MÉTHODE DE QUENOUILLE (C5, H)
(03 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **méthode de QUENOUILLE** est une méthode générale de réduction du **biais** d'un **estimateur** biaisé.

(i) Soit $(\Omega, \mathcal{T}, P_\theta)_{\theta \in \Theta}$ un **modèle statistique**, $(\mathcal{X}_0, \mathcal{B}_0)$ un **espace d'observation** et $\xi : \Omega \mapsto \mathcal{X}_0$ une **va** dont l'une des **lois** possibles est P_θ^ξ (avec $\theta \in \Theta$).

On observe un **échantillon iid** $X = (X_1, ..., X_N)$ (ie constitué de **copies** indépendantes de la **variable parente** ξ), ce qui permet de définir le **modèle d'échantillonnage** à distance finie $(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}$, avec $\mathcal{X} = \mathcal{X}_0^N$, $\mathcal{B} = \mathcal{B}_0^{\otimes N}$ et $P_\theta^X = (P_\theta^\xi)^{\otimes N}$.

Soit $g : \Theta \mapsto \mathbf{R}^L$ une fonction **mesurable** et $\tau = g(\theta)$ un **paramètre** d'intérêt. On note, $\forall N \in \mathbf{N}^*$, $t_N : \mathcal{X} \mapsto \mathbf{R}^L$ une fonction mesurable définissant un **estimateur** T_N (obtenu eg par la **méthode du maximum de vraisemblance** ou par la **méthode des moments**) de τ. La **suite** des estimateurs $(T_N)_{N \in \mathbf{N}^*}$ est donc fondée sur la suite $(t_N)_{n \in \mathbf{N}^*}$.

Enfin, suppose que T_N est un estimateur biaisé de τ (ie $E\, T_N - \tau \neq 0$).

(ii) La **méthode de M.H. QUENOUILLE**, ou parfois **méthode de M.H. QUENOUILLE - J.W. TUKEY**, (en anglais : « *jack-knife method* ») consiste à définir :

(a) des **pseudo-estimateurs**, ou **pseudo-valeurs** :

(1) $D_{Nn} = N\, T_N - (N-1)\, T_{N-1,n}$,

expression dans laquelle $T_{N-1,n}$ désigne, $\forall n \in N_N^*$, l'estimateur de τ de même « type » que T_N mais basé sur le (N-1)-échantillon déduit de X en lui otant la coordonnée X_n : cet estimateur est donc basé sur une fonction mesurable $t_{N-1,n} : \mathcal{X}_{N-1,n} \mapsto \mathbf{R}^L$, où $\mathcal{X}_{N-1,n} = \Pi_{\alpha \neq n}\, \mathcal{X}_\alpha$ et $\mathcal{X}_\alpha = \mathcal{X}_0$, $\forall \alpha$;

(b) l'**estimateur de M.H. QUENOUILLE** comme **moyenne arithmétique** simple des pseudo-estimateurs :

(2) $J_N = N^{-1} \Sigma_{n=1}^N D_{Nn}$.

Autrement dit, on calcule la moyenne des estimateurs de même type que t_N (ou T_N), mais basés sur N-1 observations X_n, puis on calcule l'estimateur résultant selon (2). La procédure précédente revient donc à remplacer l'estimateur d'ensemble T_N par une **combinaison linéaire convexe** d'estimateurs partiels D_{Nn}.

(iii) Les principales propriétés de l'estimateur obtenu sont les suivantes :

(a) sous certaines **conditions de régularité**, le biais de J_N est moindre que celui de T_N. Ainsi, lorsque $\Theta = \mathbf{R}^L$, $g = id_\Theta$ et $L = 1$, si le biais de $t_N(X) = T_N$ est de la forme :

(3) $B_\theta\, t_N(X) = \Sigma_{\alpha=1}^{+\infty}\, b_\alpha(\theta) / N^\alpha$,

avec $b_1(\theta) \neq 0$, on montre que le biais de J_N est de la forme :

(4) $B_\theta\, J_N = \Sigma_{\alpha=2}^{+\infty}\, c_\alpha(\theta) / N^\alpha$.

Autrement dit, si $B_\theta\, T_N = O(1/N)$ (grand zéro), on a $B_\theta\, J_N = O(1/N^2)$;

(b) sous des hypothèses assez larges, J_N suit asymptotiquement une **loi normale** (cf **normalité asymptotique**) dont la **matrice de covariance** peut être estimée par :

(5) $(V\, J_N)^\# = N^{-1}\,(N-1)^{-1}\,(D_{Nn} - J_N)\,(D_{Nn} - J_N)'$;

(c) dans le cas scalaire (ie $\Theta = \mathbf{R}^L$, L = 1 et $g = id_\Theta$), on montre que :

(6) $u_N = N^{1/2}\{(J_N - \theta) / s_N\} \rightarrow^{\mathcal{L}} \mathcal{S}_{N-1}$ (**loi de STUDENT** à N - 1 **dl**),

où $s_N^2 = (N-1)^{-1}\,\Sigma_{n=1}^{N}\,(D_{Nn} - J_N)^2$;

(d) la méthode peut s'appliquer à l'estimateur de QUENOUILLE lui-même, donc s'itérer ad libitum. Si J_N possède un biais d'ordre N^{-1}, alors l'itéré d'ordre j de J_N, soit $J_N^{(j)}$, possède un biais d'ordre $N^{-(j+1)}$.

(iv) Plusieurs variantes ou extensions ont été étudiées. Par exemple :

(a) le cas où les coordonnées X_n de X ne forment pas une **suite iid** selon P^ξ ;

(b) de même, si N = H . K, on peut calculer des pseudo-estimateurs d'ordre K sur les H groupes extraits de X (**partition** de X en H sous-échantillons).

(v) La procédure du « ***couteau de Jack*** » (ou ***« couteau suisse »***) est aussi utilisée en **Statistique non paramétrique** et dans les problèmes de **robustesse**. Ainsi :

(a) l'estimation de la **variance** théorique σ^2 par la variance d'échantillon S_N^2 est souvent incorrecte (biais) lorsque l'**hypothèse** de **normalité** $P^\xi = \mathcal{N}_K(\mu, \Sigma)$ de la population n'est pas vérifiée ;

(b) de même, si $\mathcal{T}$ désigne une **famille** de **fr** (associée à la famille des lois initiales) et $\phi : \mathcal{T} \mapsto \mathbf{R}$ une **fonctionnelle**, alors un estimateur « naturel » de T est $T^\# = \phi(F_N)$, où F_N est la **fr empirique** associée à l'**échantillon iid** X. La méthode de QUENOUILLE permet d'estimer le biais de $T^\#$, ie $E\,T^\# - T = E\,\phi(F_N) - \phi(F)$, ainsi que son écart-type, sans avoir à définir un **modèle paramétrique**.

Les pseudo-valeurs peuvent encore être utilisées pour détecter des **aberrations** X_n ainsi que leur influence sur T_N (estimateur initial) ou sur J_N (estimateur de QUENOUILLE) (cf **courbe d'influence**).