STATISTIQUE COMPLÈTE (G1)
(22 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $(\Omega, \mathcal{T}, \mathcal{P})$ un **modèle statistique**, $(\mathcal{X}, \mathcal{B})$ un **espace d'observation**, $(\mathcal{S}, \mathcal{O})$ un **espace probabilisable** auxiliaire, $X : \Omega \mapsto \mathcal{X}$ une **variable aléatoire** (**échantillon**) et $S = s(X)$ une **statistique** définie à partir d'une **application mesurable** $s : \mathcal{X} \mapsto \mathcal{S}$.

On dit que S est une **statistique complète**, ou une **statistique totale**, ssi la famille $\mathcal{P}^S$ des **lois de probabilité** $P^S = s(P^X) = (s \circ X)(P)$ (où P parcourt $\mathcal{P}$) est une **famille complète de lois**.

Autrement dit, $\mathcal{P}^S$ est tq, pour tout couple (φ, ψ) de fonctions mesurables $\mathcal{S} \mapsto \mathbf{R}^Q$ et toute **mesure de probabilité** $P \in \mathcal{P}$:

(1) $\quad E_P\, \varphi(S) = E_P\, \psi(S) \quad \Rightarrow \quad \varphi(S) = \psi(S), \quad \mu\text{-p.p.}.$

Ceci suppose que $\mathcal{P}^S$ est une **famille de lois dominée** pr à une **mesure positive** μ.

(ii) La définition (1) équivaut à la suivante :

(1)' $\quad \forall\, \varphi \in \mathcal{M}(\mathcal{S}, \mathbf{R}^Q), \forall\, P \in \mathcal{P}, \quad E_P\, \varphi(S) = 0 \Rightarrow \varphi(S) = 0, \quad \mu\text{-p.p.}.$

Dans ce qui précède, on note eg (cf **théorème de transfert des mesures**) :

(2) $\quad E_P\, \varphi(S) = \int_\Omega \varphi(s \circ X)\, dP = \int_{\mathcal{S}} \varphi\, dP^S = 0 = \int_{\mathcal{X}} \varphi(s(x))\, dP^X(x).$

(iii) De même, on dit que S est une **statistique bornée complète**, ou une **statistique bornée totale**, ssi on remplace les fonction mesurables (φ, ψ) précédentes par des fonctions bornées. On parle aussi, dans ce cas, de **statistique quasi-complète**, ou de **famille quasi-complète de lois de probabilité**.

(iv) Soit $\mathcal{C}$ une sous-**tribu** de $\mathcal{B}$. On note $\mathcal{P}^X{}_{/\mathcal{C}}$ la **restriction** de $\mathcal{P}^X$ (familles des lois de X) à la **tribu de parties** $\mathcal{C}$, ie la famille des restrictions $P^X{}_{/\mathcal{C}}$ des lois P^X à $\mathcal{C}$.

On dit alors que le sous-modèle $(\mathcal{X}, \mathcal{C}, \mathcal{P}^X{}_{/\mathcal{C}})$ est un **modèle complet** ssi toute statistique centrée et $\mathcal{C}$-mesurable s est $\mathcal{P}^X{}_{/\mathcal{C}}$ -équivalente à zéro, ie :

(3) $\quad E\, s(X) = 0 \Rightarrow s(X) = 0,$

où l'**espérance mathématique** E est calculée avec les lois $P^X{}_{/\mathcal{C}}$.

Une statistique S est complète ssi elle induit un **modèle complet**.

(v) Ce qui précède se transpose directement à une famille paramétrée $\mathscr{P} = (P_\theta)_{\theta \in \Theta}$. On note $\mathscr{P}^X = (P_\theta^X)_{\theta \in \Theta}$ la famille des lp de X et $\mathscr{P}^S = (P_\theta^S)_{\theta \in \Theta}$ celle des lois de S = s o X.

On dit alors que S est une **statistique complète**, ou une **statistique totale**, pour le paramètre $\theta \in \Theta$ ssi :

(1)" $\forall \varphi \in \mathscr{M}(\mathcal{S}, \mathbf{R}^Q), \forall \theta \in \Theta, \; E_\theta \varphi(S) = 0 \;\Rightarrow\; \varphi(S) = 0,$ μ-p.p.,

où l'on note E_θ l'espérance calculée avec P_θ (ou P_θ^X, ou encore P_θ^S).

(vi) A titre d'exemple :

(a) la famille $(\mathscr{B}(n, p))_{n \in \mathbf{N}^*}$ des **lois binômiales** (où $p \in]0, 1[$ et n est fixé) est une famille complète. Il en va de même de la famille $(\mathscr{U}(0, \theta))$ (où $\theta \in \mathbf{R}_+^*$) des **lois uniformes continues** dont le **support** dépend du paramètre θ (cf **support d'une probabilité**) ;

(b) par contre, la famille des **lois normales** $\mathscr{N}_Q(\mu\, e_Q, \Sigma)$, dans laquelle $\mu \in \mathbf{R}$, $e_Q = (1, ..., 1)' \in \mathbf{R}^Q$ est le premier vecteur bissecteur et Σ est une **matrice définie positive**, n'est pas complète.

(vii) Une **statistique exhaustive** n'est pas nécessairement complète :

(a) soit S = s (X) une statistique exhaustive pour la famille $(P^X)_{P \in \mathscr{P}}$. On note $(P^S)_{P \in \mathscr{P}}$ la famille de ses lois, ie des images $P^S = (s \circ X)(P)$ de P par S, ou des images $P^S = s(P^X)$ de P^X par s.

Si $(P^S)_{P \in \mathscr{P}}$ est complète, on dit que S est une **statistique exhaustive complète** ;

(b) dans le cas « paramétré », soit S = s (X) une statistique exhaustive pour la famille $(P_\theta^X)_{\theta \in \Theta}$. On note $(P_\theta^S)_{\theta \in \Theta}$ la famille des lp de S, ie la famille des images $P_\theta^S = (s \circ X)(P_\theta)$ de P_θ par S, ou des images $P_\theta^S = s(P_\theta^X)$ de P_θ^X par s.

Si $(P_\theta^S)_{\theta \in \Theta}$ est complète, on dit que S est une **statistique exhaustive complète**.