Estimation ponctuelle

Charles Vin

Date

1 Introduction

Diapo osef un peu

2 Modèle Statistique

— On collecte les données :
— On note $x_1, x_2, \dots, x_n$ les réalisations
— des variables aléatoire $X_1, X_2, ..., X_n$
— $x_1 = X_1, ..., x_n = X_n$

Hypothèse : On suppose $X_1, ..., X_n$ indépendante et de même loi P (on dit "i.i.d.")
Objectif : Estimer P (difficile) à partir de $X_1, ...X_n$

Définition 2.1 (Modèle statistique)**.** On appelle modèle statistique la donnée du triplet (Ω, F, P_θ) (θ est un paramètre.) où Ω = espace des observations, F tribu sur Ω, $P_{\theta\in\Theta}$ est une famille de probabilité définie sur (Ω, F), θ = espace des paramètres.

Exemple 2.1. On souhaite tester l'efficacité d'un médicament. On traite n patients (individus statistique). A l'issue de l'étude 72 patients sont guéris.
Modélisation Statistique : On note $x_i = \begin{cases} 1 \\ 0 \end{cases}$ Si guéris ou non. x_i est une réalisation d'une VA de Bernouilli $\mathcal{B}(p), p \in]0, 1[, \Omega = \{0, 1\}, F = \mathcal{P}(\{0, 1\})$ et la famille de probabilité est $\mathcal{B}(p), p(= \theta) \in]0, 1[(= \Theta)$
Objectif : Reconstruire ou estimer ce paramètre p à l'aide de $X_1, ..., X_n$

Exemple 2.2 (Nombre de voiture au feu)**.** C'est un modèle de Poisson avec :

$$\Omega = N, F = \mathcal{P}(N), (P(\lambda), \lambda > 0) \text{ avec } P_\lambda[X = k] = e^-\lambda\frac{\lambda^k}{k!}.$$

Le paramètre est $\lambda \in R_+^*$ R+* est le Ω

Exemple 2.3 (Durée du trajet)**.** Modèle gaussien :

$$\Omega = R, F = \mathcal{B}(R), (\mathcal{N}(m, \sigma^2), (m, \sigma^2) \in \mathcal{R}x\mathcal{R}_+^*).$$

3 Estimateur et propriétés

Définition 3.1 (n-échantillon)**.** On appelle n-échantillon (échantillon de taille n) associé au modèle $(\Omega, F, (P_\theta)_{\theta in \Theta})$ une suite $X_1, ..., X_n$ de va i.i.d. de même loi P_θ

Définition 3.2 (moyenne empirique)**.** On appelle moyenne empirique la va

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

Définition 3.3 (Variance empirique)**.** On appelle variance empirique la va

$$V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X_n}^2.$$

Définition 3.4 (Statistique)**.** On appelle statistique toute application de Ω^n dans R^p

Définition 3.5 (Estimateur)**.** On appelle estimateur du paramètre θ toute statistique à valeur dans Θ

Remarque. Une statistique est une VA qui ne dépend que de $X_1, ..., X_n$ et a sa propre loi de probabilité.

Exemple 3.1. Exemple du traitement

- $\hat{p_1}$ est un estimateur de p.
- $\hat{p_2} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur de p (car entre 0 et 1)
- $S_n = \sum_{i)1}^{n} X_i$ est une statistique (car dans R)

Théorème 3.1. *$X_1, ..;, X_n$ i.i.d. de loi commune $(\mathcal{P}_\theta)_{\theta\in\Theta}$. Soit $m_\theta = \mathbb{E}_\theta[X_1]$ (espérance de X_1 pour P_θ) , $\sigma_\theta^2 = Var_{\Theta(X_1)}$ (variance de X_1 pour $\mathcal{P}_\Theta$).*
La moyenne empirique $\bar{X}_n = \frac{1}{n}$ est un estimateur de m_θ.
La variance empirique $V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ est un estimateur de σ_Θ^2

Définition 3.6 (Le moment empirique simple)**.** Le moment empirique simple d'ordre $k, k \in N$ noté $\hat{m_k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ est un estimateur du moment théorique d'ordre $k \in N^*$, $m_k(\theta) = \mathbb{E}_\theta(X_1^k)$

Définition 3.7 (Le moment empirique centré d'ordre k)**.** Le moment empirique centré d'ordre k se note $\hat{\mu_k} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X}_n)^k$ est un estimateur de $\mu_k(\Theta) = \mathbb{E}_\theta((X_1 - \mathbb{E}_\theta(X_1)^k))$

Note. La question est de savoir quel estimateur choisir !

4 Critère de performance

Soit T_n un estimateur de θ.

Définition 4.1. T_n est un estimateur sans biais du paramètre θ si $\mathbb{E}_\theta(T_n) = \theta$ (en moyenne on tombe sur le vrais paramètre).
T_n est un estimateur biaisé de θ si $\mathbb{E}_\theta(T_n) \neq \theta$
La quantité $\mathcal{B}_\theta(T_n) = \mathbb{E}_\theta(T_n) - \theta$ est appelé **biais de l'estimateur** en θ
Un estimateur T_n de θ est dit asymptotiquement sans biais si $\lim_{n\to+\infty} \mathbb{E}_\Theta(T_n) = \theta$.

Distribution d'un estimateur T_n sans biais de θ
Distribution d'un estimateur biaisé T_n de θ VOIR DIAPO

Cadre général : $X_1, X_2, ..., X_n$ i.i.d. de loi $P_\theta, \theta \in \Theta$ avec $m_\theta = E_\theta = (X_1)$ et $\sigma_\theta^2 = var_\theta(X_1)$.

Théorème 4.1. *La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur sans biais de m_θ . En effet :*

$$\begin{aligned} E_\theta(\frac{1}{n}\sum_{i=1}^{n} X_i) &= \frac{1}{n}\sum_{i=1}^{n} E_\theta(X_i) \\ &= \frac{1}{n} n E_\theta(X_1) \\ &= m_\theta \end{aligned}$$

Exemple 4.1. $X_1, \ldots, X_n$ i.i.d. de loi $Exp(\lambda)$ avec $\lambda > 0$. On a
- $E_\lambda(X_1) = \frac{1}{\lambda}$
- $Var_\lambda(X_1) = \frac{1}{\lambda^2}$
- λ inconnu

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur sans biais de $\frac{1}{\lambda}$ ($\frac{1}{\bar{X}_n}$ est un estimateur de λ)

Exemple 4.2. $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{P}(\lambda)$ avec $\lambda > 0$.

$$\forall k \in N, P_\lambda(X_1 = k) = e^- k\frac{\lambda^k}{k!}, E_\lambda(X_1) = \lambda.$$

Théorème 4.2. *La **variance empirique** $V_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X_n})^2$ est un estimateur biaisé de σ_θ^2.*
*La **variance empirique corrigée** $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X_n})^2$ est un estimateur sans biais de σ_θ^2*

Preuve : On doit montrer que $E_\theta(V_n^2) \neq \theta$. On a

$$E_\theta(V_n^2) = E_\theta(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X_n})^2).$$

Or

$$\begin{aligned}
V_n^2 &= \frac{1}{n}\sum_{i=1}^n (X_i^2 - 2X_i\bar{X_n} + \bar{X_n}^2) \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\bar{X_n}(\frac{1}{n}\sum_{i=1}^n X_i) + \bar{X_n}^2 \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X_n}^2
\end{aligned}$$

$$\begin{aligned}
E_\theta(V_n^2) &= E_\theta(\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X_n^2}) \\
&= \frac{1}{n}\sum_{i=1}^n E_\theta(X_i^2) - E_\theta(\bar{X_n}^2) \\
&= \frac{1}{n} * nE_\theta(X_1^2) - E_\theta(\bar{X_n}^2) \text{ car mes Xi ont la même loi} \\
&= E_\theta(V_n^2) = E_\theta(X_1^2) - E_\theta(\bar{X_n^2})
\end{aligned}$$

Finaly :

$$Var_\theta(X_1) = E_\theta(X_1^2) - E_\theta^2(X_1) \Leftrightarrow \sigma_\theta^2 = E_\theta(X_1^2) - m_\theta^2.$$
$$\Leftrightarrow E_\theta(X_1^2) = \sigma_\theta^2 + m_\theta^2.$$

De la même façon, on a

$$E(X_n^2) = var(\bar{X_n}) + E^2(\bar{X_n}).$$

et

$$\begin{aligned}
Var_\theta(\bar{X_n}) &= Var_\theta(\frac{1}{n}\sum_{i=1}^n X_i) \\
&= (\frac{1}{n})^2 Var_\theta(\sum_{i=1}^n X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^n Var(X_i) \text{ car les Xi sont indépendante} \\
&= \frac{1}{n^2} nVar_\theta(X_1) \text{ car les Xi ont la même loi}
\end{aligned}$$

Conclusion :

$$Var_\theta(\bar{X_n}) = \frac{Var_\theta(X_1)}{n} = \frac{\sigma_\theta^2}{n}.$$

On obtient

$$\begin{aligned}
E_\theta(V_n^2) &= E_\theta(X_1^2) - (Var_\theta(X_n) + E_\theta^2(\bar{X_n})) \\
&= (\sigma_\theta^2 + m_\theta^2) - (\frac{\sigma_\theta^2}{n} + m_\theta^2) \\
&= \sigma_\theta^2 - \frac{\sigma_\theta^2}{n} \neq \sigma_\theta^2
\end{aligned}$$

(V_n^2 est un estimateur biaisé de σ_θ^2)

Le biais de V_n^2 en $B_\theta(V_n^2) = E_\theta(V_n^2) - \sigma_\theta^2 = -\frac{\sigma_\theta^2}{n}$.

On corrige l'estimateur V_n^2 :

$$E_\theta(V_n^2) = \frac{n-1}{n}\sigma_\theta^2$$
$$E_\theta(\frac{n}{n-1}V_n^2) = \sigma_\theta^2$$

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

est un estimateur sans biais de σ_θ^2 .
V_n^2 est un estimateur asymptotiquement sans biaisés de σ_θ^2 car :

$$\lim_{n\to\infty} E_\theta(V_n^2) = \sigma_\theta^2.$$

□

Exemple 4.3. $X_1, \dots, X_n$ i.i.d. de loi $Exp(\lambda), \lambda > 0$.
On a $\sigma_\lambda^2 = Var_\lambda(X_1) = \frac{1}{\lambda^2}$
$S_n^2 = \frac{1}{n-1}\sum(X_i - \bar{X_n})^2$ est i, estimateur san biais de $\frac{1}{\lambda^2}$
V_n^2 est un estimateur biaisé de $\frac{1}{\lambda^2}$

Définition 4.2 (Risque quadratique)**.** — Soit T_n un estimateur de θ, on appelle **Risque quadratique** de l'estimateur T_n (=erreur quadratique moyenne) la quantité définie par

$$\mathcal{R}(\theta, T_n) = E_\theta[(T_n - \theta)^2].$$

— Soit T_n et T'_n deux estimateur de θ, T_n est dit **préférable** à T'_n si

$$\mathcal{R}(\theta, T_n) < \mathcal{R}(\theta, T'_n), \forall \theta in \Theta.$$

Remarque. On a la décomposition :

$$\mathcal{R}(\theta, T_n) = Var_\theta(T_n) + (E_\theta(T_n) - \theta)^2.$$

Lorsque T_n est sans biais, sont risque quadratique se réduit à sa variance :

$$\mathcal{R}(\theta, T_n) = Var_\theta(T_n).$$

Preuve : On a :

$$\begin{aligned}\mathcal{R}(\theta, T_n) = E_\theta[(T_n - \theta)^2] &= E_\theta[(T_n - E_\theta(T_n) + E_\theta(T_N) - \theta)^2] \\ &= E_\theta[(T_n - E_\theta(T_n))^2] + (E_\theta - \theta)^2 + 2E_\theta[(T_n - E_\theta(Tn)(E_\theta(T_n) - \theta))] \\ &= Var_\theta(T_n) + \mathcal{B}_\theta^2(T_n) + 2\mathcal{B}_\theta(T_n)E_\theta[(T_n - E_\theta(T_n))] \\ &= Var_\theta(T_n) + \mathcal{B}_\theta^2(T_n)\end{aligned}$$

□

Exemple 4.4 (Avec une loi binomiale)**.** Soient $X_1, \dots, X_n$ i.i.d. $\mathcal{B}(\theta)$ avec $\theta \in]0;1[$. L'estimateur empirique $\bar{X_n}$ est sans biais pour θ et son risque quadratique

$$\begin{aligned}\mathcal{B}(\theta, \bar{X_n}) &= E_\theta[(\bar{X_n} - \theta)^2] \\ &= Var_\theta(\bar{X_n}) = \frac{Var_\theta(X_1)}{n} \\ &= \frac{\theta(1-\theta)}{n}\end{aligned}$$

$\mathcal{R}(\theta, \bar{X_n})$ tend vers 0 lorsque la taille n augmente.
X_1 est un estimateur sans biais de θ car $E_\theta(X_1) = \theta$ et $\mathcal{R}(\theta, X_1) = Var_\theta(X_1) = \theta(1-\theta)$.

Définition 4.3. T_n est un estimateur de θ **convergeant en moyenne quadratique** si

$$\mathcal{R}(\theta, T_n) \longrightarrow_{n\to+\infty} 0.$$

Exemple 4.5. $X_1, \ldots, X_n$ i.i.d. avec $X_1 \sim B(\theta), \theta \in]0,1[$
Fréquence empirique

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

est un estimateur sans biais de θ. Son risque quadratique :

$$\mathcal{R}(\theta, \bar{X_n}) = \mathbb{E}[(\bar{X_n} - \theta)^2] = Var_\theta(\bar{X_n}) = \frac{Var_\theta(X_1)}{n} = \frac{\theta(1-\theta)}{n}.$$

$$\lim_{n\to\infty} \mathcal{R}(\theta, \bar{X_n}) = \lim_{n\to\infty} \frac{\theta(1-\theta)}{n} = 0 \text{ (on gagne en précision lorsque n augmente).}$$

soit X_1 comme estimateur de θ, X_1 est un estimateur sans biais de θ car $\mathbb{E}_\theta(X_1) = \theta$ et $\mathcal{R}(\theta, X_1) = Var_\theta(X_1) = \theta(1-\theta)$.
Un estimateur T_n est dit convergent en moyenne quadratique vers θ si $\lim_{n\to\infty} \mathcal{R}(\theta, T_n) = 0$

Exemple 4.6 (Exemple précédant : modèle de Bernouilli)**.** $\bar{X_n}$ converge en moyenne quadratique car $\lim_{n\to\infty} \mathcal{R}(\theta, \bar{X_n})$

Définition 4.4 (Estimateur faiblement consistent)**.** Soit T_n un estimateur de θ on dira que T_n est **faiblement consistent** si T_n converge en probabilité vers θ, signifie

$$\forall \theta \in \Theta, \forall \epsilon > 0, \lim_{n\to\infty} \mathbb{P}_\theta(|T_n - \theta| > \epsilon) = 0.$$

Définition 4.5 (Estimateur fortement consistent)**.** T_n est un **estimateur fortement consistent** de θ si T_n converge presque sûrement vers θ signifie

$$\forall \theta \in \Theta, \mathbb{P}(\{\omega \text{ tq. } \lim_{n\to\infty} T_n(\omega) = \theta\}) = 1.$$

Proposition 4.3. *Si T_n converge en moyenne quadratique vers θ alors T_n converge en probabilité vers θ .*
En effet, en appliquant l'inégalité de Tchebychev, on a pour $\epsilon > 0$

$$\mathbb{P}(|T_n - \theta| > \epsilon) \neq \frac{\mathbb{E}((T_n - \theta)^2)}{\epsilon^2}.$$

Exemple 4.7 (du modèle de Bernouilli)**.** $X_1, \ldots, X_n$ i.i.d. avec $X_1 \sim \mathcal{B}(\theta), \theta \in]0,1[$

— En appliquant la loi faible des grands nombre, on a

$$\bar{X_n} \to^{\mathbb{P}}_{n\to\infty} \mathbb{E}_\theta(X_1) = \theta.$$

($\bar{X_n}$ est faiblement consistante pour θ)

— En appliquant la loi forte des grand nombre, on a

$$\bar{X_n} \to^{p.s}_{n\to\infty} \mathbb{E}(X_1) = \theta.$$

($\bar{X_n}$ est fortement consistant pour θ)

— Autre façon de montrer le premier point. On a

$$\mathcal{R}(\theta, \bar{X_n}) = Var(\bar{X_n}) = \frac{\theta(1-\theta)}{n} \to_{n\to\infty} 0.$$

D'après la propriété précédente

$$\bar{X_n} \to^{\mathbb{P}}_{n\to\infty} \theta.$$

La variance $V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2$ est un estimateur consistent de $Var_\theta(X_1)$ car on a

$$V_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2.$$

En appliquant la loi forte des grands nombres, on a

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \rightarrow_{n\to\infty}^{p.s} \mathbb{E}(X_1^2).$$

$$\text{et } \bar{X}_n \rightarrow_{n\to\infty}^{p.s} \mathbb{E}_\theta(X_1).$$

D'où

$$V_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 \rightarrow_{n\to\infty}^{p.s} \mathbb{E}_\theta(X_1^2) - \mathbb{E}(X_1) = Var_\theta(X_1).$$

5 Méthode de maximum vraisemblance

Voir diapo pour plus de précision sur l'exemple :

Exemple 5.1. $X_i = \begin{cases} 1 \text{ si la qualité de l'air dépasse le niveau 8 le jour } i \\ 0 \text{ sinon} \end{cases}$

$\theta = \mathbb{P}(X_i = 1)$ inconnu.
On observe $(x_1, x_2, \ldots, x_10) = (0, 1, 1, 0, 1, 1, 1, 0, 0, 1)$

$$\begin{aligned} \mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_10 = x_10) &= P_\theta(X_1 = 0)P(X_2 = 1)\ldots P(X_10 = 1)\text{(indépendance)} \\ &= \theta^6(1-\theta)^4, \theta \in]0,1[\end{aligned}$$

La valeur de θ qui rend maximale la probabilité d'observer $(0, 1, 1, \ldots, 1)$ est $\hat{\theta} = \frac{6}{10}$

Cadre : $X_1, \ldots, X_n = \mathbf{X}$ i.i.d. de loi $\mathbb{P}_\theta, \theta \in \Theta$. On note $f(x, \theta)$ la densité de X_1 si la loi de X_1 est continues $f(x, \theta) = \mathbb{P}(X_1 = x)$ si la loi de X_1 est discrète. On dispose des données $(X_1, \ldots, X_n)$

Objectif : Trouver θ qui rendre la réalisation $(x_1, \ldots, x_n)$ la plus probable .

Définition 5.1 (Vraisemblance d'un échantillon)**.** On appelle **vraisemblance** de l'échantillon, la fonction L définie $\forall\theta \in \Theta$, par

$$\begin{aligned} L(x_1, \ldots, x_n, \theta) &= \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) \\ &= P(X_1 = x_1) * \cdots * P(X_n = x_n) \text{ si de loi discrète} \\ &= f(x_1, \theta)f(x_2, \theta)\ldots f(x_n, \theta) \text{ si X a une loi continue} \end{aligned}$$

Remarque.
— La vraisemblance est aléatoire. Une fois qu'on a une réalisation de l'échantillon, la vraisemblance ne depend que de θ.
— Si pour deux valeurs du paramètre, on a $L(x_1, \ldots, x_n, \theta_1) > L(x_1, \ldots, x_n, \theta_2)$. On a θ_1 est plus vraisemblable pour θ_2 étant donnée la réalisation $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$

Exemple 5.2 (modèle de Bernouilli)**.** $X_1, \ldots, X_n$ avec $X_1 \sim \mathcal{B}(\theta)$. Données : $x_1, \ldots, x_n$ avec $x_i \in 0, 1$.

$$\begin{aligned} &\mathbb{P}(X_i = 1) = \theta \\ &\mathbb{P}(X_i = 0) = 1 - \theta \\ &\mathbb{P}(X_i = x_i) = \theta^{x_i}(1-\theta)^{x_i}, x_i \in \{0, 1\} \end{aligned}$$

$$\begin{aligned} L(x_1, \ldots, x_n, \theta) &= P(X_1 = x_1, \ldots, X_n = x_n) \\ &= \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i) \text{ car indépendance} \\ &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ L(x_1, \ldots, x_n, \theta) &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

Retour à ATMO : $L(0, 1, \ldots, 1, 0) = \theta^6(1-\theta)^4$

Exemple 5.3. Modèle de Poisson
$(x_1, x_2, \dots, x_n)$ réalisation de $(X_1, \dots, X_n)$ échantillon avec $X_1 \sim \mathcal{P}(\theta)$

$$\forall x \in N, P_\theta(X = x) = e^-\theta\frac{\theta^x}{x!}.$$

La vraisemblance s'écrit

$$\begin{aligned} L(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n P_\theta(X_i = x_i) \\ &= \prod_{i=1}^n e^-\theta\frac{\theta^{x_i}}{x_i!} \\ &= e^-\theta\frac{\theta^{x_1}}{x_1!}e^-\theta\frac{\theta^{x_2}}{x_2!}\dots e^-\theta\frac{\theta^{x_n}}{x_n!} \end{aligned}$$

Exemple 5.4. $(x_1, x_2, \dots, x_n)$ réalisation de $(X_1, \dots, X_n)$ échantillon avec $X_1 \sim \mathcal{N}(\theta, 1)$ ($E_\theta(X) = \theta, var(X) = 1$); La densité de X_i sécrit

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x_\theta)^2}.$$

LA vraisemblance

$$\begin{aligned} L(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n f(x_i, \theta) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x_\theta)^2} \\ &= \frac{1}{(\sqrt{2\pi})^n}e^{-\frac{1}{2}\sum_{i=1}^n(x_i-\theta)^2} \end{aligned}$$

Exemple 5.5. $(x_1, x_2, \dots, x_n)$ réalisation de $(X_1, \dots, X_n)$ échantillon avec X_1 de densité $f(x, \theta) = \frac{\theta}{x^2}\mathbb{1}_{x \geq \theta}()$.
La vraisemblance s'écrit

$$\begin{aligned} L(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n f(x_i, \theta) \\ &= \frac{\theta}{x_1^2}\mathbb{1}_{x_1 \geq \theta} \cdots \frac{\theta}{x_n^2}\mathbb{1}_{x_n \geq \theta} \\ &= \frac{\theta^n}{\prod_{i=1}^n x_i^2} \end{aligned}$$

Définition 5.2 (Estimateur du maximum de vraisemblance)**.** On appelle estimateur pas maximum de vraisemblance (EMV) du paramètre θ un estimateur $\theta_n \in \Theta$ défini par

$$L((X_1, \dots, X_n, \theta)) = \sup_{\theta \in \Theta} L(X_1 \dots X_n, \theta).$$

Remarque. Comme la fonction $x \mapsto \log(x), x > 0$ est croissante strictement

$$\sup_{\theta \in \Theta} L(x_1 \dots x_n, \theta) = \sup_{\theta \in \Theta} \log L(x_1 \dots x_n, \theta) = \begin{cases} \sup_{\theta \in \Theta} \sum_{i=1}^n \log P_\theta(X_i = x_i) \text{ cas discret} \\ \sup_{\theta \in \Theta} \sum_{i=1}^n \log f(x_i\, \theta) \text{ cas continue} \end{cases}.$$

Définition 5.3 (Log-vraisemblance)**.** La fonction définie par

$$l(x_1, \dots, x_n, \theta) = \log L(\mathbf{X}, \theta) = \begin{cases} \sup_{\theta \in \Theta} \sum_{i=1}^n \log P_\theta(X_i = x_i) \text{ cas discret} \\ \sup_{\theta \in \Theta} \sum_{i=1}^n \log f(x_i\, \theta) \text{ cas continue} \end{cases}.$$

est appelée **log-vraisemblance**

Exemple 5.6 (Modèle de Bernouilli). $L(x_1,\dots,x_n,\theta)=\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$.
La log-vraisemblance vait

$$\begin{aligned}
l(x_1,\dots,x_n,\theta) &= \log L(x_1,\dots,x_n,\theta)\\
&= \log(\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i})\\
&= \log\theta^{\sum_{i=1}^n x_i}+\log((1-\theta)^{n-\sum_{i=1}^n x_i})\\
l(x_1,\dots,x_n,\theta) &= (\sum_{i=1}^n x_i)\log\theta+(n-\sum_{i=1}^n x_i)\log(1-\theta)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l(x_1,\dots,x_n,\theta)}{\partial\theta} &= \frac{\sum_{i=1}^n x_i}{\theta_n}-\frac{n-\sum_{i=1}^n x_i}{1-\theta}\\
&= \frac{(1-\theta)\sum_{i=1}^n x_i-(n-\sum_{i=1}^n x_i)\theta}{\theta(1-\theta)}\\
&= \frac{\sum_{i=1}^n x_i(n\theta)}{\theta(1-\theta)}
\end{aligned}$$

$$\frac{\partial l(x_1,\dots,x_n,\theta)}{\partial\theta}=0\Leftrightarrow\sum_{i=1}^n x_i(n\theta)=0$$

$$\theta^*=\frac{1}{n}\sum_{i=1}^n x_i$$

l est maximale lorsque $\theta=\theta^*=\frac{1}{n}\sum_{i=1}^n x_i$
$\hat{\theta_n}=\frac{1}{n}\sum_{i=1}^n X_i$ (variance empirique) est l'ENV de θ (variance empirique)

Exemple 5.7 (Modèle de Poisson). La vraisemblance $L(x_1,\dots,x_n,\theta)=e^{-n\theta}\frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n(x_i!)},x_i\in N$.
La log-vraisemblance :

$$\begin{aligned}
l(x_1,\dots,x_n,\theta) &= \log L(x_1,\dots,x_n,\theta)\\
&= \log e^{-n\theta}+\log\theta^{\sum_{i=1}^n x_i}-\log\prod_{i=1}^n(x_i!)\\
&= -n\theta+\sum_{i=1}^n x_i\log\theta-\log\prod_{i=1}^n(x_i!)
\end{aligned}$$

Dérivé :

$$\begin{aligned}
\frac{\partial l(x_1,\dots,x_n,\theta)}{\partial\theta} &= -n+\frac{\sum_{i=1}^n x_i}{\theta}\\
&= \frac{-n\theta+\sum_{i=1}^n x_i}{\theta}
\end{aligned}$$

$$\frac{\partial l(x_1,\dots,x_n,\theta)}{\partial\theta}=0\Leftrightarrow-n\theta+\sum_{i=1}^n x_i=0$$

$$\theta^*=\frac{1}{n}\sum_{i=1}^n x_i$$

Et :

$$\frac{\partial^2 l(x_1,\dots,x_n,\theta)}{\partial\theta^2}=\frac{1}{\theta^2}\sum_{i=1}^n x_i<0 \text{ car } x_i\in N.$$

l est maximale lorsque $\theta=\frac{1}{n}\sum_{i=1}^n x_i$
$\Rightarrow\hat{\theta_n}=\frac{1}{n}\sum_{i=1}^n x_i$ (moyenne empirique) est l'ENV pour θ

Exemple 5.8 (Loi normale)**.** $(x_1, x_2, \dots, x_n)$ réalisation de $(X_1, \dots, X_n)$ échantillon avec $X_1 \sim \mathcal{N}(\theta, 1)$.
La vraisemblance :
$$L(x_1, \dots, x_n, \theta) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2}.$$
La log-vraisemblance s'écrit :
$$l(x_1, \dots, x_n, \theta) = \log((2\pi)^n) - \frac{1}{2}\sum_{i=1}^n (x_i - \theta)^2.$$
$$\begin{aligned}\frac{\partial l(x_1, \dots, x_n, \theta)}{\partial \theta} &= +\frac{1}{2} * 2 \sum_{i=1}^n (x_i - \theta) \\ &= \sum_{i=1}^n (x_i) - n\theta \\ &= 0 \Leftrightarrow \theta^* = \frac{1}{n}\sum_{i=1}^n x_i\end{aligned}$$
$\theta_n = \frac{1}{n}\sum_{i=1}^n X_i$ est l'EMV pour θ

Exemple 5.9 (Loi random de toute à l'heure)**.** $(x_1, x_2, \dots, x_n)$ réalisation de $(X_1, \dots, X_n)$ échantillon avec X_1 de densité $f(x, \theta) = \frac{\theta}{x^2}\mathbb{1}_{x \geq \theta}()$.
La vraisemblance vaut
$$\frac{\theta^n}{\prod_{i=1}^n x_i^2}\mathbb{1}_{\min(x_i) \geq \theta} x.$$
Note. Si on dérive on vas tombé sur une dérivé partiel type $\frac{n}{\theta}$ dont on ne peux pas dérivé en theta encore

L atteint son maximum lorsque $\theta^* = \min_{1 \leq i \leq n}(x_i)$, $\hat{\theta_n} = \min_{1 \leq i \leq n}(X_i)$ est l'ENV pour θ

Exemple 5.10 (Modèle Gaussien)**.** $X_1, \dots, X_n$ échantillon de taille n où $X_i \sim \mathcal{N}(1, \sigma^2), E(X_i) = 1, var(X_i) = \sigma^2 = \theta$.
La densité de X_i vaut
$$f(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1-1)^2}.$$
La vraisemblance vaut
$$\begin{aligned}L(x_1, \dots, x_n, \sigma^2) = \prod_{i=1}^n f(x_i, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1-1)^2} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-1)^2}\end{aligned}$$
La log vraisemblance
$$\begin{aligned}l(x_1, \dots, x_n, \sigma^2) &= \log L((x_1, \dots, x_n, \sigma^2)) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - 1)^2\end{aligned}$$
La dérivé
$$\begin{aligned}\frac{\partial l}{\partial \sigma^2}(x_1, \dots, x_n, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i - 1)^2 \\ &= \frac{-n\theta^2 + \sum_{i=1}^n (x_i-1)^2}{2\sigma^4} \\ \frac{\partial l}{\partial \sigma^2}(x_1, \dots, x_n, \sigma^2) = 0 &\Leftrightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - 1)^2 = 0 \\ &\Leftrightarrow \sigma^{*2} = \frac{1}{n}\sum_{i=1}^n (x_i - 1)^2\end{aligned}$$
l est maximale lorsque $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - 1)^2$
Conclusion $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - 1)^2$ est l'EMV de σ^2

Exemple 5.11. $X_1, \ldots, X_n$ échantillon de taille n où $X_i \sim \mathcal{N}(m, \sigma^2), E(X_i) = m, var(X_i) = \sigma^2 = \theta$ avec m et σ^2 inconnu. $\theta = (m, \sigma^2)$
La densité de X_i vaut :

$$f(x, m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1 - m)^2}.$$

La vraisemblance vaut

$$\begin{aligned} L(x_1, \ldots, x_n, \sigma^2) = \prod_{i=1}^n f(x_i, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1 - m)^2} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m)^2} \end{aligned}$$

La log vraisemblance

$$\begin{aligned} l(x_1, \ldots, x_n, \sigma^2) &= \log L((x_1, \ldots, x_n, \sigma^2)) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m)^2 \end{aligned}$$

Trouver les points critiques :

$$\begin{cases} \frac{\partial l}{\partial m}(x_1 \ \ldots \ x_n \ \sigma^2) = 0 \\ \frac{\partial l}{\partial \sigma^2}(x_1 \ \ldots \ x_n \ \sigma^2) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n (x_i - m) = 0 \\ \frac{-n\sigma^2 + \sum_{i=1}^n (x_i - m)^2}{\sigma^4} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n (x_i - m) = 0 \\ \sum_{i=1}^n (x_i - m)^2 - n\sigma^2 = 0 \end{cases} \Leftrightarrow \begin{cases} m^* = \frac{1}{n}\sum_{i=1}^n x_i \\ \sigma^{*2} = \frac{1}{n}\sum_{i=1}^n (x_i - m^*)^2 \end{cases}.$$

Pour vérifier que (m^*, σ^{*2}) est un maximum, il faut montrer que

$$\det \begin{pmatrix} \frac{\partial^2 l(m*, \sigma^{*2})}{\partial m^2} & \frac{\partial^2 l(m*, \sigma^{*2})}{\partial m \partial \sigma^2} \\ \frac{\partial^2 l(m*, \sigma^{*2})}{\partial \sigma^2 \partial m} & \frac{\partial^2 l(m*, \sigma^{*2})}{\partial \sigma^4} \end{pmatrix} > 0 \text{ avec } \frac{\partial^2 l(m*, \sigma^{*2})}{\partial m^2} < 0, \frac{\partial^2 l(m*, \sigma^{*2})}{\partial \sigma^4} < 0.$$

$$\frac{\partial^2 l(m, \sigma^2)}{\partial m^2} = -\frac{n}{\sigma^2}.$$

$$\begin{aligned} \frac{\partial^2 l(m, \sigma^2)}{\partial \sigma^4} &= \frac{\partial}{\partial \sigma^2}[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i - m)^2] \\ &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (x_i - m)^2 \\ &= \frac{n}{2\sigma^{*4}} - \frac{1}{\sigma^{*6}}\sum_{i=1}^n (x_i - m^*)^2 \\ &= \frac{n}{2\sigma^{*4}} - \frac{1}{\sigma^{*6}} n\sigma^{*2} \\ &= \frac{n}{2\sigma^{*4}} - \frac{n\sigma^{*2}}{\sigma^{*6}} \\ &= \frac{n}{2\sigma^{*4}} - \frac{n}{\sigma^{*4}} \\ &= -\frac{n}{2\sigma^{*4}} \end{aligned}$$

$$\frac{\partial^2 l}{\partial m \partial \sigma^2}(m*; \sigma^{*2}) = 0 \text{ (A vérifier !).}$$

Conclusion

$$Hess(m*, \sigma^{*2}) = \begin{pmatrix} -\frac{n}{\sigma^{*2}} & 0 \\ 0 & -\frac{n}{2\sigma^{*4}} \end{pmatrix}, \det Hess(m^*, \sigma^{*2}) = \frac{n}{\sigma^{*2}} \frac{n}{2\sigma^{*4}} > 0, -\frac{n}{\sigma^{*2}} < 0 \text{ et } -\frac{n}{2\sigma^{*4}} < 0.$$

$\Rightarrow$ (m^*, σ^{*2}) est un maximum **local** de l
$(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2) = (\bar{X}_n, V_n^2)$ est l'ENV de (m, σ^2)

Remarque. 1. L'ENV peut être **biaisé**
2. L'ENV n'est pas toujours **unique**
3. L'ENV peut ne pas exister