1 Introduction

Objectif : Présenter quelques élements de la théorue de l'estimation statistique.

1.1 Problématique

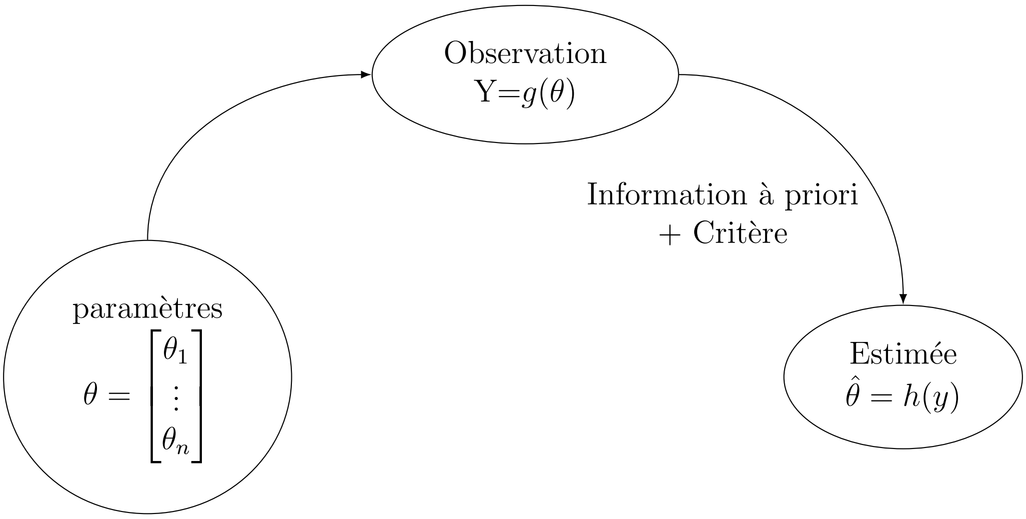

FIGURE 1 – Méthode d'estimation classique

Le raisonnement se transpose alors sur la figure suivante :

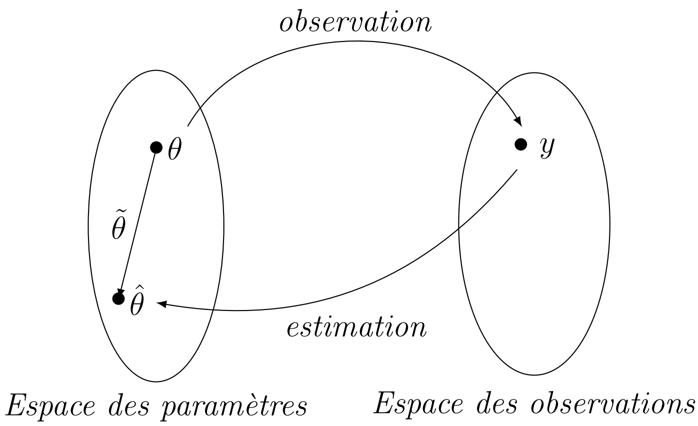

FIGURE 2 – Raisonnement en espace algébrique

On défini les index suivants :

m nombre d'expérience réalisée (taille de y)

n nombre de paramètres (taille de θ)

Estimateurs statistiques On observe une réalisation $y = g(\theta)$ où θ est une VA. et on détermine $\hat{\theta} = h(Y)$ estimée.

Exemple

Exemple 1 Θ tension constante.
$y(t) = \theta + b(t)$. soit $y_i = \theta + b_i$
On défini donc Y et Θ VA et on a $Y = A\Theta + B$ -¿ régression linéaire.

Exemple 2 filtre RC $y(t) = (1 - e^{-t/\tau})u(t) + b(t)$, $\Theta = \tau$. modèle non linéaire, traité en TD.

1.2 Performance-Qualité d'une estimation

Proposition *(Grandeurs utiles)*

- erreur d'estimation
$$\tilde{\theta} = \hat{\theta} - \theta$$
- moment d'ordre 1 :
$$E_{Y|\Theta}[\tilde{\theta}] = E_{Y|\Theta}[\hat{\theta}] - \theta$$
- Biais moyen :
$$E[\tilde{\theta}] = E_{Y\Theta}[\tilde{\theta}] = E[\hat{\theta}] - \theta$$
- moment d'ordre 2 :
  - covariance de l'erreur d'estimation
  $$C_{\tilde{\theta}\tilde{\theta}} = E[(\tilde{\theta} - m_{\tilde{\theta}})(.)^T]$$
  - Corrélation de l'erreur d'estimation
  $$\Gamma_{\tilde{\theta}\tilde{\theta}} = E[\tilde{\theta}\tilde{\theta}^T]$$
  - Puissance :(Estimateur Quadratique moyen)
  $$P_{\tilde{\theta}} = E[\|\tilde{\theta}\|^2] = tr(\Gamma_{\tilde{\theta}\tilde{\theta}})$$

1.3 Caractérisation des estimateurs

Définition

- Borne de Cramer Rao: borne minimale du biais de variance (qui dépend de l'estimateur choisi)
- Estimateur non biaisé: $E[\tilde{\theta}] = 0$
- Estimateur efficace: Borne de Cramer-Rao atteinte.
- Estimateur consistent: $E[\tilde{\theta}] \xrightarrow[N_{obs}\to\infty]{} 0$ et $V[\tilde{\theta}] \xrightarrow[N_{obs}\to\infty]{} 0$
- Estimateur robuste:
  Les performances de l'estimateur ne sont pas trop dégradé si on s'écarte un peu des hypothèses sous laquelle l'estimateur a été établi.
- Complexité de l'estimateur:
  sur l'o btention des connaissances et mise en oeuvre de l'estimateur.

2 Théorie classique de l'estimation

2.1 Estimateur des moindres carrés

Définition

Pour Y une VA de moyenne $m_y = m_{Y|\theta}$ on défini le critère :

$$J_{MC} = (Y - m_y)^T M (Y - m_y)$$

Avec M matrice symétrique définie positive et alors:

$$\hat{\theta}_{MC} = \arg\min_{\theta} J_{MC}(Y, \theta)$$

2.1.1 Condition nécessaire d'existance

Si $J_{MC}(y, \theta)$ est dérivable et pas de contrainte sur θ.

$$\nabla_J(\theta)|_{\hat{\theta}_{MC}} = \frac{\partial J_{MC}}{\partial \theta} = 0 \quad \text{Gradien}$$

Il faut ensuite vérifié que c'est un minimum absolu :

$$\nabla_J^2(\theta) = \frac{\partial^2 J_{MC}}{\partial \theta \partial \theta^T} > 0 \quad \text{Hessien}$$

Application $Y = A\theta + B$, avec B une VA. le critère des moindres carrés est alors :

$$J_{MC} = (Y - A\theta - m_B)^T M (Y - A\theta - m_B)$$

On a une forme quadratique positive car $A^T M A \geq 0$. (dans le cas > 0 on a une CNS sur ce qui suit)

Méthode 1

$$\nabla_J(\theta)|_{\hat{\theta}_{MC}} = 0 = -2A^T M(Y - A\theta - m_B)$$

Donc

$$A^T M A\theta = A^T M(Y - m_B)$$

Soit

$$\boxed{\hat{\theta}_{MC} = \underbrace{(A^T M A)^{-1} A M}_{D}(Y - m_B)}$$

On remarque que $DA = I_n$.

Méthode 2 Pour $A^T M A > 0$.

$$J_{MC} = \underbrace{(D(Y - m_B) - \Theta)^T A^T M A(D(Y - m_B) - \theta)}_{J_1(Y,\theta)} + \underbrace{(Y - m_B)^T (M - D^T A^T M A D)(Y - m_B)}_{J_2(Y)}$$

Alors $\nabla J_{MC} = 0 \implies J_1 = 0 \implies D(Y - m_B) = \hat{\theta}_{MC}$

2.1.2 Caractéristique de l'estimateur

- Estimateur non biaisé

$$\begin{aligned}
\tilde{\theta}_{MC} &= \hat{\Theta} - \theta \\
&= D(Y - m_B) - \theta \\
&= D(B - m_B)
\end{aligned}$$

  Donc $E[\widehat{\theta_{MC}}] = 0$
- moment d'ordre 2 :

$$C_{\tilde{\theta}\tilde{\theta}} = E[(\tilde{\theta} - m_{\tilde{\theta}})(.)^T] = DE[(B - m_B)(B - m_B)^T]D^T = DC_{BB}D^T$$

  - Cas MC ordinaire ($M = I_n$)

$$C_{\tilde{\theta}\tilde{\theta}} = (A^T A)^{-1} A^T C_{BB} A (A^T A)^{-1}$$

  - Cas MC pondéré ($M = C_{BB}^{-1}$)

$$C_{\tilde{\theta}\tilde{\theta}} = (A^T C_{BB}^{-1} A)^{-1}$$

- Cas θ scalaire $Y_i = \theta + B_i$ donc :

$$C_{BB} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix} \text{ et } A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

- Cas MCO : $A^T A = m$

$$\hat{\theta_{MC}} = \frac{\Sigma(y_i - m_{bi})}{m} \quad \text{et} \quad \sigma^2_{\hat{\theta}} = \frac{\Sigma\sigma_i^2}{m^2}$$

- cas MCP pour $M = C_{BB}^{-1} = diag(\sigma_1^{-2}, \ldots, \sigma_m^{-2})$

$$A^T C_{BB} A = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \quad \text{donc} \quad \hat{\theta}_{MCP} = \frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{Y_i - mB_i}{\sigma_i^2}$$

  - $\hat{\theta_{MCP}}$ défini un barycentre
  - Pour $\sigma_i = \sigma$ on a $M = \sigma I \implies MCO = MCP$

- Comparaison MCO et MCP (avec $M = C_{BB}$)

$$\begin{aligned} \sigma^2_{MCO} &\leq \sigma^2_{MCP} \\ \frac{1}{\sum \sigma_i^{-2}} &\leq \frac{1}{m^2} \sum \sigma_i^2 \\ m^2 &\leq \frac{1}{\sum \sigma_i^{-2}} \sum \sigma_i^2 \end{aligned}$$

2.2 Estimateur du maximum de vraisemblance

Définition

On considère $f_Y(y)$ ddp de y paramétrée par θ. On a $f_{Y|\theta}(y) = V(Y, \theta)$. on pose également $L(Y, \theta) = \ln(V(Y, \theta))$.
on défini alors:

$$\hat{\theta}_{MV} = \arg\min f_{Y|\theta}(y) = \arg\min L(Y, \theta)$$

GRAPHE

Exemple Modèle avec bruit additif gaussien.

Proposition

Dans le cas d'un brui Gaussien et pour $M = C_{BB}^{-1}$

$$\hat{\theta}_{MCP} = \hat{\theta}_{MV}$$

Remarque L'estimateur de MV n'est pas nécessairement efficace mais si un estimateur sans biais existe et est efficace c'est celui-ci.

Si $m \to \infty$ on montre que le MV est asymptotiquement efficace. (loi des grands nombres)

3 Théorie générale de l'estimation

3.1 Estimateur linéaire en moyenne quadratique (ELMQ)

Définition

Un ELMQ fourni une estimée de la forme

$$\hat{\theta} = HY + C$$

à partir de l'erreur quadratique moyenne $E[\|\tilde{\theta}\|^2] = E[\tilde{\theta}\tilde{\theta}^T] = P_{\tilde{\theta}}$

Concept H et C tel que $P_{\tilde{\theta}}$ minimal.

$$(1) \quad \frac{\partial P_{\tilde{\theta}}}{\partial H} = 0 \quad \text{et} \quad (2) \quad \frac{\partial P_{\tilde{\theta}}}{\partial C} = 0$$

Proposition

1)
$$\frac{\partial P_{\tilde{\theta}}}{\partial H} = 2E[HY + C - \theta] = 2E[\tilde{\theta}] = 0$$
L'ELMQ est un estimateur non biaisé.

et donc :

$$\begin{aligned} C &= -Hm_Y + m_\theta \\ \hat{\theta} &= H(Y - m_y) + m_\theta \\ \tilde{\theta} &= H(Y - m_y) - (\theta - m_\theta) \end{aligned}$$

Proposition

2)
$$\frac{\partial P_{\tilde{\theta}}}{\partial C} = 2E[(HY + C - \theta)Y^T] = 2E[\tilde{\theta}Y^T] = 0$$
$\tilde{\theta} \perp Y$ quand la puissance est minimale, $\tilde{\theta}$ et Y sont décorrélées, on a extrait toute l'information commune.

De plus :

$$\begin{aligned} E[\tilde{\theta}Y^T] &= E[\tilde{\theta}(Y - m_Y)^T] \\ &= E[(H(Y - m_Y) - \theta - m_\theta)(Y - m_y)^T] \\ &= HC_{yy} - C_{\theta Y} = 0 \implies H = C_{\theta Y}C_{YY}^{-1} \end{aligned}$$

on a donc

$$\boxed{\hat{\theta} = C_{\theta Y}C_{YY}^{-1}(Y - m_Y) + m_\theta}$$

Remarque L'ELMQ nécessite des connaissances du premier et du second ordre sur θ et Y.

Proposition

$$C_{\tilde{\theta}\tilde{\theta}} = C_{\theta\theta} - C_{\theta Y} C_{YY}^{-1} C_{Y\theta}$$

La corrélation entre θ et Y permet de diminuer l'ELMQ.

3.2 Estimateur Bayésiens

3.2.1 Fonction coût/pénalité

Définition

On appelle fonction de coût ou fonction de pénalité une fonction qui mesure l'erreur entrainée par la prise de la valeur $\hat{\theta}$ pour θ.

$$C(\hat{\theta}, \theta) \geq 0 \quad \text{ou encore} \quad C(\tilde{\theta}) \geq 0$$

On prendra le plus souvent une « bonne » fonction (continue, paire , croissante ...)

Exemple de coût

- quadratique
- Valeur absolue
- uniforme

Définition

On appelle estimateur bayésiens l'estimateur qui minimise le coût moyen :

$$\begin{aligned} E_{\theta,Y}[C(\hat{\theta}, \theta)] &= \int_{\mathbb{R}^{m+n}} C(\hat{\theta}, \theta) f_{\theta Y}(\theta, y) d\theta dy \\ &= \int_{\mathbb{R}^m} \left(\underbrace{\int_{\mathbb{R}^n} C(\hat{\theta}, \theta) f_{\theta|Y=y}(\theta) d\theta}_{E_{\theta|Y}[C(\hat{\theta},\theta)]} \right) f_Y(y) dy \end{aligned}$$

On minimise donc $E_{\theta|Y}[C(\hat{\theta}, \theta)]$ à coût conditionnel donné

$$\hat{\theta}_B = \arg\min_{\hat{\theta}} E_{\theta|Y}[C(\hat{\theta}, \theta)]$$

3.2.2 Estimateur du maximum a posteriori (MAP)

On considère un cout uniforme.

Définition

En prenant:

$$E_{\theta|Y}[C(\hat{\theta},\theta)] = \int_{\mathbb{R}^m} (1-\Pi_\Delta(\tilde{\theta}))f_{\theta|Y=y}(\theta)d\theta \;\; = 1 - \int_{\hat{\theta}-\Delta/2}^{\hat{\theta}+\Delta/2} f_{\theta|Y=y}(\theta)d\theta \simeq 1 - \Delta^n f_{\theta|Y=y}(\hat{\theta})$$

Soit

$$\hat{\theta}_{MAP} = \arg\max_\theta f_{\theta|Y=y}(\theta)$$

Lien MAP-MV on a $f_{\theta|Y=y}(\theta)f_Y(y) = f_{\theta Y}(\theta,y)$. Avec $f_\theta(\theta) = C^{ste}$ quand $f_{\theta Y}(\theta,y)$ à une valeur significative (ie $C_{\theta\theta}$ grand / σ_θ grand) alors :

$$\arg\max f_{\theta|Y=y}(\theta) \simeq \arg\max f_{Y|\Theta=\theta}(y)$$

On considère alors que θ est un paramètre aléatoire mais très mal connu. (ddp uniforme sur un interval tres grand, peu d'infos sur θ).

cf. TD « file d'attente »

Exemple et Application On considère θ scalaire aléatoire avec : $Y_i = \theta + B_i$ Avec :

$$\begin{cases} B \hookrightarrow \mathcal{N}(0, C_{BB}) \\ \Theta \hookrightarrow \mathcal{N}(m_\theta, \sigma_\theta^2) \\ B \perp \Theta \end{cases}$$

Rappel MC=MV avec : $\begin{cases} m_B = 0 \\ \hat{\theta}_{MV} = \hat{\theta}_{MC} = \frac{\sum_{i=1}^m Y_i}{m} \\ E[\hat{\theta}_{MV}] = E[\theta] = m_\theta \text{ et } \sigma_{\tilde{\theta}_{MV}} = \frac{\sigma_B}{m} \end{cases}$

On a donc :

$$f_{Y|\theta}(y) = f_B(Y - A\theta) = \prod_{i=1}^m f_{B_i}(Y_i - \theta) = C_1 \exp\left(-\frac{1}{2}\frac{\sum(Y_i-\theta)^2}{\sigma_B^2}\right)$$

Or

$$f_{\theta|Y=y}(\theta) = \frac{f_{Y|\theta}(y)f_\theta(\theta)}{f_Y(y)} = C_2 \exp\left(-\frac{1}{2}\underbrace{\left[\frac{\sum(Y_i-\theta)^2}{\sigma_B^2} + \frac{(\theta-m_\theta)^2}{\sigma_\theta^2}\right]}_{J_{MAP}}\right)$$

Le critère est ici une forme quadratique, donc :

$$\hat{\theta}_{MAP} = \arg\max f_{\theta|Y=y}(\theta) = \arg\min J_{MAP}(\theta, Y)$$

Alors on a la CNS :

$$\frac{\mathrm{d}J_{MAP}}{\mathrm{d}\theta} = 0 = 2\left[-\sum_{i=1}^{m}\frac{Y_i-\theta}{\sigma_b^2} + \frac{(\theta-m_\theta)^2}{\sigma_\theta^2}\right]$$

Soit une expression barycentrique :

$$\hat{\theta}_{MAP} = \frac{\frac{m}{\sigma_B^2}\sum\frac{Y_i}{m} + \frac{m_\theta}{\sigma_\theta^2}}{\frac{m}{\sigma_B^2}+\frac{1}{\sigma_\theta^2}}$$

Donc :

Proposition

$$E[\hat{\theta}_{MAP}] = m_\theta$$

L'estimateur est non biaisé. De plus :

$$\sigma^2_{\tilde{\theta}_{MAP}} = \frac{1}{\frac{1}{\sigma_{MV}}+\frac{1}{\sigma_\theta^2}} < \begin{cases}\sigma_\theta^2\\ \sigma_{MV}^2\end{cases}$$

On a fait mieux en prenant en compte toutes les sources d'informations.

Remarque

- Si $\sigma_\theta >> \sigma_{MV}$ alors $\hat{\theta}_{MAP} \simeq \hat{\theta}_{MV}$ (ce qui arrive pour σ_B ou m grand)
- Si $\sigma_\theta << \sigma_{MV}$ et $\hat{\theta}_{MAP} \simeq m_\theta$ (l'obersavation apporte peu d'info)

3.2.3 Estimateur en moyenne quadratique (EQM)

Définition

On le cout moyen de l'EQM:

$$C(\hat{\theta},\theta) = (\hat{\theta}-\theta)^T M(\hat{\theta}-\theta)$$

Avec $M > 0$. On cherche a minimiser le cout moyen mais sans contrainte de linéarité avec une matrice de pondération qui peux prendre en compte des facteurs d'echelles ou des unités différentes.

Etude de l'estimateur On veut minimiser $E_{\Theta|Y}[C(\hat{\theta},\theta)]$

$$\begin{aligned}
\nabla_{\hat{\theta}} E_{\Theta|Y}[C(\hat{\theta},\theta)] &= 0 \\
E_{\theta|Y}[2M(\hat{\theta}-\theta)] &= 0 \\
2ME_{\theta|Y}[\underbrace{\hat{\theta}}_{h(y)}] - E_{\theta|Y}[\theta] &= 0 \\
2M(\hat{\theta} - E_{\theta|Y}[\theta]) &= 0 \\
&\boxed{\hat{\theta}_{MQ} = E_{\theta|Y}[\theta]} \\
&= \int_{\mathbb{R}^n} \theta f_{\theta|y}(\theta) d\theta = h(Y=y)
\end{aligned}$$

Par conséquent : $E[\hat{\theta}_{MQ}] = E[\theta]$. on a un estimateur non biaisé.

Remarque Si $f_{\theta|Y}$ possède un axe de symétrie (ex : gaussienne) :

FIGURE . ($\hat{\theta}_{MQ} = \hat{\theta}_{MAP}$ dans le cas gaussien. Différent avec deux bosses.)

Dans le cas général la contrainte de linéarité pour l'ELMQ conduit à une valeur plus grande qu'avec l'EQM. Dans le cas gaussien : $\hat{\theta}_{ELMQ} = \hat{\theta}_{MQ}$, mais $\hat{\theta}_{MQ}$ nécessite plus de connaissance (ddp).

3.2.4 Estimateur en valeur absolu

Définition

on s'interesse au cas $n = 1$ (un paramètre) On choisit le cout moyen :

$$C(\hat{\theta},\theta) = |\hat{\theta} - \theta|$$

Alors :

$$E_{\theta|Y}[C(\hat{\theta},\theta)] = \int_{-\infty}^{\hat{\theta}} (\hat{\theta}-\theta) f_{\theta|Y=y}(\theta) d\theta - \int_{\hat{\theta}}^{+\infty} (\hat{\theta}-\theta) f_{\theta|Y=y}(\theta) d\theta$$

Donc :

$$\begin{aligned}
0 =& \nabla_{\hat{\theta}} E_{\theta|Y}[C(\hat{\theta},\theta)] \\
=& \dots \\
=& \int_{-\infty}^{\hat{\theta}} f_{\theta|Y=y}(\theta) d\theta - \int_{\hat{\theta}}^{+\infty} f_{\theta|Y=y}(\theta) d\theta
\end{aligned}$$

Proposition

L'estimée est alors $\hat{\theta}_{VA}$ tel que :

$$\int_{-\infty}^{\hat{\theta}} f_{\theta|Y=y}(\theta)d\theta = \int_{\hat{\theta}}^{+\infty} f_{\theta|Y=y}(\theta)d\theta$$

On parle de médiane a posteriori. Le résultat se généralise pour tout n.

Remarque Dans le cas où $f_{\theta|Y=y}(\theta)$ possède un axe de symétrie (ex gaussienne) on a :

$$\hat{\theta}_{VA} = \hat{\theta}_{MV} \overset{\overset{\max}{\downarrow}}{=} \hat{\theta}_{MAP}$$

Exemple Localisation d'un véhicule / Ellipsoïde de confiance (cf poly).

4 Conclusion

- L'estimateur statistique dépend des connaissances a priori, de la complexité des calculs et de la robustesse attendue.
- Dans certains cas particuliers/ limites on retrouve des estimateurs intuitifs /empirique.
- La loi normale joue un rôle important (hypothèses qui se justifie par la loi des grands nombres) : les calculs sont simplifiés et conduisent au même résultat.