Sur la répartition du rapport de deux variables aléatoires et sa simulation par Monte Carlo

Il est bien connu qu'une grande partie des mesures effectuées consiste, à vrai dire, à déterminer un rapport entre deux quantités. Si les deux grandeurs en question se déterminent indépendamment l'une de l'autre, ce cas ne pose aucun problème sérieux pour la détermination de l'écart-type du résultat. Cependant, pour les mesures qui demandent l'accumulation de données pendant un temps parfois assez long (p. ex. comptage d'impulsions), il peut être impossible de dépasser un certain temps de mesure puisque - en présence d'influences perturbatrices - la précision acquise risquerait facilement d'être trompeuse. Dans cette situation on est amené à se contenter de courts intervalles de mesure, où l'on détermine alternativement les deux quantités. En raison de dérivées possibles cela a souvent peu de sens d'arriver à des valeurs moyennes qui pourraient d'abord paraître indispensables pour trouver le rapport à déterminer. Cependant, il est toujours justifié de former des rapports consécutifs des résultats de deux mesures différentes. En général, ces rapports auront une précision assez basse, mais on pourra s'attendre à ce qu'ils ne soient pas biaisés. D'autre part, une interprétation soigneuse de telles données demande évidemment quelques connaissances antérieures concernant la répartition de ces rapports. Malheureusement, les traités traditionnels sur la propagation des erreurs ne s'occupent que d'opérations linéaires appliquées aux distributions statistiques - et pour cause, car c'est seulement pour celles-ci que l'opération de convolution résoud tous les problèmes de manière si élégante et si simple. Pour toutes les autres opérations, cependant, cette méthode de calcul ne donne guère que des résultats approximatifs.

Quant au cas du rapport de deux distributions, le seul qui nous préoccupe dans ce qui suit, nous partons d'un résultat donné par Feller [1]. Celui-ci indique, après une légère généralisation, que la densité de probabilité du rapport $V = x_1 / x_2$ est donnée par

$$P(V) = \int_{-\infty}^{\infty} |x_2| \cdot f_1(V x_2) \cdot f_2(x_2) \, dx_2 , \qquad (1)$$

où $f_1(x_1)$ et $f_2(x_2)$ sont les densités des quantités x_1 et x_2, respectivement.

Pour arriver à des résultats qui sont plus proches des applications pratiques, nous nous limiterons au cas spécial où x_1 et x_2 suivent une distribution normale

$$f_{1,2}(x_{1,2}) = \frac{1}{s_{1,2}\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x - m_{1,2}}{s_{1,2}}\right)^2} \qquad (2)$$

$m_{1,2}$ représente donc les espérances mathématiques et $s_{1,2}^2$ les variances de x_1 et x_2, respectivement. Pour $P(V)$ on obtient maintenant

$$P(V) = \frac{1}{2\pi s_1 s_2} \int_{-\infty}^{\infty} |x_2| \cdot e^{-\frac{1}{2}\left\{\left(\frac{Vx_2 - m_1}{s_1}\right)^2 + \left(\frac{x_2 - m_2}{s_2}\right)^2\right\}} .$$

Si nous posons $m_1/V = \mu_1$, $s_1/V = \sigma_1$

ainsi que $m_2 = \mu_2$, $s_2 = \sigma_2$ et $x_2 = x$,

l'exposant prend la forme simplifiée

$$\{\ldots\} = \left(\frac{x - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x - \mu_2}{\sigma_2}\right)^2 = (Ax - B)^2 + C$$

avec

$$A = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} ,$$

$$B = \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1 \sigma_2 \sqrt{\sigma_1^2 + \sigma_2^2}} ,$$

$$C = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} .$$

Avec ces abréviations, l'intégrale qui en résulte, c'est-à-dire

$$P(V) = \frac{1}{2\pi s_1 s_2} \int_{-\infty}^{\infty} |x| \cdot e^{-\frac{1}{2}\left\{(Ax - B)^2 + C\right\}} dx ,$$

n'est pas trop difficile à évaluer et nous donne le résultat final

$$P(V) = \frac{s_1 s_2}{\sqrt{2\pi}} \cdot \frac{e^{-C/2}}{s_1^2 + (Vs_2)^2} \cdot \left[B\,\Phi(B) + 2\,\varphi(B) \right] \tag{3}$$

avec $$B = \frac{m_1 V s_2^2 + m_2 s_1^2}{s_1 s_2 \sqrt{s_1^2 + (Vs_2)^2}}$$ et

$$C = \frac{(m_1 - Vm_2)^2}{s_1^2 + (Vs_2)^2} .$$

φ et Φ représentent la fonction d'erreur et son intégrale, c'est-à-dire

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} , \qquad \Phi(x) = \int_{-x}^{x} \varphi(t)\, dt ,$$

que l'on trouve bien tabulés pour cette normalisation p. ex. dans [2].

Il découle de la formule (3) que la répartition du rapport V pour deux variables normales correspond à la distribution de Cauchy (du type $1/(1+x^2)$) si $B = 0$ et $C = 0$, c'est-à-dire pour $m_1 = m_2 = 0$. Ce simple cas spécial, d'ailleurs, est déjà connu depuis longtemps (voir p.ex. [3]). D'autre part, la loi de Cauchy (ou courbe de résonance) est bien connue pour ses caractéristiques pathologiques, car elle ne possède aucun moment à part celui de l'ordre zéro qui ne détermine que la normalisation.

Pour vérifier ce comportement et examiner la répartition pour le cas d'un intérêt plus général où les valeurs moyennes de x_1 et de x_2 ne disparaissent plus, une étude directe à partir de (3) semble assez difficile. Nous avons donc préféré profiter de cette occasion pour nous familiariser avec la méthode dite de Monte Carlo, en simulant expérimentalement la répartition des rapports et en choisissant au hasard des valeurs pour x_1 et x_2, compte tenu de leurs densités (2). Ainsi il devrait être possible de vérifier (ou non) le facteur

$$F = \sqrt{\frac{\pi}{2}} \cdot \left[B \cdot \Phi (B) + 2 \varphi (B) \right] \cdot e^{-C/2} \tag{4}$$

qui distingue, d'après (3), la répartition générale du cas particulier avec $m_1 = m_2 = 0$ et nous permet d'écrire aussi

$$P(V) = \frac{s_1 s_2}{\pi} \cdot \frac{F}{s_1^2 + (V s_2)^2} \quad . \tag{3'}$$

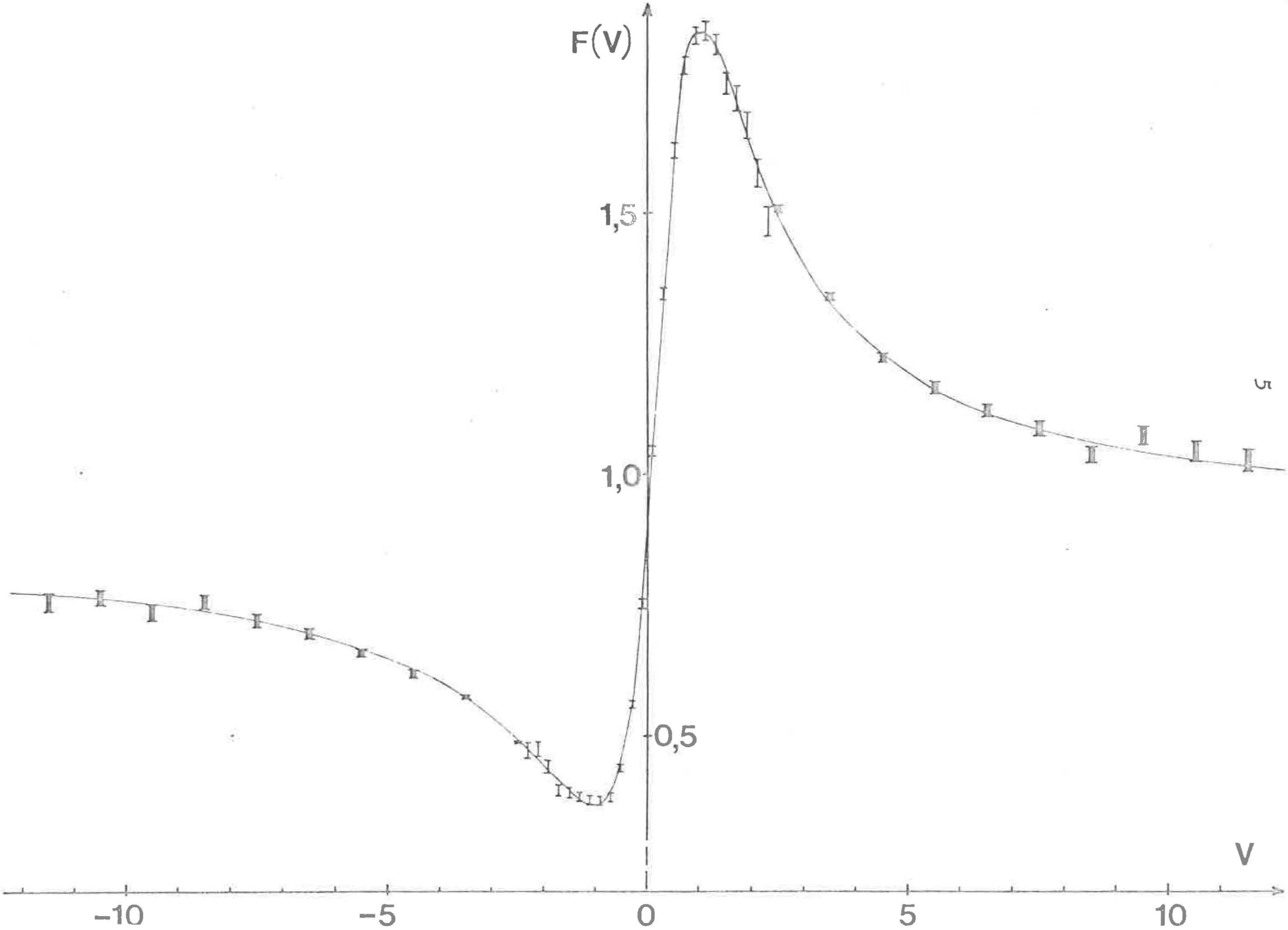
Dans ce qui suit, nous nous inspirons largement de l'excellent petit traité [4] sur la technique de Monte Carlo. Comme toujours dans ces cas, le problème mathématique se divise en deux parties. La première consiste à produire une série de nombres pseudo-aléatoires distribués uniformément entre 0 et 1. A cet effet, nous nous servons d'une relation à recurrence de la forme [5]

$$z_i = (A\, z_{i-1} + B) \quad \text{modulo } M,$$

où $A = 1$ modulo 4 et B est un nombre impair.

La série $R_i = z_i/M$ donne alors des nombres pseudo-aléatoires entre 0 et 1 avec une période M. Comme nombre entier, M était fixé pour notre ordinateur à $2^{15} \cong 33\,000$. Pour la plupart des applications de ce genre, cependant, cette période serait beaucoup trop courte. Mr. Carré a aimablement mis à notre disposition un sous-programme spécial écrit en langage assembleur, où deux mots sont employés pour un nombre entier. Cela nous permet d'arriver d'un seul coup à une période de $M = 2^{30} \cong 10^9$.

La deuxième partie consiste maintenant à transformer les R_i de telle façon que la nouvelle quantité se distribue conformément à la densité voulue, dans notre cas donc d'après une distribution de Gauss. Au lieu de nous servir de procédés approximatifs qui se basent sur le théorème central limite, nous préférons une méthode directe [6]. A partir d'une paire de valeurs R_i on obtient une paire de valeurs G_i qui représentent un échantillon d'une population normale, en formant


F(V)
1,5
1,0
0,5
V
-10
-5
0
5
10

$$G_1 = \sqrt{-2\cdot\ln R_1}\cdot\sin(2\pi\cdot R_2)$$
$$G_2 = \sqrt{-2\cdot\ln R_1}\cdot\cos(2\pi\cdot R_2), \qquad (6)$$

et qui seraient indépendantes. Mais puisque nous voulons former ensuite des rapports du type $V = G_1/G_2$, il nous paraît préférable de changer l'ordre des valeurs R_i, car autrement on obtiendrait simplement $V = \mathrm{tg}\,(2\pi\cdot R_2)$, où R_1 serait sans aucune influence. Les premiers calculs faits avec le choix

$$G_1 = \sqrt{-2\cdot\ln R_1}\cdot\sin(2\pi\cdot R_3)$$
$$G_2 = \sqrt{-2\cdot\ln R_4}\cdot\cos(2\pi\cdot R_2),$$

montrent après 10^5 essais une fidèle reproduction de la distribution de Cauchy pour $m_1 = m_2 = 0$ et servent ainsi de contrôle valable. Des tentatives pour déterminer empiriquement le facteur F dans l'expression générale sont actuellement en cours et semblent bien confirmer les prévisions théoriques (4).

Références

[1] W. Feller: "An Introduction to Probability Theory and its Applications, Volume II" (Wiley, New York; 1966), p. 48

[2] "Tables of Normal Probability Functions" AMS 21 (NBS, Washington; 1953)

[3] Ch. Eisenhart and M. Zelen: "Elements of Probability", in "Handbook of Physics" (ed. Condon and Odishaw), (McGraw-Hill, New York; 1958)

[4] J.M. Hammersley and D.C. Handscomb: "Monte Carlo Methods" (Methuen, London; 1964), (traduction française chez Dunod)

[5] M. Greenberger: "Notes on a new pseudorandom number generator", J. Assoc. Comp. Mach. 8, 163 (1961)

[6] G.E.P. Box and M.E. Muller: "A note on the generation of random normal deviates", Ann. Math. Stat. 29, 610 (1958)

Juin 1968 (J.W.M.)