Chapitre 1

Rappels et compléments de probabilités

Chapitre 2

Exhaustivité, complétion, information de Fischer et estimation sans biais

Chapitre 3

Estimation ponctuelle

En language commun, estimer une quantité c'est en donner une valeur approchée. En statistique les quantités d'intêret sont liées à la loi de probabilité d'une variable aléatoire observable et peuvent être le (ou les) paramètre(s) de cette loi ou des fonctions de ce paramètre (moyenne, variance etc.). On supposera que des observations de cette variable sont disponibles sur la base desquelles sera faite l'inférence statistique.

3.1 Estimation ponctuelle : généralités

3.1.1 Définition

Soit $(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta\in\Theta})$ un modèle statistique, on veut estimer θ ou une fonction $g(\theta)$ où g est une fonction définie sur Θ.

Définition 3.1. On appelle estimateur de $g(\theta)$ toute statistique à valeurs dans $g(\Theta)$:

$$\begin{aligned} T \quad : \quad \mathfrak{X} &\longrightarrow g(\Theta) \\ x &\longrightarrow T(x) \end{aligned}$$

$T(x)$ est dite estimation de $g(\theta)$

Définition 3.2. Un estimateur T de $g(\theta)$ est dit sans biais si

$$E_\theta(T) = g(\theta).$$

La différence $E_\theta(T) - g(\theta)$ est dite biais de T.

Un estimateur est donc sans biais si en moyenne, il est égal à la quantité qu'il estime.

Définition 3.3. Une suite d' estimateurs $(T_n)_{n\in N}$ de $g(\theta)$ est dite asymptotiquement sans biais si

$$\lim_{n\to+\infty} E_\theta(T_n) = g(\theta)$$

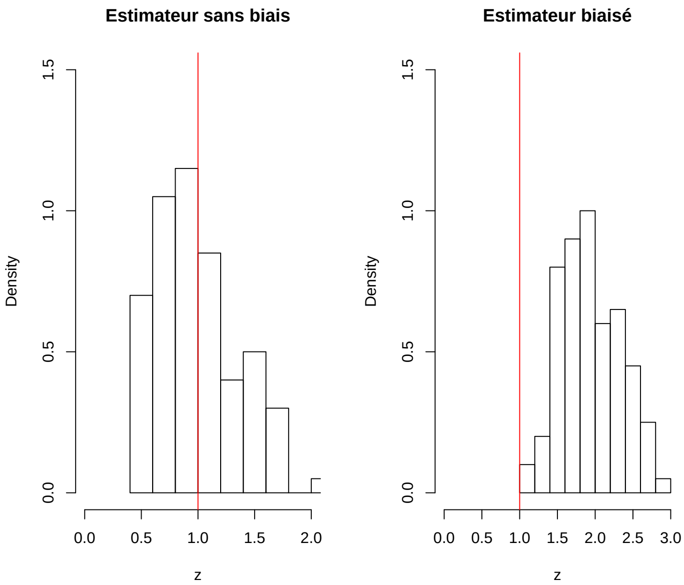

FIGURE 3.1 – Estimateurs sans biais et avec biais

Définition 3.4. Une suite d' estimateurs $(T_n)_{n\in N}$ de $g(\theta)$ est dite convergente en probabilité (ou consistente) si

$$T_n \underset{n\longrightarrow+\infty}{\overset{\Pr oba}{\longrightarrow}} g(\theta).$$

Elle sera dite convergente presque sûrement (ou fortement consistente) si

$$T_n \underset{n\longrightarrow+\infty}{\overset{P.S}{\longrightarrow}} g(\theta).$$

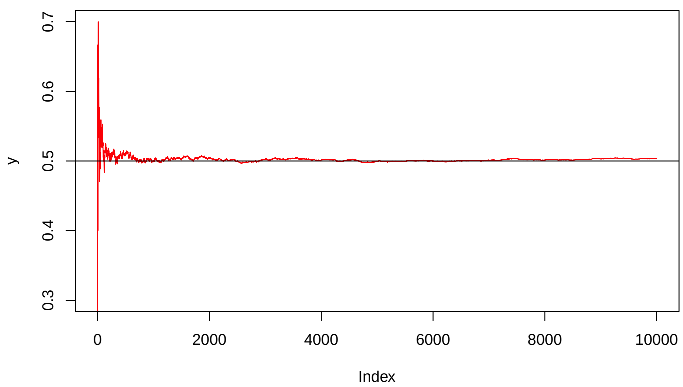

FIGURE 3.2 – Estimateur convergent

3.1.2 Quelques problèmes d'estimation

Soit $X_1, X_2, .., X_n)$ un n-échantillon d'une v.a X, les moments de cette variable et les probabilités du type $P(X \in A)$ peuvent être estimées de manière simple, et dans un cadre général (i.e sans faire aucune hypothèse sur la loi de X). Les estimateurs obtenus possèdent de bonnes propriétés.

Théorème 3.1. *Soit $X_1, X_2, .., X_n)$ un n-échantillon d'une v.a X, admettant un moment d'ordre k, alors le moment emirique d'ordre k $m_k = \frac{\sum_{i=1}^{n} X_i^k}{n}$ est un estimateur sans biais et convergent presque sûrement de $E(X^k)$.*

Démonstration. Application immédiate de la loi forte des grands nombres (voir le chapitre 1) □

Corollaire 3.2. *$\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$ est un estimateur sans biais et convergent presque sûrement de $E(X)$.*

Corollaire 3.3. *$S_n'^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X_n})^2}{n-1}$ est un estimateur sans biais et convergent presque sûrement de $Var(X)$.*

Démonstration. On a

$$S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X_n})^2}{n} = \frac{\sum_{i=1}^{n} {X_i}^2}{n} - \overline{X_n}^2$$

$$E(S_n^2) = E(\frac{\sum_{i=1}^{n} {X_i}^2}{n}) - E(\overline{X_n}^2) = E({X_i}^2) - E(\overline{X_n}^2)$$

d'autre part

$$\begin{aligned} E(\overline{X_n}^2) &= E((\frac{\sum_{i=1}^{n} X_i}{n})^2) = E(\frac{\sum_{i=1}^{n} {X_i}^2}{n^2}) - E(\frac{\sum\sum_{i \neq j} X_i X_j}{n^2}) \\ &= \frac{E({X_i}^2)}{n} - \frac{n(n-1)}{n^2} E(X_i X_j) \\ &= \frac{E({X_i}^2)}{n} - \frac{(n-1)}{n} E(X_i)^2 \end{aligned}$$

on déduit

$$E(S_n^2) = \frac{(n-1)}{n}(E({X_i}^2) - E(X_i)^2) = \frac{(n-1)}{n} Var(X)$$

et

$$E(S_n'^2) = E(\frac{n}{n-1}S_n^2) = Var(X)$$

D'autre part :

$$S_n^2 = \frac{\sum\limits_{i=1}^{n} {X_i}^2}{n} - \overline{X_n}^2 \underset{n \longrightarrow +\infty}{\overset{P.S}{\longrightarrow}} E(X^2) - E(X)^2 = Var(X)$$

et

$$S_n'^2 = \frac{n}{n-1}S_n^2 \underset{n \longrightarrow +\infty}{\overset{P.S}{\longrightarrow}} Var(X)$$

d'après la loi forte des grands nombres.

□

Corollaire 3.4. $S_{X,Y}'^2 = \frac{\sum\limits_{i=1}^{n} (X_i - \overline{X_n})(Y_i - \overline{Y_n})}{n-1}$ *est un estimateur sans biais et convergent presque sûrement de* $Cov(X,Y)$.

Démonstration identique que pour le corollaire 16.

Corollaire 3.5. *Soit A un évènement, lié à une espérience aléatoire et soit* $F_n(A) = \frac{\sum\limits_{i=1}^{n} 1_{\{X_i \in A\}}}{n}$ *la fréquence de l'évènement* $\{X \in A\}$ *(le nombre de réalisations de* $\{X \in A\}$*, lors de n répétitions de cet expérience dans des conditions indépendantes).* $F_n(A)$ *est un estimateur sans biais et convergent presque sûrement de* $P(X \in A)$.

Démonstration. $\forall i$, $1_{\{X_i \in A\}} \sim B(P(A))$ et d'après le corollaire 15, $\frac{\sum\limits_{i=1}^{n} 1_{\{X_i \in A\}}}{n}$ est un estimateur sans biais et convergent de $P(A)$.

□

Définition 3.5. Soit $(x_1, x_2, ..., x_n)$ un n-échantillon d'observations de X, on appelle fonction de répartition empirique de X la fonction F_n de R dans [0,1] définie par :

$$F_n(x) = \frac{\sum\limits_{i=1}^{n} 1_{\{x_i \leq x\}}}{n}$$

Corollaire 3.6. $F_n(x) = \frac{\sum\limits_{i=1}^{n} 1_{\{x_i \leq x\}}}{n}$ *est un estimateur sans biais et convergent de* $F(x)$*, la fonction de répartition de X*

Démonstration. $F(x) = P(X < x)$, le corollaire 18 permet de déduire le résultat.

□

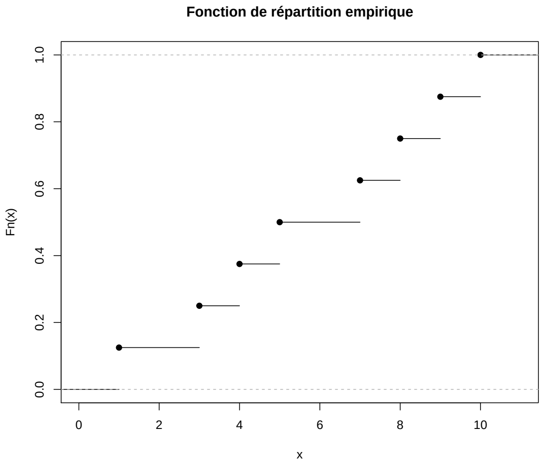

FIGURE 3.3 – Fonction de répartition empirique

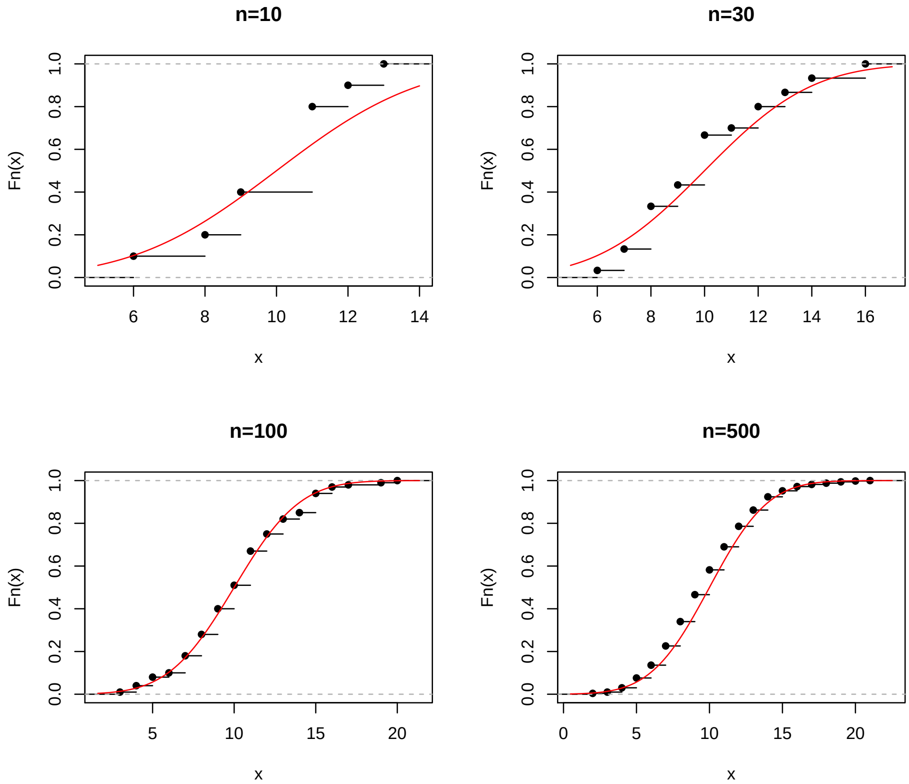

FIGURE 3.4 – Evolution des f.r empiriques vers la f.r théorique

Pour certaines quantités, moments, variances, covariances, probabilités il a été vu qu'il existe des estimateurs simples ayant de bonnes qualités : sans biais et convergents. Pour des paramètres quelconques, il est nécessaire d'avoir des méthodes d'estimation conduisant à des estimateurs de bonne qualité. Il existe plusieurs grandes méthodes d'estimation

1. Méthode des moments
2. Méthode du maximum de vraisemblance
3. Recherche des estimateurs sans biais de variance minimale

Il existe d'autres méthodes importantes comme l'estimation par les moindres carrés qui sera utilisée dans le contexte de la régression linéaire et l'estimation bayésienne qui ne sera pas abordée ici.

3.2 Méthode des moments

Les moments empiriques sont des estimateurs sans biais et convergents des moments théoriques. La méthode des moments consiste à égaler les uns aux autres. Soit le modèle d'échantillon $(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta\in\Theta})^{\prime n)}$ associé à n observations $(x_1, x_2, ..., x_n)$ d,'une v.a de loi P_θ.

On suppose $\theta = (\theta_1, \theta_2, ..., \theta_p)$. On veut estimer θ i.e p paramètres. On obtient un système de p équations à p inconnues en écrivant :

$$E(X) = \frac{\sum\limits_{i=1}^{n} X_i}{n} \ (1)$$

$$E(X^2) = \frac{\sum\limits_{i=1}^{n} X_i^2}{n} \ (2)$$

....

....

$$E(X^p) = \frac{\sum\limits_{i=1}^{n} X_i^p}{n} (p)$$

La résolution de ce système permet d(obtenir les estimateurs $\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_p$.

Exemple 3.1. 1. Soit le modèle d'échantillon poissonien $(N, \mathcal{P}(N), \{P(\lambda), \lambda \in R_+^*\})^{\prime n)}$ associé à n observations $(x_1, x_2, ..., x_n)$ d,'une v.a de loi $P(\lambda)$.

Estimer λ par la méthode des moments.

$$E(X) = \lambda = \overline{x} \implies \widehat{\lambda} = \overline{x}$$

2. Soit le modèle d'échantillon gaussien $(R, \mathcal{P}(R), \{N(m, \sigma^2), m \in R\})^{\prime n)}$. On suppose σ^2 connu.

   Estimer m.

$$E(X) = m = \overline{x} \implies \widehat{m} = \overline{x}$$

3. Soit le modèle d'échantillon gaussien $(R, \mathcal{P}(R), \{N(m, \sigma^2), m \in R\})^{\prime n)}$

   Estimer m et σ^2

$$\begin{aligned} E(X) &= m = \overline{x} \\ E(X^2) &= \sigma^2 + m^2 = m_2 \\ \widehat{m} &= \overline{x} \text{ et } \widehat{\sigma^2} = m_2 - \overline{x}^2 = S_n^2 \end{aligned}$$

La méthode des moments, outre sa simplicité, donne des estimateurs, qui sous certaines hypothèses, ont de bonnes propriètés asymptôtiques.

3.3 Estimateurs du maximum de vraisemblance

Considérons une pièce de monnaie pour laquelle la probabilité d'obtenir pile est inconnue. Cette pièce est lancée 10 fois et on obtient 6 piles. Une méthode d'estimation intuitive est de considérer que l'évènement réalisée, ici 6 piles et 4 faces, a une forte probabilité (puiqu'il s'est réalisé contrairemant aux autres).
On estime alors p par la valeur $\widehat{p}$ de [0,1] qui attribue à cet évènement, la plus forte probabité, ou en d'autres termes par la valeur $\widehat{p}$ qui maximise la vraisemblance.

On a

$$L(p, \widetilde{x}) = p^6(1-p)^4$$

$$\frac{\partial}{\partial p} L(p, \widetilde{x}) = 0 \iff \frac{\partial}{\partial p} \log L(p, \widetilde{x}) = 0$$

ce qui donne

$$\widehat{p} = \frac{6}{10}$$

Définition 3.6. Soit $\widetilde{x} = (x_1, x_2, ..., x_n)$ un n-échantillon d'une v.a X de loi $(f_\theta)_{\theta \in \Theta}$. On appelle estimateur du maximum de vraisemblance de θ, s'il existe, la statistique $\widehat{\theta}$ telle que

$$L(\widehat{\theta}, \widetilde{x}) = \max_{\theta \in \Theta} L(\theta, \widetilde{x})$$

La maximisation de la fonction de vraisemblance peut se faire de différentes manières

Cas 3.7. *θ réel et $L(\theta, \widetilde{x})$ dérivable par rapport à θ*
$\widehat{\theta}$ est alors solution du système d'équations dit équations du maximum de vraisemblance

$$\begin{cases} \dfrac{\partial}{\partial \theta} L(\theta, X) = 0 & \textit{: condition du premir ordre} \\ \dfrac{\partial^2}{\partial \theta^2} L(\theta, X) < 0 & \textit{: condition du deuxième ordre} \end{cases}$$

ou de manière équivalente :

$$\begin{cases} \dfrac{\partial}{\partial \theta} \log L(\theta, X) = 0 & \textit{: condition du premir ordre} \\ \dfrac{\partial^2}{\partial \theta^2} \log L(\theta, X) < 0 & \textit{: condition du deuxième ordre} \end{cases}$$

Car la fonction log *est continue et strictement monotone.*

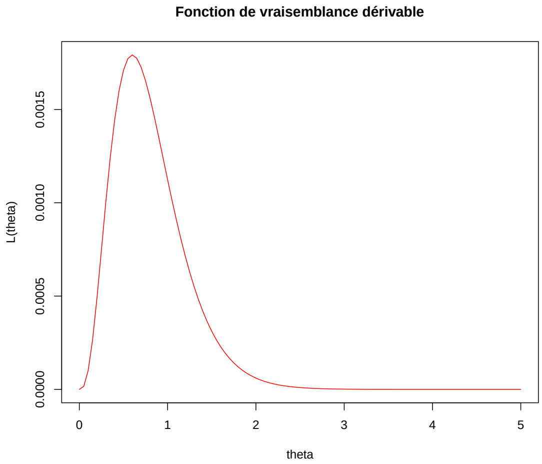

FIGURE 3.5 – Fonction de vraisemblance

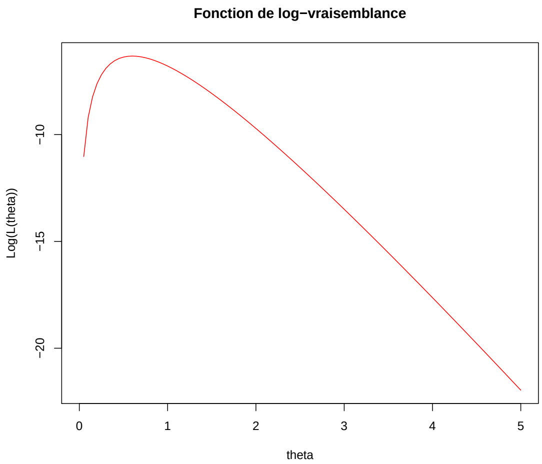

FIGURE 3.6 – Fonction de log-vraisemblance

Exemple 3.2. Estimation du paramètre λ d'une loi de Poisson .

$$L(\lambda, \widetilde{x}) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!}$$

$$l(\lambda, \widetilde{x}) = \log L(\lambda, \widetilde{x}) = -n\lambda + (\sum_{i=1}^{n} x_i) \log \lambda - \log(\prod_{i=1}^{n} x_i!)$$

$$\frac{\partial}{\partial \lambda} \log L(\lambda, \widetilde{x}) = -n + \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} = 0$$

$$\widehat{\lambda} = \overline{x}$$

De plus on a :

$$\frac{\partial^2}{\partial \lambda^2} \log L(\widehat{\lambda}, \widetilde{x}) = -\frac{\sum\limits_{i=1}^{n} x_i}{\overline{x}^2} < 0$$

On conclut que $\widehat{\lambda} = \overline{x}$ est l'EMV de λ.

Cas 3.8. *: θ réel et $L(\theta, \widetilde{x})$ non dérivable par rapport à θ*
Il faut maximiser la vraisemblance directement

Exemple 3.3. Estimation du paramètre θ d'une loi $\mathcal{U}_{[0,\theta]}$.

La fonction de vraisemblance $L(m, \sigma^2, \widetilde{x}) = \dfrac{1}{\theta^n} 1_{[0,\theta]}(\max(x_i))$ n'est pas dérivable en θ, car discontinue au point $\theta = \max(x_i)$ (voir Figure). On voit que le maximum est atteint au point $\widehat{\theta} = \max(x_i)$.

Cas 3.9. *: $\theta = (\theta_1, \theta_2, .., \theta_p)$ vectoriel et $L(\theta, \widetilde{x})$ dérivable par rapport à θ*
Les conditions du premier ordre s'écrivent, si on note par ∇ l'opérateur gradient :

$$\nabla L(\theta, X) = 0 \iff \nabla \log L(\theta, X) = 0$$

c'est à dire :

$$\left\{\begin{array}{l} \frac{\partial}{\partial \theta_1} L(\theta, X) = 0 \\ \frac{\partial}{\partial \theta_2} L(\theta, X) = 0 \\ . \\ . \\ \frac{\partial}{\partial \theta_p} L(\theta, X) = 0 \end{array}\right. \iff \left\{\begin{array}{l} \frac{\partial}{\partial \theta_1} \log L(\theta, X) = 0 \\ \frac{\partial}{\partial \theta_2} \log L(\theta, X) = 0 \\ . \\ . \\ \frac{\partial}{\partial \theta_p} \log L(\theta, X) = 0 \end{array}\right.$$

La condition du deuxième ordre s'exprime à l'aide de la matrice hessienne qui doit être définie négative.

Exemple 3.4. Estimation des paramètres m et σ^2 d'une loi normale.

On a

$$L(m, \sigma^2, \widetilde{x}) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i - m)^2}{\sigma^2}} \right) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2}\sum\limits_{i=1}^{n}\frac{(x_i - m)^2}{\sigma^2}}$$

$$\log L(m, \sigma^2, \widetilde{x}) = \log \left(\frac{1}{\sqrt{2\pi}} \right)^n - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - m)^2}{\sigma^2}$$

Les EMV de sont solution du système d'équations :

$$\frac{\partial}{\partial m} \log L(m, \sigma^2, \widetilde{x}) = \sum_{i=1}^{n} \frac{(x_i - m)}{\sigma^2} = 0$$

$$\frac{\partial}{\partial\sigma^2}\log L(m,\sigma^2,\widetilde{x}) = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-m)^2}{\sigma 4} = 0$$

ce qui donne

$$\widehat{m} = \overline{x} \ \text{ et } \widehat{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-\overline{x}\ )^2}{n} = S^2$$

La vérification de la condition du second ordre est omise.

Remarque 1. Les équations du maximum de vraisemblance ne sont pas toujours résolubles analytiquement, et on doit alors le faire numériquement pour avoir des valeurs approchées des solutions. C'est le cas par exemple pour les paramètres de la loi gamma ou pour ceux de la loi de Weibull.

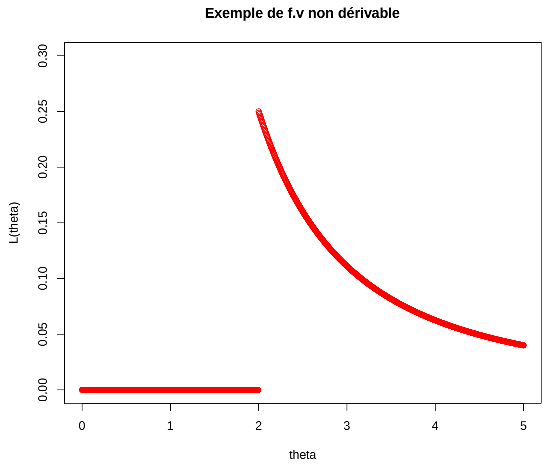

FIGURE 3.7 – Fonction de vraisemblance

Théorème 3.10. *Soit $\widetilde{x} = (x_1, x_2, ..., x_n)$ un n-échantillon d'une v.a X de loi $(f_\theta)_{\theta\in\Theta}$. S'il existe une statistique exhaustive T pour θ alors l'estimateur du maximum de vraisemblance de θ est fonction de T.*

Conséquence du théorème de factorème de factorisation

Théorème 3.11 (Invariance fonctionnelle)**.** *Soit $\widehat{\theta}$ l'estimateur du maximum de vraisemblance de θ alors l'estimateur du maximum de vraisemblance de $g(\theta)$ est $g(\widehat{\theta})$ pour toute fonction g.*

3.4 Proprétés asymptôtiques des EMV

Les estimateurs du maximum de vraisemblance n'ont pas de propriétés connus pour des échantillons de taille finie (sauf la propriété d'invariance fonctionnelle).

Cependant ils ont de trés bonnes propriètés asymptotiques.

Théorème 3.12. *Soit $\widehat{\theta_n}$ l'EMV de θ basé sur un n-échantillon et $I_n(\theta)$ l'information de Fischer ramenée par les observations sur θ. (On suppose las conditions de régularité de Fischer vérifiées). On a*

$$\frac{\widehat{\theta_n} - \theta}{\sqrt{\dfrac{1}{I_n(\theta)}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

Théorème 3.13. *Soit g une fonction dérivable et $g(\widehat{\theta})$ l'EMV de $g(\theta)$ On a sous les hypothèses du théorème précédent*

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\dfrac{g'(\theta)^2}{I_n(\theta)}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

Les estimateurs du maximum de vraisemblance sont donc asymptotiquement gaussiens, asymptotiquement sans biais et asymptotiquement efficaces.

On a si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\theta)^2}{I_n(\theta)}\right)$$

Pour les applications le fait que la variance soit inconnue (dépend de θ) est génant. On a cependant les théorèmes suivants obtenues en remplaçant la variance par la variance estimée.

Théorème 3.14. *On a sous certaines hypothèses de régularité*

$$\frac{\widehat{\theta_n} - \theta}{\sqrt{\dfrac{1}{I_n(\widehat{\theta_n})}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

Théorème 3.15. *On a sous certaines hypothèses de régularité*

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\dfrac{g'(\widehat{\theta_n})^2}{I_n(\widehat{\theta_n})}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

On peut alors écrire si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\widehat{\theta}_n)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\widehat{\theta}_n)^2}{I_n(\widehat{\theta}_n)}\right)$$

Les théorèmes 3.14 et 3.15 sont trés utiles pour la construction d'intervalles de confiance asymptôtiques (voir le Chapitre 5)

3.5 Estimateurs sans biais de variance minimale

Les deux méthodes d'estimation précédentes sont des méthodes basées, soit sur le bon sens (EMV), soit sur des proprétés probabilistes comme la loi des grands nombres (méthode des moments). Les estimateurs obtenus n'ont aucune propriété d'optimalité pour un échantillon de taille finie, ils ont cependant de bonnes propriètés asymptôtiques. La méthode suivante compare les estimateurs, selon un critère précis, l'erreur commise en estimant $g(\theta)$ et plus présisément, en comparant le carré des erreurs. Le meilleur estimateur étant celui qui conduit à la plus petite erreur.

Définition 3.7. Soit T un estimateur de $g(\theta)$, on appelle risque quadratique de l'estimateur T, l'erreur quadratique moyenne de T, i.e

$$R_\theta(T) = E_\theta((T(X) - g(\theta))^2)$$

Proposition 3.16. *On a*

$$R_\theta(T) = Var(T) + biais(T)^2$$

Démonstration. On a

$$\begin{aligned} R_\theta(T) &= E_\theta\left((T(X) - E_\theta(T(X) + E_\theta(T(X) - g(\theta)^2\right) \\ &= E_\theta((T(X) - E_\theta(T(X)))^2 + (E_\theta(T(X)) - g(\theta))^2 \\ &= Var_\theta(T) + biais(T)^2 \end{aligned}$$

Le passage de la première ligne à la deuxième se faisant en développant le carré et en remarquant que le terme correspondant au double produit $2E_\theta\left((T(X) - E_\theta(T(X))(E_\theta(T(X) - g(\theta))\right)$ est nul. □

Définition 3.8. Un estimateur T_1 est dit meilleur qu'un autre estimateur T_2 au sens du risque quadratique si

$$R_\theta(T_1) \leq R_\theta(T_2) \qquad \forall\theta \in \Theta$$

(uniformément en θ)

Remarque 2. Parmi tous les estimateurs sans biais le meilleur au sens du risque quadratique est celui qui a la plus petite variance. T_1 sera optimal si

$$Var_\theta(T_1) \leq Var_\theta(T_2) \qquad \forall\theta \in \Theta$$

pour tout T_2 autre estimateur sans biais.

Si un tel estimateur existe, il sera dit estimateur sans biais de variance minimale (ou uniformément minimale) : ESBVUM en abrégé.

Exemple 3.5. $\widetilde{x} = (x_1, x_2, ..., x_n)$ n-échantillon de X$\sim \mathcal{U}_{[0,\theta]}$ de densité $f_\theta(x) = \frac{1}{\theta}1_{[0,\theta]}(x)$

On a $E(\overline{X}) = \frac{\theta}{2}$ et donc $T_1(\widetilde{X}) = 2\overline{X}$ est un estimateur sans biais de θ.

On peut aussi vérifier que $T_2(\widetilde{X}) = \frac{n+1}{n} \max\limits_{1\leq i\leq n}(X_i)$ est aussi un estimateur sans biais de θ.

En effet

$$f_{\max\limits_{1\leq i\leq n}(X_i)}(x) = \frac{nx^{n-1}}{\theta^n}1_{[0,\theta]}(x) \implies E(T_2(\widetilde{X})) = \frac{n}{n+1}\theta$$

On a :

$$Var(T_1) = Var(2\overline{X_n}) = \frac{4Var(X)}{n} = \frac{\theta^2}{3n}$$

et

$$Var(T_2) = \left(\frac{n+1}{n}\right)^2 E(S^2) - \theta^2 = \left(\frac{n+1}{n}\right)^2 E(S^2) - \theta^2 = \frac{(n+1)^{\ 2}}{n(n+2)}\theta^2 - \theta^2 = \frac{\theta^2}{n(n+2)}$$

On conclut que $Var(T_2) < Var(T_1)$ dés que $n > 1$, et donc que T_2 est un meilleur estimateur sans biais de θ que T_1 au sens du risque quadratique.

Théorème 3.17 (Théorème de Rao-Blackwell)**.** *Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive pour θ. Alors*
1) $E_\theta(T/S)$ est un estimateur sans biais de $g(\theta)$
2) $Var_\theta(E_\theta(T/S)) \leq Var_\theta(T)$

Ce théorème est donc un théorème d'amélioration des estimateurs sans biais. $E_\theta(T/S)$ est un meilleur estimateur sans biais de $g(\theta)$ que T.

Démonstration. 1) $E_\theta(T/S)$ est un estimateur de $g(\theta)$ car ne dépend pas de θ.

2) $E_\theta(E_\theta(T/S)) = E_\theta(T) = g(\theta)$ (Théorème de l'espérance conditionnelle)

3) $Var_\theta(T) = Var_\theta(E_\theta(T/S)) + E_\theta(Var_\theta(T/S))$ (Théorème de la variance conditionnelle)

ce qui implique que

$$Var_\theta(E_\theta(T/S)) = Var_\theta(T) - E_\theta(Var_\theta(T/S)) \leq Var_\theta(T)$$

□

Cas particulier : Si T est une fonction S, i.e si $T = h(S)$ alors $E_\theta(h(S)/S) = h(S) = T$
Dans ce cas T est un estimateur inaméliorable.

Exemple 3.6. Dans l'exemple de l'estimation du paramètre d'une loi de Poisson, $T(\widetilde{x}) = \overline{x}$ est un estimateur sans biais de λ. Comme $S(\widetilde{x}) = \sum\limits_{i=1}^n x_i$ est une statistique exhaustive pour λ, $E\left(\overline{X}/\sum\limits_{i=1}^n X_i\right) = \overline{X}$ est un ESBVUM de λ.

Théorème 3.18 (Théorème de Lehmann-Scheffé). *Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive et complète pour θ. Alors $E_\theta(T/S)$ est l'unique estimateur sans biais de variance minimale de $g(\theta)$.*

Démonstration. Supposons qu'il y a deux estimateurs sans biais de variance mimimale T_1 et T_2. Ils sont forcément fonctions de S, car sinon $E_\theta(T_1/S)$ et $E_\theta(T_1/S)$ sont meilleurs (Rhéorème de Rao-Blackwell). Posons $T_1 = h_1(S)$ et $T_2 = h_2(S)$. On a

$$E_\theta(T_1 - T_2) = E_\theta(h_1(S) - h_2(S)) = 0 \implies h_1 - h_2 = 0 \quad P_\theta - p.s.$$

□

Exemple 3.7 (Estimation du paramètre d'une loi de Poisson (suite)). Il a été montré que $T(\widetilde{X}) = \overline{X}$, est un ESBVUM de λ. Comme la statistique exhaustive $S(\widetilde{X}) = \sum\limits_{i=1}^{n} X_i$ est aussi complète, $T(\widetilde{X}) = \overline{X}$ est l'unique ESBVUM de λ.

Exemple 3.8 (Estimation du paramètre d'une loi uniforme $\mathcal{U}_{[0,\theta]}$ (suite)). La statistique $\max\limits_{1\leq i\leq n}(X_i)$ est exhaustive (théorème de factorisation), on peut aussi vérifier qu'elle est complète. $T_2(\widetilde{X}) = \dfrac{n+1}{n} \max\limits_{1\leq i\leq n}(X_i)$ est donc l'unique ESBVUM de θ.

Théorème 3.19 (Théorème de Rao-Cramer). *Soit T un estimateur sans biais de $g(\theta)$. On suppose que g est dérivable et que les conditions de régularité de Fischer sont vérifiées, alors*

$$Var_\theta(T) \geq \frac{(g'(\theta))^2}{I(\theta)}$$

La quantité $\dfrac{(g'(\theta))^2}{I(\theta)}$ est dite borne de Rao-Cramer (BRC).

Démonstration. On rappelle que

$$|\rho_{X,Y}| = \left|\frac{Cov(X,Y)}{\sigma_X \sigma_Y}\right| \leq 1 \iff Cov(X,Y)^2 \leq \sigma_X^2 \sigma_Y^2$$

On a

$$E_\theta(T(X)) = \int_{\mathcal{X}} T(x) f_\theta(x) dx = \int_{\mathcal{X}} T(x) L(\theta, x) dx = g(\theta)$$

$$\begin{aligned}
\frac{\partial}{\partial\theta}(E_\theta(T(X)) &= \int_{\mathcal{X}} T(x) \frac{\partial}{\partial\theta}(L(\theta,x)) dx = g'(\theta) \\
&= \int_{\mathcal{X}} T(x) \frac{\partial}{\partial\theta}(\log L(\theta,x)) L(\theta,x) dx = Cov(T(X), \frac{\partial}{\partial\theta}(\log L(\theta, X))
\end{aligned}$$

On déduit que :

$$Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta, X)) \leq Var_\theta(T(X)) Var(\frac{\partial}{\partial\theta}(\log L(\theta, X))$$

et donc

$$Var_\theta(T(X)) \geq \frac{Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta, X))}{Var(\frac{\partial}{\partial\theta}(\log L(\theta, X))} = \frac{g'(\theta)^2}{I(\theta)}$$

□

Cas particulier : $g(\theta) = \theta$
$BCR = \dfrac{1}{I(\theta)}$

Définition 3.9. Un estimateur sans biais T de $g(\theta)$ est dit efficace si

$$Var_\theta(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

Exemple 3.9. Estimation du paramètre d'une loi de Poisson (suite)
On a

$$Var(\overline{X}) = \frac{\lambda}{n}$$

et

$$BCR = \frac{1}{nI(\lambda)} = \frac{1}{\frac{n}{\lambda}} = \frac{\lambda}{n}.$$

Par conséquent $\overline{X}$ est un estimateur efficace de λ.

Remarque 3. En reprenant la démonstration du Théorème de Rao-Cramer, on remarque que pour que T soit efficace, il est nécessaire et suffisant que

$$Var_\theta(T(X)) = \frac{Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta, X))}{Var(\frac{\partial}{\partial\theta}(\log L(\theta, X))} = \frac{g'(\theta)^2}{I(\theta)}.$$

Cet égalité signifie que le coefficient de corrélation entre $T(X)$ et $\frac{\partial}{\partial\theta}(\log L(\theta, X)$ est égal à 1, ce qui alieu si et seulement s'il existe des constantes $k(\theta)$ et $l(\theta)$ (qui peuvent dépendre de θ) telles que :

$$\frac{\partial}{\partial\theta}(\log L(\theta, X) = k(\theta)T(X) + l(\theta)$$

cela implique :

$$E_\theta(\frac{\partial}{\partial\theta}(\log L(\theta, X)) = k(\theta)E_\theta(T(X)) + l(\theta)$$

$$E_\theta(T(X)) = g(\theta) = -\frac{l(\theta)}{k(\theta)}$$

$$\frac{\partial}{\partial\theta}(\log L(\theta, X) = k(\theta)(T(X) - g(\theta))$$

Cette dernière équation constitue une condition nécessaire et suffisante d'existence d'un estimateur efficace pour une certaine fonction $g(\theta)$ de θ, l'estimateur étant alors forcément $T(X)$. On peut remarquer qu'il ne peut exister qu'une seule fonction $g(\theta)$ de θ qui peut-être estimée efficacement.

3.6 Estimation sans biais : cas vectoriel

Lorsque le paramètre est vectoriel, i.e $\theta = (\theta_1, \theta_2, ..., \theta_p)$, les théorèmes de Rao-Blackwell et de Lehmann-Schéffé restent vraies. Le théorème de Rao-Cramer nécessite une adaptation. Les variances et quantités d'information invoquées ne sont plus des nombres, mais des matrices. On rappelle que si A et B sont deux matrices semi-définies positives, on dit que A est supérieure ou égale à B si $A - B$ est semi-définie positive i.e si

$$Q_A(x) = x^t Ax \geq Q_B(x) = x^t Bx$$

et on note $A \geq B$ ou $A - B \geq 0$.

Théorème 3.20 (Théorème de Rao-Cramer : cas vectoriel)**.** *Soit T un estimateur sans biais de $g(\theta)$ (à valeurs dans R^p). On suppose les conditions de régularité de Fischer vérifiées, alors*

$$\mathcal{V}_\theta(T) \geq \nabla(g(\theta)^t \mathcal{I}(\theta)^{-1} \nabla(g(\theta)$$

où $\mathcal{V}_\theta(T)$ est la matrice de variances-covariances de T.

Remarque 4.
1. T est dit estimateur efficace si $\mathcal{V}_\theta(T) = \nabla(g(\theta)^t \mathcal{I}(\theta)^{-1} \nabla(g(\theta)$
2. Si $g(\theta) = \theta$ alors l'inégalité de de Rao-Cramer devient $\mathcal{V}_\theta(T) \geq \mathcal{I}(\theta)^{-1}$

Exemple 3.10 (Estimation simultanée de la moyenne et de la variance de la loi $N(m, \sigma^2)$)**.** On sait que la statistique

$$T(\tilde{x}) = (\overline{x_n}, S_n'^2) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} x_i^2}{n-1} - n(n-1) \left(\frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right)$$

est un estimateur sans biais de (m, σ^2), qui est par ailleurs l'unique ESBVUM puisque $\left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2 \right)$ est exhaustive et complète. Pour qu'il soit de plus efficace on doit avoir $\mathcal{V}_\theta(T) = \mathcal{I}_n^{-1}(m, \sigma^2)$. On a :

$$\mathcal{V}_\theta(T) = \begin{pmatrix} Var(\overline{X}_n) & Cov(\overline{X_n}, S_n'^2) \\ Cov(\overline{X}_n, S_n'^2) & Var(S_n'^2) \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{n-1}{2\sigma^4} \end{pmatrix}$$

puisque $\overline{X_n}$ et $S_n'^2$ sont indépendants et que $\frac{(n-1)S_n'^2}{\sigma^2} \sim \chi^2_{n-1}$. D'autre part

$$\mathcal{I}_n(m, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix} \text{ et donc } \mathcal{I}_n^{-1}(m, \sigma^2) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix}.$$

On conclut que $\mathcal{V}_\theta(T) \neq \mathcal{I}_n^{-1}(m, \sigma^2)$ et que T n'est pas un estimateur efficace de (m, σ^2).