Estimation ponctuelle

En statistique, comme dans la théorie des probabilités le hasard intervient fortement. Mais dans la théorie des probabilités, on suppose la loi connue précisément et on cherche à donner les caractéristiques de la variable qui suit cette loi. L'objectif de la statistique est le contraire : à partir de la connaissance de la variable, que peut-on dire de la loi de cette variable ?

1. Définitions

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

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

Définition 1. *Un* n-échantillon *de X est un n-uplet $(X_1, X_2, \ldots, X_n)$ tel que les X_k ont la même loi que X et sont indépendantes.*
Une réalisation de l'échantillon *est alors un n-uplet $(x_1, x_2, \ldots, x_n)$ de valeurs prises par l'échantillon.*

Définition 2. *Une* statistique *de l'échantillon est une variable aléatoire $\varphi(X_1, X_2, \ldots, X_n)$ où φ est une application de $\mathbb{R}^n$ dans $\mathbb{R}$.*
Un estimateur *T de θ est une statistique à valeurs dans I. Une* estimation *est la valeur de l'estimateur correspondant à une réalisation de l'échantillon.*

Exemple : $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ est un estimateur de l'espérance mathématique.

Définition 3. *Le* biais *de l'estimateur T de θ est $\mathbb{E}[T] - \theta_0$. S'il est nul, on dit que T est un estimateur sans biais.*
L'estimateur T_n est asymptotiquement sans biais *si $\lim \mathbb{E}[T_n] = \theta_0$.*

On note souvent le biais $b_\theta(T)$.

Définition 4. *L'estimateur est dit* convergent *si la suite (T_n) converge en probabilité vers θ_0 :*

$$\forall \varepsilon > 0, \mathbb{P}(|T_n - \theta_0| > \varepsilon) \underset{n \to +\infty}{\longrightarrow} 0.$$

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

D'après Bienaymé-Tchebychev pour qu'un estimateur asymptotiquement sans biais soit convergent il suffit que

$$\mathrm{Var}(T_n) \underset{n \to +\infty}{\longrightarrow} 0.$$

2. Critères de comparaison d'estimateurs

Un bon critère de comparaison est le *risque quadratique.*

Définition 5. *Soient T un estimateur de θ. Le risque quadratique est défini par*

$$R(T,\theta) = \mathbb{E}[(T-\theta)^2]$$

On peut alors comparer deux estimateurs.

Définition 6. *On dit que T_1 est un* meilleur estimateur *que T_2 si*

$$\forall \theta \in I, \quad R(T_1,\theta) \leq R(T_2,\theta)$$

et

$$\exists \theta \in I, \quad R(T_1,\theta) < R(T_2,\theta).$$

Un estimateur est dit *admissible* s'il n'existe pas d'estimateur meilleur.

L'erreur quadratique moyenne de T se décompose en deux termes, le carré du biais et la variance de T :

$$\mathbb{E}[(T-\theta)^2] = b_\theta^2(T) + \mathrm{Var}(T).$$

Cette décomposition permet de se ramener à une discussion sur la variance pour les estimateurs sans biais de θ.

Définition 7. *Soient T_1 et T_2 deux estimateurs sans biais de θ. On dit que T_1 est un* plus efficace *que T_2 si*

$$\forall \theta \in I, \quad \mathrm{Var}(T_1) \leq \mathrm{Var}(T_2)$$

et

$$\exists \theta \in I, \quad \mathrm{Var}(T_1) < \mathrm{Var}(T_2).$$

On parle d'estimateur à variance minimale si seul le premier point est vérifié, c'est-à-dire :

$$\mathrm{Var}(T_1) \leq \mathrm{Var}(T_2).$$

3. Exemples fondamentaux

Soit X une variable aléatoire telle que $\mathbb{E}[X] = m$ et $\mathrm{Var}(X) = \sigma^2$.

3.a. Estimation de m.

Théorème 8.

La moyenne empirique $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ est un estimateur sans biais et convergent de m.

On a

$$\mathbb{E}[\overline{X_n}] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] = m \quad \text{et} \quad \mathrm{Var}(\overline{X_n}) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}[X_k] = \frac{\sigma^2}{n} \underset{n\to+\infty}{\longrightarrow} 0.$$

D'après la loi forte des grands nombres $\overline{X_n}$ est même fortement convergent.

Il est possible de déterminer la loi asymptotique de la moyenne empirique.

Proposition 9. *Si n est assez grand on peut utiliser l'approximation normale (lorsque X admet un moment d'ordre 2)*

$$\overline{X_n} \overset{\mathcal{L}}{\sim} \mathcal{N}(m, \sigma^2/n).$$

C'est une conséquence du TCL qui nous assure que

$$\sqrt{n}(\overline{X_n} - m) \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0, \sigma^2).$$

3.b. Estimation de σ^2 en supposant m connu.

Théorème 10.

Lorsque m est connu

$$S_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2$$

est un estimateur sans biais et convergent de σ^2.

On a

$$\mathbb{E}[S_n^2] = \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2\right] = \frac{1}{n}\sum_{k=1}^{n}\mathrm{Var}(X_k) = \sigma^2$$

Par ailleurs, les variables $(X_k - m)^2$ étant indépendantes :

$$\mathrm{Var}(S_n^2) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}((X_k - m)^2) = \frac{1}{n}\left(\mathbb{E}[(X-m)^4] - \mathbb{E}[(X-m)^2]^2\right) = \frac{1}{n}\left(\mu_4 - \sigma^4\right)$$

avec $\mu_k = \mathbb{E}((X-m)^k)$.

Donc S_n^2 est un estimateur convergent. La loi forte des grands nombres appliquée aux variables $(X_k - m)^2$ entraîne même la convergence presque sûre vers σ^2.

Comme dans le cas de la moyenne empirique le TCL nous permet de déterminer la loi asymptotique de S_n^2 ; on a lorsque n est assez grand :

$$S_n^2 \overset{\mathcal{L}}{\sim} \mathcal{N}(\sigma^2, (\mu_4 - \sigma^4)/n).$$

3.c. Estimation de σ^2 lorsque m est inconnu.

En général on ne connaît pas m ; on le remplace par un estimateur et on introduit la variance empirique associée :

$$\overline{S_n^2} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X_n})^2.$$

Théorème 11.

La variance empirique $\overline{S_n^2}$ est un estimateur biaisé et convergent de σ^2. Il est asymptotiquement sans biais.

On a

$$\mathbb{E}[\overline{S_n^2}] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k^2) - \mathbb{E}[\overline{X_n}^2] = \frac{1}{n}(n(m^2+\sigma^2)) - (m^2 + \frac{\sigma^2}{n}) = \frac{n-1}{n}\sigma^2.$$

D'autre part, on peut montrer que :

$$\mathrm{Var}(\overline{S_n^2}) = \frac{1}{n}\left(\mu_4 - \sigma^4\right) - \frac{2}{n^2}\left(\mu_4 - 2\sigma^4\right) + \frac{1}{n^3}\left(\mu_4 - 3\sigma^4\right) \to 0$$

avec $\mu_k = \mathbb{E}((X-m)^k)$. L'estimateur est donc convergent.

Le résultat précédent et le lemme de Slutsky (*Probabilité 2, Jean-Yves Ouvrard, p. 347*) permet de déterminer la loi asymptotique de $\overline{S_n^2}$:

$$\overline{S_n^2} \overset{\mathcal{L}}{\sim} \mathcal{N}(\sigma^2, (\mu_4 - \sigma^4)/n).$$

Théorème 12.

La variance empirique corrigée

$$\widehat{S_n^2} = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X_n})^2.$$

est un estimateur sans biais et convergent de σ^2.

Cela se montre facilement en remarquant que

$$\widehat{S_n^2} = \frac{n}{n-1}\overline{S_n^2}.$$

4. Cas particulier de la loi normale

On suppose dans ce paragraphe que X suit la loi normale $\mathcal{N}(m, \sigma^2)$. On sait que $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ suit alors la loi normale $\mathcal{N}(m, \sigma^2/n)$, ce qui confime que c'est un estimateur sans biais, convergent de m.

Les résultats obtenus au paragraphe précédent pour l'estimation de σ^2 sont encore valables ; en particulier on a :

$$\mathbb{E}(S_n^2) = \sigma^2 \quad \text{et } \mathrm{Var}(S_n^2) = \frac{2(n-1)}{n^2}\sigma^4$$

En effet, calculons μ_k

$$\begin{aligned}
\mu_k &= \mathbb{E}((X-m)^k) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-m)^k \exp\left(-\frac{(x-m)^2}{2\sigma}\right)dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(\sqrt{2}\sigma u)^k \exp\left(-u^2\right)\sqrt{2}\sigma du \quad \text{ en posant } x = m\sqrt{2}\sigma u \\
&= 0 \quad \text{si } k \text{ est impair.}
\end{aligned}$$

Lorsque $k = 2p$ est pair on obtient

$$\begin{aligned}
\mu_{2p} &= \frac{2^p\sigma^{2p}}{\sqrt{\pi}}\int_{-\infty}^{+\infty} u^{2p}\exp\left(-u^2\right)du = \frac{2^{p+1}\sigma^{2p}}{\sqrt{\pi}}\int_{0}^{+\infty} u^{2p}\exp\left(-u^2\right)du \\
&= \frac{2^p\sigma^{2p}}{\sqrt{\pi}}\int_{0}^{+\infty} v^{p-1/2}\exp\left(-v\right)dv \quad \text{ en posant } u = \sqrt{v} \\
&= \frac{2^p\sigma^{2p}}{\sqrt{\pi}}\Gamma(p+1/2) = \frac{(2p)!}{2^p(p!)}\sigma^{2p}
\end{aligned}$$

et donc

$$\mathrm{Var}(S_n^2) = \frac{1}{n}\left(\mu_4 - \sigma^4\right) - \frac{2}{n^2}\left(\mu_4 - 2\sigma^4\right) + \frac{1}{n^3}\left(\mu_4 - 3\sigma^4\right) = \frac{2(n-1)}{n^2}\sigma^4$$

Définition13. *Soient* X_1 $,\ldots,X_n$, *n variables aléatoires indépendantes identiquement distribuées de loi* $\mathcal{N}(0,1)$. *La loi du* χ^2 *à n degrés de liberté est la loi de la variable aléatoire*

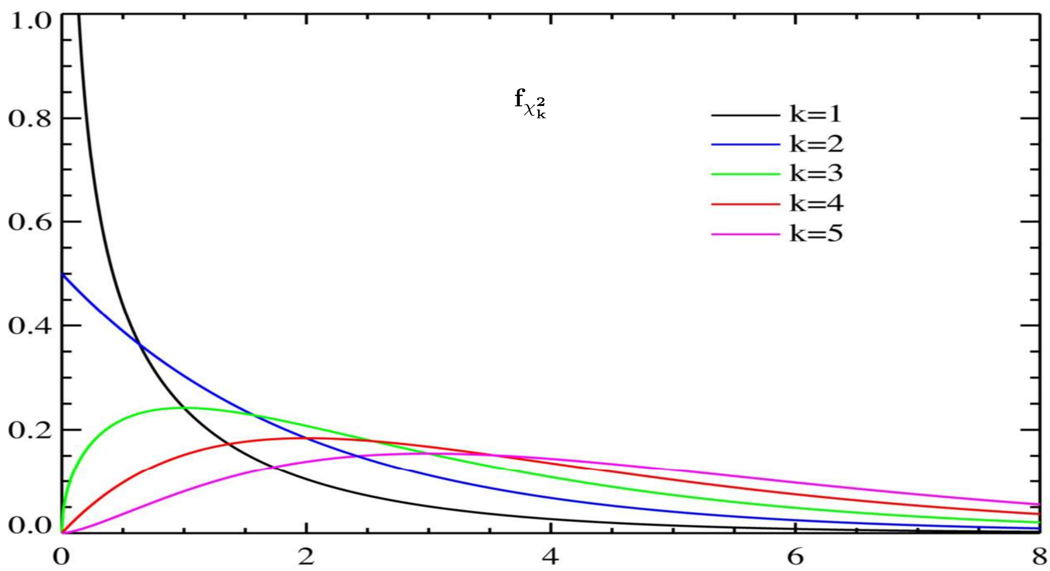
$$\chi_n^2=\sum_{k=1}^n X_k^2.$$

La densité de cette loi est donnée par :

$$f_{\chi_n^2}(u)=\frac{1}{2\Gamma(n/2)}\left(\frac{u}{2}\right)^{n/2-1}\exp\left(-\frac{u}{2}\right)1_{u>0}$$

et sa fonction caractéristique par

$$\phi_{\chi_n^2}(t)=\frac{1}{(1-2it)^{n/2}}$$



Pour déterminer la densité on peut remarquer que : si U suit une loi $\mathcal{N}(0,1)$ alors on a pour $t>0$

$$\mathcal{P}(U^2\le t)=\mathcal{P}(-t\le U\le t)=F_U(\sqrt{t})-F_U(-\sqrt{t})$$

et par conséquent

$$f_{U^2}(t)=\frac{1}{2\sqrt{t}}f_U(\sqrt{t})+\frac{1}{2\sqrt{t}}f_U(-\sqrt{t})=\frac{1}{\sqrt{t}}f_U(\sqrt{t})=\frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{t}{2}\right)$$

Ensuite on obtient le résultat général par récurence.

Théorème 14.

Soit $(X_1,\ldots,X_n)$ *un* $n-$*échantillon de le loi* $\mathcal{N}(0,1)$. *Les variables aléatoires*

$$\sqrt{n}\,\overline{X}_n \quad et \quad \sum_{k=1}^n (X_k-\overline{X}_n)^2=n\overline{S_n^2}=(n-1)\widehat{S_n^2}$$

sont indépendantes et suivent respectivement la loi normale réduite et la loi du χ^2 *à* $(n-1)$ *degrés de liberté.*

Démonstration. Montrons que $\overline{X}_n$ et $\sum_{k=1}^{n}(X_k-\overline{X}_n)^2$ sont indépendantes. On a

$$\begin{pmatrix} \overline{X}_n \\ X_1-\overline{X}_n \\ \vdots \\ X_n-\overline{X}_n \end{pmatrix} = A \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad \text{où} \quad A = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}$$

Le vecteur aléatoire $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ est gaussien de loi $\mathcal{N}(0,I_n)$ où I_n est la matrice identité d'ordre n.

Par conséquent, le vecteur $\begin{pmatrix} \overline{X}_n \\ X_1-\overline{X}_n \\ \vdots \\ X_n-\overline{X}_n \end{pmatrix} = A \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ est également gaussien de loi $\mathcal{N}(0,AI_nA^t) = \mathcal{N}(0,AA^t)$. Or

$$AA^t = \begin{pmatrix} \frac{1}{n} & 0 & 0 & \cdots & 0 \\ 0 & \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ 0 & -\frac{1}{n} & \frac{n-1}{n} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -\frac{1}{n} \\ 0 & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}$$

donc la variable $\overline{X}_n$ est indépendante du vecteur $\begin{pmatrix} X_1-\overline{X}_n \\ \vdots \\ X_n-\overline{X}_n \end{pmatrix}$ et donc de $\sum_{k=1}^{n}(X_k-\overline{X}_n)^2$.

Comme $\overline{X}_n$ suit la loi $\mathcal{N}(0,1/n)$ on en déduit que $\sqrt{n}\overline{X}_n$ suit la loi $\mathcal{N}(0,1)$.

Montrons $n\overline{S_n^2} = \sum_{k=1}^{n}(X_k-\overline{X}_n)^2$ suit la loi du χ^2 à $(n-1)$ degrés de liberté. On a

$$\begin{pmatrix} X_1-\overline{X}_n \\ \vdots \\ X_n-\overline{X}_n \end{pmatrix} = B \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad \text{où} \quad B = \begin{pmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}$$

Comme B est une matrice symétrique, il existe une matrice orthogonale U et une matrice diagonale D telle que

$$B = UDU^t$$

Or les valeurs propres de B sont :

- la valeur propre simple 0 dont le sous-espace propre associé a pour équation $x_1 = \cdots = x_n$;
- la valeur propre d'ordre $(n-1)$ égale à 1 dont le sous-espace propre associé a pour équation $x_1+x_2+\cdots+x_n = 0$

En ordonnant convenablement la base de vecteurs propres on peut choisir

$$D = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

On a $Y = BX = UDU^tX$ et

$$\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 = Y^tY = X^tUDU^tUDU^tX = (U^tX)^tD(U^tX)$$

Or le vecteur aléatoire $Z = U^tX$ est gaussien de loi $\mathcal{N}(0, U^tI_nU) = \mathcal{N}(0, U^tU) = \mathcal{N}(0, I_n)$. D'où

$$\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 = Z^tDZ = \sum_{k=1}^{n-1} Z_i^2$$

qui suit la loi du χ^2 à $(n-1)$ degrés de liberté. □

On en déduit immédiatement que si $(X_1, \dots, X_n)$ est un échantillon d'une variable aléatoire $\mathcal{N}(m, \sigma^2)$ la variable aléatoire

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$$

suit la loi du χ^2 à $(n-1)$ degrés de liberté.
Il suffit de poser $Y_k = X_k - m/\sigma$. Alors, comme on a

$$\overline{Y}_n = \frac{\overline{X}_n - m}{\sigma} \quad \text{et} \quad \frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 = \sum_{k=1}^{n}(Y_k - \overline{Y}_n)^2$$

le résultat découle de ce qui précède.

5. Construction d'estimateur par la méthode du maximum de vraisemblance

5.a. Cas discret. On suppose donnée une observation X tirée selon une loi $\mathbb{P}_\theta$, $\theta \in \Theta$. On supposera ici que $\mathbb{P}_\theta$ est discrète et on pose :

$$f_\theta(x) = \mathbb{P}_\theta(X = x).$$

On appelle alors *fonction de vraisemblance* la fonction $L_X(\theta) = f_\theta(X)$. Quand on dispose d'un $n-$échantillon $(X_1, \dots, X_n)$ de loi $\mathbb{P}_\theta$, la vraisemblance s'écrit alors

$$L_{X_1,\dots,X_n}(\theta) = \prod_{i=1}^{n} f_\theta(X_i).$$

Lorsque la fonction de vraisemblance admet un unique maximum atteint en $\hat{\theta} = g_n(X_1, \dots, X_n)$, on peut utiliser cette valeur pour estimer θ. On dit alors que

$$T = g_n(X_1, \dots, X_n)$$

est l'*estimateur par maximum de vraisemblance* de θ.

Cet estimateur est naturel puisqu'il conduit à privilégier la valeur de θ la "plus probable" au vu de l'observation. Il possède en général de bonnes propriétés. L'inconvénient est que ce maximum peut ne pas exister ou ne pas être unique et il peut être difficile à exhiber.

En pratique, la recherche de ce maximum se fait par dérivation de L relativement à θ. On peut de manière équivalente maximiser le logarithme de la vraisemblance (la fonction logarithme étant croissante, maximiser la vraisemblance et la log-vraisemblance revient au même, mais souvent les calculs sont plus simples).

Exemple : Estimation du paramètre d'une loi de Bernoulli.

Ici on suppose $\Theta =]0, 1[$ et les X_i suivent une loi de Bernoulli de paramètre $\theta \in \Theta$. On a

$$f_\theta(x) = \mathbb{P}_\theta(X = x) = \begin{cases} \theta & \text{si } x = 1 \\ 1 - \theta & \text{si } x = 0 \\ 0 & \text{sinon} \end{cases}$$

Posons $S_n = X_1 + \cdots + X_n$. Ainsi S_n est le nombre de 1 dans l'échantillon et $n - S_n$ le nombre de 0. La vraisemblance et la log-vraisemblance s'écrivent alors :

$$L_{X_1,\ldots,X_n}(\theta) = \theta^{S_n}(1-\theta)^{n-S_n} \qquad lnL_{X_1,\ldots,X_n}(\theta) = S_n ln\theta + (n - S_n)ln(1-\theta).$$

Un calcul montre alors que le maximum est atteint en

$$\hat{\theta} = n^{-1}S_n.$$

Par conséquent l'estimateur de θ par maximum de vraisemblance est

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

qui est également l'estimateur de la moyenne.

5.b. Cas à densité. On suppose donnée une observation X tirée selon une loi $\mathbb{P}_\theta$, $\theta \in \Theta$. On supposera ici que $\mathbb{P}_\theta$ admet une densité par rapport à la mesure de Lebesgue notée f_θ.

On appelle alors *fonction de vraisemblance* la fonction $L_X(\theta) = f_\theta(X)$. Quand on dispose d'un $n-$échantillon $(X_1, \ldots, X_n)$ on a la vraisemblance

$$L_{X_1,\ldots,X_n}(\theta) = \prod_{i=1}^{n} f_\theta(X_i).$$

Ensuite, on procède comme dans le cas discret.

Exemple : On cherche à estimer le paramètre θ inconnu d'une loi exponentielle. La vraisemblance s'écrit :

$$L_{X_1,\ldots,X_n}(\theta) = \exp\left(-\sum_{i=1}^{n} X_i/\theta\right)\theta^{-\sum_{i=1}^{n} X_i}$$

Le maximum est atteint en un unique point $\hat{\theta}_n = \overline{X}_n$.