

## QUOTIENT DE VARIABLES ALÉATOIRES (C2, M4)

(17 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Notion simple, qui intervient fréquemment en **Statistique**, notamment entre des **va** positives : pourcentages, règles de trois, etc (cf **estimateur par règle de trois**).

(i) Soit  $(\Omega, \mathcal{F}, P)$  un **espace probabilisé** et  $\zeta = (\xi, \eta) : \Omega \mapsto \mathbf{R}^2$  un **couple aléatoire** réel dont la **loi**  $P^\zeta = P^{(\xi, \eta)}$  est une **loi absolument continue** pr à  $\lambda_2$  et possède une **densité** notée  $f$ .

On appelle **quotient**, ou **rapport**, de  $\eta$  pr à  $\xi$  la va  $\rho$  définie selon :

$$(1) \quad \rho = \begin{cases} \eta / \xi & \text{si } \xi \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

(ii) Si  $\rho \in L^1_{\mathbf{R}}(\Omega, \mathcal{F}, P)$ , on a :

$$(2) \quad E \rho = E (\eta / \xi) = (E \xi)^{-1} E \eta - (E \xi)^{-1} C (\eta / \xi, \xi),$$

où  $C(\alpha, \beta)$  dénote la covariance entre deux **va**  $\alpha$  et  $\beta$ .

Autrement dit, si  $\eta$  (resp  $\xi$ ) est un **estimateur sans biais** de  $E \eta$  (resp de  $E \xi$ ), alors  $E(\eta / \xi)$  diffère de  $E \eta / E \xi$  d'un terme qui dépend de  $C(\rho, \xi)$ .

Ce résultat est utilisé eg en théorie des sondages (cf **estimateur par le quotient**).

Lorsque les variables sont dans  $L_{\mathbf{R}}^2$ , on a (**inégalité de CAUCHY-SCHWARZ** conditionnelle) :

$$(1) \quad (C \eta / \xi, \xi) \leq \sigma (\eta / \xi) \cdot \sigma (\xi) \quad (\text{où } \sigma (\cdot) \text{ désigne l'écart-type}).$$

On en déduit que le **biais** de  $\eta / \xi$  pr à  $E \eta / E \xi$  est majoré selon :

$$(3) \quad E (\eta / \xi - E \eta / E \xi) \leq \sigma (\eta / \xi) \cdot \sigma (\xi) / |E \xi|.$$

On montre que la densité  $h$  de la va  $\rho$  s'écrit :

$$(4) \quad h(r) = \int_A y f(r y, y) dy - \int_B y f(r y, y) dy, \quad \forall r \in \mathbf{R},$$

où  $A$  désigne  $\mathbf{R}_+$  et  $B$  désigne  $\mathbf{R}_-$ .

Si  $\xi \geq 0$  et si  $E \xi < \infty$ ,  $h$  peut se calculer selon :

$$(5) \quad h(r) = (2 \pi i)^{-1} \int (d \varphi(u, v) / d u) u = -r v dv, \quad \forall r \in \mathbf{R},$$

où  $\varphi$  désigne la **fonction caractéristique** de la loi du couple  $(\xi, \eta)$ .

(iii) A titre d'exemple, si  $\xi \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2)$  et  $\eta \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$  sont des normales indépendantes, la densité  $h$  pr à  $\lambda_1$  de  $\rho$  est de la forme (cf **méthode de la fonction muette**) :

$$(6) \quad h(r) = a(r) \cdot [b(r) + c(r) \cdot d(r)], \quad \forall r \in \mathbf{R},$$

avec :

$$(6)' \quad \begin{aligned} a(r) &= (2\pi)^{-1} \cdot \exp \{ (1/2) (r^2 + 1)^{-1} (A(r))^2 - B^2 \}, \\ b(r) &= 2 (r^2 + 1)^{-1} \cdot \exp \{ - (1/2) (r^2 + 1)^{-1} (A(r))^2 \}, \\ c(r) &= (2\pi)^{1/2} (r^2 + 1)^{-3/2} \cdot A(r), \\ d(r) &= 2 \cdot \Phi \{ (r^2 + 1)^{-1/2} \cdot A(r) \} - 1, \end{aligned}$$

où  $A(r) = (\mu_\xi / \sigma_\xi) r + (\mu_\eta / \sigma_\eta)$ ,  $B^2 = (\mu_\xi / \sigma_\xi)^2 + (\mu_\eta / \sigma_\eta)^2$  et  $\Phi$  désigne la **fr** de la **loi normale**  $\mathcal{N}(0,1)$ .

On peut remplacer les rapports  $\mu / \sigma$  précédents par les inverses des **coefficients de variation** propres à  $\eta$  et à  $\xi$ .

(iv) Si  $\xi$  est à valeurs dans  $\mathbf{R}_+^*$  (ie si  $P(\xi > 0) = 1$ ), si  $r = \eta \xi^{-1} \in L_{\mathbf{R}}^p$  et s'il existe  $\alpha > 0$  tq  $(\xi, \eta)$  admette une **fonction génératrice**  $g(u, v) = E e^{u\xi + v\eta}$ ,  $\forall v$  tq  $|v| < \alpha$  et  $\forall u$  tq  $-\infty < u < \alpha$ , alors les **moments** (algébriques) de  $\rho$  s'écrivent :

$$(7) \quad E r_j = \{ (j-1)! \}^{-1} \int t^{j-1} (j / d v j g(-t, v)) v = 0 dt, \quad \forall j \in N_p^*.$$

(v) Si  $\zeta$  prend ses valeurs (ie si  $P^\zeta$  a son **support**) dans  $\mathbf{R}_+^2$  et si  $\zeta$  est de carré intégrable, avec pour espérance  $E \zeta$  et pour variance  $V\zeta$ , et si  $|\rho_{\xi\eta}| \leq 1$  (**coefficient de corrélation linéaire** entre  $\xi$  et  $\eta$ ), on peut estimer le quotient (nombre certain) :

$$(9) \quad \kappa = \mu_\eta / \mu_\xi$$

à l'aide de l'**estimateur par le quotient**, ou **estimateur du quotient** :

$$(10) \quad K_N = \bar{Y}_N / \bar{X}_N,$$

avec  $\bar{X}_N = N^{-1} \sum_{n=1}^N X_n$  et  $\bar{Y}_N = N^{-1} \sum_{n=1}^N Y_n$ , où  $Z = ((X_1, Y_1), \dots, (X_N, Y_N))$  désigne un **échantillon iid** issu de la **variable parente**  $\zeta = (\xi, \eta)$ .

On établit que  $K_N$  est un estimateur (non paramétrique), en général biaisé, avec :

$$(11) \quad E K_N = (\mu_\eta / \mu_\xi) + (\sigma_\xi^2 / N) (\mu_\eta / \mu_\xi^3) - (N^{-1} \rho_{\xi\eta}) \mu_\xi^{-2} (\sigma_\xi \sigma_\eta) + O(N^{-2}).$$

(vi) La **méthode de QUENOUILLE** permet de réduire des biais de cette nature.

(vii) En **théorie des sondages**, la méthode d'**estimation par le quotient** constitue une façon simple d'intégrer une information a priori.