

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieure 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 :

Hafri Djouhaina

Titre : **Estimation par la méthode du maximum de vraisemblance**

Membres du Comité d'Examen :

Pr. Necir Abdelhakim UMKB Encadreur

Dr. Cherfaoui Mouloud UMKB Président

Dr. Chine Amel UMKB Examineur

September 2020

DÉDICACE

*A ma très chère **maman***

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 très cher **papa***

Tes sacrifices et tes qualités humaines m'ont permis de vivre ce jour.

A mes frères et sœurs

A mes amis

REMERCIEMENTS

Tout d'abord, je remercie "Allah Le Tout Puissant" de m'avoir aidé et donné la santé et la volonté pour arriver à ce stade.

Mes vifs remerciements sont adressés à mon encadreur Pr. Necir Abdelhakim pour ses précieux conseils, ses orientations pertinentes et sa patience tout au long de la réalisation de ce mémoire.

Je tiens à remercier Monsieur Cherfaoui Mouloud et Mademoiselle Chine Amel qui m'ont fait l'honneur de faire partie du jury de soutenance.

Je remercie tous les enseignants qui ont contribué à ma formation, ainsi que tous les employés du département de Mathématiques.

Je remercie tout particulièrement mes parents, pour leur encouragement et soutien sur tous les aspects, ainsi que toute ma famille. Je n'oublie pas l'ensemble de mes amis proches et aussi mes collègues d'études.

A ceux qui ont contribué, de près ou de loin, à la réalisation de ce modeste travail : un grand merci à vous tous.

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste des figures	vi
Introduction	1
1 Généralités	3
1.1 Lois de probabilité usuelles	3
1.1.1 Lois discrètes	3
1.1.2 lois continues	7
1.2 Echantillonnage	10
1.2.1 Moyenne et variance empiriques	10
1.3 Estimation paramétrique ponctuelle	11
1.3.1 Estimateur ponctuel	12
1.3.2 Propriétés des estimateurs	12
1.4 Convergence de variables aléatoires	14
2 Estimation par la méthode du maximum de vraisemblance	17
2.1 Historique la méthode du maximum de vraisemblance	17
2.2 Estimateurs du maximum de vraisemblance : première approche	18

2.2.1	Méthode du maximum de vraisemblance	18
2.2.2	Estimateurs du maximum de vraisemblance : cas X var discrète . . .	18
2.2.3	Méthode du maximum de vraisemblance :cas X var à densité	19
2.3	Estimateurs du maximum de vraisemblance pour (x_1, \dots, x_n)	20
2.3.1	Fonction de vraisemblance	20
2.3.2	Fonction de vraisemblance : contexte théorique	21
2.3.3	Estimateurs du maximum de vraisemblance pour (x_1, \dots, x_n)	21
2.3.4	Fonction de Log-vraisemblance	21
2.3.5	équation de vraisemblance	22
2.3.6	Expression analytique de l'EMV	22
2.3.7	Pratique de l'EMV	22
2.4	Estimateurs du maximum de vraisemblance (aléatoire)	23
2.4.1	Estimateur du maximum de vraisemblance (aléatoire)	23
2.4.2	Propriétés théoriques de l'emv	26
2.4.3	Lien entre l'emv et l'information de Fisher	28
2.5	Estimation des parametres de la loi loi généralisée des valeurs extrêmes (GEV) par la méthode du maximum de vraisemblance	29
2.5.1	Loi GEV avec $k = 0$ (EV1, loi Gumbel)	29
2.5.2	Loi GEV avec $k \neq 0$ (EV2 et EV3)	31
2.6	Estimation d'un modèle de regression linéaire par la méthode du maximum de vraisemblance	35
2.6.1	Estimateur du maximum de vraisemblance (ML)	36
2.7	Conclusion	45
3	Implémentation d'estimation par le maximum de vraisemblance en R	46
3.1	Introduction	46
3.2	Génération des données	47
3.3	Modélisation des données générées	48
3.4	Estimation par le maximum de vraisemblance	48

Bibliographie

52

Table des figures

1	Taux de mortalité du COVID-19 par âge en Angleterre [1].	1
3.1	Distribution de données générées par la loi de poisson.	47
3.2	Probabilité du premier point de données sur 20 valeurs de λ	49
3.3	Probabilité d'un seul point de données dans l'espace Log.	50
3.4	Estimation par maximum de vraisemblance de la valeur de λ	51

Introduction

Le hasard fait partie de notre quotidien. Chaque jour, nous faisons des jugements basés sur la probabilité, par exemple :

- En Angleterre, 11.6% des patients de 75 ans et plus, meurent s'ils ont contracté le Corona virus, contre 3.1% de 65 ans et 0.5% de 45. Ce qui explique la décision des différents gouvernements pour la distanciation sociale et la minimisation des visites aux personnes âgées [1]



FIGURE 1 – Taux de mortalité du COVID-19 par âge en Angleterre [1].

La théorie des probabilités a été développée à partir de l'étude mathématique des jeux de hasard par Fermat et Pascal [2]. Cette théorie traite les résultats possibles d'un événement, comme le lancé d'un dé.

Il est toujours possible d'associer à un phénomène aléatoire une probabilité et définir ainsi une loi de probabilité. Mais, lorsque le nombre d'épreuves augmente indéfiniment, il devient très difficile à traiter les différentes probabilités de l'ensemble de données. C'est pour cela, il est essentiel d'identifier la loi de probabilité suivie par une variable aléatoire donnée, car cela caractérise la liste des valeurs possibles avec leurs probabilités associées. Cependant, la tâche d'identification de la loi de probabilité et de ses paramètres reste une

tâche difficile et un domaine de recherche actif. Différentes méthodes ont été proposées dans la littérature pour estimer les différents paramètres d'une loi de probabilité. Parmi ces approches, on trouve la méthode de maximum de vraisemblance.

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.

Ce manuscrit est organisé en 3 chapitres :

- Le premier chapitre rappelle des généralités nécessaires sur la probabilité ainsi que les différents types de lois de probabilités.
- Le deuxième chapitre présente en détaille la méthode d'estimation par le maximum de vraisemblance dans le cas discret et continue, ainsi que les différentes approches d'estimation.
- Le dernier chapitre illustre l'approche d'estimation par le maximum de vraisemblance par un exemple implémenté en langage R .

Chapitre 1

Généralités

Définition 1.1 Une variable aléatoire (v.a) réelle X est une fonction définie sur un espace de probabilité (Ω, \mathcal{F}, P) à valeur dans \mathbb{R} est mesurable par rapport aux tribus \mathcal{F} et $\mathcal{B}_{\mathbb{R}}$ (tribu borélienne de \mathbb{R}).

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X(\omega) = x. \end{aligned}$$

1.1 Lois de probabilité usuelles

1.1.1 Lois discrètes

Définition 1.2 Une loi de probabilité concentrée sur un ensemble discret E est dite une loi discrète

Fonction de répartition d'une loi discrète

Si X est une variable aléatoire telle que $X(\Omega) = \{x_1, \dots, x_n\}$, sa fonction de répartition est égale à[4]

$$F_X(x) = P(X \leq x) = \sum_{\substack{1 \leq i \leq n \\ x_i \leq x}} P(X = x_i) \quad (1.1)$$

Espérance et variance dans le cas discret

Si X est une variable aléatoire discrète,

$$E(X) = \sum_{k=0}^{+\infty} P(X = k) \quad (1.2)$$

$$E(X^2) = \sum_{k=0}^{+\infty} k^2 P(X = k) \quad (1.3)$$

$$V(X) = E(X^2) - E(X)^2 \quad (1.4)$$

principales lois discrètes

Loi de Bernoulli de paramètre p

Définition 1.3 *La variable aléatoire X définie sur (Ω, \mathcal{F}, P) suit une loi de Bernoulli si*

$$X(\Omega) = \{0, 1\} \begin{cases} X = 0 \text{ échec} \\ X = 1 \text{ succès} \end{cases} \quad (1.5)$$

Loi de probabilité de X :

$$\begin{aligned} P(X = 1) &= p, \\ P(X = 0) &= (1 - p) = q \end{aligned} \quad (1.6)$$

on note

$$\begin{aligned} X &\rightarrow \mathcal{B}(p) \\ X &\rightarrow \mathcal{B}(1, p) \end{aligned} \quad (1.7)$$

et

$$\begin{aligned}
 E(X) &= p \\
 V(X) &= p(1-p) = pq \\
 \sigma(X) &= \sqrt{p(1-p)} = \sqrt{pq}
 \end{aligned}
 \tag{1.8}$$

Exemple 1.1 on lance une pièce de monnaie, soit X la variable aléatoire définie comme suit :

$$X = \begin{cases} 1 & \text{si pile est apparu} \\ 0 & \text{si face est apparu} \end{cases}
 \tag{1.9}$$

On a

$$P(X=1) = P(X=0) = \frac{1}{2} \quad \left(p = \frac{1}{2} \right)
 \tag{1.10}$$

$$E(X) = p = \frac{1}{2},
 \tag{1.11}$$

$$V(X) = p(1-p) = \frac{1}{4}, \quad \sigma_X = \frac{1}{2}
 \tag{1.12}$$

$$X \rightarrow \mathcal{B}\left(\frac{1}{2}\right)
 \tag{1.13}$$

Remarque 1.1 Si on lance la pièce de monnaie n fois, la probabilité d'avoir pile, la probabilité d'avoir face = $\frac{1}{2}$ dans chaque lancer, d'où on peut définir une autre loi dite loi Binomiale.

Loi Binomiale $\mathcal{B}(n; p)$

Définition 1.4 Une urne contient N boules, r boules blanches et $(N - r)$ non blanches. On pose : $p = \frac{r}{N}$ (probabilité de tirer une boule blanche de l'urne)

On tire une boule de l'urne, on note sa couleur et on la remet dans l'urne.

On note X la v.a qui représente le nombre de boules blanches tirées.

Déterminer la loi de probabilité de la v.a X :

On a :

$$\Omega = \left\{ (B, B, \bar{B}, \bar{B}, B, \dots)_{n\text{-uplet}}, \dots \right\}, \quad \text{card}(\Omega) = N^n$$

$$X : \Omega \rightarrow X(\Omega), \quad X(\Omega) = \{0, 1, \dots, n\}$$

(\cdot, \dots, \cdot)_{n-uplet} nombre de boules blanches

Ainsi la loi de probabilité de X est :

$$P(X = 0) = \frac{(N - r)^n}{N^n} = (1 - p)^n$$

$$P(X = 1) = \frac{r(N - r)^{n-1}}{NN^{n-1}} C_n^1 = C_n^1 p (1 - p)^{n-1}$$

$$P(X = 2) = \frac{r^2 (N - r)^{n-2}}{N^2 N^{n-2}} C_n^2 = C_n^2 p^2 (1 - p)^{n-2}$$

$$\vdots$$

$$P(X = k) = \frac{r^k (N - r)^{n-k}}{N^k N^{n-k}} C_n^k = C_n^k p^k (1 - p)^{n-k} \quad (1.14)$$

Donc

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k \in \{0, 1, \dots, n\} \quad (1.15)$$

et

$$E(X) = np \quad (1.16)$$

$$V(X) = np(1 - p) = npq \quad (1.17)$$

Loi de poisson

Définition 1.5 La variable aléatoire X est dite suivre la loi de Poisson de paramètre λ ($\lambda > 0$) si

$X(\Omega) = \{0, 1, 2, \dots, t\}$ et

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (1.18)$$

On note

$$X \rightarrow P(\lambda) \tag{1.19}$$

On commence par prouver que $P(X = k)$ est bien une loi de probabilité.

1. $\forall k \in \mathbb{R}^+; P(X = k) > 0$ c'est évident
2. $\sum_{k=0}^{+\infty} P(X = k) = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$

Donc la loi de Poisson est bien une loi de probabilité

et

$$E(X) = \lambda$$

$$V(X) = \lambda$$

Exemple 1.2 *une suspension bactérienne contient 5000 bactéries/litrer. On ensemence à partir de cette suspension, 50 boites de pétri, à raison d' 1 cm³ par boite*

Si X représente le nombre de colonies par boite, alors la loi de probabilité de X est :

$$X \rightarrow P(\lambda = 5) \tag{1.20}$$

Rappel : 1litre = 1000cm³

Donc ici le nombre moyen de bactéries par boite est 5. On suppose aussi que le nombre de colonies par boite est le même que le nombre moyenne de bactéries par boites

$$P(X = 0) = \frac{5^0 e^{-5}}{0!} = 0,0067 \text{ soit approximativement } 0,67\% \text{ de chance}$$

La probabilité qu'il n'y ait au moins une colonie sur la boite de pétri est : $P(X > 0) = 1 - P(X = 0) = 1 - 0,0067 = 0,9933$ soit 99,3% de chance d'avoir au moins une colonie bactérienne qui se développe dans la boite de pétri (voir événement)

1.1.2 lois continues

Définition 1.6 *on dit qu'une variable aléatoire X suit la loi de densité f lorsque*

- f est une fonction de \mathbb{R} dans \mathbb{R}^+
- $\int_{-\infty}^{+\infty} f(t) dt = 1$

Fonction de répartition d'une loi continue

Si X est une variable aléatoire de densité f , sa fonction de répartition est égale à :

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad (1.21)$$

on a alors

$$P(X > x) = 1 - F_X(x) \quad (1.22)$$

est sa densité vaut

$$f(x) = \frac{dF_X(x)}{dx}. \quad (1.23)$$

Espérance et variance dans le cas continu

Si X est une variable aléatoire continue de densité f ,

$$E(X) = \int_{-\infty}^{+\infty} xf(x) dx$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2f(x) dx$$

$$V(X) = E(X^2) - E(X)^2$$

Loi normale(gaussienne)

Définition 1.7 Une v.a X suit une loi normale ou gaussienne (loi de Laplace-Gauss) de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ si elle admet pour densité de probabilité la fonction f définie, pour tout nombre réel x , par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right\}. \quad (1.24)$$

Cette distribution, notée $\mathcal{N}(\mu, \sigma^2)$, représente un bon modèle pour beaucoup de phénomènes naturels. De plus le (théorème central limite), la rend la plus importante en statistique et donc la plus populaire. On a

$$E(X) = \mu \text{ et } \text{Var}(X) = \sigma^2 \quad (1.25)$$

Lorsque l'espérance est nulle, X est dite v.a normale centrée. Si de plus la variance vaut 1, on parle alors de ce que l'on appelle loi normale standard ou centrée réduite. Dans ce cas, la fonction de densité de probabilité est paire et a la forme simple suivante :

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, z \in \mathbb{R} \quad (1.26)$$

Ainsi, sa représentation graphique, appelée cloche de Gauss, est symétrique par rapport à l'axe des ordonnées.

Le passage du cas générale au cas particulier se fait par un simple changement de variable. En effet on a

$$X \sim \mathcal{N}(\mu, \sigma^2) \iff Z := \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1). \quad (1.27)$$

loi uniforme

Définition 1.8 On dit une v.a continue X suit une loi **uniforme** sur un intervalle $[a, b]$, où a et b sont deux réels tels que $a < b$, si elle admet pour densité de probabilité

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) \quad (1.28)$$

On écrit $X \sim \mathcal{U}([a, b])$ et on a

$$E(X) = \frac{a+b}{2} \text{ et } V(X) = \frac{(b-a)^2}{12} \quad (1.29)$$

Le cas particulier $a = 0$ et $b = 1$ donne naissance la loi uniforme standard, pour laquelle

on a $E(X) = 1/2$ et $V(X) = 1/12$.

1.2 Echantillonnage

Nous allons étudier comment se comporte un échantillon (éléments pris au hasard) dans une population X dont on connaît les caractéristiques statistiques (lois,...) considérée X . Dans ce cas, prendre un échantillon aléatoire de taille n consiste à considérer n réalisations de X ou encore considérer n variables aléatoires X_1, \dots, X_n indépendantes, de même loi que X [3].

Définition 1.9 Soit X une variable aléatoire sur un espace probabilité (Ω, F, p) . Un échantillon de X de taille n est un n -uplet (X_1, \dots, X_n) de variables aléatoires indépendantes 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, \dots, x_n) où $X_i(\omega) = x_i$.

1.2.1 Moyenne et variance empiriques

Définition 1.10 On appelle **statistique** sur un n -échantillon (X_1, \dots, X_n) toute fonction réel X_i fonction mesurable des $X_i, i = 1, \dots, n$.

Moyenne empirique

Définition 1.11 On appelle **moyenne empirique** (échantillonnage, expérimentale) la statistique notée \bar{X}_n définie par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.30)$$

Soit X une v.a de moyenne μ et de variance σ^2 . Alors, on a

$$E(\bar{X}_n) = \mu \text{ et } \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \quad (1.31)$$

Variance empirique

Définition 1.12 On appelle *variance empirique*, la statistique

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (1.32)$$

Proposition 1.1 Si X est une v.a de variance σ^2 , alors[5]

$$E(S_n^2) = \frac{n-1}{n} \sigma^2. \quad (1.33)$$

Fréquence

Soit $(X_i)_{i=1\dots n}$ un échantillon aléatoire de taille n ayant une loi de Bernoulli de paramètre p comme **loi mère**. Alors

$$F = \frac{X_1 + \dots + X_n}{n} \quad (1.34)$$

est la fréquence de la valeur 1 dans l'échantillon et nF suit une loi binomiale de paramètres n et p . Ainsi

$$E(F) = p \quad \text{et} \quad \text{Var}(F) = \frac{pq}{n}. \quad (1.35)$$

Donc, quand n tend vers l'infini, F converge en loi vers $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

1.3 Estimation paramétrique ponctuelle

Cette fois il s'agit d'estimer certaines caractéristiques statistiques de la loi (moyenne, variance...) au travers d'une série d'observations x_1, x_2, \dots, x_n . C'est la problématique inverse de l'échantillonnage.

à partir des caractéristiques d'un échantillon, que peut-on déduire des caractéristiques de la population dont il est issu ?

L'estimation consiste à donner des valeurs approximatives aux paramètres d'une population à l'aide d'un échantillon de n observations issues de cette population. On peut se tromper sur la valeur exacte, mais on donne la "meilleure valeur" possible que l'on peut

supposer.

1.3.1 Estimateur ponctuel

Définition 1.13 On souhaite estimer un paramètre θ d'une population (cela peut être sa moyenne μ , son écart-type σ , une proportion p). Un estimateur de θ est une statistique T (donc une fonction de (X_1, \dots, X_n)) dont la réalisation est envisagée comme une "bonne valeur" du paramètre θ . On parle d'estimation de θ associée à cet estimateur la valeur observée lors de l'expérience, c'est-à-dire la valeur prise par la fonction au point observé (x_1, \dots, x_n) .

Définition 1.14 Soit X une variable aléatoire dont la densité de probabilité $f(x, \theta)$ dépend d'un paramètre θ appartenant à $I \subset \mathbb{R}$. A l'aide d'un échantillon issu de X , il s'agit de déterminer au mieux la vraie valeur θ_0 de θ . On pourra utiliser deux méthodes :

- **estimation ponctuelle** : on calcule une valeur vraisemblable $\hat{\theta}$ de θ_0
- **estimation par intervalle** : on cherche un intervalle dans lequel θ_0 se trouve avec une probabilité élevée[3].

1.3.2 Propriétés des estimateurs

Estimateur sans biais

Définition 1.15 Nous appelons **biais** d'un estimateur T_n au point la fonction

$$\theta \longrightarrow b_{T_n}(\theta) = E(T_n) - \theta \quad \forall \theta \in \Theta \quad (1.36)$$

On dit que l'estimateur est

- un estimateur **sans biais** si

$$\forall \theta \in \Theta \quad b_{T_n}(\theta) = 0 \implies E(T_n) = \theta \quad (1.37)$$

— un estimateur **asymptotiquement sans biais** si

$$\lim_{n \rightarrow \infty} b_{T_n}(\theta) = 0 \quad \forall \theta \in \Theta \quad (1.38)$$

Estimateur convergent

Définition 1.16 L'estimateur est dit **convergent (ou consistant)** si la suite T_n converge en probabilité vers θ ($T_n \xrightarrow{P} \theta$)

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|T_n - \theta| > \varepsilon) = 0 \quad (1.39)$$

On dit que l'estimateur est fortement convergent lorsqu'on a la convergence **presque sure (p.s)**

Proposition 1.2 Pour valider la convergence d'un estimateur, il suffit de montrer que :

$$\forall \theta \in \Theta, \lim_{n \rightarrow \infty} E(T_n) = \theta \text{ et } \lim_{n \rightarrow \infty} \text{Var}(T_n) = 0 \quad (1.40)$$

Pour vérifier ce critère, il faut connaître la loi de X et celle de T_n pour calculer l'espérance et la variance.

Estimateur asymptotiquement normale

Définition 1.17 Soit T_n un estimateur du paramètre de la loi P_θ d'une variable aléatoire observée X . Nous supposons qu'il existe deux fonctions $a = a(\theta, n)$ et $b = b(\theta, n)$ telle que :

$$\begin{aligned} \left(\frac{T_n - a}{b} \right) &\rightharpoonup Z \\ n &\longrightarrow \infty \\ Z &\rightharpoonup \mathcal{N}(0, 1) \end{aligned}$$

Nous disons alors que T_n est un **estimateur asymptotiquement normale**

Remarque 1.2 *La loi asymptotique sera d'autant plus utile que les fonctions a et b ont une expression simple.*

1.4 Convergence de variables aléatoires

Vu l'utilité de la notion de convergence dans l'estimation statistique, on rappelle, dans cette section, les définitions et résultats essentiels relatifs aux différents modes de convergence. Soit $(X_n)_{n \geq 1}$ une suite de v.a. [3]

Définition 1.18 (Convergence en probabilité) *On dit que $(X_n)_{n \geq 1}$ converge en probabilité*

vers une v.a. X , et on écrit $X_n \xrightarrow{P} X$, si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0. \quad (1.41)$$

Définition 1.19 (Convergence presque sûre) *On dit que $(X_n)_{n \geq 1}$ converge presque sûrement vers une v.a. X , et on écrit $X_n \xrightarrow{ps} X$, si*

$$P(\lim_{n \rightarrow \infty} X_n \neq X) = 0. \quad (1.42)$$

Définition 1.20 (Convergence en loi) *On dit que $(X_n)_{n \geq 1}$ converge en loi vers une v.a. X , et on écrit $X_n \xrightarrow{\mathcal{L}} X$, si*

$$\lim_{n \rightarrow \infty} F_n(X) = F(X). \quad (1.43)$$

Définition 1.21 (Convergence en moyenne quadratique) *On dit que $(X_n)_{n \geq 1}$ converge en moyenne quadratique vers une v.a. X , et on écrit $X_n \xrightarrow{mq} X$, si*

$$\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0. \quad (1.44)$$

1. Les relations 1.41, 1.42, 1.43, 1.44 restent valables si on remplace la v.a. X par une constante réelle l .

2. La convergence en moyenne quadratique se généralise à la convergence en moyenne d'ordre $k \geq 2$.
3. Les implications suivantes permettent le passage entre certains types de convergence.

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{\mathcal{L}} X$$

$$X_n \xrightarrow{ps} X \implies X_n \xrightarrow{P} X$$

$$X_n \xrightarrow{mq} X \implies X_n \xrightarrow{P} X$$

Avant d'énoncer les deux théorèmes limites fondamentaux en statistique (1.1 et 1.2), on commence par introduire, dans la 1.3, deux inégalités remarquables

Proposition 1.3

- **Inégalité de Markov** : Pour tout réel $a > 0$ et toute fonction mesurable positive sur \mathbb{R} g on a

$$P(g(X) \geq a) \leq \frac{1}{a} E(g(X)) \tag{1.45}$$

- **Inégalité de Chebychev** : Soit Y une v.a de variance finie σ^2 : Alors

$$\forall \varepsilon > 0, P(|Y - E(Y)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{1.46}$$

Théorème 1.1 (Lois des grands nombres) Soient X_1, \dots, X_n des v.a iid d'espérance μ et de variance σ^2 (finie). Alors, on a :

- **loi faible** :

$$\bar{X}_n \xrightarrow{P} \mu. \tag{1.47}$$

- **loi forte** :

$$\bar{X}_n \xrightarrow{ps} \mu. \tag{1.48}$$

Théorème 1.2 (théorème centrale limite) Soit $(X_n)_{n \geq 1}$ une suite de v.a iid, d'espérance et de

variance σ^2 finies. Alors

$$\frac{U_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \quad (1.49)$$

où $U_n := \sum_{i=1}^n X_i$ désigne la somme des n premières v.a. [3][9]

Chapitre 2

Estimation par la méthode du maximum de vraisemblance

2.1 Historique la méthode du maximum de vraisemblance

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.[6]

2.2 Estimateurs du maximum de vraisemblance : première approche

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

- 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 : $\theta = (\theta_1, \theta_2, \theta_3 \dots)^t$,
- l'aspect numérique accessible grâce à l'application de méthodes d'optimisation connues.

Elle permet de :

- construire des estimateurs performants,
- mettre en œuvre des tests statistiques "puissants".

2.2.2 Estimateurs du maximum de vraisemblance : cas X var 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 notre objectif est d'estimer θ à partir de (x_1, \dots, x_n) , comme 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 :

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

$$L_n(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n)). \quad (2.1)$$

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

$$L_n(\theta) = \mathbb{P}_\theta\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n \mathbb{P}_\theta(X = x_i). \quad (2.2)$$

2.2.3 Méthode du maximum de vraisemblance :cas X var 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 : $\{(X_1, \dots, X_n) \in [x_1, x_1 + \epsilon_1[\times \dots \times [x_n, x_n + \epsilon_n[\}$ avec $\epsilon_1, \dots, \epsilon_n$ assez petits.

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

$$L_n^{(\epsilon_1, \dots, \epsilon_n)}(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \in [x_1, x_1 + \epsilon_1[\times \dots \times [x_n, x_n + \epsilon_n[). \quad (2.3)$$

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

$$L_n^{(\epsilon_1, \dots, \epsilon_n)}(\theta) = \int_{x_1}^{x_1 + \epsilon_1} \dots \int_{x_n}^{x_n + \epsilon_n} f_\theta(t_1, \dots, t_n) dt_1 \dots dt_n, \quad (2.4)$$

Cependant, déterminer un θ qui maximise $L_n^{(\epsilon_1, \dots, \epsilon_n)}(\theta)$ n'est pas chose aisée :

- l'expression analytique d'une fonction intégrale comme $L_n^{(\epsilon_1, \dots, \epsilon_n)}(\theta)$ n'existe pas toujours,
- il dépend de $\epsilon_1, \dots, \epsilon_n$ dont la petitesse reste subjective.

Une solution est de considérer une version idéalisée de $L_n^{(\epsilon_1, \dots, \epsilon_n)}(\theta)$ qui est

$$L_n(\theta) = f_\theta(x_1, \dots, x_n). \quad (2.5)$$

Cette expression est une conséquence du résultat suivant : en notant F_θ la fonction de répartition de (X_1, \dots, X_n) , il vient

$$\lim_{(\epsilon_1, \dots, \epsilon_n) \rightarrow (0, \dots, 0)} \frac{L^{(\epsilon_1, \dots, \epsilon_n)}(\theta)}{\epsilon_1 \times \dots \times \epsilon_n} = \frac{\partial^n}{\partial \epsilon_1 \dots \partial \epsilon_n} F_\theta(x_1, \dots, x_n) = f_\theta(x_1, \dots, x_n). \quad (2.6)$$

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 :

$$L_n(\theta) = f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i). \quad (2.7)$$

2.3 Estimateurs du maximum de vraisemblance pour (x_1, \dots, x_n)

2.3.1 Fonction de vraisemblance

Dans ce document, on appelle fonction de vraisemblance pour (x_1, \dots, x_n) la fonction de θ :

$$L_n(x_1, \dots, x_n; \theta) = f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta) \quad (2.8)$$

On rappelle que

$$f(x; \theta) = \begin{cases} \mathbb{P}_\theta(X = x), & \text{si } X \text{ est une var discrète,} \\ f_\theta(x_i), & \text{si } X \text{ est une var à densité de densité } f_\theta. \end{cases} \quad (2.9)$$

La fonction de vraisemblance n'est intéressante que si θ et x_i vérifient $f(x_i; \theta) \neq 0$ pour tout $i \in \{1, \dots, n\}$, sinon on peut d'ores et déjà remettre en cause l'hypothèse que X suit la loi $\mathcal{L}(\theta)$.

2.3.2 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) .

2.3.3 Estimateurs du maximum de vraisemblance pour (x_1, \dots, x_n)

On appelle estimateur du maximum de vraisemblance de θ (emv) pour (x_1, \dots, x_n) un réel θ^* qui maximise la fonction de vraisemblance $L_n(x_1, \dots, x_n; \theta)$ en θ , i.e. pour tout θ ,

$$L_n(x_1, \dots, x_n; \theta) \leq L_n(x_1, \dots, x_n; \theta^*) \quad (2.10)$$

Une expression alternative est :

$$\theta^* \in \arg \max_{\theta} L_n(x_1, \dots, x_n; \theta) \quad (2.11)$$

où $\arg \max$ désigne l'argument du maximum qui est l'ensemble des points en lesquels une expression atteint sa valeur maximale.

Puisqu'il dépend de x_1, \dots, x_n , θ^* est une estimation ponctuelle de θ . Un tel estimateur n'existe pas toujours et peut ne pas être unique.

Dans ce document, pour simplifier, on suppose qu'il existe un unique emv θ^* de θ pour (x_1, \dots, x_n) .

2.3.4 Fonction de Log-vraisemblance

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

$$l_n(x_1, \dots, x_n; \theta) = \ln(L_n(x_1, \dots, x_n; \theta)). \quad (2.12)$$

Elle n'a de sens que si θ vérifie $L_n(x_1, \dots, x_n; \theta) > 0$. La fonction logarithme népérien étant croissante, l'emv θ^* de θ pour (x_1, \dots, x_n) vérifie :

$$\theta^* \in \arg \max_{\theta} L_n(x_1, \dots, x_n; \theta) = \arg \max_{\theta} l_n(x_1, \dots, x_n; \theta). \quad (2.13)$$

2.3.5 équation de vraisemblance

On appelle équation de vraisemblance l'équation en θ :

$$\frac{\partial}{\partial \theta} l_n(x_1, \dots, x_n; \theta) = 0. \quad (2.14)$$

2.3.6 Expression analytique de l'EMV

Pour envisager d'avoir une expression analytique de l'emv θ^* de θ pour (x_1, \dots, x_n) , une idée est d'exprimer $L_n(x_1, \dots, x_n; \theta)$ en fonction de produits de termes exponentiels/puissances, puis de considérer la fonction de log-vraisemblance $l_n(x_1, \dots, x_n; \theta)$. Si cette dernière est dérivable en θ , une condition nécessaire que doit vérifier θ^* est d'être solution de l'équation de vraisemblance. Il faut ensuite vérifier que θ^* est bien un maximum pour $l_n(x_1, \dots, x_n; \theta)$. :

- soit en étudiant les variations de $l_n(x_1, \dots, x_n; \theta)$,
- soit en montrant que

$$\frac{\partial^2}{\partial \theta^2} l_n(x_1, \dots, x_n; \theta^*) < 0. \quad (2.15)$$

2.3.7 Pratique de l'EMV

Bien souvent, la résolution de l'équation de vraisemblance n'amène pas une expression analytique pour l'emv θ^* de θ pour (x_1, \dots, x_n) . En tant que problème d'optimisation, on

peut quand même approcher la valeur de θ^* avec précision à l'aide d'algorithmes itératifs efficaces.

Il y a notamment :

- l'algorithme de Newton-Raphson,
- l'algorithme de Gauss-Newton,
- le score de Fisher.

Exemple 2.1 *partant de l'équation de vraisemblance et d'un développement de Taylor au premier ordre de $\frac{\partial}{\partial \theta} l_n(x_1, \dots, x_n; \theta)$, l'algorithme de Newton-Raphson est caractérisé par la m -ème itération :*

$$\theta^{(m+1)} = \theta^{(m)} - \left(\frac{\partial^2}{\partial \theta^2} l_n(x_1, \dots, x_n; \theta^{(m)}) \right)^{-1} \left(\frac{\partial}{\partial \theta} l_n(x_1, \dots, x_n; \theta^{(m)}) \right). \quad (2.16)$$

Dès qu'il y a stabilisation pour un m_* , i.e. $\theta^{(m_*+1)} = \theta^{(m_*)}$, on considère la valeur $\theta^* = \theta^{(m_*)}$.

2.4 Estimateurs du maximum de vraisemblance (aléatoire)

2.4.1 Estimateur du maximum de vraisemblance (aléatoire)

On appelle estimateur du maximum de vraisemblance (emv) un estimateur (aléatoire) $\hat{\theta}_n$ de θ défini par

$$\hat{\theta}_n \in \arg \max_{\theta} L_n(X_1, \dots, X_n; \theta). \quad (2.17)$$

On a substitué x_i par X_i pour tout $i \in \{1, \dots, n\}$ dans la définition de θ^* .

Dans ce document, pour simplifier, on suppose qu'il existe un unique emv $\hat{\theta}_n$ de θ . Si on peut exprimer θ^* sous la forme $\theta^* = g_n(x_1, \dots, x_n)$ où g_n désigne une fonction connue, l'emv (aléatoire) de θ est

$$\hat{\theta}_n = g_n(X_1, \dots, X_n) \quad (2.18)$$

C'est la performance de $\hat{\theta}_n$ qui permet d'évaluer la précision de θ^* dans l'estimation de θ .

Exemple 2.2 *Avec une loi discrète*

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

$$f(x; \lambda) = P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

La fonction de vraisemblance s'écrit

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-\lambda n} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}. \quad (2.19)$$

Il est plus simple d'utiliser le logarithme, la vraisemblance étant positive :

$$\ln L(x_1, \dots, x_n; \lambda) = \ln e^{-\lambda n} + \ln \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} = -\lambda n + \sum_{i=1}^n \ln \frac{\lambda^{x_i}}{x_i!} = -\lambda n + \ln \lambda \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!). \quad (2.20)$$

La dérivée première

$$\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^n x_i}{\lambda} \quad (2.21)$$

s'annule pour $\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$. La dérivée seconde

$$\frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial \theta^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2} \quad (2.22)$$

est toujours négative ou nulle. Ainsi l'estimation donnée par $\Lambda = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ conduit à un estimateur du maximum de vraisemblance égal à $\hat{\lambda} = \bar{X}$. Il est normal de retrouver la moyenne empirique qui est le meilleur estimateur possible pour le paramètre λ (qui représente aussi l'espérance d'une loi de Poisson).

Exemple 2.3 Avec une loi continue

On souhaite estimer les paramètres μ et σ d'une loi normale à partir d'un n-échantillon.

La loi normale $N(\mu, \sigma)$ a pour fonction densité

$$f(x; \mu, \sigma) = f_{(\mu, \sigma)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (2.23)$$

Ecrivons la fonction de vraisemblance pour une réalisation d'un échantillon de n variables indépendantes :

$$f(x_1, \dots, x_n; \mu, \sigma) = \prod_{i=1}^n f(x_i; \mu, \sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right). \quad (2.24)$$

Or (théorème de König) $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$, où \bar{x} représente la moyenne de l'échantillon. Ainsi la fonction de vraisemblance peut être écrite sous la forme

$$f(x_1, \dots, x_n; \mu, \sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right).$$

$$\frac{\partial}{\partial \mu} \ln L = \frac{\partial}{\partial \mu} \left(\ln\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} - \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right) \right) = 0 - \frac{2n(\bar{x} - \mu)}{2\sigma^2}$$

On obtient donc l'estimateur par le maximum de vraisemblance de l'espérance :

$$\hat{\mu} = \bar{X} = \sum_{i=1}^n \frac{X_i}{n}. \quad (2.25)$$

Pour le second paramètre, on calcule

$$\frac{\partial}{\partial \sigma} \ln L = \frac{\partial}{\partial \sigma} \left(\frac{n}{2} \ln\left(\frac{1}{2\pi\sigma^2}\right) - \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right) = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{\sigma^3} \quad (2.26)$$

Donc

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i - \hat{\mu})^2}{n}. \quad (2.27)$$

que l'on peut traduire par

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (2.28)$$

On vérifie que c'est bien des maxima locaux :

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\sigma^2} \leq 0 \quad (2.29)$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = -\frac{n}{\sigma^2} - \frac{3}{\sigma^4} (\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2). \quad (2.30)$$

Au point $\hat{\sigma}$,

$$\frac{\partial^2 \ln L}{\partial \sigma^2}(\hat{\sigma}^2) = -\frac{n}{\hat{\sigma}^2} - \frac{3}{\sigma^4} (n\hat{\sigma}^2 + n(\bar{x} - \mu)^2). \quad (2.31)$$

La méthode fournit un estimateur non biaisé de la moyenne $\mathbb{E}(\hat{\mu}) = \mu$ mais par contre, l'estimateur de la variance est biaisé $\mathbb{E}(\hat{\sigma}^2) = \frac{n}{n-1}\sigma^2$. Néanmoins l'estimateur est asymptotiquement sans biais.

2.4.2 Propriétés théoriques de l'emv

Scari, l'emv θ_n est "au pire"

- asymptotiquement sans biais,
- fortement consistant,
- asymptotiquement efficace,
- asymptotiquement normal.

C'est donc un estimateur performant.

Théorème sur l'efficacité

Scari, s'il existe un réel $a_n(\theta)$ dépendant de θ tels que l'emv θ_n de θ vérifie la factorisation :

$$\frac{\partial}{\partial \theta} l_n(X_1, \dots, X_n; \theta) = a_n(\theta)(\hat{\theta}_n - \theta), \quad (2.32)$$

alors

— $\hat{\theta}_n$ est un estimateur sans biais : $\mathbb{E}(\hat{\theta}_n) = \theta$,

— on a

$$\mathbb{V}(\hat{\theta}_n) = \frac{1}{a_n(\theta)}, \quad (2.33)$$

— on a

$$I_n(\theta) = \frac{1}{\mathbb{V}(\hat{\theta}_n)}. \quad (2.34)$$

En particulier, le premier et le troisième points entraînent que $\hat{\theta}_n$ est un estimateur efficace de θ .

Emv et efficacité

S'il existe un estimateur efficace de θ , alors c'est l'emv $\hat{\theta}_n$ de θ (en revanche, un emv n'est pas nécessairement efficace).

Sur la normalité asymptotique de l'emv

Scrici, l'emvb $\hat{\theta}_n$ de θ est asymptotiquement normal ; la suite de var :

$$(\sqrt{I_n(\theta)}(\hat{\theta}_n - \theta))_{n \in \mathbb{N}^*} \quad (2.35)$$

converge en loi vers une var Z suivant la loi normale $N(0, 1)$.

Résultat asymptotique central

Scrici, l'emv $\hat{\theta}_n$ de θ vérifie : la suite de var :

$$(\sqrt{I_n(\hat{\theta}_n)}(\hat{\theta}_n - \theta))_{n \in \mathbb{N}^*} \quad (2.36)$$

converge en loi vers une var Z suivant la loi normale $N(0, 1)$; on a substitué θ à $\hat{\theta}_n$ dans la définition de l'information de Fisher du résultat de normalité asymptotique de $\hat{\theta}_n$.

Ce dernier résultat est central pour la construction d'intervalles de confiance pour θ , de tests statistiques simples ou complexes ...

2.4.3 Lien entre l'emv et l'information de Fisher

Le lien existant entre l'emv et l'information de Fisher peut s'expliquer ainsi :

- Plus les valeurs de $(\frac{\partial}{\partial \theta} l_n(X_1, \dots, X_n; \theta))^2$ sont petites, moins l'emv θ^* de θ pour (x_1, \dots, x_n) qui annule cette fonction ne se distingue de ses valeurs voisines. Dans ce contexte, (x_1, \dots, x_n) ne nous apportent pas assez d'information pour estimer précisément θ .
- A contrario, plus les valeurs de $(\frac{\partial}{\partial \theta} l_n(X_1, \dots, X_n; \theta))^2$ sont grandes, plus l'emv θ^* de θ pour (x_1, \dots, x_n) qui annule cette fonction se singularise. Dans ce contexte, (x_1, \dots, x_n) nous apportent suffisamment d'information pour estimer précisément θ .

Ainsi, une mesure de la quantité moyenne d'information apportée par (X_1, \dots, X_n) sur θ est

$\mathbb{E}((\frac{\partial}{\partial \theta} l_n(X_1, \dots, X_n; \theta))^2)$ Après calculs, celle-ci correspond à l'information de Fisher $I_n(\theta)$.

Cela est en adéquation avec :

- les performances théoriques de l'emv $\hat{\theta}_n$ principalement garanties dans un cadre asymptotique :

quand n est assez grand, $I_n(\theta) = nI_1(\theta)$ prend une large étendue de valeurs éloignées de 0,

- la "borne minimale du risque quadratique de $\hat{\theta}_n$ " :

$$R(\hat{\theta}_n, \theta) \geq \frac{1}{I_n(\theta)} (1 + \frac{\partial}{\partial \theta} B(\hat{\theta}_n, \theta))^2 + (B(\hat{\theta}_n, \theta))^2, \quad (2.37)$$

qui met en lumière la difficulté d'estimer θ par $\hat{\theta}_n$ suivant la grandeur de $I_n(\theta)$ dans un cadre non-asymptotique.[6]

2.5 Estimation des parametres de la loi loi généralisée des valeurs extrêmes (GEV) par la méthode du maximum de vraisemblance

2.5.1 Loi GEV avec $k = 0$ (EV1, loi Gumbel)

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

$$f(x) = \frac{1}{\alpha} \exp \left[-\frac{(x-u)}{\alpha} - \exp \left(-\frac{(x-u)}{\alpha} \right) \right], \text{ si } k = 0 \quad (2.38)$$

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 :

$$\ln L(\alpha, u) = -n \ln(\alpha) - \sum_{i=1}^n \frac{x_i - u}{\alpha} - \sum_{i=1}^n \exp \left[-\frac{x_i - u}{\alpha} \right] \quad (2.39)$$

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

$$n - \sum_{i=1}^n \frac{x_i - u}{\alpha} + \sum_{i=1}^n \frac{x_i - u}{\alpha} \exp \left[-\frac{x_i - u}{\alpha} \right] = 0 \quad (2.40)$$

$$-n + \sum_{i=1}^n \exp \left[-\frac{x_i - u}{\alpha} \right] = 0 \quad (2.41)$$

On peut toutefois simplifier ce système en écrivant l'équation (2.41) sous la forme :

$$u = -\alpha \ln \left\{ \frac{1}{\alpha} \sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha} \right] \right\} \quad (2.42)$$

et en remplaçant u dans (2.40) par (2.42) on obtient :

$$\alpha = \bar{x} - \left\{ \sum_{i=1}^n x_i \exp \left[-\frac{x_i}{\alpha} \right] \right\} \left\{ \sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha} \right] \right\}^{-1} \quad (2.43)$$

Les équations (2.42) et (2.43) sont équivalentes au système (2.40) et (2.41). On peut ainsi,

pour un échantillon donné x_1, x_2, \dots, x_n , déterminer $\hat{\alpha}$ par (2.43) et ensuite connaissant $\hat{\alpha}$ en déduire \hat{u} par (2.42).

Pour résoudre (2.6), on doit employer une méthode itérative comme celle de Newton-Raphson.

Toutefois, 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. [7]

En particulier, les estimateurs \hat{u} et $\hat{\alpha}$ obtenus dans le cas de la loi EV1 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

$$\alpha^* = \frac{n}{n - 0.8} \hat{\alpha} \quad (2.44)$$

et

$$u^* = \alpha^* \ln \left\{ n \left(\sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha^*} \right] \right)^{-1} \right\} - \frac{0.7}{n} \alpha^* \quad (2.45)$$

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 :

$$Var(\hat{\alpha}) = 0.608 \frac{\alpha^2}{n} \quad (2.46)$$

$$\text{Var}(\hat{u}) = 1.109 \frac{\alpha^2}{n} \quad (2.47)$$

$$\text{Cov}(\hat{u}, \hat{\alpha}) = 0.257 \frac{\alpha^2}{n} \quad (2.48)$$

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

$$\text{Var}(\alpha^*) = \left[\frac{n}{n-0.8} \right]^2 \text{Var}(\hat{\alpha}) \quad (2.49)$$

$$\text{Var}(u^*) = \left(\frac{\alpha^2}{n} \right) \left[1.109 - \frac{0.360}{n-0.8} + \frac{0.298}{n^2} \right] \quad (2.50)$$

et

$$\text{Cov}(u^*, \alpha^*) = \left(\frac{\alpha^2}{n} \right) \left[\frac{0.257n}{n-0.8} - \frac{0.426}{n} \right] \quad (2.51)$$

Remarque 2.1 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 $\hat{\alpha}$.

2.5.2 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

$$f(x) = \frac{1}{\alpha} \left[1 - \frac{k}{\alpha} (x - u) \right]^{1/k-1} \exp \left[- \left(1 - \frac{k}{\alpha} (x - u) \right)^{1/k} \right], \text{ si } k \neq 0 \quad (2.52)$$

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

$$y(x) = -\frac{1}{k} \ln \left[1 - \frac{k}{\alpha} (x - u) \right] \quad (2.53)$$

ce qui est équivalent à :

$$1 - \frac{k}{\alpha} (x - u) = e^{-ky(x)} \quad (2.54)$$

la relation (2.52) devient :

$$f(x) = \frac{1}{\alpha} \left(e^{-ky(x)} \right)^{1/k-1} \exp \left[- \left(e^{-ky(x)} \right)^{1/k} \right] \quad (2.55)$$

ce qui s'écrit encore :

$$f(x) = \frac{1}{\alpha} e^{(k-1)y(x)} \exp \left(-e^{-y(x)} \right) \quad (2.56)$$

Le logarithme de la fonction de vraisemblance est alors :

$$\ln L(x; u, \alpha, k) = -n \ln(\alpha) - (1 - k) \sum_{i=1}^k y(x_i) - \sum_{i=1}^n e^{-y(x_i)} \quad (2.57)$$

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 :

$$- \frac{\partial \ln L(x; u, \alpha, k)}{\partial u} = \frac{Q}{\alpha} \quad (2.58)$$

$$- \frac{\partial \ln L(x; u, \alpha, k)}{\partial \alpha} = \frac{1}{\alpha} \frac{P + Q}{k} \quad (2.59)$$

$$- \frac{\partial \ln L(x; u, \alpha, k)}{\partial k} = \frac{1}{k} \left(R - \frac{P + Q}{k} \right) \quad (2.60)$$

où

$$P = n - \sum_{i=1}^n e^{-y(x_i)} \quad (2.61)$$

$$Q = \sum_{i=1}^n e^{-y(x_i)+y(x_i)} - (1-k) \sum_{i=1}^n e^{-ky(x_i)} \quad (2.62)$$

$$R = n - \sum_{i=1}^n y(x_i) + \sum_{i=1}^n y(x_i) e^{-y(x_i)} \quad (2.63)$$

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 jième itération.

Les éléments du vecteur

$$-\Delta \ln L(\theta^{(j)}) = \left(-\frac{\partial \ln L(x; u, \alpha, k)}{\partial u}, -\frac{\partial \ln L(x; u, \alpha, k)}{\partial \alpha}, -\frac{\partial \ln L(x; u, \alpha, k)}{\partial k} \right)^T \quad (2.64)$$

à la jième itération peuvent être développés en séries de Taylor autour de $\hat{\alpha}$ (voir NERC, 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 :

$$\theta^{(j+1)} = \theta^{(j)} + V(\hat{\theta})^{-1} \Delta \ln L(\theta^{(j)}) \quad (2.65)$$

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

$$V(\hat{\theta}) = \begin{pmatrix} -\frac{\partial^2 \ln L}{\partial u^2} & -\frac{\partial^2 \ln L}{\partial u \partial \alpha} & -\frac{\partial^2 \ln L}{\partial u \partial k} \\ -\frac{\partial^2 \ln L}{\partial \alpha \partial u} & -\frac{\partial^2 \ln L}{\partial \alpha^2} & -\frac{\partial^2 \ln L}{\partial \alpha \partial k} \\ -\frac{\partial^2 \ln L}{\partial k \partial u} & -\frac{\partial^2 \ln L}{\partial k \partial \alpha} & -\frac{\partial^2 \ln L}{\partial k^2} \end{pmatrix} \quad (2.66)$$

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, Prescott et Walden (1983) mentionnent que la convergence vers $\hat{\theta}$ est plus rapide en utilisant $V(\theta^{(j)})$ plutôt que $I_f(\theta^{(j)})$. Les éléments de V sont calculées à partir des équations (2.21) à (2.26) 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.

Les variances et covariances asymptotiques des estimateurs du maximum de vraisemblance 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 :

$$I_f^{-1} = \begin{pmatrix} \text{Var}(\hat{u}) & \text{Cov}(\hat{u}, \hat{\alpha}) & \text{Cov}(\hat{u}, \hat{k}) \\ \text{Cov}(\hat{u}, \hat{\alpha}) & \text{Var}(\hat{\alpha}) & \text{Cov}(\hat{\alpha}, \hat{k}) \\ \text{Cov}(\hat{u}, \hat{k}) & \text{Cov}(\hat{\alpha}, \hat{k}) & \text{Var}(\hat{k}) \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \alpha^2 b & \alpha^2 h & \alpha f \\ \alpha^2 h & \alpha^2 a & \alpha g \\ \alpha f & \alpha g & c \end{pmatrix} \quad (2.67)$$

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 :

$$E \left[-\frac{\partial^2 \ln L}{\partial u^2} \right] = \frac{n}{\alpha^2} p \quad (2.68)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial \alpha^2} \right] = \frac{n}{\alpha^2 k^2} [1 - 2\Gamma(2 - k) + p] \quad (2.69)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial k^2} \right] = \frac{n}{k^2} \left[\frac{\pi^2}{6} - \left(1 - C - \frac{1}{k} \right)^2 + \frac{2q}{p} + \frac{p}{k^2} \right] \quad (2.70)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial u \partial \alpha} \right] = \frac{n}{\alpha^2 k} [p - \Gamma(2 - k)] \quad (2.71)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial u \partial k} \right] = -\frac{n}{\alpha k} \left(q + \frac{p}{k} \right) \quad (2.72)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial \alpha \partial k} \right] = \frac{n}{\alpha^2 k^2} \left\{ 1 - C - \frac{[1 - \Gamma(2 - k)]}{k} - q - \frac{p}{k} \right\} \quad (2.73)$$

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

$$p = (1 - k)^2 \Gamma(1 - 2k) \quad (2.74)$$

$$q = \Gamma(2 - k) \left[\Psi(1 - k) - \frac{(1 - k)}{k} \right] \quad (2.75)$$

2.6 Estimation d'un modèle de regression linéaire par la méthode du maximum de vraisemblance

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 re-

prendre 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.[8]

2.6.1 Estimateur du maximum de vraisemblance (ML)

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

Considérons le MRLS suivant :

$$Y_t = a_0 + a_1 X_t + u_t \quad (2.76)$$

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 « $u_t \sim N(0; \sigma^2) \rightarrow E(u_t) = 0$ », 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'ou, 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 :

$$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2) = f(Y_1 | a_0 + a_1 X_t, \sigma_u^2) * f(Y_2 | a_0 + a_1 X_t, \sigma_u^2) * \dots * f(Y_t | a_0 + a_1 X_t, \sigma_u^2) \quad (2.77)$$

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) :

$$f(Y_t) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp \left[-\frac{1}{2} \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right] \quad (2.78)$$

Fonction de vraisemblance et fonction log-vraisemblance

Fonction de vraisemblance Elle est obtenue lorsqu'on remplace la fonction de densité de la loi normale (2.78) dans la fonction de densité de probabilité jointe (2.77), en supposant connues les « Y_1, Y_2, \dots, Y_t » :

$$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2) = \frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \exp \left[-\frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right] \quad (2.79)$$

ou encore :

$$L(a_0, a_1, \sigma_u^2) = \frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \exp \left[-\frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right] \quad (2.80)$$

Fonction log-vraisemblance Lorsqu'on effectue une transformation logarithmique de la fonction de vraisemblance cidessus, 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 :

$$\ln L = \ln \left(\frac{1}{(\sqrt{2\pi})^t \sigma_u^t} \right) + \ln \left(\exp \left[-\frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right] \right) \quad (2.81)$$

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

$$\ln\left(\frac{1}{(\sqrt{2\pi})^t \sigma_u^t}\right) = \ln(1) - [\ln(\sqrt{2\pi})^t + \ln \sigma_u^t] = -\frac{t}{2} \ln(2\pi) - t \ln \sigma_u \quad (2.82)$$

$$\ln\left(\exp\left[-\frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right]\right) = -\frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2 \quad (2.83)$$

$$\ln L = -\frac{t}{2} \ln(2\pi) - t \ln \sigma_u - \frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2 \quad (2.84)$$

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

$$\ln L = -\frac{t}{2} \ln(2\pi) - \frac{t}{2} \ln \sigma_u^2 - \frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2 \quad (2.85)$$

Estimation des paramètres « \hat{a}_0, \hat{a}_1 et σ_u^2 » Pour estimer les paramètres « $\hat{a}_0, \hat{a}_1, \sigma_u^2$ » par le maximum de vraisemblance, la démarche va consister à maximiser la fonction log-vraisemblance ci-dessus (2.85), ce qui revient à annuler ses dérivées premières par rapport aux arguments « \hat{a}_0, \hat{a}_1 et σ_u^2 » comme suit :

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

$$\frac{\partial \ln L}{\partial a_0} = -\frac{1}{2} * 2 * \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^{2-1} (-1) = -\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right) (-1) = 0 \dots (a) \quad (2.86)$$

$$\frac{\partial \ln L}{\partial a_1} = -\frac{1}{2} * 2 * \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^{2-1} (-X_t) = -\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right) (-X_t) = 0 \dots (b) \quad (2.87)$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \frac{\partial}{\partial \sigma_u^2} \left(-\frac{t}{2} \ln \sigma_u^2\right) - \frac{1}{2} * \frac{\partial}{\partial \sigma_u^2} \left(\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right) = 0 \dots (c) \quad (2.88)$$

ona

$$\frac{\partial}{\partial \sigma_u^2} \left(-\frac{t}{2} \ln \sigma_u^2 \right) = -\frac{t}{2} * \frac{1}{\sigma_u^2} = -\frac{t}{2\sigma_u^2} \dots (d) \quad (2.89)$$

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

$$\frac{\partial}{\partial \sigma_u^2} [(\sum (\bullet))^2] = \frac{\partial}{\partial \sigma_u} \left(\frac{\partial}{\partial \sigma_u} [(\sum (\bullet))^2] \right) \quad (2.90)$$

$$\frac{\partial}{\partial \sigma_u} [(\sum (\bullet))^2] = \sum [2 * [(Y_t - a_0 - a_1 X_t) * \sigma_u^{-1}]^{2-1} * (-1) (Y_t - a_0 - a_1 X_t) * \sigma_u^{-1-1}] \quad (2.91)$$

$$\frac{\partial}{\partial \sigma_u} [(\sum (\bullet))^2] = \sum -2 * \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right) * (Y_t - a_0 - a_1 X_t) * \sigma_u^{-2} \quad (2.92)$$

$$\frac{\partial}{\partial \sigma_u} \left(\frac{\partial}{\partial \sigma_u} [(\sum (\bullet))^2] \right) = \sum -2 * (-2) * \frac{(Y_t - a_0 - a_1 X_t)^2}{\sigma_u * \sigma_u^3} = 4 \sum \left[\frac{Y_t - a_0 - a_1 X_t}{\sigma_u^2} \right]^2 \dots (e) \quad (2.93)$$

Remplaçons les expressions (e) et (d) dans (c), l'on a :

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{t}{2\sigma_u^2} + \frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u^2} \right)^2 \dots (c)^* \quad (2.94)$$

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

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{t}{2\sigma_u^2} + \frac{1}{2} * \frac{1}{\sigma_u^4} \sum (Y_t - a_0 - a_1 X_t)^2 = -t + \frac{2\sigma_u^2}{2\sigma_u^4} \sum (Y_t - a_0 - a_1 X_t)^2 = 0 \quad (2.95)$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -t + \frac{1}{\sigma_u^2} \sum (Y_t - a_0 - a_1 X_t)^2 = 0 \dots (c)^{**} \quad (2.96)$$

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

$$\frac{\partial \ln L}{\partial a_0} = 0 \rightarrow \sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV} X_t) = 0 \rightarrow \sum Y_t = T \hat{a}_{0MV} + \hat{a}_{1MV} \sum X_t \quad (2.97)$$

$$\frac{\partial \ln L}{\partial a_1} = 0 \rightarrow \sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV} X_t) X_t = 0 \rightarrow \sum X_t Y_t = \hat{a}_{0MV} \sum X_t + \hat{a}_{1MV} \sum X_t^2 \quad (2.98)$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = 0 = -t + \frac{1}{\sigma_{uMV}^2} \sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV} X_t)^2 = 0 \rightarrow \sigma_{uMV}^2 = \frac{1}{T} \sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV} X_t)^2. \quad (2.99)$$

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 : $\hat{a}_{0MV} = \hat{a}_{0MCO}$ et $\hat{a}_{1MV} = \hat{a}_{1MCO}$. Par contre, le développement de la dernière expression donne un estimateur du maximum de vraisemblance « σ_{uMV}^2 » de la variance de l'erreur « σ_u^2 » différent de l'estimateur MCO :

$$\hat{\sigma}_{uMV}^2 = \frac{1}{T} \sum (Y_t - \hat{a}_0 - \hat{a}_1 X_t)^2 = \frac{1}{T} \sum e_t^2 \neq \hat{\sigma}_{uMCO}^2 = \frac{1}{(T-2)} \sum e_t^2. \quad (2.100)$$

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

$$Y_t = a_0 + a_1 X_t + a_2 X_{2t} + \dots + a_i X_{it} + \dots + a_k X_{kt} + u_t \quad (2.101)$$

ou encore (sans constante).

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + \dots + a_i X_{it} + \dots + a_k X_{kt} + u_t \quad (2.102)$$

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 (2.101) peut encore s'écrire comme suit :

$$\underbrace{Y}_{(T,1)} = \underbrace{X}_{(T,k+1)} \underbrace{a}_{(k+1,1)} + \underbrace{u}_{(T,1)} \rightarrow Y = Xa + u \quad (2.103)$$

Avec :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \\ \vdots \\ Y_T \end{pmatrix}; X = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1t} & X_{2t} & \cdots & X_{kt} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1T} & X_{2T} & \cdots & X_{kT} \end{pmatrix}; a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}; \text{et } u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \\ \vdots \\ u_T \end{pmatrix}. \quad (2.104)$$

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 (2.102) est (les formats de X et a changent) :

$$\underbrace{Y}_{(T,1)} = \underbrace{X}_{(T,k)} \underbrace{a}_{(k,1)} + \underbrace{u}_{(T,1)} \rightarrow Y = Xa + u \quad (2.105)$$

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_u^2 »), est une variable normalement distribuée de paramètre (moyenne et va-

riance) :

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

$$V(Y_t) = \sigma_u^2$$

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

$$f(Y_t|a, \sigma_u^2) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left[-\frac{1}{2\sigma_u^2} (Y_t - X_t a)^2\right] \quad (2.106)$$

Fonction de densité de probabilité jointe ou conjointe de Y_1, Y_2, \dots, Y_t Connaissant les paramètres « a et σ_u^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 :

$$f(Y_1, Y_2, \dots, Y_t|a, \sigma_u^2) = f(Y_1|a, \sigma_u^2) * f(Y_2|a, \sigma_u^2) * \dots * f(Y_t|a, \sigma_u^2) = \prod_{t=1}^T f(Y_t|a, \sigma_u^2) \quad (2.107)$$

$$\begin{aligned} \rightarrow \prod_{t=1}^T f(Y_t|a, \sigma_u^2) &= \left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T \exp\left[-\frac{1}{2\sigma_u^2} \sum_{t=1}^T (Y_t - X_t a)^2\right] \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T \exp\left[-\frac{1}{2\sigma_u^2} (Y - Xa)'(Y - Xa)\right] \end{aligned}$$

Fonction de vraisemblance et fonction log-vraisemblance

Fonction de vraisemblance La fonction de vraisemblance, notée « L », correspond à la relation «(2.107)» en supposant connus les « Y_1, Y_2, \dots, Y_t » et inconnus les paramètres « a et σ_u^2 » (à trouver), ce qui revient à écrire :

$$L(a, \sigma_u^2) = f(a, \sigma_u^2|Y_1, Y_2, \dots, Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T \exp\left[-\frac{1}{2\sigma_u^2} (Y - Xa)'(Y - Xa)\right] \quad (2.108)$$

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 :

$$\ln L = \ln \left[\left(\frac{1}{\sqrt{2\pi\sigma_u^2}} \right)^T \right] \ln \left(\exp \left[-\frac{1}{2\sigma_u^2} (Y - Xa)'(Y - Xa) \right] \right) \quad (2.109)$$

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

$$\begin{aligned} \rightarrow \ln \left[\left(\frac{1}{\sqrt{2\pi\sigma_u^2}} \right)^T \right] &= \ln \left[(2\pi\sigma_u^2)^{-T/2} \right] = -\frac{T}{2} [\ln \pi + \ln (2\sigma_u^2)] = -\frac{T}{2} \ln \pi - \frac{T}{2} \ln (2\sigma_u^2) \\ \rightarrow \ln \left(\exp \left[-\frac{1}{2\sigma_u^2} (Y - Xa)'(Y - Xa) \right] \right) &= -\frac{1}{2\sigma_u^2} (Y - Xa)'(Y - Xa) \\ \rightarrow (Y - Xa)'(Y - Xa) &= Y'Y - Y'Xa - X'a'Y + X'Xa'a = Y'Y - 2X'a'Y + X'Xa'a^2 \end{aligned}$$

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

$$\begin{aligned} \ln L(a, \sigma_u^2) &= -\frac{T}{2} \ln \pi - \frac{T}{2} \ln (2\sigma_u^2) - \frac{1}{2\sigma_u^2} (Y - Xa)'(Y - Xa) \\ &= -\frac{T}{2} \ln \pi - \frac{T}{2} \ln (2\sigma_u^2) - \frac{1}{2\sigma_u^2} (Y'Y - 2X'a'Y + X'Xa'a^2) \end{aligned}$$

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

Estimation des paramètres « a et σ_u^2 » Pour estimer les paramètres inconnus « a et σ_u^2 » par la méthode du maximum de vraisemblance, la démarche va consister à maximiser la fonction log-vraisemblance cidessus (2.109), ce qui revient à annuler ses dérivées premières par rapport aux arguments « a et σ_u^2 »

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

$$\frac{\partial \ln L}{\partial a} = -\frac{1}{2\sigma_u^2} (-2XY + 2XXa) = 0 \dots \dots (i)$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \frac{\partial}{\partial \sigma_u^2} \left(-\frac{T}{2} \ln(2\sigma_u^2) \right) + \frac{\partial}{\partial \sigma_u^2} \left[-\frac{1}{2\sigma_u^2} (Y^2 - 2XY + XXa^2) \right] = 0 \dots \dots (ii)$$

$$\rightarrow \frac{\partial}{\partial \sigma_u^2} \left(-\frac{T}{2} \ln(2\sigma_u^2) \right) = -\frac{T}{2\sigma_u^2}$$

$$\rightarrow \frac{\partial}{\partial \sigma_u^2} \left[-\frac{1}{2} (\sigma_u^2)^{-1} (Y - Xa)'(Y - Xa) \right] = -\frac{1}{2} * (-1) * (\sigma_u^2)^{-1-1} (Y - Xa)'(Y - Xa)$$

$$= \frac{1}{2\sigma_u^4} (Y - Xa)'(Y - Xa) = \frac{éé}{2\sigma_u^4}$$

Considérant ces expressions, (ii) s'écrit :

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{T}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} (Y - Xa)'(Y - Xa) = -\frac{T}{2\sigma_u^2} + \frac{éé}{2\sigma_u^4} = \frac{-T\sigma_u^2 + éé}{2\sigma_u^4} = 0 \dots \dots (ii)$$

(2.110)

Les expressions (i) et (ii) 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 « $\hat{\sigma}_{uMV}^2$ », soit respectivement :

$$\hat{a}_{MV} = (XX)^{-1} XY \dots \dots$$

$$\hat{\sigma}_{uMV}^2 = \frac{éé}{T} \dots \dots$$

Au regard des résultats obtenus, soient les expressions , 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 « $\hat{\sigma}_{uMV}^2$ » de la variance de l'erreur

« σ_u^2 », étant biaisé, est différent de l'estimateur MCO, car (avec $k = K - 1$ ou $K = k + 1$) :

$$E\left(\widehat{\sigma}_{uMV}^2\right) = \frac{1}{T}E\left(\sum_{t=1}^T e_t^2\right) \neq E\left(\widehat{\sigma}_{MCO}^2\right) = \frac{1}{(T-k-1)}E\left(\sum_{t=1}^T e_t^2\right) = \frac{1}{(T-K)} * (T-K) \sigma_u^2 = \sigma_u^2 \quad (2.111)$$

En effet :

$$E\left(\sum_{t=1}^T e_t^2\right) = (T-K) \sigma_u^2 \rightarrow E\left(\widehat{\sigma}_{uMV}^2\right) = \frac{1}{T}(T-K) \sigma_u^2 = \frac{(T-K)}{T} \sigma_u^2 = \sigma_u^2 - \frac{K}{T} \sigma_u^2 \quad (2.112)$$

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_u^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 croit indéfiniment), $\widehat{\sigma}_{uMV}^2$ est aussi non biaisé :

$$\lim E\left(\widehat{\sigma}_{uMV}^2\right) = \sigma_u^2 \text{ ou } \uparrow T \rightarrow \widehat{\sigma}_{uMV}^2 \simeq \widehat{\sigma}_{MCO}^2 \quad (2.113)$$

2.7 Conclusion

Dans ce chapitre , j'ai introduit les différentes types de probabilité ainsi que leurs utilisation. Dans le chapitre suivant , Je vais présenter l'estimation par la méthode de vraisemblance et sa propre utilisation pour résoudre différents problèmes et modèles .

Chapitre 3

Implémentation d'estimation par le maximum de vraisemblance en R

3.1 Introduction

La distribution de Poisson est très utilisée pour modéliser le nombre de fois qu'un événement se produit dans une période de temps. Par exemple, on peut modéliser le nombre de cas de **covid19** par jour en **Algérie** comme une distribution de Poisson.

La distribution de Poisson est une distribution simple avec un seul paramètre et dans ce chapitre nous allons l'utiliser pour illustrer les principes de l'estimation du maximum de vraisemblance.

Les différentes étapes de notre expérimentation sont présentées ci-dessous :

- Génération des données à partir d'une distribution de poisson
- Modélisation de données générées à partir de la distribution de poisson avec un paramètre inconnu,
- Estimation de ce paramètre en utilisant l'estimation du maximum de vraisemblance,
- Comparaison de résultat d'estimation avec la vérité terrain.

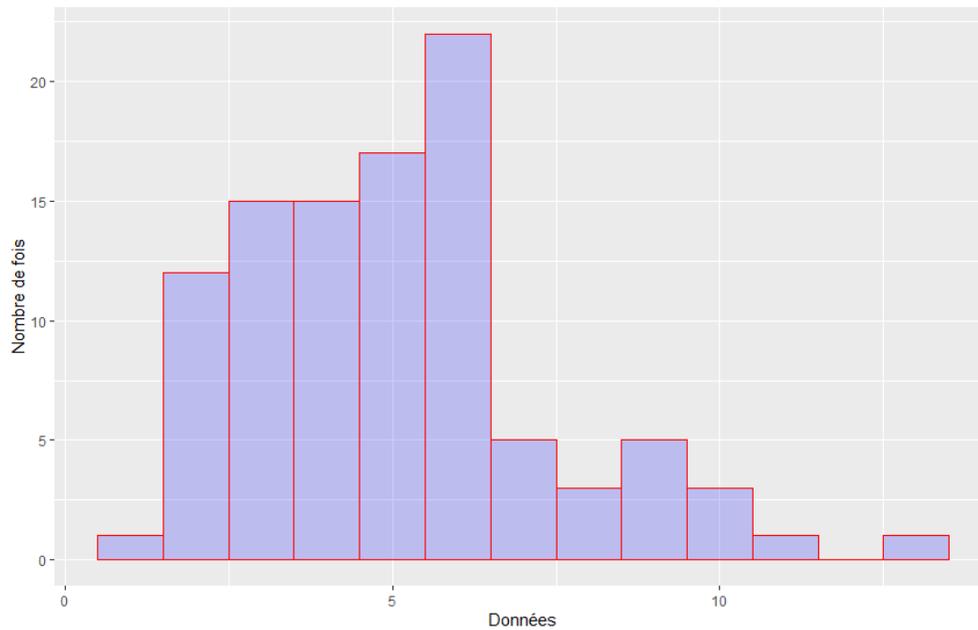


FIGURE 3.1 – Distribution de données générées par la loi de poisson.

3.2 Génération des données

La première étape consiste à générer des données à partir de la distribution de poisson. Dans le langage de R, nous pouvons générer facilement des nombres aléatoires à partir d'une distribution de probabilité spécifique. Cela est réalisé en utilisant la fonction `rpois`. Le code suivant démontre la génération de 100 points de données à partir de la distribution de Poisson avec le paramètre $\lambda = 5$ en utilisant la fonction `rpois`.

```
# Generation des donnees a partir de la distribution de poisson avec un
parametre lambda
data <- rpois(n=100, lambda=5)
# Sauver ces donnees dans une data frame pour les plotter
df_Poisson <- data.frame(data=data)
# Visualization de la distrubition des donnees
qplot(data, geom="histogram", binwidth = 1, bins = 20,
main = "Distrubition de poisson", xlab = "Donnees",
ylab = "Nombre de fois", fill=I("blue"),
col=I("red"), alpha=I(.2))
```

3.3 Modélisation des données générées

Lorsque on a un ensemble des données $x : x_1, x_2, \dots, x_n$ d'une distribution de probabilité avec un paramètre λ , On peut modéliser (écrire) la fonction de densité de probabilité de x comme $f(x; \lambda)$.

Comme était illustré dans le chapitre précédent, la densité de probabilité $f(x; \lambda)$ est appelé une fonction de vraisemblance. En partant d'un ensemble de données connu, La méthode d'estimation de maximum de vraisemblance cherche la valeur de paramètre λ pour laquelle la probabilité est plus élevée.

Si chaque point de donnée observée de l'ensemble est indépendante l'un de l'autre, nous pouvons simplifier la fonction de densité de probabilité conjointe comme :

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = f(x_1; \lambda) * f(x_2; \lambda) \dots f(x_n; \lambda) \quad (3.1)$$

Dans le cas de notre modèle (un seul paramètre λ), donc nous pouvons calculer la probabilité pour un plage de valeurs du paramètre λ et choisir manuellement la valeur qui a la probabilité la plus élevée.

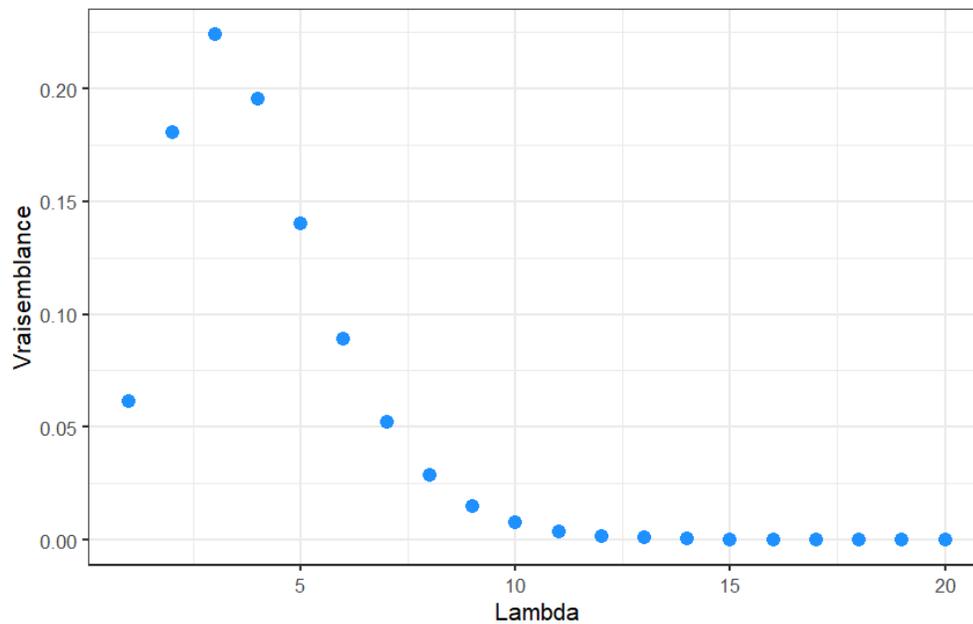
3.4 Estimation par le maximum de vraisemblance

Dans notre exemple, nous avons définie $f(x, \lambda)$ comme une fonction de densité de poisson. Donc, on peut utiliser la méthode `dpois()` de langage **R** pour calculer la vraisemblance de premier point de données en utilisant le paramètre `lambda`

```
dpois(data[1], lambda=1)
```

```
[1] 0.06131324
```

N'oublions pas que le paramètre `lambda` est inconnu et qu'il s'agit d'un paramètre de la fonction de vraisemblance. Le langage **R** étant un langage de programmation fonctionnel, nous pouvons facilement calculer les probabilités pour un tas de paramètres en même temps, comme ceci :

FIGURE 3.2 – Probabilité du premier point de données sur 20 valeurs de λ .

```
vraisemblance <- dpois(data[1], lambda=seq(20))
```

Voici le graphe généré de la probabilité d'un seul point de données sur 20 valeurs de λ . En analysant le graphe ci-dessus pour un seul point de données, nous pouvons voir qu'à partir $\lambda = 10$, beaucoup de probabilités sont assez proches de zéro. Cela signifie que, lors du calcul de la probabilité de tous l'ensemble de données, nous pourrions souvent finir par multiplier la probabilité très faible (proche de zéro). Cela entraînera des erreurs de calcul à la fin.

Pour remédier ce problème, nous utilisons l'espace *log*. Dans ce cas, nous allons remplacer les multiplications de l'éq 3.1 par addition. Dans le langage **R**, c'est facile à faire ça :

```
log_vraisemblance <- dpois(data[1], lambda=seq(20), log=TRUE)
```

Finalement, pour calculer la log-vraisemblance de tous l'ensemble de données, il suffit juste de sommer les log-vraisemblance de tous les points de données de cet ensemble.

Le code suivant présente l'utilisation de la méthode de maximum de vraisemblance pour trouver le paramètre λ . Rappelons qu'au début nous avons générer ces données avec une loi de poisson avec le paramètre $\lambda = 5$. Donc le but est de trouver un λ égale ou

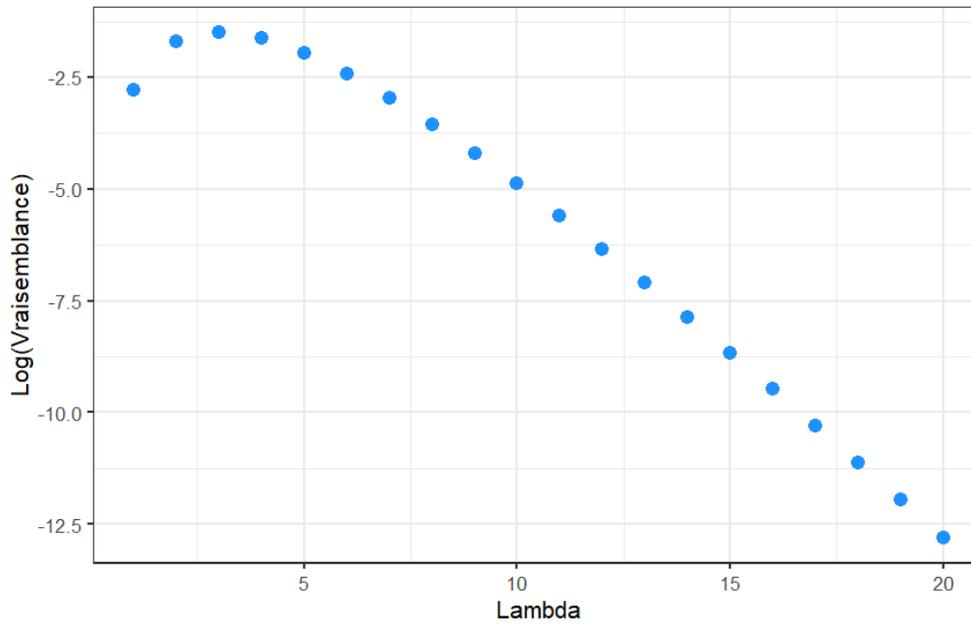


FIGURE 3.3 – Probabilité d’un seul point de données dans l’espace Log.

très proche à la valeur 5.

```
vraisemblance_poisson <- function(lambda, y){
# Sommer les vraisemblances de chaque element
vraisemblance <- sum(dpois(y, lambda, log=TRUE))
return(vraisemblance)
}

# Définir un plage de lambda de 1 a 15 avec un pas de 0.5
lambdas <- seq(1,15, by=0.5)

# Calculer le log-vraisemblance de tous les valeurs de lambda
MV <- sapply(lambdas, function(x){vraisemblance_poisson(x, data)})
df <- data.frame(MV=MV, lambda=lambdas)
ggplot(aes(x=lambda, y=MV))+
geom_point(size=4, color="dodgerblue")+
xlab("Lambda") + xlim(0,15) +
ylab("Log-Vraisemblance")+
theme_bw(base_size = 15) +
geom_vline(xintercept = lambdas[which.max(l1)], color="red", size=2)
```

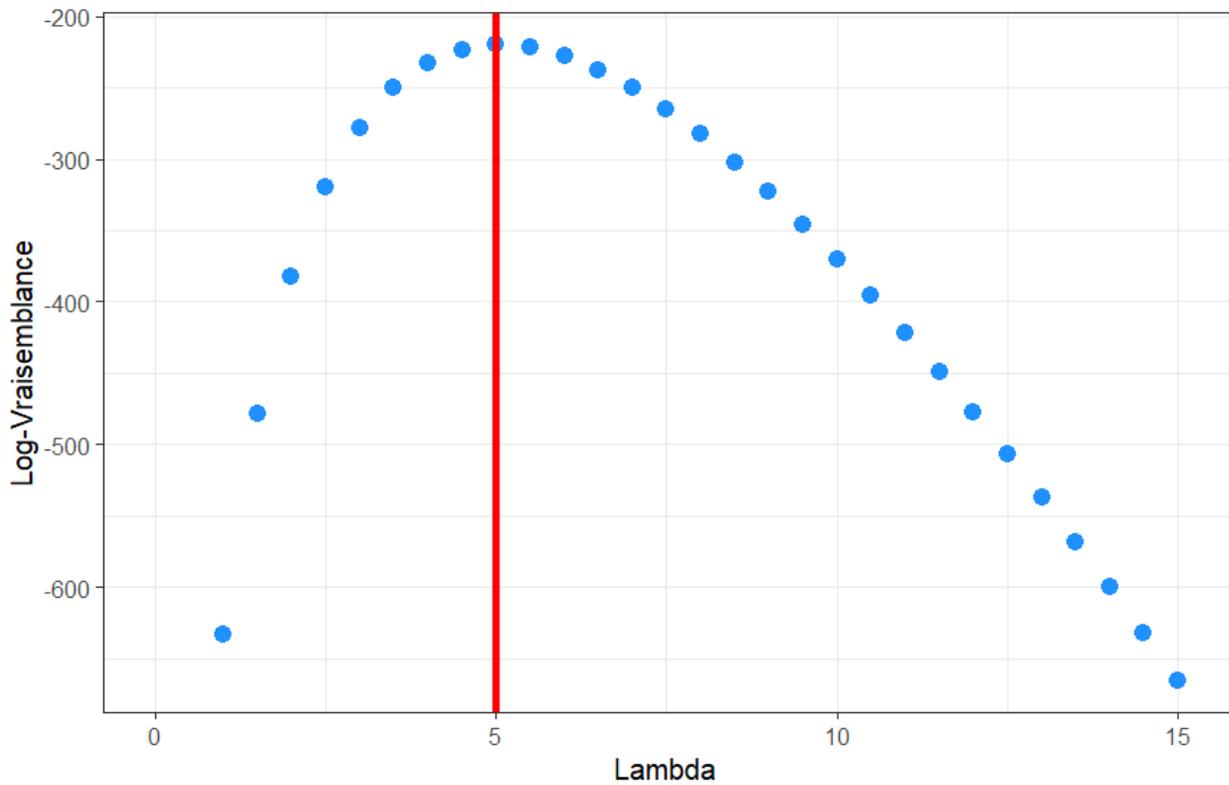


FIGURE 3.4 – Estimation par maximum de vraisemblance de la valeur de λ .

En utilisant le modèle de Poisson, la méthode de maximum de vraisemblance a fonctionné et nous avons estimé le paramètre $\lambda = 5$ qui est la valeur de la vérité terrain.

Bibliographie

- [1] M. S, "The coronavirus is most deadly if you are older and male — new data reveal the risks," *Nature*, pp. 585(7823) :16–17, 2020.
- [2] B. Pascal, "Les lettres de blaise pascal : accompagnées de lettres de ses correspondants," *G. CRÈS & C*, pp. 585(7823) :16–17, 1623-1662.
- [3] P. Dusart, "Cours de statistiques inférentielles," *Université de Limoges.*, 2015.
- [4] M. Chevalier, "Statistiques 4,lois de probabilité usuelles (rappels)," *Université de lyon*, 2015.
- [5] G. Saporta, "Probabilités, analyse des données et statistique." *Technip*, 2006.
- [6] C. Chesneau, "Sur l'estimateur du maximum de vraisemblance (emv)," *Université de Caen*, 2017.
- [7] L. Perreault and B. Bobée, "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.
- [8] J. K. Kuma, "Estimation par la méthode du maximum de vraisemblance : Éléments de théorie et pratiques sur logiciel," *Université de Kinshasa*, 2019.
- [9] B. JOURDAIN, "Probabilités et statistique pour l'ingénieur," *Ecole des Ponts, Paris Tech.*, 2018.

المخلص

في هذا العمل ، درسنا طريقة الاحتمال الأقصى هي إحدى الطرق الأكثر استخداما في الإحصاء الذي تم الحصول عليه بطريقة الاحتمالية القصوى له بعض الخصائص المثيرة للاهتمام ، على وجه الخصوص التقارب ، الكفاءة ، الحالة الطبيعية المقاربة

تم تطوير هذه الطريقة من قبل الإحصائي رونالد أيلمر فيشر في عام 1922

Résumé

Dans ce travail, nous avons étudié la méthode du maximum de vraisemblance. Elle est l'une des méthodes la plus utilisée en statistique. L'estimateur obtenue par la méthode du maximum de vraisemblance possède des propriétés intéressantes, notamment : la convergence, l'efficacité, la normalité asymptotique.

Cette méthode a été développée par le statisticien [Ronald Aylmer Fisher](#) en 1922

Abstract

In this work, we have studied the maximum likelihood method. It is one of the most widely used methods in statistics. The estimator obtained by the maximum likelihood method has interesting properties, in particular: convergence, efficiency, asymptotic normality.

This method was developed by statistician [Ronald Aylmer Fisher](#) in 1922