M1-HAX710X **Année : 2024-2025**

Estimation Ponctuelle

Ludovic Menneteau

1 Généralités

Problématique 1 *On considère une v.a.* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{X}, \mathcal{B})$ *de loi* $\mathbb{P}_X$ *inconnue. On suppose que* $\mathbb{P}_X$ *appartient à une famille de loi* $(P_\theta)_{\theta\in\Theta}$ *indexée par un ensemble* Θ *et entièrement spécifiée, i.e. qu'il existe* $\theta_v \in \Theta$ *(*$\theta_v : \theta$ *vrai) tel que* $\mathbb{P}_X = P_{\theta_v}$*. Notre objectif est d'étudier des procédures permettant à partir de* X*, d'obtenir de l'information sur* $\mathbb{P}_X$ *en estimant* θ_v *(ou une fonction de* θ_v*). Comme* θ_v *peut prendre n'importe quelle valeur dans* Θ*, les procédures d'estimations devront être évaluées uniformément en* θ*.*

Définition 2 *On dit que* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *est un* ***modèle statistique*** *si* $(\mathcal{X}, \mathcal{B})$ *est un espace mesurable (i.e.* $\mathcal{X}$ *est un ensemble,* $\mathcal{B}$ *est une tribu sur* $\mathcal{X}$*) et* $(P_\theta)_{\theta\in\Theta}$ *est une famille de probabilités sur* $(\mathcal{X}, \mathcal{B})$ *indéxée par un ensemble* Θ *(appelé* ***ensemble des paramètres*** *du modèle).*
Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{X}, \mathcal{B})$ *est une v.a. telle que* $\mathbb{P}_X$ *appartient à* $(P_\theta)_{\theta\in\Theta}$*, on dira que* X *est* ***associée*** *au modèle statistique* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$*.*

Définition 3 *Soient* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique et* $(\mathcal{Y}, \mathcal{D})$*, un espace mesurable.*

1. *On appelle* ***statistique*** *à valeurs dans* $(\mathcal{Y}, \mathcal{D})$ *toute fonction mesurable* $s : \mathcal{X} \to \mathcal{Y}$ *ne dépendant pas de* θ*.*
2. *Soit* $\varphi : \Theta \to \varphi(\Theta)$ *une fonction telle que* $\varphi(\Theta) \subset \mathcal{Y}$*. Une statistique* $\widehat{\varphi} : \mathcal{X} \to \mathcal{Y}$ *est appelée un* ***estimateur*** *de* $\varphi(\theta)$*.*

Notations 4 *Soit* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique.*

1. *Pour* $p \geq 1$*, on dit que* $h \in \mathcal{L}^p\left((P_\theta)_{\theta\in\Theta}\right)$ *si* $h \in \mathcal{L}^p(P_\theta)$ *pour tout* $\theta \in \Theta$*.*
2. *Soient* $h : \mathcal{X} \to \mathbb{R}$ *et* $\theta \in \Theta$*.*
   (a) *Si* $h \in \mathcal{L}^1(P_\theta)$*, on définit :*
   $$E_\theta(h) := \int_{\mathcal{X}} h \, dP_\theta.$$
   (b) *Si* $h \in \mathcal{L}^2(P_\theta)$*, on définit :*
   $$V_\theta(h) := E_\theta\left((h - E_\theta(h))^2\right).$$
   (c) *Si* g *et* $h \in \mathcal{L}^2(P_\theta)$*, on définit :*
   $$\mathrm{Cov}_\theta(g, h) := E_\theta\left((g - E_\theta(g))(h - E_\theta(h))\right).$$
3. *Soient* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{X}, \mathcal{B})$ *une v.a. associée au modèle statistique* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *et* $\theta \in \Theta$*.*

(a) *Si* $h \in \mathcal{L}^1(P_\theta)$, *et si on suppose que* $\theta_v = \theta$, *alors l'espérance de* $h(X) : \mathbb{E}(h(X)) := \int_\Omega h(X) \ d\mathbb{P}$ *sera notée* $\mathbb{E}_\theta(h(X))$.

(b) *Si* $h \in \mathcal{L}^2(P_\theta)$, *et si on suppose que* $\theta_v = \theta$, *alors la variance de* $h(X) :\ \mathbb{V}(h(X)) := \mathbb{E}\left((h(X)) - \mathbb{E}(h(X))^2\right)$ *sera notée* $\mathbb{V}_\theta(h(X))$.

(c) *Si* g *et* $h \in \mathcal{L}^2(P_\theta)$, *et si on suppose que* $\theta_v = \theta$, *alors la covariance de* $g(X)$ *et* $h(X)$:

$$\mathbb{C}ov(g(X), h(X)) := \mathbb{E}((g(X)) - \mathbb{E}(g(X))\ \ (h(X)) - \mathbb{E}(h(X)))$$

sera notée $\mathbb{C}ov_\theta(g(X), h(X))$.

4. *NB : Par le Théorème de transfert, il est clair que :*

$$\mathbb{E}_\theta(h(X)) = E_\theta(h),\ \mathbb{V}_\theta(h(X)) = V_\theta(h) \text{ et } \mathbb{C}ov_\theta(g(X), h(X)) = \mathrm{Cov}_\theta(g, h).$$

Définition 5 *Soit* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique.*
Pour $n \geq 1$, *le modèle produit*

$$\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)^n := \left(\mathcal{X}^n, \mathcal{B}^{\otimes n}, (P_\theta^{\otimes n})_{\theta\in\Theta}\right)$$

est appelé ***modèle d'échantillonage*** *(d'ordre n) associé à* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$.

Rappels 6 *Soient* $(\mathcal{X}, \mathcal{B})$ *est un espace mesurable et* ν *et* μ *deux mesures* σ*-finies sur* $(\mathcal{X}, \mathcal{B})$.

1. *On dit que* ν *est* ***absolument continue*** *par rapport à* μ *et on note* $\nu \ll \mu$ *si l'une des deux conditions équivalentes (par le Théorème de Radon Nycodym) suivantes est vérifiée :*
   i) $[N \in \mathcal{B} \text{ et } \mu(N) = 0] \Rightarrow [\nu(N) = 0]$.
   ii) Il existe $f : \mathcal{X} \to \mathbb{R}^+$ *mesurable telle que* $\nu = f.\mu$, *i.e. pour tout* $B \in \mathcal{B}$:

$$\nu(B) = \int_{\mathcal{X}} \mathbf{1}_B(x)\ \ f(x)\ \ d\mu(x).$$

   On dit que f *est une* ***densité*** *de* ν *par rapport à* μ.

2. *En général,* f *n'est pas unique puisque toutes autre fonction mesurable positive* $\mu - pp$ *égale à* f *convient aussi. On note parfois* $\frac{d\nu}{d\mu}$ *la classe d'équivalence des densités de* ν *par rapport à* μ, *on écrira* $f \in \frac{d\nu}{d\mu}$ *pour dire que* f *est une densité de* ν *par rapport à* μ.

3. *Si* $\nu \ll \mu$ *et* $\mu \ll \nu$ *les mesures sont dites* ***équivalentes*** *et on note* $\mu \sim \nu$. *Il est facile de voir que cela revient à dire que* $\frac{d\nu}{d\mu}$ *contient une version* f *strictement positive (et qu'alors* $\frac{1}{f}$ *est une version de* $\frac{d\mu}{d\nu}$*).*

Définition 7 *Soit* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique.*

1. *On dit que le modèle est **dominé** s'il existe une mesure σ-finie μ sur $(\mathcal{X}, \mathcal{B})$ telle que , $\forall\theta \in \Theta : P_\theta \ll \mu$.*
   *μ est appelée une mesure **dominante** du modèle.*
   *Pour $\theta \in \Theta$, on note f_θ une version de la densité $\frac{dP_\theta}{d\mu}$ et on appelle **fonction de vraisemblance** du modèle par rapport à la mesure dominante μ la fonction :*
   $$f : (x, \theta) \in \mathcal{X} \times \Theta \mapsto f_\theta(x) \in \mathbb{R}^+.$$
2. *On dit que le modèle est **homogène** si toutes les mesures P_θ sont équivalentes.*

Remarques 8 *1) Comme les f_θ, f dépend non seulement du modèle mais aussi de μ. De plus, comme les densités, la fonction de vraisemblance n'est pas unique en tant que fonction puisque si g est une fonction définie sur $\mathcal{X} \times \Theta$ telle que pour tout $\theta \in \Theta,\ g(.,\theta) = f(.,\theta) \quad \mu-p.p.$ (en x), alors g est aussi une fonction de vraisemblance du modèle par rapport à μ (f et g sont deux versions de la même vraisemblance). Dans la suite, lorsque l'on écrit soit f la vraisemblance (resp. la densité), il faut comprendre "une version de la vraisemblance (resp. la densité)".*
2) Sans perte de généralité, on peut supposer (si besoin) que μ est une probabilité. En effet, comme μ est σ-finie, on peut trouver $\{A_k : k \in K\}$ une partition mesurable de $\mathcal{X}$ telle que K est dénombrable et $\forall k \in K,\ \mu(A_k) \in]0, \infty[$.
Il est facile de voir que, si $(p_k)_{k\in K}$ est une famille de réels strictement positifs telle que $\sum_{k\in K} p_k = 1$, la mesure

$$\nu : B \in \mathcal{B} \mapsto \sum_{k\in K} p_k \frac{\mu(A_k \cap B)}{\mu(A_k)}$$

est une probabilité équivalente à μ et peut donc être choisie comme une mesure dominante du modèle.

Proposition 9 *Soit $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ un modèle statistique dominé par une mesure μ. Alors, pour tout $n \geq 1$, le modèle statistique d'échantillonage $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)^n$ est dominé par $\mu^{\otimes n}$. De plus si, pour $\theta \in \Theta,\ P_\theta = f_\theta \cdot \mu$ alors $P_\theta^{\otimes n} = f_{\theta,n} \cdot \mu^{\otimes n}$ avec*

$$f_{\theta,n} : (x_1, ..., x_n) \in \mathcal{X}^n \mapsto \prod_{1\leq j\leq n} f_\theta(x_j).$$

Preuve. *On suppose que $P_\theta = f_\theta.\mu$.*

Pour $(B_1, ..., B_n) \in \mathcal{B}^n$,

$$
\begin{aligned}
P_\theta^{\otimes n}\left(\prod_{1\le j\le n} B_j\right) &= \prod_{1\le j\le n} P_\theta(B_j) \\
&= \prod_{1\le j\le n} \int_{B_j} f_\theta(x_j)\ d\mu(x_j) \\
&= \int_{\prod_{1\le j\le n} B_j} \underbrace{\left(\prod_{1\le j\le n} f_\theta(x_j)\right)}_{:=f_{\theta,n}(x_1,\dots,x_n)} d\mu^{\otimes n}((x_1, ..., x_n))
\end{aligned}
$$

Les mesures finies $P_\theta^{\otimes n}$ *et* $f_{\theta,n}.\mu^{\otimes n}$ *coïncidant sur le* π *système* $\left\{\prod_{1\le j\le n} B_j : (B_1, ..., B_n) \in \mathcal{B}^n\right\}$, *elles coïncident aussi sur* $\mathcal{B}^{\otimes n} = \sigma\left\{\prod_{1\le j\le n} B_j : (B_1, ..., B_n) \in \mathcal{B}^n\right\}$. ■

Le résultat suivant montre que, si $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ est un modèle dominé, il existe des mesures dominantes particulières pouvant s'exprimer comme des combinaisons convexes dénombrables des éléments de $(P_\theta)_{\theta\in\Theta}$, ce qui aura un intérêt pour établir certains résultats (par exemple le Théorème 21).

Théorème 10 *Soit* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique dominé. Alors :*

1. *Il existe un sous ensemble dénombrable* $T \subset \Theta$ *tel que :*

$$[B \in \mathcal{B} \text{ et } \forall t \in T : P_t(B) = 0] \Longrightarrow [\forall \theta \in \Theta : P_\theta(B) = 0].$$

2. *On pose*

$$P^* := \sum_{t\in T} c(t) P_t$$

où $(c(t))_{t\in T}$ *est une famille de nombres strictement positifs telle que* $\sum_{t\in T} c(t) = 1$. *Alors, pour tout* $\theta \in \Theta$ *et pour toute mesure* μ *dominant le modèle, on a :*

$$P_\theta \ll P^* \ll \mu,$$

(i.e. P^* *est une probabilité qui domine le modèle et est dominée par* μ*).*
P^* *est appelée une* ***mesure dominante privilégiée*** *du modèle.*
(NB : P^* *est un suprémum de* $(P_\theta)_{\theta\in\Theta}$ *pour le semi ordre partiel induit par* $\ll$ *sur l'espace des mesures* σ*-finies* μ *sur* $(\mathcal{X}, \mathcal{B})$*).*

Preuve. 1) a) Construction de l'ensemble T :
Soit μ_0 une mesure dominante du modèle. Sans perte de généralités (c.f. la Remarque 8-2) , on peut supposer que μ_0 est une probabilité (une mesure bornée suffirait pour l'argument que l'on va développer). Pour tout $\theta \in \Theta$, on considère f_θ une densité de P_θ par rapport à μ_0. Soit

$$\mathcal{C}_0 = \{\{f_\theta > 0\} : \theta \in \Theta\}$$

et
$$\mathcal{C} = \left\{ \text{unions dénombrables d'éléments de } \mathcal{C}_0 \right\}.$$

On a $\mathcal{C}_0 \subset \mathcal{C} \subset \mathcal{B}$. Montrons que $m := \sup_{C \in \mathcal{C}} \mu_0(C)$ est atteint.
Comme μ_0 est bornée, m est fini. De plus, par définition du sup, il existe, pour tout $n \geq 1$, $C_n \in \mathcal{C}$ tel que :

$$m - \frac{1}{n} \leq \mu_0(C_n) \leq m.$$

On pose alors $C^* := \bigcup_{n \geq 1} C_n$. Une union dénombrable d'unions dénombrables étant dénombrable, on a $C^* \in \mathcal{C}$ et donc $\mu_0(C^*) \leq m$. Par ailleurs, pour tout $n \geq 1$, $C_n \subset C^*$ donc

$$\forall n \geq 1 : m - \frac{1}{n} \leq \mu_0(C^*) \leq m \ \text{ i.e. } \mu_0(C^*) = m.$$

Comme $C^* \in \mathcal{C}$, il existe $T \subset \Theta$ tel que T est dénombrable et $C^* = \bigcup_{t \in T} \{f_t > 0\}$.
b) Montrons que T convient :
Soit $B \in \mathcal{B}$, tel que $\forall t \in T : P_t(B) = 0$. Montrons que $\forall \theta \in \Theta : P_\theta(B) = 0$.
Pour $\theta \in \Theta$, on a :

$$\begin{aligned} P_\theta(B) &= P_\theta(B \cap C^*) + P_\theta(B \cap (C^*)^c) \\ &\leq \left(\sum_{t \in T} P_\theta(B \cap \{f_t > 0\}) \right) + P_\theta((C^*)^c). \end{aligned}$$

Il suffit donc de montrer que $P_\theta((C^*)^c) = 0$ et que,

$$\forall t \in T : P_\theta(B \cap \{f_t > 0\}) = 0.$$

-) Pour montrer que $P_\theta((C^*)^c) = 0$, observons que

$$\begin{aligned} P_\theta((C^*)^c) &= \int_{\mathcal{X}} \mathbf{1}_{(C^*)^c} \ f_\theta \ d\mu_0 = \int_{\{f_\theta > 0\}} \mathbf{1}_{(C^*)^c} \ f_\theta \ d\mu_0 \\ &= P_\theta((C^*)^c \cap \{f_\theta > 0\}). \end{aligned}$$

Donc, comme $P_\theta \ll \mu_0$, il suffit de montrer que

$$\mu_0((C^*)^c \cap \{f_\theta > 0\}) = 0.$$

Or, (comme $C^* \cup \{f_\theta > 0\} = C^* \uplus ((C^*)^c \cap \{f_\theta > 0\})$),

$$\mu_0(C^* \cup \{f_\theta > 0\}) = \mu_0(C^*) + \mu_0((C^*)^c \cap \{f_\theta > 0\})$$

de ce fait comme μ_0 est bornée :

$$\mu_0((C^*)^c \cap \{f_\theta > 0\}) = \mu_0(C^* \cup \{f_\theta > 0\}) - \mu_0(C^*).$$

De plus, $C^* \cup \{f_\theta > 0\} \in \mathcal{C}$, d'où par définition de C^* :

$m = \mu_0 (C^*) \leq \mu_0 (C^* \cup \{f_\theta > 0\}) \leq m$ i.e. $\mu_0 (C^* \cup \{f_\theta > 0\}) = \mu_0 (C^*) = m$,

et donc :

$$\mu_0 ((C^*)^c \cap \{f_\theta > 0\}) = m - m = 0.$$

-) Montrons que, $\forall t \in T : P_\theta (B \cap \{f_t > 0\}) = 0$.

$$\begin{aligned}
[P_t (B) = 0] \Leftrightarrow \left[\int_{\mathcal{X}} \mathbf{1}_B \ f_t \ d\mu_0 = 0\right] &\Leftrightarrow [\mathbf{1}_B \ f_t = 0 \ \mu_0 \text{ p.p.}] \\
&\Leftrightarrow [\mu_0 (\{\mathbf{1}_B \ f_t \neq 0\}) = 0] \Leftrightarrow [\mu_0 (B \cap \{f_t > 0\}) = 0] \\
&\Longrightarrow [P_\theta (B \cap \{f_t > 0\}) = 0] \qquad (\text{car } P_\theta \ll \mu_0).
\end{aligned}$$

2) Comme $(c(t))_{t \in T}$ est une famille de nombres strictement positifs telle que $\sum_{t \in T} c(t) = 1$, $P^* = \sum_{t \in T} c(t) P_t$ est clairement une probabilité sur $(\mathcal{X}, \mathcal{B})$. De plus :

$$[P^* (B) = 0] \Leftrightarrow [\forall t \in T : P_t (B) = 0] \underset{\text{par 1)}}{\Longrightarrow} [\forall \theta \in \Theta : P_\theta (B) = 0],$$

donc, $\forall \theta \in \Theta : P_\theta \ll P^*$. Enfin, soit μ une mesure dominante du modèle. Comme μ domine $(P_\theta)_{\theta \in \Theta}$:

$$[\mu (B) = 0] \Longrightarrow [\forall t \in T : P_t (B) = 0] \Leftrightarrow [P^* (B) = 0]$$

donc $P^* \ll \mu$. ■

2 Estimation d'une fonction réelle du paramètre

2.1 Biais. Risque quadratique. Comparaison d'estimateurs.

Définition 11 *Soient $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta \in \Theta}\right)$ un modèle statistique et $\varphi : \Theta \to \varphi(\Theta)$ une fonction.telle que $\varphi(\Theta) \subset \mathbb{R}$.*

1. *Si $\widehat{\varphi}$ est un estimateur de $\varphi(\theta) \in \mathbb{R}$ tel que $\widehat{\varphi} \in \mathcal{L}^1\left((P_\theta)_{\theta \in \Theta}\right)$, alors on définit le* ***biais*** *de $\widehat{\varphi}$ par rapport à $\varphi(\theta)$:*

$$B_\theta^\varphi (\widehat{\varphi}) := E_\theta (\widehat{\varphi}) - \varphi(\theta).$$

   Si $\forall \theta \in \Theta$, $B_\theta^\varphi (\widehat{\varphi}) = 0$, on dit que $\widehat{\varphi}$ est un estimateur ***sans biais*** *de $(\varphi(\theta))_{\theta \in \Theta}$.*

2. *Si $\widehat{\varphi} \in \mathcal{L}^2\left((P_\theta)_{\theta \in \Theta}\right)$, on définit le* ***risque (ou écart) quadratique moyen*** *entre $\widehat{\varphi}$ et $\varphi(\theta)$ par*

$$R_\theta^\varphi (\widehat{\varphi}) := E_\theta \left((\widehat{\varphi} - \varphi(\theta))^2\right).$$

3. *Si* $\widehat{\varphi}$ *et* $\widetilde{\varphi} \in \mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right)$, *on dit que* $\widehat{\varphi}$ *est (uniformément) préférable à* $\widetilde{\varphi}$ *si* $\forall\theta \in \Theta$, $R_\theta^\varphi(\widehat{\varphi}) \leq R_\theta^\varphi(\widetilde{\varphi})$.
   Cette relation de préférence introduit une relation de semi ordre partiel sur $\mathcal{E}_\varphi$.

4. *Si* $\mathcal{F}_\varphi \subset \mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right)$, *on dit qu'un estimateur* $\widehat{\varphi}$ *est* ***optimal*** *sur* $\mathcal{F}_\varphi$ *si* $\widehat{\varphi} \in \mathcal{F}_\varphi$ *et s'il est préférable à tout autre estimateur appartenant à* $\mathcal{F}_\varphi$ *(en général, l'existence d'un tel estimateur n'est pas garantie).*

5. *On note*

$$\mathcal{E}_\varphi^0 = \left\{\widehat{\varphi} \in \mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right) : \forall\theta \in \Theta,\ E_\theta(\widehat{\varphi}) = \varphi(\theta)\right\}$$

   la classe des ***estimateurs sans biais*** *de* $(\varphi(\theta))_{\theta\in\Theta}$ *dans* $\mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right)$.

Proposition 12 *(Décomposition biais-variance du risque quadratique)*
Considérons un modèle statistique $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *avec* $\Theta \subset \mathbb{R}$, *une fonction* $\varphi : \Theta \to \varphi(\Theta) \subset \mathbb{R}$ *et* $\widehat{\varphi} \in \mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right)$. *Alors :*

1. *Pour tout* $\theta \in \Theta$:
$$R_\theta^\varphi(\widehat{\varphi}) = B_\theta^\varphi(\widehat{\varphi})^2 + V_\theta(\widehat{\varphi}).$$

2. *Si* $\widehat{\varphi} \in \mathcal{E}_\varphi^0$ *alors* $R_\theta^\varphi(\widehat{\varphi}) = V_\theta(\widehat{\varphi})$.
   Un estimateur optimal dans $\mathcal{E}_\varphi^0$ *(s'il existe) est donc un estimateur de* $(\varphi(\theta))_{\theta\in\Theta}$ *sans biais et de variance minimale (**esbvm**).*

Preuve. Pour tout $\theta \in \Theta$:

$$E_\theta\left((\widehat{\varphi} - \varphi(\theta))^2\right)$$
$$= E_\theta\left(((\widehat{\varphi} - E_\theta(\widehat{\varphi})) + (E_\theta(\widehat{\varphi}) - \varphi(\theta)))^2\right)$$
$$= \underbrace{E_\theta\left((\widehat{\varphi} - E_\theta(\widehat{\varphi}))^2\right)}_{=V_\theta(\widehat{\varphi})} + 2\ (E_\theta(\widehat{\varphi}) - \varphi(\theta))\ \underbrace{E_\theta((\widehat{\varphi} - E_\theta(\widehat{\varphi})))}_{=0} + \underbrace{E_\theta\left((E_\theta(\widehat{\varphi}) - \varphi(\theta))^2\right)}_{=B_\theta^\varphi(\widehat{\varphi})^2}.$$

■

En général, il n'existe pas forcément d'esbvm de $(\varphi(\theta))_{\theta\in\Theta}$ (ni même d'esb, voir par exemple en cours du modèle binomial $(B(n,\theta))_{\theta\in\Theta}$).
Le résultat suivant fournit une cns d'existence d'esbvm et montre que, s'il existe un esbvm, il est essentiellement unique.

Proposition 13 *Soient* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique ,* $\varphi : \Theta \to \varphi(\Theta) \subset \mathbb{R}$ *et* $\widehat{\varphi} \in \mathcal{E}_\varphi^0$. *Alors :*

1. *Les deux assertions suivantes sont équivalentes :*
   i) $\widehat{\varphi}$ *est un esbvm de* $(\varphi(\theta))_{\theta\in\Theta}$.
   ii) $\forall\widetilde{\varphi} \in \mathcal{E}_\varphi^0$ *et* $\forall\theta \in \Theta : \mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) = 0$.

2. Si $\widehat{\varphi}$ et $\widetilde{\varphi}$ sont deux esbvm de $(\varphi(\theta))_{\theta\in\Theta}$, alors

$$\forall\theta\in\Theta : \widetilde{\varphi} = \widehat{\varphi} \quad P_\theta\text{−p.p.}$$

Preuve. 1) $(i) \Rightarrow (ii)$: On suppose que $\widehat{\varphi}$ est un esbvm de $(\varphi(\theta))_{\theta\in\Theta}$.
Soit $\widetilde{\varphi} \in \mathcal{E}^0_\varphi$. Pour $\alpha \in \mathbb{R}$, on pose

$$\widetilde{\varphi}_\alpha := (1-\alpha)\ \widehat{\varphi} + \alpha\ \widetilde{\varphi} = \widehat{\varphi} + \alpha(\widetilde{\varphi} - \widehat{\varphi}).$$

Pour tout $\theta \in \Theta$, $\widetilde{\varphi}_\alpha \in \mathcal{L}^2(P_\theta)$ et

$$E_\theta(\widetilde{\varphi}_\alpha) = (1-\alpha)E_\theta(\widehat{\varphi}) + \alpha E_\theta(\widetilde{\varphi}) = (1-\alpha)\varphi(\theta) + \alpha\varphi(\theta) = \varphi(\theta),$$

donc $\widetilde{\varphi}_\alpha \in \mathcal{E}^0_\varphi$. De ce fait, par optimalité de $\widehat{\varphi}$ on a

$$\forall\theta\in\Theta : V_\theta(\widetilde{\varphi}_\alpha) \geq V_\theta(\widehat{\varphi}).$$

Or

$$V_\theta(\widetilde{\varphi}_\alpha) = V_\theta(\widehat{\varphi} + \alpha\ (\widetilde{\varphi} - \widehat{\varphi})) = V_\theta(\widehat{\varphi}) + 2\alpha\ \mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) + \alpha^2 V_\theta(\widetilde{\varphi} - \widehat{\varphi}).$$

On en déduit que : $\forall\theta\in\Theta, \forall\alpha\in\mathbb{R}$:

$$2\alpha\ \mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) + \alpha^2 V_\theta(\widetilde{\varphi} - \widehat{\varphi}) \geq 0. \qquad (*)$$

Si $\alpha > 0$, $(*)$ conduit à

$$2\ \mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) + \alpha V_\theta(\widetilde{\varphi} - \widehat{\varphi}) \geq 0.$$

et en faisant $\alpha \downarrow 0^+$, on obtient que $\mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) \geq 0$.
Si $\alpha < 0$, $(*)$ conduit à

$$2\ \mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) + \alpha V_\theta(\widetilde{\varphi} - \widehat{\varphi}) \leq 0.$$

En faisant $\alpha \uparrow 0^-$, on obtient que $\mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) \leq 0$.
On a donc bien $\mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) = 0$.
2) $(ii) \Rightarrow (i)$: On suppose (ii).
Soient $\widetilde{\varphi} \in \mathcal{E}^0_\varphi$ et $\theta \in \Theta$. On a

$$\begin{aligned} V_\theta(\widetilde{\varphi}) &= V_\theta(\widehat{\varphi} + (\widetilde{\varphi} - \widehat{\varphi})) \\ &= V_\theta(\widehat{\varphi}) + 2\underbrace{\mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi})}_{=0 \text{ par } (ii)} + \underbrace{V_\theta(\widetilde{\varphi} - \widehat{\varphi})}_{\geq 0} \geq V_\theta(\widehat{\varphi}), \end{aligned}$$

donc $\widehat{\varphi}$ est bien un esbvm.
2) Si $\widehat{\varphi}$ et $\widetilde{\varphi}$ sont deux esbvm de $\varphi(\theta)$, alors $\forall\theta\in\Theta$:

$$\begin{aligned} V_\theta(\widetilde{\varphi} - \widehat{\varphi}) &= \mathrm{Cov}_\theta(\widetilde{\varphi} - \widehat{\varphi},\ \widetilde{\varphi} - \widehat{\varphi}) \\ &= \mathrm{Cov}_\theta(\widetilde{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) - \mathrm{Cov}_\theta(\widehat{\varphi}, \widetilde{\varphi} - \widehat{\varphi}) \\ &= 0 - 0 = 0 \qquad (\text{par } 1)) \end{aligned}$$

Donc il existe $c(\theta) \in \mathbb{R}$ tel que $\widetilde{\varphi} - \widehat{\varphi} = c(\theta) \quad P_\theta$−p.p. et comme

$$E_\theta(\widetilde{\varphi} - \widehat{\varphi}) = \varphi(\theta) - \varphi(\theta) = 0, \text{ on obtient que } c(\theta) = 0,$$

et donc $\widetilde{\varphi} = \widehat{\varphi} \quad P_\theta$−p.p. ■

2.2 Estimateur du maximum de vraisemblance et estimateur des moments

Dans ce court paragraphe, on définit deux procédures d'estimations classiques.

Définition 14 *Soient* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$*, un modèle statistique dominé par une mesure* μ *et*

$$f : (x,\theta) \in \mathcal{X} \times \Theta \mapsto f_\theta(x) \in \mathbb{R}^+$$

une fonction de vraisemblance associée et soit $(X_1, ..., X_n)$ *un échantillon i.i.d de v.a. associé à ce modèle (i.e.* $(X_1, ..., X_n)$ *est associé au modèle produit* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)^n$*). On dit que* $\widehat{\Theta}_n$ *,est un* ***estimateur du maximum de vraisemblance*** *(emv) de* $(\theta)_{\theta\in\Theta}$ *si*

$$\widehat{\Theta}_n \in \underset{\theta\in\Theta}{\arg\max} \prod_{1\leq j\leq n} f(X_j, \theta).$$

Remarques 15 *1) L'existence et l'unicité d'un emv n'est pas garantie (voir TD 2).*
2) On a aussi

$$\widehat{\Theta}_n \in \underset{\theta\in\Theta}{\arg\max} \sum_{1\leq j\leq n} \ln f(X_j, \theta)$$

ce qui est parfois plus maniable au niveau des calculs.
3) si $\widehat{\Theta}_n$ *est un emv de* $(\theta)_{\theta\in\Theta}$, $\varphi\left(\widehat{\Theta}_n\right)$ *est appelé emv de* $(\varphi(\theta))_{\theta\in\Theta}$.

Définition 16 *Soient* $\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), (P_\theta)_{\theta\in\Theta}\right)$*, un modèle statistique et soit* $(X_1, ..., X_n)$ *un échantillon i.i.d de v.a. associé à ce modèle On suppose qu'il existe une fonction* Ψ *et* $\ell \geq 1$ *tels que* $\forall\theta \in \Theta,\ \theta = \Psi\left(\mathbb{E}_\theta\left(X_1^k\right) : 1 \leq k \leq \ell\right)$ *et, pour* $1 \leq k \leq \ell$*, on pose*

$$\overline{X}_n^{(k)} := \frac{1}{n} \sum_{1\leq j\leq n} X_j^k.$$

Alors

$$\widetilde{\Theta}_n := \Psi\left(\overline{X}_n^{(k)} : 1 \leq k \leq \ell\right)$$

est appelé ***estimateur des moments*** *de* $(\theta)_{\theta\in\Theta}$.

Voir le TD2 pour de multiples exemples.

2.3 Exhaustivité. Amélioration d'estimateurs sans biais.

Rappels 17 *Soient* $(\mathcal{X}, \mathcal{B}, P)$ *un espace probabilisé,* $(\mathcal{Y}, \mathcal{D})$*, un espace mesurable et* $s : \mathcal{X} \to \mathcal{Y}$ *une fonction mesurable.*

1. *La sous tribu de* $\mathcal{B}$

$$\sigma(s) := \left\{ s^{-1}(D) : D \in \mathcal{D} \right\}$$

   est appelée ***tribu engendrée par*** s*.*

2. *On montre qu'une fonction* $w : \mathcal{X} \to \mathbb{R}$ *est* $\sigma(s) - \mathcal{B}(\mathbb{R})$ *mesurable ssi il existe* $g : \mathcal{Y} \to \mathbb{R}$ *telle que* g *est* $\mathcal{D} - \mathcal{B}(\mathbb{R})$ *mesurable et* $w = g \circ s$*.*

3. *Pour tout* $h \in \mathcal{L}^1(P)$*, on peut montrer qu'il existe* $w : \mathcal{X} \to \mathbb{R}$ *fonction* $\sigma(s) - \mathcal{B}(\mathbb{R})$ *mesurable telle que, pour tout* $D \in \mathcal{D}$ *:*

$$E_P\left(w \ \mathbf{1}_{s^{-1}(D)}\right) = E_P\left(h \ \mathbf{1}_{s^{-1}(D)}\right)$$

   (i.e. il existe $g : \mathcal{Y} \to \mathbb{R}$ *telle que pour tout* $D \in \mathcal{D}$*,*

$$E_P\left((g \circ s) \ \mathbf{1}_{s^{-1}(D)}\right) = E_P\left(h \ \mathbf{1}_{s^{-1}(D)}\right)).$$

   De plus, $w \in \mathcal{L}^1(P)$ *et* w *est unique modulo l'identification des fonctions* P*-p.s. égales.* $w(= g \circ s)$ *est notée* $E_P(h \mid s)$ *et appelée l'****espérance conditionnelle*** *(sous* P*) sachant* s *de* h*.*
   Pour $y \in \mathcal{Y}$*,* $g(y)$ *est aussi notée* $E_P(h \mid s = y)$*.*

4. *On montre que :*

   (a) *L'application*

$$h \in \mathcal{L}^1(P) \mapsto E_P(h \mid s)$$

   est linéaire et croissante.

   (b) $\forall h \in \mathcal{L}^1(P)$*,* $\forall v$ *fonction* $\sigma(s) - \mathcal{B}(\mathbb{R})$ *mesurable telle que* $v \ h \in \mathcal{L}^1(P)$ *:*

$$E_P(v \ h \mid s) = v \ E_P(h \mid s).$$

   (c) $\forall h \in \mathcal{L}^1(P)$ *:*

$$E_P(E_P(h \mid s)) = E_P(h).$$

Définition 18 *Soient* $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta \in \Theta}\right)$ *un modèle statistique,* $(\mathcal{Y}, \mathcal{D})$*, un espace mesurable et* $s : \mathcal{X} \to \mathcal{Y}$ *une fonction mesurable. On dit que* s *est une* ***statistique exhaustive*** *(par rapport au modèle) si, pour tout* $h \in \mathcal{L}^1\left((P_\theta)_{\theta \in \Theta}\right)$*,* $E_\theta(h \mid s)$ *ne dépend pas de* θ *(comme d'habitude, on note* E_θ *pour* E_{P_θ}*).*
Dans la suite, on notera $E(h \mid s)$ *la valeur commune des* $E_\theta(h \mid s)$*.*

Remarques 19 *1) Intuitivement, une statistique exhaustive est une transformation des données (i.e. de x) qui conserve toute l'information disponible sur* θ *(voir le Théorème* **??** *pour plus de précisions à ce propos).*
2) L'identité $\mathrm{id}_{\mathcal{X}} : x \in \mathcal{X} \mapsto x$ *est évidemment toujours exhaustive (mais n'a aucun intérêt en temps que résumé de l'information).*
3) Si s *est exhaustive et* $b : \mathcal{Y} \to \mathcal{Y}$ *est bijective, alors* $b \circ s$ *est exhaustive (car* $\sigma(b \circ s) = \sigma(s)$*).*

On va maintenant montrer comment, à partir d'une statistique exhaustive, il est possible d'améliorer un estimateur sans biais initial (via le Théorème 20), et même, parfois, d'obtenir l'esbvm (via le Théorème 24). Voir le TD 4 pour des applications de cette démarche.

Théorème 20 *(de Rao Blackwell)*
Soient $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique,* $\widehat{\varphi}$ *un estimateur sans biais de* $\varphi(\theta) \in \mathbb{R}$*,* $(\mathcal{Y}, \mathcal{D})$ *un espace mesurable et* $s : \mathcal{X} \to \mathcal{Y}$ *une statistique exhaustive. Alors* $\widehat{\varphi}_{|s} := E(\widehat{\varphi} \mid s)$ *est un estimateur sans biais de* $\varphi(\theta)$*, de plus,* $\widehat{\varphi}_{|s}$ *est préférable à* $\widehat{\varphi}$ *i.e.* $\forall \theta \in \Theta$:

$$E_\theta\left(\left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)^2\right) \leq E_\theta\left(\left(\widehat{\varphi} - \varphi(\theta)\right)^2\right).$$

Preuve. Comme s est exhaustive, $\widehat{\varphi}_{|s} = E_\theta(\widehat{\varphi} \mid s)$ ne dépend pas de θ : c'est bien une statistique. De plus, $\forall \theta \in \Theta$:

$$\begin{aligned}
E_\theta\left(\left(\widehat{\varphi} - \varphi(\theta)\right)^2 \mid s\right) &= E_\theta\left(\left(\widehat{\varphi} - \widehat{\varphi}_{|s}\right)^2 \mid s\right) \\
&\quad + 2E_\theta\left(\left(\left(\widehat{\varphi} - \widehat{\varphi}_{|s}\right)\left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)\right) \mid s\right) \\
&\quad + E_\theta\left(\left(\left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)\right)^2 \mid s\right) \\
(*) \qquad &= E_\theta\left(\left(\widehat{\varphi} - \widehat{\varphi}_{|s}\right)^2 \mid s\right) + \left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)^2 \\
&\geq \left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)^2,
\end{aligned}$$

avec $(*)$ car $\widehat{\varphi}_{|s} - \varphi(\theta)$ est $\sigma(s)$ mesurable donc

$$\begin{aligned}
&E_\theta\left(\left(\left(\widehat{\varphi} - \widehat{\varphi}_{|s}\right)\left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)\right) \mid s\right) \\
&= \left(\widehat{\varphi}_{|s} - \varphi(\theta)\right) \; E_\theta\left(\left(\widehat{\varphi} - \widehat{\varphi}_{|s}\right) \mid s\right) \\
&= \left(\widehat{\varphi}_{|s} - \varphi(\theta)\right) \; \left(\widehat{\varphi}_{|s} - \widehat{\varphi}_{|s}\right) = 0.
\end{aligned}$$

On en déduit que :

$$\begin{aligned}
E_\theta\left(\left(\widehat{\varphi} - \varphi(\theta)\right)^2\right) &= E_\theta\left(E_\theta\left(\left(\widehat{\varphi} - \varphi(\theta)\right)^2 \mid s\right)\right) \\
&\geq E_\theta\left(\left(\widehat{\varphi}_{|s} - \varphi(\theta)\right)^2\right)..
\end{aligned}$$

■

Le résultat suivant fournit un critère maniable pour établir l'exhaustivité d'une statistique.

Théorème 21 *(Critère de factorisation de Fisher Neyman)*
Soient $\left(\mathcal{X}, \mathcal{B}, (P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique dominé,* $(\mathcal{Y}, \mathcal{D})$, *un espace mesurable et* $s : \mathcal{X} \to \mathcal{Y}$ *une fonction mesurable.*
Les 2 assertions suivantes sont équivalentes :
i) s *est exhaustive.*
ii) Si μ *est une dominante du modèle, alors* $\forall\theta \in \Theta$, *il existe des fonctions mesurables*

$$g_\theta : y \in \mathcal{Y} \mapsto g_\theta(y) \in \mathbb{R}^+$$

et

$$u : x \in \mathcal{X} \mapsto u(x) \in \mathbb{R}^+$$

telles que $P_\theta = f_\theta \cdot \mu$. *avec*

$$f_\theta : x \in \mathcal{X} \mapsto u(x) \;\; g_\theta(s(x)).$$

Preuve. 1) $(i) \Rightarrow (ii)$: On suppose que s est exhaustive et pour h mesurable positive et $\theta \in \Theta$, on note $E(h \mid s)$ la valeur commune des $E_\theta(h \mid s)$.
a) Soit $P^* := \sum_{t\in T} c(t) \; P_t$ une dominante privilégiée du modèle (voir Théorème 10) Pour tout $\theta \in \Theta$, on considère $f_\theta^* \in \frac{dP_\theta}{dP^*}$ et on veut montrer que f_θ^* est (i.e. admet une version) $\sigma(s)$ −mesurable.
Pour h mesurable positive et tout $s^{-1}(D) \in \sigma(s)$:

$$\begin{aligned} E_*\left(h \; \mathbf{1}_{s^{-1}(D)}\right) &= \sum_{t\in T} c(t) \;\; E_t\left(h \; \mathbf{1}_{s^{-1}(D)}\right) \\ &= \sum_{t\in T} c(t) \;\; E_t\left(E_t(h \mid s) \;\; \mathbf{1}_{s^{-1}(D)}\right) \\ &= \sum_{t\in T} c(t) \;\; E_t\left(E(h \mid s) \;\; \mathbf{1}_{s^{-1}(D)}\right) \\ &= E_*\left(E(h \mid s) \;\; \mathbf{1}_{s^{-1}(D)}\right), \end{aligned}$$

donc $\forall\theta \in \Theta$:

$$E_*(h \mid s) = E(h \mid s) = E_\theta(h \mid s). \qquad (*)$$

Et, $\forall B \in \mathcal{B}$, $\forall\theta \in \Theta$:

$$\begin{aligned} E_*(f_\theta^* \; \mathbf{1}_B) &= E_\theta(\mathbf{1}_B) = E_\theta(E_\theta(\mathbf{1}_B \mid s)) = E_\theta(E_*(\mathbf{1}_B \mid s)) \qquad (\text{par } (*)) \\ &= E_*(f_\theta^* \; E_*(\mathbf{1}_B \mid s)) = E_*(E_*(f_\theta^* \; E_*(\mathbf{1}_B \mid s) \mid s)) \\ &= E_*(E_*(f_\theta^* \mid s) \;\; E_*(\mathbf{1}_B \mid s)) = E_*(E_*(E_*(f_\theta^* \mid s) \;\; \mathbf{1}_B \mid s)) \\ &= E_*(E_*(f_\theta^* \mid s) \;\; \mathbf{1}_B), \end{aligned}$$

donc $E_*(f_\theta^* \mid s) \in \frac{dP_\theta}{dP^*}$ et elle est bien $\sigma(s)$ −mesurable i.e. il existe une fonction mesurable

$$g_\theta : y \in \mathcal{Y} \mapsto g_\theta(y) \in \mathbb{R}^+$$

telle que $P_\theta = (g_\theta \circ s) \cdot P^*$.
b) Soit μ est une dominante quelconque.

$\forall\theta \in \Theta$, on a $P^* \ll \mu$ (c.f. Théorème 10) donc il existe une fonction $u : \mathcal{X} \to \mathbb{R}^+$ telle que

$$P^* = u \cdot \mu$$

donc, par a), $P_\theta = (u \times (g_\theta \circ s)) \cdot \mu$.
$(ii) \Rightarrow (i)$: Soient μ est une dominante du modèle,

$$g_\theta : y \in \mathcal{Y} \mapsto g_\theta(y) \in \mathbb{R}^+ \quad (\theta \in \Theta)$$

et

$$u : x \in \mathcal{X} \mapsto u(x) \in \mathbb{R}^+$$

des fonctions mesurables telles que $P_\theta = f_\theta \cdot \mu$. avec $f_\theta = u \times (g_\theta \circ s)$.
a) On considère $P^* := \sum_{t\in T} c(t)\ \ P_t$ une dominante privilégiée. Montrons que $\frac{dP_\theta}{dP^*}$ admet une version f_θ^* $\sigma(s)$ $-$mesurable.
Si on pose

$$g : y \in \mathcal{Y} \mapsto \sum_{t\in T} c(t)\ \ g_t(y),$$

il est facile de voir que $P^* = f^* \cdot \mu$. avec

$$f^* : x \in \mathcal{X} \mapsto u(x)\ \ g(s(x)).$$

Pour $\theta \in \Theta$, définissons la fonction $f_\theta^* : \mathcal{X} \longrightarrow \mathbb{R}^+$ telle que :

$$\begin{aligned} f_\theta^*(x) &= \frac{g_\theta(s(x))}{g(s(x))} \quad \text{si} \quad g(s(x)) > 0 \\ &= 0 \quad \text{sinon} \end{aligned}$$

Il est clair que f_θ^* est $\sigma(s)$ $-$mesurable. On va montrer que $f_\theta^* \in \frac{dP_\theta}{dP^*}$.
Pour tout $B \in \mathcal{B}$:

$$E_*\left(f_\theta^*\ \mathbf{1}_B\right) = E_*\left(f_\theta^*\ \mathbf{1}_{B\cap\{g\circ s>0\}}\right) + E_*\left(f_\theta^*\ \mathbf{1}_{B\cap\{g\circ s=0\}}\right).$$

Or $E_*\left(f_\theta^*\ \mathbf{1}_{B\cap\{g\circ s=0\}}\right) = 0$ car $P^*(\{g\circ s = 0\}) \leq P^*(\{f^* = 0\}) = 0$.
On a donc

$$\begin{aligned} E_*\left(f_\theta^*\ \mathbf{1}_B\right) = E_*\left(f_\theta^*\ \mathbf{1}_{B\cap\{g\circ s>0\}}\right) &= \int_{\mathcal{X}} \mathbf{1}_{B\cap\{g\circ s>0\}}\ f_\theta^*\ f^*\ d\mu \\ &= \int_{\mathcal{X}} \mathbf{1}_{B\cap\{g\circ s>0\}}\ \frac{g_\theta\circ s}{g\circ s}\ \ u\ g\circ s\ d\mu \\ &= \int_{\mathcal{X}} \mathbf{1}_{B\cap\{g\circ s>0\}}\ \ u\ g_\theta\circ s\ d\mu \\ &= \int_{\mathcal{X}} \mathbf{1}_{B\cap\{g\circ s>0\}}\ \ f_\theta\ d\mu \\ &= P_\theta(B\cap\{g\circ s>0\}) \underset{(*)}{=} P_\theta(B) \quad \text{i.e. } f_\theta^* \in \frac{dP_\theta}{dP^*}. \end{aligned}$$

Où, au niveau de $(*)$ on a utilisé le fait que $P_\theta\left(\{g\circ s=0\}\right)=0$
(car $P^*\left(\{g\circ s=0\}\right)=0$ et $P_\theta \ll P^*$).
b) Soit h une fonction mesurable positive, on va montrer que $\forall\theta\in\Theta$:

$$E_\theta\left(h\mid s\right)=E_*\left(h\mid s\right)\quad \text{(et donc ne dépend pas de } \theta\text{)},$$

ce qui établira l'exhaustivité de s.
Pour tout $\theta\in\Theta$ et tout $s^{-1}(D)\in\sigma(s)$:

$$\begin{aligned}
E_\theta\left(h\ \mathbf{1}_{s^{-1}(D)}\right) &= E_*\left(f_\theta^*\ h\ \mathbf{1}_{s^{-1}(D)}\right)\\
&= E_*\left(E_*\left(f_\theta^*\ h\ \mathbf{1}_{s^{-1}(D)}\mid s\right)\right)\\
&= E_*\left(E_*\left(h\mid s\right)\ f_\theta^*\mathbf{1}_{s^{-1}(D)}\right)\quad \text{(car } f_\theta^* 1_{s^{-1}(D)} \text{ est } \sigma(s)-\text{mesurable)}\\
&= E_\theta\left(E_*\left(h\mid s\right)\ \mathbf{1}_{s^{-1}(D)}\right)
\end{aligned}$$

i.e. $E_\theta\left(h\mid s\right)=E_*\left(h\mid s\right)$ c.q.f.d. ■

Définition 22 *Soient* $\left(\mathcal{X},\mathcal{B},(P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique,* $(\mathcal{Y},\mathcal{D})$ *un espace mesurable et* $s:\mathcal{X}\to\mathcal{Y}$ *une statistique. On dit que* s *est* ***complète*** *(ou totale) si :*

$$\left[w\in\mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right),\ \sigma(s)-\mathcal{B}(\mathbb{R})\ \text{ mesurable et }\ E_\theta(w)=0,\ \forall\theta\in\Theta\right]\Rightarrow\left[w=0\quad P_\theta-\text{p.s. }\forall\theta\in\Theta\right],$$

i.e. si

$$\left[g\circ s\in\mathcal{L}^2\left((P_\theta)_{\theta\in\Theta}\right)\quad\text{et}\quad E_\theta(g\circ s)=0,\ \forall\theta\in\Theta\right]\Rightarrow\left[g\circ s=0\quad P_\theta-\text{p.s. }\forall\theta\in\Theta\right].$$

Remarques 23 *Dans la définition précédente, on peut remplacer les* $\mathcal{L}^2$ *par toutes classes de parties denses dans* $\mathcal{L}^2$*.*

Théorème 24 *(de Lehmann Scheffé)*
Soient $\left(\mathcal{X},\mathcal{B},(P_\theta)_{\theta\in\Theta}\right)$ *un modèle statistique,* $\widehat{\varphi}$ *un estimateur sans biais de* $\varphi(\theta)\in\mathbb{R}$*,* $(\mathcal{Y},\mathcal{D})$ *un espace mesurable et* $s:\mathcal{X}\to\mathcal{Y}$ *une statistique exhaustive et complète.*
Alors $\widehat{\varphi}_{|s}:=E\left(\widehat{\varphi}\mid s\right)$ *est l'estimateur sans biais optimal de* $\left(\varphi(\theta)\right)_{\theta\in\Theta}$.

Preuve. Soit $\widetilde{\varphi}$ un estimateur sans biais de $\left(\varphi(\theta)\right)_{\theta\in\Theta}$. Comme s est exhaustive, le Théorème de Rao Blackwell assure que $E\left(\widetilde{\varphi}\mid s\right)$ est un esb préférable à $\widetilde{\varphi}$.
De plus $E\left(\widehat{\varphi}\mid s\right)-E\left(\widetilde{\varphi}\mid s\right)$ est $\sigma(s)$ mesurable et, comme, pour tout $\theta\in\Theta$,

$$\begin{aligned}
E_\theta\left(E\left(\widehat{\varphi}\mid s\right)-E\left(\widetilde{\varphi}\mid s\right)\right) &= E_\theta\left(E\left(\widehat{\varphi}-\widetilde{\varphi}\mid s\right)\right)\\
&= E_\theta\left(E_\theta\left(\widehat{\varphi}-\widetilde{\varphi}\mid s\right)\right)\\
&= E_\theta\left(\widehat{\varphi}-\widetilde{\varphi}\right)=\varphi(\theta)-\varphi(\theta)=0,
\end{aligned}$$

la complétude de s conduit à $E\left(\widehat{\varphi}\mid s\right)-E\left(\widetilde{\varphi}\mid s\right)=0$ i.e. $E\left(\widehat{\varphi}\mid s\right)=E\left(\widetilde{\varphi}\mid s\right)$ et $E\left(\widehat{\varphi}\mid s\right)$ est donc préférable à $\widetilde{\varphi}$. ■