Statistique inférentielle

Estimation

A. Godichon-Baggioni

Objectifs

On dispose de n données $x_1, \ldots, x_n$ qui sont des mesures d'une variable quantitative, et plus précisément des réalisations de variables aléatoires indépendantes et identiquement distribuées $X_1, \ldots, X_n$.

Objectifs :

- Estimer une caractéristique θ de X_1 (moyenne, variance,...)
- Estimer un paramètre θ de X_1 si cette variable est paramétrée (loi de Bernoulli, normale, exponentielle,...)

Exemples :

- Lancer de pièce : On s'intéresse à l'estimation de $\theta = \mathbb{E}[X_1]$.
- Loi exponentielle : On s'intéresse à l'estimation du paramètre $\theta = (\mathbb{E}[X_1])^{-1}$.

I. Définitions

Modèle statistique

Dans ce qui suit, on considère un objet aléatoire $\mathbf{X}$.

Définition (Modèle statistique)

Une expérience statistique est la donnée d'un objet aléatoire $\mathbf{X}$ *à valeurs dans un espace mesurable* $(E, \mathcal{E})$*, et d'une famille de loi* $(P_\theta)_{\theta \in \Theta}$ *supposée contenir la vraie loi de* $\mathbf{X}$*, et appelée modèle statistique pour la loi de* $\mathbf{X}$*.*

Exemples

Exemple 1 : lancer de pièce. Soit $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d, et $X_1 \sim \mathcal{B}(\theta)$ avec $\theta \in (0,1)$. On a $E = \{0,1\}^n$ et

$$(P_\theta)_{\theta \in \Theta} = \left(\mathcal{B}(\theta)^{\otimes n}\right)_{\theta \in (0,1)}.$$

Exemple 2 : taille des hommes adultes. La taille d'un homme adulte est modélisée par une loi normale de paramètres μ, σ^2 inconnus. Ici, $E = \mathbb{R}^n$, $\mathcal{E} = \mathcal{B}(\mathbb{R}^n)$, $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d avec $X_1 \sim \mathcal{N}\left(\mu, \sigma^2\right)$. On a donc

$$(P_\theta)_{\theta \in \Theta} = \left(\mathcal{N}\left(\mu, \sigma^2\right)^{\otimes n}\right)_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*}.$$

Statistique et estimateur

Définition

Une statistique $T(\mathbf{X})$ est une fonction mesurable de l'objet aléatoire $\mathbf{X}$ ne dépendant pas de θ (mais dépendant éventuellement de paramètres connus). Un estimateur de θ est une statistique $\hat{\theta} = \theta(\mathbf{X})$ destinée à approcher θ.

Exemple : Lancer de pièce

Attention ! Ne pas confondre estimateur et estimation !

Notation : Dans ce qui suit, on considère $\hat{\theta}_n$ un estimateur de θ.

Erreur quadratique moyenne

Définition

On suppose que θ est à valeurs dans $\Theta \subset \mathbb{R}$. L'erreur quadratique moyenne (ou risque quadratique) de l'estimateur $\hat{\theta}_n$ est définie pour tout $\theta \in \Theta$ par

$$EQM\left(\hat{\theta}_n, \theta\right) = \mathbb{E}\left[\left(\hat{\theta}_n - \theta\right)^2\right].$$

Remarquons que grâce à l'inégalité de Markov, pour tout $c > 0$,

$$\mathbb{P}\left[\left|\hat{\theta}_n - \theta\right| \geq c\right] \leq \frac{\mathrm{EQM}\left(\hat{\theta}_n, \theta\right)}{c^2}.$$

Biais d'un estimateur

Définition

On appelle biais d'un estimateur $\hat{\theta}_n$ de θ la quantité

$$B\left(\hat{\theta}_n, \theta\right) = \mathbb{E}\left[\hat{\theta}_n\right] - \theta.$$

1. *S'il est nul, on dit que l'estimateur est sans biais ou non biaisé*
2. *Si*

$$B\left(\hat{\theta}_n, \theta\right) \xrightarrow[n \to +\infty]{} 0,$$

on dit que l'estimateur est asymptotiquement sans biais.

Exemple : lancer de pièce. *L'estimateur $\hat{\theta}_n$ est sans biais.*

Décomposition Bais-Variance

Proposition

Soit $\hat{\theta}_n$ un estimateur de θ, on a

$$EQM\left(\hat{\theta}_n, \theta\right) = B\left(\hat{\theta}_n, \theta\right)^2 + \mathbb{V}\left[\hat{\theta}_n\right].$$

Exemple : lancer de pièce Comme $\hat{\theta}_n$ est un estimateur sans biais de θ, on a

$$\mathrm{EQM}\left(\hat{\theta}_n, \theta\right) = \mathbb{V}\left[\hat{\theta}_n\right] = \frac{\theta(1-\theta)}{n}.$$

Convergence, consistance et normalité asymptotique

Définition

On dit que l'estimateur $\hat{\theta}_n$ est

1. *convergent ou consistant si*

$$\hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta,$$

2. *fortement consistant si*

$$\hat{\theta}_n \xrightarrow[n\to+\infty]{p.s} \theta,$$

3. *asymptotiquement normal si il existe $\sigma^2 > 0$ tel que*

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right).$$

II. Estimation de la moyenne et de la variance

Estimation de la moyenne

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées. Un estimateur naturel de la moyenne $\mu = \mathbb{E}[X_1]$ est donc

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Proposition

1. *$\overline{X}_n$ est un estimateur sans biais et (fortement) consistant de μ.*
2. *Si $\sigma^2 = \mathbb{V}[X_1] < +\infty$, alors $\overline{X}_n$ est asymptotiquement normal et*

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right)$$

$$EQM\left(\overline{X}_n, \mu\right) = \frac{\sigma^2}{n}.$$

Estimation de la moyenne

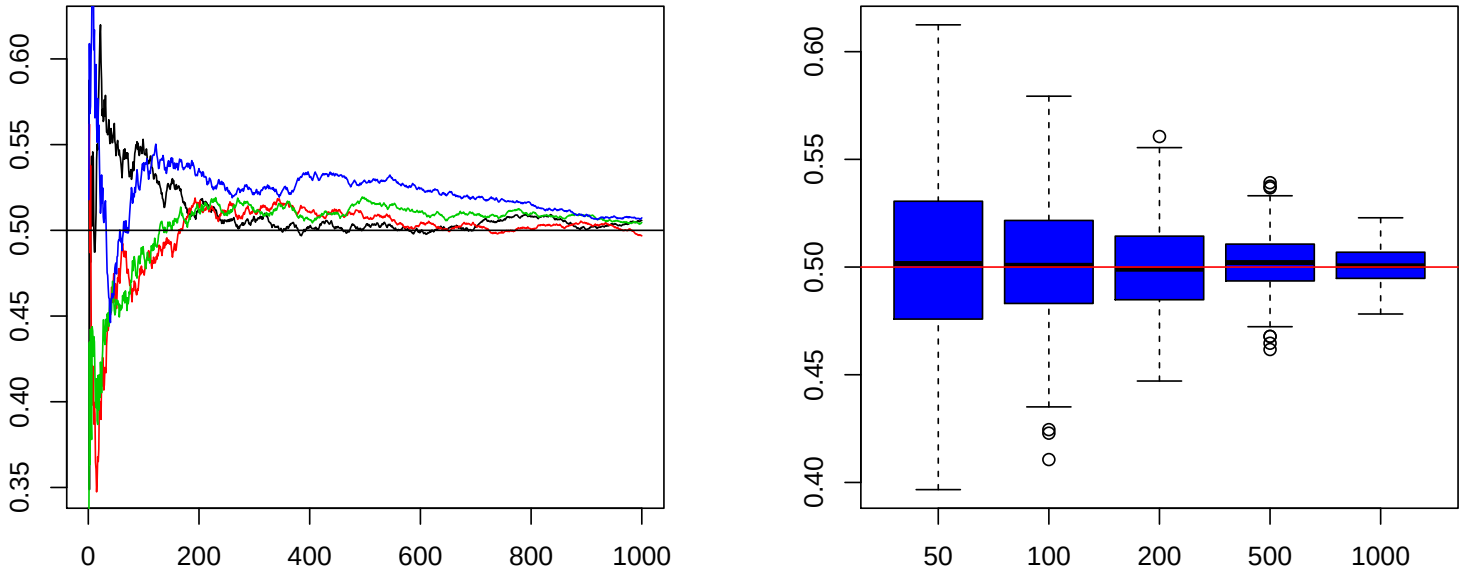

FIGURE – Evolution de $\overline{x}_n$ par rapport à n pour 4 échantillons (à gauche) et boxplots pour les $\overline{x}_n$ obtenus pour $n = 50, 100, 200, 500, 1000$ à l'aide de 4000 échantillons (à droite).

Estimation de la variance

Lorsque μ est connu, un estimateur naturel de la variance est

$$\hat{V}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

$\hat{V}_n$ est sans biais et fortement consistant.
Si μ est inconnue, on a la variance empirique

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}_n^2.$$

Estimateur sans biais de la variance

Proposition

$\hat{\sigma}_n^2$ est un estimateur biaisié de σ^2 mais asymptotiquement sans biais. Plus précisément, on a

$$\mathbb{E}\left[\hat{\sigma}_n^2\right] = \frac{n-1}{n}\sigma^2.$$

Définition

L'estimateur sans biais de la variance S_n^2 est défini par

$$S_n^2 = \frac{n}{n-1}\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2\right).$$

Estimateur sans biais de la variance

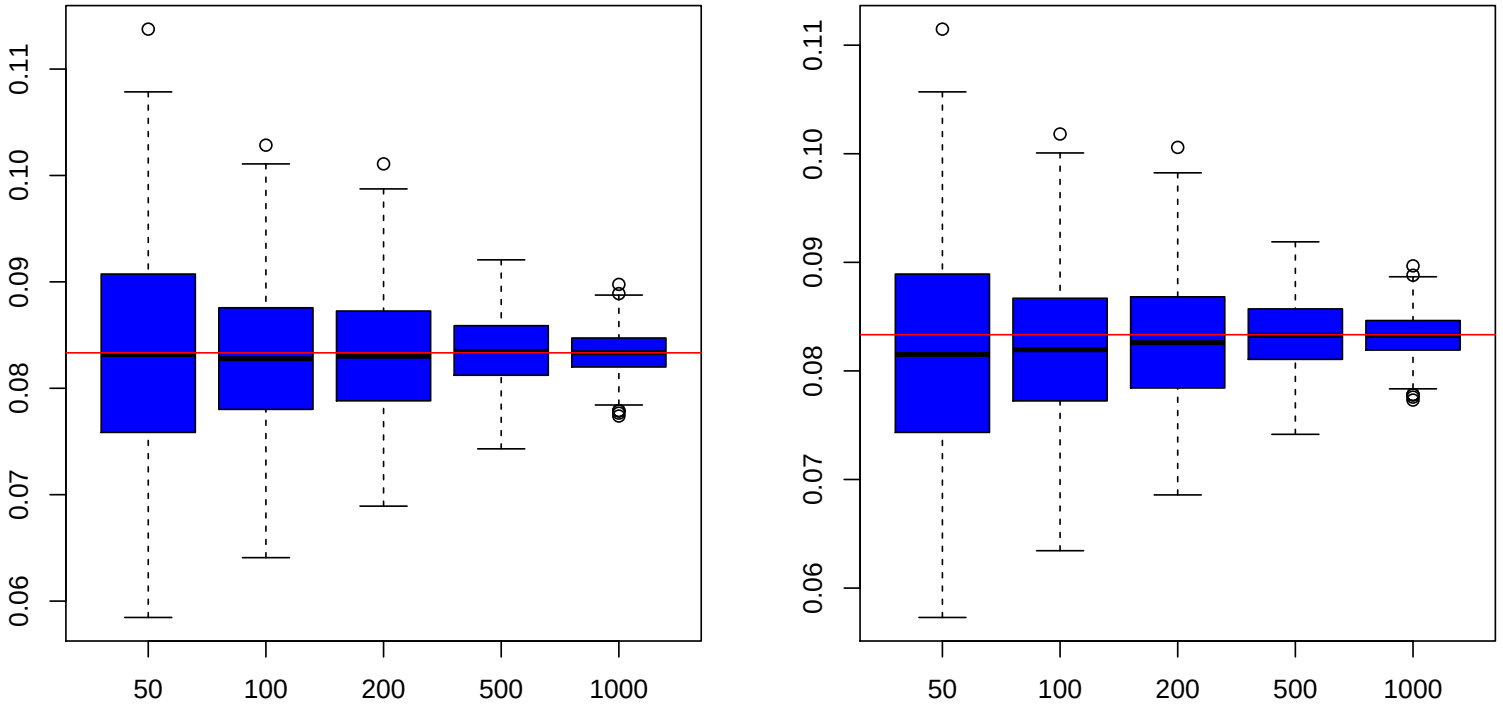

FIGURE – Boxplots pour l'estimateur non biaisé (à gauche) et biaisé (à droite)

Convergence

Proposition

Les estimateurs $\hat{\sigma}_n^2$ et S_n^2 sont consistants.

Proposition

Si X_1 admet un moment d'ordre 4, alors $\hat{\sigma}_n^2$ et S_n^2 sont asymptotiquement normaux, et on a

$$\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \tau^4 - \sigma^4\right)$$

$$\sqrt{n}\left(S_n^2 - \sigma^2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \tau^4 - \sigma^4\right)$$

où

$$\tau^4 = \mathbb{E}\left[(X_1 - \mu)^4\right].$$

III. Méthode des moments

Méthode des moments

Proposition

Soit Θ un intervalle ouvert de $\mathbb{R}$ et $\theta \in \Theta$. Soit φ un C^1-difféomorphisme de Θ dans $\varphi(\Theta)$. Soit $\hat{\varphi}_n$ un estimateur consistant de $\varphi(\theta)$, alors $\hat{\theta}_n = \varphi^{-1}(\hat{\varphi}_n)$ est un estimateur consistant de θ, i.e

$$\hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta.$$

De plus, si $\hat{\varphi}_n$ est un estimateur asymptotiquement normal de $\varphi(\theta)$, i.e si il existe $\sigma^2 > 0$ tel que

$$\sqrt{n}\left(\hat{\varphi}_n - \varphi(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right),$$

et si $\varphi'(\theta) \neq 0$, alors $\hat{\theta}_n$ est un estimateur asymptotiquement normal de θ et

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{\left(\varphi'(\theta)\right)^2}\right).$$

Méthode des moments

La méthode des moments consiste à trouver un C^1-difféomorphisme φ et un moment k tel que $\mathbb{E}\left[X_1^k\right] = \varphi(\theta)$. Comme un estimateur de m_k est donné par

$$\hat{m}_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

on obtient l'estimateur

$$\hat{\theta}_n = \varphi^{-1}\left(\hat{m}_{n,k}\right)$$

Exemples

Exemple : la loi uniforme. On considère des variables aléatoires i.i.d $X_1, \ldots, X_n$ suivant une loi uniforme sur $\left[0, \theta^2\right]$, avec $\theta > 0$, i.e de densité f_θ définie pour tout $x \in \mathbb{R}$ par

$$f_\theta(x) = \frac{1}{\theta^2} \mathbf{1}_{[0,\theta^2]}(x).$$

Exemple : la loi exponentielle. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d suivant une loi exponentielle de paramètre $\theta > 0$, i.e de densité f_θ définie pour tout $x \in \mathbb{R}$,

$$f_\theta(x) = \theta \exp\left(-\theta x\right) \mathbf{1}_{\mathbb{R}_+}(x).$$

Remarques

Remarque : Pour éviter les erreurs, il est plus judicieux (et cela revient à peu près au même) d'écrire θ comme une fonction de $\mathbb{E}\left[X^k\right]$.

Remarque : Attention ! Il peut arriver qu'une variable aléatoire n'admette pas de moment d'ordre 1. Il faut alors essayer d'être malin !

Exemple : Soit $\theta > 0$, on considère une variable aléatoire X de densité définie pour tout $x \in \mathbb{R}$ par

$$f_\theta(x) = \frac{\theta}{x^2}\mathbf{1}_{x \geq \theta}.$$

IV. Méthode du Maximum de Vraisemblance

(LOG)-VRAISEMBLANCE

On considère un objet aléatoire $\mathbf{X}$ et un modèle statistique $(P_\theta)_{\theta\in\Theta}$ où Θ est un ouvert de $\mathbb{R}$. On suppose que pour tout $\theta \in \Theta$, P_θ admet une densité g_θ.

Définition (Vraisemblance et log-vraisemblance)

La vraisemblance de $\mathbf{X}$ *est définie pour tout* $\theta \in \Theta$ *par*

$$L_{\mathbf{X}}(\theta) = g_\theta(\mathbf{X}).$$

La log-vraisemblance de $\mathbf{X}$ *est définie pour tout* $\theta \in \Theta$ *par*

$$l_{\mathbf{X}}(\theta) = \log\left(g_\theta(\mathbf{X})\right).$$

Attention ! La (log-)vraisemblance est aléatoire.

Estimateur du maximum de vraisemblance

Définition (Estimateur du maximum de vraisemblance (EMV))

Le maximum de vraisemblance, si il existe, est un élément $\hat{\theta}_n$ de Θ qui vérifie

$$L_{\mathbf{X}}\left(\hat{\theta}_n\right) = \sup_{\theta \in \Theta} L_{\mathbf{X}}(\theta).$$

De manière équivalente, l'estimateur du maximum de vraisemblance, si il existe, vérifie

$$l_{\mathbf{X}}\left(\hat{\theta}_n\right) = \sup_{\theta \in \Theta} l_{\mathbf{X}}(\theta).$$

Remarques

Attention ! L'EMV est généralement aléatoire.

Attention ! Ni l'existence ni l'unicité de l'EMV ne sont assurées.

Remarque : A noter que si $\hat{\theta}_n$ est un EMV de θ, alors $\varphi\left(\hat{\theta}_n\right)$ est un EMV de $\varphi(\theta)$.

Remarque : On notera également

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_{\mathbf{X}}(\theta) \qquad \text{ou} \qquad \hat{\theta}_n = \arg\max_{\theta\in\Theta} l_{\mathbf{X}}(\theta)$$

(LOG-)VRAISEMBLANCE D'UN ÉCHANTILLON

Dans ce qui suit, on considère $\mathbf{X} = (X_1, \ldots, X_n)$ ou les X_i sont des variables aléatoires i.i.d et X_1 admet une certaine densité f_θ.

Proposition (Vraisemblance et log-vraisemblance d'un échantillon)

La vraisemblance de $\mathbf{X} = (X_1, ..., X_n)$ *est définie pour tout* $\theta \in \Theta$ *par*

$$L_{\mathbf{X}}(\theta) = \prod_{j=1}^{n} f_\theta\left(X_j\right).$$

La log vraisemblance de $\mathbf{X}$ *est définie pour tout* $\theta \in \Theta$ *par*

$$l_{\mathbf{X}}(\theta) = \log\left(L_{\mathbf{X}}(\theta)\right) = \log\left(\prod_{i=1}^{n} f_\theta\left(X_i\right)\right).$$

Remarques

Remarque : Dans ce qui suit, on notera également

$$L_{\mathbf{X}}(\theta) = L_n(\theta) \qquad \text{et} \qquad l_{\mathbf{X}}(\theta) = l_n(\theta)$$

Remarque : Dans le cas discret, on peut voir la réalisation de la vraisemblance comme étant la probabilité d'observer cet échantillon pour un θ fixé.

Exemples

Exemple 1 : cas discret. On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et suivent une loi de Poisson de paramètre $\theta \in \mathbb{R}_+^*$, i.e la densité de X_1 est définie pour tout $x \in \mathbb{N}$ par

$$f_\theta(x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

Exemple 2 : cas continu. On considère $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et $X_1 \sim \mathcal{N}(\theta, 1)$.

Comment trouver l'EMV

Soit l'EMV peut être donné "explicitement" par la vraisemblance, soit il est souvent plus facile de maximiser la log-vraisemblance. Pour cela, voici quelques options possibles :

- Dresser le tableau de variations de la log vraisemblance
- Chercher les zéros de la dérivée, i.e résoudre

$$l'_{\mathbf{X}}(\theta) = 0.$$

- Vérifier qu'il(s) maximise(nt) la log-vraisemblance (étude de la concavité, ...)

Exemples

Cas discret : Soit $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d avec $X_1 \sim \mathcal{B}(\theta)$ avec $\theta \in (0, 1)$.

Cas continu : Soit $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d avec $X_1 \sim \mathcal{N}\left(\mu, \sigma^2\right)$ avec $\mu \in \mathbb{R}$ inconnu et $\sigma^2 > 0$ connu.

Cas intéressant : Soit $\mathbf{X} = (X_1, \ldots, X_n)$ où les X_i sont i.i.d et admettent une densité f_θ définie pour tout $\theta > 0$ et pour tout $x \in \mathbb{R}$ par

$$f_\theta(x) = \frac{\theta}{x^2} \mathbf{1}_{x \geq \theta}.$$

Un autre exemple intéressant

Exemple : Soit $X_1, \ldots, X_n$ des variables aléatoires i.i.d suivant une loi uniforme sur $[0, \theta]$, avec $\theta > 0$.

- Méthode des moments ?
- Méthode du maximum de vraisemblance ?