ESTIMATEUR A DISTANCE MINIMUM (H3)
(05 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'expression **estimateur à distance minimum** désigne un estimateur obtenu en minimisant une distance. Cette méthode de génération d'un **estimateur** consiste, en effet, à déduire celui-ci d'un problème d'optimisation (minimisation) associé à une distance : eg **estimateur de la fonction de répartition** ou **estimateur de la densité**, **estimateur des moindres carrés généralisés**, **estimateur à distance minimale** de ZELLNER.

L'exemple de l'estimation de la **fr** est typique de l'approche.

(i) Soit $(\Omega, \mathcal{T}, P_\theta)_{\theta \in \Theta}$, où $\Theta \subset \mathbf{R}^Q$, un **modèle statistique** et $\xi : \Omega \mapsto \mathbf{R}$ une **vars** dont l'une des **fr** possibles est F_θ (fr associée à P_θ), $\forall \theta \in \Theta$.

Soit $X = (X_1, ..., X_N)$ un **échantillon iid** issu de ξ et dont la **fonction de répartition empirique** F_N est définie selon :

(1) $$F_N(x) = N^{-1} . \text{Card}\, \{n \in N_N^* : X_n \leq x\}, \quad \forall x \in \mathbf{R},$$

ce qui se note aussi :

(1) $$F_N(x) = N^{-1} . \#\, \{n \in N_N^* : X_n \leq x\}, \quad \forall x \in \mathbf{R}.$$

On munit l'ensemble $\mathcal{F}$ des fr sur **R** d'une **distance** d, ce qui définit un **espace métrique** $(\mathcal{F}, d)$.

On appelle alors **estimateur à distance minimum** de F_θ (ou de θ) tout estimateur $T_N = t_N(X)$, défini par l'**application mesurable** (ou **statistique**) $t_N : \mathbf{R}^N \mapsto \Theta$ et tq (en notant T(N) pour désigner T_N) :

(2) $$d(F_{T(N)}, F_N) = \inf_{\theta \in \Theta} d(F_\theta, F_N).$$

(ii) Si $\Theta \subset \mathbf{R}^Q$, si $L \subset \mathbf{R}^S$ (avec $S \in \mathbf{N}^*$) et si $g : \Theta \mapsto L$ est une **application surjective** ($g(\Theta) = L$), alors la statistique $g(T_N)$ est un estimateur à distance minimum du « nouveau » paramètre $g(\theta)$ (**propriété d'invariance**).

Les distances suivantes sont très utilisées, notamment en **Statistique non paramétrique** :

(a) **distance de H. CRAMER - R. von MISES** :

(3) $$d(F_\theta, F_N) = \int_{[0,1]} F_N \{F_\theta^{-1}(p) - p\}^2\, dp,$$

où F_θ^{-1} est la **fonction quantile** associée à F_θ ;

(b) **distance de A.N. KOLMOGOROV** (en notant **R**Q pour désigner $\mathbf{R}^Q$) :

(4) $$d(F_\theta , F_N) = \sup_{x \in \mathbf{R}Q} |F_\theta (x) - F_N (x)| \;;$$

(c) **distance de N.H. KUIPER** :

(5) $$d(F_\theta , F_N) = \sup_{(x,y) \in \mathbf{R}Q \times \mathbf{R}Q} |\{F_\theta (y) - F_\theta (x)\} - \{F_N (y) - F_N (x)\}|.$$

(iii) Les notions précédentes s'étendent à plusieurs variables. Il existe aussi des tests (généralement de même nom) attachés à ces distances.