J. HAAG

Sur un problème de probabilité

Nouvelles annales de mathématiques 4e série, tome 19 (1919), p. 58-69

<http://www.numdam.org/item?id=NAM_1919_4_19__58_1>

[J2e]

SUR UN PROBLÈME DE PROBABILITÉ;

PAR M. J. HAAG.

1. Soient deux quantités indépendantes x et y, obéissant à la loi de Gauss, la valeur moyenne de chacune d'elles étant zéro. On sait que leur somme obéit aussi à la loi de Gauss, avec un écart unitaire se déduisant par combinaison quadratique des écarts unitaires de x et de y. C'est là une propriété dont on fait

fréquemment usage dans le calcul des probabilités et particulièrement dans la théorie des erreurs.

On peut se demander s'il existe une propriété analogue pour le produit xy et, si cette propriété n'existe pas, à quelle loi de probabilité obéit ce produit. Indépendamment de son intérêt pratique, cette question conduit, comme on va le voir, à des développements mathématiques intéressants et constitue un bel exemple d'application numérique de diverses théories d'analyse.

2. Appelons a et b les écarts unitaires de x et de y et cherchons :

1° *La probabilité* dP *pour que le produit* $z = xy$ *soit compris, en valeur absolue, entre* z *et* $z + dz$;
2° *La probabilité* P *pour que cette valeur absolue soit inférieure à* z.

Si l'on considère le point M(x, y), cela revient à calculer la probabilité entre les deux hyperboles voisines

$$(1) \qquad xy = z \qquad \text{et} \qquad xy = z + dz,$$

ainsi que la probabilité dans la région du plan extérieure à la première de ces courbes.

Faisons le changement de variables

$$(2) \qquad x = \omega a \cos\varphi, \qquad y = \omega b \sin\varphi.$$

Il faut calculer, dans chacune des régions ci-dessus, l'intégrale double

$$\frac{2}{\pi}\int\int e^{-\omega^2}\omega\, d\omega\, d\varphi,$$

Ce calcul est élémentaire et conduit, en tenant compte des symétries, aux deux formules suivantes :

$$dP = \frac{2\,dz}{\pi\,ab}\int_0^{\frac{\pi}{2}} e^{-\frac{z}{ab\sin\varphi\cos\varphi}}\frac{d\varphi}{\sin\varphi\cos\varphi},$$

$$P = 1 - \frac{2}{\pi}\int_0^{\frac{\pi}{2}} e^{-\frac{z}{ab\sin\varphi\cos\varphi}}\,d\varphi.$$

Posons

(3) $$\frac{2z}{ab} = t, \qquad 2\varphi = \psi:$$

il vient

(4) $$dP = \frac{2}{\pi}Q\,dt, \qquad Q = \int_0^{\frac{\pi}{2}} e^{-\frac{t}{\sin\psi}}\frac{d\psi}{\sin\psi};$$

(5) $$P = 1 - \frac{2}{\pi}\int_0^{\frac{\pi}{2}} e^{-\frac{t}{\sin\psi}}\,d\psi.$$

La formule (4) se déduit d'ailleurs de la formule (5), par différentiation sous le signe $\int$.

3. Les intégrales (4) et (5) ne se ramènent pas aux fonctions élémentaires. On pourrait les calculer, pour différentes valeurs de t, par la methode des trapèzes. Ce procédé est surtout avantageux pour P, pour les grandes valeurs de t, car l'exponentielle est, dans ce cas, très petite et l'intégrale (5) peut être evaluee, avec une precision suffisante, au moyen de quelques trapèzes seulement. Ceci correspond, du reste, au fait que P tend vers 1, pour $t = \infty$, comme il est evident *a priori*.

Pour les petites valeurs de t, la methode des trapèzes paraît longue et peu pratique. Nous allons en indiquer une autre, qui sera beaucoup plus propice aux calculs numériques.

4. Cherchons, si possible, une équation différentielle simple, à laquelle satisfasse la fonction Q de t.

Remarquons d'abord que cette fonction n'a de sens que si t est positif. Pour $t \leqq 0$, elle est infinie. Il est donc entendu que nous ferons varier t dans l'intervalle $(t_0, +\infty)$, où t_0 désigne un nombre positif quelconque, arbitrairement petit. Dans un tel intervalle, on peut différentier Q sous le signe $\int$ autant qu'on veut.

Nous avons

$$Q' = -\int_0^{\frac{\pi}{2}} e^{-\frac{t}{\sin\psi}} \frac{d\psi}{\sin^2\psi}$$

$$= \left[e^{-\frac{t}{\sin\psi}} \cot\psi\right]_0^{\frac{\pi}{2}} - t\int_0^{\frac{\pi}{2}} e^{-\frac{t}{\sin\psi}} \frac{\cos^2\psi\, d\psi}{\sin^3\psi};$$

$$Q' = -t\int_0^{\frac{\pi}{2}} e^{-\frac{t}{\sin\psi}} \frac{d\psi}{\sin^3\psi} + tQ;$$

$$Q = \int_0^{\frac{\pi}{2}} e^{-\frac{t}{\sin\psi}} \frac{d\psi}{\sin^3\psi};$$

$$Q' = -tQ'' + tQ.$$

Donc, la fonction Q satisfait à l'équation différentielle linéaire du second ordre

$$tQ'' + Q' - tQ = 0. \tag{6}$$

Au changement près de t en it, on reconnaît un cas particulier de l'*équation de Bessel*. Du reste, on sait que l'intégrale générale de cette équation peut s'exprimer par des intégrales définies, suivant la méthode de Laplace. L'intégrale (4) en est un cas particulier, comme on le voit en faisant le changement de variable

$$\frac{1}{\sin\psi} = u.$$

On trouve, en effet,

$$Q = \int_1^{+\infty} \frac{e^{-tu}}{\sqrt{u^2-1}} du, \tag{7}$$

forme classique des intégrales auxquelles nous venons de faire allusion.

5. L'intégrale générale de l'équation de Bessel peut être développée en série au voisinage du point $t = 0$. Pour le cas particulier de l'équation (6), on sait que le développement est de la forme

$$Q = CJ - C'(J \log t - I), \tag{8}$$

en posant

$$J = \sum_{n=0}^{\infty} \frac{\left(\frac{t^2}{4}\right)^n}{(n!)^2}, \qquad I = \sum_{n=1}^{\infty} \left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right) \frac{\left(\frac{t^2}{4}\right)^n}{(n!)^2}, \tag{9}$$

$$C,\ C' = \text{const.}$$

Ces séries convergent avec une grande rapidité et, par suite, se prêtent bien au calcul numérique. Il est donc intéressant pour nous de les utiliser pour l'intégrale particulière à laquelle nous avons affaire. Et pour cela, il nous suffit de *calculer les valeurs des constantes* C et C' qui conviennent à cette intégrale.

6. On a tout d'abord

$$C' = \lim_{t=0} \frac{Q}{\log t}. \tag{10}$$

Pour calculer cette limite, qui se présente sous la forme $\frac{\infty}{\infty}$, partons de la formule (7), en y faisant le changement de variable

$$v = ut, \tag{11}$$

ce qui donne

$$Q = \int_t^{+\infty} \frac{e^{-v}}{\sqrt{v^2 - t^2}}\, dv. \tag{12}$$

Désignant par t' un nombre positif quelconque, nous pouvons écrire

$$Q = \int_t^{t'} \frac{e^{-v}}{\sqrt{v^2 - t^2}}\, dv + \int_{t'}^{+\infty} \frac{e^{-v}}{\sqrt{v^2 - t^2}}\, dv = Q_1 + Q_2.$$

Laissons fixe t' et faisons tendre t vers zéro. La seconde intégrale, qui est, comme on le voit aisément, uniformément convergente quand t varie dans tout intervalle intérieur à l'intervalle $(-t', +t')$, a pour limite

$$(13) \qquad R_2 = \int_{t'}^{+\infty} \frac{e^{-v}}{v}\, dv.$$

Quant à la première, elle peut s'écrire, d'après le théorème de la moyenne,

$$Q_1 = e^{-t''}\left[\log\left(t' + \sqrt{t'^2 - t^2}\right) - \log t\right],$$

t'' désignant un certain nombre compris entre t et t'. Nous avons alors

$$\frac{Q}{\log t} = -e^{-t''} + \frac{Q_2 - e^{-t''}\log\left(t' + \sqrt{t'^2 - t^2}\right)}{\log t}.$$

Si, laissant t' fixe, nous faisons tendre t vers zéro, le second membre tend vers $-e^{-\theta}$, θ désignant la limite de t''. Or, cette limite est certainement comprise entre 0 et t', et, comme t' peut être choisi à l'avance aussi petit qu'on veut, θ ne peut être que nul. Donc, la limite de $\frac{Q}{\log t}$ est égale à -1 et l'on a

$$(14) \qquad C' = -1.$$

Nous avons maintenant

$$(15) \qquad C = \lim_{t=0}(Q + \log t).$$

Cette limite se présente sous la forme $\infty - \infty$. D'après ce qui précède, elle est égale à

$$R_2 + \log 2t' - \lim_{t=0} (t'' \log t).$$

Mais, nous ne connaissons pas la limite de $t'' \log t$, qui est certainement dépendante de t', puisqu'il en est ainsi de la somme $R_2 + \log 2t'$.

On peut songer à chercher la limite de cette limite, quand t' tend à son tour vers zéro. Mais il est plus simple de reprendre la question de la manière suivante.

Donnons-nous un nombre positif quelconque $k > 1$. Nous pouvons écrire

$$Q_1 = \int_t^{kt} \frac{e^{-v}}{\sqrt{v^2 - t^2}}\,dv + \int_{kt}^{t'} \frac{e^{-v}}{\sqrt{v^2 - t^2}}\,dv = Q_3 + Q_4.$$

Or

$$Q_3 = e^{-t'''} \log(k + \sqrt{k^2 - 1}), \qquad t < t''' < kt;$$

$$\lim_{t=0} Q_3 = \log(k + \sqrt{k^2 - 1}).$$

Puis

$$Q_4 = \int_{kt}^{t'} \frac{e^{-v} - 1}{v}\,dv + \int_{kt}^{t'} \frac{dv}{v} + \int_{kt}^{t'} e^{-v}\left[\frac{1}{\sqrt{v^2 - t^2}} - \frac{1}{v}\right] dv;$$

$$Q_4 + \log t = \log t' - \log k + \int_{kt}^{t'} \frac{e^{-v} - 1}{v}\,dv + \int_{kt}^{t'} e^{-v}\left[\frac{1}{\sqrt{v^2 - t^2}} - \frac{1}{v}\right] dv.$$

Pour $t = 0$, la première intégrale devient

$$R_1 = \int_0^{t'} \frac{e^{-v} - 1}{v}\,dv. \tag{16}$$

La seconde s'écrit

$$e^{-t''}\left[\log\left(1 + \sqrt{1 - \frac{t^2}{t'^2}}\right) - \log\left(1 + \sqrt{1 - \frac{1}{k^2}}\right)\right],$$

t'' étant compris entre kt et t'. Sa limite, pour $t = 0$, est, en appelant θ la limite de t'',

$$e^{-\theta}\left[\log 2 - \log\left(k + \sqrt{k^2 - 1}\right) + \log k\right].$$

Finalement, nous avons

$$(17)\quad \begin{aligned} C = {} & R_2 - \log\left(k + \sqrt{k^2 - 1}\right) + \log t' - \log k + R_1 \\ & + e^{-\theta}\left[\log 2 - \log\left(k + \sqrt{k^2 - 1}\right) + \log k\right], \end{aligned}$$

où θ est un certain nombre, dont nous savons seulement qu'il est compris entre o et t'. La première ligne de (17) est indépendante de t', comme on le voit en prenant la dérivée. Il en est donc de même de la seconde et, par suite de θ. Comme θ est compris entre o et t' et qu'on peut choisir t' arbitrairement petit, on a nécessairement $\theta = 0$. Nous avons finalement

$$(18)\qquad C = R_1 + R_2 + \log 2t'.$$

7. Reste à calculer la valeur numérique de cette constante.

Nous avons d'abord

$$(19)\qquad R_1 = \sum_{n=1}^{\infty} (-1)^n \frac{t'^n}{n.n!}.$$

Cette série converge quel que soit t', mais d'autant plus vite que t' est plus petit. Pour calculer R_2, appliquons *la formule asymptotique de Stieltjes :*

$$(20)\quad R_2 = \frac{e^{-t'}}{t'}\left[1 - \frac{1}{t'} + \frac{2!}{t'^2} - \frac{3!}{t'^3} + \ldots + (-1)^n \frac{n!}{t'^n} + \ldots\right].$$

Cette série est divergente, quel que soit t'; néanmoins, si l'on prend la somme de ses n premiers termes, on sait que l'on obtient une valeur approchée de R_2, l'erreur commise étant inférieure en valeur absolue au dernier terme pris et de signe contraire à ce terme.

Or, si $t' \geqq 2$, la série commence par converger, pour diverger ensuite. Le terme minimum est égal, si t' est entier, à $e^{-t'}\frac{(t'-1)!}{t'^{t'}}$.

On peut le rendre aussi petit qu'on veut, en prenant t' assez grand. Par conséquent, en arrêtant le développement à ce terme minimum, on obtient une valeur aussi approchée qu'on le veut de R_2.

Prenons $t' = 4$. Nous trouvons les deux valeurs approchées par excès et par défaut

$$0,004007 \quad \text{et} \quad 0,00358.$$

En prenant la moyenne, nous obtenons

$$R_2 = 0,00379,$$

à 0,0002 près.

En calculant ensuite R_1 par la formule (19), on trouve, à 0,0001 près,

$$R_1 = -1,96732.$$

Enfin

$$\log 8 = 2,07944.$$

Finalement, nous avons

$$C = +0,1159,$$

à environ 0,0003 près.

8. Nous connaissons maintenant le développement de Q. En intégrant la formule (4) de 0 à t, nous en déduisons

$$P = \frac{2}{\pi}\int_0^t Q\,dt.$$

Or

$$Q = -J\log t + I + CJ.$$

D'où

$$\frac{\pi P}{2} = -\int_0^t J\log t\,dt - \int_0^t I\,dt + C\int_0^t J\,dt.$$

Si l'on pose

$$J_1 = \int_0^t J\,dt,$$

on a

$$\int_0^t J \log t\,dt = J_1 \log t - \int_0^t J_1 \frac{dt}{t}.$$

Donc

$$P = \frac{2}{\pi}\left[J_1 (C - \log t) + \int_0^t \left(1 - \frac{J_1}{t}\right) dt\right].$$

En tenant compte des formules (9) et intégrant terme à terme, on trouve finalement

$$(21) \qquad P = \frac{2t}{\pi}\,[A + B\,(0,1159 - \log t)],$$

en posant

$$(22) \qquad A = \sum_{n=0}^{\infty} \alpha_n u_n, \qquad B = \sum_{n=0}^{\infty} u_n,$$

$$(23) \qquad u_n = \frac{1}{2n+1}\,\frac{\left(\frac{t^2}{4}\right)^n}{(n!)^2}, \qquad \alpha_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n} + \frac{1}{2n+1}.$$

Si l'on appelle a_n et b_n les coefficients de $\left(\frac{t^2}{4}\right)^n$ dans A et B, leurs valeurs pour les premiers termes sont les suivantes :

$a_0 = 1$, $\quad a_1 = \frac{4}{9} = 0,444\ldots$, $\quad a_2 = 0,085$, $\quad a_3 = 0,007842$,
$a_4 = 0,0004233$, $\quad a_5 = 0,00001499$, $\quad a_6 = 0,000000375$;

$b_0 = 1$, $\quad b_1 = \frac{1}{3} = 0,333\ldots$, $\quad b_2 = 0,05$, $\quad b_3 = 0,003968$,
$b_4 = 0,0001929$, $\quad b_5 = 0,00000631$, $\quad b_6 = 0,000000148$.

Ces termes sont suffisants pour calculer P jusqu'à $t = 4$.

9. Nous avons fait le calcul numérique de P au

moyen de la formule (21), pour des valeurs de t échelonnées de 0,2 en 0,2 jusqu'à $t = 2$ et de 0,4 en 0,4 jusqu'à $t = 4$. Nous avons poussé l'approximation jusqu'au $\frac{1}{1000}$.

A titre de contrôle, nous avons calculé P par les trapèzes, pour $t = 2,5$; 3 et 4. Nous avons retrouvé exactement les trois mêmes nombres que par la première méthode.

Nous avons ensuite construit une courbe, d'où nous avons déduit le Tableau ci-dessous, qui donne la loi de probabilité cherchée.

t.	P.	t.	P.
0,02......	0,060	1,9......	0,931
0,04......	0,110	2,0......	0,938
0,06......	0,151	2,1....	0,945
0,08......	0,186	2,2......	0,951
0,10......	0,218	2,3......	0,957
0,2.......	0,348	2,4......	0,962
0,3.......	0,445	2,5......	0,966
0,4.......	0,525	2,6......	0,970
0,5.......	0,590	2,7......	0,973
0,6.......	0,644	2,8......	0,975
0,7.......	0,689	2,9	0,977
0,8.......	0,728	3,0 .. .	0,980
0,9.......	0,762	3,1......	0,982
1,0.......	0,792	3,2......	0,983
1,1.......	0,817	3,3.	0,985
1,2.......	0,838	3,4......	0,986
1,3.......	0,857	3,5	0,988
1,4	0,874	3,6.. ..	0,989
1,5	0,888	3,7......	0,991
1,6.......	0,901	3,8	0,992
1,7	0,912	3,9......	0,993
1,8.......	0,922	4,0......	0,994

La probabilité P a une allure très différente de la loi de Gauss, c'est-à-dire de la fonction Θ.

Elle est égale à $0,5$ pour $t = 0,363$. On peut comparer la valeur correspondante de z, qui joue le rôle d'*écart médian*, au produit des écarts médians α et β de x et de y. On a

$$t = \frac{2z}{ab} = \frac{2z}{\alpha\beta}\overline{0,4769}^2 = \frac{z}{\alpha\beta} \times 0,4548.$$

Pour $t = 0,363$, on trouve $z = \alpha\beta \times 0,798$. Donc, l'écart médian en z est à peu près les $\frac{8}{10}$ du produit des écarts médians en x, y. Mais, tandis que, dans la loi de Gauss, la probabilité augmente rapidement au delà de $0,5$, elle augmente beaucoup plus lentement avec la loi actuelle. Par exemple, si z est égal à 4 écarts médians, la probabilité P est seulement de $0,882$, au lieu de $0,993$ avec la loi de Gauss. Par contre, pour les petites valeurs de t, la croissance de P est plus rapide que celle de la fonction Θ. La dérivée $\frac{dP}{dt}$ est d'ailleurs infinie pour $t = 0$. Autrement dit, la courbe qui, dans le problème actuel, remplace la courbe classique dite « en chapeau de gendarme », présente une asymptote pour $t = 0$. Le sommet du chapeau est à l'infini. Toutefois, cet infini est atteint très lentement; car il est de l'ordre de $\log t$.