Chapitre 3 : Estimation ponctuelle

- Définition d'un estimateur
- L'Estimateur par la Méthode des Moments (EMM)
- L'Estimateur de Maximum de Vraisemblance (EMV)
- Qualité d'un estimateur
- Propriétés des EMM et EMV

Introduction

Hypothèse 1 : les données $x_1, \ldots, x_n$ sont les réalisations de variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi.

Hypothèse 2 : les techniques de statistique descriptive ont permis d'adopter une famille de lois de probabilité précise pour la loi des X_i, mais que la valeur du ou des paramètres θ de cette loi est inconnue.

Introduction

Hypothèse 1 : les données $x_1, \dots, x_n$ sont les réalisations de variables aléatoires $X_1, \dots, X_n$ indépendantes et de même loi.

Hypothèse 2 : les techniques de statistique descriptive ont permis d'adopter une famille de lois de probabilité précise pour la loi des X_i, mais que la valeur du ou des paramètres θ de cette loi est inconnue.

Estimer θ : donner, au vu des observations $x_1, \dots, x_n$, une approximation ou une évaluation de θ que l'on espère la plus proche possible de la vraie valeur inconnue.

- **estimation ponctuelle** : une unique valeur vraisemblable pour θ
- **estimation ensembliste** ou **région de confiance** ensemble de valeurs vraisemblables pour θ

Introduction

Hypothèse 1 : les données $x_1, \ldots, x_n$ sont les réalisations de variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi.

Hypothèse 2 : les techniques de statistique descriptive ont permis d'adopter une famille de lois de probabilité précise pour la loi des X_i, mais que la valeur du ou des paramètres θ de cette loi est inconnue.

Estimer θ : donner, au vu des observations $x_1, \ldots, x_n$, une approximation ou une évaluation de θ que l'on espère la plus proche possible de la vraie valeur inconnue.

- **estimation ponctuelle** : une unique valeur vraisemblable pour θ
- **estimation ensembliste** ou **région de confiance** ensemble de valeurs vraisemblables pour θ

Notations : $F(x;\theta) =$ fonction de répartition des X_i.
v.a. discrètes : probabilités élémentaires $P(X = x;\theta)$
v.a. continues : densité $f(x;\theta)$

Définition d'un estimateur

Une **statistique** t est une fonction des observations $x_1, \ldots, x_n$:

$$t : \quad \begin{array}{l} \mathbf{R}^n \to \mathbf{R}^m \\ (x_1, \ldots, x_n) \mapsto t\,(x_1, \ldots, x_n) \end{array}$$

Exemples : $\bar{x}_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i,\ x_1^*,\ (x_1, x_3 + x_4, 2 \ln x_6)$ sont des statistiques.

Définition d'un estimateur

Une **statistique** t est une fonction des observations $x_1, \dots, x_n$:

$$\begin{aligned} t : \quad & \mathbb{R}^n \to \mathbb{R}^m \\ & (x_1, \dots, x_n) \mapsto t(x_1, \dots, x_n) \end{aligned}$$

Exemples : $\bar{x}_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i, \; x_1^*, \; (x_1, x_3 + x_4, 2 \ln x_6)$ sont des statistiques.

$t_n = t(x_1, \dots, x_n)$ est une réalisation de la variable aléatoire $T_n = t(X_1, \dots, X_n)$.

Définition d'un estimateur

Une **statistique** t est une fonction des observations $x_1, \ldots, x_n$:

$$t : \quad \begin{array}{l} \mathbb{R}^n \to \mathbb{R}^m \\ (x_1, \ldots, x_n) \mapsto t\,(x_1, \ldots, x_n) \end{array}$$

Exemples : $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i,\ x_1^*,\ (x_1, x_3 + x_4, 2 \ln x_6)$ sont des statistiques.

$t_n = t\,(x_1, \ldots, x_n)$ est une réalisation de la variable aléatoire $T_n = t\,(X_1, \ldots, X_n)$.

Un **estimateur** d'une grandeur θ est une statistique T_n à valeurs dans l'ensemble des valeurs possibles de θ.

Une **estimation** de θ est une réalisation t_n de l'estimateur T_n.

L'Estimateur par la Méthode des Moments (EMM)

Principe : estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

L'Estimateur par la Méthode des Moments (EMM)

Principe : estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Si $\theta = E[X]$, alors l'EMM de θ est $\tilde{\theta}_n = \bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

L'Estimateur par la Méthode des Moments (EMM)

Principe : estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Si $\theta = E[X]$, alors l'EMM de θ est $\tilde{\theta}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Pour $\theta \in \mathbb{R}$, si $E[X] = \varphi(\theta)$, où φ est une fonction inversible, alors $\theta = \varphi^{-1}(E[X])$ et l'EMM de θ est $\tilde{\theta}_n = \varphi^{-1}\left(\bar{X}_n\right)$.

L'Estimateur par la Méthode des Moments (EMM)

Principe : estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Si $\theta = E[X]$, alors l'EMM de θ est $\tilde{\theta}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Pour $\theta \in \mathbb{R}$, si $E[X] = \varphi(\theta)$, où φ est une fonction inversible, alors $\theta = \varphi^{-1}(E[X])$ et l'EMM de θ est $\tilde{\theta}_n = \varphi^{-1}\left(\bar{X}_n\right)$.

Si la loi des X_i a deux paramètres θ_1 et θ_2 tels que $(E[X], Var[X]) = \varphi(\theta_1, \theta_2)$, où φ est une fonction inversible, alors les EMM de θ_1 et θ_2 sont $\left(\tilde{\theta}_{1n}, \tilde{\theta}_{2n}\right) = \varphi^{-1}\left(\bar{X}_n, S_n^2\right)$, où $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2$.

L'Estimateur par la Méthode des Moments (EMM)

Principe : estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Si $\theta = E[X]$, alors l'EMM de θ est $\tilde{\theta}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Pour $\theta \in \mathrm{R}$, si $E[X] = \varphi(\theta)$, où φ est une fonction inversible, alors $\theta = \varphi^{-1}(E[X])$ et l'EMM de θ est $\tilde{\theta}_n = \varphi^{-1}\left(\bar{X}_n\right)$.

Si la loi des X_i a deux paramètres θ_1 et θ_2 tels que $(E[X], Var[X]) = \varphi(\theta_1, \theta_2)$, où φ est une fonction inversible, alors les EMM de θ_1 et θ_2 sont $\left(\tilde{\theta}_{1n}, \tilde{\theta}_{2n}\right) = \varphi^{-1}\left(\bar{X}_n, S_n^2\right)$, où $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2$.

etc...

Fonction de vraisemblance

Quand les observations sont toutes discrètes ou toutes continues, on appelle **fonction de vraisemblance** (ou plus simplement vraisemblance) pour l'échantillon $x_1, \ldots, x_n$, la fonction du paramètre θ :

$$\mathcal{L}(\theta; x_1, \ldots, x_n) = \begin{cases} P(X_1 = x_1, \ldots, X_n = x_n; \theta) & \text{si les } X_i \text{ sont discrètes} \\ \\ f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) & \text{si les } X_i \text{ sont continues} \end{cases}$$

Fonction de vraisemblance

Quand les observations sont toutes discrètes ou toutes continues, on appelle **fonction de vraisemblance** (ou plus simplement vraisemblance) pour l'échantillon $x_1, \ldots, x_n$, la fonction du paramètre θ :

$$\mathcal{L}(\theta; x_1, \ldots, x_n) = \begin{cases} P(X_1 = x_1, \ldots, X_n = x_n; \theta) & \text{si les } X_i \text{ sont discrètes} \\ \\ f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) & \text{si les } X_i \text{ sont continues} \end{cases}$$

Quand les X_i sont indépendantes et de même loi, la vraisemblance s'écrit :

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

GRENOBLE INP UGA Ensimag

Maximum de vraisemblance : exemple introductif

$n = 1$. On sait que X_1 est de loi binomiale $\mathcal{B}(15, p)$, avec p inconnu.

On observe $x_1 = 5$ et on cherche à estimer p.

La fonction de vraisemblance est la probabilité d'avoir observé un 5 quand la valeur du paramètre est p :

$$\mathcal{L}(p; 5) = P(X_1 = 5; p) = \binom{15}{5}\ p^5\,(1-p)^{15-5}$$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\mathcal{L}(p;5)$	0.01	0.10	0.21	0.19	0.09	0.02	0.003	10^{-4}	$2\,10^{-7}$

Maximum de vraisemblance : exemple introductif

$n = 1$. On sait que X_1 est de loi binomiale $\mathcal{B}(15, p)$, avec p inconnu.

On observe $x_1 = 5$ et on cherche à estimer p.

La fonction de vraisemblance est la probabilité d'avoir observé un 5 quand la valeur du paramètre est p :

$$\mathcal{L}(p;5) = P(X_1 = 5; p) = \binom{15}{5}\, p^5\, (1-p)^{15-5}$$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\mathcal{L}(p;5)$	0.01	0.10	0.21	0.19	0.09	0.02	0.003	10^{-4}	$2\,10^{-7}$

La valeur la plus vraisemblable de p est celle qui maximise la fonction de vraisemblance.

$\Rightarrow \hat{p}_n = 1/3$.

L'Estimateur de Maximum de Vraisemblance (EMV)

On fait comme si c'était l'éventualité la plus probable qui s'était produite au cours de l'expérience.

L'**estimation de maximum de vraisemblance** de θ est la valeur $\hat{\theta}_n$ de θ qui rend maximale la fonction de vraisemblance $\mathcal{L}(\theta; x_1, \ldots, x_n)$.

L'**estimateur de maximum de vraisemblance (EMV)** de θ est la variable aléatoire correspondante.

L'Estimateur de Maximum de Vraisemblance (EMV)

On fait comme si c'était l'éventualité la plus probable qui s'était produite au cours de l'expérience.

L'**estimation de maximum de vraisemblance** de θ est la valeur $\hat{\theta}_n$ de θ qui rend maximale la fonction de vraisemblance $\mathcal{L}(\theta; x_1, \ldots, x_n)$.
L'**estimateur de maximum de vraisemblance (EMV)** de θ est la variable aléatoire correspondante.

Quand $\theta = (\theta_1, \ldots, \theta_d) \in \mathbf{R}^d$ et que toutes les dérivées partielles ci-dessous existent, $\hat{\theta}_n$ est solution du système d'équations appelées **équations de vraisemblance** :

$$\forall j \in \{1, \ldots, d\}, \quad \frac{\partial}{\partial \theta_j} \ln \mathcal{L}(\theta; x_1, \ldots, x_n) = 0$$

Exemple des durées de vie d'ampoules

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
91.6	35.7	251.3	24.3	5.4	67.3	170.9	9.5	118.4	57.1

Hypothèse : $x_1, \ldots, x_n$ sont des réalisations de variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi $\exp(\lambda)$.

EMM = EMV :

- L'**estimateur** de λ est $\hat{\lambda}_n = 1/\overline{X}_n$.
- L'**estimation** de λ pour les durées de vie d'ampoules est $\hat{\lambda}_n = 1/\overline{x}_n = 0.012$.

Qualité d'un estimateur

On estime un paramètre $\theta \in \mathbb{R}$ à l'aide d'un estimateur T_n.

- Le **biais** de T_n est :

$$Biais[T_n] = E[T_n - \theta] = E[T_n] - \theta$$

Qualité d'un estimateur

On estime un paramètre $\theta \in \mathbb{R}$ à l'aide d'un estimateur T_n.

- Le **biais** de T_n est :

$$Biais[T_n] = E[T_n - \theta] = E[T_n] - \theta$$

- L'**erreur quadratique moyenne** est :

$$EQM(T_n) = E\left[(T_n - \theta)^2\right] = Var[T_n] + Biais[T_n]^2$$

Qualité d'un estimateur

On estime un paramètre $\theta \in \mathbb{R}$ à l'aide d'un estimateur T_n.

- Le **biais** de T_n est :

$$Biais[T_n] = E[T_n - \theta] = E[T_n] - \theta$$

- L'**erreur quadratique moyenne** est :

$$EQM(T_n) = E\left[(T_n - \theta)^2\right] = Var[T_n] + Biais[T_n]^2$$

T_n est un **estimateur sans biais** de θ si et seulement si $E[T_n] = \theta$.

Qualité d'un estimateur

On estime un paramètre $\theta \in \mathbb{R}$ à l'aide d'un estimateur T_n.

- Le **biais** de T_n est :

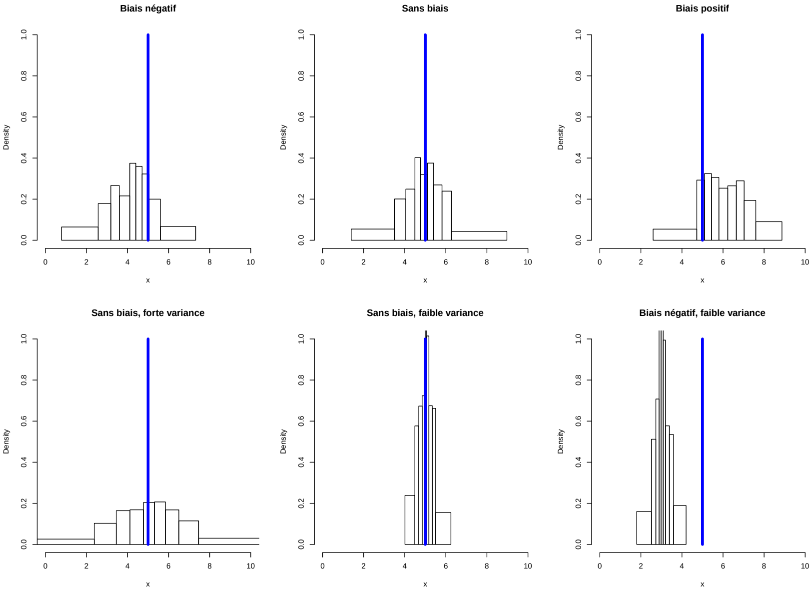
$$Biais[T_n] = E[T_n - \theta] = E[T_n] - \theta$$

- L'**erreur quadratique moyenne** est :

$$EQM(T_n) = E\left[(T_n - \theta)^2\right] = Var[T_n] + Biais[T_n]^2$$

T_n est un **estimateur sans biais** de θ si et seulement si $E[T_n] = \theta$.

On considèrera que le meilleur estimateur possible de θ est un **estimateur sans biais et de variance minimale (ESBVM)**.



Biais négatif
Sans biais
Biais positif
Sans biais, forte variance
Sans biais, faible variance
Biais négatif, faible variance
Density
x
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10

Convergences 1

T_n est un **estimateur convergent** de θ si et seulement si T_n converge, en un certain sens, vers θ quand n tend vers l'infini.

Convergences 1

T_n est un **estimateur convergent** de θ si et seulement si T_n converge, en un certain sens, vers θ quand n tend vers l'infini.

Une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ **converge presque sûrement** vers la variable aléatoire X si et seulement si

$$P\left(\left\{\omega;\lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Convergences 1

T_n est un **estimateur convergent** de θ si et seulement si T_n converge, en un certain sens, vers θ quand n tend vers l'infini.

Une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ **converge presque sûrement** vers la variable aléatoire X si et seulement si

$$P\left(\left\{\omega; \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ **converge en probabilité** vers la variable aléatoire X si et seulement si

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

Convergences 2

Une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ **converge en loi** vers la loi de probabilité de fonction de répartition F si et seulement si

$$\forall x, \lim_{n\to\infty} F_{X_n}(x) = F(x)$$

Convergences 2

Une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ **converge en loi** vers la loi de probabilité de fonction de répartition F si et seulement si

$$\forall x, \lim_{n\to\infty} F_{X_n}(x) = F(x)$$

Une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ de L^2 **converge en moyenne quadratique** vers la variable aléatoire X si et seulement si

$$\lim_{n\to\infty} E\left[|X_n - X|^2\right] = 0$$

Théorème Central-Limite

Soit $\{X_n\}_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, d'espérance $E[X]$ et d'écart-type $\sigma[X] = \sqrt{Var[X]}$ finis. Pour tout $n \geq 1$, on pose :

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nE[X]}{\sqrt{nVar[X]}} = \sqrt{n}\,\frac{\bar{X}_n - E[X]}{\sigma[X]}$$

Alors la suite $\{Z_n\}_{n \geq 1}$ converge en loi vers la loi normale centrée-réduite, ce qui s'écrit :

$$\sqrt{n}\,\frac{\bar{X}_n - E[X]}{\sigma[X]} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Théorème Central-Limite

Concrètement :

- La loi de toute variable aléatoire égale à la somme d'un nombre "suffisamment grand" de variables aléatoires indépendantes et de même loi est approximativement une loi normale.
- Pour n grand, $\sum_{i=1}^{n} X_i$ est approximativement de loi $\mathcal{N}(nE[X], nVar[X])$.

Loi des grands nombres

Soit $\{X_n\}_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, d'espérance $E[X]$. Soit $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

Alors la suite $\{\bar{X}_n\}_{n\geq 1}$ converge presque sûrement vers $E[X]$, ce qui s'écrit :

$$\bar{X}_n \xrightarrow{ps} E[X]$$

Loi des grands nombres

Soit $\{X_n\}_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, d'espérance $E[X]$. Soit $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Alors la suite $\{\bar{X}_n\}_{n\geq 1}$ converge presque sûrement vers $E[X]$, ce qui s'écrit :

$$\bar{X}_n \xrightarrow{ps} E[X]$$

Concrètement :

- Quand on fait un très grand nombre d'expériences identiques et indépendantes, la moyenne des réalisations de la variable aléatoire à laquelle on s'intéresse tend vers l'espérance de sa loi.
- Ce résultat permet de justifier l'idée naturelle d'estimer une espérance par une moyenne empirique et une probabilité par une proportion.

Quantité d'information

Pour $\theta \in \mathbf{R}$, si la loi des observations vérifie certaines conditions de régularité, on appelle **quantité d'information** (de Fisher) sur θ apportée par l'échantillon $x_1, \ldots, x_n$, la quantité :

$$\mathcal{I}_n(\theta) = Var\left[\frac{\partial}{\partial\theta} \ln \mathcal{L}(\theta; X_1, \ldots, X_n)\right]$$

Quantité d'information

Pour $\theta \in \mathbf{R}$, si la loi des observations vérifie certaines conditions de régularité, on appelle **quantité d'information** (de Fisher) sur θ apportée par l'échantillon $x_1, \ldots, x_n$, la quantité :

$$\mathcal{I}_n(\theta) = Var\left[\frac{\partial}{\partial\theta} \ln \mathcal{L}(\theta; X_1, \ldots, X_n)\right]$$

Inégalité de Fréchet-Darmois-Cramer-Rao (FDCR) : Si la loi des observations vérifie les conditions de régularité, alors pour tout estimateur T_n de θ, on a :

$$Var(T_n) \geq \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta)}$$

Efficacité d'un estimateur

On appelle **efficacité** d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta) Var\left[T_n\right]}$$

Efficacité d'un estimateur

On appelle **efficacité** d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta) Var\left[T_n\right]}$$

- $0 \leq Eff(T_n) \leq 1$.

Efficacité d'un estimateur

On appelle **efficacité** d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta) Var[T_n]}$$

- $0 \leq Eff(T_n) \leq 1$.
- T_n est dit un estimateur **efficace** si et seulement si $Eff(T_n) = 1$.

Efficacité d'un estimateur

On appelle **efficacité** d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta) Var[T_n]}$$

- $0 \leq Eff(T_n) \leq 1$.
- T_n est dit un estimateur **efficace** si et seulement si $Eff(T_n) = 1$.
- T_n est dit **asymptotiquement efficace** si et seulement si $\lim_{n \to +\infty} Eff(T_n) = 1$.

Efficacité d'un estimateur

On appelle **efficacité** d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta) Var[T_n]}$$

- $0 \leq Eff(T_n) \leq 1$.
- T_n est dit un estimateur **efficace** si et seulement si $Eff(T_n) = 1$.
- T_n est dit **asymptotiquement efficace** si et seulement si $\lim_{n\to+\infty} Eff(T_n) = 1$.
- La quantité $\frac{1}{\mathcal{I}_n(\theta)}$ est appelée la **borne de Cramer-Rao**.

Efficacité d'un estimateur

On appelle **efficacité** d'un estimateur T_n la quantité :

$$Eff(T_n) = \frac{\left[\frac{\partial}{\partial\theta} E[T_n]\right]^2}{\mathcal{I}_n(\theta) Var[T_n]}$$

- $0 \leq Eff(T_n) \leq 1$.
- T_n est dit un estimateur **efficace** si et seulement si $Eff(T_n) = 1$.
- T_n est dit **asymptotiquement efficace** si et seulement si $\lim\limits_{n \to +\infty} Eff(T_n) = 1$.
- La quantité $\dfrac{1}{\mathcal{I}_n(\theta)}$ est appelée la **borne de Cramer-Rao**.
- **Si un estimateur sans biais est efficace, sa variance est égale à la borne de Cramer-Rao, donc c'est forcément un ESBVM.**

Propriétés des EMM - Moyenne empirique

On estime l'espérance $E[X]$ par la moyenne empirique $\bar{X}_n$.

La loi des grands nombres dit que $\bar{X}_n$ converge presque sûrement vers $E[X]$.

Propriétés des EMM - Moyenne empirique

On estime l'espérance $E[X]$ par la moyenne empirique $\bar{X}_n$.

La loi des grands nombres dit que $\bar{X}_n$ converge presque sûrement vers $E[X]$.

$$E[\bar{X}_n] = E[X] \qquad\qquad Var[\bar{X}_n] = \frac{Var[X]}{n}$$

Propriétés des EMM - Moyenne empirique

On estime l'espérance $E[X]$ par la moyenne empirique $\bar{X}_n$.

La loi des grands nombres dit que $\bar{X}_n$ converge presque sûrement vers $E[X]$.

$$E[\bar{X}_n] = E[X] \qquad\qquad Var[\bar{X}_n] = \frac{Var[X]}{n}$$

Donc la moyenne empirique $\bar{X}_n$ est un estimateur sans biais et convergent en moyenne quadratique de $E[X]$.

Propriétés des EMM - Variance empirique

On estime la variance $Var[X]$ par la variance empirique

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

Propriétés des EMM - Variance empirique

On estime la variance $Var[X]$ par la variance empirique

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

$$E[S_n^2] = \frac{n-1}{n}\, Var[X] \neq Var[X]$$

Donc la variance empirique S_n^2 est un estimateur biaisé de $Var[X]$.

Propriétés des EMM - Variance empirique

On estime la variance $Var[X]$ par la variance empirique

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

$$E[S_n^2] = \frac{n-1}{n}\, Var[X] \neq Var[X]$$

Donc la variance empirique S_n^2 est un estimateur biaisé de $Var[X]$.

Mais ${S'}_n^2 = \frac{n}{n-1}\, S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ est un estimateur sans biais et convergent en moyenne quadratique de $Var[X]$.

${S'}_n^2$ est appelée **variance estimée** de l'échantillon.

Propriétés des EMV

- $\hat{\theta}_n$ converge presque sûrement vers θ.

Propriétés des EMV

- $\hat{\theta}_n$ converge presque sûrement vers θ.
- $\sqrt{\mathcal{I}_n(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ ou $\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \dfrac{1}{\mathcal{I}_1(\theta)}\right)$.

Propriétés des EMV

- $\hat{\theta}_n$ converge presque sûrement vers θ.
- $\sqrt{\mathcal{I}_n(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ ou $\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \dfrac{1}{\mathcal{I}_1(\theta)}\right)$.

  $\Rightarrow$ quand n est grand, $\hat{\theta}_n$ est approximativement de loi $\mathcal{N}\left(\theta, \dfrac{1}{\mathcal{I}_n(\theta)}\right)$.

  $\Rightarrow$ $\hat{\theta}_n$ est asymptotiquement gaussien, sans biais et efficace.

Propriétés des EMV

- $\hat{\theta}_n$ converge presque sûrement vers θ.
- $\sqrt{\mathcal{I}_n(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ ou $\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \dfrac{1}{\mathcal{I}_1(\theta)}\right)$.

  $\Rightarrow$ quand n est grand, $\hat{\theta}_n$ est approximativement de loi $\mathcal{N}\left(\theta, \dfrac{1}{\mathcal{I}_n(\theta)}\right)$.

  $\Rightarrow$ $\hat{\theta}_n$ est asymptotiquement gaussien, sans biais et efficace.
- Si $\hat{\theta}_n$ est l'EMV de θ, alors $\varphi(\hat{\theta}_n)$ est l'EMV de $\varphi(\theta)$.

Propriétés des EMV

- $\hat{\theta}_n$ converge presque sûrement vers θ.
- $\sqrt{\mathcal{I}_n(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ ou $\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \dfrac{1}{\mathcal{I}_1(\theta)}\right)$.

  $\Rightarrow$ quand n est grand, $\hat{\theta}_n$ est approximativement de loi $\mathcal{N}\left(\theta, \dfrac{1}{\mathcal{I}_n(\theta)}\right)$.

  $\Rightarrow$ $\hat{\theta}_n$ est asymptotiquement gaussien, sans biais et efficace.
- Si $\hat{\theta}_n$ est l'EMV de θ, alors $\varphi(\hat{\theta}_n)$ est l'EMV de $\varphi(\theta)$.
- **Méthode delta** : si φ est dérivable, on a :

$$\sqrt{n}\left[\varphi(\hat{\theta}_n) - \varphi(\theta)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\varphi'(\theta)^2}{\mathcal{I}_1(\theta)}\right)$$

  $\Rightarrow$ quand n est grand, $\varphi(\hat{\theta}_n)$ est approximativement de loi

$$\mathcal{N}\left(\varphi(\theta), \frac{\varphi'(\theta)^2}{\mathcal{I}_n(\theta)}\right)$$