ESTIMATEUR PAR QUOTIENT (H, M)

(i) Dans le même cadre que celui définissant l'**estimateur par différence**, on suppose que le vecteur Y satisfait à une hypothèse bayésienne de la forme (cf **superpopulation**) :

(1) $\quad E\,Y = b_1 \,.\, X, \qquad$ avec $V\,Y = \sigma^2 \,.\, X^{\wedge}$,

où X est un vecteur à coordonnées positives, défini sur Ω, et $X^{\wedge}$ est la matricialisée de X (cf **matricialisation**).

Autrement dit, on considère deux **caractères** $\eta : \Omega \mapsto \mathbf{R}$ et $\xi : \Omega \mapsto \mathbf{R}_+$, et ceux-ci sont supposés en relation de proportionnalité et d'**hétéroscédasticité** selon (1).

(ii) Pour estimer le total $T = e_M'\,Y$, la méthode de prédiction bayésienne (cf **prédicteur bayésien**) conduit à estimer b_1 à l'aide du modèle :

(2) $\quad E\,y = b_1 \,.\, x, \qquad$ avec $V\,y = \sigma^2 \,.\, x^{\wedge}$,

où y (resp x) est l'**échantillon** des N observations relatives au caractère η (resp ξ). On note $b_{1,g}^{\wedge}$ l'**estimateur des moindres carrés généralisés** ainsi obtenu.

On appelle **estimateur par quotient** (W.G. COCHRAN) de T l'estimateur :

(3) $\quad T_N' = (\,\bar{y} / \bar{x}\,) \,.\, e_M'\,X$

(cf **quotient**, **quotient de variables aléatoires**).

L'estimateur par quotient de la **moyenne** théorique $\bar{Y}_M = T / M$ s'en déduit selon :

(4) $\quad T_N'' = T_N' / M = (\,\bar{y} / \bar{x}\,) \,.\, \bar{X}_M$,

où $\bar{y}_N = e_N'\,y / N$ (resp $\bar{x}_N = e_N'\,x / N$) est la **moyenne empirique** de η (resp de ξ) calculée sur l'échantillon y (resp x), et où $\bar{X}_M$ est la moyenne théorique (supposée connue) du caractère ξ.

(iii) Le nom de la méthode vient de ce que, pour estimer T (ou $\bar{Y}_M$), on utilise l'hypothèse de proportionnalité (1) des caractères η et ξ dans la population Ω. Ceci revient à estimer le quotient théorique $\bar{Y}_M / \bar{X}_M$ par le quotient empirique $\bar{y}_N / \bar{x}_N$.

(iv) Si l'échantillon des unités A est tiré dans Ω selon un plan sans remise et équiprobable Π (cf **plan de sondage**, **sondage sans remise**), on montre que T_N' (resp T_N'') est un estimateur biaisé de T (resp de $\bar{Y}_M$), ie que :

(5) $\quad E_\Pi\,(r - R) = -\,(M\,N)^{-1}\,(M - N)\,(1 / \bar{X}_M^2)\,(S_{XY} - R \,.\, S_{XX}) + O\,(1 / N)$,

avec $r = \bar{y}_N / \bar{x}_N$, $R = \bar{Y}_M / \bar{X}_M$, ainsi que les **moments corrigés** suivants :

(6)

$$S_{XY} = (M - 1)^{-1} . \Sigma_{m=1}^{M} (X_m - \bar{X}_M) (Y_m - \bar{Y}_M) \quad \text{(\textbf{covariance} théorique),}$$

$$S_{XX} = (M - 1)^{-1} . \Sigma_{m=1}^{M} (X_m - \bar{X}_M)^2 \quad \text{(\textbf{variance} théorique).}$$

Asymptotiquement, ce biais est nul.

Sous les mêmes hypothèses et notations, on montre que :

(7) $$V_{\Pi} (r - R) = (M N)^{-1} (M - N) (1 / \bar{X}_M^2) \{S_{YY} + R^2 S_{XX} - 2 R S_{XY}\} + O (1 / N^2).$$

On en déduit la variance de T_N' (resp de T_N''). Celle-ci est alors estimée en remplaçant, dans (7), chaque **caractéristique** théorique ($\bar{X}_M$, S_{XX} , ...) par son analogue empirique ($\bar{x}_N$, s_{xx} , ...).