BIAIS (C5, H1, H5)
(20 / 12 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) Soit $(\Omega, \mathscr{T}, P)$ un **espace probabilisé**, $\xi : \Omega \mapsto \mathbf{R}^K$ un **vecteur aléatoire** réel tq $\xi \in L^1_{\mathbf{R}K} (\Omega, \mathscr{T}, P)$ et $\alpha \in \mathbf{R}^K$ un point donné.

On appelle **biais** (**au sens de l'espérance**), ou **erreur moyenne**, de ξ pr à α le vecteur :

(1) $\quad B(\xi, \alpha) = E(\xi - \alpha) = E\,\xi - \alpha.$

Ce vecteur est aussi noté $B_\alpha\,\xi$, et représente l'**écart** moyen entre ξ et $\alpha \in \mathbf{R}^K$.

(ii) Soit $(\Omega, \mathscr{T}, P_\theta)_{\theta \in \Theta}$ un **modèle statistique paramétrique** (avec $\Theta \subset \mathbf{R}^Q$) et X une **va** à valeurs dans un **espace d'observation** $(\mathscr{X}, \mathscr{B})$. On note $(\mathscr{X}, \mathscr{B}, P_\theta^X)_{\theta \in \Theta}$ le **modèle image** associé et l'on considère un **estimateur** $t : \mathscr{X} \mapsto \Theta$ (resp $t : \mathscr{X} \mapsto \mathbf{R}^Q$) de θ (resp de $g(\theta)$, où $g : \Theta \mapsto \mathbf{R}^Q$ est un **reparamétrage** de θ).

Si la fonction :

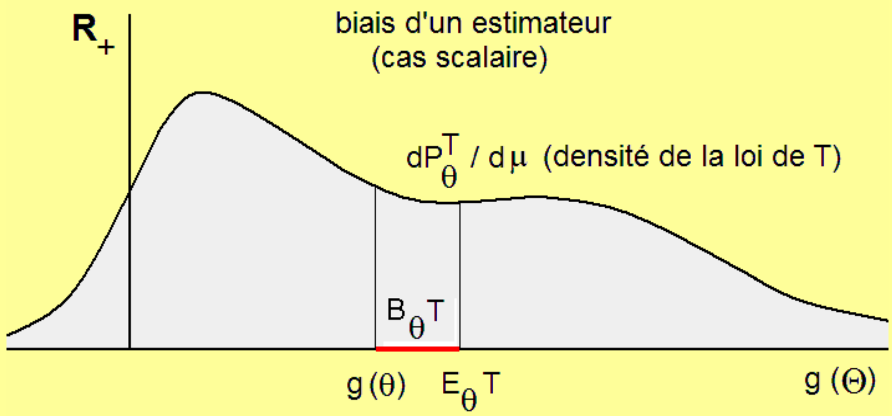
(2) $\quad T = t \circ X : \Omega \mapsto \Theta \quad$ (resp $T = t \circ X : \Omega \mapsto \mathbf{R}^Q$)

appartient à $L^1_{\mathbf{R}K} (\Omega, \mathscr{T}, P_\theta)$, $\forall\ \theta \in \Theta$, on appelle **biais de l'estimateur** T de θ (resp de $g(\theta)$) le vecteur :

$(3)_a \quad B(T, \theta) = E_\theta (T - \theta) = E_\theta T - \theta,$

$(3)_b \quad$ resp $B_\theta (T, g(\theta)) = E_\theta (T - g(\theta)) = E_\theta T - g(\theta),$

qui est aussi noté $B_\theta T$ (resp $B_{g(\theta)}\ T$, ou même encore $B_\theta T$) (cf graphique ci-dessous dans le cas scalaire).



On dit qu'un estimateur vectoriel est un **estimateur sans biais** ssi son biais est nul, quel que soit $\theta \in \Theta$.

(iii) La notion de **biais** peut, plus généralement, s'apprécier **selon une caractéristique de centralité** (ou selon un **paramètre de position**) autre que l'**espérance mathématique** (cf aussi **régression**).

Ainsi, dans le cadre précédent, si les probabilités P_θ admettent toutes une caractéristique de centralité, notée sous forme d'opérateur C_θ , le biais (au sens de C_θ) peut se définir comme :

(4) $B_\theta T = C_\theta (T - \theta)$ (resp $C_\theta (T - g(\theta))$).

Lorsque $\Theta = \mathbf{R}$, on dit ainsi que T est **sans biais** (**au sens de la médiane**) ssi :

(5) $P_\theta (T < \theta) = 1/2 = P_\theta (T \geq \theta), \quad \forall \theta \in \Theta,$

ie ssi $Q_{\theta,1/2} T = 0$ ($Q_{\theta,1/2} T$ représentant, $\forall \theta \in \mathbf{R}$, la **médiane** de la **loi** P_θ^T de T.

De même, si $\Theta = \mathbf{R}^Q$, on dit que T est **sans biais** (**au sens du mode**) ssi :

(6) $S_\theta (T - \theta) = 0$ (resp $S_\theta (T - g(\theta)) = 0$), $\forall \theta \in \Theta,$

$S_\theta T$ étant le **mode** (supposé unique) de la loi P_θ^T de T.

La notion de biais connaît une extension asymptotique (cf **biais asymptotique**).