

Remise à niveau - Statistique Estimation

Claire Boyer

1. Modèle statistique

2. Estimateurs et propriétés

3. Méthode des moments

4. Quantiles

Une expérience statistique est la donnée :

- ▶ d'un objet aléatoire X à valeurs dans $(E, \mathcal{E})$
- ▶ d'une famille de lois $(P_\theta)_{\theta \in \Theta}$ supposée contenir la loi de X et appelée **modèle statistique** pour la loi de X

Exemples

- ▶ Pile ou face : $X = (X_1, \dots, X_n)$ avec $X_i \sim B(\theta)$. Modèle : $(B(\theta)^{\otimes n})_{\theta \in [0,1]}$
- ▶ Taille d'hommes adultes : $X = (X_1, \dots, X_n)$ avec $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Modèle : $(\mathcal{N}(\mu, \sigma^2)^{\otimes n})_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*}$

Modèle paramétrique vs non paramétrique

- ▶ Si $\Theta \subseteq \mathbb{R}^d$, modèle **paramétrique**
- ▶ Sinon, modèle **non paramétrique**

Exemples

- ▶ Pile ou face : paramétrique ($\Theta \subseteq \mathbb{R}$).
- ▶ Taille $\sim \mathcal{N}(\mu, \sigma^2)$: paramétrique ($\subseteq \mathbb{R}^2$).
- ▶ Loi f sur $[0.5, 2.5]$ inconnue : non paramétrique.

Un modèle $(P_\theta)_{\theta \in \Theta}$ est dit **identifiable** si l'application $\theta \mapsto P_\theta$ est injective, i.e.,

$$P_\theta = P_{\theta'} \implies \theta = \theta'.$$

Exemples :

- ▶ $(B(\theta))_{\theta \in [0,1]}$ est identifiable.
- ▶ $(\mathcal{N}(\mu, \sigma^2))_{(\mu, \sigma^2)}$ est identifiable.
- ▶ Mais $(\mathcal{N}(\mu, \sigma^2))_{(\mu, \sigma)}$ ne l'est pas (symétrie en σ).

1. Modèle statistique

2. Estimateurs et propriétés

3. Méthode des moments

4. Quantiles

Soit $\theta \in \Theta$, le paramètre d'une loi $\mathbb{P}_\theta$.

Soit $\mathbb{X} = (X_1, \dots, X_n)$ un échantillon i.i.d. issu de cette loi.

- ▶ Une **statistique** $T(X_1, \dots, X_n)$ est une variable aléatoire, où T est une fonction mesurable de l'échantillon $X_1, X_2, \dots, X_n$.
- ▶ Un **estimateur** de θ est une statistique $\hat{\theta} = \hat{\theta}(\mathbb{X})$ destinée à approcher θ .

👉 Différence entre estimateur et estimation

👉 Dans la suite, on considère $\hat{\theta}_n$ un estimateur de θ .

Définition (Erreur quadratique moyenne ou risque)

$$\mathcal{R}(\hat{\theta}_n) = \mathbb{E}[(\hat{\theta}_n - \theta)^2]$$

(l'espérance est prise sous $\mathbb{P}_\theta$)

Inégalité de Markov : pour tout $c > 0$, $\mathbb{P}(|\hat{\theta}_n - \theta| \geq c) \leq \frac{\mathcal{R}(\hat{\theta}_n)}{c^2}$

Biais : $B(\hat{\theta}_n, \theta) = \mathbb{E}[\hat{\theta}_n] - \theta$

- ▶ Estimateur **sans biais** si $B(\hat{\theta}_n, \theta) = 0$
- ▶ Estimateur **asymptotiquement sans biais** si $B(\hat{\theta}_n, \theta) \xrightarrow[n \rightarrow +\infty]{} 0$

Décomposition Biais-Variance :

$$\mathcal{R}(\hat{\theta}_n) = \left[B(\hat{\theta}_n, \theta) \right]^2 + \text{Var}[\hat{\theta}_n]$$

- ▶ On dit qu'un estimateur est **consistant** si

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

- ▶ **fortement consistant** si

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

- ▶ **asymptotiquement normal** s'il existe $\sigma^2 > 0$ tel que

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

1. Modèle statistique

2. Estimateurs et propriétés

3. Méthode des moments

4. Quantiles

- ▶ Moment d'ordre k : $m_k = \mathbb{E} [X_1^k]$
- ▶ Moment empirique d'ordre k : $\hat{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
- ▶ Le moment empirique est **fortement consistant** si $\mathbb{E} [|X_1|^k]$ existe

Méthode des moments

1. Trouver un C^1 -difféomorphisme φ et un ordre k de moments tels que

$$\mathbb{E} [X_1^k] = \varphi(\theta)$$

2. Remplacer chaque moment par son estimateur empirique (méthode **plug-in**)

$$\hat{\theta}_n = \varphi^{-1}(\hat{m}_k)$$

1. **Loi normale** : on considère un échantillon $(X_1, \dots, X_n)$ de loi normale $\mathcal{N}(\mu, \sigma^2)$
2. **Loi uniforme** : on considère un échantillon $(X_1, \dots, X_n)$ de loi uniforme sur $[0, \theta^2]$ avec $\theta > 0$, i.e., de densité f_θ définie par

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

3. **Loi exponentielle** : on considère un échantillon $(X_1, \dots, X_n)$ de loi exponentielle de paramètre $\theta > 0$, i.e., de densité f_θ définie par

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

Soit X un objet aléatoire et $(P_\theta)_{\theta \in \Theta}$ un modèle statistique dominé par une mesure ν et de densité $g_\theta = \frac{dP_\theta}{d\nu}$

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

$$L_X(\theta) = g_\theta(X)$$

La **log-vraisemblance** est $\ell_X(\theta) = \log L_X(\theta) = \log g_\theta(X)$

► **Cas générique** : si $\mathbb{X} = (X_1, \dots, X_n) \sim P_\theta^{\otimes n}$, alors

$$L_{\mathbb{X}}(\theta) = L_n(\theta) = \prod_{i=1}^n g_\theta(X_i), \quad \ell_{\mathbb{X}}(\theta) = \ell_n(\theta) = \sum_{i=1}^n \ln g_\theta(X_i).$$

Cas Bernoulli (discret)

Si $\mathbb{X} \sim \mathcal{B}(\theta)^{\otimes n}$, $\theta \in (0, 1)$,

$$L_n(\theta) = \theta^{\sum_i X_i} (1 - \theta)^{n - \sum_i X_i}.$$

Cas exponentiel (continu)

Si $\mathbb{X} \sim \text{Exp}(\theta)^{\otimes n}$, $\theta > 0$:

$$L_n(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n X_i\right).$$

EMV

Un **EMV** $\hat{\theta}_n$ est un élément de Θ tel que :

$$L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta), \quad \ell_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} \ell_n(\theta).$$

Exemple 1 : cas discret. $X \sim \text{Bernoulli}(\theta)^{\otimes n}$: $\hat{\theta}_n = \bar{X}_n$.

Exemple 2 : cas continu. $X \sim \text{Exp}(\theta)^{\otimes n}$: $\hat{\theta}_n = \frac{1}{\bar{X}_n}$.

Remarque : Si $\hat{\theta}$ est un EMV de θ , alors $\varphi(\hat{\theta})$, avec φ une bijection, est un EMV pour $\varphi(\theta)$

Souvent plus facile de maximiser la log-vraisemblance $\ell_{\mathcal{X}}(\theta)$

Pour trouver l'EMV

► **Méthode artisanale**

- Étudier son tableau de variations ;
- Étudier ses points critiques, i.e., résoudre $\frac{\partial}{\partial \theta} \ell_{\mathcal{X}}(\theta) = 0$;
- Vérifier que le point maximise bien $\ell_{\mathcal{X}}$ (concavité, etc.)

► **Méthode itérative**, par exemple "montée" de gradient, ou algorithme de Newton

1. Choisir un point initial θ_0
2. Pour $t \in \mathbb{N}^*$:

$$\theta_{t+1} = \theta_t - \left(\ell_n''(\theta_t) \right)^{-1} \ell_n'(\theta_t) \quad \text{Newton step}$$

3. Arrêter selon un critère de convergence

1. Modèle statistique
2. Estimateurs et propriétés
3. Méthode des moments
4. Quantiles

Soit $X_1, \dots, X_n$ un échantillon.

- ▶ Les **statistiques d'ordre** sont

$$X_{(1)} \leq \dots \leq X_{(n)}.$$

- ▶ La **fonction de répartition empirique** est

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} = \frac{1}{n} \sup \{i : X_{(i)} \leq x\}.$$

De manière équivalente,

$$F_n(x) = \frac{1}{n} \text{Card} \{i, X_i \leq x\} = \frac{1}{n} \sup \{i, X_{(i)} \leq x\}$$

Soient $X_1, \dots, X_n$ i.i.d. de fonction de répartition F . Alors pour tout x :

- Loi

$$nF_n(x) \sim \mathcal{B}(n, F(x))$$

- Convergence

$$F_n(x) \xrightarrow[n \rightarrow \infty]{p.s.} F(x)$$

- Normalité asymptotique

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, F(x)(1 - F(x)))$$

Inverse généralisée Soit F une fonction de répartition. On appelle inverse généralisée de F la fonction définie pour tout $u \in [0, 1]$ par

$$F^{-1}(u) = \inf \{x \in \mathbb{R}, F(x) \geq u\}$$

Propriétés Soit F une f.d.r, de F^{-1} son inverse généralisée. Alors

1. $F^{-1}(0) = -\infty$
2. F^{-1} est croissante
3. F^{-1} est continue à gauche
4. Pour tout $u \in [0, 1]$,

$$F(x) \geq u \iff x \geq F^{-1}(u)$$

5. Pour tout $u \in [0, 1]$, $F \circ F^{-1}(u) = u$
6. Si F est continue, $F \circ F^{-1} = Id$
Mais si F n'est pas injective, il existe x_0 tel que $F^{-1} \circ F(x_0) < x_0$
7. Si F est injective, alors $F^{-1} \circ F = Id$
Mais si F n'est pas continue, alors il existe u_0 tel que $F^{-1} \circ F(u_0) > u_0$

Exemple 1 Soit $X \sim \mathcal{U}([0, 1])$, alors sa fonction de répartition est continue mais pas injective et

$$F^{-1} \circ F(2) = 1 < 2$$

Exemple 2 Soient $Y \sim \mathcal{N}([0, 1])$, $B \sim \mathcal{B}(1/2)$, et $X = BY$. La fonction de répartition de X n'est pas continue en 0 et

$$F \circ F^{-1}(1/2) = F(0) = 3/4 > 1/2$$

Proposition (Universalité de la loi uniforme).

- ▶ Si $U \sim \mathcal{U}([0, 1])$, alors la v.a. $F^{-1}(U)$ a pour fonction de répartition F
- ▶ Si X a pour fonction de répartition F et que F est continue, alors $F(X)$ suit une loi uniforme sur $[0, 1]$

Quantile d'ordre u correspond à $F^{-1}(u)$

Quantile empirique Pour $p \in [0, 1]$, le quantile empirique est :

$$x_p^{(n)} = F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}$$

Exemple de la médiane empirique

$$x_{1/2}^{(n)} = \begin{cases} X_{(n/2)} & \text{si } n \text{ pair,} \\ X_{((n+1)/2)} & \text{si } n \text{ impair.} \end{cases}$$

👉 Le lemme quantile est au cœur des techniques de prédiction conforme, méthode de quantification d'incertitude pour la prédiction agnostique au modèle, aux données

👉 Extension de la notion de quantile au cas multivarié avec des outils de transport optimal

Théorème

Soient $(X_1, \dots, X_n)$ un échantillon i.i.d. de fonction de répartition F , $p \in (0, 1)$, et x_p le p -quantile de F .

1. (Convergence des quantiles empiriques) Si F est strictement croissante en x_p , alors

$$x_p^{(n)} = F_n^{-1}(p) \xrightarrow[n \rightarrow +\infty]{p.s.} F^{-1}(p) = x_p$$

2. (Normalité asymptotique) Si F est dérivable en x_p de dérivée $f(x_p) > 0$, alors

$$\sqrt{n} \left(\underbrace{x_p^{(n)} - x_p}_{F_n^{-1}(p) - F^{-1}(p)} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{p(1-p)}{(f(x_p))^2} \right)$$

Exemple. Soit $X \sim \text{Exp}(\theta)$. Alors $\hat{\theta}_n = \frac{\ln(2)}{x_{1/2}^{(n)}}$ est un estimateur consistant de θ , et $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \frac{\theta^2}{(\ln(2))^2} \right)$