

Statistique & Apprentissage

Paul-Henry Cournède

Amphi 2

(Rappels)

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable.

V.A., Donnée, Loi de probabilité d'une V.A., Echantillon I.I.D.

- Une **variable aléatoire** X est une application mesurable de Ω dans \mathcal{X} .

- Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une **réalisation** de la v.a. aussi appelée **observation** ou **donnée**.

- P_X mesurable, est la **loi de probabilité de X** , $X \sim P_X$:

$$P_X : \mathcal{A} \rightarrow [0; 1] \\ A \mapsto \mathbb{P}(X^{-1}(A))$$

- Si P_X est dominée, P_X admet une densité p_X .

Variables aléatoires **continues** \implies mesure de référence μ = mesure de **Lebesgue**.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x) \lambda_d(dx)$$

Variables aléatoires **discrètes** \implies mesure de référence μ = mesure de **comptage**.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x) K(dx) = \sum_{\alpha \in A} P_X(\alpha)$$

- Si X_1, \dots, X_N sont de variables aléatoires indépendantes de même loi P_X , (X_1, \dots, X_N) est dit d'**échantillon aléatoire indépendant identiquement distribué** (i.i.d.).

Le vecteur d'observations associées (x_1, \dots, x_N) est dit échantillon d'observations.

I.2 - La Méthode Empirique

Statistique, mesure empirique, méthode empirique

- Soit (X_1, \dots, X_N) , un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.

Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.

Alors $T_N = T(X_1, \dots, X_N)$ est une **variable aléatoire appelée statistique de l'échantillon**.

- Soit (X_1, \dots, X_N) , un échantillon aléatoire **i.i.d.** de loi P_X .

La **mesure empirique** est une statistique de l'échantillon à valeurs dans l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ définie par :

$$\hat{P}(X_1, \dots, X_N) = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$$

où δ_{X_i} est la mesure de Dirac en X_i .

- La **méthode empirique** vise à approcher les caractéristiques de P_X inconnue, à partir d'un échantillon d'observations (x_1, \dots, x_N) pour des v.a. i.i.d. de loi P_X .

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors : la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \dots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

Exemples : L'espérance $\mathbb{E}(X)$ est approchée par la moyenne empirique $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$; les moments empiriques ; la variance empirique S^2 ; la fonction de répartition empirique...

(Rappels)

I.3 - Modèles Statistiques Paramétriques

• Un **modèle statistique paramétrique** sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilité sur $(\mathcal{X}, \mathcal{A})$ paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$:

$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$, ou $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$ si densités

II - Estimation Paramétrique

• Soit $X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$

Soit (X_1, \dots, X_N) un échantillon i.i.d. pour une loi P_{θ^*} . Un **estimateur** de θ^* est une statistique $T(X_1, \dots, X_N)$ à valeurs dans Θ visant à estimer θ^* .

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

On considère la fonctionnelle G sur l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \dots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit (x_1, \dots, x_N) un échantillon d'observations i.i.d. pour X , on calcule G en la mesure empirique

$\hat{g} = G(\hat{P}(x_1, \dots, x_N)) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$.

II.1.b - Maximum de vraisemblance

• $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X .

La fonction : $\theta \in \Theta, \theta \mapsto p_\theta(x)$ est appelée **vraisemblance du paramètre θ en x** , on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations (x_1, x_2, \dots, x_N) pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_N) = \prod_{i=1}^N p_\theta(x_i),$$

• Pour une observation ou un échantillon d'observations x , si $\hat{\theta}(x)$ maximise la vraisemblance $\mathcal{L}(\theta; x)$ sur Θ $\hat{\theta}(x)$ sera appelée **estimation du maximum de vraisemblance** et l'estimateur associé $\hat{\theta}(X)$, estimateur du maximum de vraisemblance (EMV).

On note :

$$\hat{\theta}(x) = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta; x).$$

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient
 $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$:
les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions
 $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois
continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient
 $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$:
les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions
 $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois
continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

On suppose pour la suite que \mathcal{M}_Θ vérifie ces conditions de régularité pour simplifier la présentation. (Elles ne sont cependant pas tout le temps nécessaires).

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$: les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

On suppose pour la suite que \mathcal{M}_Θ vérifie ces conditions de régularité pour simplifier la présentation. (Elles ne sont cependant pas tout le temps nécessaires).

Définition

Soit $X \sim p_{\theta^*}, p_{\theta^*} \in \mathcal{M}_\Theta$, vérifiant les conditions de régularité. On appelle Score du modèle la variable aléatoire :

$$S_\theta(X) = \nabla_\theta \ln(p_\theta(X)) = \left(\frac{\partial \ln(p_\theta(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_\theta(X))}{\partial \theta_p} \right)^T$$

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$: les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

On suppose pour la suite que \mathcal{M}_Θ vérifie ces conditions de régularité pour simplifier la présentation. (Elles ne sont cependant pas tout le temps nécessaires).

Définition

Soit $X \sim p_{\theta^*}, p_{\theta^*} \in \mathcal{M}_\Theta$, vérifiant les conditions de régularité. On appelle Score du modèle la variable aléatoire :

$$S_\theta(X) = \nabla_\theta \ln(p_\theta(X)) = \left(\frac{\partial \ln(p_\theta(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_\theta(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_\theta(S_\theta(X)) = 0$.

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$: les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

On suppose pour la suite que \mathcal{M}_Θ vérifie ces conditions de régularité pour simplifier la présentation. (Elles ne sont cependant pas tout le temps nécessaires).

Définition

Soit $X \sim p_{\theta^*}, p_{\theta^*} \in \mathcal{M}_\Theta$, vérifiant les conditions de régularité. On appelle Score du modèle la variable aléatoire :

$$S_\theta(X) = \nabla_\theta \ln(p_\theta(X)) = \left(\frac{\partial \ln(p_\theta(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_\theta(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_\theta(S_\theta(X)) = 0$.

[Remarque Notation : $\mathbb{E}_\theta \Rightarrow$ Espérance sous la loi P_θ : en intégrant par rapport à la mesure P_θ , ici $p_\theta(x)\nu(dx)$]

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$: les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

On suppose pour la suite que \mathcal{M}_Θ vérifie ces conditions de régularité pour simplifier la présentation. (Elles ne sont cependant pas tout le temps nécessaires).

Définition

Soit $X \sim p_{\theta^*}, p_{\theta^*} \in \mathcal{M}_\Theta$, vérifiant les conditions de régularité. On appelle Score du modèle la variable aléatoire :

$$S_\theta(X) = \nabla_\theta \ln(p_\theta(X)) = \left(\frac{\partial \ln(p_\theta(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_\theta(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_\theta(S_\theta(X)) = 0$.

[Remarque Notation : $\mathbb{E}_\theta \Rightarrow$ Espérance sous la loi P_θ : en intégrant par rapport à la mesure P_θ , ici $p_\theta(x)\nu(dx)$]

Pour chaque composante : $[S_\theta(X)]_k = \frac{\partial \ln(p_\theta(X))}{\partial \theta_k}$:

II.2 - Information de Fisher

La qualité de l'estimation va dépendre de l'information contenue dans les données pour le modèle statistique.

⇒ L'information de Fisher est une mesure de cette information.

Conditions de régularité

Soit un modèle statistique $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

- (C1) Θ est ouvert et les lois de \mathcal{M}_Θ vérifient $p_\theta(x) > 0 \Leftrightarrow p_{\theta'}(x) > 0, \forall x \in \mathcal{X}, \forall \theta, \theta' \in \Theta$: les lois ont même support, noté \mathcal{S} .
- (C2) $\forall \theta \in \Theta, \forall k, 1 \leq k \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} p_\theta(x) \nu(dx) = \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) .$$

- (C3) Pour ν presque tout $x \in \mathcal{S}$, les fonctions $\theta \mapsto p_\theta(x)$ et $\theta \mapsto \ln p_\theta(x)$ sont deux fois continûment différentiables sur Θ .
- (C4) $\forall \theta \in \Theta, \forall i, j, 1 \leq i, j \leq p$, nous pouvons dériver sous l'intégrale :

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_j} \nu(dx) = \int_{\mathcal{S}} \frac{\partial^2 p_\theta(x)}{\partial \theta_i \partial \theta_j} \nu(dx) .$$

On suppose pour la suite que \mathcal{M}_Θ vérifie ces conditions de régularité pour simplifier la présentation. (Elles ne sont cependant pas tout le temps nécessaires).

Définition

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$, vérifiant les conditions de régularité. On appelle **Score** du modèle la variable aléatoire :

$$S_\theta(X) = \nabla_\theta \ln(p_\theta(X)) = \left(\frac{\partial \ln(p_\theta(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_\theta(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_\theta(S_\theta(X)) = 0$.

[Remarque Notation : $\mathbb{E}_\theta \Rightarrow$ Espérance sous la loi P_θ : en intégrant par rapport à la mesure P_θ , ici $p_\theta(x)\nu(dx)$]

Pour chaque composante : $[S_\theta(X)]_k = \frac{\partial \ln(p_\theta(X))}{\partial \theta_k}$:

$$\begin{aligned} \mathbb{E}_\theta([S_\theta(X)]_k) &= \int_{\mathcal{S}} \frac{\partial \ln(p_\theta(x))}{\partial \theta_k} p_\theta(x) \nu(dx) \\ &= \int_{\mathcal{S}} \frac{\partial p_\theta(x)}{\partial \theta_k} \nu(dx) \\ &= \frac{\partial}{\partial \theta_k} \left(\underbrace{\int_{\mathcal{S}} p_\theta(x) \nu(dx)}_{=1} \right) = 0 \end{aligned}$$

□

Définition

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_{\Theta}$, vérifiant les conditions de régularité. On appelle **Score** du modèle la variable aléatoire :

$$S_{\theta}(X) = \nabla_{\theta} \ln(p_{\theta}(X)) = \left(\frac{\partial \ln(p_{\theta}(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$.

[Remarque Notation : $\mathbb{E}_{\theta} \Rightarrow$ Espérance sous la loi P_{θ} : en intégrant par rapport à la mesure P_{θ} , ici $p_{\theta}(x)\nu(dx)$]

Pour chaque composante : $[S_{\theta}(X)]_k = \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_k}$:

$$\begin{aligned} \mathbb{E}_{\theta}([S_{\theta}(X)]_k) &= \int_{\mathcal{S}} \frac{\partial \ln(p_{\theta}(x))}{\partial \theta_k} p_{\theta}(x) \nu(dx) \\ &= \int_{\mathcal{S}} \frac{\partial p_{\theta}(x)}{\partial \theta_k} \nu(dx) \\ &= \frac{\partial}{\partial \theta_k} \underbrace{\left(\int_{\mathcal{S}} p_{\theta}(x) \nu(dx) \right)}_{=1} = 0 \end{aligned}$$

□

Définition

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_{\Theta}$, vérifiant les conditions de régularité. On appelle **Score** du modèle la variable aléatoire :

$$S_{\theta}(X) = \nabla_{\theta} \ln(p_{\theta}(X)) = \left(\frac{\partial \ln(p_{\theta}(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$.

[Remarque Notation : $\mathbb{E}_{\theta} \Rightarrow$ Espérance sous la loi P_{θ} : en intégrant par rapport à la mesure P_{θ} , ici $p_{\theta}(x)\nu(dx)$]

Pour chaque composante : $[S_{\theta}(X)]_k = \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_k}$:

$$\begin{aligned} \mathbb{E}_{\theta}([S_{\theta}(x)]_k) &= \int_{\mathcal{S}} \frac{\partial \ln(p_{\theta}(x))}{\partial \theta_k} p_{\theta}(x) \nu(dx) \\ &= \int_{\mathcal{S}} \frac{\partial p_{\theta}(x)}{\partial \theta_k} \nu(dx) \\ &= \frac{\partial}{\partial \theta_k} \underbrace{\left(\int_{\mathcal{S}} p_{\theta}(x) \nu(dx) \right)}_{=1} = 0 \end{aligned}$$

□

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Définition

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_{\Theta}$, vérifiant les conditions de régularité. On appelle **Score** du modèle la variable aléatoire :

$$S_{\theta}(X) = \nabla_{\theta} \ln(p_{\theta}(X)) = \left(\frac{\partial \ln(p_{\theta}(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$.

[Remarque Notation : $\mathbb{E}_{\theta} \Rightarrow$ Espérance sous la loi P_{θ} : en intégrant par rapport à la mesure P_{θ} , ici $p_{\theta}(x)\nu(dx)$]

Pour chaque composante : $[S_{\theta}(X)]_k = \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_k}$:

$$\begin{aligned} \mathbb{E}_{\theta}([S_{\theta}(x)]_k) &= \int_{\mathcal{S}} \frac{\partial \ln(p_{\theta}(x))}{\partial \theta_k} p_{\theta}(x) \nu(dx) \\ &= \int_{\mathcal{S}} \frac{\partial p_{\theta}(x)}{\partial \theta_k} \nu(dx) \\ &= \frac{\partial}{\partial \theta_k} \underbrace{\left(\int_{\mathcal{S}} p_{\theta}(x) \nu(dx) \right)}_{=1} = 0 \end{aligned}$$

□

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta} [S_{\theta}(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta} \left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j} \right]$$

Définition

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_{\Theta}$, vérifiant les conditions de régularité. On appelle **Score** du modèle la variable aléatoire :

$$S_{\theta}(X) = \nabla_{\theta} \ln(p_{\theta}(X)) = \left(\frac{\partial \ln(p_{\theta}(X))}{\partial \theta_1}, \dots, \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_p} \right)^T$$

Proposition

Le Score est un vecteur aléatoire centré : $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$.

[Remarque Notation : $\mathbb{E}_{\theta} \Rightarrow$ Espérance sous la loi P_{θ} : en intégrant par rapport à la mesure P_{θ} , ici $p_{\theta}(x)\nu(dx)$]

Pour chaque composante : $[S_{\theta}(X)]_k = \frac{\partial \ln(p_{\theta}(X))}{\partial \theta_k}$:

$$\begin{aligned} \mathbb{E}_{\theta}([S_{\theta}(x)]_k) &= \int_{\mathcal{S}} \frac{\partial \ln(p_{\theta}(x))}{\partial \theta_k} p_{\theta}(x) \nu(dx) \\ &= \int_{\mathcal{S}} \frac{\partial p_{\theta}(x)}{\partial \theta_k} \nu(dx) \\ &= \frac{\partial}{\partial \theta_k} \underbrace{\left(\int_{\mathcal{S}} p_{\theta}(x) \nu(dx) \right)}_{=1} = 0 \end{aligned}$$

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta} [S_{\theta}(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta} \left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j} \right]$$

Proposition

Soient $X \sim p_{\theta}$, $Y \sim q_{\theta}$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_{\theta} : (x, y) \mapsto p_{\theta}(x)q_{\theta}(y)$.

□

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta}[S_{\theta}(X)]]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta} \left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j} \right]$$

Proposition

Soient $X \sim p_{\theta}$, $Y \sim q_{\theta}$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_{\theta} : (x, y) \mapsto p_{\theta}(x)q_{\theta}(y)$.

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta}[S_{\theta}(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta}\left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j}\right]$$

Proposition

Soient $X \sim p_{\theta}$, $Y \sim q_{\theta}$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_{\theta} : (x, y) \mapsto p_{\theta}(x)q_{\theta}(y)$.

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta}[S_{\theta}(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta} \left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j} \right]$$

Proposition

Soient $X \sim p_{\theta}$, $Y \sim q_{\theta}$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_{\theta} : (x, y) \mapsto p_{\theta}(x)q_{\theta}(y)$.

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

\Rightarrow La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta}[S_{\theta}(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta}\left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j}\right]$$

Proposition

Soient $X \sim p_{\theta}$, $Y \sim q_{\theta}$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_{\theta} : (x, y) \mapsto p_{\theta}(x)q_{\theta}(y)$.

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

\Rightarrow La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = N I(\theta)$$

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_\theta (S_\theta (X))$$

Comme $\mathbb{E}_\theta (S_\theta (X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_\theta [S_\theta (X) S_\theta (X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_\theta [\nabla_\theta [S_\theta (X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \ln p_\theta(X)}{\partial \theta_i \partial \theta_j} \right]$$

Proposition

Soient $X \sim p_\theta$, $Y \sim q_\theta$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_\theta : (x, y) \mapsto p_\theta(x)q_\theta(y)$.

En effet :

$$\begin{aligned} S_\theta (X, Y) &= \nabla_\theta \ln(f_\theta(X, Y)) = \nabla_\theta \ln(p_\theta(X)q_\theta(Y)) \\ &= \nabla_\theta \ln(p_\theta(X)) + \nabla_\theta \ln(q_\theta(Y)) \\ &= S_\theta (X) + S_\theta (Y) \end{aligned}$$

$S_\theta (X)$ et $S_\theta (Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

\Rightarrow La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_θ et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = N I(\theta)$$

\Rightarrow N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_\theta (S_\theta (X))$$

Comme $\mathbb{E}_\theta (S_\theta (X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_\theta [S_\theta (X) S_\theta (X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_\theta [\nabla_\theta [S_\theta (X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \ln p_\theta(X)}{\partial \theta_i \partial \theta_j} \right]$$

Proposition

Soient $X \sim p_\theta$, $Y \sim q_\theta$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_\theta : (x, y) \mapsto p_\theta(x)q_\theta(y)$.

En effet :

$$\begin{aligned} S_\theta (X, Y) &= \nabla_\theta \ln(f_\theta(X, Y)) = \nabla_\theta \ln(p_\theta(X)q_\theta(Y)) \\ &= \nabla_\theta \ln(p_\theta(X)) + \nabla_\theta \ln(q_\theta(Y)) \\ &= S_\theta (X) + S_\theta (Y) \end{aligned}$$

$S_\theta (X)$ et $S_\theta (Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

\Rightarrow La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_θ et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = N I(\theta)$$

\Rightarrow N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_\alpha(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0;1[$

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_\theta (S_\theta(X))$$

Comme $\mathbb{E}_\theta(S_\theta(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_\theta[S_\theta(X)S_\theta(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_\theta[\nabla_\theta [S_\theta(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \ln p_\theta(X)}{\partial \theta_i \partial \theta_j} \right]$$

Proposition

Soient $X \sim p_\theta$, $Y \sim q_\theta$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_\theta : (x, y) \mapsto p_\theta(x)q_\theta(y)$.

En effet :

$$\begin{aligned} S_\theta(X, Y) &= \nabla_\theta \ln(f_\theta(X, Y)) = \nabla_\theta \ln(p_\theta(X)q_\theta(Y)) \\ &= \nabla_\theta \ln(p_\theta(X)) + \nabla_\theta \ln(q_\theta(Y)) \\ &= S_\theta(X) + S_\theta(Y) \end{aligned}$$

$S_\theta(X)$ et $S_\theta(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

\Rightarrow La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_θ et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = N I(\theta)$$

\Rightarrow N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_\alpha(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0;1[$

$$S_\alpha(X) = \frac{\partial \ln(p_\alpha(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

Définition

On appelle **Information de Fisher** au point θ la matrice de variance-covariance du vecteur score, notée $I(\theta)$:

$$I(\theta) = \mathbb{V}_{\theta}(S_{\theta}(X))$$

Comme $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$, nous avons directement :

$$I(\theta) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T]$$

Proposition

(Sous les conditions de régularité), nous avons

$$I(\theta) = -\mathbb{E}_{\theta}[\nabla_{\theta}[S_{\theta}(X)]^T]$$

c'est à dire pour tout $1 \leq i, j \leq p$:

$$I_{ij}(\theta) = -\mathbb{E}_{\theta}\left[\frac{\partial^2 \ln p_{\theta}(X)}{\partial \theta_i \partial \theta_j}\right]$$

Proposition

Soient $X \sim p_{\theta}$, $Y \sim q_{\theta}$ **deux v.a. indépendantes**, alors :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta), \quad \forall \theta \in \Theta,$$

où $I_{(X,Y)}(\theta)$ est l'information de Fisher en θ pour la variable aléatoire (X, Y) de densité $f_{\theta} : (x, y) \mapsto p_{\theta}(x)q_{\theta}(y)$.

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

\Rightarrow La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = N I(\theta)$$

\Rightarrow N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0;1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

$$R(T, \theta) = \mathbb{E}_{\theta} \left(\|T - \theta\|^2 \right)$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendants (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. \square

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendantes (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. □

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

$$R(T, \theta) = \mathbb{E}_{\theta} \left(\|T - \theta\|^2 \right)$$

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendantes (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. □

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_{\theta}(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_{\theta}(T) = \mathbb{E}_{\theta}(T) - \theta$.

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendantes (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. □

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_{\theta}(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_{\theta}(T) = \mathbb{E}_{\theta}(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_{\theta}(T) = 0, \text{ c'est à dire } \mathbb{E}_{\theta}(T) = \theta.$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendantes (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. □

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_{\theta}(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_{\theta}(T) = \mathbb{E}_{\theta}(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_{\theta}(T) = 0, \text{ c'est à dire } \mathbb{E}_{\theta}(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_{\theta}(\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_{\theta}(T)) + \|b_{\theta}(T)\|^2,$$

En effet :

$$\begin{aligned} S_{\theta}(X, Y) &= \nabla_{\theta} \ln(f_{\theta}(X, Y)) = \nabla_{\theta} \ln(p_{\theta}(X)q_{\theta}(Y)) \\ &= \nabla_{\theta} \ln(p_{\theta}(X)) + \nabla_{\theta} \ln(q_{\theta}(Y)) \\ &= S_{\theta}(X) + S_{\theta}(Y) \end{aligned}$$

$S_{\theta}(X)$ et $S_{\theta}(Y)$ sont indépendantes (images par des applications mesurables de v.a. indépendantes).

⇒ La variance de la somme de deux variables aléatoires indépendantes est la somme des variances de chacune, d'où le résultat. □

Proposition

Soit (X_1, \dots, X_N) un échantillon i.i.d. de loi P_{θ} et $I(\theta)$ l'information de Fisher associée à chaque variable individuellement.

Alors l'information de Fisher de l'échantillon (X_1, \dots, X_N) noté $I_N(\theta)$ vaut :

$$I_N(\theta) = NI(\theta)$$

⇒ N observations indépendantes apportent N fois plus d'information qu'une seule observation...

Exemple

$X \sim \text{Bernoulli}(\alpha) : p_{\alpha}(x) = \alpha^x(1-\alpha)^{1-x}$, pour $\alpha \in]0; 1[$

$$S_{\alpha}(X) = \frac{\partial \ln(p_{\alpha}(X))}{\partial \alpha} = \frac{X}{\alpha} - \frac{1-X}{1-\alpha} = \frac{X-\alpha}{\alpha(1-\alpha)}$$

$$I(\alpha) = \mathbb{V}(S_{\alpha}(X)) = \frac{1}{[\alpha(1-\alpha)]^2} \mathbb{V}(X - \alpha) = \frac{1}{\alpha(1-\alpha)} \quad \square$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_{\theta}(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_{\theta}(T) = \mathbb{E}_{\theta}(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta :$

$$b_{\theta}(T) = 0, \text{ c'est à dire } \mathbb{E}_{\theta}(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_{\theta}(\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_{\theta}(T)) + \|b_{\theta}(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_{\theta}((T - \theta)^2) = \mathbb{V}_{\theta}(T) + b_{\theta}(T)^2.$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_{\theta}(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_{\theta}(T) = \mathbb{E}_{\theta}(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_{\theta}(T) = 0, \text{ c'est à dire } \mathbb{E}_{\theta}(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_{\theta} (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_{\theta}(T)) + \|b_{\theta}(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_{\theta} ((T - \theta)^2) = \mathbb{V}_{\theta}(T) + b_{\theta}(T)^2.$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_\theta(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_\theta(T) = \mathbb{E}_\theta(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_\theta(T) = 0, \text{ c'est à dire } \mathbb{E}_\theta(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_\theta (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_\theta(T)) + \|b_\theta(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_\theta ((T - \theta)^2) = \mathbb{V}_\theta(T) + b_\theta(T)^2.$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned} \mathbb{E}_\theta [(T - \theta)^2] &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2] \\ &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T))^2] + \mathbb{E}_\theta [(\mathbb{E}_\theta(T) - \theta)^2] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0} \end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_\theta(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_\theta(T) = \mathbb{E}_\theta(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_\theta(T) = 0, \text{ c'est à dire } \mathbb{E}_\theta(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_\theta (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_\theta(T)) + \|b_\theta(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_\theta ((T - \theta)^2) = \mathbb{V}_\theta(T) + b_\theta(T)^2.$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned} \mathbb{E}_\theta [(T - \theta)^2] &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2] \\ &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T))^2] + \mathbb{E}_\theta [(\mathbb{E}_\theta(T) - \theta)^2] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0} \end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_\theta(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_\theta(T) = \mathbb{E}_\theta(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_\theta(T) = 0, \text{ c'est à dire } \mathbb{E}_\theta(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_\theta (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_\theta(T)) + \|b_\theta(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_\theta ((T - \theta)^2) = \mathbb{V}_\theta(T) + b_\theta(T)^2.$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned} \mathbb{E}_\theta [(T - \theta)^2] &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2] \\ &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T))^2] + \mathbb{E}_\theta [(\mathbb{E}_\theta(T) - \theta)^2] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0} \end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$.

L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_\theta(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_\theta(T) = \mathbb{E}_\theta(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_\theta(T) = 0, \text{ c'est à dire } \mathbb{E}_\theta(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_\theta (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_\theta(T)) + \|b_\theta(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_\theta ((T - \theta)^2) = \mathbb{V}_\theta(T) + b_\theta(T)^2.$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned} \mathbb{E}_\theta [(T - \theta)^2] &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2] \\ &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T))^2] + \mathbb{E}_\theta [(\mathbb{E}_\theta(T) - \theta)^2] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0} \end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$.

L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit \mathcal{M}_Θ , $\Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_\theta(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_\theta(T) = \mathbb{E}_\theta(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_\theta(T) = 0, \text{ c'est à dire } \mathbb{E}_\theta(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_\theta (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_\theta(T)) + \|b_\theta(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_\theta ((T - \theta)^2) = \mathbb{V}_\theta(T) + b_\theta(T)^2.$$

En effet (dans le cas $\Theta = \mathbb{R}$:

$$\begin{aligned} \mathbb{E}_\theta [(T - \theta)^2] &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0} \end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit \mathcal{M}_Θ , $\Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.

II.3 - Propriétés des estimateurs

II.3.a - Risque, biais, variance

Définition

Pour un estimateur T du paramètre θ , on appelle **risque quadratique** :

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

Définition

T est un **meilleur estimateur** que T' si, pour tout $\theta \in \Theta$ on a

$$R(T, \theta) \leq R(T', \theta),$$

et si il existe un $\theta' \in \Theta$ tel que

$$R(T, \theta') < R(T', \theta')$$

Définition

Si l'estimateur T de θ est intégrable, $\mathbb{E}_\theta(\|T\|) < +\infty$, et on appelle **biais** de T la quantité $b_\theta(T) = \mathbb{E}_\theta(T) - \theta$.

T est dit **estimateur sans biais** de θ si $\forall \theta \in \Theta$:

$$b_\theta(T) = 0, \text{ c'est à dire } \mathbb{E}_\theta(T) = \theta.$$

Proposition : Décomposition biais-variance

Lorsque T est de carré intégrable, on peut décomposer le risque quadratique :

$$\mathbb{E}_\theta (\|T - \theta\|^2) = \text{Tr}(\mathbb{V}_\theta(T)) + \|b_\theta(T)\|^2,$$

Si $\Theta \subset \mathbb{R}$, nous avons simplement :

$$\mathbb{E}_\theta ((T - \theta)^2) = \mathbb{V}_\theta(T) + b_\theta(T)^2.$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned} \mathbb{E}_\theta [(T - \theta)^2] &= \mathbb{E}_\theta [(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2 \underbrace{(\mathbb{E}_\theta(T) - \theta) \mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0} \end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit \mathcal{M}_Θ , $\Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2 \underbrace{(\mathbb{E}_\theta(T) - \theta) \mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$.
L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2 \underbrace{(\mathbb{E}_\theta(T) - \theta) \mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2 \underbrace{(\mathbb{E}_\theta(T) - \theta) \mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2 \underbrace{(\mathbb{E}_\theta(T) - \theta) \mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$.
L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \underbrace{\mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right]}_{=0} \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_{\theta} \left[(T - \theta)^2 \right] &= \mathbb{E}_{\theta} \left[(T - \mathbb{E}_{\theta}(T) + \mathbb{E}_{\theta}(T) - \theta)^2 \right] \\ &= \mathbb{E}_{\theta} \left[(T - \mathbb{E}_{\theta}(T))^2 \right] + \underbrace{\mathbb{E}_{\theta} \left[(\mathbb{E}_{\theta}(T) - \theta)^2 \right]}_{=0} \\ &\quad + 2(\mathbb{E}_{\theta}(T) - \theta) \underbrace{\mathbb{E}_{\theta} [T - \mathbb{E}_{\theta}(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_{\theta}(T) \leq \mathbb{V}_{\theta}(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_{\Theta}, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_{\theta}(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_{\theta}(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_{\alpha} \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_{\alpha} \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\alpha}(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_{\alpha} \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_{\alpha} \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_{\alpha}(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \underbrace{\mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right]}_{=0} \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T') .$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \underbrace{\mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right]}_{=0} \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \underbrace{\mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right]}_{=0} \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

• Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = I(\alpha; X_1, \dots, X_N)^{-1}$$

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

• Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = I(\alpha; X_1, \dots, X_N)^{-1}$$

\Rightarrow L'estimateur est **efficace**

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

• Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = I(\alpha; X_1, \dots, X_N)^{-1}$$

\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5 N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.

En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

• Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = I(\alpha; X_1, \dots, X_N)^{-1}$$

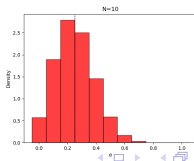
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



En effet (dans le cas $\Theta = \mathbb{R}$) :

$$\begin{aligned}\mathbb{E}_\theta \left[(T - \theta)^2 \right] &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(T - \mathbb{E}_\theta(T))^2 \right] + \mathbb{E}_\theta \left[(\mathbb{E}_\theta(T) - \theta)^2 \right] \\ &\quad + 2(\mathbb{E}_\theta(T) - \theta) \underbrace{\mathbb{E}_\theta [T - \mathbb{E}_\theta(T)]}_{=0}\end{aligned}$$

(Décomposition de Huygens), d'où le résultat. \square

II.3.b - Efficacité des estimateurs sans biais, Borne de C-R

La décomposition biais-variance permet de se concentrer sur une comparaison des variances pour les estimateurs sans biais.

Définition

Soient T, T' , deux **estimateurs sans biais** de θ pour $\theta \in \Theta$. L'estimateur T est dit **plus efficace** que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta$$

et

$$\exists \theta' \in \Theta, \mathbb{V}_{\theta'}(T) < \mathbb{V}_{\theta'}(T').$$

Théorème : Borne de Cramér-Rao

Soit $\mathcal{M}_\Theta, \Theta \subset \mathbb{R}^p$, un **modèle régulier**.

Soit T un estimateur **sans biais** de θ pour $\theta \in \Theta$, avec T de carré sommable et tel que l'on puisse différencier son espérance sous le signe intégrale.

Alors on a :

$$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$$

La quantité $I^{-1}(\theta)$ est appelée **borne de Cramér-Rao**.

[Pour A, B symétriques, on note $A < B$ si $B - A$ est définie positive, et $A \leq B$ si $B - A$ est semi-définie positive.]

Définition

Un estimateur sans biais T tel que $\mathbb{V}_\theta(T) = I^{-1}(\theta)$ est dit **efficace**, ou meilleur estimateur sans biais.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

• Calcul du biais :

$$\mathbb{E}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{E}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\alpha(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

• Calcul de la variance de l'estimateur :

$$\mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = \mathbb{V}_\alpha \left(\frac{1}{N} \sum_{i=1}^N X_i \right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_\alpha(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

• Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_\alpha \left(\hat{\alpha}^{MV} \right) = I(\alpha; X_1, \dots, X_N)^{-1}$$

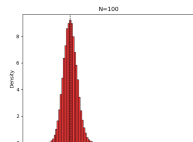
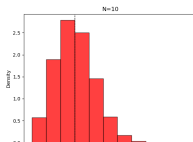
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{E}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\alpha}(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{V}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_{\alpha}(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = I(\alpha; X_1, \dots, X_N)^{-1}$$

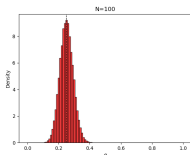
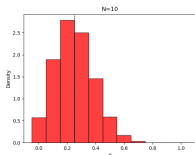
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{E}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\alpha}(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{V}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_{\alpha}(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = I(\alpha; X_1, \dots, X_N)^{-1}$$

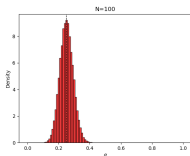
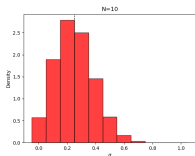
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N -échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{E}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\alpha}(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{V}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_{\alpha}(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = I(\alpha; X_1, \dots, X_N)^{-1}$$

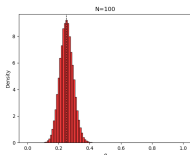
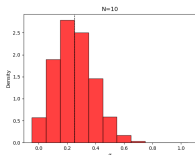
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.
 Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{E}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\alpha}(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{V}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_{\alpha}(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = I(\alpha; X_1, \dots, X_N)^{-1}$$

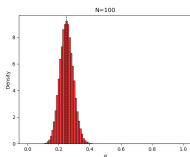
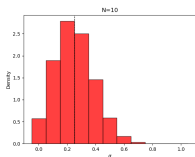
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, (X_1, \dots, X_N) N-échantillon i.i.d.

Nous avons l'estimateur du maximum de vraisemblance :

$$\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

- Calcul du biais :

$$\mathbb{E}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{E}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\alpha}(X_i) = \alpha$$

$\Rightarrow \hat{\alpha}^{MV}$ est un **estimateur non biaisé** de α .

- Calcul de la variance de l'estimateur :

$$\mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = \mathbb{V}_{\alpha}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}_{\alpha}(X_i) = \frac{\alpha(1-\alpha)}{N}$$

D'après la décomposition biais-variance, c'est aussi le risque quadratique.

- Information de Fisher : $I(\alpha; X_1) = \frac{1}{\alpha(1-\alpha)}$

donc Information de Fisher de l'échantillon :

$$I(\alpha; X_1, \dots, X_N) = \frac{N}{\alpha(1-\alpha)}$$

$$\Rightarrow \mathbb{V}_{\alpha}(\hat{\alpha}^{MV}) = I(\alpha; X_1, \dots, X_N)^{-1}$$

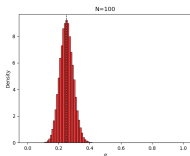
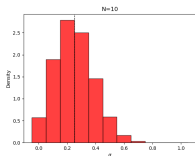
\Rightarrow L'estimateur est **efficace**

Pour une valeur théorique $\alpha = 0.25$: simulation de 10^5

N-échantillons :

$$z_1 = (x_1^1, \dots, x_N^1), \dots, z_{100000} = (x_1^{100000}, \dots, x_N^{100000}),$$

puis calcul de $\hat{\alpha}^{MV}(z_i)$ pour $i = 1, \dots, 10^5$ et tracé histogramme.



II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si
$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si
$$\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0$,
$$\mathbb{P}(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0$$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si
$$\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$$
 en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Théorème de Slutsky

Soient deux suites de variables aléatoires réelles Y_N et Z_N , Y une variable aléatoire et $c \in \mathbb{R}$, tels que

$$Y_N \xrightarrow{\mathcal{L}} Y \text{ et } Z_N \xrightarrow{\mathcal{L}} c.$$

Alors

$$Y_N + Z_N \xrightarrow{\mathcal{L}} Y + c \text{ et } Y_N Z_N \xrightarrow{\mathcal{L}} cY$$

Plus généralement, si f est continue,

$$f(X_N, Y_N) \xrightarrow{\mathcal{L}} f(X, c).$$

II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{L} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{L} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{L} \mathcal{N}(0, 1)$$

Théorème de Slutsky

Soient deux suites de variables aléatoires réelles Y_N et Z_N , Y une variable aléatoire et $c \in \mathbb{R}$, tels que

$$Y_N \xrightarrow{L} Y \text{ et } Z_N \xrightarrow{L} c.$$

Alors

$$Y_N + Z_N \xrightarrow{L} Y + c \text{ et } Y_N Z_N \xrightarrow{L} cY$$

Plus généralement, si f est continue,

$$f(X_N, Y_N) \xrightarrow{L} f(X, c).$$

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et $c \in \mathbb{R}^d$ tels que $Y_N \xrightarrow{L} c$, alors $Y_N \xrightarrow{\mathbb{P}} c$.

II.3.c - Rappels sur les lois limites

Définition

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles.

- Convergence presque sûre : $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$
- Convergence quadratique : $X_n \xrightarrow{L^2} X$ si $\mathbb{E}\left(|X_n - X|^2\right) \xrightarrow{n \rightarrow +\infty} 0$
- Convergence en probabilité : $X_n \xrightarrow{p.s.} X$ si $\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = 1$
- Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$ si $\mathbb{P}(X_n \leq x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \leq x)$ en tout point de continuité de la fonction de répartition $\mathbb{P}(X \leq x)$.

Proposition

- Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$
- Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$

Théorème : Loi forte des grands nombres

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $\mathbb{E}(|X_1|) < +\infty$. On a alors :

$$\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X_1),$$

où on a noté $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$.

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Théorème de Slutsky

Soient deux suites de variables aléatoires réelles Y_N et Z_N , Y une variable aléatoire et $c \in \mathbb{R}$, tels que

$$Y_N \xrightarrow{\mathcal{L}} Y \text{ et } Z_N \xrightarrow{\mathcal{L}} c.$$

Alors

$$Y_N + Z_N \xrightarrow{\mathcal{L}} Y + c \text{ et } Y_N Z_N \xrightarrow{\mathcal{L}} cY$$

Plus généralement, si f est continue,

$$f(X_N, Y_N) \xrightarrow{\mathcal{L}} f(X, c).$$

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et $c \in \mathbb{R}^d$ tels que $Y_N \xrightarrow{\mathcal{L}} c$, alors $Y_N \xrightarrow{\mathbb{P}} c$.

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d . S'il existe une variable aléatoire Z , une suite réelle $(a_N)_N$, $\lim_{N \rightarrow +\infty} a_N = +\infty$ et $c \in \mathbb{R}^d$ tels que :

$$a_N(Y_N - c) \xrightarrow{\mathcal{L}} Z$$

$$\text{alors : } Y_N \xrightarrow{\mathbb{P}} c.$$

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Théorème de Slutsky

Soient deux suites de variables aléatoires réelles Y_N et Z_N , Y une variable aléatoire et $c \in \mathbb{R}$, tels que

$$Y_N \xrightarrow{\mathcal{L}} Y \text{ et } Z_N \xrightarrow{\mathcal{L}} c.$$

Alors

$$Y_N + Z_N \xrightarrow{\mathcal{L}} Y + c \text{ et } Y_N Z_N \xrightarrow{\mathcal{L}} cY$$

Plus généralement, si f est continue,

$$f(X_N, Y_N) \xrightarrow{\mathcal{L}} f(X, c).$$

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et $c \in \mathbb{R}^d$ tels que $Y_N \xrightarrow{\mathcal{L}} c$, alors $Y_N \xrightarrow{\mathbb{P}} c$.

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d . S'il existe une variable aléatoire Z , une suite réelle $(a_N)_N$,

$\lim_{N \rightarrow +\infty} a_N = +\infty$ et $c \in \mathbb{R}^d$ tels que :

$$a_N(Y_N - c) \xrightarrow{\mathcal{L}} Z$$

$$\text{alors : } Y_N \xrightarrow{\mathbb{P}} c.$$

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Théorème de Slutsky

Soient deux suites de variables aléatoires réelles Y_N et Z_N , Y une variable aléatoire et $c \in \mathbb{R}$, tels que

$$Y_N \xrightarrow{\mathcal{L}} Y \text{ et } Z_N \xrightarrow{\mathcal{L}} c.$$

Alors

$$Y_N + Z_N \xrightarrow{\mathcal{L}} Y + c \text{ et } Y_N Z_N \xrightarrow{\mathcal{L}} cY$$

Plus généralement, si f est continue,

$$f(X_N, Y_N) \xrightarrow{\mathcal{L}} f(X, c).$$

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et $c \in \mathbb{R}^d$ tels que $Y_N \xrightarrow{\mathcal{L}} c$, alors $Y_N \xrightarrow{\mathbb{P}} c$.

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d . S'il existe une variable aléatoire Z , une suite réelle $(a_N)_N$,

$\lim_{N \rightarrow +\infty} a_N = +\infty$ et $c \in \mathbb{R}^d$ tels que :

$$a_N(Y_N - c) \xrightarrow{\mathcal{L}} Z$$

$$\text{alors : } Y_N \xrightarrow{\mathbb{P}} c.$$

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

$$(i) \quad Y_N \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_N) \xrightarrow{\mathcal{L}} h(Y)$$

$$(ii) \quad Y_N \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_N) \xrightarrow{\mathbb{P}} h(Y)$$

$$(iii) \quad Y_N \xrightarrow{p.s.} Y \Rightarrow h(Y_N) \xrightarrow{p.s.} h(Y)$$

Théorème Central Limite

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. telle que $0 < \mathbb{V}(X_1) < +\infty$. On a alors :

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X_1)}{\sqrt{\mathbb{V}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Théorème de Slutsky

Soient deux suites de variables aléatoires réelles Y_N et Z_N , Y une variable aléatoire et $c \in \mathbb{R}$, tels que

$$Y_N \xrightarrow{\mathcal{L}} Y \text{ et } Z_N \xrightarrow{\mathcal{L}} c.$$

Alors

$$Y_N + Z_N \xrightarrow{\mathcal{L}} Y + c \text{ et } Y_N Z_N \xrightarrow{\mathcal{L}} cY$$

Plus généralement, si f est continue,

$$f(X_N, Y_N) \xrightarrow{\mathcal{L}} f(X, c).$$

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et $c \in \mathbb{R}^d$ tels que $Y_N \xrightarrow{\mathcal{L}} c$, alors $Y_N \xrightarrow{\mathbb{P}} c$.

Proposition

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d . S'il existe une variable aléatoire Z , une suite réelle $(a_N)_N$,

$\lim_{N \rightarrow +\infty} a_N = +\infty$ et $c \in \mathbb{R}^d$ tels que :

$$a_N(Y_N - c) \xrightarrow{\mathcal{L}} Z$$

$$\text{alors : } Y_N \xrightarrow{\mathbb{P}} c.$$

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

$$(i) \quad Y_N \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_N) \xrightarrow{\mathcal{L}} h(Y)$$

$$(ii) \quad Y_N \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_N) \xrightarrow{\mathbb{P}} h(Y)$$

$$(iii) \quad Y_N \xrightarrow{p.s.} Y \Rightarrow h(Y_N) \xrightarrow{p.s.} h(Y)$$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe \mathcal{C}^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

$$(i) \quad Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$$

$$(ii) \quad Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$$

$$(iii) \quad Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y_n) \xrightarrow{p.s.} h(Y)$$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

$$(i) \quad Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$$

$$(ii) \quad Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$$

$$(iii) \quad Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y_n) \xrightarrow{p.s.} h(Y)$$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

$$(i) \quad Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$$

$$(ii) \quad Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$$

$$(iii) \quad Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y_n) \xrightarrow{p.s.} h(Y)$$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

- (i) $Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$
- (ii) $Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$
- (iii) $Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y) \xrightarrow{p.s.} h(Y)$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

- (i) $Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$
- (ii) $Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$
- (iii) $Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y_n) \xrightarrow{p.s.} h(Y)$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.
 T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

- (i) $Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$
- (ii) $Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$
- (iii) $Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y_n) \xrightarrow{p.s.} h(Y)$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.
 T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

- (i) $Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$
- (ii) $Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$
- (iii) $Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y_n) \xrightarrow{p.s.} h(Y)$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.
 T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.

Théorème de Continuité

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R}^d et h une application continue. Alors :

- (i) $Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$
- (ii) $Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow h(Y_n) \xrightarrow{\mathbb{P}} h(Y)$
- (iii) $Y_n \xrightarrow{p.s.} Y \Rightarrow h(Y) \xrightarrow{p.s.} h(Y)$

Finalement, la méthode delta permet en particulier d'extrapoler les résultats du théorème central limite à des transformées de variables aléatoires.

Méthode delta

Soit Y_N une suite de variables aléatoires à valeurs dans \mathbb{R} , $m \in \mathbb{R}$, et Z une variable aléatoire tels que :

$$\sqrt{N}(Y_N - m) \xrightarrow{\mathcal{L}} Z$$

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^1 et $h'(m) \neq 0$, alors :

$$\sqrt{N} \left(\frac{h(Y_N) - h(m)}{h'(m)} \right) \xrightarrow{\mathcal{L}} Z$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

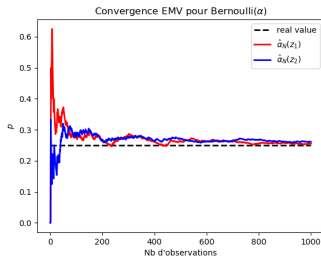
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

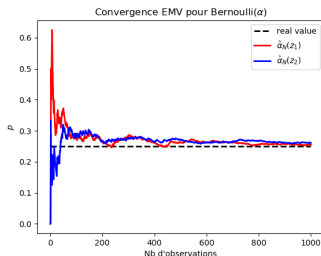
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

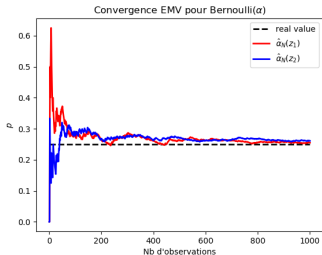
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



Proposition

Si T_N est asymptotiquement sans biais et si $\text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- T_N est consistant.

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

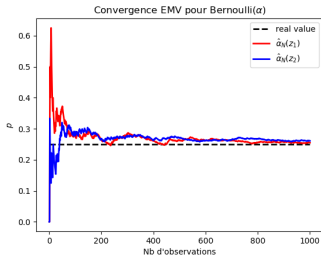
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta, T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit fortement consistant ou fortement convergent si
 $\forall \theta \in \Theta, T_N \xrightarrow{p.s.} \theta$.

Exemple

$$X \sim \text{Bernoulli}(\alpha), \hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N.$$

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

$$(i) \quad R(T_N, \theta) = \mathbb{E} \left(\|T_N - \theta\|^2 \right) \xrightarrow{N \rightarrow +\infty} 0. \quad T_N$$

converge en moyenne quadratique vers θ :

(ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + \text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow[N \rightarrow +\infty]{} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$.

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

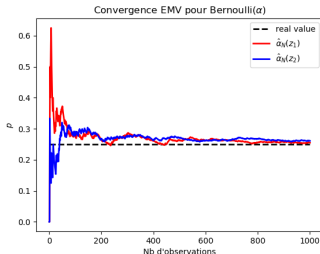
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit **fortement consistant ou fortement convergent** si $\forall \theta \in \Theta$, $T_N \xrightarrow{p.s.} \theta$.

Exemple

$X \sim \text{Bernoulli}(\alpha)$, $\hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N$.

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



Proposition

Si T_N est asymptotiquement sans biais et si $\text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- (ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + \text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$. □

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow +\infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

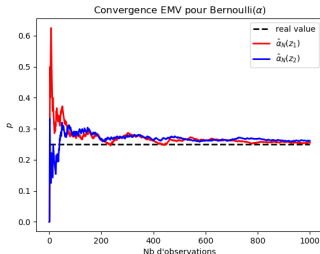
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta, T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit fortement consistant ou fortement convergent si
 $\forall \theta \in \Theta, T_N \xrightarrow{p.s.} \theta$.

Exemple

$$X \sim \text{Bernoulli}(\alpha), \hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N.$$

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E} \left(\|T_N - \theta\|^2 \right) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
(ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + \text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow[N \rightarrow +\infty]{} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$.

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow +\infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

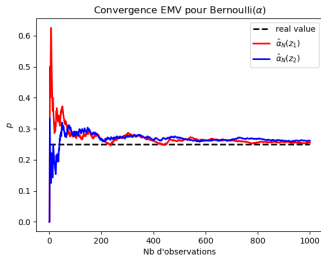
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta, T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit fortement consistant ou fortement convergent si
 $\forall \theta \in \Theta, T_N \xrightarrow{p.s.} \theta$.

Exemple

$$X \sim \text{Bernoulli}(\alpha), \hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N.$$

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E} \left(\|T_N - \theta\|^2 \right) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
(ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + \text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow[N \rightarrow +\infty]{} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$.

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow +\infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur** normal T_N .

II.3.d - Propriétés asymptotiques des estimateurs

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ pour X_1, \dots, X_N i.i.d. de loi inconnue dans \mathcal{M}_Θ .

Définition

Si pour tout $\theta \in \Theta$, $\lim_{N \rightarrow +\infty} b_\theta(T_N) = 0$, l'estimateur est dit **asymptotiquement sans biais**.

Définition

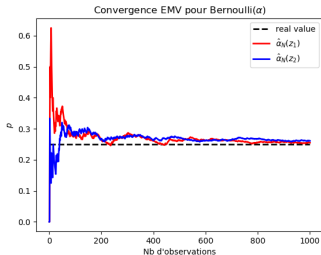
T_N est dit **consistant ou convergent** si $\forall \theta \in \Theta, T_N \xrightarrow{\mathbb{P}} \theta$.

T_N est dit fortement consistant ou fortement convergent si
 $\forall \theta \in \Theta, T_N \xrightarrow{p.s.} \theta$.

Exemple

$$X \sim \text{Bernoulli}(\alpha), \hat{\alpha}^{MV}(X_1, \dots, X_N) = \bar{X}_N.$$

D'après la loi forte des grands nombres, $\bar{X}_N \xrightarrow{p.s.} \mathbb{E}(X) = \alpha$.
 \Rightarrow convergence forte de l'estimateur.



Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

$$(i) \quad R(T_N, \theta) = \mathbb{E} \left(\|T_N - \theta\|^2 \right) \xrightarrow[N \rightarrow +\infty]{} 0. \quad T_N$$

converge en moyenne quadratique vers θ :

(ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + \text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow[N \rightarrow +\infty]{} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$.

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow +\infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur** normal T_N .

Proposition

Si un estimateur est asymptotiquement normal, alors il est convergent.

Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- (ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$. □

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow \infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur** normal T_N .

Proposition

Si un estimateur est asymptotiquement normal, alors il est convergent.

Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- (ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$. □

Définition

Soit T_N un estimateur de θ consistant. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow \infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur** normal T_N .

Proposition

Si un estimateur est asymptotiquement normal, alors il est convergent.

Définition

Soient T_N et T'_N deux estimateurs de θ asymptotiquement normaux, de variances asymptotiques associées Σ et Σ' . Alors T_N est dit **asymptotiquement plus efficace** que T'_N si on a :

$$\Sigma(\theta) \leq \Sigma'(\theta), \quad \forall \theta \in \Theta \text{ et } \exists \theta' \in \Theta, \Sigma(\theta') < \Sigma'(\theta').$$

Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- (ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$. □

Définition

Soit T_N un estimateur de θ consistant. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow \infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur** normal T_N .

Proposition

Si un estimateur est asymptotiquement normal, alors il est convergent.

Définition

Soient T_N et T'_N deux estimateurs de θ asymptotiquement normaux, de variances asymptotiques associées Σ et Σ' . Alors T_N est dit **asymptotiquement plus efficace** que T'_N si on a :

$$\Sigma(\theta) \leq \Sigma'(\theta), \quad \forall \theta \in \Theta \text{ et } \exists \theta' \in \Theta, \Sigma(\theta') < \Sigma'(\theta').$$

Définition

Un estimateur est **asymptotiquement efficace** lorsqu'il est asymptotiquement normal et que sa matrice de variance asymptotique associé $\Sigma(\theta) = I^{-1}(\theta)$.

Proposition

Si T_N est asymptotiquement sans biais et si $Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- (ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + Tr(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$. □

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$\lim_{N \rightarrow \infty} a_N = +\infty$ et $a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z$,
(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur normal** T_N .

Proposition

Si un estimateur est asymptotiquement normal, alors il est convergent.

Définition

Soient T_N et T'_N deux estimateurs de θ asymptotiquement normaux, de variances asymptotiques associées Σ et Σ' . Alors T_N est dit **asymptotiquement plus efficace** que T'_N si on a :

$$\Sigma(\theta) \leq \Sigma'(\theta), \quad \forall \theta \in \Theta \text{ et } \exists \theta' \in \Theta, \Sigma(\theta') < \Sigma'(\theta').$$

Définition

Un estimateur est **asymptotiquement efficace** lorsqu'il est asymptotiquement normal et que sa matrice de variance asymptotique associé $\Sigma(\theta) = I^{-1}(\theta)$.

Rappel : si l'estimateur était efficace, nous aurions :

$$\forall N, \mathbb{V}_\theta(T_N) = \frac{1}{N} I(\theta)^{-1}, \text{ c'est à dire :}$$

$$\mathbb{V}_\theta(\sqrt{N}(T_N - \theta)) = I(\theta)^{-1}.$$

Proposition

Si T_N est asymptotiquement sans biais et si $\text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$, alors :

- (i) $R(T_N, \theta) = \mathbb{E}(\|T_N - \theta\|^2) \xrightarrow{N \rightarrow +\infty} 0$. T_N converge en moyenne quadratique vers θ ;
- (ii) T_N est consistant.

En effet : d'après la décomposition de la variance

$$\mathbb{E}(\|T_N - \theta\|^2) = b_\theta^2(T_N) + \text{Tr}(\mathbb{V}_\theta(T_N)) \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc $T_N \xrightarrow{L^2} \theta$, et donc $T_N \xrightarrow{\mathbb{P}} \theta$. □

Définition

Soit T_N un estimateur de θ **consistant**. S'il existe une suite réelle (a_N) strictement positive et Z v.a. non dégénérée telles que :

$$\lim_{N \rightarrow \infty} a_N = +\infty \text{ et } a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z,$$

(a_N) s'appelle **vitesse de convergence** de l'estimateur T_N .

Définition

Un estimateur T_N de θ est **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de variance Σ ne dépend pas de N . On l'appellera **variance asymptotique associée à l'estimateur** normal T_N .

Proposition

Si un estimateur est asymptotiquement normal, alors il est convergent.

Définition

Soient T_N et T'_N deux estimateurs de θ asymptotiquement normaux, de variances asymptotiques associées Σ et Σ' . Alors T_N est dit **asymptotiquement plus efficace** que T'_N si on a :

$$\Sigma(\theta) \leq \Sigma'(\theta), \quad \forall \theta \in \Theta \text{ et } \exists \theta' \in \Theta, \Sigma(\theta') < \Sigma'(\theta').$$

Définition

Un estimateur est **asymptotiquement efficace** lorsqu'il est asymptotiquement normal et que sa matrice de variance asymptotique associé $\Sigma(\theta) = I^{-1}(\theta)$.

Rappel : si l'estimateur était efficace, nous aurions :

$$\forall N, \mathbb{V}_\theta(T_N) = \frac{1}{N} I(\theta)^{-1}, \text{ c'est à dire :}$$

$$\mathbb{V}_\theta(\sqrt{N}(T_N - \theta)) = I(\theta)^{-1}.$$

Théorème

Soit X_1, \dots, X_N i.i.d. pour p_θ , $p_\theta \in \mathcal{M}_\Theta$. Sous certaines hypothèses techniques de régularité (généralement vérifiées...), l'estimateur du maximum de vraisemblance $\hat{\theta}^{MV}$ est :

- convergent,
- asymptotiquement normal,
- asymptotiquement efficace.

$$\sqrt{N}(\hat{\theta}^{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\theta)^{-1})$$