.1	Théorème de Rao-Blackwell	13
1.7.2	Théorème de Lehmann-Scheffe	14
1.8	Quelques estimateurs classiques	14
2	Estimation par maximum des vraisemblance	15
2.1	Introduction	15
2.2	Approche générale	16
2.2.1	Méthode du maximum de vraisemblance	16
2.2.2	Cas de variable de discrète	16
2.2.3	Cas de variable continue à densité	17
2.3	Fonction de vraisemblance	19
2.3.1	Fonction de vraisemblance : contexte théorique	20
2.4	Fonction de log-vraisemblance	21
2.5	Estimateurs du maximum de vraisemblance	21
2.5.1	Cas estimation ponctuelle d'un paramètre	21
2.5.2	Cas estimation simultanée de plusieurs paramètres	22

2.6 Exemples de estimateurs du maximum de vraisemblance	24
2.6.1 Loi uniforme	24
2.6.2 Loi Bernoulli	24
2.6.3 loi Géométrique	25
2.6.4 loi Exponentielle	26
2.6.5 loi Pareto	27
2.7 Quantité d'information de Fisher	28
2.7.1 Information de Fisher d'un échantillon	32
2.7.2 Modèle régulier	32
2.7.3 Inégalité de Fréchet-Darmois-Cramer-Rao (FDCR)	33
2.8 Propriétés du maximum de vraisemblance	34
2.9 Estimation d'un modèle de régression linéaire	35
2.9.1 Estimateur du maximum de vraisemblance (ML)	36
2.10 Estimation des paramètres de la loi généralisée des valeurs extrêmes 47	
2.10.1 Loi GEV avec $k = 0$ (EVI, loi Gumbel) :	47
2.10.2 Loi GEV avec $k \neq 0$ (EV2 et EV3) :	50
2.11 Estimation par la méthode des moments	54
2.11.1 Application (simulation)	55
Conclusion	59
Bibliographie	61

Liste des tableaux

2.1	Comparaison entre l'EMV et l'EMM de la loi béta	57
2.2	Comparaison entre l'EMV et l'EMM de la loi exponentielle	58

Introduction

En 1912, au moment où Ronald Aylmer Fisher rédige son premier article consacré au maximum de vraisemblance, les deux méthodes statistiques les plus utilisées sont la méthode des moindres carrés et la méthode des moments. Dans son article de 1912, il propose l'estimateur du maximum de vraisemblance qu'il appelle à l'époque le critère absolu. Il prend l'exemple d'une loi normale.

En 1921, il applique la même méthode à l'estimation d'un coefficient de corrélation.

En 1912, un malentendu a laissé croire que le critère absolu pouvait être interprété comme un estimateur bayésien avec une loi a priori uniforme. Fisher réfute cette interprétation en 1921.

En 1922, il utilise la loi binomiale pour illustrer son critère et montre en quoi il est différent d'un estimateur bayésien. C'est aussi en 1922, qu'il donne le nom de maximum de vraisemblance à sa méthode.

La méthode du maximum de vraisemblance est une méthode d'estimation paramétrique qui doit sa popularité à

- ★ la simplicité de son approche,
- ★ sa faculté d'adaptation à une modélisation complexe, i.e. une loi $\mathcal{L}(\theta)$ où θ symbolise une multitude de paramètres inconnus : $\theta = (\theta_1, \theta_2, \dots, \theta_n)^t$,

★ l'aspect numérique accessible grâce à l'application de méthodes d'optimisation connues.

Elle permet de

- construire des estimateurs performants,
- construire des intervalles de confiance précis,
- mettre en œuvre des tests statistiques "puissants".

Dans la plupart des cas d'intérêt pratique, la loi P_θ , et donc aussi la vraisemblance, ont une expression dérivable par rapport à θ . Pour calculer le maximum de la vraisemblance, il faut déterminer les valeurs pour lesquelles la dérivée de la vraisemblance s'annule.

Or par définition, la vraisemblance est un produit de probabilités ou de densités, qui peut être assez compliqué à dériver.

Il est préférable de dériver une somme, et c'est pourquoi on commence par remplacer la vraisemblance par son logarithme.

La fonction logarithme étant croissante, il est équivalent de maximiser $\log(L(x_1, x_2, \dots, x_n; \theta))$ ou $L(x_1, x_2, \dots, x_n; \theta)$.

Une fois déterminée une valeur de θ pour laquelle la dérivée s'annule, il faut s'assurer à l'aide de la dérivée seconde que ce point est bien un maximum.

Alors, Donc ce travail est déroulé en deux parties :

- Chapitre 1 : appelé " Estimation paramétrique", dans la première partie de ce chapitre nous commençons par un rappel sur le modèle paramétrique, les propriétés des estimateurs, l'efficacité et optimalité d'un estimateur, et nous introduisons aussi des définitions sur l'exhaustivité d'un estimateur avec le théorème de Darmois et la famille exponentielle et amélioration. A la fin nous avons présenté quelques estimateurs classiques.

- Chapitre 2 : appelé " Estimation par maximum de vraisemblance", dans la première partie nous allons en posé cette méthode dans les deux cas, le cas discret et le cas densité.

Une première approche porte sur l'estimation de paramètre et plusieurs paramètre par cette méthode, en plus nous donnons quelques exemples par la maximum de vraisemblance. Dans la seconde approche nous présentons la quantité d'information de Fisher apportée par un échantillon et modèle régulier et étudions l'inégalité de Frechet-Darmois-Cramer-Rao (FDCR) et dans la troisième approche nous intéressé aux propriétés de estimation par maximum de vraisemblance on définie la convergence et la asymptotiques normale de l'EMV en plus invariance du M.V.et la quatrièmement approche, nous montrons comment estimer les paramètres d'un modèle de régression linéaire simple. Enfin, dans la dernière section de ce chapitre on s'intéresse aux estimations des paramètres de loi généralisée des valeurs extrêmes (GEV) par la méthode de maximum vraisemblance pour les deux cas. Enfin, une comparaison, sera faire entre les deux méthodes ML et la méthode des moments pour quelques lois de probabilités.

Chapitre 1

Estimation paramétrique

L'estimation paramétrique qui considère que les modèles sont connues avec des paramètres inconnus notés.

L'ensemble des valeurs possibles pour, appelé espace paramétrique, sera noté Θ , lequel est inclus dans $\mathbb{R}^{\{d\}}$ où d est la dimension du paramètre θ .

La loi de la variable étudiée est supposée appartenir à une famille de lois pouvant être caractérisée par une forme fonctionnelle connu.

1.1 Modèle paramétrique

Soit $X = (x_1, \dots, x_n)$ un échantillon des données où les variable aléatoires X_i sont indépendante et identiquement distribuées.

Définition 1.1.1 *On définit loi commune est dans une famille de probabilités*

$$P = \{P_\theta, \theta \in \Theta\} \text{ où } \theta \in \mathbb{R}^d.$$

avec : P_θ est la loi de probabilité de X si le paramètre vaut θ .

-Le modèle Gaussien :

$$P = \{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

-Le modèle Bernoulli :

$$P = \{B(p), p \in [0, 1]\}.$$

-Le modèle exponentiel :

$$P = \{\exp(\lambda), \lambda \in \mathbb{R}_+^*\}, \text{ avec } \theta = \lambda \text{ et } \Theta = \mathbb{R}_+^*.$$

1.2 Estimateur

Définition 1.2.1 *Un estimateur de θ est une application T_n de E dans F qui à un échantillon (X_1, \dots, X_n) de la loi P_θ associe une variable aléatoire réelle,*

$$T_n : E \longrightarrow F$$

$$X_1, X_2, \dots, X_n \longmapsto T_n(X_1, X_2, \dots, X_n).$$

On a $\theta = E(X)$, c'est-à-dire la moyenne théorique de la loi, et on retient donc très logiquement comme estimateur du paramètre θ la moyenne empirique ou moyenne de l'échantillon :

$$T_n(X_1, X_2, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Exemple 1.2.1 *Si X suit une loi de Bernoulli $B(1, \theta)$, T_n sera un estimateur strict si $0 \leq T_n(x_1, x_2, \dots, x_n) \leq 1$ pour tout échantillon observé (x_1, x_2, \dots, x_n)*

1.3 Propriétés des estimateurs

Les propriétés peuvent bien sûr s'étendre au cas d'un paramètre multidimensionnel, alors nous allons étudier des propriétés ci-après où supposerons pour simplifier que θ est un paramètre réel, c'est-à-dire que $\theta \in \mathbb{R}$.

1.3.1 Estimateur avec biais

Définition 1.3.1 On dit biais d'un estimateur T_n au variable aléatoire :

$$\forall \theta \in \Theta \quad B_{T_n}(\theta) = E(T_n) - \theta.$$

1.3.2 Estimateur sans biais

Définition 1.3.2 On dit que T_n est un estimateur sans biais de θ si l'espérance mathématique de l'estimateur est égale à la vraie valeur du paramètre θ :

$$\forall \theta \in \Theta \quad E(T_n) = \theta.$$

alors :

$$B_{T_n}(\theta) = 0.$$

1.3.3 Estimateur asymptotiquement sans biais

Définition 1.3.3 On dit qu'un estimateur T_n de θ est asymptotiquement sans biais ssi :

$$\lim_{n \rightarrow \infty} E_{\theta}(T_n) = \theta$$

avec $\lim_{n \rightarrow \infty} B_{T_n}(\theta) = 0 \quad \forall \theta \in \Theta.$

1.3.4 Convergence estimateur

Définition 1.3.4 Un estimateur T_n est convergent si pour tout $\varepsilon > 0$ la suite de variables aléatoire (T_n) converge en probabilité vers la valeur du paramètre :

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|T_n - \theta| \leq \varepsilon) = 1$$

$$\Leftrightarrow \lim_{n \rightarrow \infty} P(|T_n - \theta| < \varepsilon) = 0 \quad .$$

Définition 1.3.5 Une suite de variables aléatoires $\{T_n\}_{n \geq 1}$ converge en loi vers la loi de probabilité de fonction de répartition F si et seulement si

$$\lim_{n \rightarrow \infty} F_{T_n}(x) = F(x).$$

en tout point x où F est continue.

Définition 1.3.6 Une suite de variables aléatoires $\{T_n\}_{n \geq 1}$ de converge en moyenne quadratique vers la variable aléatoire X si et seulement si

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

1.4 Estimateur efficace

Définition 1.4.1 On appelle efficacité d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial \theta} E[T_n]\right]^2}{I_n(\theta) Var_{\theta}(T_n)}.$$

On a $0 \leq Eff(T_n) \leq 1$

- T_n est dit un estimateur efficace si et seulement si $Eff(T_n) = 1$.

- T_n est dit asymptotiquement efficace si et seulement si $\lim_{n \rightarrow \infty} Eff(T_n) = 1$.

- Si T_n est un estimateur sans biais de θ ,

$$Eff(T_n) = \frac{1}{I_n(\theta) Var_{\theta}(T_n)}.$$

1.5 Estimateur optimal

1.5.1 Ecart quadratique moyenne

Définition 1.5.1 On appelle l'écart quadratique de T_n par rapport à θ , définie par :

$$Eqm_{\theta}(T_n) = E_{\theta} [(T_n - \theta)^2].$$

Théorème 1.5.1 Soit T_n un estimateur du paramètre θ à étudier. On a :

$$E_{\theta} [(T_n - \theta)^2] = [B_{\theta}(T_n)]^2 + V_{\theta}(T_n).$$

En effet

$$\begin{aligned} E_{\theta} [(T_n - \theta)^2] &= E_{\theta} [\{T_n - E_{\theta}(T_n) + E_{\theta}(T_n) - \theta\}^2] \\ &= E_{\theta} [\{T_n - E_{\theta}(T_n)\}^2] + [E_{\theta}(T_n) - \theta]^2 + 2E_{\theta}[T_n - E_{\theta}(T_n)][E_{\theta}(T_n) - \theta] \\ &= [B_{\theta}(T_n)]^2 + V_{\theta}(T_n). \end{aligned}$$

car le terme $E_{\theta}[T_n - E_{\theta}(T_n)]$ est nul.

1. Ainsi pour rebdre l'écart quadratique moyen $E_{\theta} [(T_n - \theta)^2]$ le plus petit possible.
2. Il faut que :
 - $E(T_n) = 0$, donc choisir un estimateur sans biais.
 - La variance $V_{\theta}(T_n)$ soit faible.

1.5.2 Estimateur à variance minimal

Définition 1.5.2 *Un estimateur de θ est dit de variance minimale si, parmi tous les estimateurs possibles de θ il a la plus petite variance :*

Considérant qu'on se place dorénavant dans la classe des estimateurs sans biais, on pourra comparer deux estimateurs T_n et S^{*2} de même classe par leur variance, qui mesure alors leur dispersion par rapport au paramètre, qui est leur espérance commune. Nous dirons que l'estimateur T_n est plus efficace que S^{*2} si pour tout θ et pour une taille d'échantillon $n \succ N$:

$$V_{\theta}(T_n) \leq V_{\theta}(S^{*2}) \quad \forall \theta \in \Theta.$$

avec :

$$T_n = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2 \text{ et } S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Preuve. Montrons que m est connu l'estimateur $T_n = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ est meilleur que S^{*2} :

En effet :

$$\begin{aligned} V_\theta(T_n) &= \frac{1}{n^2} V_\theta\left(\sum_{i=1}^n (X_i - m)^2\right) \\ &= \frac{1}{n} V_\theta[(X - m)^2] \\ &= \frac{1}{n} \left[E(X - m)^4 - [E(X - m)^2]^2 \right] \\ &= \frac{1}{n} [\mu_4 - \sigma^4]. \end{aligned}$$

et :

$$\begin{aligned} V_\theta(S^{*2}) &= \left(\frac{n}{n-1}\right)^2 V_\theta(S^2) \\ &= \left(\frac{n}{n-1}\right)^2 \frac{n-1}{n^3} [(n-1)\mu_4 - (n-3)\sigma^4] \\ &= \frac{1}{n} \left[\mu_4 - \frac{n-3}{n-1}\sigma^4 \right]. \end{aligned}$$

donc $V_\theta(T_n) \leq V_\theta(S^{*2})$. ■

Théorème 1.5.2 *Si nous considérons tous les estimateurs possibles dont les distributions d'échantillonnage ont la même moyenne, l'estimateur le meilleur est celui dont la variance est minimum, à effectif égal de l'échantillon.*

1.6 Exhaustivité d'un estimateur

S'agissant d'estimer θ , certaines statistiques peuvent être exclues du fait qu'elles n'utilisent pas de façon exhaustive toute l'information contenue dans l'échantillon X_1, X_2, \dots, X_n . A l'inverse on peut s'attendre à ce qu'un «bon» estimateur soit une statistique qui ne retienne que ce qui est utile de l'échantillon.

Les notions d'exhaustivité et d'exhaustivité minimale viennent préciser cela.

Dans cette section Θ pourra être de dimension quelconque tout comme les statistiques considérées.

Définition 1.6.1 *On dit que la statistique T_n est exhaustive pour θ si la loi conditionnelle de $X = (X_1, X_2, \dots, X_n)$ sachant $T_n(X) = t$ n'est pas une fonction du paramètre θ :*

Soit E_θ l'ensemble de définition :

$$E_\theta = [f(x; \theta) \quad \forall x \in \mathbb{R} \quad \text{et} \quad \forall \theta \in \mathbb{R}] .$$

que l'on notera E_θ s'il ne dépend pas de θ .

Les variables aléatoires (X_i) étant indépendantes, la densité de l'échantillon X :

$$L(x_i, \theta) = \prod_{i=1}^n f(x_i; \theta) \quad \forall \theta \in \mathbb{R} \quad \text{et} \quad \forall x_i \in \mathbb{R}^n .$$

où $x_i = (x_1, x_2, \dots, x_n)$ est une réalisation de l'échantillon X_i .

Définition 1.6.2 *On dit que la statistique T_n est exhaustive si et seulement si la fonction de densité jointe peut être mise sous la forme :*

$$L(x_i, \theta) = g(T_n, \theta) h(x_1, x_2, \dots, x_n) .$$

où $g(T_n, \theta)$ dépend uniquement de $T_n(X_1, X_2, \dots, X_n)$ et θ ;

$h(x_1, x_2, \dots, x_n)$ est une fonction des observations et ne dépend pas de θ .

Théorème 1.6.1 *Toute statistique exhaustive et complète est minimale.*

1.6.1 Théorème de Darmais

Définition 1.6.3 *on suppose que l'ensemble de définition E_θ ne dépend pas de θ .*

Le théorème de Darmais donne les conditions d'existence d'une statistique exhaustive.

S'il existe un entier $n > 1$ tel que l'échantillon X admette une statistique exhaustive pour le paramètre θ , la fonction $f(x; \theta)$ est de la forme :

$$f(x; \theta) = \exp [a(x) \alpha(\theta) + b(x) + \beta(\theta)] \quad \forall \theta \in \mathbb{R} \quad \text{et} \quad \forall x \in E .$$

avec a, b, α, β des fonctions.

• Si f est de la forme exponentielle précédente et si l'application :

$$x_j \longrightarrow t = \sum_{i=1}^n a(x_i) .$$

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 particulière pour le paramètre θ .

1.6.2 Famille exponentielle

Définition 1.6.4 *La famille des lois exponentielles joue un rôle très important en statistique car elle possède un certain nombre de propriétés intéressantes.*

Il s'agit des lois dont la densité peut s'écrire sous la forme :

$$f(x; \theta) = a(\theta) + b(x) \exp \left[\sum_{j=1}^k a_j(\theta) T_j \right].$$

1.7 Amélioration d'un estimateur

L'existence d'une statistique exhaustive pour le paramètre permet d'améliorer, en utilisant le critère de la variance, un estimateur sans biais.

1.7.1 Théorème de Rao-Blackwell

Soient T un estimateur sans biais du paramètre θ et U une statistique exhaustive pour ce paramètre.

Alors $T^* = E_\theta(T/U)$ est un estimateur sans biais de θ au moins aussi bon que T .

Donc il est prouvé par :

- T^* est un estimateur de θ . Cette proposition est non triviale car il faut montrer que T^* dépend seulement des X_i et non de θ .

Puisque U est exhaustive, la densité conditionnelle de l'échantillon sachant U ne dépend pas de θ .

- T^* est sans biais. D'après le théorème de l'espérance totale :

$$E(T^*) = E[E(T/U)] = E(T) = \theta.$$

– T^* est au moins aussi bon que T . D'après le théorème de la variance totale :

$$V(T) = V(E(T/U)) + E(V(T/U))$$

$$V(T) = V(T^*) + E(V(T/U)).$$

– Comme $E(V(T/U))$ est positif ou nul on a $V(T) \geq V(T^*)$.

– De plus si $E(V(T/U)) = 0$ c'est que presque sûrement $T = f(U)$, il y a relation fonctionnelle entre T et U .

Ce théorème fournit une méthode pour améliorer un estimateur sans biais donné.

1.7.2 Théorème de Lehmann-Scheffe

Soit T^* un estimateur sans biais du paramètre θ dépendant d'une statistique exhaustive complète U . T^* est l'unique estimateur sans biais de variance minimale. En particulier, si on connaît un estimateur T sans biais, T^* est donné par $T^* = E_\theta(T/U)$.

1.8 Quelques estimateurs classiques

1. $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $E(\bar{X}) = m \implies \bar{X}$: est un estimateur sans biais de m .
2. $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, $E(S^2) = \frac{n-1}{n} \sigma^2 \implies S^2$ estimateur biais de σ^2 .
3. $S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} S^2$, $E(S^{*2}) = \frac{n}{n-1} E(S^2) = \sigma^2 \implies S^{*2}$ est un estimateur sans biais et consistant de σ^2 . Son estimation est $S^{*2} = \frac{n}{n-1} \sigma_e^2$ et où σ_e^2 est l'écart-type observé dans une réalisation de l'échantillon.
4. Si p est la fréquence d'un caractère, $F = \frac{S_n}{n}$, $E(F) = p \implies F$ constitue un estimateur sans biais et consistant de p . Son estimation est notée f .

Chapitre 2

Estimation par maximum des vraisemblance

2.1 Introduction

Pour obtenir un ensemble complet d'estimateurs, nous suivons une méthode intuitive représentée à la manière des moments comme on l'appelle M-estimateurs, en minimisant une fonction dépendant des données et des paramètres du modèle.

En 1912, le statisticien "Ronald Aylmer Fisher" a proposé une méthode simple à mettre en œuvre et efficace asymptotiquement, il est appelé par "L'estimateur du maximum de vraisemblance".

Cette méthode consiste à rechercher l'estimation du paramètre inconnu qui rend le plus probable ou le plus vraisemblance l'échantillon observé.

2.2 Approche générale

2.2.1 Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance est une méthode d'estimation paramétrique qui doit sa popularité à

- a la simplicité de son approche,
- b sa faculté d'adaptation à une modélisation complexe, i.e. une loi $\mathcal{L}(\theta)$ où θ symbolise une multitude de paramètres inconnus : $\theta = (\theta_1, \theta_2, \dots, \theta_n)^t$,
- c l'aspect numérique accessible grâce à l'application de méthodes d'optimisation connues.

Elle permet de :

- construire des estimateurs performants,
- construire des intervalles de confiances précis,
- mettre en œuvre des tests statistiques "puissants".

2.2.2 Cas de variable de discrète

Dans un premier temps, on suppose que X est une var discrète suivant la loi $\mathcal{L}(\theta)$ avec θ un paramètre inconnu. On rappelle que l'on veut estimer θ à partir des données x_1, \dots, x_n , le vecteur des données (x_1, \dots, x_n) étant une réalisation d'un n-échantillon (X_1, \dots, X_n) de X . La méthode du maximum de vraisemblance repose sur l'idée suivante :[\[4\]](#)

- o "le fait d'avoir observé les valeurs x_1, \dots, x_n n'est pas surprenant",

soit encore :

- o "l'hypothèse d'observer les valeurs x_1, \dots, x_n plutôt que d'autres était la plus

vraisemblable".

Dès lors, on considère θ comme une variable réelle et on s'intéresse aux valeurs de θ qui

○ "rendent l'observation des valeurs x_1, \dots, x_n la plus vraisemblable possible",

soit encore :

○ "maximisent les chances de réalisation de l'événement $\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\}$ ",

soit encore :

○ "maximisent la probabilité $\mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n))$ ".

Une telle valeur est donc solution d'un problème d'optimisation visant à maximiser une fonction de θ caractérisant la vraisemblance d'avoir obtenu x_1, \dots, x_n .

Ainsi, θ^* est un estimateur ponctuel du paramètre inconnu θ appelé estimateur du maximum de vraisemblance (emv).

Aussi, précisons que l'on peut définir la vraisemblance des données x_1, \dots, x_n par la fonction de θ :

$$L_n(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n)).$$

Comme (X_1, \dots, X_n) est un n-échantillon, par l'indépendance et la distribution identique des var X_1, \dots, X_n , on peut aussi écrire :

$$L_n(\theta) = \mathbb{P}_\theta\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n \mathbb{P}_\theta(X = x_i).$$

2.2.3 Cas de variable contenue à densité

Si X est une var à densité suivant la loi $\mathcal{L}(\theta)$ de densité f_θ , le raisonnement développé précédemment tient toujours.

Toute fois, la vraisemblance des données ne peut plus être mesurée par la fonction $L_n(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n))$ car elle est désormais nulle.

Au lieu de $\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\}$, une idée est d'introduire un événement non négligeable proche :

$(X_1, \dots, X_n) \in [x_1, x_1 + \varepsilon_1[\times \dots \times [x_n, x_n + \varepsilon_n[$ avec $\varepsilon_1, \dots, \varepsilon_n$ assez petits. [4](#)

Cela nous amène à mesurer la vraisemblance des données par la fonction de θ :

$$L_n^{(\varepsilon_1, \dots, \varepsilon_n)}(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \in [x_1, x_1 + \varepsilon_1[\times \dots \times [x_n, x_n + \varepsilon_n[).$$

En notant $f_\theta(t_1, \dots, t_n)$ une densité de (X_1, \dots, X_n) , on a

$$L_n^{(\varepsilon_1, \dots, \varepsilon_n)}(\theta) = \int_{x_1}^{x_1 + \varepsilon_1} \dots \int_{x_n}^{x_n + \varepsilon_n} f_\theta(t_1, \dots, t_n) dt_1 \dots dt_n,$$

Cependant, déterminer un θ qui maximise $L_n^{(\varepsilon_1, \dots, \varepsilon_n)}(\theta)$ n'est pas chose aisée :

- l'expression analytique d'une fonction intégrale comme $L_n^{(\varepsilon_1, \dots, \varepsilon_n)}(\theta)$ n'existe pas toujours,
- il dépend de $\varepsilon_1, \dots, \varepsilon_n$ dont la petitesse reste subjective.

Une solution est de considérer une version idéalisée de $L_n^{(\varepsilon_1, \dots, \varepsilon_n)}(\theta)$ qui est

$$L_n(\theta) = f_\theta(x_1, \dots, x_n).$$

Cette expression est une conséquence du résultat suivant : en notant F_θ la fonction de répartition de

(X_1, \dots, X_n) , il vient

$$\lim_{(\varepsilon_1, \dots, \varepsilon_n) \rightarrow (0, \dots, 0)} \frac{L_n^{(\varepsilon_1, \dots, \varepsilon_n)}(\theta)}{\varepsilon_1 \times \dots \times \varepsilon_n} = \frac{\partial^n}{\partial \varepsilon_1 \dots \partial \varepsilon_n} F_\theta(x_1, \dots, x_n) = f_\theta(x_1, \dots, x_n).$$

En notant θ^* un maximum de $L_n(\theta)$, θ^* est un estimateur du maximum de vraisemblance (emv) de θ correspondant aux données.

Comme (X_1, \dots, X_n) est un n-échantillon, par l'indépendance et la distribution identique des var X_1, \dots, X_n , on peut aussi écrire :

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

2.3 Fonction de vraisemblance

On appelle fonction de vraisemblance pour (x_1, \dots, x_n) la fonction de θ :

Cas continu :

Si X est une variable aléatoire continue une densité $f(x_i; \theta)$, alors :

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

Exemple 2.3.1 Soit l'échantillon (X_1, \dots, X_n) de la loi probabilité normale ; $X_i \sim N(\mu, \sigma^2)$ et la densité :

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

La fonction de vraisemblance s'écrit alors :

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

Cas discrète :

Si X est une variable aléatoire discrète, alors :

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

Exemple 2.3.2 Soit l'échantillon (X_1, \dots, X_n) de la loi probabilité de Poisson de la densité :

$$\mathbb{P}(X = x_i; \lambda) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}.$$

La fonction de vraisemblance s'écrit alors :

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

2.3.1 Fonction de vraisemblance : contexte théorique

Dans un contexte théorique, il est possible que seule la modélisation inhérente à X soit décrite, sans mention des données x_1, \dots, x_n . Dès lors, on appelle fonction de vraisemblance la fonction de vraisemblance d'une réalisation quelconque (x_1, \dots, x_n) de (X_1, \dots, X_n) .

2.4 Fonction de log-vraisemblance

On appelle fonction de log-vraisemblance de θ pour une réalisation donnée (x_1, \dots, x_n) de l'échantillon, la fonction de θ :

$$l_n : \quad \Theta \times \mathbb{R}^n \longrightarrow \mathbb{R}^+.$$

$$(x_1, \dots, x_n; \theta) \longrightarrow \ln (L_n(x_1, \dots, x_n; \theta)).$$

Avec :

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

$$l_n(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \ln f(x_i; \theta).$$

La fonction l se nomme la fonction log-vraisemblance de l'échantillon (x_1, \dots, x_n) ; elle est considérée comme de fonction de θ avec $\theta \in \Theta$.

la fonction de log-vraisemblance n'a de sens que si θ vérifie $L_n(x_1, \dots, x_n, \theta) > 0$.

la fonction de log-vraisemblance étant croissante, l'emv θ^* de θ pour (x_1, \dots, x_n) vérifie :

$$\theta^* \in \arg \min_{\theta \in \Theta} l_n(x_1, \dots, x_n; \theta) = \arg \min_{\theta \in \Theta} L_n(x_1, \dots, x_n; \theta).$$

2.5 Estimateurs du maximum de vraisemblance

2.5.1 Cas estimation ponctuelle d'un paramètre

Dans les cas usuels, la fonction $\theta \longrightarrow l_n(x_1, \dots, x_n; \theta)$ est de classe \mathcal{C}^2 et l'estimateur du maximum de vraisemblance existe sous deux conditions dites condition

nécessaire (CN) et condition suffisante(CS).

La condition nécessaire(CN) :

La condition nécessaire du programme de minimisation est l'existence d'un point critique, noté θ^* , pour la fonction $l(x_1, \dots, x_n; \theta)$ c'est-à-dire qu'il existe $\theta^* \in \text{int}(\Theta)$ ($\text{int}(\Theta)$ est le plus grand ouvert inclus dans Θ) tel que :

$$CN : \left. \frac{\partial l_n}{\partial \theta}(x_1, \dots, x_n; \theta) \right|_{\theta=\theta^*} = 0.$$

Ou de façon équivalente :

$$CN : \left. \frac{\partial \ln l_n}{\partial \theta}(x_1, \dots, x_n; \theta) \right|_{\theta=\theta^*} = 0.$$

La condition suffisante (CS) :

La condition suffisante permet de vérifier que le point critique θ^* est un maximum ce qui est vrai si :

$$CS : \left. \frac{\partial^2 l_n}{\partial^2 \theta}(x_1, \dots, x_n; \theta) \right|_{\theta=\theta^*} < 0.$$

Ou de façon équivalente :

$$CS : \left. \frac{\partial^2 \ln l_n}{\partial^2 \theta}(x_1, \dots, x_n; \theta) \right|_{\theta=\theta^*} < 0.$$

2.5.2 Cas estimation simultanée de plusieurs paramètres

Nous supposons que la loi de X ne dépend plus d'un seul paramètre $\theta^{\mathbb{R}}$ mais de plusieurs $\theta_1, \dots, \theta_d$. Nous devons donc estimer le vecteur $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$.

La méthode du maximum de vraisemblance conduit en général à la résolution du

système des condition :

La condition nécessaire (CN) :

La condition nécessaire (CN) porte maintenant sur les dérivées partielles par rapport aux différents paramètres :

$$CN : \left. \frac{\partial l_n}{\partial \theta_i}(x_1, \dots, x_n; \theta) \right|_{\theta=\theta^*} = 0, i = 1, \dots, d.$$

Ou de façon équivalente :

$$CN : \left\{ \begin{array}{l} \frac{\partial \ln l_n}{\partial \theta_1}(x_1, \dots, x_n; \theta) = 0 \\ \frac{\partial \ln l_n}{\partial \theta_2}(x_1, \dots, x_n; \theta) = 0 \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial l_n}{\partial \theta_i}(x_1, \dots, x_n; \theta) = 0. \end{array} \right.$$

La condition suffisante (CS) :

La condition suffisante (CS) porte sur la matrice hessienne des dérivées seconde :

$$CS : \left[\left. \frac{\partial^2 \ln l_n}{\partial \theta_i \partial \theta_j}(x_1, \dots, x_n; \theta) \right|_{\theta=\theta^*} \right]_{i,j=1,\dots,d}, \text{ matrice définie vigative.}$$

2.6 Exemples de estimateurs du maximum de vraisemblance

2.6.1 Loi uniforme

Supposons que X_1, \dots, X_n *i.i.d.* $\sim U_{(0;\theta)}$; de densité

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

La vraisemblance est la fonction qui à n valeurs x_1, \dots, x_n et à une valeur positive θ associe :

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n \frac{1}{\theta} 1_{[0;\theta]}(x_i) \\ &= \frac{1}{\theta^n} 1_{[0;\theta]}(x_1, \dots, x_n) \\ &= \frac{1}{\theta^n} 1_{[\max\{x_i\}, +\infty[}(\theta). \end{aligned}$$

Vue comme fonction de θ , la fonction vraisemblance est nulle si θ est inférieur à la plus grande des valeurs observées, elle vaut $\frac{1}{\theta^n}$ sinon. Elle est donc maximale pour :

$$\hat{\theta}_n = \max \{x_1, \dots, x_n\}.$$

2.6.2 Loi Bernoulli

Soit x_1, \dots, x_n un échantillon observé fixé ($\forall i \in [1; n], x_i \in \{0; 1\}$), remarquons que si $X \sim Ber(p)$, on a

$$\mathbb{P}(X = x_i; p) = p^{x_i} (1 - p)^{1 - x_i}.$$

On a la fonction de vraisemblance est :

$$L(x_1, \dots, x_n; p) = \prod_{i=1}^n \mathbb{P}(X_i = x_i; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}.$$

La fonction de log-vraisemblance vaut :

$$l(x_1, \dots, x_n; p) = \sum_{i=1}^n (x_i) \ln(p) + (n - \sum_{i=1}^n x_i) \ln(p-1).$$

Calculons les dérivées première par rapport à p .

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

Calculons les dérivées seconde par rapport à p .

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

Or $\frac{\partial \ln L(x_1, \dots, x_n; p)}{\partial p} = 0 \iff p = \frac{\sum_{i=1}^n x_i}{n}$ et on constate que $\frac{\partial^2 \ln L(x_1, \dots, x_n; p)}{\partial p^2} < 0$.

Donc la fonction de vraisemblance est maximale en $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}$.

2.6.3 loi Géométrique

Soit x_1, \dots, x_n un échantillon observé fixé ($\forall i \in [1; n], x_i \in \mathbb{N}^*$), remarquons que si

$X \sim G(p)$.

On a la fonction de vraisemblance est :

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

La fonction de log-vraisemblance vaut :

$$l(x_1, \dots, x_n; p) = n \ln(p) - n \ln(1-p) + \sum_{i=1}^n x_i \ln(1-p).$$

Calculons les dérivées première par rapport à p .

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

Calculons les dérivées seconde par rapport à p .

$$\begin{aligned} \frac{\partial^2 \ln L(x_1, \dots, x_n; p)}{\partial p^2} &= -\frac{n}{p^2} + \frac{n}{(1-p)^2} - \frac{\sum_{i=1}^n x_i}{(p-1)^2} \\ &= -\frac{n}{p^2(1-p)^2} ((1-p)^2 + \bar{x} - 1). \end{aligned}$$

Or $\frac{\partial \ln L(x_1, \dots, x_n; p)}{\partial p} = 0 \iff p = \frac{n}{\sum_{i=1}^n x_i}$ et on constate que $\frac{\partial^2 \ln L(x_1, \dots, x_n; p)}{\partial p^2} < 0$.

Donc la fonction de vraisemblance est maximale en $\hat{p} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$.

2.6.4 loi Exponentielle

Soit x_1, \dots, x_n un échantillon observé fixé ($\forall i \in [1; n], x_i > 0$), remarquons que si $X \sim \mathcal{E}(\lambda)$.

On a la fonction de vraisemblance est :

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

La fonction de log-vraisemblance vaut :

$$l(x_1, \dots, x_n; \lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i.$$

Calculons les dérivées première par rapport à λ .

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

Calculons les dérivées seconde par rapport à λ .

$$\frac{\partial^2 \ln L(x_1, \dots, x_n; \lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2}.$$

Or $\frac{\partial \ln L(x_1, \dots, x_n; \lambda)}{\partial \lambda} = 0 \iff \lambda = \frac{n}{\sum_{i=1}^n x_i}$ et on constate que $\frac{\partial^2 \ln L(x_1, \dots, x_n; \lambda)}{\partial \lambda^2} < 0$.

Donc la fonction de vraisemblance est maximale en $\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$.

2.6.5 loi Pareto

Supposons que X_1, \dots, X_n *i.i.d.* et $(\forall i \in [1; n], x_i \succ 1)$ de densité :

$$f(x; \gamma) = \frac{1}{\gamma} x^{-\frac{1}{\gamma}-1}.$$

On a la fonction de vraisemblance est :

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

La fonction de log-vraisemblance vaut :

$$l(x_1, \dots, x_n; \gamma) = -n \ln(\gamma) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^n \ln(x_i).$$

Calculons les dérivées première par rapport à γ .

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

Calculons les dérivées seconde par rapport à γ .

$$\frac{\partial^2 \ln L(x_1, \dots, x_n; \gamma)}{\partial \gamma^2} = -\frac{n}{\gamma^2} - \frac{2}{\gamma^3} \sum_{i=1}^n \ln(x_i).$$

Or $\frac{\partial \ln L(x_1, \dots, x_n; \gamma)}{\partial \gamma} = 0 \iff \gamma = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$ et on constate que $\frac{\partial^2 \ln L(x_1, \dots, x_n; \gamma)}{\partial \gamma^2} < 0$.

Donc la fonction de vraisemblance est maximale en $\hat{\gamma} = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$.

2.7 Quantité d'information de Fisher

La quantité d'information de Fisher est un outil précieux pour évaluer la qualité d'un estimateur. Elle n'est définie que sous certaines conditions de régularité.

Définition 2.7.1 *Pour $\theta \in \mathbb{R}$, si la loi des observations vérifie les conditions de régularité, on appelle quantité d'information (de Fisher) sur θ apportée par*

l'échantillon x_1, \dots, x_n , la quantité :

$$I_n(\theta) = E \left[\left(\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} \right)^2 \right].$$

Remarque 2.7.1 La définition de la quantité d'information ci-dessus est une définition générale, applicable quelle que soit la nature des variables aléatoires observées.

$$I_n(\theta) = Var \left(\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} \right) = -E \left[\frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial \theta^2} \right].$$

Preuve. L'étant une densité

$$\int_{\mathbb{R}^n} L(X; \theta) = 1.$$

En dérivant les deux membres par rapport à θ et on remarquant que :

$$\frac{\partial L(X; \theta)}{\partial \theta} = L(x_1, \dots, x_n; \theta) \frac{\partial \ln L(X; \theta)}{\partial \theta}.$$

il vient :

$$\int_{\mathbb{R}^n} \frac{\partial \ln L(X; \theta)}{\partial \theta} L(X; \theta) dX = 0.$$

ce qui prouve que la variable $\frac{\partial \ln L(X; \theta)}{\partial \theta}$ est centrée et que $I_n(\theta) = Var \left(\frac{\partial \ln L(X; \theta)}{\partial \theta} \right)$.

On dérivans une deuxième fois :

$$\int_{\mathbb{R}^n} \frac{\partial^2 \ln L(X; \theta)}{\partial \theta^2} L(X; \theta) dX + \int_{\mathbb{R}^n} \frac{\partial \ln L(X; \theta)}{\partial \theta} \frac{\partial L(X; \theta)}{\partial \theta} L(X; \theta) dX = 0.$$

en utilisant à nouveau la remarque sur $\frac{\partial L(X;\theta)}{\partial \theta}$, il vient :

$$\int_{\mathbb{R}^n} \frac{\partial^2 \ln L(X;\theta)}{\partial \theta^2} L(X;\theta) dX + \int_{\mathbb{R}^n} \left(\frac{\partial \ln L(X;\theta)}{\partial \theta} \right)^2 L(X;\theta) dX = 0.$$

Alors :

$$E \left[\left(\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial \theta^2} \right].$$

■

Propriété de $I_n(\theta)$:

La définition de la quantité d'information ci-dessus est une définition générale, applicable quelle que soit la nature des variables aléatoires observées.

Quand celles-ci sont indépendantes et de même loi, il est facile de voir que

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

Preuve. pour des variables aléatoires continues de densité f

$$\begin{aligned} I_n(\theta) &= \text{Var} \left(\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} \right) \\ &= \text{Var} \left(\frac{\partial}{\partial \theta} \ln \prod_{i=1}^n f(x_i; \theta) \right) \\ &= \text{Var} \left(\frac{\partial}{\partial \theta} \ln \sum_{i=1}^n f(x_i; \theta) \right) \\ &= \text{Var} \left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right) \\ &= \sum_{i=1}^n \text{Var} \left(\frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right) = nI_1(\theta). \end{aligned}$$

■

Propriété additive :

On dite que l'information de Fisher est additive ,si X et Y deux variables aléatoires indépendantes dans des modèles paramétriques au paramètre θ commun et de densité $f(x; \theta)$ et $g(y; \theta)$ et soient $I_X(\theta)$ et $I_Y(\theta)$ les quantités d'informations qui apportées par X et Y respectivement sur θ Soit $I_{(X;Y)}(\theta)$ la quantité d'information de $(X; Y)$:

$$I_{(X;Y)}(\theta) = I_X(\theta) + I_Y(\theta) \quad \forall \theta \in \Theta.$$

Preuve. Soit $h(x; y; \theta)$ la densité de $(x; y)$: $h(x; y; \theta) = f(x; \theta)g(y; \theta)$, alors :

$$\begin{aligned} \ln h(x; y; \theta) &= \ln f(x; \theta) + \ln g(y; \theta) \\ \frac{\partial^2 \ln h(x; y; \theta)}{\partial \theta^2} &= \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} + \frac{\partial^2 \ln g(y; \theta)}{\partial \theta^2} \\ I_{(X;Y)}(\theta) &= -E \left[\frac{\partial^2}{\partial \theta^2} \ln (f(x; \theta).g(y; \theta)) \right] \\ &= -E \left[\frac{\partial^2}{\partial \theta^2} (\ln f(x; \theta) + \ln g(y; \theta)) \right] \end{aligned}$$

Soit par linéarité de l'espérance mathématique

$$\begin{aligned} I_{(X;Y)}(\theta) &= -E \left[\frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) \right] - E \left[\frac{\partial^2}{\partial \theta^2} \ln g(y; \theta) \right] \\ &= I_X(\theta) + I_Y(\theta). \end{aligned}$$

■

2.7.1 Information de Fisher d'un échantillon

On appelle information de Fisher au point θ la matrice notée $I_n(\theta)$ vérifiant :

$$I_n(\theta) = \begin{pmatrix} E \left[\left(\frac{\partial \ln f(x;\theta)}{\partial \theta_1} \right)^2 \right] & \dots & E \left[\frac{\partial \ln f(x;\theta)}{\partial \theta_1} \frac{\partial \ln f(x;\theta)}{\partial \theta_d} \right] \\ \vdots & \ddots & \vdots \\ E \left[\frac{\partial \ln f(x;\theta)}{\partial \theta_1} \frac{\partial \ln f(x;\theta)}{\partial \theta_d} \right] & \dots & E \left[\left(\frac{\partial \ln f(x;\theta)}{\partial \theta_d} \right)^2 \right] \end{pmatrix}.$$

L'information de Fisher est donc une fonction qui à toute valeur du paramètre inconnu $\theta \in \Theta \subseteq \mathbb{R}^d$ associe une matrice de taille $d \times d$. Sous les hypothèses du modèle régulier,

$$I(\theta) = \begin{pmatrix} -E \left[\frac{\partial^2 \ln f(x;\theta)}{\partial^2 \theta_1} \right] & \dots & -E \left[\frac{\partial^2 \ln f(x;\theta)}{\partial \theta_1 \partial \theta_d} \right] \\ \vdots & \ddots & \vdots \\ -E \left[\frac{\partial^2 \ln f(x;\theta)}{\partial \theta_1 \partial \theta_d} \right] & \dots & -E \left[\frac{\partial^2 \ln f(x;\theta)}{\partial^2 \theta_d} \right] \end{pmatrix}.$$

pour tout $1 \leq i, j \leq d$:

$$I_{i,j}(\theta) = -E \left[\frac{\partial^2 \ln f(x;\theta)}{\partial \theta_i \partial \theta_j} \right].$$

2.7.2 Modèle régulier

Soit (P_θ, Θ) un modèle paramétrique. On note $f(x; \theta)$ la densité de P_θ , et si Θ est un ouvert .

On dit que le modèle (P_θ, Θ) est régulier, si les quatre hypothèses ci-dessus sont vérifiées :

(H1) Le support des lois $P_\theta : \Delta = \{x \in X ; f(x; \theta) > 0\}$ est indépendant de $\theta \in \Theta$.

(H2) La fonction de vraisemblance est deux fois continûment dérivables sur Θ , et existent pour tous $x \in \Delta$.

(H3) On suppose que les fonctions $\frac{\partial f}{\partial \theta}$ et $\frac{\partial^2 f}{\partial \theta^2}$ sont intégrables pour tout $\theta \in \Theta$, et que les dérivées peuvent s'effectuer sous le signe somme. Ainsi pour tout $\theta \in \Theta$ et $A \subset X$ borélien, on a :

$$\int_X \frac{\partial}{\partial \theta} f(x; \theta) dx = \frac{\partial}{\partial \theta} \int_X f(x; \theta) dx = 0.$$

et

$$\int_X \frac{\partial^2}{\partial \theta^2} f(x; \theta) dx = \frac{\partial^2}{\partial \theta^2} \int_X f(x; \theta) dx = 0.$$

(H4) L'information de Fisher $I_n(\theta)$ existe et vérifier :

$$I_n(\theta) = E \left[\left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 \right] > 0, \quad \forall \theta \in \Theta.$$

2.7.3 Inégalité de Fréchet-Darmois-Cramer-Rao (FDCR)

(a) Si la loi des observations vérifie les conditions de régularité, alors pour tout estimateur $\hat{\theta}_n$ de θ , on a :

$$Var(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} \left[\frac{\partial E(\hat{\theta}_n)}{\partial \theta} \right]^2.$$

(b) Si le support de X ne dépende pas de θ et que l'information de Fisher existe alors, pour tout estimateur sans biais $\hat{\theta}_n$

de θ , c'est-à-dire $E(\hat{\theta}_n) = \theta$ on obtient l'inégalité de Cramer-Rao s'écrit :

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

- (c) Si $\hat{\theta}_n$ est un estimateur sans biais d'une fonction $h(\theta)$; donc $E(\hat{\theta}_n) = h(\theta)$;
 l'inégalité de Cramer-Rao s'écrit :

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

2.8 Propriétés du maximum de vraisemblance

pour étudier ces propriétés nous allons poser des hypothèses sur la distribution de la variable d'intérêt X . Ces hypothèses sont qualifiées d'hypothèses de régularité.

Les hypothèses de régularité sont au nombre de trois :

- la fonction $\ln f(x_i; \theta)$ est trois fois différentiable par rapport à θ . Ses dérivées sont continues et finies pour toute valeur de x et de θ .
- Les espérances des dérivées première et seconde de $\ln f(x_i; \theta)$ par rapport à θ existent.
- la vraie valeur de θ , notée θ_0 , appartient à un ensemble compacte Θ .

Sous ces hypothèses de régularité, on peut montrer que l'estimateur du maximum de vraisemblance présente de bonnes propriétés :

(h1) convergence :

soit $\theta \in \Theta$ converge presque sûrement vers θ lorsque n tend vers l'infini :

$$\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{p.s.} \theta_0$$

(h2) asymptotiquement normal :

Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance $\hat{\theta}$

est :

$$\sqrt{n} \left(\hat{\theta}_n - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{loi} N \left(0, \frac{1}{I_1(\theta_0)} \right).$$

où θ_0 désigne la vraie valeur du paramètre et $I(\theta_0)$ correspond à la quantité d'information moyenne de Fisher évaluée au point θ_0 .

(h3) Invariance du maximum de vraisemblance :

Si $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ , alors $\varphi \left(\hat{\theta}_n \right)$ est l'EMV de $\varphi(\theta)$. De plus, si φ est dérivable, on a :

$$\sqrt{n} \left(\varphi \left(\hat{\theta}_n \right) - \varphi(\theta) \right) \xrightarrow[n \rightarrow \infty]{loi} N \left(0, \frac{\varphi'(\theta)^2}{I_1(\theta_0)} \right).$$

2.9 Estimation d'un modèle de régression linéaire

A côté de la méthode des moindres carrés ordinaires (MCO), la méthode du maximum de vraisemblance/MV (en anglais « Maximum Likelihood method » ou ML method) permet aussi d'estimer les paramètres d'un modèle de régression, sous l'hypothèse que la vraie (loi) distribution desdits paramètres est connue(1). Si le principe pour les MCO est de trouver le paramètre qui minimise la somme des carrés des erreurs, la méthode du maximum de vraisemblance cherche par contre à trouver le paramètre à même (ayant une forte probabilité) de reproduire les vraies valeurs de l'échantillon (celles réellement observées), soit trouver la valeur la plus vraisemblable du paramètre d'une population partant d'un échantillon donné (lire Bosonga Bofeki L. JP., 2019, p. 131). Autrement dit, pour reprendre les propos (de la même veine) chers à Kintambu Mafuku, « la méthode du MV est basée sur l'idée que si nous nous trouvons en présence des possibles valeurs différentes pour un paramètre, nous choisirons la valeur avec laquelle le modèle générerait avec

plus de probabilité l'échantillon observé » (Kintambu Mafuku E.G., 2004, p. 76).

L'on notera aussi que, sous l'hypothèse que les erreurs sont normalement distribuées, les estimateurs des MCO et ceux du maximum de vraisemblance sont identiques comme démontré plus bas.

Par ailleurs, il tient de préciser que l'estimateur de maximum de vraisemblance sert de base à certains tests statistiques, notamment : le test de Wald, celui du ratio de vraisemblance et celui du multiplicateur de Lagrange.

Dans les lignes qui suivent, nous montrons comment estimer les paramètres d'un modèle de régression linéaire simple, autant pour un modèle de régression multiple, par la méthode du maximum de vraisemblance ; ensuite, nous présentons les trois tests d'hypothèses construits sur base de l'estimateur du maximum de vraisemblance. [\[9\]](#)

2.9.1 Estimateur du maximum de vraisemblance (ML)

a) Estimateur ML d'un modèle de régression linéaire simple (MRLS) :

Considérons le MRLS suivant :

$$Y_t = a_0 + a_1 X_t + u_t.$$

Y_t , linéairement dépendant du terme d'erreur « u_t », est une variable normalement distribuée de paramètres (moyenne et variance) : $E(Y_t) = a_0 + a_1 X_t$ et $V(Y_t) = \sigma_u^2$.

En effet, si « $u_t \sim N(0; \sigma^2) \longrightarrow E(u_t) = 0$ »,

alors: $E(Y_t) = E(a_0 + a_1 X_t + u_t) = a_0 + a_1 X_t$.

Puisque : $Y_t - E(Y_t) = (a_0 + a_1 X_t + u_t) - (a_0 + a_1 X_t) = u_t$, alors la variance de Y_t

est : $V(Y_t) = E[Y_t - E(Y_t)]^2 = E(u_t)^2 = \sigma_u^2$.

D'où, la distribution de Y_t : $Y_t \sim [N(a_0 + a_1 X_t); \sigma_u^2]$.

Fonction de densité de probabilité jointe Elle s'écrit :

$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2)$. Si les « Y_t » sont indépendantes, cette fonction peut s'écrire aussi :

$$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2) = f(Y_1 | a_0 + a_1 X_t, \sigma_u^2) * f(Y_2 | a_0 + a_1 X_t, \sigma_u^2) * \dots * f(Y_t | a_0 + a_1 X_t, \sigma_u^2). \quad (2.1)$$

Fonction de densité de la loi normale générale En général, cette fonction se présente comme suit (a_0 et a_1 sont considérés comme des coefficients estimés) :

$$f(Y_t) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp \left[-\frac{1}{2} \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right]. \quad (2.2)$$

Fonction de vraisemblance et fonction log-vraisemblance :

Fonction de vraisemblance : Elle est obtenue lorsqu'on remplace la fonction de densité de la loi normale (2.2) dans la fonction de densité de probabilité jointe (2.1), en supposant connues les « Y_1, Y_2, \dots, Y_t » :

$$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2) = \frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \exp \left[-\frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right].$$

ou encore :

$$L(a_0, a_1, \sigma_u^2) = \frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \exp \left[-\frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right].$$

Fonction log-vraisemblance : Lorsqu'on effectue une transformation logarithmique de la fonction de vraisemblance ci-dessus, l'on obtient la fonction dite log-vraisemblance qui servira de base à l'estimation des paramètres « \hat{a}_0, \hat{a}_1 ». La fonction log-vraisemblance s'écrit :

$$\ln L = \ln \left(\frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \right) + \ln \left(\exp \left[-\frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right] \right).$$

En effet $\ln 1 = 0$; $\ln(\exp)(a) = \ln e^a = a$; $\ln x^2 = 2 \ln x$; $\sqrt{2\pi} = (2\pi)^{\frac{1}{2}}$.

$$\ln \left(\frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \right) = \ln(1) - [\ln(\sqrt{2\pi})^t + \ln \sigma_u^t] = -\frac{1}{2} t \ln(2\pi) - t \ln \sigma_u.$$

$$\ln \left(\exp \left[-\frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right] \right) = -\frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2.$$

$$\ln L = -\frac{1}{2} t \ln(2\pi) - t \ln \sigma_u - \frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2.$$

Considérant que : $\ln \sigma_u^2 = 2 \ln \sigma_u \longrightarrow \frac{1}{2} \ln \sigma_u^2 = \ln \sigma_u$, alors l'on peut écrire (fonction logvraisemblance retenue pour l'estimation des paramètres) :

$$\rightarrow \ln L = -\frac{1}{2} t \ln(2\pi) - \frac{t}{2} \ln \sigma_u^2 - \frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2. \quad (2.3)$$

Estimation des paramètres « \hat{a}_0, \hat{a}_1 et σ_μ^2 » Pour estimer les paramètres « \hat{a}_0, \hat{a}_1 et σ_μ^2 » par le maximum de vraisemblance, la démarche va consister à maximiser la fonction log-vraisemblance ci-dessus (2.3), ce qui revient à annuler ses dérivées premières par rapport aux arguments « \hat{a}_0, \hat{a}_1 et σ_μ^2 » comme suit :

En effet : $([G[Y(x)]]^n)' = nG^{n-1}Y(x)'$.

$$\frac{\partial \ln L}{\partial a_0} = -\frac{1}{2} * 2 * \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^{2-1} (-1) = -\sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right) (-1) = 0 \quad (2.4)$$

$$\frac{\partial \ln L}{\partial a_1} = -\frac{1}{2} * 2 * \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^{2-1} (-X_t) = -\sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right) (-X_t) = 0 \quad (2.5)$$

$$\frac{\partial \ln L}{\partial \sigma_\mu^2} = \frac{\partial}{\partial \sigma_\mu^2} \left(-\frac{t}{2} \ln \sigma_u^2 \right) - \frac{1}{2} * \frac{\partial}{\partial \sigma_\mu^2} \left(\sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right) = 0 \quad (2.6)$$

On a :

$$\frac{\partial}{\partial \sigma_\mu^2} \left(-\frac{t}{2} \ln \sigma_u^2 \right) = -\frac{t}{2} * \frac{1}{\sigma_\mu^2} = -\frac{t}{2\sigma_\mu^2} \quad (2.7)$$

Appelons : $\sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^2 = \sum (\bullet)$

$$\frac{\partial}{\partial \sigma_\mu^2} [(\sum (\bullet))^2] = \frac{\partial}{\partial \sigma_\mu} \left(\frac{\partial}{\partial \sigma_\mu} [(\sum (\bullet))^2] \right)$$

$$\frac{\partial}{\partial \sigma_\mu} [(\sum (\bullet))^2] = \sum \left[2 * \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right)^{2-1} * (-1) (Y_1 - a_0 - a_1 X_t) * \sigma_\mu^{-1-1} \right]$$

$$\longrightarrow \frac{\partial}{\partial \sigma_\mu} [(\sum (\bullet))^2] = \sum -2 * \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_u} \right) * (Y_1 - a_0 - a_1 X_t) * \sigma_\mu^{-2}$$

$$\begin{aligned} \frac{\partial}{\partial \sigma_\mu} \left(\frac{\partial}{\partial \sigma_\mu} [(\sum (\bullet))^2] \right) &= \sum -2 * (-2) * \frac{(Y_1 - a_0 - a_1 X_t)^2}{\sigma_u * \sigma_\mu^3} \\ &= 4 \sum \left[\frac{Y_1 - a_0 - a_1 X_t}{\sigma_\mu^2} \right]^2 \end{aligned} \quad (2.8)$$

Remplaçons les expressions (2.8) et (2.7) dans (2.6), l'on a :

$$\frac{\partial \ln L}{\partial \sigma_\mu^2} = -\frac{t}{2\sigma_\mu^2} + \frac{1}{2} \sum \left(\frac{Y_1 - a_0 - a_1 X_t}{\sigma_\mu^2} \right)^2 = 0 \quad (2.9)$$

Si l'on développe cette expression, l'on obtient :

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma_\mu^2} &= -\frac{t}{2\sigma_\mu^2} + \frac{1}{2} * -\frac{1}{\sigma_\mu^4} \sum (Y_1 - a_0 - a_1 X_t)^2 \\ &= -t + \frac{2\sigma_\mu^2}{2\sigma_\mu^4} \sum (Y_1 - a_0 - a_1 X_t)^2 = 0 \\ \frac{\partial \ln L}{\partial \sigma_\mu^2} &= -t + \frac{1}{\sigma_\mu^2} \sum (Y_1 - a_0 - a_1 X_t)^2 = 0\end{aligned}\quad (2.10)$$

Egalisées à zéro, les expressions (2.4), (2.5) et (2.10) s'écrivent (avec \hat{a}_{0MV} = estimateur du maximum de vraisemblance de a_0) :

$$\begin{aligned}\frac{\partial \ln L}{\partial a_0} = 0 &\longrightarrow \sum (Y_1 - \hat{a}_{0MV} - \hat{a}_{1MV} X_t) = 0 \longrightarrow \sum Y_1 = T\hat{a}_{0MV} + \hat{a}_{1MV} \sum X_t \\ \frac{\partial \ln L}{\partial a_1} = 0 &\longrightarrow \sum (Y_1 - \hat{a}_{0MV} - \hat{a}_{1MV} X_t) X_t = 0 \longrightarrow \sum Y_1 X_t = \hat{a}_{0MV} \sum X_t + \hat{a}_{1MV} \sum X_t^2 \\ \frac{\partial \ln L}{\partial \sigma_\mu^2} = 0 &= -t + \frac{1}{\sigma_{\mu MV}^2} \sum (Y_1 - \hat{a}_{0MV} - \hat{a}_{1MV} X_t)^2 = 0 \longrightarrow \sigma_{\mu MV}^2 = \frac{1}{T} \sum (Y_1 - \hat{a}_{0MV} - \hat{a}_{1MV} X_t)^2\end{aligned}$$

Les deux premières expressions étant identiques aux équations normales fournies par les MCO, l'on déduit que les estimateurs MCO des paramètres « a_0 et a_1 » et les estimateurs du maximum de vraisemblance sont égaux ou les mêmes :

$$\hat{a}_{0MV} = \hat{a}_{0MCO} \text{ et } \hat{a}_{1MV} = \hat{a}_{1MCO}.$$

Par contre, le développement de la dernière expression donne un estimateur du maximum de vraisemblance « $\sigma_{\mu MV}^2$ » de la variance de l'erreur « σ_μ^2 » différent de l'estimateur MCO :

$$\sigma_{\mu MV}^2 = \frac{1}{T} \sum (Y_1 - \hat{a}_{0MV} - \hat{a}_{1MV} X_t)^2 = \sigma_{\mu MV}^2 = \frac{1}{T} \sum e_t^2 \neq \sigma_{\mu MCO}^2 = \frac{1}{(T-2)} \sum e_t^2.$$

Notons que, l'estimateur MCO de la variance des erreurs étant sans biais, l'esti-

mateur du maximum de vraisemblance est biaisé, mais il reste convergent. Cette dernière propriété garantit la minimisation du biais avec l'accroissement de la taille de l'échantillon.

b) Estimateur ML d'un modèle de régression linéaire multiple (MRLM) :

Forme fonctionnelle d'un MRLM

Un modèle de régression linéaire multiple ou modèle linéaire général, soit une généralisation de la régression simple au cas multivarié (où on a k variables explicatives, avec $k > 1$), s'écrit :

$$Y_t = a_0 + a_1X_t + a_2X_{2t} + \dots + a_iX_{it} + \dots + a_kX_{kt} + u_t. \quad (2.11)$$

Ou encore (sans constante) :

$$Y_t = a_1X_t + a_2X_{2t} + \dots + a_iX_{it} + \dots + a_kX_{kt} + u_t. \quad (2.12)$$

Avec : $t = 1, \dots, T$ (les observations) ;

Y_t = la variable dépendante observée au temps t ;

X_{1t}, \dots, X_{kt} = les k variables explicatives ou « régresseurs » ;

a_0 = un paramètre du modèle ou terme constant (constante) ;

a_1, \dots, a_k = les paramètres réels et inconnus du modèle ;

u_t = le terme d'erreur.

Sous forme matricielle, la relation (2.11) peut encore s'écrire comme suit :

$$\underbrace{Y}_{(T,1)} = \underbrace{X}_{(T,k+1)} \underbrace{a}_{(k+1,1)} + \underbrace{u}_{(T,1)} \longrightarrow Y = Xa + u$$

avec :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \\ \vdots \\ Y_T \end{pmatrix}; X = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1t} & X_{2t} & \cdots & X_{kt} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1T} & X_{2T} & \cdots & X_{kT} \end{pmatrix}; a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}; u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \\ \vdots \\ u_T \end{pmatrix}.$$

Les «1» sur la première colonne de la matrice X captent les constantes « a_0 » dans le modèle.

Par contre, la notation matricielle de la relation sans constante (2.12) est (les formats de X et a changent) :

$$\underbrace{Y}_{(T,1)} = \underbrace{X}_{(T,k)} \underbrace{a}_{(k,1)} + \underbrace{u}_{(T,1)} \longrightarrow Y = Xa + u$$

Y_t , linéairement dépendant du terme d'erreur « u » (les erreurs sont supposées indépendantes et normalement distribuées, avec une espérance nulle ou $E(u) = 0$ et une variance constante « σ_μ^2 »), est une variable normalement distribuée de paramètre (moyenne et variance) :

$$E(Y_t) = a_0 + a_1 X_t + a_2 X_{2t} + \dots + a_k X_{kt} = X_a.$$

$$V(Y_t) = \sigma_\mu^2.$$

Densité de probabilité de Y_t : La densité de probabilité de Y_t , connaissant les paramètres « a et σ_μ^2 », est :

$$f(Y_t|a, \sigma_u^2) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp \left[-\frac{1}{2\sigma_\mu^2} \sum (Y_t - X_t a)^2 \right].$$

Fonction de densité de probabilité jointe ou conjointe de Y_1, Y_2, \dots, Y_t :

Connaissant les paramètres « a et σ_μ^2 », cette fonction s'écrit : $f(Y_1, Y_2, \dots, Y_t|a, \sigma_u^2)$.

Si les « Y_t » sont indépendantes, cette fonction peut s'écrire aussi :

$$f(Y_1, Y_2, \dots, Y_t|a, \sigma_u^2) = f(Y_1|a, \sigma_u^2) * f(Y_2|a, \sigma_u^2) * \dots * f(Y_t|a, \sigma_u^2) = \prod_{t=1}^T f(Y_t|a, \sigma_u^2)$$

$$\rightarrow \prod_{t=1}^T f(Y_t|a, \sigma_u^2) = \left(\frac{1}{\sqrt{2\pi\sigma_\mu^2}} \right)^T \exp \left[-\frac{1}{2\sigma_\mu^2} \sum_{t=1}^T (Y_t - X_t a)^2 \right] \quad (2.13)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_\mu^2}} \right)^T \exp \left[-\frac{1}{2\sigma_\mu^2} (Y - Xa)' (Y - Xa) \right] \quad (2.14)$$

Fonction de vraisemblance et fonction log-vraisemblance :

Fonction de vraisemblance : La fonction de vraisemblance, notée « L », correspond à la relation (2.14) en supposant connus les « Y_1, Y_2, \dots, Y_t » et inconnus les paramètres « a et σ_μ^2 » (à trouver), ce qui revient à écrire :

$$L(a, \sigma_\mu^2) = f(a, \sigma_u^2|Y_1, Y_2, \dots, Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma_\mu^2}} \right)^T \exp \left[-\frac{1}{2\sigma_\mu^2} (Y - Xa)' (Y - Xa) \right].$$

Fonction log-vraisemblance : La transformation logarithmique de la fonction de vraisemblance ci-dessus donne la fonction dite « log-vraisemblance » qui servira de base à l'estimation des paramètres « a ». La fonction log-vraisemblance

s'écrit :

$$\ln L = \ln \left[\left(\frac{1}{\sqrt{2\pi\sigma_\mu^2}} \right)^T \right] \ln \left(\exp \left[-\frac{1}{2\sigma_\mu^2} (Y - Xa)'(Y - Xa) \right] \right). \quad (2.15)$$

En effet : $\ln(uv) = \ln u + \ln v$; $\ln \exp(a) = \ln e^a = a$; $\ln x^2 = 2 \ln x$; $\frac{1}{\sqrt{2\pi\sigma_\mu^2}} = (2\pi\sigma_\mu^2)^{-1/2}$.

$$\longrightarrow \ln \left[\left(\frac{1}{\sqrt{2\pi\sigma_\mu^2}} \right)^T \right] = \ln \left[\left(\sqrt{2\pi\sigma_\mu^2} \right)^{-\frac{T}{2}} \right] = -\frac{T}{2} [\ln \pi + \ln (2\sigma_\mu^2)] = -\frac{T}{2} \ln \pi - \ln (2\sigma_\mu^2)$$

$$\longrightarrow \ln \left(\exp \left[-\frac{1}{2\sigma_\mu^2} (Y - Xa)'(Y - Xa) \right] \right) = -\frac{1}{2\sigma_\mu^2} (Y - Xa)'(Y - Xa)$$

$$\longrightarrow (Y - Xa)'(Y - Xa) = Y'Y - Y'Xa - X'a'Y + X'Xa'a = Y'Y - 2X'aY - X'Xa^2$$

Avec : $Y'Xa = X'aY$; $a' = a$ Etant donné ces expressions, la fonction log-vraisemblance (2.15) se réécrit finalement :

$$\begin{aligned} L(a, \sigma_\mu^2) &= -\frac{T}{2} \ln \pi - \ln (2\sigma_\mu^2) - \frac{1}{2\sigma_\mu^2} (Y - Xa)'(Y - Xa) \\ &= -\frac{T}{2} \ln \pi - \ln (2\sigma_\mu^2) - \frac{1}{2\sigma_\mu^2} (Y'Y - 2X'aY - X'Xa^2). \end{aligned} \quad (2.16)$$

L'expression (2.16) est celle retenue pour l'estimation des paramètres.

Estimation des paramètres « a et σ_μ^2 » : Pour estimer les paramètres inconnus « a et σ_μ^2 » par la méthode du maximum de vraisemblance, la démarche va consister à maximiser la fonction log-vraisemblance ci-dessus (2.16), ce qui revient à annuler ses dérivées premières par rapport aux arguments « a et σ_μ^2 » comme suit :

En effet : $([G[Y(x)]]^n)' = nG^{n-1}Y(x)'$; $\frac{d}{dX} [X^n] = nX^{n-1}$

$$\frac{\partial \ln L}{\partial a} = -\frac{1}{2\sigma_\mu^2} (Y'Y - 2X'aY - X'Xa^2) = 0 \quad (2.17)$$

$$\frac{\partial \ln L}{\partial \sigma_\mu^2} = \frac{\partial}{\partial \sigma_\mu^2} \left(-\frac{T}{2} \ln(2\sigma_\mu^2) \right) + \frac{\partial}{\partial \sigma_\mu^2} \left[-\frac{1}{2\sigma_\mu^2} (Y'Y - 2X'aY - X'Xa^2) \right] = 0 \quad (2.18)$$

$$\begin{aligned} &\longrightarrow \frac{\partial}{\partial \sigma_\mu^2} \left(-\frac{T}{2} \ln(2\sigma_\mu^2) \right) = -\frac{T}{2\sigma_\mu^2} \\ &\longrightarrow \frac{\partial}{\partial \sigma_\mu^2} \left[-\frac{1}{2} (\sigma_\mu^2)^{-1} (Y - Xa)'(Y - Xa) \right] = -\frac{1}{2} * (-1) * (\sigma_\mu^2)^{-1-1} (Y - Xa)'(Y - Xa) \\ &= \frac{1}{2\sigma_\mu^4} (Y - Xa)'(Y - Xa) = \frac{e'e}{2\sigma_\mu^4} \end{aligned}$$

Considérant ces expressions, (2.18) s'écrit :

$$\begin{aligned} \frac{\partial \ln L}{\partial \sigma_\mu^2} &= -\frac{T}{2\sigma_\mu^2} + \frac{1}{2\sigma_\mu^4} (Y - Xa)'(Y - Xa) \\ &= -\frac{T}{2\sigma_\mu^2} + \frac{e'e}{2\sigma_\mu^4} = \frac{T\sigma_\mu^2 + e'e}{2\sigma_\mu^4} = 0 \end{aligned} \quad (2.19)$$

Les expressions (2.17) (2.19) mises ensemble forment un système de « $k + 1$ » équations dont la solution donne l'estimateur du maximum de vraisemblance des coefficients « \hat{a}_{MV} » et celui de la variance de l'erreur « $\hat{\sigma}_{\mu MV}^2$ », soit respectivement :

$$\hat{a}_{MV} = (X'X)^{-1} X'Y. \quad (2.20)$$

$$\hat{\sigma}_{\mu MV}^2 = \frac{e'e}{T}. \quad (2.21)$$

Au regard des résultats obtenus, soient les expressions (2.20) (2.21), l'on note ce qui suit :

- L'estimateur MCO des paramètres « a » et l'estimateur du maximum de vraisemblance sont égaux : $\hat{a}_{MV} = \hat{a}_{MCO}$.
- Par contre, l'estimateur du maximum de vraisemblance « $\hat{\sigma}_{\mu MV}^2$ » de la variance de l'erreur « σ_μ^2 », étant biaisé, est différent de l'estimateur MCO, car (avec : $K = k - 1$ ou $K = k + 1$) :

$$E(\hat{\sigma}_{\mu MV}^2) = \frac{1}{T} E\left(\sum_{t=1}^T e_t^2\right) \\ \neq E(\hat{\sigma}_{\mu MCO}^2) = \frac{1}{(T - k - 1)} E\left(\sum_{t=1}^T e_t^2\right) = \frac{1}{(T - K)} (T - K) \sigma_\mu^2 = \sigma_\mu^2.$$

En effet :

$$E\left(\sum_{t=1}^T e_t^2\right) = (T - K) \sigma_\mu^2 \longrightarrow E(\hat{\sigma}_{\mu MV}^2) = \frac{1}{T} (T - K) \sigma_\mu^2 = \left(\frac{T - K}{T}\right) \sigma_\mu^2 = \sigma_\mu^2 - \frac{K}{T} \sigma_\mu^2.$$

On constate que, l'estimateur MCO de la variance des erreurs étant sans biais, l'estimateur du maximum de vraisemblance de la variance des erreurs est biaisé vers le bas (en moyenne, il sous-estime la valeur réelle de σ_μ^2), mais il reste convergent. Cette dernière propriété garantit la minimisation du biais avec l'accroissement de la taille de l'échantillon ; c'est-à-dire que, asymptotiquement (si T croît indéfiniment), $\hat{\sigma}_{\mu MV}^2$ est aussi non biaisé :

$$\lim_{T \rightarrow \infty} E(\hat{\sigma}_{\mu MV}^2) = \sigma_\mu^2 \text{ ou } \uparrow T \longrightarrow \hat{\sigma}_{\mu MV}^2 \simeq \hat{\sigma}_{\mu MCO}^2.$$

2.10 Estimation des parametres de la loi généralisée des valeurs extrêmes

2.10.1 Loi GEV avec $k = 0$ (EVI, loi Gumbel) :

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes de fonction de densité de probabilité

$$f(x) = \frac{1}{\alpha} \exp \left[-\frac{(x-u)}{\alpha} - \exp \left(-\frac{(x-u)}{\alpha} \right) \right], \text{ si } k = 0. \quad (2.22)$$

La vraisemblance logarithmique de l'échantillon x_1, x_2, \dots, x_n , une réalisation de X_1, X_2, \dots, X_n , est donnée par :

$$\ln L(\alpha, u) = -n \ln(\alpha) - \sum_{i=1}^n \frac{(x_i - u)}{\alpha} - \sum_{i=1}^n \exp \left[-\frac{(x_i - u)}{\alpha} \right]. \quad (2.23)$$

En prenant la dérivée de $\ln L(\alpha, u)$ par rapport à a et u , on obtient le système d'équations qui permet de maximiser cette fonction

$$n - \sum_{i=1}^n \frac{(x_i - u)}{\alpha} + \sum_{i=1}^n \frac{(x_i - u)}{\alpha} \exp \left[-\frac{(x_i - u)}{\alpha} \right] = 0. \quad (2.24)$$

$$-n - \sum_{i=1}^n \exp \left[-\frac{(x_i - u)}{\alpha} \right] = 0. \quad (2.25)$$

On peut toutefois simplifier ce système en écrivant l'équation (2.25) sous la forme :

$$u = -\alpha \ln \left\{ \frac{1}{n} \sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha} \right] \right\}. \quad (2.26)$$

et en remplaçant u dans (2.24) par (2.26) on obtient :

$$\alpha = \bar{x} - \left\{ \sum_{i=1}^n x_i \exp \left[-\frac{x_i}{\alpha} \right] \right\} \left\{ \sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha} \right] \right\}^{-1}. \quad (2.27)$$

Les équations (2.26) et (2.27) sont équivalentes au système (2.24) et (2.25). On peut ainsi, pour un échantillon donné X_1, X_2, \dots, X_n , déterminer $\hat{\alpha}$ par (2.27) et ensuite connaissant $\hat{\alpha}$ en déduire \hat{u} par (2.26).

Pour résoudre (2.27), on doit employer une méthode itérative comme celle de Newton-Raphson.

Toute fois, Phien (1987) souligne que la méthode numérique proposée par Clarke (1973) converge plus rapidement. Cette méthode est expliquée en détail dans Phien (1987) et est utilisée dans le logiciel AJUSTE.

Les estimateurs du maximum de vraisemblance sont asymptotiquement non-biaisés. Par contre, pour une taille d'échantillon finie, ils sont généralement biaisés. En particulier, les estimateurs \hat{u} et $\hat{\alpha}$ obtenus dans le cas de la loi EVI possèdent un biais (Lowery et Nash, 1970). Une bonne correction réduisant ce biais a été proposée par Fiorentio et Gabriele (1984), et les estimateurs corrigés résultants sont donnés par : [14]

$$\hat{\alpha} = \frac{n}{n - 0.8} \alpha.$$

et

$$u^* = \alpha^* \ln \left\{ n \left(\sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha^*} \right] \right)^{-1} \right\} - \frac{0.7}{n} \alpha^*.$$

Cette correction, qui permet de réduire significativement le biais, a été introduite dans le logiciel AJUSTE.

Les variances et la covariance asymptotiques des estimateurs du maximum de vraisemblance sont obtenues en inversant la matrice d'information de Fisher définie à

l'annexe B.

Nous obtenons pour la loi Gumbel (voir Phien, 1987) des expressions explicites des variances et de la covariance de \hat{u} et $\hat{\alpha}$. Elles sont données par :

$$Var(\hat{\alpha}) = 0.608 \frac{\alpha^2}{n}.$$

$$Var(\hat{u}) = 1.109 \frac{\alpha^2}{n}.$$

$$Cov(\hat{u}, \hat{\alpha}) = 0.257 \frac{\alpha^2}{n}.$$

De ces trois expressions, on déduit aisément les variances et covariances asymptotiques des estimateurs corrigés (Phien, 1987) :

$$Var(\alpha^*) = \left[\frac{n}{n-0.8} \right]^2 Var(\hat{\alpha}).$$

$$Var(u^*) = \left(\frac{\alpha^2}{n} \right) \left[1.109 - \frac{0.360}{n-0.8} + \frac{0.298}{n^2} \right].$$

et

$$Cov(u^*, \alpha^*) = \left(\frac{\alpha^2}{n} \right) \left[\frac{0.257n}{n-0.8} - \frac{0.426}{n} \right].$$

Remarque 2.10.1 *Remarquons que comme on peut s'y attendre, pour de grandes valeurs de n , les variances et la covariance de u^* et α^* sont équivalentes à celles de \hat{u} et $\hat{\alpha}$.*

2.10.2 Loi GEV avec $k \neq 0$ (EV2 et EV3) :

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes dont la fonction de densité de probabilité est donnée par :

$$f(x) = \frac{1}{\alpha} \left[1 - \frac{k}{\alpha} (x - u) \right]^{\frac{1}{k}-1} \exp \left[- \left(1 - \frac{k}{\alpha} (x - u) \right)^{\frac{1}{k}} \right], \text{ si } k \neq 0. \quad (2.28)$$

et une réalisation x_1, x_2, \dots, x_n de ces variables. Si l'on considère la variable auxiliaire :

$$y(x) = -\frac{1}{k} \ln \left[1 - \frac{k}{\alpha} (x - u) \right].$$

ce qui est équivalent à :

$$1 - \frac{k}{\alpha} (x - u) = e^{-ky(x)}.$$

la relation (2.28) devient :

$$f(x) = \frac{1}{\alpha} (e^{-ky(x)})^{\frac{1}{k}-1} \exp \left[- (e^{-ky(x)})^{\frac{1}{k}} \right].$$

ce qui s'écrit encore :

$$f(x) = \frac{1}{\alpha} e^{(k-1)y(x)} \exp \left(-e^{-y(x)} \right).$$

Le logarithme de la fonction de vraisemblance est alors :

$$\ln L(x; u, \alpha, k) = -n \ln(\alpha) - (1 - k) \sum_{i=1}^n y(x_i) - \sum_{i=1}^n \exp e^{-y(x_i)}.$$

Les estimateurs du maximum de vraisemblance sont obtenus en maximisant cette fonction, c'est-à-dire en annulant simultanément les trois dérivées partielles de

$\ln L$ par rapport aux 3 paramètres. Selon la notation de Jenkinson (1969), on peut montrer que le système d'équation à résoudre est :

$$-\frac{\partial \ln L(x; u, \alpha, k)}{\partial u} = \frac{Q}{\alpha}. \quad (2.29)$$

$$-\frac{\partial \ln L(x; u, \alpha, k)}{\partial \alpha} = \frac{1}{\alpha} \frac{P + Q}{k}. \quad (2.30)$$

$$-\frac{\partial \ln L(x; u, \alpha, k)}{\partial k} = \frac{1}{k} \left(R - \frac{P + Q}{k} \right). \quad (2.31)$$

Où

$$P = n - \sum_{i=1}^n \exp e^{-y(x_i)}. \quad (2.32)$$

$$Q = \sum_{i=1}^n \exp e^{-y(x_i)+ky(x_i)} - (1 - k) \sum_{i=1}^n e^{-y(x_i)}. \quad (2.33)$$

$$R = n - \sum_{i=1}^n y(x_i) + \sum_{i=1}^n y(x_i) e^{-y(x_i)}. \quad (2.34)$$

Ce système d'équations ne peut être résolu explicitement. C'est pourquoi nous devons utiliser une méthode itérative pour trouver la solution.

Supposons que l'estimateur du maximum de vraisemblance du vecteur $\theta = (u, \alpha, k)^T$ soit $\hat{\theta} = (\hat{u}, \hat{\alpha}, \hat{k})^T$, et que $\theta^{(j)}$ est la solution du système obtenue à la j ème itération. Les éléments du vecteur

$$-\Delta \ln L(\theta^{(j)}) = \left(-\frac{\partial \ln L(u, \alpha, k)}{\partial u}, -\frac{\partial \ln L(u, \alpha, k)}{\partial \alpha}, -\frac{\partial \ln L(u, \alpha, k)}{\partial k} \right)^T.$$

à la j ème itération peuvent être développés en séries de Taylor autour de $\hat{\alpha}$ (voir NERe, 1975).

On obtient, en ne retenant que les trois premiers termes (dérivées secondes), la formule itérative qui s'exprime sous forme matricielle de la façon suivante :

$$\theta^{(j+1)} = \theta^{(j)} + V(\hat{\theta})^{-1} \Delta \ln L (\theta^{(j)}) .$$

où $\Delta \ln L (\theta^{(j)})$ est le vecteur des dérivées premières évaluées en $\theta^{(j)}$ et $V(\hat{\theta})$ est la matrice suivante :

$$V(\hat{\theta}) = \begin{pmatrix} -\frac{\partial^2 \ln L}{\partial u^2} & -\frac{\partial^2 \ln L}{\partial u \partial \alpha} & -\frac{\partial^2 \ln L}{\partial u \partial k} \\ -\frac{\partial^2 \ln L}{\partial \alpha \partial u} & -\frac{\partial^2 \ln L}{\partial \alpha^2} & -\frac{\partial^2 \ln L}{\partial \alpha \partial k} \\ -\frac{\partial^2 \ln L}{\partial k \partial u} & -\frac{\partial^2 \ln L}{\partial k \partial \alpha} & -\frac{\partial^2 \ln L}{\partial k^2} \end{pmatrix} .$$

dont les éléments sont les dérivées secondes de la fonction de vraisemblance logarithmique affectées du signe moins et évaluées en $\hat{\theta}$ Puisque l'estimateur du maximum de vraisemblance $\hat{\theta}$ n'est pas connu, une approximation de cette matrice est nécessaire pour appliquer la procédure itérative.

Jenkinson (1969) propose d'approximer la matrice $V(\hat{\theta})$ par la matrice d'information de Fisher $I_f (\theta^{(j)})$ dont les éléments, les espérances mathématiques des dérivées secondes de la fonction de vraisemblance logarithmique affectées du signe moins, sont évaluées en $\theta^{(j)}$.

Toutefois, Perscott et Walden (1983) mentionnent que la convergence vers θ est plus rapide en utilisant $V (\theta^{(j)})$ plutôt que $I_f (\theta^{(j)})$.

Les éléments de V sont calculés à partir des équations (2.29) à (2.34) et ils sont donnés explicitement dans Prescott et Walden (1983). Hosking (1985) donne un programme FORTRAN qui permet l'application de la procédure itérative présentée plus haut. Le logiciel AJUSTE utilise ce programme presque intégralement. [14]

Les variances et covariances asymptotiques des estimateurs du maximum de vrai-

semblance sont obtenues en inversant la matrice d'information de Fisher I_f (voir annexe B).

En utilisant la notation de Jenkinson (1969) la matrice I_f^{-1} s'écrit :

$$I_f^{-1} = \begin{pmatrix} Var(\hat{u}) & Cov(\hat{u}, \hat{\alpha}) & Cov(\hat{u}, \hat{k}) \\ Cov(\hat{u}, \hat{\alpha}) & Var(\hat{\alpha}) & Cov(\hat{\alpha}, \hat{k}) \\ Cov(\hat{u}, \hat{k}) & Cov(\hat{\alpha}, \hat{k}) & Var(\hat{k}) \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \alpha^2 b & \alpha^2 h & \alpha f \\ \alpha^2 h & \alpha^2 a & \alpha g \\ \alpha f & \alpha g & c \end{pmatrix}.$$

où les valeurs de a, b, c, j, g et h ne dépendent que de k et ont été tabulées (Jenkinson, 1969) pour un certain nombre de valeurs de ce paramètre. Une interpolation est donc nécessaire en général.

Pour éviter cela, il est préférable d'évaluer exactement les termes de la matrice I_f laquelle est finalement inversée pour obtenir les variances et covariances des estimateurs.

Prescott et Walden (1980) donnent les expressions théoriques des éléments de I_f en termes de fonctions gamma (Γ) et digamma (Ψ). Nous reproduisons ici les équations fournies par ces auteurs :

$$E \left[-\frac{\partial^2 \ln L}{\partial u^2} \right] = \frac{n}{\alpha^2} p.$$

$$E \left[-\frac{\partial^2 \ln L}{\partial \alpha^2} \right] = \frac{n}{\alpha^2 k^2} [1 - 2\Gamma(2 - k) + p].$$

$$E \left[-\frac{\partial^2 \ln L}{\partial k^2} \right] = \frac{n}{k^2} \left[\frac{\pi^2}{6} - \left(1 - C - \frac{1}{k}\right)^2 + \frac{2q}{k} + \frac{p}{k^2} \right].$$

$$E \left[-\frac{\partial^2 \ln L}{\partial u \partial \alpha} \right] = \frac{n}{\alpha^2 k} [p - \Gamma(2 - k)].$$

$$E \left[-\frac{\partial^2 \ln L}{\partial u \partial k} \right] = -\frac{n}{\alpha k} \left(q + \frac{p}{k} \right).$$

$$E \left[-\frac{\partial^2 \ln L}{\partial \alpha \partial k} \right] = \frac{n}{\alpha^2 k^2} \left\{ 1 - C - \frac{[1 - \Gamma(2 - k)]}{k} - q - \frac{p}{k} \right\}.$$

où C est la constante d'Euler(0.5772),

$$p = (1 - k)^2 \Gamma(1 - 2k).$$

$$q = \Gamma(2 - k) \left[\Psi(1 - k) - \frac{(1 - k)}{k} \right].$$

2.11 Estimation par la méthode des moments

La méthode des moments consiste à estimer les paramètres inconnus en utilisant les moments d'ordre d , et les moments théoriques $E(X)$, $E(X^2)$, ..., $E(X^d)$ en fonction des paramètres, par la suite, nous noterons, pour simplifier :

$$m_d = E(X^d). \quad 1 \leq d \leq p$$

et les moments empiriques :

$$\hat{m}_d = \frac{1}{n} \sum_{i=1}^n X_i^d.$$

Il s'agit de résoudre le système d'équations en égalant les moments théoriques aux moments empiriques en fonctions des paramètres inconnus. La solution des équations si elle existe et est unique, sera appelée estimateur obtenu par la méthode des moments.

Le principe général de la méthode des moments est le suivant :

supposons qu'il existe un entier non nul k et une fonction $h : \mathbb{R} \rightarrow \mathbb{R}$ connue telle

que le paramètre θ que l'on souhaite estimer vérifie

$$\theta = h [E (X^d)] .$$

Au plan pratique, on exprime les m_d en fonction des moments de X des paramètre θ :

$$\left\{ \begin{array}{l} E (X^1) = m_1 = g_1 (\theta_1, \theta_2, \dots, \theta_p) \\ E (X^2) = m_2 = g_2 (\theta_1, \theta_2, \dots, \theta_p) \\ \cdot \\ \cdot \\ \cdot \\ E (X^d) = m_d = g_p (\theta_1, \theta_2, \dots, \theta_p) \end{array} \right. \iff \left\{ \begin{array}{l} \theta_1 = h_1 (m_1, m_2, \dots, m_p) \\ \theta_2 = h_2 (m_1, m_2, \dots, m_p) \\ \cdot \\ \cdot \\ \cdot \\ \theta_p = h_p (m_1, m_2, \dots, m_p) \end{array} \right.$$

On remplaçant les valeurs théoriques m_d par les moments empiriques \hat{m}_d , on obtient l'estimateur des moments est l'estimateur

$$\left\{ \begin{array}{l} \hat{\theta}_1 = h_1 (\hat{m}_1, \hat{m}_2, \dots, \hat{m}_p) \\ \hat{\theta}_2 = h_1 (\hat{m}_1, \hat{m}_2, \dots, \hat{m}_p) \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_p = h_1 (\hat{m}_1, \hat{m}_2, \dots, \hat{m}_p) \end{array} \right.$$

2.11.1 Application (simulation)

Exemple 2.11.1 *loi de béta*

Une variable aléatoire réelle X , prenant ses valeurs dans l'intervalle $[0, 1]$, notée

$B(\alpha, \beta)$, de paramètres positifs α et β

$$B(\alpha, \beta) = \int_0^1 x^{(\alpha-1)} (1-x)^{(\beta-1)} dx = B(\beta, \alpha).$$

Elle est reliée à la fonction gamma par la relation :

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

X suit une loi bêta de paramètre α et β , telles que

$$E[X] = \frac{\alpha}{\alpha + \beta}.$$

$$Var[X] = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}.$$

Donc on peut exprimer les deux paramètres α et β en fonction de l'esperance et la variance :

$$\left\{ \begin{array}{l} E[X] = \frac{\alpha}{\alpha + \beta} \\ Var[X] = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \end{array} \right. \implies \left\{ \begin{array}{l} \alpha = E[X] \left(\frac{E[X](1-E[X])}{Var[X]} - 1 \right) \\ \beta = (1 - E[X]) \left(\frac{E[X](1-E[X])}{Var[X]} - 1 \right) \end{array} \right.$$

Et les estimateurs par la méthode des moments sont :

$$\left\{ \begin{array}{l} \hat{\alpha} = \bar{X} \left(\frac{\bar{X}(1-\bar{X})}{S^2} - 1 \right) \\ \hat{\beta} = (1 - \bar{X}) \left(\frac{(1-\bar{X})(1-\bar{X})}{S^2} - 1 \right) \end{array} \right.$$

avec : $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$

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

Application numérique

On génère un échantillon de taille $n = 1000$ suit la loi beta de paramètre $\alpha = 2$ et $\beta = 4$ respectivement, on appliquant la méthode du maximum de vraisemblance et la méthode des moments pour calculer l'estimateur de α et β , on calcule le biais et l'erreur quadratique moyenne. Pour cela on prend $m = 50$.

Le but de cette application est de comparer entre deux méthodes d'estimation ponctuelle la méthode du maximum de vraisemblance et la méthode des moment. [1]

La méthode du moment			La méthode du maximum de vraisemblance		
EMM	biais	EQM	EMV	biais	EQM
2.0661	0.0661	0.0392	2.0607	0.0342	0.1121
4.0164	0.0102	0.1102	4.0086	-0.0199	0.0545

TAB. 2.1 – Comparaison entre l'EMV et l'EMM de la loi béta

Exemple 2.11.2 loi exponentielle

la loi exponentielle de paramètre λ , l'espérance empirique et la variance empirique sont :

$$\begin{cases} E[X] = \frac{1}{\lambda} \\ Var[X] = \frac{1}{\lambda^2} \end{cases}$$

La méthode des moments consiste à trouver $\hat{\lambda}_{MM}$ l'estimateur de paramètre λ , donc l'estimateur par la méthode des moments est :

$$\hat{\lambda}_{MM} = \frac{1}{\bar{X}}.$$

Application numérique

On génère un échantillon de taille $n = 1000$ suit la loi exponentielle de paramètre $\lambda = 2$, on appliquant la méthode du maximum de vraisemblance et la méthode des moments pour calculer l'estimateur de λ , et le biais et l'erreur quadratique moyenne. Pour cela on prend $m = 50$

Le but de cette application est de comparer entre deux méthodes d'estimation ponctuelle la méthode du maximum de vraisemblance et la méthode des moments. [\[1\]](#)

La méthode du moment			La méthode du maximum de vraisemblance		
EMM	biais	EQM	EMV	biais	EQM
2.0191	0.0191	0.0004	2.0191	-0.0031	0.0051

TAB. 2.2 – Comparaison entre l'EMV et l'EMM de la loi exponentielle

Conclusion : On peut conclure que la méthode du maximum de vraisemblance est meilleur que celle des moments.

Conclusion

En statistique, l'estimateur du maximum de vraisemblance est un estimateur statistique utilisé pour inférer les paramètres de la loi de probabilité d'un échantillon donné en recherchant les valeurs des paramètres maximisant la fonction de vraisemblance.

On cherche à trouver le maximum de cette vraisemblance pour que les probabilités des réalisations observées soient aussi maximum ceci est un problème d'optimisation. On utilise généralement le fait que si L (la vraisemblance de θ) est dérivable (ce qui n'est toujours le cas) et si L admet un maximum global en une valeur $\hat{\theta} = \theta$ et que la dérivée seconde est strictement négative en $\hat{\theta} = \theta$, alors $\hat{\theta} = \theta$ est un maximum local de $L(x_1, \dots, x_n; \theta)$. 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).

Dans ce mémoire, nous avons essayé d'exposer cette méthode d'où on a défini la fonction de vraisemblance, la fonction de log vraisemblance et l'estimateur du maximum de vraisemblance qui est bien un estimateur asymptotiquement sans biais, fortement consistant, asymptotiquement efficace et asymptotiquement nor-

Conclusion

male c'est donc un estimateur performant.

De plus, on a aussi applique cette méthode sur quelques lois de probabilités et estimé leurs paramètres.

Bibliographie

- [1] ABBA,M.(2019).Sur la théorie de l'estimation paramétrique,Mémoire master,Université Mohamed Khider,Biskra
- [2] Adoum Kamata,M;R ;J.(2018).Statistiques inférentielle,Université de Bngui.
- [3] Boulay,J-P.(2010).Statistique mathématique applications commentées.Paris.ISBN 978-2-7298-5602-1.
- [4] Bouraine et Berdjoudj,L,(2013 ;2014).Statistique inférentielle,Université A. MIRA - Béjaia.
- [5] Cantoni, E. Huber,P. Ronchetti, E.(2006).Maitriser l'aléatoire, Exercices résolus de probabilités et statistique. ISBN-10 , 2-287-34069-6. ISBN-13 , 978-2287-34069-7 Springer Paris Berlin Heidelberg New York.
- [6] Chesneau.C.(2017).Sur l'estimateur du maximum de vraisemblance (emv),Université de Caen.
- [7] Delmas, J. (2010).Introduction au calcul des probabilité et à la statistique. Paris. ISBN 978-2-7225-0922-1.
- [8] Dusart,P .Cours de statistiques inférentielle.
- [9] Keribin,C.(2021).Modélisation statistique.Université Paris-Saclay.STA201.
- [10] Gaudoin,O.Principes et Méthodes Statistiques.

- [11] Kuma.J. K.(2019) .Estimation par la méthode du maximum de vraisemblance :éléments de théorie et pratiques sur logiciel", Université de Kinshasa.
- [12] Lecoutre, J-P.(2008).Statistique et probabilités. Dunod, Paris, 2005, ISBN 978-2-10-054431-8.
- [13] Lejeune, M. (2010).Statistique la théorie et ses applications.Springer-Verlag France, Paris.
- [14] Mario ,H.(1993).modélisation de variables de précipitation par des distributions statistiques application à la région du québec et du nouveau-brunswick ,Université de québec 2800,rue Einstein,Suite 205.
- [15] O. Wintenberger.Statistique Mathématique.
- [16] Perreault.L et Bobée.B.Loi généralisée des valeurs extrêmes :Propriétés mathématiques et statistiques :Estimation des paramètres et des quantiles X_t de période de retour T, Université du Québec, no. R350, 1992.
- [17] Ribereau,P.(2016).Cours de Statistiques Inférentielles.
- [18] Saporta, G. (2006).Probabilités analyse des données et statistique.Editions Technip, Paris.
- [19] Scholler,J.(2018 ;2019).Module 2-outils quantitatitifs statistiques et analyse de données 1,compléments de probabilités et statistique inférentielle.
- [20] Veysseyre, R. (2001 ;2006).Aide mémoire, Statistique et probabilités pour l'ingé-nieur.Dunod, Paris, ISBN 2 10 049994 7.

ملخص

في هذه المذكرة، نقدم طريقة تقدير، وهي قابلة للتطبيق على نطاق أوسع بكثير من التقنيات المستخدمة سابقًا، ولكنها تتطلب أيضًا فرضيات قوية إلى حد ما، ألا وهي طريقة الاحتمال الأقصى.

حيث قمنا بشرح هذه الأخيرة و تعرضنا إلى خصائص المقدر مع بعض الأمثلة.

Abstract

In this brief we introduce a method of estimation, which is much more widely applicable than the techniques used previously, but which also requires rather strong assumptions.

This is the estimation by the method of the maximum of true semblance.

We have explained the principle of the latter as well as the properties of its estimator is some examples of application

Résumé

Dans ce mémoire nous introduisons une méthode d'estimation, qui est beaucoup plus largement applicable que les techniques utilisées précédemment, mais qui nécessite également d'assez fortes hypothèses.

Il s'agit de l'estimation par la méthode du maximum de vrais semblance, ou ML.

On a exposé le principe de cette dernière ainsi que les propriétés de son estimateur est quelques exemples d'application