SONDAGE CONVERGENT (M6)
(16 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $\Omega = \{\omega_1 ,..., \omega_M\}$ un ensemble fini (**population**), Π un **plan de sondage** sur Ω, $(\mathcal{Y}, \mathcal{C})$ un **espace d'observation**, $\eta : \Omega \mapsto \mathcal{Y}$ une **variable** (ou un **caractère**) **observable** et $Y = (Y_1 ,..., Y_M)$ l'ensemble des valeurs prises par η dans la population Ω, ie $Y_m = \eta (\omega_m)$, $\forall$ m = 1 ,..., M.

On note $A = \{a_1 ,..., a_N\}$ un N-**échantillon aléatoire** tiré dans Ω selon Π et $y = (y_1 ,..., y_N)$ les valeurs prises par η sur l'échantillon A, ie $y_n = \eta (a_n)$, $\forall$ n = 1 ,..., N.

Soit $(\mathcal{Z}^L, \mathcal{D}^{\otimes L})$ un **espace mesurable**, $g : \mathcal{Y}^M \mapsto \mathcal{Z}^L$ $(L \leq N)$ une fonction donnée, et g (Y) un **paramètre d'intérêt**, déduit du « **paramètre** » Y.

Soit $T_N = t_N (y)$ une **statistique** permettant de réaliser une inférence (eg estimation) sur g (Y). On étudie la situation dans laquelle $N \to +\infty$.

(ii) Lorsque le sondage de A dans Ω est sans remise (cf **sondage exhaustif**, **tirage exhaustif**), l'étude des **propriétés asymptotiques** de T_N nécessite que $M \to +\infty$, puisque $1 \leq N \leq M$. On peut donc définir une fonction φ : **N*** $\mapsto$ **N*** tq $M = \varphi (N)$ et que $N \to +\infty \Rightarrow \varphi (N) \to +\infty$. Une autre optique consiste seulement à faire tendre N vers M, M restant fixe (optique à « distance finie » $N \leq M$, ou « avec grand échantillon » si $M >> 0$) ;

(ii) Lorsque le sondage est avec remise (cf **sondage bernoullien**, **tirage bernoullien**), l'étude asymptotique de T_N ne suppose pas nécessairement que $M \to +\infty$, puisque les inégalités $1 \leq N \leq M$ ne sont plus nécessaires. On peut cependant étendre le cas précédent en choisissant une fonction φ constante.

(iii) Les deux **sondages élémentaires** précédents montrent que l'étude asymptotique d'un sondage quelconque (Π, T_N) doit supposer l'existence d'une fonction adéquate φ : **N*** $\mapsto$ **N*** tq $\varphi (N) = M$.

Lorsque N ou M varie, on note Ω_M ou $\Omega_{\varphi (N)}$ au lieu de Ω, Π_M ou $\Pi_{\varphi (N)}$ au lieu de Π, et Y (M) ou $Y_{\varphi (N)}$ au lieu de Y. De même, on note A (N) au lieu de A et y (N) au lieu de y. Comme y varie avec N, on note aussi $T_N = t_N (y (N))$ au lieu de $t_N (y)$. Enfin, comme Y varie avec N selon Y (φ (N)), la fonction g varie aussi avec N, et l'on note alors $g_N (Y (\varphi (N)))$ le « paramètre » à estimer, paramètre dont la « dimension » $L \leq N$ peut même, elle aussi, varier avec N (eg lorsqu'il existe une fonction ϕ : **N*** $\mapsto$ **N*** tq $1 \leq L \leq \phi (N) \leq N$).

Par suite, l'analyse asymptotique d'un sondage consiste à étudier la **suite** (aléatoire) $(T_N)_{N \in N^*}$ constituée d'estimateurs T_N du « paramètre » $g_N (Y (\varphi (N)))$; ce dernier définit lui-même une suite (non aléatoire, dans un cadre non bayésien) notée $(g_N (Y (\varphi (N))))_{N \in N^*}$.

On dit alors que la suite $(\Pi_{\varphi(N)}, T_N)_{N \in N^*}$ est une **suite de sondages convergente**, ou simplement un **sondage convergent**, pour la suite $(g_N)_{N \in N^*}$ lorsque N tend vers l'infini ssi (**convergence en probabilité** généralisée) :

(1) $\forall\, \varepsilon > 0, \forall\, \eta > 0, \exists\, N_{\varepsilon,\eta}$ tq $N \geq N_{\varepsilon,\eta} \;\Rightarrow\; \Pi_{\varphi(N)}(\|T_N - g_N(Y(\varphi(N)))\| > \varepsilon) < \eta$,

où $(\mathcal{Z}^L, \|.\|)$ est supposé être un **espace normé**. Autrement dit, pour tout $\varepsilon > 0$, l'expression $\Pi_{\varphi(N)}(\|T_N - g_N(Y(\varphi(N)))\| > \varepsilon)$ tend vers 0 lorsque N tend vers l'infini.

(iv) On définit de même la **convergence en moyenne quadratique** du sondage (ou de la suite) précédent(e) en remplaçant (1) par :

(2) $\forall\, \varepsilon > 0, \forall\, \eta > 0, \exists\, N_{\varepsilon,\eta}$ tq $N \geq N_{\varepsilon,\eta} \;\Rightarrow\; E_N \|T_N - g_N(Y(\varphi(N)))\|^2 < \varepsilon$.

Autrement dit, $E_N \|T_N - g_N(Y(\varphi(N)))\|^2 \to 0$ lorsque $N \to +\infty$. Ici, E_N désigne l'**espérance mathématique** calculée avec le plan $\Pi_{\varphi(N)}$.