CM2 : Le Cadre Paramétrique et Méthodes d'Estimation

1 Le Cadre Paramétrique

Dans ce chapitre, nous abordons l'estimation statistique dans un contexte paramétrique.

Definition 1.1 (Modèle statistique paramétrique)**.** On dispose d'une observation $(X_1, \ldots, X_n)$ constituant un échantillon de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de loi commune P appartenant à une famille de lois de probabilités paramétrée :

$$\{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$$

Remark 1.2. Si l'espace des paramètres Θ est de dimension infinie, on parle alors de modèle **non-paramétrique**.

Problématique : Estimer la loi P revient à estimer le paramètre inconnu $\theta \in \mathbb{R}^p$.

Example 1.3. Voici quelques exemples classiques de modèles paramétriques :

- Loi de Bernoulli : $\mathrm{Bern}(\theta)$
- Loi Normale : $\mathcal{N}(\mu, \sigma^2)$
- Loi de puissance : $f_\theta(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$
- Loi Exponentielle : $\mathrm{Exp}(\theta)$

Notations :

- $E_\theta[h(X_1, \ldots, X_n)]$ désigne l'espérance sous la loi P_θ.
- La loi du vecteur $(X_1, \ldots, X_n)$ est notée $P_\theta^{\otimes n}$.

Definition 1.4 (Estimateur)**.** Un estimateur de θ est une fonction des observations :

$$\hat{\theta} = \hat{\theta}_n = h(X_1, \ldots, X_n)$$

Qualité d'un estimateur : Pour évaluer la performance d'un estimateur, on utilise principalement deux critères :

1. **Risque quadratique :** $R(\hat{\theta}, \theta) = E\left[(\hat{\theta} - \theta)^2\right]$
2. **Consistance :** L'estimateur est dit consistant si $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta$ (convergence en probabilité).

Definition 1.5 (Modèle identifiable)**.** Un modèle est dit identifiable si l'application $\theta \to P_\theta$ est injective.

2 La Méthode des Moments (MM)

Definition 2.1. Soit $k \geq 1$. On définit :

- Le **moment théorique** d'ordre k de la loi des X_i : $M_k = E[X_i^k]$.
- Le **moment empirique** d'ordre k de la loi des X_i : $\hat{M}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$.

D'après la **Loi des Grands Nombres (LGN)**, on a la convergence suivante :

$$\hat{M}_k \xrightarrow[n\to\infty]{P} M_k$$

Principe de la méthode : Si l'on peut exprimer le paramètre d'intérêt θ (ou une fonction $g(\theta)$) comme une fonction φ des k premiers moments théoriques :

$$\theta = \varphi(M_1, \ldots, M_k)$$

Alors, l'estimateur par la méthode des moments est obtenu en remplaçant les moments théoriques par les moments empiriques :

$$\hat{\theta} = \varphi(\hat{M}_1, \ldots, \hat{M}_k)$$

Example 2.2 (Loi de Bernoulli)**.** Soit $X_i \sim \mathrm{Ber}(\theta)$. On a $\theta = P(X_i = 1) = E[X_i]$. Le moment théorique est $M_1 = E[X_i]$. L'estimateur est donc :

$$E[X_i] \rightsquigarrow \frac{1}{n}\sum_{i=1}^n X_i = \bar{X} \implies \hat{\theta} = \bar{X}$$

Example 2.3 (Loi Exponentielle)**.** Soit $X_i \sim \mathrm{Exp}(\theta)$ de densité $f_\theta(x) = \theta e^{-\theta x}\nVdash_{x\geq 0}$. On sait que $M_1 = E[X_1] = \frac{1}{\theta}$, d'où $\theta = \frac{1}{M_1}$. L'estimateur des moments est :

$$\hat{\theta} = \frac{1}{\hat{M}_1} = \frac{1}{\bar{X}}$$

Example 2.4 (Loi de puissance)**.** Soient $X_1, \ldots, X_n$ i.i.d. de loi P_θ de densité $f_\theta(x) = \theta x^{\theta-1}\nVdash_{x\in[0,1]}$ avec $\theta > 0$. Calculons le premier moment théorique :

$$\begin{aligned} M_1 = E[X] &= \int_0^1 x \cdot \theta x^{\theta-1} dx = \theta \int_0^1 x^\theta dx \\ &= \theta \left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1} \end{aligned}$$

On isole θ :

$$\begin{aligned}(\theta+1)M_1 = \theta &\iff \theta M_1 + M_1 = \theta \\ &\iff \theta(1-M_1) = M_1 \\ &\iff \theta = \frac{M_1}{1-M_1}\end{aligned}$$

L'estimateur des moments est :

$$\hat{\theta}_M = \frac{\bar{X}}{1-\bar{X}}$$

Note : Si $\bar{X} = 1$, l'estimateur n'est pas défini. Cependant, $P_\theta(\bar{X} = 1) = P_\theta(X_1 = X_2 = \cdots = X_n = 1) = 0$.

3 Le Lemme d'Application Continue (LAC)

Une question fondamentale est de savoir si l'estimateur des moments est consistant. Pour cela, on utilise le Lemme d'Application Continue.

Lemma 3.1 (Lemme d'Application Continue - LAC)**.** Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires. Si $X_n \xrightarrow{P} X$ et si g est une fonction continue, alors :

$$g(X_n) \xrightarrow{P} g(X)$$

De même, si $X_n \xrightarrow{\mathcal{L}} X$, alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

Remark 3.2. Une condition suffisante de continuité en présence de points de discontinuité D_g est $P(X \in D_g) = 0$.

Example 3.3. Pour l'exemple précédent $g(x) = \frac{x}{1-x}$. Par la LGN, $\bar{X} \xrightarrow{P} E[X]$. Par le LAC, si g est continue en $E[X]$, alors $g(\bar{X}) \xrightarrow{P} g(E[X]) = \theta$. L'estimateur est donc consistant.

LAC pour des couples : Soient (X_n, Y_n) des suites de variables aléatoires.

- Si $(X_n, Y_n) \xrightarrow{P} (X, Y)$ et si $g : \mathbb{R}^2 \to \mathbb{R}$ est continue, alors $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$.
- Si $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, alors $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, Y)$.

Remark 3.4. La convergence des couples est équivalente à la convergence de chaque composante pour la probabilité :

$$(X_n, Y_n) \xrightarrow{P} (X, Y) \iff X_n \xrightarrow{P} X \text{ et } Y_n \xrightarrow{P} Y$$

Attention : Cette réciproque est fausse pour la convergence en loi !

4 Étude de la Variance Empirique

Si les X_i admettent une espérance μ et une variance σ^2, on appelle **variance empirique** :

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

En développant, on obtient :

$$\begin{aligned}\hat{\sigma}_n^2 &= \frac{1}{n}\sum X_i^2 + \frac{1}{n}\sum \bar{X}^2 - \frac{2}{n}\sum X_i\bar{X} \\ &= \frac{1}{n}\sum X_i^2 + \bar{X}^2 - 2\bar{X}\bar{X} \\ &= \frac{1}{n}\sum X_i^2 - (\bar{X})^2 = \tilde{\sigma}^2\end{aligned}$$

Estimateur des moments : Puisque $\sigma^2 = E[X^2] - (E[X])^2$, on remplace les moments théoriques par les moments empiriques :

$$\hat{\sigma}_{MM}^2 = \frac{1}{n}\sum X_i^2 - (\bar{X})^2$$

Consistance : D'après la LGN :

- $\bar{X} \xrightarrow{P} E[X]$
- $\frac{1}{n}\sum X_i^2 \xrightarrow{P} E[X^2]$

Ainsi, le couple des moments empiriques converge :

$$\begin{pmatrix} \bar{X} \\ \frac{1}{n}\sum X_i^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} E[X] \\ E[X^2] \end{pmatrix}$$

En utilisant le LAC avec $g(x, y) = y - x^2$ (qui est continue de $\mathbb{R}^2 \to \mathbb{R}$), on en déduit que :

$$\hat{\sigma}_n^2 \xrightarrow{P} E[X^2] - (E[X])^2 = \text{Var}(X)$$

L'estimateur $\hat{\sigma}_n^2$ est donc consistant pour la variance σ^2.

Remark 4.1. Exercice : Calculer le biais et le risque de $\hat{\sigma}_n^2$.

5 Méthode du Maximum de Vraisemblance (MV)

Definition 5.1 (Modèle dominé)**.** Un modèle $(P_\theta)_{\theta\in\Theta}$ est dit dominé s'il existe une mesure ν (positive σ-finie) telle que pour tout θ, P_θ admette une densité f_θ par rapport à ν.

En pratique :

- Si l'espace E est au plus dénombrable, ν est la **mesure de comptage**. Si $\exists\{a_1, a_2, \dots\}$ tel que $\sum_{k\geq 1} P_\theta(X_i = a_k) = 1$, alors $\nu = \sum \delta_{a_k}$ avec $\delta_a(\{a\}) = 1$ (mesure de Dirac). On écrira $f_\theta(x) = P_\theta(X_i = x)$.
- Si $E = \mathbb{R}^p$, alors f_θ est la densité usuelle (par rapport à la mesure de Lebesgue).

Definition 5.2 (Vraisemblance)**.** On appelle **vraisemblance** de l'échantillon $(X_1, \dots, X_n)$ (Likelihood) la fonction :

$$\theta \to L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i)$$

C'est une variable aléatoire car elle dépend de l'échantillon.

Definition 5.3 (Estimateur du Maximum de Vraisemblance)**.** Un estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$ est défini tel que :

$$\forall\theta \in \Theta, \quad L_n(\theta) \leq L_n(\hat{\theta}_{MV})$$

On travaille souvent avec la **log-vraisemblance** car elle transforme le produit en somme, ce qui facilite les calculs :

$$\log L_n(\theta) = \sum_{i=1}^{n} \ln f_\theta(X_i)$$

Puisque le logarithme est une fonction croissante, maximiser L_n revient à maximiser $\log L_n$:

$$\log L_n(\hat{\theta}) = \sup_{\theta\in\Theta} \log L_n(\theta)$$

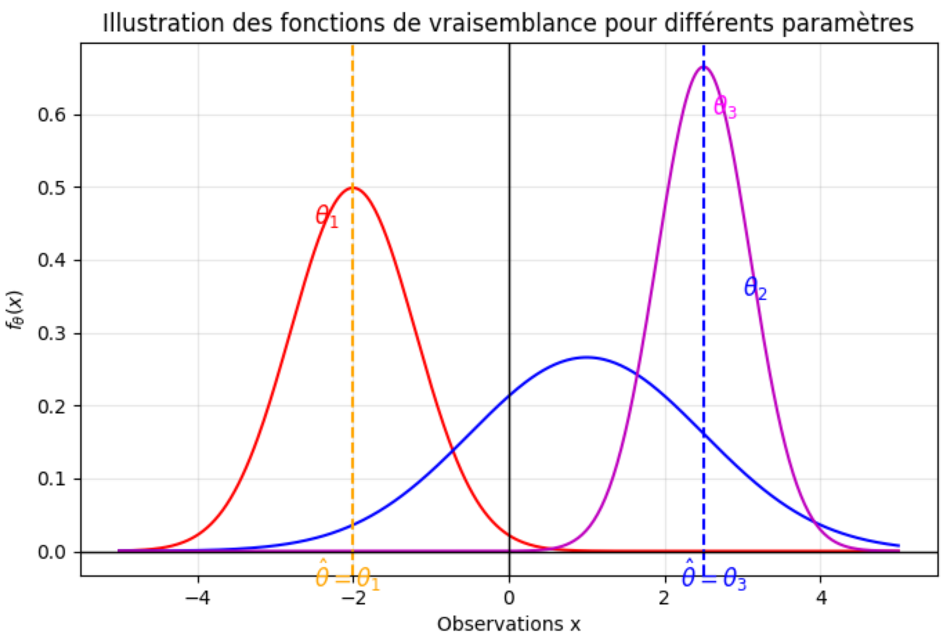

Figure 1: Comparaison des vraisemblances pour différentes valeurs de θ. L'estimateur du maximum de vraisemblance choisit le paramètre qui maximise la probabilité d'observer les données.

6 Mise en œuvre du MV : Exemple de la loi de Bernoulli

Soit $X_i \sim \text{Ber}(\theta)$, la densité est $f_\theta(x) = \theta^x(1-\theta)^{1-x}$ pour $x \in \{0, 1\}$. La vraisemblance est :

$$L_n(\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$$

La log-vraisemblance s'écrit :

$$\log L_n(\theta) = \left(\sum X_i\right)\ln\theta + \left(n - \sum X_i\right)\ln(1-\theta)$$

Condition du premier ordre (Point critique) : On dérive par rapport à θ :

$$(\log L_n)'(\theta) = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta}$$

Recherchons le point critique où la dérivée s'annule :

$$\begin{aligned}
(\log L_n)'(\theta) = 0 &\iff \frac{\sum X_i}{\theta} = \frac{n - \sum X_i}{1-\theta} \\
&\iff (1-\theta)\sum X_i = \theta(n - \sum X_i) \\
&\iff \sum X_i - \theta \sum X_i = n\theta - \theta \sum X_i \\
&\iff \sum X_i = n\theta \\
&\iff \theta = \frac{\sum X_i}{n} = \bar{X}
\end{aligned}$$

On peut aussi réécrire la dérivée sous la forme :

$$(\log L_n)'(\theta) = \frac{\sum X_i - n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}(\bar{X} - \theta)$$

La dérivée change de signe en $\bar{X}$ (de positif à négatif), donc on a bien un maximum en $\hat{\theta}_{MV} = \bar{X}$.

Condition du second ordre : Pour confirmer qu'il s'agit d'un maximum global, on vérifie la concavité de la log-vraisemblance en calculant la dérivée seconde :

$$(\log L_n)''(\theta) = -\frac{\sum X_i}{\theta^2} - \frac{n - \sum X_i}{(1-\theta)^2}$$

Puisque $\sum X_i \geq 0$ et $n - \sum X_i \geq 0$, on a $(\log L_n)''(\theta) < 0$ pour tout $\theta \in]0,1[$. La fonction est donc strictement concave, ce qui garantit que le maximum est unique et global.

Conclusion : Par unicité du point critique et concavité de la fonction, $\hat{\theta} = \bar{X}$ est l'unique Estimateur du Maximum de Vraisemblance (E.M.V.).