

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

MOHAMDI Nabila

Titre :

Analyse de la fonction de vraisemblance pour l'estimation d'un paramètre.

Membres du Comité d'Examen :

| | | |
|------------------------------|------|--------------|
| Pr. NECIR Abdelhakim | UMKB | Président |
| Pr. CHERFAOUI Mouloud | UMKB | Encadreur |
| Dr. SOLTANE Louiza | UMKB | Examinatrice |

Juin 2023

DÉDICACE

Je dédie ce modeste travail à

Que Dieu les protège et leurs donne tout le bonheur du monde.

Mon père : Maseoud qui m'a indiqué la bonne voie, qui m'a encouragé durant
toutes mes années d'études,

Ma chère mère

Ma chère sœur SOURIA

Ma chère sœur HADJAR

Mon frère M^{ed} Amine

Tous mes amis(es) avec qui j'ai partagé les meilleurs moments.

Mon promoteur et tous les enseignants et le personnels de l'université de Biskra.

REMERCIEMENTS

*J*e remercie ALLAH le Tout-puissant de m'avoir donné le courage, la volonté et la patience de mener à terme cet humble travail.

*J'*aimerais remercier tout particulièrement mon encadreur Mr CHERFAOUI Mouloud. Pour ses orientations, sa confiance et sa patience qui ont constitué un apport considérable sans lesquels ce travail n'aurait pu être mené au bon terme.

Mes respects et remerciements vont également aux membres du jury qui m'ont fait l'honneur d'évaluer mon travail.

Je tiens à exprimer mes sincères remerciements à tous les enseignants qui nous ont formé et qui par leurs compétences nous ont motivé pour poursuivre nos études.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Table des matières

| | |
|--|----------|
| Remerciements | ii |
| Table des matières | ii |
| Table des figures | v |
| Introduction générale | 1 |
| 1 Introduction à la théorie d'estimation paramétrique | 3 |
| Introduction | 3 |
| 1.1 Quelques définitions | 3 |
| 1.2 Estimation paramétrique et estimateur | 5 |
| 1.3 Propriétés d'un estimateur | 6 |
| 1.3.1 Le biais d'un estimateur | 6 |
| 1.3.2 Estimateur convergent | 8 |
| 1.3.3 Choix d'un estimateur | 11 |
| 1.4 Statistique exhaustive | 13 |
| 1.4.1 Théorème de factorisation | 15 |
| 1.4.2 Théorème de Darmais | 16 |
| 1.5 Estimateur efficace | 18 |

| | | |
|----------|---|-----------|
| 1.5.1 | Information de Fisher | 18 |
| 1.5.2 | Estimateur sans biais de variance minimale | 20 |
| 1.5.3 | Inégalité de Fréchet-Darmonis-Cramer-Rao(FDCR) | 20 |
| 1.5.4 | Théorème sur l'efficacité | 21 |
| | Conclusion | 22 |
| 2 | Estimation par maximum de vraisemblance | 23 |
| | Introduction | 23 |
| 2.1 | Méthode "maximum de vraisemblance"(M.V) | 24 |
| 2.2 | Définitions | 24 |
| 2.3 | Exemples usuels d'EMV | 25 |
| 2.3.1 | Cas de variables discrètes | 26 |
| 2.3.2 | Cas de variables continues | 32 |
| | Conclusion | 36 |
| 3 | Implémentation de la méthode MV sur ordinateur | 38 |
| | Introduction | 38 |
| 3.1 | Position du problème l'implémentation de la méthode | 39 |
| 3.2 | Exemples illustratifs du problème | 43 |
| 3.2.1 | Exemple 01 : Cas de la loi Normale | 43 |
| 3.2.2 | Exemple 02 : Cas de la loi Exponentielle | 44 |
| 3.2.3 | Exemple 03 : Illustration par simulation | 47 |
| | Conclusion | 47 |
| | Conclusion générale | 50 |

Table des figures

| | | |
|-----|---|----|
| 3.1 | Précision des calculs sous le logiciel Matlab et le logiciel <i>R</i> | 42 |
| 3.2 | Code source programme Matlab pour le calcul de borne de majoration | 46 |
| 3.3 | Exemples de variation de la borne de majoration | 46 |
| 3.4 | Variation de la fonction de vraisemblance cas de loi Normale | 48 |
| 3.5 | Variation de la fonction de vraisemblance cas de loi Exponentielle . . . | 48 |
| 3.6 | Variation de la fonction de vraisemblance cas de loi de Poisson | 48 |

Introduction générale

En mathématiques, un estimateur est une statistique permettant d'évaluer un paramètre inconnu relatif à une loi de probabilité (comme son espérance ou sa variance). Il peut, par exemple, servir à estimer certaines caractéristiques d'une population totale à partir de données obtenues sur un échantillon. La définition et l'utilisation de tels estimateurs constituent la statistique inférentielle. La qualité des estimateurs s'exprime par leur convergence, leur biais, leur efficacité et leur robustesse.

Dans la littérature diverses méthodes ont été introduites et qui permettent d'obtenir des estimateurs de qualités différentes. Les plus populaires sont la méthode des moments et celle du maximum de vraisemblance avec une légère préférence pour MV, le fait permet la construction des estimateurs ayant de bonnes propriétés.

Dans le présent mémoire nous allons s'intéressé à l'application de la méthode du MV. Particulièrement, nous focalisons sur les limites de l'implémentation de la technique en question sur ordinateur.

Pour comprendre et mettre au clair le problème que nous avons abordé, la suite du présent document est organisée comme suit :

- Dans le Chapitre 1, nous rappelons quelques notions sur la théorie de l'estimation paramétrique.

- Dans le Chapitre 2, nous allons présenter la méthode du maximum de vraisemblance qui sera appuyée par des exemples d'applications.
- Dans le chapitre 3, et qui forme notre contribution effectif dans ce mémoire, via des exemples concrets nous avons expliqué pourquoi l'implémentation de la méthode du maximum de vraisemblance n'est pas toujours possible.

Enfin, en plus de la conclusion générale et les perspectives présentées, le document a été complété et enrichi par une liste bibliographique.

Chapitre 1

Introduction à la théorie d'estimation paramétrique

Introduction

Dans ce chapitre, nous allons introduire dans un premier lieu quelques définitions liées à la théorie d'estimation paramétrique. Par la suite, les critères de mesure de la qualité d'un estimateur seront abordés.

1.1 Quelques définitions

Dans cette section nous allons présenter quelques définitions qui nous permettront de comprendre le langage de la théorie d'estimation en générale et l'estimation paramétrique en particulier.

Définition 1.1.1 (*Variable aléatoire*) Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé. On appelle variable aléatoire X toute fonction mesurable définie sur Ω dans \mathbb{R} et on note :

$$X : \Omega \rightarrow \mathbb{R}.$$

Remarque 1.1.1 *On peut distinguer dans la littérature et la pratique deux types de variables aléatoires à savoir des variables aléatoires discrètes et des variables aléatoires continues.*

Définition 1.1.2 *(Echantillon) Soit X une variable aléatoire sur un référentiel Ω . Un échantillon de X de taille n est un n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires et identiquement distribuées de même loi que X . La loi de X sera appelée loi mère. Une réalisation de cet échantillon est un n -uplet de réels (x_1, x_2, \dots, x_n) .*

Définition 1.1.3 *Un modèle statistique est une description mathématique approximative du mécanisme qui a généré les observations, que l'on suppose être un processus stochastique et non un processus déterministe. Il s'exprime généralement à l'aide d'une famille de distributions (ensemble de distributions) et d'hypothèses sur les variables aléatoires X_1, \dots, X_n . Chaque membre de la famille est une approximation possible de F : l'inférence consiste donc à déterminer le membre qui s'accorde le mieux avec les données.*

Définition 1.1.4 *(Modèle paramétrique) Un modèle paramétrique est un modèle où l'on suppose le type de loi de X est connu, mais qu'il dépend d'un paramètre inconnu, de dimension n .*

Exemple 1.1.1 *Les modèles suivants Sont des modèles paramétriques.*

1. *Le modèle gaussien $\{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$.*
2. *Le modèle exponentiel $\{\varepsilon(\lambda), \lambda > 0\}$.*

1.2 Estimation paramétrique et estimateur

L'approche paramétrique suppose que les données sont issues d'une loi de probabilité de forme connue dont seuls les paramètres sont inconnus. L'objectif est de connaître une approximation de ces paramètres. Cette "approximation" est appelée désormais, dans le langage statistique, l'estimateur des paramètres inconnus.

Définition 1.2.1 *On cherche à connaître un paramètre θ qui dépend de la loi de X . Pour cela, on définit un estimateur comme une variable aléatoire mesurable par rapport à un échantillon à n éléments de la variable aléatoire X . En d'autres termes, un estimateur est une fonction qui fait correspondre à chaque réalisation possible (x_1, x_2, \dots, x_n) de l'échantillon X_1, \dots, X_n à n éléments la valeur $\hat{\theta}$ que l'on nomme "la valeur estimée" ou "estimation".*

$$\hat{\theta}_n = f(x_1, x_2, \dots, x_n).$$

Formellement, un estimateur ne peut prendre qu'un nombre fixe n d'arguments. En pratique, on considère généralement une suite d'estimateurs $\hat{\theta}_n$ pour chaque taille n de l'échantillon, qu'on appelle également estimateur. Un estimateur ne doit évidemment jamais dépendre de θ , il ne dépend que des observations empiriques c'est-à-dire de la réalisation de l'échantillon.

Exemple 1.2.1 *Les paramètres d'une loi de normale μ, σ^2 peut-être estimés respectivement*

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

et

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

1.3 Propriétés d'un estimateur

1.3.1 Le biais d'un estimateur

Soit T_n désigne l'estimateur du paramètre θ et soulignons que tout estimateur T_n peut donner lieu à l'écriture suivante :

$$E(T_n) = \theta + B(n, \theta),$$

où $B(n, \theta)$ est appelé le biais de T_n .

Ainsi, selon la fonction $B(n, \theta)$ on peut distinguer trois types d'estimateurs. Ci-dessus les trois situations possibles illustrées par des exemples d'applications.

Définition 1.3.1 (*Estimateur sans biais*) On dit que T_n est un estimateur sans biais de θ si et seulement si :

$$B(n, \theta) = 0,$$

c'est-à-dire

$$E(T_n) = \theta.$$

Exemple 1.3.1 Soit (X_1, X_2, \dots, X_n) un n -échantillon *i.i.d.* issu de la variable aléatoire X dont $E(X) = m$. L'estimateur empirique $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ de m est un estimateur sans biais. En effet, on a

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n m = m.$$

Définition 1.3.2 (*Estimateur avec biais*) On dit que T_n est un estimateur avec biais ou biaisé de θ si seulement si :

$$B(n, \theta) = E(T_n - \theta) = E(T_n) - \theta \neq 0,$$

c'est-à-dire

$$E(T_n) \neq \theta.$$

Exemple 1.3.2 Soit (X_1, X_2, \dots, X_n) un n échantillon, l'estimateur empirique S^2 est un estimateur biaisé de σ^2 et ceci le fait que

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

Définition 1.3.3 (*Estimateur asymptotiquement sans biais*) On dit que l'estimateur biaisé T_n de θ est un estimateur asymptotiquement sans biais si et seulement :

$$\lim_{n \rightarrow \infty} B(n, \theta) = 0,$$

c'est-à-dire

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

Exemple 1.3.3 Soit (X_1, X_2, \dots, X_n) un n -échantillon *i.i.d.* issu de la variable aléatoire X dont $E(X) = m$. L'estimateur empirique S^2 de $\sigma^2 < \infty$ est un estimateur sans biais. En effet, d'après l'exemple ??varAvecBiais), on a

$$E(S^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2. \quad (1.1)$$

D'après cette dernière équation, il est clair que l'estimateur S^2 est un estimateur avec biais mais asymptotiquement il est sans biais, car :

$$\lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2$$

1.3.2 Estimateur convergent

Dans cette section nous allons introduire quelques types de convergence d'une suite de variables aléatoires qui peuvent être adoptés à l'étude de convergence d'un estimateur T_n qui est lui aussi vu comme une suite de variables aléatoires.

Pour cela, considérant dans tout au long de cette section X_n une suite de *v.a.*

La convergence en loi

On note F_n et F les fonctions de répartition de X_n et X respectivement. on dit que X_n converge en loi vers X et on note :

$$X_n \xrightarrow{l} X,$$

si la suivante limite a lieu :

$$\lim_{n \rightarrow \infty} F_n(X) = F(X).$$

La convergence en probabilité

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

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

Ou encore,

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

Ainsi, on dit que T_n est un estimateur convergent de θ si T_n tend vers θ en probabilité quand n tend vers ∞ c'est-à-dire

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

et on note

$$T_n \xrightarrow[n \rightarrow \infty]{p} \theta.$$

La convergence presque sûre

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

$$P\left(\forall \omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)\right) = 1$$

\Leftrightarrow

$$P\left(\forall \omega : \left| \lim_{n \rightarrow \infty} X_n(\omega) - X(\omega) \right| = 0\right) = 1,$$

ou

$$\forall \varepsilon > 0, \sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty,$$

c'est-à-dire la série converge.

Remarque 1.3.1 *On parle d'estimateur fortement convergent lorsqu'on a convergence presque sûre de cet estimateur.*

Proposition 1.3.1 *soit T_n un estimateur convergent du paramètre θ , et soit ϕ une fonction de $\mathbb{R} \rightarrow \mathbb{R}$ continue au point θ . Alors $\phi(T_n)$ est un estimateur convergent de $\phi(\theta)$.*

La convergence en EMQ

Introduisant d'abord la définition suivante :

Définition 1.3.4 (*Erreur Quadratique Moyenne : EQM*) *On appelle erreur quadratique moyenne d'un estimateur T_n par rapport à θ la quantité :*

$$\begin{aligned} EQM(T_n, \theta) &= E[(T_n - \theta)^2] \\ &= V(T_n) + [B(n, \theta)]^2. \end{aligned}$$

avec $V(T_n)$ est la variance de l'estimateur tandis que $B(n, \theta)$ est son biais.

On dit que l'estimateur T_n en moyenne quadratique si

$$\lim_{n \rightarrow \infty} EQM(T_n, \theta) = 0.$$

Proposition 1.3.2 *Si un estimateur est sans biais ou asymptotiquement sans biais et si sa variance tend vers 0 lorsque n tend vers l'infini, alors il est convergent en EQM.*

Exemple 1.3.4 *L'estimateur $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur convergent de $m = E(X)$ en EQM, car :*

1. \bar{X} est un estimateur sans biais de m , (voir l'exemple 1.3.1),

$$2. V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n} \text{ et } \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0.$$

1.3.3 Choix d'un estimateur

Lors de l'estimation d'un paramètre il est possible qu'on ait à notre disposition plus d'un estimateur. Dans ce cas, il est de toute évidence de choisir le meilleur d'entre eux. Les deux définitions suivantes nous fournissent un premier aperçu sur sélection d'un estimateur.

Définition 1.3.5 Soient T_1 et T_2 deux estimateurs sans biais de θ , on dit que T_1 est plus efficace que T_2 s'il est préférable au sens de la variance c'est-à-dire

$$V(T_1) \leq V(T_2).$$

Définition 1.3.6 Soient T_1 et T_2 deux estimateurs de θ , on dit que T_1 est meilleur (plus performant) que T_2 au sens de l'EQM, si la variance c'est-à-dire

$$EQM(T_1) \leq EQM(T_2).$$

Exemple 1.3.5 Soit X_1, \dots, X_n un n échantillon *i.i.d* et

$$T_1 = \frac{X_1 + X_2}{2}$$
$$T_2 = \bar{X} + \frac{X_1 - X_2}{2}$$

deux estimateurs de $m = E(X)$.

On a pour T_1 ,

$$\begin{aligned}
 E(T_1) &= E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}E(X_1 + X_2) \\
 &= \frac{1}{2}(E(X_1) + E(X_2)) = E(X_1) \\
 &= m
 \end{aligned} \tag{1.2}$$

et

$$\begin{aligned}
 V(T_1) &= V\left(\frac{X_1 + X_2}{2}\right) \\
 &= \frac{1}{4}(V(X_1) + V(X_2)) = \frac{1}{4}2V(X_1) \\
 &= \frac{\sigma^2}{2}.
 \end{aligned} \tag{1.3}$$

Et pour T_2 on a :

$$\begin{aligned}
 E(T_2) &= E\left(\frac{1}{n}\sum_{i=1}^n X_i + \frac{X_1 - X_2}{2}\right) = E(\bar{X}) + \frac{1}{2}(E(X_1) - E(X_2)) \\
 &= E(\bar{X}) \\
 &= m \text{ (voir exemple 1.3.1),}
 \end{aligned} \tag{1.4}$$

et

$$\begin{aligned}
 V(T_2) &= V\left(\bar{X} + \frac{X_1 - X_2}{2}\right) \\
 &= V(\bar{X}) + \frac{1}{4}(V(X_1) + V(X_2)) \\
 &= V(\bar{X}) + \frac{1}{4}2V(X_1) \\
 &= \frac{1}{n}\sigma^2 + \frac{1}{2}\sigma^2 \quad (\text{voir exemple 1.3.2}) \\
 &= \left(\frac{2+n}{2n}\right)\sigma^2.
 \end{aligned} \tag{1.5}$$

On a d'une part, d'après (1.2) et (1.4) on remarque que les deux estimateurs T_1 et T_2 sont des estimateur sans biais de m . D'autre part, d'après (1.3) et (1.5) on $V(T_1) \leq V(T_2)$,. Ainsi, on conclut que l'estimateur T_1 est plus efficace que l'estimateur T_2 .

1.4 Statistique exhaustive

Soit X_1, X_2, \dots, X_n un n -échantillon d'une variable aléatoire X . On notera $L(x_1, x_2, \dots, x_n; \theta)$ la densité de (X_1, X_2, \dots, X_n) si X est une variable absolument continue, et la probabilité conjointe $P(X_1 = x_1 \cap \dots \cap X_n = x_n)$ si X est une variable discrète.

La fonction $L(x; \theta)$, considéré comme fonction de θ seul, est appelé "vraisemblance" de θ .

Soit également, T une statistique en fonction de X_1, X_2, \dots, X_n de loi $g(t; \theta)$.

Définition 1.4.1 La statistique T sera dite exhaustive, si l'on a $L(x, \theta) = g(t, \theta)h(x)$. En d'autres termes si la densité conditionnelle de l'échantillon est indépendante du paramètre θ .

$$P(X = x/T = t) \text{ ne dépend pas de } \theta.$$

La définition ci-dessous signifie qu'une fois T est connu, aucune valeur de l'échantillon

ni aucune autre statistique ne nous apportera de renseignements supplémentaire sur θ :

Exemple 1.4.1 Soit X une v.a qui suit la loi de Poisson de paramètre θ , soit (x_1, \dots, x_n) une réalisation d'un n -échantillon iid, et soit la statistique

$$T = \sum_{i=1}^n X_i.$$

On montre que la statistique T est une statistique exhaustive pour θ .

On a :

$$P(X_1, \dots, X_n / T = t) = \frac{P(X_1, \dots, X_n, T = t)}{P(T = t)}$$

La statistique T suit une loi de Poisson de paramètre $n\theta$, alors

$$P(T = t) = \frac{\exp(-n\theta) (n\theta)^t}{t!},$$

et

$$\begin{aligned} P(x_1, \dots, x_n, T = t) &= P\left(x_1, \dots, x_{n-1}, T - \sum_{i=1}^{n-1} x_i\right) \\ &= \prod_{i=1}^n \left(\frac{\exp(-\theta) (\theta)^{x_i}}{(x_i)!} \right) \frac{\exp(-\theta) (\theta)^{T - \sum_{i=1}^{n-1} x_i}}{(T - \sum_{i=1}^{n-1} x_i)!} \\ &= \frac{\exp(-n\theta) (\theta)^t}{(x_1)! (x_2)! (x_{n-1})! (T - \sum_{i=1}^{n-1} x_i)!} \end{aligned}$$

d'où

$$P(x_1, \dots, x_n / T = t) = \frac{t!}{(x_1)! (x_2)! (x_{n-1})! (T - \sum_{i=1}^{n-1} x_i)!}.$$

La probabilité conditionnelle ne dépend pas de θ , donc T est bien une statistique exhaustive pour θ .

Dans l'exemple ci-dessus, on a vu que le calcul de la probabilité conditionnelle est loin d'être immédiat. Le théorème 1.4.1 va nous simplifier la tâche.

1.4.1 Théorème de factorisation

Soit un n -échantillon d'une variable aléatoire X . On note $L(x, \theta)$ la densité jointe de $X = (x_1, \dots, x_n)$, elle est appelée fonction de vraisemblance de $\theta \in \Theta$ (ouvert de $\mathbb{R}^d, d \geq 1$).

Théorème 1.4.1 (*Théorème de factorisation*) *La statistique $T(X)$ est une statistique exhaustive pour θ si et seulement s'il existe deux fonctions mesurables g et h à valeurs positives dans \mathbb{R}_+ telle que la densité jointe se met sous la forme :*

$$L(x, \theta) = h(x) \times g(T(x), \theta), \theta \in \Theta; x \in \mathbb{R}^n,$$

où

$$L(x, \theta) = \begin{cases} \prod_{i=1}^n f(x_i, \theta), & \text{si } X \text{ est absolument continue;} \\ \prod_{i=1}^n P(x_i, \theta), & \text{si } X \text{ est discrète.} \end{cases}$$

avec $\theta \in \Theta, x \in \mathbb{R}^n, g(T(x), \theta)$ est la densité de la statistique T et $h(x)$ ne dépend pas de θ .

Exemple 1.4.2 *Soit $X \rightsquigarrow U[0, \theta], \theta > 0$, et $X = (x_1, \dots, x_n)$ un échantillon,*

$$T = \sup_{1 \leq i \leq n} x_i.$$

D'après le théorème de factorisation, la statistique T est exhaustive car :

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & \text{si } 0 \leq x \leq \theta; \\ 0, & \text{sinon.} \end{cases}$$

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

La fonction de répartition de T :

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

Par conséquent, la densité de T est donnée par :

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

On constate que

$$h(x) = \frac{L(X, \theta)}{g(t, \theta)} = \frac{1}{nt^{n-1}} \quad 0 \leq x \leq \theta$$

ne dépend pas de θ .

Remarque 1.4.1 Le principe de factorisation nous donne un moyen de reconnaître si une statistique est exhaustive, mais ne permet pas de la construire ou même de savoir s'il en existe une.

1.4.2 Théorème de Darmais

Soit X une v.a. dont le domaine de définition ne dépend pas de θ . Le théorème de Darmais stipule que une condition nécessaire et suffisante pour que l'échantillon

(X_1, X_2, \dots, X_n) admette une statistique exhaustive est que la densité $f(x, \theta)$ de la variable aléatoire X appartienne à une famille exponentielle.

La définition suivante nous mis au claire la notion "loi de famille exponentielle".

Définition 1.4.2 *On appelle famille exponentielle à paramètre θ toute loi de probabilité (discrète ou continue) dont la vraisemblance peut se mettre sous la forme :*

$$f(x; \theta) = \begin{cases} \exp[\alpha(\theta)\beta(x) + \gamma(\theta) + \delta(x)] & \text{si } x \in \chi \\ 0 & \text{sinon} \end{cases}$$

- α est une fonction deux fois différentiables à valeurs dans \mathbb{R}^d , $d \geq 1$.
- β est une fonction mesurable à valeurs dans \mathbb{R}^d , $d \geq 1$.
- γ est une fonction réelle deux fois différentiables et ne dépend pas de x .
- δ est une fonction borélienne ne dépend pas de θ .

Exemple 1.4.3 *Soit X v.a suit une loi γ de paramètre θ inconnu alors :*

$$\begin{aligned} f(x, \theta) &= \frac{1}{\Gamma(\theta)} \exp(-x) x^{\theta-1} \\ &= \exp \left[\ln \left(\frac{1}{\Gamma(\theta)} \exp(-x) x^{\theta-1} \right) \right] \\ &= \exp [-\ln \Gamma(\theta) - x + (\theta - 1) \ln(x)] \end{aligned}$$

tel que $\Gamma(\theta)$ est la fonction gamma définie par

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

Alors,

$$\begin{cases} \beta(x) = \ln(x) \\ \alpha(\theta) = \theta - 1 \\ \gamma(\theta) = -\ln \Gamma(\theta) \\ \delta(x) = -x \end{cases}$$

et on a l'application

$$x_i \rightarrow \sum_{i=1}^n \beta(x_i) = \sum_{i=1}^n \ln(x_i),$$

est bijective et continûment différentiable pour tout i . Donc, la statistique exhaustive est :

$$\sum_{i=1}^n \ln(X_i) = \ln\left(\prod_{i=1}^n X_i\right).$$

1.5 Estimateur efficace

Précédemment nous avons abordé le problème du choix d'un estimateur lorsque plusieurs statistiques ont été proposées. Dans cette section nous allons introduire deux nouvelles notions (Information de Fisher et Borne de Fréchet) qui nous permettra de vérifier si un estimateur T est efficace ou non.

1.5.1 Information de Fisher

Définition 1.5.1 On appelle quantité d'information de Fisher $I_n(\theta)$ apportée par un n échantillon sur le paramètre θ , la quantité positive où nulle suivante :

$$I_n(\theta) = E\left(\left(\frac{\partial \ln L(X, \theta)}{\partial \theta}\right)^2\right) \quad (1.6)$$

où :

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

Pour beaucoup de modèle statistique le développement de l'expression (3.8) en sa forme brute est très complexe. Le théorème suivant nous fournis une version équivalente à l'expression (3.8) dont le développement est moins complexe.

Théorème 1.5.1 Si le domaine de définition de X ne dépend pas du paramètre θ

alors :

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

Exemple 1.5.1 Soit X une v.a tel que $X \rightsquigarrow N(m, \sigma^2)$ alors on a :

$$\left\{ \begin{array}{l} L(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right), \\ \ln L(x, \theta) = -\frac{1}{2} \ln(2\pi) - \ln(\sigma) - \frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2, \\ \frac{\partial \ln L(x, \theta)}{\partial m} = \frac{1}{\sigma^2} (x - m), \\ \frac{\partial^2 \ln L(x, \theta)}{\partial m^2} = -\frac{1}{\sigma^2}, \end{array} \right.$$

Ainsi, l'information de Fisher pour cette v.a est donnée par :

$$I(m) = -E \left(\frac{\partial^2 \ln L(x, \theta)}{\partial m^2} \right) = \frac{1}{\sigma^2}.$$

Propriétés de l'information de Fisher

La quantité d'information de Fisher vérifiée les propriétés suivantes :

1. **La positivité** : On a

$$I_n(\theta) = V(S(X, \theta)) \geq 0, \quad \forall \theta.$$

2. **L'additivité** : Une autre propriété fondamentale de l'information de Fisher est son additivité. L'information résultant de deux variables aléatoires indépendantes est la somme de leurs informations de Fisher. Soient X, Y deux v.a indépendantes. On note $I_X(\theta)$ et $I_Y(\theta)$ les informations au point θ respectivement fournies par X, Y . L'information $I_{(X,Y)}(\theta)$ au point θ du couple (X, Y)

peut être quantifier par :

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

1.5.2 Estimateur sans biais de variance minimale

Il existe plusieurs théorèmes qui montrent que l'estimateur de variance minimale est lié à l'existence d'une statistique exhaustive.

Théorème 1.5.2 (unicité) *S'il existe un estimateur de θ sans biais de variance minimale, il est unique.*

Théorème 1.5.3 (Rao-Blackwell) *Soit T un estimateur sans biais de θ quelconque et U une statistique exhaustive pour θ . Alors*

$$T^* = E(T/U),$$

est un estimateur sans biais de θ au moins aussi bon que T .

Théorème 1.5.4 *S'il existe une statistique exhaustive U de θ , alors l'estimateur T sans biais de θ de variance minimale est unique et ne dépend que de U .*

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

Le résultat suivant nous indique que la variance d'un estimateur ne peut être inférieure à une certaine borne, qui dépend de la quantité d'information de Fisher apportée par l'échantillon sur le paramètre θ : Si le domaine de définition de X ne dépend pas de θ , on a pour tout estimateur T sans biais de θ :

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

et si T est un estimateur sans biais de $k(\theta)$:

$$V(T) \geq \frac{[k'(\theta)]^2}{I_n(\theta)},$$

où k est une fonction dérivable.

Définition 1.5.2 *On dit que T est un estimateur efficace si :*

1. T est un estimateur sans biais de θ et

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

2. T est un estimateur sans biais de $k(\theta)$ et

$$V(T) = \frac{[k'(\theta)]^2}{I_n(\theta)}.$$

1.5.4 Théorème sur l'efficacité

La borne de Cramer-Rao ne peut être atteinte que si la loi de X appartient à la famille exponentielle :

$$f(x, \theta) = \exp[\beta(x) \alpha(\theta) + \delta(x) + \gamma(\theta)]$$

car T est nécessairement exhaustive pour θ . donc, il n'existe qu'une seule fonction $k(\theta)$ qui puisse être estimée efficacement :

$$h(\theta) = -\frac{\gamma'(\theta)}{\alpha'(\theta)},$$

l'estimateur de $h(\theta)$ est :

$$T_n = \frac{1}{n} \sum_{i=1}^n \beta(X_i)$$

la variance minimale est

$$\begin{aligned} V(T_n) &= -\frac{1}{n\alpha'(\theta)} \frac{\partial}{\partial \theta} \left(\frac{\gamma'(\theta)}{\alpha'(\theta)} \right) \\ &= \frac{h'(\theta)}{n\alpha(\theta)}. \end{aligned}$$

Conclusion

Dans la pratique, lorsque on dispose d'un modèle paramétrique $f(x, \theta)$, le problème est celui de l'estimation du paramètre θ grâce à laquelle le modèle en question sera entièrement connu. Il existe plusieurs façons de construire un estimateur pour ce paramètre. Les plus populaires sont la méthode des moments et celle du maximum de vraisemblance.

Dans le chapitre qui suit, nous allons s'intéresser à la méthode du maximum de vraisemblance pour l'estimation d'un paramètre.

Chapitre 2

Estimation par maximum de vraisemblance

Introduction

Le présent chapitre est consacré à l'application de la méthode du maximum de vraisemblance pour l'estimation des paramètres de quelques lois usuelles. En effet, pour comprendre au mieux le mécanisme de construction des estimations via la méthode du maximum de vraisemblance et de vérifier leurs qualité, des exemples détaillé ont été présentés. Et ceci toute en considérant le cas de variables aléatoires discrètes et de variables aléatoires continues, à savoir : l'estimation des paramètres de la loi Normale et la loi Exponentielle pour le cas de variables aléatoires continues et l'estimation des paramètres de la loi de Bernoulli et de la loi de Poisson pour le cas de variables aléatoires discrètes.

2.1 Méthode ”maximum de vraisemblance”(M.V)

En statistique, l'estimateur du maximum de vraisemblance (**EMV**) 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. Cette méthode a été développée par le statisticien Ronald Aylmer Fisher en 1922 [1, 8]. Cette méthode est la plus utilisée en pratique. Elle est optimale et asymptotiquement efficace lorsque la taille de l'échantillon est assez grande ($n \geq 30$), où elle fournit des estimateurs de très bonne qualité.

2.2 Définitions

Définition 2.2.1 Soit $L(x; \theta)$ la fonction de vraisemblance de l'échantillon $x = (x_1, x_2, \dots, x_n)$ issu de la variable aléatoire X calculée au point x en fonction de θ donnée par :

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i, \theta). \quad (2.1)$$

Définition 2.2.2 L'estimateur du maximum de vraisemblance (EMV) pour θ est la statistique $\hat{\theta}$ telle que :

$$L(x; \hat{\theta}) = \max_{\theta \in \Theta} L(x; \theta), \quad (2.2)$$

avec $L(x; \theta)$ est donné par 2.1.

Autrement dit, le principe de cette méthode est de trouver un estimateur $\hat{\theta}$ qui maximise la fonction de vraisemblance $L(x; \theta)$.

Remarque 2.2.1 Maximiser la fonction $L(x; \theta)$ revient à maximiser $\ln L(x; \theta)$, (c.à.d elles atteignent leurs maximum au même point), Donc il est plus commode

de maximiser $\ln L(x; \theta)$.

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(x; \theta) = \arg \max_{\theta \in \Theta} (\ln L(x; \theta)).$$

Il s'agit de trouver un estimateur de θ , qui maximise la fonction de vraisemblance de l'échantillon. la valeur $\hat{\theta}$ qui maximise $L(x; \theta)$ serait un bon estimateur car elle donne la plus grande probabilité pour cet échantillon. Si le domaine de X ne dépend pas de θ et si la fonction de vraisemblance est deux fois différentiable (dérivable) par rapport à θ , alors $\hat{\theta}$ est une solution du système suivant :

$$\begin{cases} \frac{\partial}{\partial \theta} \ln(L(x, \theta)) & = 0, \quad (\text{pour trouver } \hat{\theta}) \\ \frac{\partial^2}{\partial \theta^2} \ln(L(x; \theta)) \Big|_{\theta=\hat{\theta}} & < 0 \quad (\text{pour assurer l'existence du } \max_{\theta \in \Theta} (\ln L(x; \theta))) \end{cases} \quad (2.3)$$

2.3 Exemples usuels d'EMV

Quand les observations sont toutes discrètes(respectivement tous continues), on appelle fonction de vraisemblance de l'échantillon x_1, x_2, \dots, x_n , la fonction de paramètre θ :

$$\begin{cases} P(X_1 = x_1, \dots, X_n = x_n; \theta), & \text{si les } X_i \text{ sont des v.a discrètes,} \\ f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta), & \text{si les } X_i \text{ sont des v.a continues.} \end{cases} \quad (2.4)$$

L'objectif de la présente section est présenté quelques exemples d'application de la méthode en question dans le cas de variables discrètes et continues.

2.3.1 Cas de variables discrètes

On appelle fonction de vraisemblance de l'échantillon $x = (x_1, x_2, \dots, x_n)$ issu d'une variable aléatoire discrète X , la fonction de θ définie par :

$$L(x_1, x_2, \dots, x_n; \theta) = P((X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)),$$

sous l'hypothèse que les X_i soit i.i.d, alors $L(x; \theta)$ peut être réécrite comme suit

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

Loi de Poisson

Soit l'échantillon x_1, x_2, \dots, x_n avec n - réalisation issues d'une v.a de la loi de Poisson dont la densité est :

$$P(X = x_i; \lambda) = \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda),$$

avec $x_i \in \mathbb{N}$ et $\lambda > 0$.

La fonction de vraisemblance s'écrit alors :

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \lambda) &= \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) \\ &= \frac{\lambda^{\left(\sum_{i=1}^n x_i\right)}}{\prod_{i=1}^n x_i!} \exp(-n\lambda). \end{aligned}$$

étant donné que la vraisemblance est positive, on considère son logarithme népérien :

$$\begin{aligned} \ln L(x_1, x_2, \dots, x_n; \lambda) &= \ln \exp(-\lambda n) + \ln \left(\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \right) \\ &= -\lambda n + \sum_{i=1}^n \ln \left(\frac{\lambda^{x_i}}{x_i!} \right) \\ &= -\lambda n + \sum_{i=1}^n x_i \ln \lambda - \sum_{i=1}^n \ln(x_i!) \end{aligned}$$

La dérivée première s'annule quand

$$\begin{aligned} \frac{\partial \ln L(x_1, x_2, \dots, x_n; \lambda)}{\partial \lambda} &= 0, \\ -n + \frac{\sum_{i=1}^n x_i}{\lambda} &= 0, \\ \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i &= \bar{X}. \end{aligned} \tag{2.5}$$

La dérivée seconde s'écrit :

$$\frac{\partial^2 \ln L(x_1, x_2, \dots, x_n; \lambda)}{\partial \lambda^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0, \forall \lambda.$$

Donc $\hat{\lambda}$ donnée par (2.5) est un extremum maximal.

L'estimateur $\hat{\lambda}$ est-il sans biais ?

$$\begin{aligned} E(\hat{\lambda}) &= E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{n}{n} E(x) \\ &= \lambda. \end{aligned}$$

D'où $\hat{\lambda}$ est un estimateur sans biais.

Est-ce que l'estimateur $\hat{\lambda}$ est convergent ?

$$\begin{aligned} V(\hat{\lambda}) &= V\left(\frac{1}{n} \sum_{i=1}^n x_i\right), \\ &= \frac{n}{n^2} V(x), \\ &= \frac{\lambda}{n}, \end{aligned}$$

ainsi

$$\lim_{n \rightarrow \infty} V(\hat{\lambda}) = \lim_{n \rightarrow \infty} \frac{\lambda}{n} = 0.$$

Le fait que $V(\hat{\lambda})$ tend vers 0 lorsque n tend vers l'infinie alors $\hat{\lambda}$ est convergent.

Est-ce que l'estimateur $\hat{\lambda}$ est efficace ? En statistique, la borne Cramér-Rao exprime une borne inférieure sur la variance d'un estimateur sans biais, basée sur l'information de Fisher. Elle est aussi appelée borne de Fréchet-Darmois-Cramér-Rao (ou borne *BICR*) en l'honneur de Maurice Fréchet, Georges Darmois, Harald Cramér et Calyampudi Radhakrishna Rao voir [4]. Elle énonce que l'inverse de l'information de Fisher $I(\lambda)$, d'un paramètre λ , est un minorant de la variance d'un estimateur sans biais de ce paramètre (noté λ).

$$BICR = \frac{1}{I(\lambda)},$$

où

$$\begin{aligned}
 I(\lambda) &= -E\left(\frac{\partial^2 \log L(x_1, x_2, \dots, x_n, \lambda)}{\partial^2 \lambda}\right) \\
 &= -E\left(-\frac{\sum_{i=1}^n x_i}{\lambda^2}\right) \\
 &= \sum \frac{E(x)}{\lambda^2} = \frac{n}{\lambda} \\
 BICR &= V(\hat{\lambda}).
 \end{aligned}$$

Donc $\hat{\lambda}$ est un estimateur efficace, le fait qu'il est sans biais donc c'est un estimateur "MVUE" (sans biais de variance minimale).

Loi de Bernoulli

Soit X une variable aléatoire de loi de Bernoulli de paramètre p ($0 < p < 1$), on note

$$X \rightsquigarrow Ber(p).$$

La densité de probabilité de X est

$$P(X = x) = p^x (1 - p)^{1-x}, \quad \forall x \in \{0; 1\} \text{ et } \forall p \in]0, 1[$$

On suppose que p inconnue, soit (x_1, x_2, \dots, x_n) un n -échantillon (*i.i.d.*). La fonction de vraisemblance de l'échantillon est donnée comme suit :

$$\begin{aligned}
 L(x_1, x_2, \dots, x_n; p) &= \prod_{i=1}^n P(X_i = x_i, p), \\
 &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}, \\
 &= p^s (1-p)^{n-s}, \quad \text{tel que } s = \sum_{i=1}^n x_i.
 \end{aligned}$$

Calculons la dérivée première par rapport à p :

$$\begin{aligned}
 \frac{\partial L(x_1, x_2, \dots, x_n; p)}{\partial p} &= 0, \\
 \Rightarrow sp^{s-1}(1-p)^{n-s} - p^s(n-s)(1-p)^{n-s-1} &= 0, \\
 \Rightarrow sp^{s-1}(1-p)^{n-s} = p^s(n-s)(1-p)^{n-s-1}, \\
 \Rightarrow s(1-p) = p(n-s), \\
 \Rightarrow s = pn,
 \end{aligned}$$

donc

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

alors

$$\hat{p} = \bar{X}.$$

Dans le but de vérifier que \hat{p} est un extremum maximal, on calcule la dérivée seconde de $L(x; p)$ par rapport à p :

$$\frac{\partial^2 L(x_1, x_2, \dots, x_n; p)}{\partial p^2} = -\frac{s}{p^2} - \frac{n-s}{(p-1)^2} < 0,$$

Le fait que $s > 0$ et $n - s > 0$, alors on constate que l'inégalité suivante est vraie :

$$\frac{\partial^2 L(x_1, x_2, \dots, x_n; p)}{\partial p^2} < 0$$

Donc la fonction de vraisemblance est maximale au :

$$\hat{p} = \bar{X}.$$

Il est à souligner qu'avec même démarche suivie dans l'exemple d'estimation du paramètre de loi de Poisson, on peut montrer ce qui suit :

$$E(\hat{p}) = p \tag{2.6}$$

$$V(\hat{p}) = \frac{p(1-p)}{n} \tag{2.7}$$

$$\lim_{n \rightarrow \infty} V(\hat{p}) = \lim_{n \rightarrow \infty} \frac{p(1-p)}{n} = 0 \tag{2.8}$$

$$BICR = \frac{p(1-p)}{n} \tag{2.9}$$

D'après ces résultats on constate que $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$ est un estimateur

1. sans biais,
2. convergent
3. et de variance minimale (car $BICR = V(\hat{p})$).

Ainsi on conclut que l'estimateur \hat{p} est un estimateur efficace de p .

2.3.2 Cas de variables continues

On appelle fonction de vraisemblance de v.a continue X de réalisation $x = (x_1, x_2, \dots, x_n)$, la fonction de θ :

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

avec $f(x_i, \theta)$ est un densité de probabilité de la v.a X .

Loi Exponentielle

Soit X une v.a de densité exponentielle de paramètre $\lambda > 0$. On note $X \rightsquigarrow \varepsilon(\lambda)$ dont la densité est :

$$f(x) = \begin{cases} \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right), & \text{si } x > 0 \\ 0. & \text{sinon} \end{cases}$$

Soit x_1, x_2, \dots, x_n un n -échantillon (n -réalisations) issue d'une v. a de loi exponentielle de paramètre λ inconnue. Pour estimer λ par la méthode EMV, construisant d'abord sa fonction de vraisemblance.

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \lambda) &= \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\lambda} \exp\left(-\frac{x_i}{\lambda}\right) \\ &= \lambda^{-n} \exp\left(-\frac{\sum_{i=1}^n x_i}{\lambda}\right). \end{aligned}$$

Pour faciliter les calculs, il est préférable d'utiliser $\ln L$ plutôt que L , dans ce cas, on a :

$$\ln L(x_1, x_2, \dots, x_n; \lambda) = -n \ln(\lambda) - \frac{\sum_{i=1}^n x_i}{\lambda}.$$

Calculons la dérivée première par rapport à λ :

$$\begin{aligned} \frac{\partial \ln L(x_1, x_2, \dots, x_n; \lambda)}{\partial \lambda} &= \frac{-n}{\lambda} + \frac{\sum_{i=1}^n x_i}{\lambda^2} = 0, \\ \Rightarrow \frac{1}{\lambda} &= \frac{n}{\sum_{i=1}^n x_i}, \\ \Rightarrow \lambda &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}. \end{aligned}$$

Pour assurer que le point obtenu est un extremum maximal, calculons la dérivée seconde par rapport à λ :

$$\left. \frac{\partial^2 \ln L(x_1, x_2, \dots, x_n; \lambda)}{\partial \lambda^2} \right|_{\lambda=\bar{X}} = \frac{n}{\lambda^2} - \frac{2 \sum_{i=1}^n x_i}{\lambda^3} \Big|_{\lambda=\bar{X}}.$$

On constate que

$$\left. \frac{\partial^2 \ln L(x_1, x_2, \dots, x_n; \lambda)}{\partial \lambda^2} \right|_{\lambda=\bar{X}} = \frac{-n}{\bar{X}^2} < 0.$$

Par conséquent $\hat{\lambda}$ est un extremum maximal, dont l'expression est :

$$\hat{\lambda} = \bar{X}.$$

Il est à souligner qu'avec même démarche suivie dans l'exemple d'estimation du paramètre de loi de Poisson, on peut montrer ce qui suit :

$$E(\hat{\lambda}) = \lambda, \tag{2.10}$$

$$V(\hat{\lambda}) = \frac{\lambda^2}{n}, \tag{2.11}$$

$$\lim_{n \rightarrow \infty} V(\hat{\lambda}) = \lim_{n \rightarrow \infty} \frac{\lambda^2}{n} = 0, \tag{2.12}$$

$$BICR = \frac{\lambda^2}{n}, \tag{2.13}$$

D'après ces résultats on constate que $\hat{\lambda}$ est un estimateur

1. sans biais,
2. convergent
3. et de variance minimale (car $BICR = V(\hat{\lambda})$).

Ainsi on conclut que l'estimateur $\hat{\lambda}$ est un estimateur efficace de λ .

Loi Normale

On suppose que $X \rightsquigarrow N(m, \sigma^2)$, sa densité est

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

Considérons d'abord l'estimation de m :

La fonction de vraisemblance est défini à partir de la densité f comme suit :

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

La vraisemblance qu'une population de moyenne m et de variance σ^2 , généré à partir d'un échantillon observé, s'avère donc être définie par :

$$\begin{aligned} L(x_1, x_2, \dots, x_n; m) &= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x_i-m}{\sigma}\right)^2\right) \right] \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i-m}{\sigma}\right)^2\right). \end{aligned}$$

Ensuite, on a le logarithme de la vraisemblance :

$$\ln L(x_1, x_2, \dots, x_n; m) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i-m}{\sigma}\right)^2.$$

La dérivée première est donnée comme suite :

$$\frac{\partial \ln L(x_1, x_2, \dots, x_n; m)}{\partial m} = \sum_{i=1}^n \frac{x_i - m}{\sigma^2}.$$

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

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

L'estimateur $\hat{m} = \frac{1}{n} \sum_{i=1}^n x_i$ de la moyenne d'une loi normale est

1. sans biais,
2. convergent
3. et de variance minimale (car $BICR = V(\hat{p})$).
4. est un estimateur efficace

Et ceci le fait que :

$$E(\hat{m}) = m \tag{2.14}$$

$$V(\hat{m}) = \frac{m}{n} \tag{2.15}$$

$$BICR = \frac{m}{n} \tag{2.16}$$

Considérons maintenant l'EMV de σ^2 :

Soit toujours

$$\begin{aligned} \ln L(x_1, x_2, \dots, x_n; \sigma^2) &= -n \ln(\sigma \sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma} \right)^2 \\ &= -\frac{n}{2} \ln(\sigma^2) - \ln(\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^2} \end{aligned}$$

L'équation de ln- vraisemblance est :

$$\frac{\partial \ln L(x_1, x_2, \dots, x_n; m; \sigma^2)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^4}$$

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

$$\frac{-n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^4} = 0 \Rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - m)^2 = 0.$$

Donc on remplace par sa valeur devient :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2. \quad (2.17)$$

D'après l'estimateur (2.17) on constate que deux situations sont envisageables, à savoir :

1. la moyenne m est connue : dans ce cas $\hat{\sigma}^2$ un estimateur sans biais et convergent de σ^2 .
2. la moyenne m est inconnue : dans ce cas m doit être substitué par son estimateur et l'expression (2.17) sera réécrite sous la forme

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2, \quad (2.18)$$

et ce dernier est un estimateur biaisé et convergent de σ^2 . Il est asymptotiquement sans biais.

Conclusion

Les exemples d'application de la méthode du maximum de vraisemblance présentés dans ce chapitre nous laissent croire qu'il ne réside aucune difficulté dans l'implémentation de la méthode en question. Malheureusement ceci n'est pas toujours le cas. En effet, il suffit d'envisager d'estimer les paramètres du modèle de Weibull ou Gamma pour se rendre compte de l'existence de difficulté de construction des estimations à l'aide de la méthode du maximum de vraisemblance. Certes le problème qu'on rencontre dans les deux modèles précédent (Weibull et Gamma) peut être surpassé à l'aide de l'outil informatique (logiciels). Cependant, le recours à l'outil informatique n'est pas toujours une solution. La justification de ce dernier fera l'objet du chapitre suivant.

Chapitre 3

Implémentation de la méthode

MV sur ordinateur

Introduction

Dans ce chapitre à travers des exemples d'application nous allons montrer le problème de la mise en œuvre de la méthode du maximum de vraisemblance sur l'outil informatique. Plus précisément en prenant la loi normale et la loi exponentielle comme exemples illustratifs nous allons montrer que l'implémentation de la méthode de vraisemblance sur machine dépend des paramètres de la loi considéré et la taille de l'échantillon, particulièrement nous allons montrer que l'application n'est possible dans certaines situations.

3.1 Position du problème l'implémentation de la méthode

Pour la plupart des lois de probabilité usuelles, l'estimateur du maximum de vraisemblance est défini d'une façon unique, et se calcule explicitement. Sur le plan théorique, il présente de nombreux avantages. Sous des hypothèses vérifiées par de nombreux modèles courants, on démontre qu'il est asymptotiquement sans biais et convergent. Cependant, dans le cadre de lois non usuelles c'est-à-dire dans le cadre de lois générales, la construction (l'obtention) de la statistique du paramètre $\theta(\theta \in \Omega \subseteq \mathbb{R}^d, d \geq 1)$ manuellement (théoriquement) n'est pas toujours évidente d'où, pour la mise en œuvre de cette technique, le recours à l'outil informatique est inévitable. Ainsi, on dispose de deux possibilités pour faire face à ce dernier problème.

1. dans le cas où le calcul des deux premières dérivés de la vraisemblance $L(x, \theta)$ par rapport à θ de l'échantillon $x = (x_1, x_2, \dots, x_n)$ est possible(existe), on peut utiliser les algorithmes des résolutions des systèmes d'équations non linéaires (pour plus de détails sur ces algorithmes le lecteur peut se référer par exemple aux [3, 2]), où le système à résoudre est :

$$\frac{\partial L(x, \theta)}{\partial \theta_k} = 0, \quad \text{pour } k = 1, 2, \dots, d. \quad (3.1)$$

Par la suite de vérifier, via la dérivée seconde de $L(x, \theta)$ par rapport à θ , que la solution du système (3.1) représente bien un point maximum.

Ce problème peut être rencontré même dans le cas de modèles usuels. En effet, l'estimation des paramètre du modèle Gamma et Weibull en est une preuve. À titre illustratif, soit l'exemple du modèle de la loi weibull suivant :

Exemple 3.1.1 (*Estimation du paramètre de la loi Weibull*)

Supposons que nous disposons d'un n-échantillon x_1, x_2, \dots, x_n i.i.d, issue d'une

loi de Weibull dont la fonction de répartition et la fonction de densité sont respectivement données comme suite :

$$F(x) = \exp(-x^\theta). \quad (3.2)$$

$$f(x; \theta) = \theta x^{\theta-1} \exp(-x^\theta). \quad (3.3)$$

La vraisemblance de l'échantillon est :

$$L(x_1, x_2, \dots, x_n; \theta) = \theta^n \prod_{i=1}^n x_i^{\theta-1} \exp\left(-\sum_{i=1}^n x_i^\theta\right).$$

En introduisant le logarithme, l'expression précédente peut être réécrite comme suite :

$$\ln L(x; \theta) = n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n x_i^\theta.$$

Après la dérivation de \ln -vraisemblance par rapport à θ on obtient :

$$\frac{\partial \ln L(x; \theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^n \ln x_i - \sum_{i=1}^n x_i^\theta \ln x_i.$$

Ainsi l'estimateur $\hat{\theta}$ du paramètre θ est déterminé par la résolution l'équation suivante :

$$\begin{aligned} \frac{\partial \ln L(x; \theta)}{\partial \theta} &= 0, \\ \Rightarrow \hat{\theta} &= \frac{n}{\sum_{i=1}^n (x_i^{\hat{\theta}} - 1) \ln x_i}. \end{aligned} \quad (3.4)$$

D'après l'équation (3.4), on ne peut pas obtenir une forme explicite pour l'estimateur de θ . Il est clair qu'elle n'est pas linéaire, elle ne peut donc se résoudre

que numériquement par approximations successives (par exemple la méthode du point fixe).

2. Dans certaines modèles l'expression de la première dérivée, de la vraisemblance de l'échantillon, est très complexe. Dans ce cas il est préférable d'utiliser directement les algorithmes d'optimisation pour la recherche du maximum de $L(x; \theta)$. Ci-dessous un exemple illustratif.

Exemple 3.1.2 (*Estimation du paramètre de lissage*)

Dans l'estimation à noyau d'une densité [5, 6, 7], l'estimateur du paramètre de lissage h par la méthode du maximum de vraisemblance est défini par :

$$\hat{h} = \arg \max_h \prod_{i=1}^n \hat{f}_i(x_i, h), \quad (3.5)$$

avec

$$\hat{f}_i(x_i, h) = \frac{1}{(n-1)h} \sum_{j=1, j \neq i}^n K\left(\frac{x_i - x_j}{h}\right), \quad (3.6)$$

où $h = h(n)$ est appelé paramètre de lissage et la fonction K est appelée noyau.

D'après l'expression (3.5) il est clair que l'expression de sa dérivée est très complexe, car on a affaire à la dérivation du produit de n fonctions.

Cependant, l'implémentation des deux solutions précédentes sur l'outil informatique (les logiciels de calculs) n'est pas toujours réalisable. Ceci est dû à la précision des calculs de logiciels utilisés. En effet, l'outil informatique ne peut prendre en considération des valeurs hors de l'intervalle $] -10^{309}; -10^{324}[\cup]10^{-324}; 10^{309}[$, où les valeurs de l'intervalle $[-10^{-324}; 10^{-324}]$ sont considérées comme un zéro absolu, tandis que les valeurs supérieures ou égales en valeur absolue à 10^{309} sont considérées infini ($\pm\infty$). Or $L(x, \theta)$ peut prendre des valeurs en dehors de l'intervalle $] -10^{309}; -10^{-324}[\cup]10^{-324}; 10^{309}[$, ce qui signifie que l'implémentation de la méthode MV n'est pas possible dans ce cas. La figure 3.1, présente un exemple illustratif de la précision du logiciel Matlab et le logiciel R .

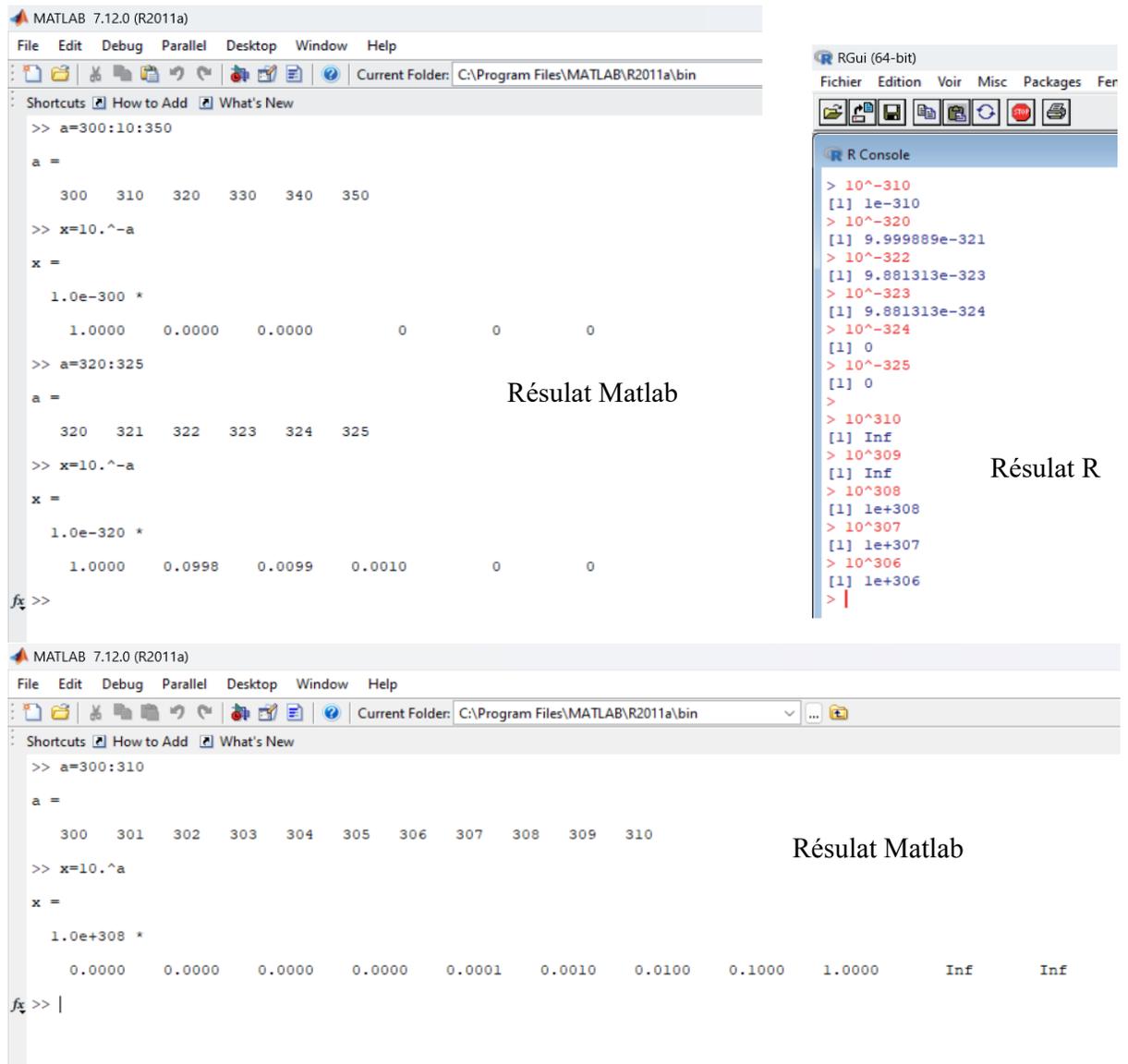


FIGURE 3.1 – Précision des calculs sous le logiciel Matlab et le logiciel *R*.

3.2 Exemples illustratifs du problème

3.2.1 Exemple 01 : Cas de la loi Normale

Soit $X \rightsquigarrow N(m, \sigma^2)$. Il est aisé de vérifier que pour tout x , m et $\sigma > 0$ l'inégalité suivante est vraie.

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

Ainsi, la vraisemblance associée à un n -échantillon issu de la v.a X peut être majoré comme suit :

$$L(x_1, \dots, x_n; m) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2\right) \leq \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n = B_n. \quad (3.7)$$

Si on suppose que $\frac{1}{\sigma\sqrt{2\pi}} < 1$, c'est-à-dire $\sigma > \frac{1}{\sqrt{2\pi}}$ alors en fait et à mesure que la taille de l'échantillon augmente la borne B_n se rapproche de zéro (voir figure 3.3 à gauche). Par conséquent,

$$\forall n \geq n_0, B_n \leq 10^{-324} \Rightarrow L(x_1, \dots, x_n; m) \leq 10^{-324},$$

tel que

$$n_0 = \left\lceil \frac{324 \ln(10)}{\ln(\sigma\sqrt{2\pi})} \right\rceil + 1.$$

avec $\lfloor A \rfloor$ est la partie entière inférieure de A (exemple : $\lfloor 10.95 \rfloor = 10$). Et dans ce cas, bien évidemment sur machine,

$$B_n = 0 \Rightarrow L(x_1, \dots, x_n; m) = 0.$$

Le constat ci-dessus se traduit par la recherche d'un maximum d'une fonction nulle sur tout son domaine de définition, et ceci n'a aucun sens. Par conséquent, l'obtention de l'estimateur recherché n'est pas possible. Autrement dit, si on dispose d'un échantillon, issu d'une loi normale de moyenne m et d'écart-type $\sigma > \frac{1}{\sqrt{2\pi}}$, de la taille est supérieure à n_0 , alors le recours à l'outil informatique pour la recherche du maximum de vraisemblance en sa forme brute est en vain.

3.2.2 Exemple 02 : Cas de la loi Exponentielle

Pour appuie le constat de l'exemple précédent introduisant un deuxième exemple dans lequel nous considérons le modèle exponentielle.

Soit $X \rightsquigarrow Exp(\lambda > 0)$ dont la densité est définie par

$$f(x, \lambda) = \begin{cases} \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right), & \text{si } x > 0 \\ 0. & \text{sinon} \end{cases}$$

Il est aisé de vérifier que

$$f(x, \lambda) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) \leq \frac{1}{\lambda}, \forall x \geq 0. \quad (3.8)$$

Considérant maintenant x_1, x_2, \dots, x_n un n -échantillon issue d'une v.a. de loi exponentielle de paramètre λ inconnue. De l'équation (3.8) on a

$$f(x_i) = \frac{1}{\lambda} \exp\left(-\frac{x_i}{\lambda}\right) \leq \frac{1}{\lambda}, \text{ pour } i = 1, 2, \dots, n.$$

Ainsi, la vraisemblance associée à un n -échantillon issu de X peut être majoré comme suit :

$$L(x_1, \dots, x_n; \lambda) = \frac{1}{\lambda^n} \exp\left(-\frac{\sum_{i=1}^n x_i}{\lambda}\right) \leq \left(\frac{1}{\lambda}\right)^n = B_n.$$

Si on suppose que $\frac{1}{\lambda} < 1$, c'est-à-dire $\lambda > 1$, alors on peut constater qu'en fur et à mesure que la taille de l'échantillon augmente la borne B_n décroît et se rapproche de zéro (voir figure 3.3 à droite). Par conséquent, l'écriture suivante aura lieu

$$\forall n \geq n_0, B_n \leq 10^{-324} \Rightarrow L(x_1, \dots, x_n; \lambda) \leq 10^{-324},$$

tel que

$$n_0 = \left\lceil \frac{324 \ln(10)}{\ln(\lambda)} \right\rceil + 1.$$

En guise de conclusion, on ne peut implémenter cette technique pour des échantillons de taille $n > n_0$, ceci le fait que sur machine on a ce qui suit : $B_n = 0 \Rightarrow L(x_1, \dots, x_n; \lambda) = 0$, et cela $\forall \lambda > 0$.

Remarque 3.2.1 *Dans la pratique $L(x_1, \dots, x_n; \theta)$ peut prendre la valeur zéro (au sens l'outil informatique) même pour des échantillons de taille inférieure à n_0 car $L(x_1, \dots, x_n; \theta) \leq B_n$.*

```

Editor - C:\Users\DELL\Desktop\ApplicationVraisemblance\borneExp.m
1 function borneExp()
2 mu=.21:.005:.25;
3 M=length(mu);
4 n=30:40;
5 N=length(n);
6 B=zeros(N,M);
7 for j=1:M
8     for i=1:N
9         B(i,j)=(mu(j)).^n(i);
10    end
11 end
12 plot(n,B)
13 xlabel('n')
14 ylabel('L_n(X_i,\mu)')
15
16 end

Editor - C:\Users\DELL\Desktop\ApplicationVraisemblance\borneNorm
1 function borneNormal()
2 sigma=0.45:0.02:.55;
3 M=length(sigma);
4 n=30:100;
5 N=length(n);
6 B=zeros(N,M);
7 for j=1:M
8     for i=1:N
9         B(i,j)=1/(sigma(j)*sqrt(2*pi)
10    end
11 end
12 plot(n,B)
13 xlabel('n')
14 ylabel('L_n(X_i,\mu)')
15
16 end
17
    
```

FIGURE 3.2 – Code source programme Matlab implémenter pour le calcul de borne de majoration de la fonction de vraisemblance : cas de loi normale (à droite) et cas de loi exponentielle (à gauche)

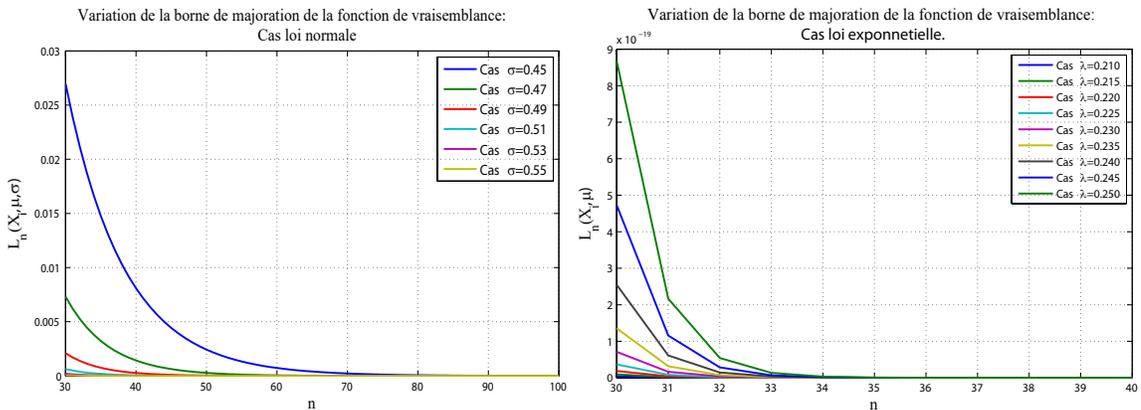


FIGURE 3.3 – Exemples de variation de la borne de majoration de la fonction de vraisemblance : cas de loi normale (à gauche) et cas de loi exponentielle (à droite)

3.2.3 Exemple 03 : Illustration par simulation

Dans cette section nous allons présenter quelques exemples de la variation de la fonction de vraisemblance en fonction de la taille de l'échantillon et le(s) paramètre(s) du modèle considéré.

Pour répondre à notre objectif nous avons considéré 3 lois à savoir la loi normale de paramètres m et σ^2 , la loi exponentielle de paramètre λ , la loi de Poisson de paramètre λ .

Pour les calculs numériques nous faisons varier la taille de l'échantillon et nous avons fixé le reste des paramètres, comme suit :

- Cas de loi normale : $\sigma = 2$ et $m = 0$ et $n \in \{10, 11, \dots, 15\}$;
- Cas de loi exponentielle : $\lambda = 3$ et $n \in \{15, 16, \dots, 20\}$;
- Cas de loi de Poisson : $\lambda = 3$ et $n \in \{10, 11, \dots, 15\}$;

Les résultats graphiques obtenus pour ces derniers paramètres sont présentés dans les figures 3.4–3.6 respectivement pour normale, exponentielle et Poisson.

D'après les résultats graphiques on constate que pour les trois modèles retenues le maximum de la vraisemblance (qui correspond au point de l'estimateur) décroît en fait et à mesure que la taille de l'échantillon augmente. Ceci signifie que pour chacun des trois modèles il existe une taille d'échantillon n_0 pour laquelle la valeur du maximum de leurs fonctions vraisemblances est inférieure ou égale à 10^{-324} . Par conséquent, pour tous autres échantillons de taille supérieure à n_0 la valeur de la vraisemblance est considérée sur machine nulle (un zéro absolu) pour toute valeur du paramètre recherché d'où la recherche du maximum est impossible.

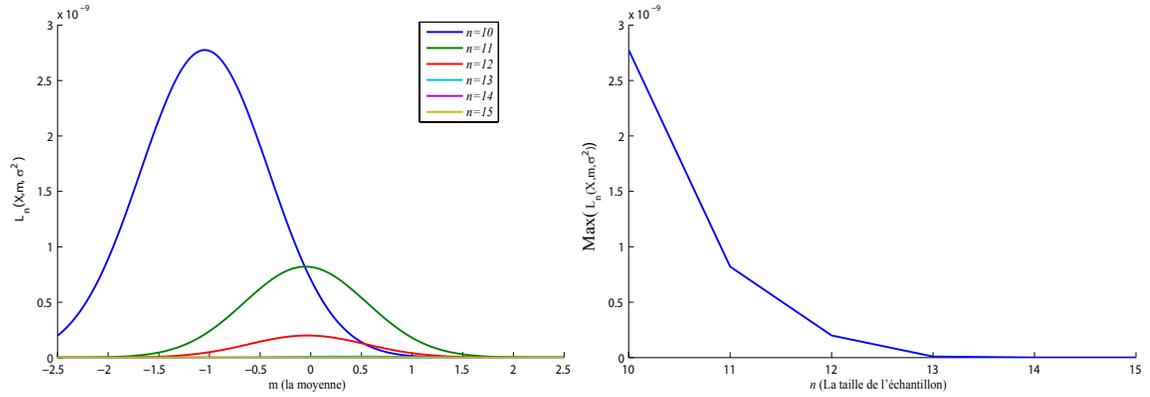


FIGURE 3.4 – Exemples de variation de la fonction de vraisemblance et son Max : cas de loi normale

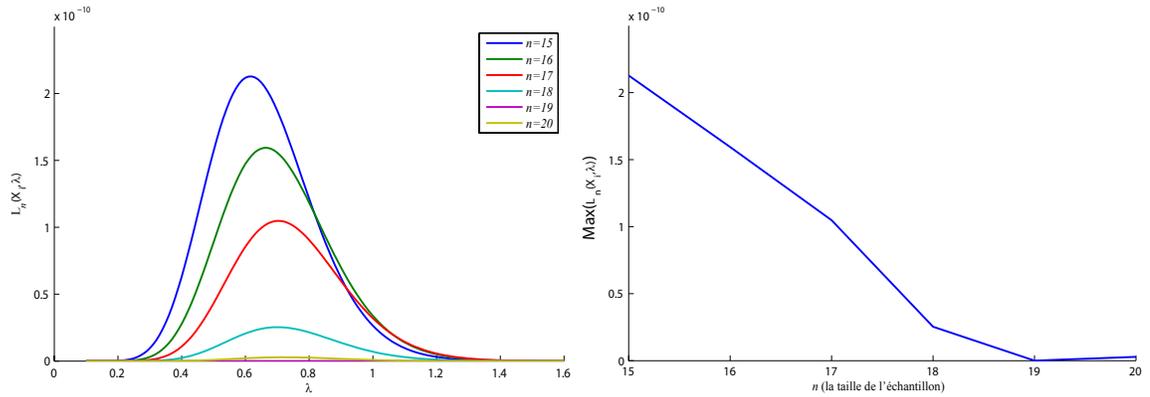


FIGURE 3.5 – Exemples de variation de la fonction de vraisemblance et son Max : cas de loi exponentielle

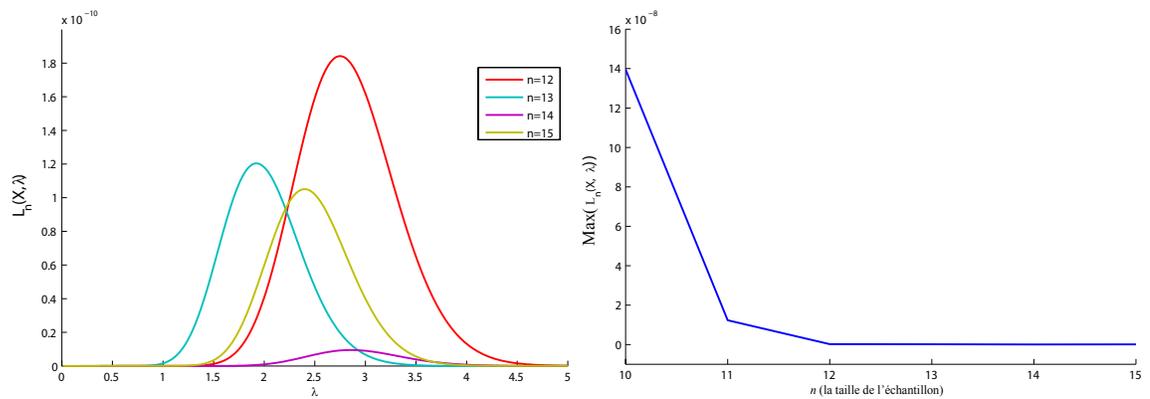


FIGURE 3.6 – Exemples de variation de la fonction de vraisemblance et son Max : cas de loi de Poisson

Conclusion

Dans ce travail, nous avons illustré via des exemples concrets le problème d'implantation de la méthode du maximum de vraisemblance sur l'outil informatique qui est engendré par la précision limitée des calculs sur l'outil informatique. De plus, les exemples traités montrent qu'en fait, à mesure que la variance de l'échantillon et/ou la taille de l'échantillon plus la chance que la méthode est applicable sur ordinateur se réduit.

Conclusion générale

La vraisemblance est une des idées de base de la statistique. Elle donne un cadre général et très puissant pour traiter toutes sortes d'application, en particulier pour :

1. trouver les estimateurs dont la variance est la plus petite possible dans les grands échantillons,
2. et construire des tests puissants.

Dans ce mémoire, nous nous sommes intéressés au cadre pratique de la méthode du maximum de vraisemblance. Plus précisément, aux contraintes de l'implémentation de cette technique sur ordinateur. Pour atteindre notre objectif, un rappel sur les notions de bases de la théorie d'estimation paramétrique a été exposé en premier lieu. Par la suite, nous nous sommes concentrée sur le principe de construction d'un estimateur à l'aide de la méthode du maximum de vraisemblance. En effet, pour mieux comprendre le mécanisme de la méthode en question et de vérifier la qualité des estimations qu'elle nous fournit, nous l'avons appliquée sur quelques modèles usuels (loi normale, loi exponentielle, loi binomiale et loi de Poisson). Les exemples traités nous laissent croire que la méthode est très efficace car elle nous fournit des estimations de très bonne qualité.

Cependant, l'analyse approfondie de la méthode montre d'une part que la construction de la statistique du paramètre recherché n'est pas toujours évidente. D'autre

part, le recours à l'outil informatique pour surpasser cette dernière difficulté n'est pas toujours possible. En effet, nous avons montrée à base de quelque exemple (loi normale, loi exponentielle et loi de Poisson) que si nous considérant la forme brute de la fonction vraisemblance et que la variance de l'échantillon dont dispose est considérable notamment pour de grandes tailles d'échantillon le recours à l'outil informatique est en vain. Ceci peut être justifié par le fait que, sous les conditions précédentes, le maximum de la fonction de vraisemblance se rapproche de zéro (indépendamment du paramètre recherché) et y a de forte chance qu'il sera inférieur à 10^{-324} que les logiciels considèrent comme un zéro absolu.

Enfin, il sera intéressant de compléter le présent travail en :

- Élargissant notre analyse à d'autres modèles et d'autres champs d'application de la vraisemblance (tests, régression,...) pour mieux comprendre le problème afin de dégager d'éventuelles solutions.
- Analysant le cas de log-vraisemblance, pour vérifier au quel point l'introduction de la fonction logarithme peut élargir le domaine d'application de la méthode de vraisemblance sur ordinateur.

Bibliographie

- [1] J. Aldrich. R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, vol. 12, no 3, 1997, p. 162–176.
- [2] K.E. Atkinson, *An Introduction to Numerical Analysis. Nonlinear Systems of Equations* Canada : John Wiley Sons, 1978, pp. 88–95.
- [3] Burden, R.L., Faires, J.D *Numerical Analysis. Numerical Solutions of Nonlinear Systems of Equations*. Belmont : Thomson Brooks/Cole. 2005, pp. 597–640.
- [4] A. Kagan. Another Look at the Cramér–Rao Inequality. *The American Statistician*, vol. 55, no 3, août 2001, p. 211–212
- [5] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(1962), 1065–1076.
- [6] M. Rosenblatt. Remarks in some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(1956), 832–837.
- [7] B. Silverman, (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.
- [8] S. Stigler. The Epic Story of Maximum Likelihood. *Statistical Science*, vol. 22, no 4, 2007.

Résumé

Dans ce travail, nous avons essayé d'illustrer à travers des exemples concrets le problème d'implantation de la méthode du maximum de vraisemblance (MV) en sa forme brute sur les logiciels, engendré principalement par la précision des calculs limitée de ces derniers. En effet, les exemples exposés montrent que la possibilité de l'implémentation de la méthode du MV sur ordinateur dépend de la variance et de la taille de l'échantillon dont on dispose. Où la possibilité de l'implémentation se réduit, en fur et à mesure que la variance de l'échantillon et/ou la taille de l'échantillon augmente.

Mots clés : Estimation, vraisemblance, logiciels, précision.

Abstract

In this work, we have illustrated through concrete examples the problem of implementing the maximum likelihood (ML) method in its raw form on software, generated mainly by the limited precision of the latter's calculations. Indeed, the examples presented show that the possibility of implementing the ML method on a computer depends on the variance and the size of the sample available. Where the possibility of implementation decreases, as the sample variance and/or sample size increases.

Key words : Estimation, likelihood, software, precision.

Résumé

Dans ce travail, nous avons essayé d'illustrer à travers des exemples concrets le problème d'implantation de la méthode du maximum de vraisemblance (MV) en sa forme brute sur les logiciels, engendré principalement par la précision des calculs limitée de ces derniers. En effet, les exemples exposés montrent que la possibilité de l'implémentation de la méthode du MV sur ordinateur dépend de la variance et de la taille de l'échantillon dont on dispose. Où la possibilité de l'implémentation se réduit, en fur et à mesure que la variance de l'échantillon et/ou la taille de l'échantillon augmente.

Mots clés : Estimation, vraisemblance, logiciels, précision.

Abstract

In this work, we have illustrated through concrete examples the problem of implementing the maximum likelihood (ML) method in its raw form on software, generated mainly by the limited precision of the latter's calculations. Indeed, the examples presented show that the possibility of implementing the ML method on a computer depends on the variance and the size of the sample available. Where the possibility of implementation decreases, as the sample variance and/or sample size increases.

Key words: Estimation, likelihood, software, precision.

ملخص

في هذا العمل حاولنا توضيح مشكلة تنفيذ طريقة المعقولة العظمى في شكلها الخام على البرامج التي تسببها بشكل رئيسي الدقة المحدودة لحسابات هذه الاخيرة. في الواقع توضح الامثلة المذكورة ان امكانية تنفيذ طريقة المعقولة العظمى على الكمبيوتر تعتمد على التباين وحجم العينة المتاحة حيث ينخفض اماكن تنفيذها تدريجيا مع زيادة تباين العينة و / او حجمها

الكلمات المفتاحية دقة-برمجيات-معقولة-تقدير