Eléments de statistique

# Estimation

A. Godichon-Baggioni

# I. Modèle statistique

# Modèle statistique

## Définition

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

# EXEMPLES

**Exemple 1 : pile ou face.** On a $E = \{0,1\}^n$. On a $\mathbf{X} = (X_1, \ldots, X_n)$ avec $X_i \sim \mathcal{B}(\theta)$ avec $\theta$ inconnu. Le modèle statistique est donc

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

**Exemple 2 : taille des hommes adultes.** La taille des hommes adultes est modélisée par une loi normale de paramètres $\mu, \sigma^2$ inconnus. Ici, $E = \mathbb{R}^n$ et $\mathcal{E} = \mathcal{B}(\mathbb{R}_n)$ et $\mathbf{X} = (X_1, \ldots, X_n)$ avec $X_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$ i.i.d, et

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

# Modèle paramétrique

Définition

*Si l'espace* $\Theta$ *des paramètres du modèle statistique* $(P_\theta)_{\theta\in\Theta}$ *est contenu dans* $\mathbb{R}^d$ *pour un certain* $d \in \mathbb{N}^*$*, on parle de modèle paramétrique. Sinon on parle de modèle non paramétrique.*

# EXEMPLES

**Exemple 1 : le lancer de pièce.** Le modèle est paramétrique car $[0,1] \subset \mathbb{R}$.

**Exemple 2 : taille des hommes adultes.** Le modèle est paramétrique car $\mathbb{R} \times \mathbb{R}_+^* \subset \mathbb{R}^2$.

**Exemple 3 : taille des hommes adultes.** On considère que la taille ne suit pas une loi normale mais une loi inconnue sur $[0.5, 2.5]$. On suppose que cette loi est à densité $f$ par rapport à la mesure de Lebesgue. Dans ce cas, $\Theta$ correspond à l'ensemble des densités sur $[0.5, 2.5]$ ce qui est clairement de dimension infinie. Le modèle est donc non paramétrique.

# Modèle identifiable

Définition

*Le modèle statistique* $(P_\theta)_{\theta\in\Theta}$ *est dit identifiable si l'application* $\theta \longmapsto P_\theta$ *est injective.*

**Exemple 1 : lancer de pièce.** Le modèle $(B(\theta))_{\theta\in[0,1]}$ est identifiable.

**Exemple 2 : taille des hommes adultes.** Le modèle $\left(\mathcal{N}\left(\mu,\sigma^2\right)^{\otimes n}\right)_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+^*}$ est identifiable mais pas le modèle $\left(\mathcal{N}\left(\mu,\sigma^2\right)^{\otimes n}\right)_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}^*}$

# Modèle dominé

## Définition

*Le modèle $(P_\theta)_{\theta\in\Theta}$ sur $(E,\mathcal{E})$ est dit dominé si il existe une mesure $\sigma$-finie $\lambda$ sur $(E,\mathcal{E})$ telle que, pour tout $\theta\in\Theta$, on a $P_\theta << \lambda$, i.e*

$$\forall A\in\mathcal{E},\quad \lambda(A)=0\Rightarrow P_\theta(A)=0.$$

*La mesure $\lambda$ est alors appelée mesure dominante.*

**Exemple 1 : le lancer de pièce.** Une mesure dominante de $Q_\theta=(1-\theta)\delta_0+\theta\delta_1$ est $\lambda=\delta_0+\delta_1$.

**Exemple 2 : la taille des hommes adultes.** Le modèle est dominé par la mesure de Lebesgue sur $\mathbb{R}$.

# II. Définitions

# Statistique et estimateur

On note $\mathbf{X} = (X_1, \ldots, X_n)$.

## Définition

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

**Exemple :** Lancer de pièce

**Attention !** Ne pas confondre estimateur et estimation !

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

# ERREUR QUADRATIQUE MOYENNE

## Définition

*On suppose que $\theta$ est à valeurs dans $\Theta \subset \mathbb{R}$. L'erreur quadratique moyenne (ou risque quadratique) de l'estimateur $\hat{\theta}_n$ est défini 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 $\theta$ 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 BIAIS-VARIANCE

**Proposition**

*Soit $\hat{\theta}_n$ un estimateur de $\theta$, 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 $\theta$, 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*

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

# III. Méthode des moments

# Méthode des moments

## Proposition

*Soit $\Theta$ un intervalle ouvert de $\mathbb{R}$ et $\theta \in \Theta$. Soit $\varphi$ un $C^1$-difféomorphisme de $\Theta$ 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 $\theta$, 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 $\theta$ 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 $\varphi$ 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 $f_\theta$ 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 $f_\theta$ définie pour tout $x \in \mathbb{R}$,

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

# IV. Méthode du Maximum de Vraisemblance

# (Log) Vraisemblance

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

## Définition

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

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

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

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

# REMARQUES ET EXEMPLES

**Remarque :** Si $\mathbf{X} = (X_1, \ldots, X_n) \sim P_\theta^{\otimes n}$, on a

$$L_{\mathbf{X}}(\theta) = \prod_{i=1}^{n} g_\theta\left(X_i\right) \qquad l_{\mathbf{X}}(\theta) = \sum_{i=1}^{n} \ln\left(g_\theta\left(X_i\right)\right)$$

et on note $L_{\mathbf{X}}(\theta) = L_n(\theta)$ et $l_{\mathbf{X}}(\theta) = l_n(\theta)$.

**Exemple 1 : cas discret.** Soit $\mathbf{X} = (X_1, \ldots, X_n) \sim \mathbb{B}(\theta)^{\otimes n}$ avec $\theta \in (0, 1)$. On a

$$L_n(\theta) = \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n}$$

**Exemple 2 : cas continu.** Soit $\mathbf{X} = (X_1, \ldots, X_n) \sim \mathcal{E}(\theta)^{\otimes n}$, avec $\theta > 0$. On a

$$L_n(\theta) = \theta^n \exp\left(-n\overline{X}_n\theta\right).$$

# Estimateur du Maximum de Vraisemblance

## Définition

*Un estimateur du maximum de vraisemblance (EMV) de $\theta$, si il existe, est un élément $\hat{\theta}_n$ de $\Theta$ tel que*

$$L_n\left(\hat{\theta}_n\right) = \sup_{\theta \in \Theta} L_n(\theta) \Leftrightarrow l_n\left(\hat{\theta}_n\right) = \sup_{\theta \in \Theta} l_n(\theta).$$

**Exemple 1 : cas discret.** Soit $\mathbf{X} = (X_1, \ldots, X_n) \sim \mathbb{B}(\theta)^{\otimes n}$ avec $\theta \in (0, 1)$. On a

$$\hat{\theta}_n = \overline{X}_n.$$

**Exemple 2 : cas continue.** Soit $\mathbf{X} = (X_1, \ldots, X_n) \sim \mathcal{E}(\theta)^{\otimes n}$, avec $\theta > 0$. On a

$$\hat{\theta}_n = \overline{X}_n^{-1}.$$

# Remarque

**Remarque :** Si $\hat{\theta}_n$ est un estimateur du maximum de vraisemblance de $\theta$, $\varphi\left(\hat{\theta}_n\right)$ est un estimateur du maximum de vraisemblance de $\varphi(\theta)$.

**Exemple : la loi exponentielle.** Soit $\mathbf{X} = (X_1, \ldots, X_n) \sim \mathcal{E}(\theta)^{\otimes n}$, avec $\theta > 0$, alors $\overline{X}_n$ est l'estimateur du maximum de vraisemblance de $\mathbb{E}\left[X_1\right] = \theta^{-1}$.

# 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

$$\frac{\partial}{\partial\theta} l_{\mathbf{x}}(\theta) = 0.$$

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

# Algorithme de Newton

Il arrive que l'on ne soit pas capable de calculer explicitement l'estimateur du maximum de vraisemblance, et ce, même si il existe. On peut alors l'approcher à l'aide d'algorithmes d'optimisation, et notamment l'algorithme de Newton :

- On choisit un point initial $\theta_0$.
- Pour tout $t \in \mathbb{N}^*$,

$$\theta_{t+1} = \theta_t - \left(l_n''\left(\theta_t\right)\right)^{-1} l_n'\left(\theta_t\right)$$

- On arrête quand un critère de convergence est satisfait.

# V. Quantiles

# STATISTIQUES D'ORDRE

## Définition

*Soit* $X_1, \ldots, X_n$ *un échantillon. Les n statistiques d'ordre* $X_{(1)}, \ldots, X_{(n)}$ *s'obtiennent en rangeant l'échantillon dans l'ordre, i.e on a*

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

# Fonction de répartition empirique

La fonction de répartition empirique $F_n$ d'un échantillon $X_1, \ldots, X_n$ est définie pour tout réel $x$ par

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{]-\infty,x]}(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{]-\infty,x]}\left(X_{(i)}\right).$$

De manière équivalente, on a

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

# CONVERGENCE

## Proposition

*Soit $\mathbf{X} = (X_1, \ldots, X_n)$ des variables aléatoires indépendantes et identiquement distribuées de fonction de répartition $F$. Pour tout réel $x$, on a*

- *Loi :*
$$nF_n(x) \sim \mathcal{B}\left(n, F(x)\right).$$
- *Convergence :*
$$F_n(x) \xrightarrow[n\to+\infty]{p.s} F(x).$$
- *Normalité asymptotique :*
$$\sqrt{n}\left(F_n(x) - F(x)\right) \xrightarrow[n\to=\infty]{\mathcal{L}} \mathcal{N}\left(0, F(x)(1-F(x))\right).$$

# Inverse généralisée

## Définition

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

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

*$F^{-1}(u)$ est appelé le quantile d'ordre $u$.*

# PROPRIÉTÉS

## Proposition

*Soit $F$ une fonction de répartition et $F^{-1}$ son inverse généralisée. Alors :*

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

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

5. *Pour tout $u \in [0, 1]$, on a $\left(F \circ F^{-1}\right)(u) \geq u$ et :*
   - 5.1 *Si $F$ est continue, $F \circ F^{-1} = Id$, mais si $F$ n'est pas injective, il existe $x_0$ tel que $\left(F^{-1} \circ F\right)(x_0) < x_0$.*
   - 5.2 *si $F$ est injective, alors $F^{-1} \circ F = Id$, mais si elle n'est pas continue, il existe $u_0$ tel que $\left(F \circ F^{-1}\right)(u_0) > u_0$.*

# EXEMPLES

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

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

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

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

# MÉTHODE INVERSE

## Proposition

*Soit $U$ une variable aléatoire suivant une loi uniforme sur $[0,1]$, $F$ une fonction de répartition et $F^{-1}$ son inverse généralisée. Alors*

- *la variable aléatoire $X = F^{-1}(U)$ a pour fonction de répartition $F$.*
- *Si $X$ a pour fonction de répartition $F$ et si $F$ est continue, alors $F(X)$ suit une loi uniforme sur $[0,1]$.*

# QUANTILES EMPIRIQUES

## Définition

*Soit $(X_1, \ldots, X_n)$ un échantillon et $F_n$ la fonction de répartition empirique associée. Pour tout $p \in [0, 1]$, on note $x_p(n)$ le quantile empirique associé, i.e*

$$x_p(n) = F_n^{-1}(p) = \inf \{x \in \mathbb{R}, F_n(x) \geq p\} = X_{(\lceil pn \rceil)},$$

*ou $\lceil . \rceil$ est la partie entière supérieure.*

**Exemple :** La médiane empirique est $X_{(n/2)}$ si $n$ est pair et $X_{(n+1)/2}$ sinon.

# CONVERGENCE

## Théorème

*Soient $(X_1, \ldots, X_n)$ i.i.d de fonction de répartition $F$, $p \in (0,1)$ et $x_p$ le $p$-quantile de $F$, alors :*

- *Si $F$ est strictement croissante en $x_p$, alors*

$$x_p(n) \xrightarrow[n \to +\infty]{p.s} x_p.$$

- *Si $F$ est dérivable en $x_p$ de dérivée $f\left(x_p\right) > 0$, alors*

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

**Exemple :** Soit $X \sim \mathcal{E}(\theta)$ avec $\theta > 0$. Alors $\hat{\theta}_n = \frac{\ln(2)}{x_n\left(\frac{1}{2}\right)}$ est un estimateur consistant de $\theta$ et

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