

Chapitre 2. Estimation ponctuelle

Claude Petit, université de Rennes - claud.petit@univ-rennes.fr

Sept. 2025

1. Estimateurs: définitions et propriétés

1,5. Interlude

2. Estimateurs plug-in

3. Estimateurs par substitution

4. Estimateurs des moments

5. Estimateurs du maximum de vraisemblance

6. Estimateurs bayésiens

1. Estimateurs: définitions et propriétés

Estimateur, biais, erreur quadratique

- $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ avec $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ modèle paramétrique et $X \sim \mathbb{P}$ v.a.

Definition

Un estimateur S (ou S_n) de θ est une statistique à valeurs dans Θ .

Doit être fonction des **seules** données et **ne pas dépendre du paramètre** θ .
Doit donner une valeur proche de la vraie valeur de θ !

Definition

Le **biais** d'un estimateur S est la quantité

$$b(\theta) = \mathbb{E}_\theta[S] - \theta \quad (1)$$

L'**erreur quadratique moyenne** de l'estimateur est

$$\mathbb{E}_\theta[(S - \theta)^2] = \|S - \theta\|_2^2 = \mathbb{V}_\theta(S) + b(\theta)^2 \quad (2)$$

Si $b(\theta) \neq 0$, on dit que l'estimateur est biaisé.

Deux exemples canoniques

- $X_1, \dots, X_n$ n -échantillon d'une va X d'espérance m et variance σ^2 .

$\bar{X} = (\sum_{i=1}^n X_i)/n$ **moyenne empirique** et **variance empirique**:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3)$$

- $\bar{X}$ et S^2 sont des **estimateurs** respectifs de la moyenne m et de la variance σ^2 de X .
- $\mathbb{E}[\bar{X}] = m$ et par calcul rapide (à faire) $\mathbb{E}[S^2] = \frac{n-1}{n} \sigma^2$.
- $\bar{X}$ estimateur sans biais de m , S^2 est un estimateur biaisé de la variance.

Propriétés asymptotiques

Soit S_n estimateur de θ (modèle d'échantillonnage). $X_1, \dots, X_n$ v.a.i.i.d. de loi $\mathbb{P}_\theta$. S_n fonction de $(X_1, \dots, X_n)$ et n .

S_n est **asymptotiquement sans biais** si

$$\lim_{n \rightarrow +\infty} \mathbb{E}_\theta[S_n] = \theta \quad (4)$$

S_n **converge en moyenne quadratique** si

$$S_n \xrightarrow{L^2} \theta \quad (5)$$

S_n est **fortement consistant** pour θ si

$$S_n \xrightarrow{\mathbb{P}_\theta - \text{p.s.}} \theta \quad (6)$$

S_n est **consistant** pour θ si

$$T_n \xrightarrow{\mathbb{P}_\theta} \theta \quad (7)$$

Rappels sur les convergences probabilistes

- S_n converge en **moyenne quadratique** vers θ si, et seulement si,

$$\lim_{n \rightarrow +\infty} \|S_n - \theta\|_2 = \lim_{n \rightarrow +\infty} \mathbb{E} [|S_n - \theta|^2] = 0 \quad (8)$$

- S_n converge **presque sûrement** vers θ si

$$\forall \theta, \mathbb{P}_\theta \left[\lim_{n \rightarrow +\infty} S_n = \theta \right] = 1 \quad (9)$$

- S_n converge **en probabilité** θ si, et seulement si,

$$\forall \epsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}_\theta [|S_n - \theta| > \epsilon] = 0 \quad (10)$$

Consistance forte $\Rightarrow$ consistance et convergence $L^2 \Rightarrow$ consistance.

- $(S_n)_n$ converge **en loi** vers une v.a. S (notation $\rightsquigarrow$) ssi, en notant F_n et F respectivement la fonction de répartition de S_n et celle de S , on a

$$S_n \rightsquigarrow S \iff \lim_{n \rightarrow +\infty} F_n(x) = F(x), \forall x \text{ où } F \text{ est continue.} \quad (11)$$

Les trois théorèmes limites -1-

Theorem (loi forte des grands nombres (LFGN))

Soient $X_1, \dots, X_n$ un n -échantillon d'une v.a. X de loi $\mathbb{P}$ où $\mathbb{P}$ admet un moment d'ordre 1 noté $\mathbb{E}[X] = m$. Alors :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}\text{-p.s.}} m \quad (12)$$

La vitesse de convergence est précisée par le théorème suivant.

Theorem (de la limite centrale (TLC))

Soient $X_1, \dots, X_n$ un n -échantillon d'une v.a. $X \sim \mathbb{P}$ où $m = \mathbb{E}[X]$ et $\sigma^2 = \mathbb{V}(X)$ existent. Alors

$$\sqrt{n}(\overline{X} - m) \rightsquigarrow \mathcal{N}(0, \sigma^2) \quad (13)$$

Les trois théorèmes limites -2-

- De façon équivalente:

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1) \quad (14)$$

- S_n est v_n -consistant et asymptotiquement normal (A.N.) si

$$v_n(S_n - \theta) \rightsquigarrow \mathcal{N}(0, \Sigma_\theta) \quad (15)$$

- La formule est donnée ici en dimension p ($\theta \in \mathbb{R}^p$). Σ_θ matrice $p \times p$ symétrique positive = matrice de covariance de la loi limite.
- Le troisième théorème limite est la loi faible des grands nombres. C'est un cas particulier de la loi forte des grands nombres qui assure que la convergence a lieu en probabilité.

Les trois théorèmes limites -3-

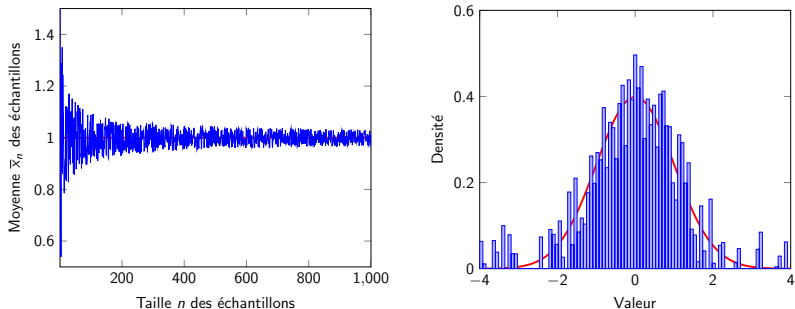


Figure 1: Gauche: illustration de la LFGN: un échantillon i.i.d. de taille n est généré et sa moyenne empirique $\bar{x}_n$ est calculée pour des valeurs de n de plus en plus grande. Droite: illustration du TLC: pour $n = 100$, comparaison de la densité théorique d'un échantillon avec la densité gaussienne standard (en rouge).

1,5. Interlude

Les différents estimateurs en une diapositive

Soit $S = S_n$ un estimateur de θ . Nous allons étudier les estimateurs...

- **Plug-in**: on remplace $\mathbb{P}$ par $\mathbb{P}_n$, loi empirique.
- Par **substitution**: $\psi(S)$ est un estimateur de $\psi(\theta)$.
- Des **moments**: on identifie moments empirique $\hat{m}_k = \frac{1}{n} \sum_i X_i^k$ et théorique $m_k = \mathbb{E}[X^k]$.
- **Maximum de vraisemblance**: $\hat{\theta}_n = \arg \max_{\theta} L(X, \theta)$.
- **Bayésien**: statistiques *a posteriori* sachant les observations:
 $\theta_n^* = \mathbb{E}[\theta | X = x]$ avec $X = (X_1, \dots, X_n)$.

On peut maintenant attaquer les exercices sans faire le reste du cours...

2. Estimateurs plug-in

Fonction de répartition empirique -1-

$\mathbb{X} = \mathbb{R}$. $X \sim \mathbb{P} \in \mathcal{P}$ n -échantillon de loi inconnue.

Trouver $\mathbb{P} \iff$ trouver F , sa fonction de répartition.

Definition (Mesure et fonction de répartition empirique)

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad (16)$$

$$F_n(x) = \mathbb{P}_n([-\infty, x]) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[X_i \leq x]} \quad (17)$$

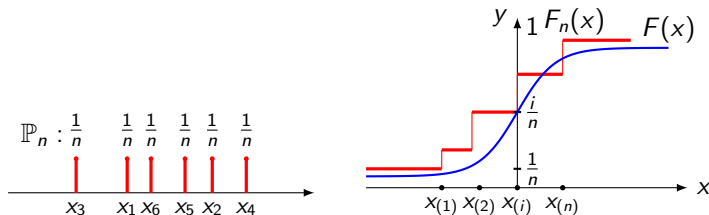
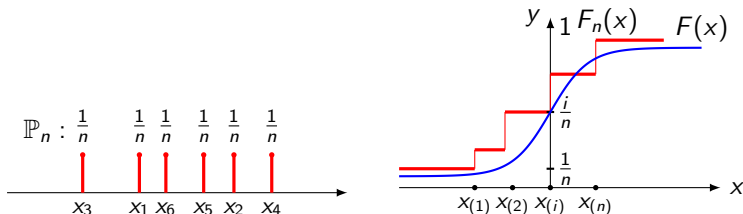


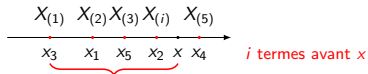
Figure 2: Mesure et fonction de répartition empirique d'un échantillon.

Fonction de répartition empirique -2-



- $\mathbb{P}_n$ **mesure aléatoire**: $\forall B \in \mathcal{B}_{\mathbb{R}}, \mathbb{P}_n(B)$ fonction des X_i , donc v.a.
- Mesure discrète: poids $1/n$ à l'observation X_i .
- F_n fonction $\ll$ càdlàg $\gg$.
- Si $X_{(i)}$ i ème stat. d'ordre de l'échantillon ($X_{(1)}$ min, $X_{(n)}$ max) :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[X_i \leq x]} = \begin{cases} 0 & \text{si } x \leq X_{(1)} \\ i/n & \text{si } X_{(i)} \leq x < X_{(i+1)} \\ 1 & \text{si } x \geq X_{(n)} \end{cases} \quad (18)$$

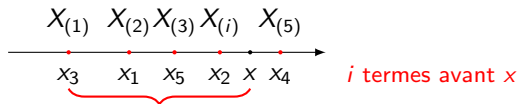


Fonction de répartition empirique -3-

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[X_i \leq x]} = \begin{cases} 0 & \text{si } x \leq X_{(1)} \\ i/n & \text{si } X_{(i)} \leq x < X_{(i+1)} \\ 1 & \text{si } x \geq X_{(n)} \end{cases} \quad (19)$$

Pour tout x, n, i , on a $[X_{(i)} \leq x] = [nF_n(x) \geq i]$.

-  Formule importante: $\mathbb{E}[\mathbb{1}_B] = \mathbb{P}(B) \Rightarrow$
- $F_n(x)$ estimateur de $F(x) = \mathbb{P}[X \leq x]$: $\mathbb{E}[F_n(x)] = F(x)$ et $\mathbb{V}(F_n(x)) = F(x)(1 - F(x))/n$.
- $nF_n(x)$ somme de n v.a.i.i.d. de Bernoulli de paramètre $F(x)$.
- $\Rightarrow$ Leur somme suit une loi binomiale de paramètres n et $F(x)$.
- **LFGN** $\Rightarrow \forall x$ fixé, $F_n(x) \rightarrow F(x)$ presque sûrement.



Théorème fondamental de la statistique -1-

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[X_i \leq x]} \quad (20)$$

$$\mathbb{E}[F_n(x)] = F(x) \text{ et } \mathbb{V}(F_n(x)) = F(x)(1 - F(x))/n \quad (21)$$

Theorem (de la limite centrale pour la loi empirique)

$$\forall x \in \mathbb{R}, \sqrt{n}(F_n(x) - F(x)) \rightsquigarrow \mathcal{N}(0; F(x)[1 - F(x)]) \quad (22)$$

La convergence a lieu pour tout x réel *fixé*.

Le théorème suivant (théorème fondamental de la statistique) assure que la CV preque sûre de F_n vers F est *uniforme en x* .

Theorem (de Glivenko-Cantelli)

$$\|F_n - F\|_{\infty} = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\mathbb{P}_{\theta} - p.s.} 0 \quad (23)$$

Théorème fondamental de la statistique -2-

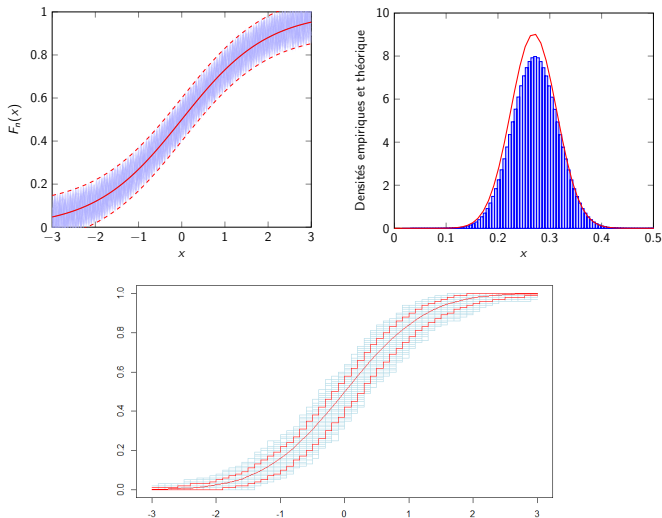


Figure 3: haut, gauche: fonctions F_n (bleu) et F (rouge) avec bande de confiance uniforme. En haut à droite: illustration de la normalité asymptotique. Bas: détail de la convergence de quelques F_n vers F (crédit Arthur Charpentier).

Definition

Une **fonctionnelle statistique** ϕ est une fonction de $\mathbb{P}$. Elle est linéaire si elle est de la forme

$$\phi(\mathbb{P}) = \int f(x) d\mathbb{P}(x) = \mathbb{E}_{\mathbb{P}}[f(X)] \quad (24)$$

$$\phi : \mathcal{P} \longrightarrow \mathbb{R}$$

$$\mathbb{P} \mapsto \phi(\mathbb{P})$$

ϕ est une **forme « linéaire » sur un ensemble de probabilités** (attention, $\mathcal{P}$ n'est pas un espace vectoriel).

Exemples: moyenne $m(\mathbb{P}) = \int x\mathbb{P}(dx)$, variance $\sigma^2(\mathbb{P}) = \int (x - m(\mathbb{P}))^2 \mathbb{P}(dx)$, médiane $m_e(\mathbb{P}) = F^{-1}(1/2)$ avec $F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) > p\}$ inverse généralisée de F .

Definition

Un estimateur « plug-in » est obtenu par substituant $\mathbb{P}_n$ à $\mathbb{P}$ dans une fonctionnelle statistique: $S_n = \phi(\mathbb{P}_n)$.

Si $\phi(\mathbb{P}) = \int f(x)\mathbb{P}(dx)$, l'estimateur plug-in associé est

$$\phi(\mathbb{P}_n) = \frac{1}{n} \sum_{i=1}^n f(X_i) \quad (25)$$

Exemples:

- Moyenne: $m(\mathbb{P}) = \int x\mathbb{P}(dx) \Rightarrow m(\mathbb{P}_n) = \bar{X}$.
- Variance: $\sigma^2(\mathbb{P}) = \int (x - m(\mathbb{P}))^2\mathbb{P}(dx) \Rightarrow \sigma^2(\mathbb{P}_n) = S_n^2$.
- Médiane: $m_e(\mathbb{P}) = F^{-1}(1/2) \Rightarrow m_e(\mathbb{P}_n) = X_{(n/2)}$ (si n pair).

- Si $f \in L^1(\mathbb{P})$, c.-à-d. si $\int |f(x)| d\mathbb{P} = \mathbb{E}[|f(X)|] < \infty$, la **LFGN** $\Rightarrow$

$$\phi(\mathbb{P}_n) \xrightarrow{\mathbb{P}\text{-p.s.}} \phi(\mathbb{P}) \quad (26)$$

- Si $f \in L^2(\mathbb{P})$, c.-à-d. si $\int |f(x)|^2 d\mathbb{P} = \mathbb{E}[f(X)^2] < \infty$, alors le **TLC** $\Rightarrow$

$$\sqrt{n}(\phi(\mathbb{P}_n) - \phi(\mathbb{P})) \rightsquigarrow \mathcal{N}(0, \tau(\mathbb{P})^2) \quad (27)$$

avec $\tau(\mathbb{P})^2 = \int (f(x) - \phi(\mathbb{P}))^2 d\mathbb{P}$.

3. Estimateurs par substitution

Definition

Soit S un estimateur de θ et $\psi : \Theta \rightarrow \mathbb{R}^d$ une fonction mesurable.
Estimateur par substitution de $\psi(\theta)$: estimateur de la forme $\psi(S)$.

Theorem (de l'application continue)

*Soit S un estimateur (fortement) consistant de θ .
Soit $\psi : \Theta \subset \mathbb{R}^p \rightarrow \mathbb{R}^d$ une fonction continue.
Alors $\psi(S)$ est (fortement) consistant pour $\psi(\theta)$.*



Estimateur par substitution $\iff$ obtenu par changement de variables.

Méthode Delta -1-

Theorem (**Méthode delta** en dimension 1)

Soit S_n un estimateur de θ tel que $v_n(S_n - \theta) \rightsquigarrow S$

Soit $\psi : \Theta \subset \mathbb{R} \longrightarrow \mathbb{R}$ dérivable. Alors

$$v_n(\psi(S_n) - \psi(\theta)) \rightsquigarrow \psi'(S) \quad (28)$$

• Si S_n A.N. avec $v_n(S_n - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2)$, alors $\psi(S_n)$ A.N. et

$$v_n(\psi(S_n) - \psi(\theta)) \rightsquigarrow \mathcal{N}(0, \Psi'(\theta)^2 \sigma^2) \quad (29)$$

• Exemple 1: $(X_1, \dots, X_n)$ n -échantillon de loi $b(\theta)$, $\theta \in]0, 1[$.

$$\sqrt{n}(\bar{X}_n - \theta) \rightsquigarrow \mathcal{N}(0, \theta(1 - \theta)) \quad (30)$$

Posons $\psi(\theta) = 2 \arcsin \sqrt{\theta} \Rightarrow \psi(\bar{X}_n)$ estimateur de $\psi(\theta)$.

ψ dérivable $\Rightarrow \psi(\bar{X}_n)$ $\sqrt{n}$ -consistant et A.N.

$$\psi'(\theta) = \frac{1}{\sqrt{\theta(1 - \theta)}} \Rightarrow \sqrt{n} \left(2 \arcsin \sqrt{\bar{X}_n} - 2 \arcsin \sqrt{\theta} \right) \rightsquigarrow \mathcal{N}(0, 1)$$

Méthode Delta: exemple -2-

- $(X_1, \dots, X_n)$ n -échantillon d'une v.a. X de loi exponentielle de paramètre $\theta > 0$ inconnu, que l'on souhaite estimer.

$$f(x) = \theta e^{-\theta x} \mathbb{1}_{[0, +\infty[}(x) \quad (31)$$

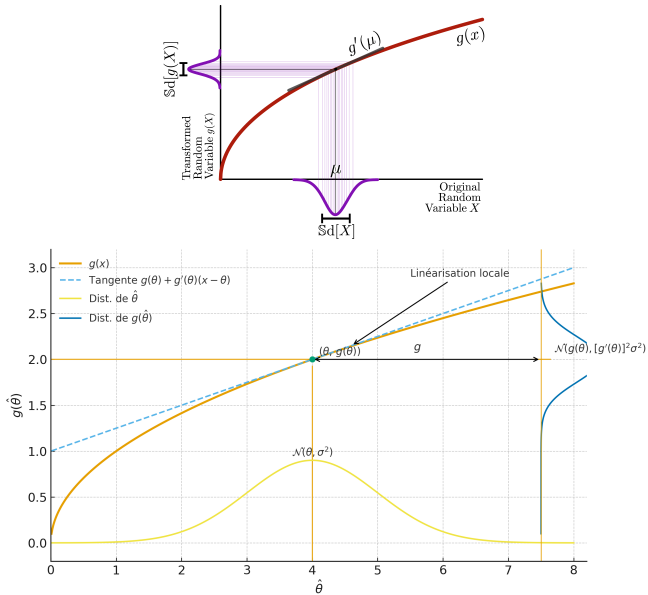
- $\bar{X}_n$ moyenne empirique de l'échantillon.
- Soit $\lambda = 1/\theta$. On sait que

$$\mathbb{E}[X] = \lambda \text{ et } \mathbb{V}(X) = \lambda^2 \quad (32)$$

- D'après le théorème de la limite centrale, $\sqrt{n}(\bar{X}_n - \lambda) \rightsquigarrow \mathcal{N}(0, \lambda^2)$
- Posons $\psi(x) = 1/x$ qui est différentiable sur $]0, +\infty[$ avec $\psi'(x)^2 = 1/x^4$. En appliquant la **méthode delta**, on en déduit que

$$\frac{\sqrt{n}}{\theta} \left(\frac{1}{\bar{X}_n} - \theta \right) \rightsquigarrow \mathcal{N}(0, 1) \quad (33)$$

Méthode Delta: illustration -3-



Méthode Delta en dimension p -3-

Theorem (**Méthode delta multidimensionnelle**)

S_n estimateur de θ / $\nu_n(S_n - \theta) \rightsquigarrow S$

$\psi : \Theta \subset \mathbb{R}^p \longrightarrow \mathbb{R}^d$ différentiable de matrice jacobienne J . Alors

$$\nu_n(\psi(S_n) - \psi(\theta)) \rightsquigarrow d\psi_\theta(S) \quad (34)$$

Si S_n est A.N. avec $\nu_n(S_n - \theta) \rightsquigarrow \mathcal{N}(0, \Sigma_\theta)$ alors $\psi(S_n)$ est A.N. et

$$\nu_n(\psi(S_n) - \psi(\theta)) \rightsquigarrow \mathcal{N}(0, J\Sigma_\theta J^T) \quad (35)$$

- $d\psi_\theta(h)$ est la différentielle de ψ au point θ , calculée en h . Lorsque $d = 1$, la différentielle est égale à $\psi'(\theta) \times h$. Si nous notons $J_\psi(\theta)$ la matrice jacobienne de ψ en θ , alors $d\psi_\theta(h) = J_\psi(\theta) \times h$.

- $J = J_{\psi(\theta)} \in \mathbb{R}^{d \times p}$ est la matrice jacobienne de ψ et $J^T = J_{\psi(\theta)}^T$ sa transposée. Dans le cas de la dimension 1, on a simplement $J\Sigma J^T = \psi'(\theta)^2 \sigma_\theta^2$, où σ_θ^2 variance de la loi normale asymptotique.

Méthode Delta: exemple en dimension $p = 4$

- Soit $S_n = (X_n, Y_n)^T$ vérifiant $\sqrt{n}(S_n - m) \rightsquigarrow \mathcal{N}(0, \Sigma)$ où $m = (0, 1)^T$ et $\Sigma = I_2$ matrice identité 2×2 .

Quelle est la loi asymptotique de $X_n Y_n$?

- On pose $\psi(x, y) = xy$.

ψ est C^1 sur $\mathbb{R}^2$, de jacobienne $J = \left(\frac{\partial}{\partial x}(xy), \frac{\partial}{\partial y}(xy) \right) = (y, x)$.

- Appliquée en m , on a $J_\psi(m) = J(m) = (1, 0)$ et par suite

$$J(m) \times \Sigma \times J(m)^T = 1 \quad (36)$$

$$\Rightarrow \sqrt{n}X_n Y_n \rightsquigarrow \mathcal{N}(0, 1).$$

Méthode delta: exemple multidimensionnel -5-

- $(X_1, \dots, X_n)$ n -échantillon de X / $\mathbb{E}[X^4] < \infty$. $\bar{X}$, $\overline{X^2}$, S_n^2 as usual...

- $\alpha_i = \mathbb{E}[|X|^i]$, $\forall i = 1, \dots, 4$ et $\alpha = (\alpha_1, \alpha_2)^T$.

- $T_n = (\bar{X}, \overline{X^2})^T$ et $\psi(x, y) = y - x^2$. $S_n^2 = \psi(\bar{X}, \overline{X^2})$. TLC $\Rightarrow$

$$\sqrt{n} \left(\begin{pmatrix} \bar{X} \\ \overline{X^2} \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right) \rightsquigarrow \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha_2 - \alpha_1^2 & \alpha_3 - \alpha_1 \alpha_2 \\ \alpha_3 - \alpha_1 \alpha_2 & \alpha_4 - \alpha_2^2 \end{pmatrix} \right)$$

ψ polynomiale donc différentiable et $d\psi_{(x,y)}(h, k) = -2xh + k$.

- **Méthode delta** appliquée à ψ donne

$$\sqrt{n} (S_n^2 - (\alpha_2 - \alpha_1^2)) \rightsquigarrow \mathcal{N} (0, -4\alpha_1^4 - \alpha_2^2 + 8\alpha_1^2 \alpha_2 - 4\alpha_1 \alpha_3 + \alpha_4)$$

loi (normale) de $-2\alpha_1 T_1 + T_2$ où $(T_1, T_2)^T = \lim \sqrt{n}(T_n - \alpha)$.

Si v.a. centrées ($\alpha_1 = 0$), le résultat devient

$$\sqrt{n} (S_n^2 - \alpha_2) \rightsquigarrow \mathcal{N} (0, \alpha_4 - \alpha_2^2)$$

4. Estimateurs des moments

Estimateur des moments: définition

X n -échantillon. $m_k(\theta) = \mathbb{E}_\theta [X^k]$ son moment théorique d'ordre k .

Moment empirique d'ordre k :

$$\hat{m}_k(\theta) = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (37)$$

Definition

L'**estimateur des moments** de $\theta \in \Theta \subset \mathbb{R}^p$ est la solution $\hat{\theta}$ du système à p équations, d'inconnue θ :

$$\begin{cases} m_1(\theta) = \hat{m}_1(\theta) \\ \vdots \\ m_p(\theta) = \hat{m}_p(\theta) \end{cases} \iff M(\theta) = \hat{M}(\theta) \quad (38)$$

$\theta \in \mathbb{R}^p$. Si $p = 1$, l'estimateur des moments est obtenu en **identifiant moyenne théorique et empirique**.

Estimateur des moments: exemples

Pour chaque exemple, $(X_1, \dots, X_n)$ n -échantillon de même loi que X et θ inconnu à estimer.

- X de loi de **Poisson** de paramètre θ . On sait que $\mathbb{E}[X] = \theta$.

L'estimateur des moments est $\bar{X} = \hat{m}_1(\theta)$.

- X de loi **exponentielle (anglo-saxonne)** de paramètre θ . On sait que $\mathbb{E}[X] = 1/\theta$. L'estimateur des moments de θ est T_n tel que

$$\bar{X} = \hat{m}_1(\theta) = \frac{1}{T_n} \Rightarrow T_n = \frac{1}{\bar{X}} \quad (39)$$

- X de loi **uniforme** sur $[0, \theta]$. $T_n = 2\bar{X}$ (à faire en exercice).

- X de loi **normale** $\mathcal{N}(m, \sigma^2)$ pour le paramètre $\theta = (m, \sigma^2)$.

L'estimateur des moments est donné par le couple $(\bar{X}, S^2)$ (à faire en exercice).

Estimateur des moments: propriétés

$M(\theta) = (m_1(\theta), \dots, m_p(\theta))^T$ vecteur des moments d'ordre 1 à p de X et $\hat{M}(\theta)$ celui des moments empiriques (dépendent de θ via les v.a. X_i).

Theorem

- Si $M : \Theta \rightarrow \mathbb{R}^p$ bijective, alors

$$\hat{\theta} = M^{-1} \circ \hat{M}(\theta) \text{ et } \hat{M}(\theta) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_{\theta-p.s.}} M(\theta) \quad (40)$$

- Si M est bijective et M^{-1} continue, alors

$$\hat{\theta} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_{\theta-p.s.}} \theta \quad (41)$$

- Si de plus M^{-1} est différentiable et $\mathbb{P}_{\theta}$ a un moment d'ordre 2, alors

$$\sqrt{n} \left(\hat{M}(\theta) - M(\theta) \right) \rightsquigarrow \mathcal{N}(0, \Sigma_{\theta}) \text{ et } \sqrt{n} \left(\hat{\theta} - \theta \right) \rightsquigarrow \mathcal{N}(0, \tilde{\Sigma}_{\theta}) \quad (42)$$

avec $\Sigma_{\theta} = (\text{cov}(X_i, X_j))_{i,j=1..p}$ et $\tilde{\Sigma}_{\theta} = J_M^{-1} \Sigma_{\theta} (J_M^{-1})^T$

Estimateur des moments: exemple

$(X_1, \dots, X_n)$ n -échantillon de $X \sim B(\alpha, \beta)$ ($\alpha, \beta > 0$), densité

$$g(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{1}_{[0,1]}(x) \quad (43)$$

L'estimateur des moments de (α, β) est solution de

$$\begin{cases} \bar{X} = \mathbb{E}[X] = \frac{\alpha}{\alpha+\beta} \\ \bar{X^2} = \mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \end{cases} \quad (44)$$

$$\psi(x, y) = \left(\frac{x}{x+y}, \frac{x(x+1)}{(x+y)(x+y+1)} \right) \quad (45)$$

est de classe C^1 et inversible sur $]0, 1]^2$. En l'inversant, on trouve l'estimateur en fonction de $\bar{X}$ et $\bar{X^2}$:

$$\hat{\alpha} = \frac{\bar{X}(\bar{X^2} - \bar{X})}{\bar{X^2} - \bar{X}^2} \text{ et } \hat{\beta} = \frac{(\bar{X^2} - \bar{X})(1 - \bar{X})}{\bar{X^2} - \bar{X}^2}, \quad (46)$$

estimateurs consistants et A.N. (à faire en exercice).

5. Estimateurs du maximum de vraisemblance

Maximum de vraisemblance : définition de l'e.m.v.

Definition

Un **estimateur du maximum de vraisemblance** de θ (on notera **e.m.v.**) est, sous réserve d'existence, une solution de l'équation

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} L(X, \theta) \quad (47)$$

- Peut ne pas exister, n'être pas unique, ou difficile à calculer.
- Déterminer l'e.m.v. $\iff$ déterminer maximum de la log-vraisemblance $l(X, \theta)$. Si L différentiable, l'e.m.v. est solution de

$$\frac{\partial}{\partial \theta} \ln L(X, \theta) = \frac{\partial}{\partial \theta} l(X, \theta) = \nabla l(X, \theta) = S(X, \theta) = 0 \quad (48)$$

- La dérivée (ou le gradient) de la log-vraisemblance est le score S .
- Condition nécessaire, non suffisante, pour être un max local. Max local: $\exists V \subset \Theta \ni \hat{\theta} / L(x, \theta) < L(x, \hat{\theta}) \forall \theta \in V$.

- Si Θ intervalle de $\mathbb{R}$ (paramètre à estimer scalaire), on a une **condition suffisante d'existence et d'unicité**:

$L(x, \theta)$ fonction de θ de classe C^2 sur Θ (x étant fixé) et

$$\left\{ \begin{array}{l} \frac{\partial l}{\partial \theta}(x, \theta) = 0 \\ \frac{\partial^2 l}{\partial \theta^2}(x, \theta) < 0 \end{array} \right. \quad (49)$$

- La solution doit par ailleurs appartenir à l'intérieur $\overset{\circ}{\Theta}$ de Θ .
- On a existence et unicité d'un maximum global si la fonction l est **strictement concave** sur Θ (en tant que fonction de Θ).

E.m.v. exemple d'un échantillon gaussien 1

- $X = (X_1, \dots, X_n)$ **échantillon gaussien** $\mathcal{N}(m, \sigma^2)$ et $\theta = (m, \sigma^2)$.
- Modèle paramétrique, dominé par mesure de Lebesgue. Vraisemblance:

$$L(x, \theta) = \prod_{i=1}^n f(x_i, \theta) = (2\pi)^{-n/2} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \ln \sigma^2 \right)$$

- Support du modèle: $\mathbb{R}^n \Rightarrow$ homogène. Vraisemblance de classe C^2 en θ (si $\sigma > 0$) $\Rightarrow$ e.m.v. solution de l'équation de vraisemblance

$$\frac{d}{d\theta} l(X, \theta) = \nabla_l(\theta) = 0 \quad (50)$$

où $l(x, \theta) = \ln L(x, \theta)$. Solution unique si matrice hessienne $H_l(\hat{\theta}) < 0$.

$$l(x, \theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) \quad (51)$$

E.m.v. exemple d'un échantillon gaussien 2

$$l(x, \theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) \quad (52)$$

$l(x, \theta)$ fonction vectorielle car $\theta = (m, \sigma^2) \in \mathbb{R}^2$.

Dérivée de l = gradient (coordonnées = dérivées partielles en m et σ^2).

$$\begin{aligned} \frac{d}{d\theta} l(x, \theta) = \nabla_l(\theta) &= \left(\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m); \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2\sigma^2} \right)^T = 0 \\ \iff \begin{cases} \sum_{i=1}^n (x_i - m) = 0 \\ \sum_{i=1}^n (x_i - m)^2 = n\sigma^2 \end{cases} &\iff \begin{cases} m = \bar{x} \\ \sigma^2 = S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases} \end{aligned}$$

$\hat{\theta} = (\bar{X}, S^2)$ point critique de l . Matrice hessienne de l en ce point:

$$H_l(\hat{\theta}) = \begin{pmatrix} -n/S^2 & 0 \\ 0 & -n/(2S^4) \end{pmatrix} < 0 \quad (53)$$

$\hat{\theta}$ e.m.v. unique (poser $S^2 = V$ pour éviter les erreurs dans les calculs).

E.m.v. exemple d'un échantillon uniforme 1

- $X = (X_1, \dots, X_n)$ n -échantillon de loi uniforme sur $[0, \theta]$ où $\theta > 0$.
- Modèle paramétrique, dominé par mesure de Lebesgue sur $\mathbb{R}^n$.

$$L(x, \theta) = \prod_{i=1}^n \left(\frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x_i) \right) = \frac{1}{\theta^n} \mathbb{1}_{[0, \theta]^n}(x) = \frac{1}{\theta^n} \mathbb{1}_{[0 \leq x_{(1)} \leq x_{(n)} \leq \theta]}(x)$$

$x_{(1)}$ min des x_i et $x_{(n)}$ max.

- Le support dépend de $\theta \Rightarrow$ modèle non homogène. Nous ne pouvons pas déterminer le maximum de vraisemblance par la méthode précédente.

On remarque par contre que, à x fixé, la fonction de vraisemblance est décroissante en θ à partir de $x_{(n)}$ (avant, elle est nulle). Le maximum est donc atteint en $\hat{\theta} = x_{(n)}$. L'e.m.v. existe et est unique:

$$\hat{\theta}_{MV} = X_{(n)} = \max_{i=1}^n X_i \quad (54)$$

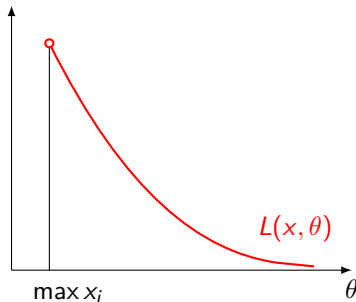


Figure 5: Vraisemblance d'un échantillon uniforme

- $\theta X_{(n)} \sim$ loi béta de paramètres $(n, 1)$. La densité du maximum est donc

$$\frac{ny^{n-1}}{\theta} \mathbb{1}_{[0, \theta]}(y) \quad (55)$$

$\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$ et l'e.m.v. a donc un biais égal à $b(\theta) = -\frac{\theta}{n+1}$.

E.m.v.: encore d'autres exemples

- Exemple: si $X_1, \dots, X_n$ n -échantillon de loi uniforme sur $[\theta, \theta + 1]$, on peut montrer (en exercice) que tout estimateur de θ compris entre $X_{(n)} - 1$ et $X_{(1)}$ est un e.m.v. (en exercice). Il n'y a alors pas unicité de cet e.m.v.
- Exemple: familles exponentielles

Famille exponentielle canonique de paramètre $\theta_1, \dots, \theta_p$, statistique canonique $T = (T_1, \dots, T_p)$ et fonction de log-partition $\beta(\theta)$.

Si β est de classe C^2 sur Θ et possède une matrice hessienne définie négative alors l'e.m.v. existe, est unique, et vérifie

$$\frac{\partial}{\partial \theta_i} \beta(\hat{\theta}) = -\frac{1}{n} \sum_{j=1}^n T_i(X_j) \quad (56)$$

Condition suffisante d'existence de l'e.m.v.

- L'e.m.v. peut ne pas exister, par exemple si L non continue en θ .

Theorem (Condition suffisante d'existence de l'e.m.v.)

Si Θ compact et l continue sur Θ , alors l'e.m.v. existe.

Theorem

Sous les trois hypothèses de régularité ($\star$) suivantes:

- *Le modèle est identifiable et homogène.*
- *Θ est un ouvert ou un compact de $\mathbb{R}^p$.*
- *Pour tout x , la fonction $L(x, \theta)$ est de classe C^2 en θ sur Θ .*

Alors e.m.v. $\hat{\theta}$ solution $\mathbb{P}_\theta$ -p.s. de

$$\begin{cases} \nabla l(x, \theta) = 0 \\ H_l(x, \theta) < 0 \end{cases} \quad (57)$$

N'assure toujours pas l'unicité de l'e.m.v. ! Une CS pour unicité: Θ espace convexe et l strictement concave sur Θ .

Reparamétrisation $\iff$ Trouver un estimateur par substitution de l'e.m.v. Quelque soit ψ , l'e.m.v. est invariant par reparamétrisation.

Theorem (de **Zehna**)

$\hat{\theta}$ e.m.v. de θ et ψ fonction *quelconque* $\Rightarrow$ l'e.m.v. de $\psi(\theta)$ est $\psi(\hat{\theta})$

Démonstration :

Soit $\eta = \psi(\theta)$. Si ψ bijective, $\tilde{L}(x, \eta) = L(x, \psi^{-1}(\eta))$. Sinon, $\tilde{L}(x, \eta) = \sup_{\theta/\psi(\theta)=\eta} L(x, \theta)$. Soit $\hat{\eta}$ e.m.v. de η .

Par définition, $\hat{\eta}$ maximise

$$\tilde{L}(x, \hat{\eta}) = \sup_{\eta} \tilde{L}(x, \eta) = \sup_{\eta} \sup_{\theta/\psi(\theta)=\eta} L(x, \theta) = L(x, \hat{\theta}) = \tilde{L}(x, \psi(\hat{\theta})).$$

Par identification, on a $\hat{\eta} = \psi(\hat{\theta})$.

□

Propriétés asymptotiques de l'e.m.v.

$\exists$ foultitude de conditions de régularité différentes $\Rightarrow$ consistance forte ou faible de l'e.m.v. Voici un énoncé dans le cadre multidimensionnel, avec des hypothèses de régularité que nous adopterons par défaut et qui sont suffisamment pratiques pour couvrir la plupart des cas classiquement étudiés.

Theorem (**Consistance forte de l'e.m.v.**)

$X_1, \dots, X_n$ modèle d'échantillonnage de v.a.i.i.d. $\sim L(x, \theta)$. Sous hypothèses de régularité ($\star\star$),

- Le modèle est identifiable et homogène.
- Θ est un ouvert ou un compact de $\mathbb{R}^p$.
- Pour tout x , la fonction $L(x, \theta)$ est différentiable en θ sur Θ .

Alors l'estimateur du maximum de vraisemblance est fortement consistant.

6. Estimateurs bayésiens

Estimateur bayésien : définition et exemples.

Cadre bayésien : modèle statistique avec paramètre muni d'une loi de probabilité *a priori*. Cette loi mesure l'information sur θ disponible avant prise en compte des observations. Inférence bayésienne construite sur loi *a posteriori* : mesure la mise à jour de l'information sur θ au fur et à mesure de la prise en compte des observations. Loi *a posteriori* par *formule de Bayes*: $L(\theta|x) \propto L(x|\theta) \times \Pi(\theta)$

$\propto = \ll \text{proportionnel à} \gg$. Exemples d'estimateurs bayésiens:

- Moyenne *a posteriori*

$$\mathbb{E}[H|X = x] = \int_{\Theta} \theta \times L(\theta|x) d\theta \quad (58)$$

- Médiane *a posteriori*

$$\mathbb{P}[H \leq \theta_{me}] = \int_{-\infty}^{\theta_{me}} L(\theta|x) d\theta = 1/2 \quad (59)$$

- Mode *a posteriori*

$$\underset{\theta}{\operatorname{argmax}} L(\theta|x) \quad (60)$$

Estimateur bayésien : échantillon gaussien.

$(X_1, \dots, X_n)$ n -échantillon de $X \sim \mathcal{N}(\theta, \sigma^2)$. paramètre inconnu θ de loi *a priori* $\mathcal{N}(a, b^2)$. σ^2 , a et b hyperparamètres connus. *a priori* informatif si b petit, c.-à-d. si θ voisin de a . Loi *a posteriori* du modèle :

$$\mathcal{N}\left(\frac{\bar{x}b^2 + a\sigma^2/n}{b^2 + \sigma^2/n}; \frac{b^2\sigma^2/n}{b^2 + \sigma^2/n}\right) \quad (61)$$

L'estimateur de θ pour la moyenne *a posteriori* est

$$\hat{\theta} = \bar{X} \frac{b^2}{b^2 + \sigma^2/n} + a \left(1 - \frac{b^2}{b^2 + \sigma^2/n}\right) \quad (62)$$

$\Rightarrow$ Moyenne pondérée de $\bar{x}$ et de la moyenne *a priori* a .

Si $n \rightarrow +\infty$, $\hat{\theta} \rightarrow \bar{X}$: approches fréquentistes et bayésiennes idem.

Si b petit (*a priori* informatif), $\hat{\theta} \approx a$. Observations presque pas prises en compte. Si *a priori* pas informatif (b grand), la loi *a priori* n'apporte pratiquement pas d'information. Toute l'information sur θ est dans $\bar{X}$.

Estimateur bayésien : échantillon de Bernoulli.

$X = (X_1, \dots, X_n)$ n -échantillon de loi de Bernoulli de paramètre θ inconnu. La vraisemblance de l'échantillon est

$$L(x, \theta) = \prod_{i=1}^n L(x_i, \theta) = \theta^s (1 - \theta)^{n-s} \quad (63)$$

avec $s = \sum_{i=1}^n x_i$. Si la loi *a priori* est une loi béta, la vraisemblance *a posteriori* suit aussi une loi béta:

$$L(\theta|x) \propto \theta^s (1 - \theta)^{n-s} \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbb{1}_{[0,1]}(\theta) \quad (64)$$

Une estimation possible de θ peut être donnée par la moyenne *a posteriori* de la loi béta:

$$\hat{\theta} = \mathbb{E}[\theta|X] = \frac{S + a}{(S + a) + (n - S + b)} \quad (65)$$

À retenir :

- ce qu'est un estimateur.
- les différents types d'estimateurs.
- la visualisation graphique de la LFGN, du TLC et du premier principe de la statistique.