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

ABBA Manar

Titre :

Sur la théorie de l'estimation paramétrique

Membres du Comité d'Examen :

Dr.	**CHERFAOUI Mouloud**	UMKB	Président
Dr.	**SAYAH Abdallah**	UMKB	Encadreur
Dr.	**ABDELLI Jihane**	UMKB	Examinateur

Juin 2019

Dédicace

Je dédie ce humble travail à mes chers parents

mes frères

mes amis.

REMERCIEMENTS

Tout d'abord, je remercie Dieu le tout puissant qui m'a donné la force et le savoir afin d'accomplir ce travail.

Un grand merci pour mon encadreur **Dr.SAYAH Abdallah** d'avoir accepté de dériger ce projet et pour la confiance

qu'il nous a accordée, et surtout ses encouragements, et ses précieux conseils.

Je tiens aussi à remercier également **Dr.CHERFAOUI Mouloud** et **ABDELLI Jihane,** membres de jury,

pour avoir fait l'honneur d'accepter de jurer ce travail.

Je remercie tous ceux qui nous ont enseignés durant toutes nos études à département de mathématique.

Finalement un grand merci à mes familles, mes chères amie et à toutes les personnes qui nous ont encouragées pendant

la réalisation de mon mémoire.

Table des matières

Table des figures

Liste des tableaux

Introduction

La théorie de l'estimation étudie les propriétés des estimateurs et des méthodes générales d'estimation. L'objectif est de comparer les lois d'échantillonage des estimateurs. Elle consiste à approximer les valeurs exacte et inconnues des paramètres d'une population statistique considéré ou d'un modèle mathématique à partir d'observation d'individus s'appelle échantillon. Le paramètre de la population est estimé à partir d'une statistique calculée sur la base d'un échantillon.

L'estimation ponctuelle d'un paramètre consiste à évaluer la valeur du paramètre de la population à l'aide d'une valeur unique prise dans un échantillon.

Pour évaluer la précision d'un estimateur, il est d'usage de construire un intervalle de confiance autour de cet estimateur.

Soit X une variable aléatoire associée à un certain phénomène aléatoire observable de facon répétée. Notre objectif est d'estimer certaines caractéristiques d'intérêt de sa loi (la moyenne, la variance, ...) sur la base d'une série d'observations $x_1, x_2, ..., x_n$.

Considérerons toujours, même si des développements analogues sont possibles dans d'autre circonstances, que $x_1, x_2, ..., x_n$ sont des réalisations d'un n échantillon aléatoire $X_1, X_2, ..., X_n$.

Cette hypothèse sur nos observations qui peut être plus ou moins réaliste est nécessaire pour étudier de facons simple, en termes probabilistes, la qualité des estimations que l'on cherche à produire.

Dans le cadre de ce mémoire, on va s'intéresser aux théories de l'estimation paramétrique, aussi nous donnons un aperçu historique sur les trois méthodes d'estimations : la méthode du maximum de vraisemblance, la méthode des moments et la méthode d'estimation par intervalle de confiance.

La méthode du maximum de vraisemblance à été introduite par Gauss, cette approche est habituellement associée au nom du statisticien anglais Fisher, qui à découvert cette méthode d'inférence et a été le premier à donner les bases d'une théorie de l'estimation paramétrique fondée sur la vraisemblance.

La méthode des moments a en premier lieu était discutée par K.Pearson, puis elle été généralisée par L.Hansen.

La méthode d'estimation par intervalle de confiance a été introduite dans la statistique par Jerzy Neyman.

Ce mémoire est partagée en deux chapitres.

Dans le premier chapitre on définit qu'est ce qu'un estimateur, les qualités d'un estimateur et recherche du meilleur estimateur.

Dans le deuxième chapitre nous aborderons trois méthodes d'estimation paramétrique (la méthode du maximum de vraisemblance, la méthode des moments et la méthode d'estimation par intervalle de confiance).

Chapitre 1

Estimation

Un estimateur est une statistique permettant d'évaluer un paramètre inconnu d'une loi de probabilité. Il peut par exemple servir à estimer un certains caractéristiques d'une population totale à partir des données obtenu sur un échantillon.

1.1 Quelques définitions

1.1.1 Echantillon

Définition 1.1.1 *On appelle* n *échantillon de loi* F, *toute suite de variable aléatoire (v.a) indépendantes et identiquements distribuées (i.i.d).*

1.1.2 Statistique

Définition 1.1.2 *Une statistique* T_n *est une fonction mesurable des v.a* $X_i, i = 1, ..., n$ *qui donne une information d'un paramètre de la population*

$$T_n = f(X_1, X_2, ..., X_n).$$

Une statistique peut être à valeurs dans $\mathbb{R}$ où $\mathbb{R}^d$; dans le cas de $\mathbb{R}^d$, on dit que T_n est une statistique vectorielle.

1.1.3 Estimateur

Définition 1.1.3 *Soit un n échantillon $(X_1, ..., X_n)$, suit une loi de probabilité de paramètre $\theta \in \Theta$ (ouvert de $\mathbb{R}^d; d \geq 1$). On appelle estimateur de θ toute v.a à fonction des X_i donnée par :*

$$T_n = f(X_1, X_2, ..., X_n),$$

où

$$f : \mathbb{R}^n \to \mathbb{R}^d; d \geq 1.$$

T_n fournit une réalisation $(x_1, x_2, ..., x_n)$ qui est une estimation ponctuelle de θ :

$$\theta_n = f(x_1, x_2, ..., x_n).$$

Remarque 1.1.1 *L'estimation est dite ponctuelle si on estime un paramètre de la population avec un seul nombre.*

Remarque 1.1.2 *Pour un même paramètre, il peut y avoir plusieurs estimateurs possibles.*

1.2 Propriétes génerales d'un estimateur

T_n désigne l'estimateur du paramètre θ.

Tout estimateur peut donner lieu à l'écriture :

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

où $B(n,\theta)$ est le biais de T_n.

1.2.1 Estimateur sans biais

Définition 1.2.1 *On dit que T_n est un estimateur sans biais de θ si :*

$$\mathbb{E}(T_n) = \theta \text{ c'est à dire (c-à-d) } B(n,\theta) = 0.$$

Exemple 1.2.1 *Soit $(X_1, X_2, ..., X_n)$ un n échantillon telles que $\mathbb{E}(X_i) = m$, $\forall i = 1, ..., n$ $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (la moyenne empirique) est un estimateur sans biais de m, en effet :*

$$\mathbb{E}\left(\overline{X}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i\right) = \frac{1}{n}n\mathbb{E}\left(X_1\right) = \frac{1}{n}nm = m.$$

1.2.2 Estimateur asymptotiquement sans biais

Définition 1.2.2 *On dit que T_n est un estimateur asymptotiquement sans biais de θ si :*

$$\lim_{n\to\infty} \mathbb{E}\left(T_n\right) = \theta \text{ c-à-d } \lim_{n\to\infty} B\left(n,\theta\right) = 0.$$

Exemple 1.2.2 *Soit $(X_1, X_2, ..., X_n)$ un n échantillon telles que $\mathbb{V}\left(X_i\right) = \sigma^2 < \infty, \forall i = 1, ..., n$.*
$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$*, la variance empirique est un estimateur asymptotiquement sans biais de σ^2, en effet :*

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

$$\lim_{n\to\infty}\mathbb{E}\left(S^2\right) = \lim_{n\to\infty}\left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2.$$

1.2.3 Estimateur avec biais (biaisé)

Définition 1.2.3 *On dit que T_n est un estimateur avec biais de θ si :*

$$\mathbb{E}(T_n) \neq \theta \quad \text{c-à-d } B(n,\theta) = \mathbb{E}(T_n - \theta) = \mathbb{E}(T_n) - \theta \neq 0$$

Exemple 1.2.3 *Soit $(X_1, X_2, ..., X_n)$ un n échantillon, S^2est un estimateur biaisé de σ^2.*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 - \left(\overline{X} - m\right)^2,$$
$$\mathbb{E}\left(S^2\right) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\left(\frac{n-1}{n}\right) \neq \sigma^2.$$

1.2.4 Estimateur convergent

La convergence en probabilité

On dit que la suite $(X_n)_{n\geq 1}$converge en probabilité vers X, et on note $X_n \underset{n\to\infty}{\overset{p}{\to}} X$ si :

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

ou,

$$\forall\varepsilon > 0, \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \leq \varepsilon\right) = 1.$$

Définition 1.2.4 *On dit que T_n est un estimateur convergent de θ si T_n tend vers θ en probabilité quand n tend vers ∞ c-à-d :*

$$\forall\varepsilon > 0, \lim_{n\to\infty}\mathbb{P}[|\, T_n - \theta \,| > \varepsilon] = 0 \quad \text{c-à-d } \; T_n \overset{p}{\to} \theta \text{ quand } n \to \infty.$$

La convergence presque sûre

On dit que la suite $(X_n)_{n\geq 1}$ converge presque sûrement vers X, et on note $X_n \underset{n\to\infty}{\overset{p.s}{\to}} X$ si :

$$\mathbb{P}\left(\forall w : X_n(w) \underset{n\to\infty}{\to} X(w)\right) = 1 \iff \mathbb{P}\left(\forall w : \left|\lim_{n\to\infty} X_n(w) = X(w)\right|\right) = 1,$$

ou

$$\forall \varepsilon > 0; \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty \text{ c-à-d la série converge.}$$

Remarque 1.2.1 *On parle d'estimateur fortement convergent lorsqu'on a convergence presque sûre.*

Proposition 1.2.1 *Soit T_n un estimateur convergent du paramètre θ, et soit ϕ une fonction de $\mathbb{R} \to \mathbb{R}$, continue au point θ. Alors $(\phi(T_n))$ est un estimateur convergent de $\phi(\theta)$.*

Définition 1.2.5 *On appelle erreur quadratique moyenne d'un estimateur T_n par rapport à θ la quantité :*

$$\mathbb{EQM}(T_n, \theta) = \mathbb{E}\left[(T_n - \theta)^2\right] = \mathbb{V}(T_n) + [B(n,\theta)]^2.$$

Proposition 1.2.2 *Si un estimateur est sans biais ou asymptotiquement sans biais et si sa variance tend vers 0, alors il est convergent.*

Exemple 1.2.4 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *est un estimateur convergent de $m = \mathbb{E}(X)$ car*
1-$\overline{X}$ est un estimateur sans biais de m,
2-$\mathbb{V}\left(\overline{X}\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(X) = \frac{\sigma^2}{n} \to 0$ *quand* $n \to \infty$,

Définition 1.2.6 *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-à-d :*

$$\mathbb{V}(T_1) \leq \mathbb{V}(T_2).$$

Exemple 1.2.5 *Soit $X_1, ..., X_n$ un n échantillon telles que*
$T_1 = \frac{X_1 + X_2}{2}$; $T_2 = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{X_1 - X_2}{2}$ *deux estimateurs sans biais de $m = \mathbb{E}(X_1)$,*

$$\mathbb{E}(T_1) = \mathbb{E}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}\mathbb{E}(X_1 + X_2) = \frac{1}{2}(\mathbb{E}(X_1) + \mathbb{E}(X_2)) = \frac{1}{2}2\mathbb{E}(X_1) = m,$$

et

$$\begin{aligned}\mathbb{V}(T_1) &= \mathbb{V}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}\left[\mathbb{V}(X_1) + \mathbb{V}(X_2) + 2cov(X_1, X_2)\right] \\ &= \frac{1}{4}\left[\mathbb{V}(X_1) + \mathbb{V}(X_2)\right] = \frac{1}{4}2\mathbb{V}(X_1) = \frac{1}{2}\sigma^2,\end{aligned}$$

on a $cov(X_1, X_2) = 0$ car X_1, X_2 sont indépendante.

$$\mathbb{E}(T_2) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{X_1 - X_2}{2}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) + \frac{1}{2}\left[\mathbb{E}(X_1) - \mathbb{E}(X_2)\right] = \frac{1}{n}n\mathbb{E}(X_1) = m,$$

et

$$\begin{aligned}\mathbb{V}(T_2) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{X_1 - X_2}{2}\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) + \frac{1}{4}\left[\mathbb{V}(X_1) + \mathbb{V}(X_2) - 2cov(X_1, X_2)\right] \\ &= \frac{1}{n^2}n\mathbb{V}(X_1) + \frac{1}{4}2\mathbb{V}(X_1) = \frac{1}{n}\sigma^2 + \frac{1}{2}\sigma^2 \\ &= \frac{2+n}{2n}\sigma^2,\end{aligned}$$

alors $\mathbb{V}(T_1) < \mathbb{V}(T_2)$, donc T_1est plus efficace que T_2.

1.3 Statistique exhaustive

Définition 1.3.1 *On considère un n échantillon $(X_1, X_2, ..., X_n)$. On dit qu'une statistique T est exhaustive pour θ si la loi conditionnelle de l'échantillon $X = (X_1, X_2, ..., X_n)$*

sachant $T=t$ n'est pas une fonction de θ c-à-d

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

Donc, on peut dire qu'une fois T connu, nous n'obtenons plus d'autre information de l'échantillon concernant θ et donc que T porte toute l'information disponible sur θ.

Exemple 1.3.1 *Soit X une v.a qui suit la loi de Poisson $\mathcal{P}\left(\theta\right)$, soit $\left(X_1, X_2, ..., X_n\right)$ un échantillon, et soit la statistique $T=\sum\limits_{i=1}^{n} X_i$.*
On montre que la statistique T est une statistique exhaustive pour θ.
On a

$$\mathbb{P}\left(X_1, X_2, ..., X_n/T=t\right) = \frac{\mathbb{P}\left(X_1, X_2, ..., X_n, T=t\right)}{\mathbb{P}\left(T=t\right)}.$$

La statistique T suit la $\mathcal{P}\left(n\theta\right)$, alors

$$\mathbb{P}\left(T=t\right) = \frac{e^{-n\theta}\left(n\theta\right)^t}{t!},$$

et

$$\mathbb{P}\left(X_1, X_2, ..., X_n, T=t\right) = \mathbb{P}\left(X_1, X_2, ..., X_{n-1}, T-\sum_{i=1}^{n-1} X_i\right),$$

$$\mathbb{P}\left(X_1, X_2, ..., X_{n-1}, T-\sum_{i=1}^{n-1} X_i\right) = \prod_{i=1}^{n-1}\left(\frac{e^{-\theta}\theta^{x_i}}{(x_i)!}\right)\frac{e^{-\theta}\theta^{t-\sum\limits_{i=1}^{n-1} x_i}}{\left(t-\sum\limits_{i=1}^{n-1} x_i\right)!} = \frac{e^{-n\theta}\theta^t}{(x_1)!\,(x_2)!...\,(x_{n-1})!\left(t-\sum\limits_{i=1}^{n-1} x_i\right)!},$$

d'où

$$\mathbb{P}\left(X_1, X_2, ..., X_n/T=t\right) = \frac{t!}{n^t\,(x_1)!\,(x_2)!...\,(x_{n-1})!\left(t-\sum\limits_{i=1}^{n-1} x_i\right)!}.$$

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

Dans cet exemple on a vu que le calcul de la probabilité conditionnelle est loin d'être immédiat. Le théorème suivant va nous simplifier la tâche.

1.3.1 Théorème de factorisation

Soit un n échantillon d'une v.a X. On note $L(x,\theta)$ la densité jointe de $X=(X_1,X_2,...,X_n)$, elle est appellée fonction de vraissemblance de $\theta\in\Theta$ $\left(\text{ouvert de } \mathbb{R}^d, d\geq 1\right)$. Alors on dit que $T(X)$ est une statistique exhaustive pour θ si et seulement s'il existe deux fonctions mesurables g et h à valeurs positives(dans $\mathbb{R}_+$) telles 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

$$L(x,\theta)=\begin{cases}\prod_{i=1}^{n}f(x_i,\theta) & \text{si } X \text{ est absolument continue}\\ \prod_{i=1}^{n}\mathbb{P}(x_i,\theta) & \text{si } X \text{ est discrète}\end{cases}\quad \text{où } \theta\in\Theta, x\in\mathbb{R}^n\ ,$$

et $g(T(x),\theta)$ est la densité de la statistique T et $h(x)$ ne dépend pas de θ.

Exemple 1.3.2 *Soit $X\sim\mathcal{U}[0,\theta]$, $\theta>0$, et $X=(X_1,X_2,...,X_n)$ un échantillon, $T=\sup\limits_{1\leq i\leq n}X_i$, d'aprés le théorème de factorisation, la statistique T est exhaustive car :*

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

$$L(x,\theta)=\begin{cases}\prod_{i=1}^{n}f(x_i,\theta)\\ \prod_{i=1}^{n}\left(\dfrac{1}{\theta}\right)=\dfrac{1}{\theta^n},\ si\ 0\leq x\leq\theta\\ 0 \qquad sinon\end{cases}$$

La fonction de répartition de T :

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

La densité de T :

$$g(t,\theta)=\begin{cases} n\dfrac{1}{\theta}(\dfrac{t}{\theta})^{n-1}=\dfrac{n}{\theta^{n}}t^{n-1} & si\ 0\leq t\leq\theta \\ 0 & sinon \end{cases}$$

Donc

$$h(x)=\frac{L}{g}=\frac{\dfrac{1}{\theta^{n}}}{\dfrac{nt^{n-1}}{\theta^{n}}}=\frac{1}{nt^{n-1}}\qquad 0\leq x\leq\theta$$

ne dépend pas de θ.

Remarque 1.3.1 *Le principe de factorisation nous donne un moyen de reconnaître si une statistique est exhaustive, mais ne permet pas de la construire où même de savoir s'il en existe une..*

1.3.2 Théorème de Darmois

Soit X une v.a dont le domaine de définition ne dépent pas de θ. Une condition nécessaire et suffisante pour que l'échantillon $(X_1, X_2, ..., X_n)$ admette une statistique exhaustive est que la forme de la densité soit :

$$f(x,\theta)=\exp\left[a(x)\alpha(\theta)+b(x)+\beta(\theta)\right], \qquad \text{(famille exponentielle)}$$

où

a est une fonction mesurable à valeurs dans $\mathbb{R}^d, d\geq 1$,

α est une fonction à valeurs dans $\mathbb{R}^d, d \geq 1$,

β est une fonction rèelle ne dépend pas de x,

b est une fonction borélienne ne dépend pas de θ.

Si f est de cette forme et si de plus l'application $x_1 \rightarrow \sum_{i=1}^{n} a(x_i)$ est bijective et continûment différentiable pour tout i, alors $T = \sum_{i=1}^{n} a(X_i)$ est une statistique exhaustive particulière.

Preuve. Voir [4], page 293-294. ■

Exemple 1.3.3 *Soit X v.a suit une loi γ de paramètre θ inconnu alors :*

$$\begin{aligned} f(x,\theta) &= \frac{1}{\Gamma(\theta)} \exp(-x)\, x^{\theta-1}, x \geq 0; \theta > 1, \\ &= \exp\left[\ln\left(\frac{1}{\Gamma(\theta)} \exp(-x)\, x^{\theta-1}\right)\right], \\ &= \exp\left[-\ln\Gamma(\theta) - x + (\theta-1)\ln x\right], \end{aligned}$$

telles que $\Gamma(\theta) = \int_0^\infty x^{\theta-1} \exp(-x)\, dx$ *et* $\Gamma(\theta+1) = \theta!$,
alors : $a(x) = \ln x, \alpha(\theta) = \theta - 1, b(x) = -x, \beta(\theta) = -\ln\Gamma(\theta)$, *et on a l'application* $x_1 \rightarrow \sum_{i=1}^{n} a(x_i) = \sum_{i=1}^{n} \ln x_i$ *est bijective et continûment différentiable pour tout i.*
Donc, la statistique exhaustive est : $\sum_{i=1}^{n} \ln X_i = \ln\left(\prod_{i=1}^{n} X_i\right)$.

1.3.3 Famille exponentielle

Définition 1.3.2 *Soit X une v.a réelle, dont la loi de probabilité dépend d'un paramètre $\theta \in \mathbb{R}^d$. On dit que la loi de X appartient à la famille exponentielle si et seulement si $f(x,\theta)$(cas continu) ou $\mathbb{P}(x,\theta)$ (cas discret) respectivement est de la forme :*

$$f(x,\theta) = c(\theta)\, h(x) \exp\left\{\sum_{j=1}^{d} a_j(\theta)\, T_j(x)\right\}.$$

D'après le critère de factorisation, il est évident que la statistique$(T_1, ..., T_d)$ *est exhaustive.*

Si l'on a un échantillon de taille n *:*

$$f(x_1, ..., x_n; \theta) = c^2(\theta) h(x_1) ... h(x_n) \exp\left\{\sum_{j=1}^{d} a_j(\theta) \sum_{i=1}^{n} T_j(x_i)\right\}.$$

La statistique $\left(\sum_{i=1}^{n} T_1(x_i), ..., \sum_{i=1}^{n} T_d(x_i)\right)$ *est exhaustive.*

Exemple 1.3.4 *Soit* X *une v.a suit une loi de Poisson* $\mathcal{P}(\lambda)$

$$\mathbb{P}(x, \lambda) = \frac{\lambda^x}{x!} \exp\{-\lambda x\}$$

par comparaison, on trouve :

$c(\lambda) = \lambda^x, h(x) = \frac{1}{x!}, a_1(\lambda) = \lambda, T_1(x) = -x$. *Donc* X *appartient à la famille exponentielle, et d'après le critère de factorisation* T_1 *est exhaustive.*

1.4 Information de Fisher

Soit $(\boldsymbol{E}, \mathcal{A}, \mathbb{P}_\theta)$ un espace de probabilité telles que :

$\boldsymbol{E}$ est l'ensemble fondamentale, $\mathcal{A}$ est une tribu sur $\boldsymbol{E}$, et $\mathbb{P}_\theta$ est une probabilité définie sur l'espace probabilisable $(\boldsymbol{E}, \mathcal{A})$.

Définition 1.4.1 *On appelle modèle statistique (paramétrique) une famille d'espaces de probabilités* $(\boldsymbol{E}, \mathcal{A}, \mathbb{P}_\theta)$, $\theta \in \Theta$.

On considère le modèle paramétrique $(\boldsymbol{E}, \mathcal{A}, \mathbb{P}_\theta), \theta \in \Theta$, de v.a X, vérifie les hypothèses suivantes :

$H_1 : \forall\theta \in \Theta$ où Θ est un ouvert de $\mathbb{R}$.

H_2 : Pour tout $\theta \in \Theta$, les lois $\mathbb{P}_\theta$ ont le même support, qui ne dépend pas de θ c-à-d $L(x, \theta) > 0 \ \forall x \in \mathbb{R}^n, \theta \in \Theta$.

H_3 : Pour tout $\theta \in \Theta$ et $x \in$ $\boldsymbol{E}$,$\dfrac{\partial L(x,\theta)}{\partial \theta}$ et $\dfrac{\partial^2 L(x,\theta)}{\partial \theta^2}$ de la vraisemblance existent et continuent c-à-d $L(x,\theta) \in C^2(\theta)$. Où $C^2(\theta)$ est la classe des fonctions continues et dérivables deux fois.

H_4 : Pour tout $\theta \in \Theta$, les fonctions $L'(x,\theta) = \dfrac{\partial L(x,\theta)}{\partial \theta}$et $L"(x,\theta) = \dfrac{\partial^2 L(x,\theta)}{\partial \theta^2}$ sont intégrables et on a : pour tout $A \in \mathcal{A}$,

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_A L(x,\theta)\, dx &= \int_A L^{'}(x,\theta)\, dx, \\ \frac{\partial^2}{\partial \theta^2} \int_A L(x,\theta)\, dx &= \int_A L^{''}(x,\theta)\, dx, \end{aligned}$$

on définit la fonction score

$$S(X,\theta) = \frac{\partial \ln L(X,\theta)}{\partial \theta} = \frac{L^{'}(X,\theta)}{L(X,\theta)},$$

sous les hypothèses précédentes, cette v.a est centrée car :

$$\begin{aligned} \mathbb{E}(S(X,\theta)) &= \int_{\boldsymbol{E}} S(X,\theta)\, L(x,\theta)\, dx = \int_{\boldsymbol{E}} \frac{\partial \ln L(X,\theta)}{\partial \theta} L(x,\theta)\, dx \\ &= \int_{\boldsymbol{E}} L^{'}(x,\theta)\, dx = \frac{\partial}{\partial \theta} \int_{\boldsymbol{E}} L(x,\theta)\, dx = 0 \end{aligned}$$

car

$$\int_{\boldsymbol{E}} L(x,\theta)\, dx = 1 \text{ par définition de la densité.}$$

Définition 1.4.2 *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 :*

$$I_n(\theta) = \mathbb{V}(S(X,\theta)) = \mathbb{E}\left[(S(X,\theta))^2\right] = \mathbb{E}\left[\left(\frac{\partial \ln L(X,\theta)}{\partial \theta}\right)^2\right].$$

Où :

$$L\left(X_1, X_2, ..., X_n; \theta\right) = \prod_{i=1}^{n} f\left(X_i, \theta\right).$$

Théorème 1.4.1 *Si le domaine de définition de X ne dépend pas du paramètre θ alors :*

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

Preuve. Voir [4], page 295-296. ■

Exemple 1.4.1 *Soit X une v.a telles que $X \sim \mathcal{N}\left(m, \sigma^2\right)$ alors l'information de Fisher pour cette v.a est :*

$$\begin{aligned}
L\left(x, \theta\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}, \\
\ln L\left(x, \theta\right) &= -\frac{1}{2}\ln\left(2\pi\right) - \ln\left(\sigma\right) - \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2, \\
\frac{\partial}{\partial m}\ln L\left(x, \theta\right) &= \frac{1}{\sigma^2}\left(x-m\right), \\
\frac{\partial^2}{\partial m^2}\ln L\left(x, \theta\right) &= -\frac{1}{\sigma^2}, \\
I\left(m\right) &= -\mathbb{E}\left[\frac{\partial^2}{\partial m^2}\ln L\left(x, \theta\right)\right] = \frac{1}{\sigma^2}.
\end{aligned}$$

1.4.1 Propriétés de l'information de Fisher

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

La positivité

On a $I_n\left(\theta\right) = \mathbb{V}\left(S\left(X, \theta\right)\right) \geq 0, \forall\theta.$

L'additivité

Soient X, Y deux v.a indépendantes, on note $f(X,\theta)$ et $g(X,\theta)$ leurs loi respectives, et on désigne par $I_X(\theta), I_Y(\theta)$ et $I_{(X,Y)}(\theta)$ les informations au point θ respectivement fournit par X, Y et le couple(X,Y).

Théorème 1.4.2 *Sous les hypothèses H_1, H_2et H_3on a :*

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

Théorème 1.4.3 *Soit un n échantillon $(X_1, ..., X_n)$. Sous les hypothèses H_1, H_2 et H_3 on a :*

$$I_n(\theta) = I_{X_1}(\theta) + I_{X_2}(\theta) + ... + I_{X_n}(\theta) = nI_{X_1}(\theta).$$

Ceci veut dire que chaque observation a la même information.

Exemple 1.4.2 *·Loi normale (espérance) : soit $X \sim \mathcal{N}(m, \sigma^2)$*
On dispose d'un échantillon $(X_1, ..., X_n)$ alors :

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

donc

$$I_n(m) = \frac{n}{\sigma^2} = nI(m).\text{avec } I(m) = \frac{1}{\sigma^2}.$$

·Loi normale (variance) : soit $X \sim \mathcal{N}(m, \sigma^2)$

On dispose d'un échantillon $(X_1, ..., X_n)$ *alors :*

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

$$\ln\left(L\left(x,\sigma^2\right)\right) = -\frac{1}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left(x-m\right)^2,$$

$$\frac{\partial}{\partial\sigma^2}\ln\left(L\left(x,\sigma^2\right)\right) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\left(x-m\right)^2,$$

$$\frac{\partial^2}{\left(\partial\sigma^2\right)^2}\ln\left(L\left(x,\sigma^2\right)\right) = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\left(x-m\right)^2,$$

$$I\left(\sigma^2\right) = -\mathbb{E}\left[\frac{\partial^2}{\left(\partial\sigma^2\right)^2}\ln\left(L\left(X,\sigma^2\right)\right)\right] = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4}\mathbb{E}\left[\left(\frac{X-m}{\sigma}\right)^2\right],$$

$\mathbb{E}\left[\left(\frac{X-m}{\sigma}\right)^2\right] = 1$ *car c'est l'esperance de la loi* $\mathcal{X}_1^2$,
donc

$$I\left(\sigma^2\right) = \frac{1}{\sigma^4} - \frac{1}{2\sigma^4} = \frac{1}{2\sigma^4},$$

alors

$$I_n\left(\sigma^2\right) = \frac{n}{2\sigma^4}.$$

1.5 Lien entre l'information de Fisher et la statistique exhaustive

La relation très étroite entre les notions de statistique exhaustive et l'information de Fisher.

Soit $(\boldsymbol{E}, \mathcal{A}, \mathbb{P}_\theta), \theta \in \Theta$ un modèle statistique, $I(\theta)$ l'information continue dans ce modèle sur θ et T une statistique$(\boldsymbol{E}, \mathcal{A}) \to (Y, \mathcal{B})$, où $(Y, \mathcal{B})$ espace mesurable.

Théorème 1.5.1 *Pour toute statistique T on a :*

$$I_T(\theta) \leq I(\theta)$$

où $I_T(\theta)$ est l'information de Fisher au point θ dans $\left(Y, \mathbb{P}_\theta^T\right)$.

Théorème 1.5.2

$$I_T(\theta) = I(\theta) \iff T \text{ exhaustive.}$$

Exemple 1.5.1 *Soient $X_1, ..., X_n$ un échantillon qui suit $\mathcal{N}(m, \sigma^2)$. Considérons la statistique*

$$T = S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2,$$

on définit

$$Y_n = (n-1)\frac{T}{\sigma^2},$$

telles que $Y_n \sim \mathcal{X}_{n-1}^2$, d'après le théorème de Fisher.
Alors

$$f_{Y_n}(y) = \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} y^{\frac{n-3}{2}} \exp\left(-\frac{y}{2}\right) I_{\{y \geq 0\}}.$$

On fait un changement de variable $Y_n \to T$, on obtient :

$$f_T\left(t, \sigma^2\right) = \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\dfrac{n-1}{2}\right)} \left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} t^{\frac{n-3}{2}} \exp\left(-\frac{(n-1)\,t}{2\sigma^2}\right) I_{\{t \geq 0\}}.$$

On calcule l'information de Fisher

$$\ln f_T\left(t, \sigma^2\right) = c(t) - \frac{(n-1)\,t}{2\sigma^2} + \frac{n-1}{2}\ln\left(\frac{n-1}{\sigma^2}\right),$$

où $c(t)$ est une constante qui ne dépend pas de σ

$$\frac{\partial^2}{\left(\partial\sigma^2\right)^2}\ln f_T\left(t,\sigma^2\right)=\frac{n-1}{2\sigma^4}-\frac{(n-1)\,t}{\sigma^6},$$

donc

$$\begin{aligned}I_T\left(\sigma^2\right)&=\frac{n-1}{\sigma^6}\mathbb{E}\left(T\right)-\frac{n-1}{2\sigma^4},\\&=\frac{n-1}{\sigma^4}-\frac{n-1}{2\sigma^4},\\&=\frac{n-1}{2\sigma^4}.\end{aligned}$$

D'autre part, on a :

$$I_n\left(\sigma^2\right)=nI\left(\sigma^2\right)=\frac{n}{2\sigma^4}.$$

Il s'en suit que pour une taille d'échantillon finie n, la variance empirique, T n'est pas exhaustive pour σ^2 puisque $I_T\left(\sigma^2\right)<I_n\left(\sigma^2\right)$.

1.6 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 1.6.1 ***(unicité)*** *S'il existe un estimateur de θ sans biais de variance minimale, il est unique presque sûrement.*

Théorème 1.6.2 ***(Rao-Blackwell)*** *Soit T un estimateur sans biais de θ quelconque et U une statistique exhaustive pour θ. Alors $T^*=\mathbb{E}\left(T/U\right)$ est un estimateur sans biais de θ au moins aussi bon que T.*

Théorème 1.6.3 *S'il existe une statistique exhaustive U de θ, alors l'estimateur T sans biais de θ de variance minimale (unique d'après le théorème précédent) ne dépend que de*

U.

Définition 1.6.1 *On dit q'une statistique exhaustive U d'un modèle statistique paramétrique est complète si pour toute fonction borélienne $h : \mathbb{R}^d \to \mathbb{R}$ telles que h est intégrable et*

$$\mathbb{E}\left[h\left(U\right)\right] = 0 \quad \forall \theta \in \Theta \implies h = 0 \ p.s.$$

Théorème 1.6.4 *(**Lehmann-Scheffé**) Si T^* est un estimateur sans biais de θ dépendant d'une statistique exhaustive complète U alors T^* est l'unique estimateur sans biais de variance minimale. En particulier si l'on dispose déja de T estimateur sans biais de θ,*

$$T^* = \mathbb{E}\left(T \diagup U\right).$$

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

Le résultat suivant nous indique qua 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 :

$$\mathbb{V}\left(T\right) \geq \frac{1}{I_n\left(\theta\right)},$$

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

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

où k est une fonction dérivable.

Preuve. Voir [4], page 302.

■

Définition 1.6.2 *On dit que T est un estimateur efficace si :*

1. T est un estimateur sans biais de θ alors

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

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

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

Théorème sur l'efficacité

·La borne de Cramer-Rao ne peut être atteinte que si la loi de X est de forme exponentielle :

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

car T est nécessairement exhaustive pour θ.

donc, il n'existe qu'une seule fonction $k(\theta)$ qui puisse être estimée efficacement :

$$k(\theta) = -\frac{\beta'(\theta)}{\alpha'(\theta)},$$

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

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

la variance minimale est :

$$\mathbb{V}(T_n) = -\frac{1}{n\alpha'(\theta)}\frac{\partial}{\partial\theta}\left(\frac{\beta'(\theta)}{\alpha'(\theta)}\right) = \frac{k'(\theta)}{n\alpha'(\theta)}.$$

Preuve. Voir [4], page 303-304. ■

Exemple 1.6.1 *Loi gamma de paramètre θ.*
On a vu que : $T = \sum_{i=1}^{n} \ln X_i$ *est exhaustive*
On a :

$$f(x,\theta) = \exp\left[(\theta-1)\ln x - x - \ln\Gamma(\theta)\right],$$

donc la loi gamma appartient à la famille exponentielle telles que $a(x) = \ln x, \alpha(\theta) = \theta - 1, \beta(\theta) = -\ln\Gamma(\theta), b(x) = -x.$
D'après le théorème de Darmois $T = \sum_{i=1}^{n} a(X_i) = \sum_{i=1}^{n} \ln X_i = \ln\left(\prod_{i=1}^{n} X_i\right)$ *est exhaustive.*
Donc d'aprés le théorème précédent la seule fonction qui puisse être estimée efficacement est :

$$k(\theta) = -\frac{\beta'(\theta)}{\alpha'(\theta)} = \frac{\Gamma'(\theta)}{\Gamma(\theta)} = \frac{\partial}{\partial\theta}\ln\Gamma(\theta),$$

l'estimateur efficace est :

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

Chapitre 2

Les méthodes d'estimation paramétrique

Dans les situations où il n'y a pas d'estimateur évident, on est amené à recourir à une méthode de construction d'un estimateur, les trois méthodes que nous présenterons ici étant celles du maximum de vraisemblance et des moments et la méthode d'estimation par intervalle de confiance.

2.1 La méthode du maximum de vraisemblance

Soit un échantillon $(X_1, ..., X_n)$ dont la loi mère appartient à la famille paramétrique de densités $f(x_i, \theta)$ (ou de fonction de probabilité $\mathbb{P}(x_i, \theta)$) $\theta \in \Theta, \Theta \subset \mathbb{R}^d$.

La vraisemblance $L(x_1, ..., x_n; \theta)$ represente la probabilité d'observer n uplet $(x_1, ..., x_n)$ pour une valeur fixée du paramètre θ. Dans la situation inverse ici où on a observé $(x_1, ..., x_n)$ sans connaître la valeur de θ, on va attribuer à θ la valeur qui parait la plus vraisemblable, compte tenu de l'observation dont on dispose, c-à-d celle qui va lui attribuer la plus forte probabilité.

Définition 2.1.1 *On appelle fonction de vraisemblance de θ pour une réalisation*

$x = (x_1, ..., x_n)$ *de l'échantillon* $X = (X_1, ..., X_n)$ *la fonction de* θ *:*

$$L(x_1, ..., x_n; \theta) = \begin{cases} \prod_{i=1}^{n} \mathbb{P}(x_i, \theta) & \text{si les } X_i \text{ sont discrètes} \\ \prod_{i=1}^{n} f(x_i, \theta) & \text{si les } X_i \text{ sont continues} \end{cases}.$$

Définition 2.1.2 *On appelle estimateur du maximum de vraisemblance (EMV) de* θ *la v.a qui maximise* $L(x, \theta)$ *et on a le notée* $\hat{\theta}$:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(x, \theta),$$

en d'autre termes :

$$L\left(x, \hat{\theta}\right) \geq L(x, \theta) \quad \forall \theta \in \Theta.$$

Caractéristiques

1. $L(x, \theta)$ n'a aucune raison d'être différentiable en θ.

   Exemple 2.1.1 *Soit* $X = (X_1, ..., X_n) \sim \mathcal{U}[0, \theta], \theta > 0$

   $$\begin{aligned} L(X; \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(X_i), \\ &= \frac{1}{\theta^n} I_{\left\{0 \leq \inf\limits_{1 \leq i \leq n} X_i \leq \sup\limits_{1 \leq i \leq n} X_i \leq \theta\right\}}, \\ &= \frac{1}{\theta^n} I_{\left\{0 \leq \inf\limits_{1 \leq i \leq n} X_i\right\}} \times I_{\left\{\sup\limits_{1 \leq i \leq n} X_i \leq \theta\right\}}. \end{aligned}$$

   Où

   $$I_{[0,\theta]}(x_i) = \begin{cases} 1 & si \;\; x_i \in [0, \theta] \\ 0 & sinon \end{cases}.$$

   L est maximum pour θ *minimum alors* $\hat{\theta} = \sup\limits_{1 \leq i \leq n} X_i$ *est l'EMV pour* θ.

2. Il n'y a aucune raison pour que l'EMV soit sans biais.

Exemple 2.1.2 *Le même exemple précédent on obtient que* $\hat{\theta} = \sup\limits_{1\leq i\leq n} X_i$ *est l'EMV de* θ. *Posons* $Y = \sup\limits_{1\leq i\leq n} X_i$ *on cherche la loi de* Y *:*

On a :

$$F(Y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(\sup_{1\leq i\leq n} X_i \leq y\right) = \mathbb{P}(X_1 \leq y, ..., X_n \leq y) = \left[\mathbb{P}(X_1 \leq y)\right]^n,$$

et

$$\mathbb{P}(X_1 \leq y) = \int_0^y \frac{1}{\theta} dt = \frac{y}{\theta},$$

alors

$$F(Y) = \left(\frac{y}{\theta}\right)^n,$$

et

$$f(y) = \frac{ny^{n-1}}{\theta^n} I_{(0\leq y\leq\theta)}.$$

Alors

$$\mathbb{E}(Y) = \int_0^\theta y f(y)\, dy = \frac{n}{\theta^n}\int_0^\theta y^n dy = \frac{n}{\theta^n}\left[\frac{y^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta,$$

donc Y *est un estimateur biaisé pour* θ.

3. L'EMV n'a aucune raison d'être unique.

Exemple 2.1.3 *Soit* $X \sim \mathcal{U}[\theta, \theta+1]$ *,*$\theta > 0$ *dont on possède un échantillon* $(X_1, ..., X_n)$ *La vraisemblance de l'échantillon en* θ *:*

$$\begin{aligned} L(x_1, ..., x_n; \theta) &= I_{\left\{\theta \leq \inf\limits_{1\leq i\leq n} X_i \leq \sup\limits_{1\leq i\leq n} X_i \leq \theta+1\right\}} \\ &= I_{\left\{\theta \leq \inf\limits_{1\leq i\leq n} X_i\right\}} \times I_{\left\{\theta \geq \sup\limits_{1\leq i\leq n} X_i - 1\right\}}. \end{aligned}$$

Tout estimateur $\hat{\theta}$ *compris entre* $\hat{\theta}_1 = \sup\limits_{1\leq i\leq n} X_i - 1$ *et* $\hat{\theta}_2 = \inf\limits_{1\leq i\leq n} X_i$ *est un EMV. Celui -ci ne sera unique que si* $\sup\limits_{1\leq i\leq n} X_i = \left(\inf\limits_{1\leq i\leq n} X_i\right) + 1$

2.1.1 Equation de vraisemblance

Soit un n échantillon $X = (X_1, ..., X_n)$ de valeurs $(x_1, ..., x_n)$, cette méthode permet de prendre comme estimation de $\theta \in \Theta$ la valeur $\hat{\theta}$ qui rend maximale la vraisemblance.

Cas d'un seul paramètre

L'estimateur du maximum de vraisemblance est une solution du système :

$$(I)\left\{\begin{array}{c} \dfrac{\partial L(x_1, ..., x_n; \theta)}{\partial \theta} = 0 \\ \dfrac{\partial^2 L(x_1, ..., x_n; \theta)}{\partial \theta^2}\left(\hat{\theta}\right) < 0 \end{array}\right. ,$$

ou bien le système :

$$(II)\left\{\begin{array}{c} \dfrac{\partial \ln L(x_1, ..., x_n; \theta)}{\partial \theta} = 0 \\ \dfrac{\partial^2 \ln L(x_1, ..., x_n; \theta)}{\partial \theta^2}\left(\hat{\theta}\right) < 0 \end{array}\right. ,$$

le système (I) est donc équivalent au système (II).

Propriétés des estimateurs du maximum de vraisemblance

Soit $\hat{\theta}$ l'EMV de θ.

Théorème 2.1.1 *S'il existe une statistique exhaustive, $\hat{\theta}$ est fonction de cette statistique exhaustive.*

Théorème 2.1.2 *S'il existe un estimateur efficace T du paramètre θ, il est identique (p.s) à l'unique EMV de θ.*

La réciproque est fausse dans le cas général c-à-d un EMV n'est pas obligatoirement efficace.

2.1.2 Application

Exemple 2.1.4 *Estimation de l'esperance d'une loi normale* $\mathcal{N}(m, \sigma^2)$; σ *connu*

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

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

d'où

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

alors

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

$$\frac{\partial \ln L(x_1, ..., x_n; m)}{\partial m} = 0 \implies \hat{m} = \overline{x},$$

et

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

$$\left.\frac{\partial^2 \ln L(x_1, ..., x_n; m)}{\partial^2 m}\right|_{m=\hat{m}} = -\frac{n}{\sigma^2} < 0,$$

alors l'EMV de m *est :*

$$\hat{m} = \overline{X}.$$

Exemple 2.1.5 *Estimation du paramètre d'une loi exponentielle* $\mathcal{E}(\lambda)$; $\lambda > 0$

$$f(x, \lambda) = \lambda \exp(-\lambda x) \;\; ; \;\; x \geq 0,$$

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

d'où

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

alors

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

et

$$\frac{\partial^2 \ln L(x_1,...,x_n;\lambda)}{\partial^2\lambda} = -\frac{n}{\lambda^2}$$
$$\left.\frac{\partial^2 \ln L(x_1,...,x_n;\lambda)}{\partial^2\lambda}\right|_{\lambda=\hat{\lambda}} = -n\overline{x}^2 < 0,$$

alors l'EMV de λ *est :*

$$\hat{\lambda} = \frac{1}{\overline{X}}.$$

2.2 La méthode des moments

Cette méthode peut être la méthode la plus ancienne de trouver des estimateurs ponctuels et la plus naturelle. L'idée de base est d'écrire les moments théoriques $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, ..., $\mathbb{E}(X^d)$ en fonction des paramètres.

Soit X une v.a qui suit une certaine loi de probabilité qui dépend d'un paramètre $\theta = (\theta_1, ..., \theta_d)$.

L'estimateur de θ par la méthode des moments (EMM) est obtenu en remplaçant les moments théoriques par les moments empiriques c-à-d :

$$\begin{cases} \mathbb{E}(X) = m_1 = f_1(\theta_1, ..., \theta_d), \\ \mathbb{E}(X^2) = m_2 = f_2(\theta_1, ..., \theta_d), \\ \quad . \\ \quad . \\ \quad . \\ \mathbb{E}(X^d) = m_d = f_d(\theta_1, ..., \theta_d), \end{cases} \Longleftrightarrow \begin{cases} \theta_1 = g_1(m_1, ..., m_d), \\ \theta_2 = g_2(m_1, ..., m_d), \\ \quad . \\ \quad . \\ \quad . \\ \theta_d = g_d(m_1, ..., m_d), \end{cases},$$

telles que

$$\begin{cases} \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \\ \hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \\ \quad . \\ \quad . \\ \quad . \\ \hat{m}_d = \frac{1}{n}\sum_{i=1}^{n} X_i^d, \end{cases},$$

alors, on obtient :

$$\begin{cases} \hat{\theta}_1 = g_1(\hat{m}_1, ..., \hat{m}_d), \\ \hat{\theta}_2 = g_2(\hat{m}_1, ..., \hat{m}_d), \\ \quad . \\ \quad . \\ \quad . \\ \hat{\theta}_d = g_d(\hat{m}_1, ..., \hat{m}_d), \end{cases}.$$

Remarque 2.2.1 *Si les $g_i; i = 1, .., d$ sont continues en $(m_1, m_2, ..., m_d)$ alors les estimateurs obtenus par la méthode des moments sont convergents presque sûrement.*

2.2.1 Applications

Exemple 2.2.1 ***Loi Gamma***

X suit la loi gamma de paramètres α et λ, telles que

$$\begin{cases} \mathbb{E}(X) = \dfrac{\alpha}{\lambda}, \\ \mathbb{V}(X) = \dfrac{\alpha}{\lambda^2}. \end{cases}$$

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

$$\begin{cases} \mathbb{E}(X) = \dfrac{\alpha}{\lambda} \\ \mathbb{V}(X) = \dfrac{\alpha}{\lambda^2} \end{cases}, \quad \Longrightarrow \begin{cases} \alpha = \dfrac{[\mathbb{E}(X)]^2}{\mathbb{V}(X)} \\ \lambda = \dfrac{\mathbb{E}(X)}{\mathbb{V}(X)} \end{cases}.$$

Si on dispose d'un échantillon $(X_1, ..., X_n)$ de la loi gamma de paramètres α et λ, la moyenne empirique $\overline{X}$ et la variance empirique S^2 sont des estimateurs convergents de $\mathbb{E}(X)$ et $\mathbb{V}(X)$ respectivement. Donc on déduit deux estimateurs convergents de α et λ :

$$\begin{cases} \hat{\alpha} = \dfrac{\overline{X}^2}{S^2} \\ \hat{\lambda} = \dfrac{\overline{X}}{S^2} \end{cases}.$$

Exemple 2.2.2 ***Loi béta***

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

$$\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta},$$
$$\mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

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

$$\begin{cases} \mathbb{E}(X) = \dfrac{\alpha}{\alpha+\beta} \\ \mathbb{V}(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{cases}, \implies \begin{cases} \alpha = \dfrac{\mathbb{E}(X)\left(\mathbb{E}(X) - \mathbb{E}(X)^2 - \mathbb{V}(X)\right)}{\mathbb{V}(X)} \\ \beta = \dfrac{\mathbb{E}(X) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^3 - \mathbb{V}(X) + \mathbb{E}(X)\mathbb{V}(X)}{\mathbb{V}(X)} \end{cases}.$$

Si on dispose d'un échantillon $(X_1, ..., X_n)$ de la loi béta de paramètres α et β, la moyenne empirique $\overline{X}$ et la variance empirique S^2 sont des estimateurs convergents de $\mathbb{E}(X)$ et $\mathbb{V}(X)$ respectivement. Donc on déduit deux estimateurs convergents de α et β :

$$\begin{cases} \hat{\alpha} = \dfrac{\overline{X}\left(\overline{X} - \overline{X}^2 - S^2\right)}{S^2}, \\ \hat{\beta} = \dfrac{\overline{X} - 2\overline{X}^2 + \overline{X}^3 - S^2 + \overline{X}S^2}{S^2}. \end{cases}$$

2.3 La méthode d'estimation par intervalle de confiance

L'estimation ponctuelle d'un paramètre θ donne une valeur numérique unique à ce paramètre, mais, n'apporte aucune information sur la précision des résultats, c-à-d qu'elle ne tient pas compte des erreurs ducs aux fluctuations d'échantillonnage, par exemple pour evaluer la confiance que l'on peut avoir en une estimation, il est nécessaire de lui associer un intervalle qui contient avec une certaine probabilité la vraie valeur du paramètre, c'est l'estimation par intervalle de confiance.

Définition 2.3.1 *Soit $(X_1, ..., X_n)$ un échantillon de loi $\mathbb{P}$. On appelle intervalle de confiance (IC) de niveau de confiance $1-\alpha$ telles que $\alpha \in [0,1]$ donné, un intervalle aléatoire $[\theta_1, \theta_2]$ où $\theta_1 \leq \theta_2$ sont deux statistiques, fonction de l'échantillon, telles que :*

$$\mathbb{P}(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha.$$

Remarque 2.3.1 *α est donc la probabilité que $[\theta_1, \theta_2]$ ne recouvre pas la vraie valeur du paramètre.*

Remarque 2.3.2 *La détermination des bornes d'un intervalle de confiance dépend de la coupure de α en α_1 et α_2 ; telles que θ_1 est le fractile d'ordre α_1 et θ_2 est le fractile d'ordre $(1-\alpha_2)$ d'une certaine loi. Cependant deux cas sont possibles*

·Recherche d'un intervalle bilatéral. Correspondant à $\alpha_1 \neq 0$ et $\alpha_2 \neq 0$. Il n'y a aucune raison d'avoir $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ sauf pour certaines lois symétriques .

· Recherche d'un intervalle unilatéral de la forme $[\theta_1, +\infty[$ associé à $\alpha_1 = \alpha$ et $\alpha_2 = 0$ ou de forme$]-\infty, \theta_2]$ associé à $\alpha_1 = 0$ et $\alpha_2 = \alpha$.

Remarque 2.3.3 *les valeurs usuelles de α sont* $0.1, 0.05$ *et* 0.01 .

2.3.1 Estimation par intervalle de confiance pour les paramètres de la loi normale

Soit $(X_1, ..., X_n)$ un échantillon d'une v.a X qui suit une loi normale $\mathcal{N}(m, \sigma^2)$,
$\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ sa moyenne empirique et $S^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ sa variance empirique.
Pour construire un intervalle de confiance relatif à l'un des deux paramètres de cette loi, l'autre paramètre étant connu ou non. Ceci correspond aux différentes situations que nous allons étudier maintenant.
Nous rassemblons ci dessous et nous admettrons, les trois résultats permettant de calculer les intervalles de confiance de la moyenne m et de la variance σ^2.

Théorème 2.3.1 *Si $(X_1, ..., X_n)$ est un échantillon de la loi $\mathcal{N}(m, \sigma^2)$, alors :*

1. $\dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$ suit la loi normale $\mathcal{N}(0, 1)$ (d'aprés le théorème centrale limite (TCL)).
2. $\dfrac{\overline{X} - m}{\frac{S}{\sqrt{n-1}}}$ suit une loi de student $\mathcal{T}_{n-1}$ de degré de liberté $(n-1)$.
3. $\dfrac{nS^2}{\sigma^2}$ suit la loi du khi-deux $\mathcal{X}^2_{n-1}$, degré de liberté $(n-1)$.

Estimation par intervalle de confiance de la moyenne

Si la variance est connue

On sait que $\overline{X} \sim \mathcal{N}\left(m, \dfrac{\sigma^2}{n}\right)$, on a aussi $\overline{X}$ est le meilleur estimateur de m, et $\dfrac{\sqrt{n}\left(\overline{X}-m\right)}{\sigma} \sim \mathcal{N}(0,1)$ alors

$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}\left(\overline{X}-m\right)}{\sigma} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha,$$
$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \left(\overline{X}-m\right) \leq u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha,$$
$$\mathbb{P}\left(\overline{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha,$$

donc l'intervalle de confiance est :

$$IC(m) = \left[\overline{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

Où $u_{1-\frac{\alpha}{2}}$ est le fractille d'ordre $\left(1-\frac{\alpha}{2}\right)$ de la loi $\mathcal{N}(0,1)$.

Exemple 2.3.1 *Après des essais antérieurs, on peut supposer que la résistance à l'éclatement d'un certain type de réservoirs est une variable aléatoire suivant une loi normale $\mathcal{N}(m,\sigma^2)$ où $\sigma = 4\ kg/cm^2$. Des essais sur un échantillon de 25 réservoirs donnent une résistance moyenne à l'éclatement égale à 300 kg/cm^2.*

Donc $n = 25, \overline{X} = 300\ kg/cm^2, \sigma = 4\ kg/cm^2$, le niveau de confiance : $1-\alpha = 0.95, u_{1-\frac{\alpha}{2}} = 1.96$.

L'intervalle de confiance

$$\mathbb{P}\left(-1.96 \leq \frac{5\left(\overline{X}-m\right)}{4} \leq 1.96\right) = 0.95,$$

donc

$$\mathbb{P}\left(300 - 1.96\frac{4}{5} \leq m \leq 300 + 1.96\frac{4}{5}\right) = \mathbb{P}(298.432 \leq m \leq 301.568) = 0.95,$$

$$IC\,(m) = [298.432, 301.568]\,.$$

l'intervalle $[298.432, 301.568]$ *a une probabilité égale à* 0.95 *de contenir la vraie valeur de la résistance à l'éclatement de ce type de réservoirs.*

Remarque 2.3.4 *Si* $n \geq 30$ *et* σ *est inconnu en remplace* σ *par* S' *telles que* $S' = \sqrt{\frac{n}{n-1}}S$.

Exemple 2.3.2 *Les mesures des diamètre d'un échantillon aléatoire de* 100 *billes présentent une moyenne de* 0.80 *cm et d'ecart type de* 0.04 *cm. Evaluer les limites de confiance à :* 99%*, pour le diamètre moyen des billes ?*
On a $n = 100 > 30$ *alors les limites de confiance à* 99% *sont :*

$$\mathbb{P}\left(\overline{X} - \frac{S'}{\sqrt{n}}u_{1-\frac{\alpha}{2}} \leq m \leq \overline{X} + \frac{S'}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right) = 1-\alpha,$$

où $1-\alpha = 0.99, u_{1-\frac{\alpha}{2}} = 2.58, \overline{X} = 0.80$ *cm*$, S = 0.04$ *cm et* $S' = \sqrt{\frac{n}{n-1}}S = 0.04$*, donc*

$$\mathbb{P}\left(0.80 - \frac{0.04}{\sqrt{100}}2.58 \leq m \leq 0.80 + \frac{0.04}{\sqrt{100}}2.58\right) = \mathbb{P}\,(0.79 \leq m \leq 0.81) = 0.99,$$

$$IC\,(m) = [0.79, 0.81]\,.$$

Si la variance est inconnue $(n < 30)$

On utilise le fait que $T = \dfrac{\overline{X} - m}{S}\sqrt{n-1} \sim \mathcal{T}_{n-1}$

$$\mathbb{P}\left(-t_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - m}{S}\sqrt{n-1} \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$
$$\mathbb{P}\left(-\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}} \leq \overline{X} - m \leq \frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$
$$\mathbb{P}\left(\overline{X} - \frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}} \leq m \leq \overline{X} + \frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

donc l'intervalle de confiance est :

$$IC\,(m) = \left[\overline{X} - t_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}}, \overline{X} + t_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}}\right].$$

Où $t_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $\left(1 - \frac{\alpha}{2}\right)$ de la loi student à $(n-1)$ degré de liberté.

Exemple 2.3.3 *Sur un échantillon de taille $n = 20$ durées de vie d'un certain modèle de lampe on a obtenu comme moments empiriques $\overline{X} = 2000\ h$ et $S = 300\ h$. L'intervalle de confiance de niveau 0.95 pour la durée de vie moyenne m est donc*

$$\mathbb{P}\left(2000 - t_{1-\frac{\alpha}{2}}\frac{300}{\sqrt{19}} \leq m \leq 2000 + t_{1-\frac{\alpha}{2}}\frac{300}{\sqrt{19}}\right) = 0.95,$$

où $t_{1-\frac{\alpha}{2}} = 2.093$ d'où l'intervalle de confiance de m au niveau de confiance 0.95 est :

$$\mathbb{P}\,(1855.95 \leq m \leq 2144.05) = 0.95,$$

donc

$$IC\,(m) = [1855.95, 2144.05]\,.$$

Si σ^2 est connue, soit $\sigma = 300$ alors l'intervalle de confiance est définie par :

$$\mathbb{P}\left(2000 - u_{1-\frac{\alpha}{2}}\frac{300}{\sqrt{20}} \leq m \leq 2000 + u_{1-\frac{\alpha}{2}}\frac{300}{\sqrt{20}}\right) = 0.95,$$

avec $u_{1-\frac{\alpha}{2}} = 1.96$ *donc l'intervalle de confiance de* m *au niveau de confiance* 0.95 *est :*

$$\mathbb{P}\left(1868.52 \leq m \leq 2131.48\right) = 0.95,$$

$$IC\left(m\right) = \left[1868.52, 2131.48\right].$$

L'intervalle obtenu est plus grand ; la connaissance du paramètre σ *conduit logiquement à un intervalle plus précis.*

Estimation par intervalle de confiance de la variance

Si la moyenne est connue

$T = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - m\right)^2$ est le meilleur estimateur de la variance σ^2et on a $\dfrac{nT}{\sigma^2} \sim \mathcal{X}_n^2$ comme somme de n carrées de $\mathcal{N}\left(0,1\right)$ indépendantes, alors :

$$\mathbb{P}\left(k_1 \leq \frac{nT}{\sigma^2} \leq k_2\right) = 1-\alpha,$$
$$\mathbb{P}\left(\frac{k_1}{nT} \leq \frac{1}{\sigma^2} \leq \frac{k_2}{nT}\right) = 1-\alpha,$$
$$\mathbb{P}\left(\frac{nT}{k_2} \leq \sigma^2 \leq \frac{nT}{k_1}\right) = 1-\alpha,$$

donc l'intervalle de confiance de σ^2 est :

$$IC\left(\sigma^2\right) = \left[\frac{nt}{k_2}, \frac{nt}{k_1}\right],$$

où k_1 est le quantile d'ordre$\frac{\alpha}{2}$ et $k_2$est le quantile d'ordre$\left(1-\frac{\alpha}{2}\right)$ de la loi $\mathcal{X}_n^2$.

Exemple 2.3.4 *Soit* X *une v.a suivant la loi normale* $\mathcal{N}\left(30,\sigma\right)$, *on prélève un échantillon de taille* $n = 25$. *Cherchons un intervalle de confiance de la variance au niveau de confiance* $1-\alpha = 0.95, n = 25$ *alors* $\alpha = 0.05;\ \dfrac{\alpha}{2} = \dfrac{0.05}{2} = 0.025$ *donc*
$k_1 = \mathcal{X}_{0.025}^2 = 13.120; k_2 = \mathcal{X}_{0.975}^2 = 40.644$ *et* $t = 15$

$$\mathbb{P}\left(\frac{25\times 15}{40.644}\leq\sigma^2\leq\frac{25\times 15}{13.120}\right)=\mathbb{P}\left(9.23\leq\sigma^2\leq 28.58\right)=0.95,$$

donc

$$IC\left(\sigma^2\right)=\left[9.23, 28.58\right].$$

Si la moyenne est inconnue

On utilise $S^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ et on sait que $\frac{nS^2}{\sigma^2}\sim\mathcal{X}_{n-1}^2,$

$$\mathbb{P}\left(l_1\leq\frac{nS^2}{\sigma^2}\leq l_2\right)=1-\alpha,$$
$$\mathbb{P}\left(\frac{l_1}{nS^2}\leq\frac{1}{\sigma^2}\leq\frac{l_2}{nS^2}\right)=1-\alpha,$$
$$\mathbb{P}\left(\frac{nS^2}{l_2}\leq\sigma^2\leq\frac{nS^2}{l_1}\right)=1-\alpha,$$

donc l'intervalle de confiance de σ^2 est :

$$IC\left(\sigma^2\right)=\left[\frac{nS^2}{l_2},\frac{nS^2}{l_1}\right].$$

où $l_1$est le quantille d'ordre $\frac{\alpha}{2}$ et l_2 d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{X}_{n-1}^2$.

Exemple 2.3.5 $n=30; S^2=12; 1-\alpha=0.90$ *alors* $l_1=\mathcal{X}_{0.05}^2=17.7$ *et* $l_2=\mathcal{X}_{0.95}^2=42.6$ *donc*

$$8.46\leq\sigma^2\leq 20.33.$$

Remarque 2.3.5 *Si* $n>30$ *donc on a les deux approximations suivantes :*

$$\sqrt{2}\mathcal{X}_p^2-\sqrt{2p-1}\rightarrow\mathcal{N}\left(0,1\right)\ \text{si } p>30$$

approximation de Fisher

et

$$\mathcal{X}_p^2 = p\left(u\sqrt{\frac{2}{9p}} + 1 - \frac{2}{9p}\right)^3$$

approximation de wilson Hilferty

valable même pour les valeurs faible de p.

Exemple 2.3.6 *L'ecart type de la durée du vie d'un échantillon de* 200 *ampoules électriques est égale à* 100 h, *calculons les limites de confiance à* 95% *pour l'écart type de toute la population.*

On a :

$$\frac{200S^2}{\sigma^2} \sim \mathcal{X}_{199}^2$$

$$\mathbb{P}\left(l_1 \leq \frac{nS^2}{\sigma^2} \leq l_2\right) = 0.95,$$

on utilisant l'approximation de Fisher :

$l_1 = \mathcal{X}_{0.025}^2 = \frac{1}{2}\left(-1.96 + \sqrt{2 \times (199) - 1}\right)^2 = 162 \; et \; l_2 = \mathcal{X}_{0.975}^2 = \frac{1}{2}\left(1.96 + \sqrt{2 \times (199) - 1}\right)^2 =$ 239

$$\mathbb{P}\left(\frac{nS^2}{l_2} \leq \sigma^2 \leq \frac{nS^2}{l_1}\right) = 0.95,$$

donc les limites de confiance à 95%*pour l'ecart type sont :*

$$91.2 \leq \sigma \leq 111.3$$

2.3.2 L'intervalle de confiance du paramètre p d'une loi binômiale

Soit X une v.a qui suit la loi $\mathcal{B}(n,p)$, la statistique $\hat{p}=\dfrac{X}{n}$ est le meilleur estimateur sans biais de p. D'après le théorème central limite :

$$\frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \underset{n\to\infty}{\overset{\mathcal{L}}{\rightarrow}} \mathcal{N}(0,1),$$

où $\overset{\mathcal{L}}{\rightarrow}$ est la convergence en loi.

Soit $\alpha\in\,]0,1[$ alors on a :

$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \leq u_{1-\frac{\alpha}{2}}\right)=1-\alpha,$$

la résolution en p peut se faire de façon directe, en posant $u_{1-\frac{\alpha}{2}}=u$:

$$\begin{aligned}\left(\frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}\right)^2 &\leq u^2\\ &\Longleftrightarrow p^2\left(1+\frac{u^2}{n}\right)-2p\left(\hat{p}+\frac{u^2}{2n}\right)+\hat{p}^2\leq 0\\ &\Longleftrightarrow p^2\left(n+u^2\right)-p\left(2n\hat{p}+u^2\right)+n\hat{p}^2\leq 0\end{aligned}$$

$$\Delta=u^2\left[\left(4n\hat{p}-4n\hat{p}^2\right)+u^2\right],$$

alors

$$p_1=\frac{2n\hat{p}+u^2+u\sqrt{4n\hat{p}(1-\hat{p})+u^2}}{2n+2u^2}=\frac{\hat{p}+\dfrac{u^2}{2n}+u\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}+\dfrac{u^2}{4n^2}}}{1+\dfrac{u^2}{n}},$$

$$n \to \infty; \frac{u^2}{4n^2}, \frac{u^2}{2n} \text{ et } \frac{u^2}{n} \text{ tend vers } 0$$

donc

$$p_1 = \hat{p} - u\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$
$$p_2 = \hat{p} + u\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

donc si n est grand, l'intervalle de confiance de p est :

$$IC\,(p) = \left[\hat{p} - u\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}, \hat{p} + u\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right]$$

où $u = u_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $\left(1 - \frac{\alpha}{2}\right)$de la loi $\mathcal{N}\,(0, 1)$.

Exemple 2.3.7 *Un sondage politique réalisé sur* 1000 *personnes, et visant à émettre un pronostic électoral a conduit à :*

$$\textit{candidat } A \;: 510$$
$$\textit{candidat } B : 490$$

$$\hat{p} = 0.51$$
$$\hat{q} = 0.49$$

Les intervalles de confiance à 95% *sont :*

$$IC\,(p) = [0.48, 0.54]$$
$$IC\,(q) = [0.46, 0.52]$$

2.3.3 Intervalle de confiance de la différence des espérances de deux lois normales

On sait que
$$\begin{aligned}
&\overline{X}_1 \to \mathcal{N}\left(m_1, \frac{\sigma_1^2}{n_1}\right) && \sqrt{n_1}\frac{\left(\overline{X}_1 - m_1\right)}{\sigma_1} \to \mathcal{N}(0,1)\\
&\overline{X}_2 \to \mathcal{N}\left(m_2, \frac{\sigma_2^2}{n_2}\right) && \sqrt{n_2}\frac{\left(\overline{X}_2 - m_2\right)}{\sigma_2} \to \mathcal{N}(0,1)\\
&\overline{X}_1 - \overline{X}_2 \to \mathcal{N}\left(m_1 - m_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)
\end{aligned}$$

Ecarts-types connus

$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

$$\mathbb{P}\left(\left(\overline{X}_1 - \overline{X}_2\right) - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq (m_1 - m_2) \leq \left(\overline{X}_1 - \overline{X}_2\right) + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha,$$

donc l'intervalle de confiance de niveau $(1-\alpha)$ est :

$$IC(m_1 - m_2) = \left[\left(\overline{X}_1 - \overline{X}_2\right) - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \left(\overline{X}_1 - \overline{X}_2\right) + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right].$$

Ecarts-types égaux et inconnus

Posons $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

On a :

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

et puisque σ est inconnu en remplace σ^2 par S_p^2. Sachant que :

on a si $n \geq 30$ en remplace σ^2 par S_p^2. et on a :

$$\frac{(n_1-1)\,S_1'^2}{\sigma^2} \sim \mathcal{X}^2_{n_1-1} \text{ et } \frac{(n_2-1)\,S_2'^2}{\sigma^2} \sim \mathcal{X}^2_{n_2-1}$$

posons : $S_p^2 = \dfrac{(n_1-1)\,S_1'^2+(n_2-1)\,S_2'^2}{n_1+n_2-2}$ et on a $S_1'^2 = \frac{1}{n_1-1}\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ et $S_2'^2 = \frac{1}{n_2-1}\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ alors

$$(n_1+n_2-2)\,\frac{S_p^2}{\sigma^2} = \frac{(n_1-1)\,S_1'^2}{\sigma^2}+\frac{(n_2-1)\,S_2'^2}{\sigma^2} \sim \mathcal{X}^2_{n_1+n_2-2}$$

alors

$$\begin{aligned}
\mathbb{V}\left((n_1+n_2-2)\,\frac{S_p^2}{\sigma^2}\right) &= 2\,((n_1+n_2-2)) \\
\mathbb{V}\left(S_p^2\right) &= \frac{2\sigma^4}{n_1+n_2-2} \\
\mathbb{V}\left(S_1'\right) &= \frac{2\sigma^4}{n_1-1} \text{ et } \mathbb{V}\left(S_2'\right) = \frac{2\sigma^4}{n_2-1}
\end{aligned}$$

d'ou

$$\begin{aligned}
\frac{\mathbb{V}\left(S_p^2\right)}{\mathbb{V}\left(S_i'^2\right)} &= \frac{n_i-1}{n_1+n_2-2} \leq 1\,(i=1,2) \\
\mathbb{V}\left(S_p^2\right) &\leq \mathbb{V}\left(S_i'^2\right)\,(i=1,2)
\end{aligned}$$

alors S_p^2 est préférable à $S_1'^2$ et $S_2'^2$ donc

$$\frac{\left(\overline{X}_1-\overline{X}_2\right)-(m_1-m_2)}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \to \mathcal{T}\,(n_1+n_2-2)$$

donc l'intervalle de confiance de niveau $(1-\alpha)$ est :

$$IC\,(m_1-m_2) = \left[\overline{X}_1-\overline{X}_2-t_{1-\frac{\alpha}{2}}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \overline{X}_1-\overline{X}_2+t_{1-\frac{\alpha}{2}}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right],$$

où $t_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $\left(1-\frac{\alpha}{2}\right)$de $\mathcal{T}\left(n_1+n_2-2\right)$.

Ecarts types quelconques

$\sigma_1^2 \neq \sigma_2^2$ inconnu, $n_1 = n_2 = n$ telles que n est grand alors on a :

$$\frac{\sqrt{n}\left(\overline{X}_1-\overline{X}_2-(m_1-m_2)\right)}{\sqrt{S_1'^2+S_2'^2}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

Donc l'intervalle de confiance de niveau $(1-\alpha)$ est :

$$IC\left(m_1-m_2\right)=\left[\overline{X}_1-\overline{X}_2-u_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{S_1'^2+S_2'^2},\overline{X}_1-\overline{X}_2+u_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{S_1'^2+S_2'^2}\right].$$

Application(**simulation**) On génère un échantillon de taille $n = 1000$ suit la loi beta de paramètre $a = 2$ et $b = 4$ respectivement, on appliquant la méthode du maximum de vraisemblance et la méthode des moments pour calculer l'estimateur de a et b, on calcule le biais et l'erreur quadratique moyenne. Pour cela on prend m=50.
Le but de cettte application est de comparer entre deux méthodes d'estimation ponctuelle la méthode du maximum de vraisemblance et la méthode des moments.

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

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

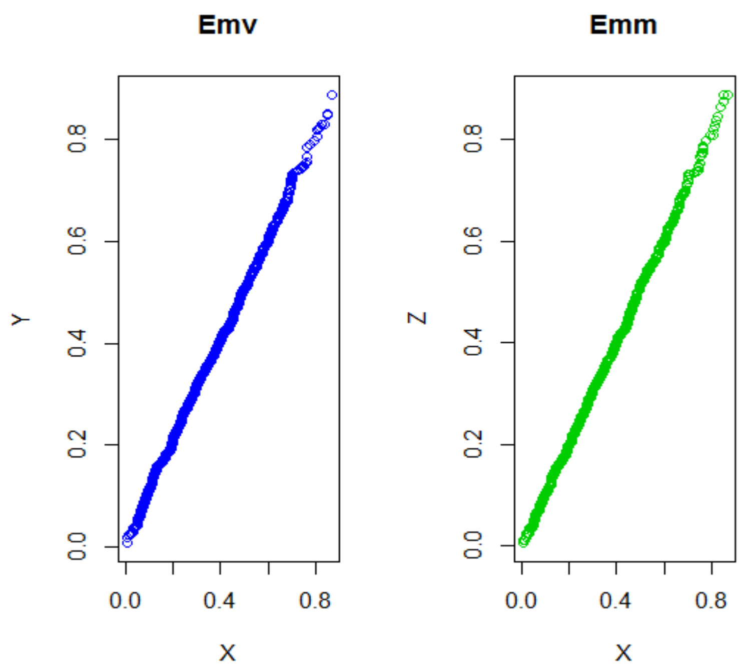

FIG. 2.1 – Comparaison entre l'EMV et l'EMM de loi beta

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

Le but de cettte application est de comparer entre deux méthodes d'estimation ponctuelle la méthode du maximum de vraisemblance et la méthode des moments.

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

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

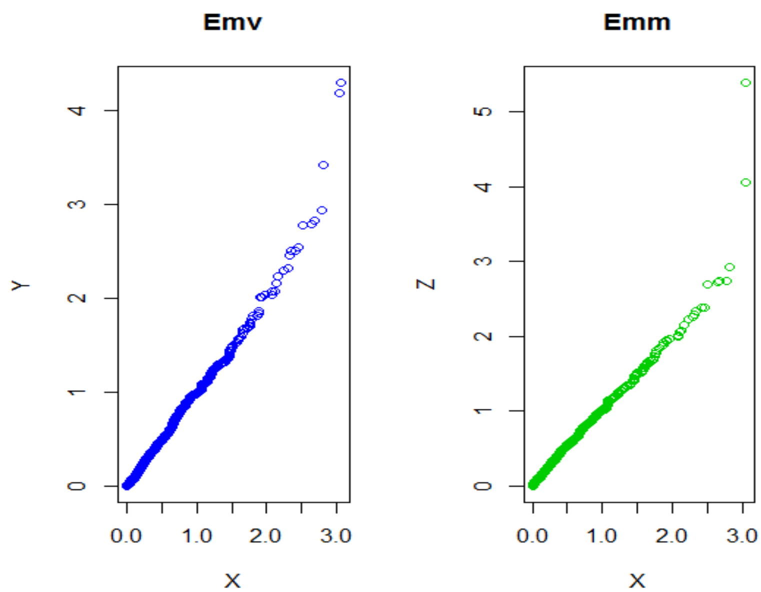

FIG. 2.2 – Comparaison entre l'EMV et l'EMM de loi exponentielle

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

Conclusion

En conclusion, il y a deux types d'estimation proposés : l'estimation ponctuelle et l'estimation par intervalle de confiance. L'estimation ponctuelle donne une seule valeur, il y a le danger qu'on oublie que l'estimateur est une v.a avec une certaine variance. En effet dans l'estimation par intervalle on détermine deux valeurs et on donne la probabilité que le paramètre de la population se trouve à l'intérieur de cet intervalle de confiance, et cette estimation a l'avantage de montrer la précision dans l'estimation du paramètre et ceci facilite la compréhension des résultats si l'on ne connait pas les méthodes statistiques. Les deux estimations sont efficaces chacun dans certain cas, nous pouvons comparer entre les méthodes d'estimation ponctuelle.

La méthode du maximum de vraisemblance et celle des moments peuvent donner des estimateurs égaux, mais la où ces deux sont différents, l'estimateur du maximum de vraisemblance est généralement meilleur car il est asymptotiquement sans biais, et sa distribution asymptotique est normale.

Finalement, le meilleur estimateur est l'estimateur sans biais à variance minimale.

Bibliographie

[1] Grammont, L. (2003). Cours de statistiques inferentielles.

[2] Lejeune,M. (2010). Statistique la théorie et ses applications. Springer-Verlag France, Paris.

[3] Pierre, D. (2017). Cours de statistique inférentielles. Licence 2-S4.SI-Mass.

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

[5] Sayah, A. (2000). Cours PS2. Université mohamed kheider Biskra, biskra.

[6] Tassi, P. (1989). Méthodes statistiques. Economica, Paris.

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

[8] Ycart, B. (2002). Estimation paramétrique tests statistique. Centre de publication universitaire, Tunis.

Annexe A : Logiciel *R*

2.4 Qu'est-ce-que le langage R ?

• Le langage **R** est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

• **R** a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Developement Core Team.

L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe B : Abréviations et Notations

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

v.a	:	Variable aléatoire.
i.i.d	:	Indépendantes identiquement distribuées.
c-à-d	:	C'est à dire.
$\mathbb{E}(X)$	:	Espérance mathématique ou moyenne du v.a X.
$\mathbb{V}(X)$	:	Variance du v.a X.
$B(n,\theta)$	:	Le biais de l'estimateur T_n.
$cov(X_1, X_2)$	:	La covariance entre deux v.a X_1 et X_2.
$\xrightarrow{p}$	:	Convergence en probabilité .
$\xrightarrow{p.s}$	:	Convergence presque sûre.
$\xrightarrow{\mathcal{L}}$	:	Convergence en loi.
$\mathbb{EQM}(T_n, \theta)$	:	L'erreur quadratique moyenne de T_n par rapport à θ.
$I_n(\theta)$	:	Information de Fisher.
EMV	:	Estimateur du maximum de vraisemblance.
EMM	:	Estimateur des moments.
IC	:	Intervalle de confiance.