REVUE DE STATISTIQUE APPLIQUÉE

M. BASTIEN

# Loi du rapport de deux variables normales

*Revue de statistique appliquée*, tome 8, nº 1 (1960), p. 45-50

<http://www.numdam.org/item?id=RSA_1960__8_1_45_0>

# LOI DU RAPPORT DE DEUX VARIABLES NORMALES

M. BASTIEN

Ingénieur au Service Statistique de l'Institut de Recherches de la Sidérurgie

*La fonction de répartition du rapport de 2 variables normales ne peut être explicitée au moyen de fonctions simples ou tabulées.*

*R.C. Geary (1930) en a donné une bonne approximation. Celle-ci semble peu connue et suffit à montrer que les conditions permettant d'approcher la loi de ce rapport par une loi normale sont plus sévères qu'il n'est généralement admis.*

*L'étude de la loi du rapport de 2 variables normales n'a guère qu'un intérêt théorique puisqu'en pratique les 2 variables ne seront que sensiblement normales et que l'on ignorera généralement la valeur des paramètres dont dépend cette loi.*

Soient x et y deux variables normales telles que :

$$\begin{cases} E(x) = \mu_1 \\ E(y) = \mu_2 \end{cases} \qquad \begin{cases} v(x) = \sigma_1^2 \\ v(y) = \sigma_2^2 \end{cases} \qquad \begin{cases} Cov\,(x,y) = \rho\sigma_1\sigma_2 \\ \rho^2 < 1 \end{cases}$$

On se propose d'étudier la valeur de l'approximation de la loi de $\frac{y}{x}$ par la loi normale.

On étudie, pour simplifier l'écriture, la loi de $a = \left(\dfrac{\frac{y}{\sigma_2}}{\frac{x}{\sigma_1}}\right)$, en ayant donc :

$$\frac{y}{x} = ka \qquad \text{où} \qquad k = \frac{\sigma_2}{\sigma_1}$$

I - <u>Si $\mu_1 = \mu_2 = 0$</u>, on montre que la variable $\dfrac{a - \rho}{\sqrt{1 - \rho^2}}$ a une distribution de Cauchy réduite.

II - <u>Si $\mu_1 \neq 0$</u>,

- en utilisant les paramètres suivants :

---------------

Référence.

GEARY (R.C.) (1930) - "The frequency distribution of the quotient of two normal variables". Jour. Roy. Statist. Soc. <u>93</u>, 442.

Nous tenons à la disposition du lecteur une annexe justifiant les résultats qui ne se trouvent pas dans l'article ci-dessus de R.C. Geary.

$$\begin{cases} e_1 = \mu_1/\sigma_1 \\ e_2 = \mu_2/\sigma_2 \\ \alpha = e_2/e_1 \quad (k\alpha = \mu_2/\mu_1) \\ S = \sqrt{\dfrac{e_1^2 - 2\rho e_1 e_2 + e_2^2}{1-\rho^2}} \end{cases}$$

(les inégalités $S^2 \geqslant e_1^2$ et $S^2 \geqslant e_2^2$ sont toujours satisfaites).

- et en posant :

$$u = \frac{|e_1|(a-\alpha)}{\sqrt{1 - 2a\rho + a^2}} \qquad (1)$$

on montre que la densité de a, f(a), se déduit des relations suivantes :

$$(*) \begin{cases} f(a)da = \pm \{ g(u) + \varepsilon(u) \} \, du, \\ du = |e_1| \dfrac{\{a(\alpha - \rho) + 1 - \rho\alpha\}}{\{1 - 2a\rho + a^2\}^{3/2}} da \\ \varepsilon(u) = \dfrac{e^{-\frac{S^2}{2}}}{\pi} \left\{ \dfrac{1}{\sqrt{S^2 - u^2}} - \dfrac{G(\sqrt{S^2 - u^2})}{g(\sqrt{S^2 - u^2})} \right\} \end{cases}$$

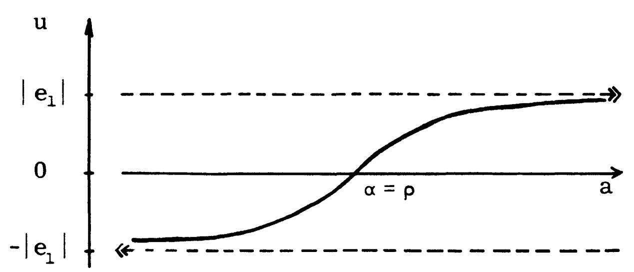
le signe à prendre étant celui de $\frac{du}{da}$.

- On distingue les 3 cas suivants, selon l'allure de la variation de u en fonction de a, pour montrer comment on peut, dans chacun des cas, encadrer assez finement la fonction de répartition de a, F(a). Les inégalités assez simples adjointes aux croquis ci-dessous, sont extraites d'inégalités plus étroites.

(1) $\underline{\alpha - \rho = 0}$

c'est-à-dire

$\boxed{-1 < \alpha = \rho < 1}$




loi de a symétrique autour de la médiane $\alpha$

Pour $a < \alpha$ : $G(u) - G(-|e_1|) \leqslant F(a) < G(u)$

Pour $a > \alpha$ : $G(u) < F(a) \leqslant G(u) + G(-|e_1|)$

---

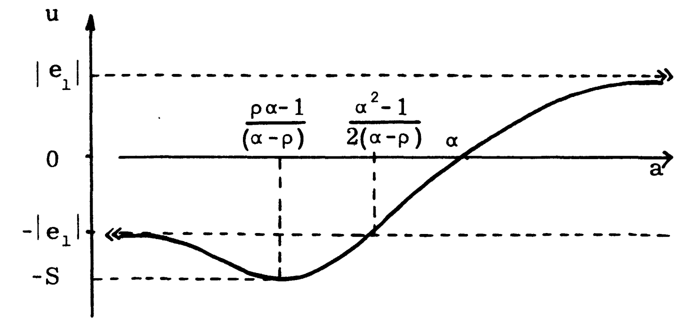
(*) Par analogie avec la forme mathématique de la densité et celle de la fonction de répartition de la variable normale centrée réduite, on conviendra de noter g(u) et G(u) les deux fonctions suivantes :

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad ; \quad G(u) = \int_{-\infty}^{u} g(t)\, dt$$

(2) $\underline{\alpha - \rho > 0}$

c'est-à-dire

$\boxed{-1 < \rho < \alpha}$



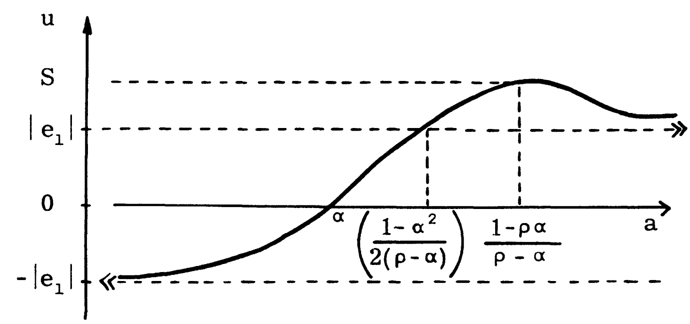

Pour $a \leqslant \frac{\rho\,\alpha - 1}{\alpha - \rho}$ : $F(a) < G(-|e_1|)$

Pour $a > \frac{\rho\,\alpha - 1}{\alpha - \rho}$ : $G(u) + G(-|e_1|) - 2G(-S) < F(a) \leqslant G(u) + G(-|e_1|)$

(3) $\underline{\alpha - \rho < 0}$

c'est-à-dire

$\boxed{\alpha < \rho < 1}$




Pour $a < \frac{1 - \rho\,\alpha}{\rho - \alpha}$ : $G(u) - G(-|e_1|) \leqslant F(a) < G(u) - G(-|e_1|) + 2G(-S)$

Pour $a \geqslant \frac{1 - \rho\,\alpha}{\rho - \alpha}$ : $F(a) > 1 - G(-|e_1|)$.

- D'une façon plus grossière, mais encore plus simple et valable pour les 3 cas et pour toute valeur de a.

En posant : $$u = \frac{|e_1|\,(a - \alpha)}{\sqrt{1 - 2a\,\rho + a^2}} \qquad (1)$$

On a : $$\begin{cases} 0 \leqslant F(a) \leqslant 1 & (2) \\ G(u) - G(-|e_1|) \leqslant F(a) \leqslant G(u) + G(-|e_1|) & (3) \end{cases}$$

Ces inégalités sont pratiquement amplement suffisantes lorsque $|e_1|$ est assez grand :

------------

(1),(2) et (3) sont les résultats de R.C. Geary.

Ainsi, par exemple, F(a) égale G(u) à moins de $10^{-3}$ par excès ou par défaut lorsque $|e_1| > 3,09$ puisque $G(-3,09) = 10^{-3}$.

Dans ces conditions, la non "normalité approchée" de la loi de a provient essentiellement de la non linéarité de a fonction de u. (Cf. croquis p. ).

Toutefois, lorsque $\left|\frac{u}{e_1}\right|$ est assez petit et d'autant mieux que $\frac{(\alpha - \rho)^2}{1 - \rho^2 + (\alpha - \rho)^2}$ est voisin de zéro (Cf. N. B. p. ) :

$$a = \alpha + \frac{u}{e_1^2 - u^2}\left\{\sqrt{(1-\rho^2)(S^2 - u^2)} + (\alpha - \rho)\,u\right\},$$

solution de (1) pour $u^2 < e_1^2$, devient :

$$a \sim \alpha + \frac{S\sqrt{1-\rho^2}}{e_1^2}\,u = \alpha + u\sqrt{\alpha^2\left(\frac{1}{e_1^2} - \frac{2\rho}{e_1 e_2} + \frac{1}{e_2^2}\right)} \qquad (4)$$

c'est-à-dire :

$$\left\{\begin{array}{l} \frac{y}{x} = ka \sim \frac{\mu_2}{\mu_1} + u\sqrt{\left(\frac{\mu_2}{\mu_1}\right)^2\left(\frac{\sigma_1^2}{\mu_1^2} - 2\rho\frac{\sigma_1\sigma_2}{\mu_1\mu_2} + \frac{\sigma_2^2}{\mu_2^2}\right)} \\ \text{où } u \text{ est sensiblement la variable normale centrée réduite puis-} \\ \text{que } F(a) \approx G(u). \end{array}\right.$$

On retrouve ainsi l'approximation normale classique de la loi de $\frac{y}{x}$, obtenue à l'aide du développement de :

$$\frac{y}{x} = \frac{\mu_2}{\mu_1}\left(1 + \frac{y - \mu_2}{\mu_2}\right)\left(1 + \frac{x - \mu_1}{\mu_1}\right)^{-1}$$

Cette approximation classique n'est donc valable que dans la mesure où l'approximation linéaire (4) l'est aussi, jusqu'à $|u|$ voisin de 2 ou 3 par exemple.

$|e_1|$ devra donc être largement supérieur à 3 et d'autant plus largement que $\frac{|\alpha - \rho|}{\sqrt{\alpha^2 - 2\alpha\rho + 1}}$ est différent de zéro.

Le graphique ci-dessous montre à droite la variation de $\left|\frac{u}{e_1}\right|$ en fonction de a, lorsque $\rho = 0$, pour les 3 valeurs $\alpha = 0$ ; $\alpha = 0,5$ ; $\alpha = 1$.

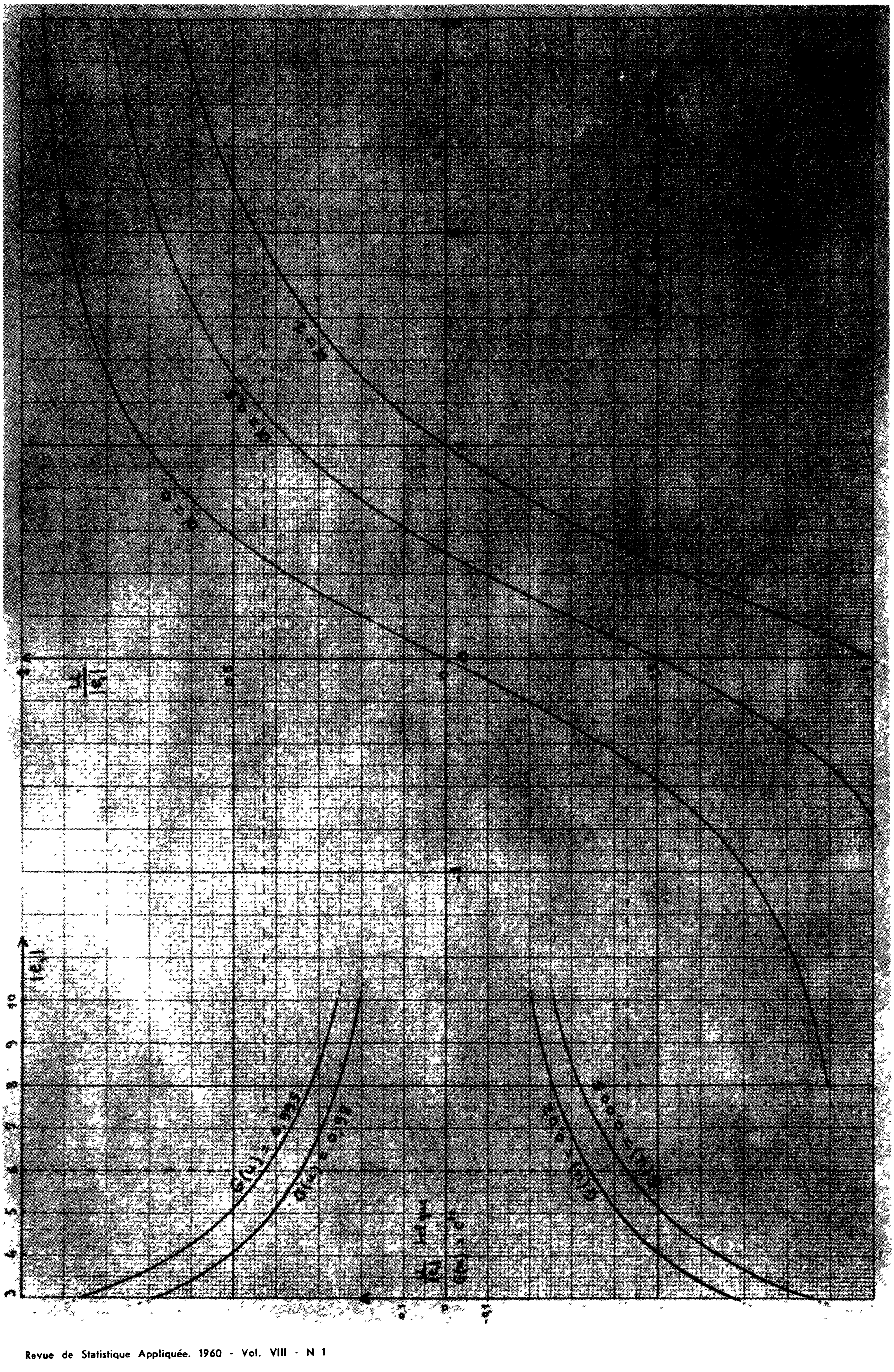
Il représente donc sensiblement (c'est-à-dire dans la mesure où $F(a) \sim G(u)$) la fonction de répartition F(a) tracée sur papier gaussien.

L'échelle des probabilités totales, en ordonnée, est fixée par G(u). Elle est donc fonction de $|e_1|$ puisque c'est $\frac{u}{|e_1|}$ au lieu de u qui est représenté.

Le point 50% de cette échelle est cependant indépendant de $|e_1|$ puisqu'il correspond à $u = 0$. <u>La médiane de a est sensiblement $\alpha$.</u>

Les courbes de gauche donnent, en fonction de $|e_1|$, les valeurs de $\frac{u}{|e_1|}$ telles que :

$$F(a) \sim G(u) = \left\{\begin{array}{l} 0,005 \\ 0,02 \\ 0,98 \\ 0,995 \end{array}\right.$$



3
4
5
6
7
8
9
10

L'ensemble du graphique permet de juger rapidement en fonction de $|e_1|$ (lorsque $\rho = 0$ et pour l'une des 3 valeurs choisies pour $\alpha$) la valeur de l'approximation normale de la loi de $\frac{y}{x}$, c'est-à-dire la linéarité de la variation de a en fonction de u dans l'intervalle, (0,005 ; 0,995) par exemple, des probabilités totales.

Ainsi, pour $|e_1| = 6$, où F(a) égale G(u) à moins de $10^{-9}$ près par excès ou par défaut :

- les courbes de gauche indiquent que $F(a) \sim G(u)$ prend les deux valeurs 0,995 et 0,005 pour $\frac{u}{|e_1|} = \pm 0,43$ ;

- les courbes de droite permettent, selon la valeur de $\alpha$ de transformer ces deux valeurs de $\frac{u}{|e_1|}$ en valeurs associées de a, c'est-à-dire de déterminer les deux quantiles $a_{0,995}$ et $a_{0,005}$ correspondant aux probabilités 0,995 et 0,005.

On voit que :

- pour $\alpha = 0$, la variation de a en fonction de u est sensiblement linéaire entre ces deux quantiles $a = \pm 0,48$. On peut accepter l'approximation normale de la loi de $\frac{y}{x}$ dans l'intervalle (0,005 ; 0,995) des probabilités totales.

- il n'en est plus de même par contre pour $\alpha = 1$, où la dissymétrie de la courbe est déjà prononcée entre les deux quantiles $a_{0,995} = 1,94$ et $a_{0,005} = 0,52$.

NOTA BENE -

- Pour $u^2 > e_1^2$, l'équation (1) a les deux solutions :

$$a = \alpha + \frac{u}{u^2 - e_1^2} \left\{ (\rho - \alpha) u \pm \sqrt{(1 - \rho^2)(S^2 - u^2)} \right\}$$

- Pour $u^2 < e_1^2$, l'équation (1) a la seule solution :

$$a = \alpha + \frac{u}{e_1^2 - u^2} \left\{ (\alpha - \rho) u + \sqrt{(1 - \rho^2)(S^2 - u^2)} \right\}$$

Tenant compte du fait que :

$$S^2 \geqslant e_1^2$$

$$\frac{\alpha - \rho}{S\sqrt{1 - \rho^2}} \leqslant \frac{1}{|e_1|} \quad \text{car} \quad \frac{(\alpha - \rho)^2}{S^2(1 - \rho)^2} = \frac{(\alpha - \rho)^2}{e_1^2\{1 - \rho^2 + (\alpha - \rho)^2\}} \leqslant \frac{1}{e_1^2}$$

on peut développer a autour de u = 0 :

$$a = \alpha + \frac{S\sqrt{1 - \rho^2}}{e_1^2} u \left\{ 1 + \frac{\alpha - \rho}{S\sqrt{1 - \rho^2}} u + \left(\frac{1}{e_1^2} - \frac{1}{2S^2}\right) u^2 + \dots \right\}$$

Ainsi, à valeur $|e_1|$ fixée, la linéarité de a autour de $a = \alpha$ sera d'autant meilleure que $\frac{|e_1|\,|\alpha - \rho|}{S\sqrt{1 - \rho^2}}$ sera plus voisin de zéro, c'est-à-dire que $\frac{(\alpha - \rho)^2}{\alpha^2 - 2\alpha\rho + 1}$ sera plus voisin de zéro.