

Estimation

Lorenzo Segoni

18 novembre 2025

Table des matières

| | | |
|----------|--|-----------|
| 1 | Méthodes d'estimation | 1 |
| 1.1 | Modèles et estimations statistiques | 1 |
| 1.1.1 | Modèles statistiques | 1 |
| 1.1.2 | Estimateur statistique | 3 |
| 1.1.3 | Propriétés asymptotiques | 7 |
| 1.2 | Propriétés asymptotiques | 9 |
| 1.2.1 | Estimation de θ à partir d'un estimateur de $g(\theta)$ | 10 |
| 1.3 | Méthode des moments | 11 |
| 2 | Méthode du maximum de vraisemblance | 15 |

Chapitre 1

Méthodes d'estimation

1.1 Modèles et estimations statistiques

1.1.1 Modèles statistiques

Première étape de la démarche statistique : la modélisation

On observe une donnée x .

Définition 1.1 (Modèle statistique). On suppose que x est la réalisation d'une variable aléatoire X de loi P_X (inconnue). On postule que P_X appartient à une famille de lois de probabilités spécifiée :

$$\{P_\theta, \theta \in \Theta\}.$$

Remarque 1.1. Un modèle est toujours faux : la réalité est trop complexe pour être décrite parfaitement. L'enjeu est que l'approximation apportée par le modèle soit suffisamment fidèle pour répondre à la question posée. En pratique, ce sont souvent les spécialistes du domaine (médecine, biologie, économie, etc.) qui proposent ou discutent le modèle.

Deuxième étape : l'inférence statistique

À partir du modèle $\{P_\theta, \theta \in \Theta\}$ et d'une observation X , on cherche à obtenir des informations sur la loi réelle P_X .

Définition 1.2 (Modèle statistique). Un modèle statistique est un triplet :

$$(E, \mathcal{E}, (P_\theta)_{\theta \in \Theta})$$

où :

- (E, \mathcal{E}) est un espace mesurable (dans ce cours, on retiendra surtout que $X \in E$) ;
- $(P_\theta)_{\theta \in \Theta}$ est une famille de lois de probabilité sur E ;
- Θ est l'ensemble des paramètres.

Si $\Theta \subset \mathbb{R}^d$ avec $d \geq 1$, le modèle est dit **paramétrique**. Sinon, il est dit **non paramétrique**.

Exemple 1.1. 1. **Jeu de pile ou face répété n fois :**

$$E = \{0, 1\}^n, \quad \Theta =]0, 1[, \quad P_\theta = (\mathcal{B}(\theta))^{\otimes n}.$$

Une observation est $X = (X_1, \dots, X_n) \in \{0, 1\}^n$, avec X_i des variables aléatoires indépendantes de loi $\mathcal{B}(\theta)$.

2. **Modélisation de la taille d'individus par une loi normale.** On observe sur un échantillon de taille n :

$$E = \mathbb{R}^n, \quad X = (X_1, \dots, X_n), \quad X_i \text{ variables aléatoires indépendantes de loi } \mathcal{N}(m, \sigma^2),$$

avec

$$\Theta = \{(m, \sigma^2) : m \in \mathbb{R}, \sigma^2 > 0\} \subset \mathbb{R}^2.$$

3. **Même population, mais modélisation par une loi à densité sur $[0.5, 2.5]$:**

$$\Theta = \{f : [0.5, 2.5] \rightarrow \mathbb{R}_+ \mid f \text{ est une densité}\}.$$

Il s'agit alors d'un **modèle non paramétrique**.

Valeur vraie du paramètre. Le point clé de la modélisation est qu'il existe une valeur vraie $\theta^* \in \Theta$ telle que :

$$P_X = P_{\theta^*}.$$

Mais θ^* est inconnue

Dans toute la suite, on notera simplement θ pour θ^* .

On utilisera les notations :

$$\mathbb{P}_\theta \quad \text{et} \quad \mathbb{E}_\theta$$

pour bien marquer la dépendance en θ . Et on a que X dépend de θ

$$\begin{aligned} P(X \in A) &= P_X(A) = P_{\theta^*}(A), \\ P_\theta(X \in A) &= P_\theta(A). \end{aligned}$$

Remarque 1.2. On considérera souvent une observation X comme un n -échantillon de variables aléatoires *iid* de loi Q_θ :

$$P_\theta = (Q_\theta)^{\otimes n}.$$

On peut alors définir le modèle statistique comme la famille $\{Q_\theta, \theta \in \Theta\}$.

Définition 1.3 (Identifiabilité). Un modèle $\{P_\theta, \theta \in \Theta\}$ est dit **identifiable** si :

$$P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2.$$

1.1.2 Estimateur statistique

On suppose que le modèle $\{P_\theta, \theta \in \Theta\}$ est identifiable. On observe X et on s'intéresse à une quantité d'intérêt $g(\theta)$, où :

$$g : \Theta \rightarrow \mathbb{R}^p, \quad p \geq 1.$$

Définition 1.4 (Statistique et estimateur). — Une **statistique** $T(X)$ est une fonction mesurable de X (éventuellement de paramètres connus), qui ne dépend pas de θ .

— Un **estimateur** de $g(\theta)$ est une statistique $\hat{g} = h(X)$ à valeurs dans \mathbb{R}^p (avec h mesurable).

Exemple 1.2. Pour un n -échantillon (X_1, \dots, X_n) , les objets suivants sont des sta-

tistiques :

$$\bar{X}_n, \quad X_1, \quad X_2^2 \cdot X_8, \quad 0.$$

Remarque 1.3. Si l'on dispose d'une réalisation x de X , alors $\hat{g}(x)$ est appelé un **estimé** de $g(\theta)$. Par abus, on note souvent \hat{g} à la fois l'estimateur (variable aléatoire) et son estimé (valeur numérique).

Notation 1.1 (Estimateur plug-in). Si $\hat{\theta}$ est un estimateur de θ , un estimateur naturel de $g(\theta)$ est :

$$g(\hat{\theta}).$$

On appelle cela un **estimateur plug-in**. Sa qualité dépend de la qualité de $\hat{\theta}$ et de la régularité de g .

Définition 1.5 (Risque quadratique). Le risque quadratique de \hat{g} est défini, pour tout $\theta \in \Theta$, par :

$$R(\hat{g}, g(\theta)) = \mathbb{E}_\theta[(\hat{g} - g(\theta))^2] \in [0, +\infty].$$

Remarque 1.4. Lien avec inégalité de Markov : Par l'inégalité de Markov,

$$\mathbb{P}_\theta(|\hat{g} - g(\theta)| > \varepsilon) \leq \frac{R(\hat{g}, g(\theta))}{\varepsilon^2}.$$

Ainsi, plus le risque quadratique est petit, plus l'estimateur est concentré autour de la valeur vraie.

Existe-t-il un estimateur optimal ? On pourrait souhaiter qu'il existe un estimateur \hat{g} meilleur que tous les autres pour tout θ , c'est-à-dire :

$$\exists \hat{g} \text{ tel que } \forall \theta \in \Theta, \forall \hat{g}' \text{ estimateur, } R(\hat{g}, g(\theta)) \leq R(\hat{g}', g(\theta))$$

La réponse est en général **non**.

Exemple 1.3 (Comparaison de deux estimateurs). Considérons le modèle :

$$\{(\mathcal{N}(\theta, 1))^{\otimes n}, \theta \in \mathbb{R}\}.$$

Deux estimateurs de θ :

$$\hat{\theta}_1 = \bar{X}_n, \quad \hat{\theta}_2 = 0.$$

On a :

$$R(\hat{\theta}_1, \theta) = \frac{1}{n}, \quad R(\hat{\theta}_2, \theta) = \theta^2.$$

De plus :

$$\mathbb{E}_\theta[\hat{\theta}_1] = \theta, \quad \text{Var}_\theta(\hat{\theta}_1) = \frac{1}{n}.$$

Conclusion :

- $\hat{\theta}_1$ est sans biais avec un risque constant $1/n$.
- $\hat{\theta}_2$ est biaisé, mais pour $|\theta| < \frac{1}{\sqrt{n}}$, son risque est inférieur à celui de $\hat{\theta}_1$.

Ainsi, même un estimateur « stupide » comme $\hat{\theta}_2$ peut être meilleur que l'estimateur usuel \bar{X}_n pour certaines valeurs de θ . Bien sûr, plus n est grand, plus l'intervalle de θ où $\hat{\theta}_2$ est meilleur devient petit.

Proposition 1.1 (Décomposition biais-variance). *Pour tout $\theta \in \Theta$, on a :*

$$R(\hat{g}, g(\theta)) = \left(\mathbb{E}_\theta[\hat{g}] - g(\theta) \right)^2 + \mathbb{E}_\theta \left[(\hat{g} - \mathbb{E}_\theta[\hat{g}])^2 \right].$$

Autrement dit :

$$R(\hat{g}, g(\theta)) = B(\hat{g}, g(\theta))^2 + \text{Var}_\theta(\hat{g}),$$

où $B(\hat{g}, g(\theta))$ est le biais de \hat{g} .

Définition 1.6 (Biais). Le biais d'un estimateur \hat{g} est défini par la fonction :

$$\theta \mapsto B(\hat{g}, g(\theta)) = \mathbb{E}_\theta[\hat{g}] - g(\theta).$$

On dit que \hat{g} est **sans biais** si, pour tout $\theta \in \Theta$:

$$\mathbb{E}_\theta[\hat{g}] = g(\theta).$$

Idée de la preuve. On écrit :

$$(\hat{g} - g(\theta))^2 = \left(\hat{g} - \mathbb{E}_\theta[\hat{g}] + \mathbb{E}_\theta[\hat{g}] - g(\theta) \right)^2,$$

puis on développe et on prend l'espérance. □

Démonstration. On part de l'identité :

$$\hat{g} - g(\theta) = (\hat{g} - \mathbb{E}_\theta[\hat{g}]) + (\mathbb{E}_\theta[\hat{g}] - g(\theta)).$$

En élevant au carré, on obtient :

$$(\hat{g} - g(\theta))^2 = (\hat{g} - \mathbb{E}_\theta[\hat{g}])^2 + 2(\hat{g} - \mathbb{E}_\theta[\hat{g}])(\mathbb{E}_\theta[\hat{g}] - g(\theta)) + (\mathbb{E}_\theta[\hat{g}] - g(\theta))^2.$$

En prenant l'espérance, le terme croisé disparaît car

$$\mathbb{E}_\theta[\hat{g} - \mathbb{E}_\theta[\hat{g}]] = 0.$$

Ainsi :

$$\mathbb{E}_\theta[(\hat{g} - g(\theta))^2] = \mathbb{E}_\theta[(\hat{g} - \mathbb{E}_\theta[\hat{g}])^2] + (\mathbb{E}_\theta[\hat{g}] - g(\theta))^2.$$

On reconnaît la variance et le carré du biais, ce qui donne :

$$R(\hat{g}, g(\theta)) = \text{Var}_\theta(\hat{g}) + B(\hat{g}, g(\theta))^2.$$

□

Remarque 1.5. Si \hat{g} est sans biais, c'est-à-dire $\mathbb{E}_\theta[\hat{g}] = g(\theta)$, alors $B(\hat{g}, g(\theta)) = 0$ et

$$R(\hat{g}, g(\theta)) = \text{Var}_\theta(\hat{g}).$$

Interprétation :

- Le **biais** mesure l'erreur systématique faite par l'estimateur (erreur d'approximation).
- La **variance** mesure les fluctuations aléatoires de l'estimateur autour de sa moyenne (erreur d'estimation).

Remarques sur les estimateurs sans biais

Si \hat{g} est sans biais :

$$R(\hat{g}, g(\theta)) = \text{Var}_\theta(\hat{g}).$$

On s'intéresse alors aux estimateurs sans biais ayant la variance minimale. On appelle un tel estimateur un **estimateur UVMB** (Uniformément de Variance Minimale parmi les estimateurs Sans Biais).

Il existe des méthodes (exhaustivité, complétude, théorème de Lehmann-Scheffé) pour caractériser directement de tels estimateurs.

Remarque 1.6. Un estimateur sans biais n'est pas toujours le meilleur choix : un estimateur **avec biais** peut avoir un risque quadratique plus petit. De plus, il n'existe pas toujours d'estimateur sans biais pour une quantité donnée.

Exemple 1.4 (Impossibilité d'un estimateur sans biais). Soit $X \sim \mathcal{E}(\theta)$ une loi exponentielle de paramètre $\theta > 0$. Supposons qu'il existe un estimateur sans biais $\hat{\theta} = h(X)$ de θ . Alors, pour tout $\theta > 0$, il faudrait :

$$\int_0^\infty h(x) \theta e^{-\theta x} dx = \theta.$$

Or, cette condition n'admet pas de solution mesurable h , donc un tel estimateur n'existe pas.

Remarque 1.7. Même si $\hat{\theta}$ est un estimateur sans biais de θ , en général $g(\hat{\theta})$ n'est pas un estimateur sans biais de $g(\theta)$. Ainsi, la propriété « sans biais » n'est pas stable par transformation.

1.1.3 Propriétés asymptotiques

On se place avec n observations. On pose \hat{g}_n un estimateur de $g(\theta)$.

Définition 1.7 (Consistance). On dit que la suite $(\hat{g}_n)_{n \in \mathbb{N}}$ est **consistante** si pour tout $\theta \in \Theta$,

$$\hat{g}_n \xrightarrow[\mathbb{P}_\theta]{n \rightarrow +\infty} g(\theta).$$

Elle est dite **fortement consistante** si pour tout $\theta \in \Theta$,

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

Remarque 1.8. On dira seulement que \hat{g}_n est consistant si $(\hat{g}_n)_{n \in \mathbb{N}}$ est consistante.

- La convergence doit avoir lieu pour tout $\theta \in \Theta$.
- La consistance est une propriété “faible”. S'il existe un estimateur consistant,

alors il en existe une infinité : si \hat{g}_n est consistant et si $(a_n)_{n \in \mathbb{N}}$ est telle que $a_n \rightarrow 1$, alors $a_n \hat{g}_n$ est consistant.

— En revanche, les estimateurs non consistants sont à exclure.

Définition 1.8 (Estimateur asymptotiquement sans biais). On suppose que pour tout $\theta \in \Theta$ et pour tout $n \in \mathbb{N}$,

$$\mathbb{E}_\theta [\|\hat{g}_n\|] < +\infty.$$

On dit que \hat{g}_n est **asymptotiquement sans biais** si pour tout $\theta \in \Theta$,

$$\mathbb{E}_\theta[\hat{g}_n] \xrightarrow{n \rightarrow +\infty} g(\theta).$$

Définition 1.9 (Normalité asymptotique). On dit que \hat{g}_n est **asymptotiquement normal** si pour tout $\theta \in \Theta$, il existe $\Sigma(\theta) > 0$ tel que

$$\sqrt{n}(\hat{g}_n - g(\theta)) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathcal{N}(0, \Sigma(\theta)).$$

Remarque 1.9. — Si \hat{g}_n est asymptotiquement normal, alors \hat{g}_n est consistant.

— Plus généralement, on peut remplacer “ \sqrt{n} ” par a_n où $a_n \xrightarrow{n \rightarrow +\infty} +\infty$. On dit alors que $\frac{1}{a_n}$ est la vitesse de convergence asymptotique.

Proposition 1.2. Soit \hat{g}_n un estimateur asymptotiquement normal de $g(\theta)$ de variance asymptotique $\Sigma(\theta)$. Soit $\hat{\Sigma}_n$ un estimateur de $\Sigma(\theta)$ consistant, c'est-à-dire pour tout $\theta \in \Theta$,

$$\hat{\Sigma}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_\theta} \Sigma(\theta).$$

Alors

$$\frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\hat{\Sigma}_n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathcal{N}(0, 1).$$

Démonstration. Conséquence du **lemme de Slutsky**.

On écrit

$$\frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\hat{\Sigma}_n}} = \frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\Sigma(\theta)}} \cdot \frac{\sqrt{\Sigma(\theta)}}{\sqrt{\hat{\Sigma}_n}},$$

avec :

$$\frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\Sigma(\theta)}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathcal{N}(0, 1) \text{ par hypothèse,}$$

$$\frac{\sqrt{\Sigma(\theta)}}{\sqrt{\hat{\Sigma}_n}} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_\theta} 1 \text{ par consistance de } \hat{\Sigma}_n.$$

Le lemme de Slutsky permet de conclure. \square

1.2 Propriétés asymptotiques

On se place avec n observations. On pose \hat{g}_n un estimateur de $g(\theta)$.

Définition 1.10 (Consistance). On dit que la suite $(\hat{g}_n)_{n \in \mathbb{N}}$ est **consistante** si pour tout $\theta \in \Theta$,

$$\hat{g}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_\theta} g(\theta).$$

Elle est dite **fortement consistante** si pour tout $\theta \in \Theta$,

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

Remarque 1.10. On dira seulement que \hat{g}_n est consistant si $(\hat{g}_n)_{n \in \mathbb{N}}$ est consistante.

- La convergence doit avoir lieu pour tout $\theta \in \Theta$.
- La consistance est une propriété “faible”. S’il existe un estimateur consistant, alors il en existe une infinité : si \hat{g}_n est consistant et si $(a_n)_{n \in \mathbb{N}}$ est telle que $a_n \rightarrow 1$, alors $a_n \hat{g}_n$ est consistant.
- En revanche, les estimateurs non consistants sont à exclure.

Définition 1.11 (Estimateur asymptotiquement sans biais). On suppose que pour tout $\theta \in \Theta$ et pour tout $n \in \mathbb{N}$,

$$\mathbb{E}_\theta [\|\hat{g}_n\|] < +\infty.$$

On dit que \hat{g}_n est **asymptotiquement sans biais** si pour tout $\theta \in \Theta$,

$$\mathbb{E}_\theta[\hat{g}_n] \xrightarrow[n \rightarrow +\infty]{} g(\theta).$$

Définition 1.12 (Normalité asymptotique). On dit que \hat{g}_n est **asymptotiquement**

normal si pour tout $\theta \in \Theta$, il existe $\Sigma(\theta) > 0$ tel que

$$\sqrt{n}(\hat{g}_n - g(\theta)) \xrightarrow[\text{loi}]{n \rightarrow +\infty} \mathcal{N}(0, \Sigma(\theta)).$$

Remarque 1.11. — Si \hat{g}_n est asymptotiquement normal, alors \hat{g}_n est consistant.

— Plus généralement, on peut remplacer “ \sqrt{n} ” par a_n où $a_n \xrightarrow{n \rightarrow +\infty} +\infty$. On dit alors que $\frac{1}{a_n}$ est la vitesse de convergence asymptotique.

Proposition 1.3. Soit \hat{g}_n un estimateur asymptotiquement normal de $g(\theta)$ de variance asymptotique $\Sigma(\theta)$. Soit $\hat{\Sigma}_n$ un estimateur de $\Sigma(\theta)$ consistant, c'est-à-dire pour tout $\theta \in \Theta$,

$$\hat{\Sigma}_n \xrightarrow[\mathbb{P}_\theta]{n \rightarrow +\infty} \Sigma(\theta).$$

Alors

$$\frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\hat{\Sigma}_n}} \xrightarrow[\text{loi}]{n \rightarrow +\infty} \mathcal{N}(0, 1).$$

Démonstration. Conséquence du **lemme de Slutsky**.

On écrit

$$\frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\hat{\Sigma}_n}} = \frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\Sigma(\theta)}} \cdot \frac{\sqrt{\Sigma(\theta)}}{\sqrt{\hat{\Sigma}_n}},$$

avec :

$$\begin{aligned} & \text{— } \frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sqrt{\Sigma(\theta)}} \xrightarrow[\text{loi}]{n \rightarrow +\infty} \mathcal{N}(0, 1) \text{ par hypothèse,} \\ & \text{— } \frac{\sqrt{\Sigma(\theta)}}{\sqrt{\hat{\Sigma}_n}} \xrightarrow[\mathbb{P}_\theta]{n \rightarrow +\infty} 1 \text{ par consistance de } \hat{\Sigma}_n. \end{aligned}$$

Le lemme de Slutsky permet de conclure. □

1.2.1 Estimation de θ à partir d'un estimateur de $g(\theta)$

Question : \hat{g}_n estimateur de $g(\theta)$. Comment estimer θ ? Et quelles propriétés pour l'estimateur?

Proposition 1.4 (Méthode delta inverse). Soit Θ un intervalle ouvert de \mathbb{R} .

— Soit \hat{g}_n estimateur de $g(\theta)$ tel que pour tout $\theta \in \Theta$,

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

- On suppose que $g : \Theta \rightarrow g(\Theta)$ est un \mathcal{C}^1 -difféomorphisme (c'est-à-dire g bijective, \mathcal{C}^1 , et $g^{-1} \in \mathcal{C}^1$). On définit :

$$\hat{\theta}_n = g^{-1}(\hat{g}_n).$$

Alors,

- $\hat{\theta}_n$ est consistant pour l'estimation de θ .
 — Pour tout $\theta \in \Theta$,

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[\text{loi}]{n \rightarrow +\infty} \frac{\sigma}{|g'(\theta)|} \mathcal{N}(0, 1).$$

Démonstration. — g continue et bijective $\Theta \rightarrow g(\Theta)$ donc $g(\Theta)$ est un intervalle ouvert.

$$\mathbb{P}(\hat{g}_n \in g(\Theta)) \xrightarrow{n \rightarrow +\infty} 1.$$

Donc $\hat{\theta}_n$ est bien défini (avec une probabilité qui tend vers 1).

- **Consistance de $\hat{\theta}_n$** : g^{-1} continue et \hat{g}_n consistant donc par composition, $\hat{\theta}_n = g^{-1}(\hat{g}_n)$ est consistant.
 — **$\hat{\theta}_n$ asymptotiquement normal** : méthode delta appliquée à g^{-1} .

On a $g^{-1}(g(\theta)) = \theta$ et pour rappel,

$$(g^{-1})'(y) = \frac{1}{g'(g^{-1}(y))}.$$

Donc

$$(g^{-1})'(g(\theta)) = \frac{1}{g'(g^{-1}(g(\theta)))} = \frac{1}{g'(\theta)}.$$

Par la méthode delta,

$$\sqrt{n} (\hat{\theta}_n - \theta) = \sqrt{n} (g^{-1}(\hat{g}_n) - g^{-1}(g(\theta))) \xrightarrow[\text{loi}]{n \rightarrow +\infty} (g^{-1})'(g(\theta)) \mathcal{N}(0, \sigma^2) = \frac{1}{g'(\theta)} \mathcal{N}(0, \sigma^2).$$

Ce qui équivaut à

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[\text{loi}]{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{\sigma^2}{(g'(\theta))^2}\right) = \frac{\sigma}{|g'(\theta)|} \mathcal{N}(0, 1).$$

□

1.3 Méthode des moments

Définition 1.13 (Estimateur par la méthode des moments). — Soit un échantillon X_1, \dots, X_n de loi \mathbb{P}_θ , $\theta \in \Theta$.

— Soit ϕ une fonction telle que $\mathbb{E}_\theta[|\phi(X_1)|] < +\infty$.

On souhaite estimer

$$g(\theta) = \mathbb{E}_\theta[\phi(X_1)].$$

On appelle **estimateur par la méthode des moments** :

$$\hat{g}_n = \frac{1}{n} \sum_{i=1}^n \phi(X_i).$$

Notation 1.2. — \hat{g}_n est consistant (fortement) par la loi des grands nombres.

— \hat{g}_n est sans biais (par linéarité de l'espérance).

— Plus généralement, si \hat{g}_n est construit par la méthode des moments comme ci-dessus, et si h est une fonction, on dira aussi que

$$h(\hat{g}_n)$$

est un estimateur par la **méthode des moments** pour l'estimation de $h(g(\theta))$.

Remarque 1.12. À priori, on ne connaît pas les propriétés de cet estimateur (biais, variance, normalité asymptotique). Il faudra étudier ces propriétés au cas par cas, éventuellement en utilisant la méthode delta.

Exemple 1.5. 1. Probabilité d'un événement

Soit A un événement. On souhaite estimer

$$g(\theta) = \mathbb{P}_\theta(X_1 \in A) = \mathbb{E}_\theta[\mathbb{1}_{X_1 \in A}],$$

où

$$\mathbb{1}_{X_1 \in A} = \begin{cases} 1 & \text{si } X_1 \in A, \\ 0 & \text{sinon.} \end{cases}$$

Par la méthode des moments :

$$\hat{g}_{A,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in A}.$$

Cet estimateur est la fréquence empirique de l'événement A .

2. Loi uniforme

Soit X_i pour $i \in [1, n]$ de loi $\mathcal{U}([\theta - 1, \theta + 1])$ avec $\theta \in \mathbb{R}$.

Alors

$$\mathbb{E}_\theta[X_1] = \theta = g(\theta).$$

Par la méthode des moments :

$$\hat{g}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

Chapitre 2

Méthode du maximum de vraisemblance

On se place dans le cadre de l'estimation de $g(\theta) = \theta$.

Définition 2.1 (Vraisemblance). On appelle **vraisemblance** associée à l'observation X , la fonction :

$$\begin{aligned} L : \Theta &\rightarrow \mathbb{R}, \\ \theta &\mapsto L(\theta; X) \end{aligned}$$

définie par :

— si X est une variable aléatoire **discrète**, alors :

$$L(\theta; X) = p_\theta(X) \quad \text{où} \quad p_\theta(x) = \mathbb{P}_\theta(X = x)$$

— si X est une variable aléatoire **à densité**, alors :

$$L(\theta; X) = f_\theta(X) \quad \text{où} \quad f_\theta \text{ est la densité de } X.$$

Remarque 2.1. Si $X = (X_1, \dots, X_n)$ est un échantillon de variables aléatoires i.i.d., alors :

— si les X_i sont discrètes :

$$L(\theta; X) = \prod_{i=1}^n p_\theta(X_i)$$

— si les X_i admettent une densité :

$$L(\theta; X) = \prod_{i=1}^n f_{\theta}(X_i)$$

Exemple 2.1. On considère une variable aléatoire $X \sim \mathcal{B}(15, p)$ avec p inconnu.

On observe $x = 5$.

La fonction de vraisemblance est :

$$L(p; 5) = \mathbb{P}_p(X = 5) = \binom{15}{5} p^5 (1-p)^{10}$$

C'est la probabilité d'observer la valeur 5 lorsque le paramètre est p .

| p | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 |
|-----------|------|------|------|------|------|
| $L(p; 5)$ | 0,01 | 0,10 | 0,21 | 0,19 | 0,09 |

— Quand $p = 0,1$, il y a environ 1 chance sur 100 d'observer $x = 5$.

— Quand $p = 0,3$, il y a environ 21 chances sur 100 d'observer $x = 5$.

Il est donc plus « vraisemblable » que la vraie valeur de p soit proche de 0,3 plutôt que de 0,1.

En suivant ce raisonnement, la valeur de p la plus vraisemblable est celle qui maximise la probabilité d'observer $x = 5$, c'est-à-dire qui maximise la vraisemblance.

On se place, pour simplifier, sur la fonction :

$$\ell(p) = \ln L(p; 5)$$

Alors :

$$\frac{d\ell(p)}{dp} = \frac{x}{p} - \frac{15-x}{1-p} = \frac{x-15p}{p(1-p)}$$

Cette dérivée s'annule en :

$$p = \frac{x}{15}$$

Ici, avec $x = 5$, on obtient :

$$\hat{p} = \frac{1}{3}$$

Ainsi, la valeur de p la plus vraisemblable au vu de l'observation $x = 5$ est :

$$\boxed{p = \frac{1}{3}}$$

Définition 2.2 (Estimateur du maximum de vraisemblance). On appelle **estimateur du maximum de vraisemblance (EMV)** de θ toute quantité satisfaisant :

$$\hat{\theta}^{MV} \in \operatorname{argmax}_{\theta \in \Theta} L(\theta; X)$$

Souvent on considère la log-vraisemblance.

$$\hat{\theta} \in \operatorname{argmax}_{\theta \in \Theta} l(\theta, X)$$

où $l(\theta, X) = \log L(\theta, X)$

Exemple 2.2. On a $P_\theta = \mathcal{N}(0, 1)$

EMV : $\hat{\theta} = \bar{X}_n$

On dira que l'EMV de θ^2 est $(\bar{X}_n)^2$

Remarque 2.2. L'inconvénient de cette méthode

- Il peut ne pas être unique
- Ne pas exister
- peut être difficile à expliciter