Sur le calcul approché des probabilités par la formule de Laplace.

par Serge Bernstein.

On trouve dans tous les Traités de calcul des probabilités la démonstration du théorème suivant, dû à Laplace:

Si p est la probabilité d'un événement A dans une expérience unique, la probabilité que le nombre m d'apparitions de l'événement A pendant n expériences satisfera à l'inégalité

$$|m-np|<z\sqrt{2np(1-p)}$$

a pour limite

$$\Phi(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-\lambda^2}d\lambda\,,$$

lorsque n augmente indéfiniment.

Cette démonstration fait entrevoir que pour des valeurs de n relativement peu considérables la formule énoncée peut être appliquée sans inquiétude, mais je n'ai nulle part rencontré de calcul rigoureux de l'erreur commise. C'est cette lacune que je me propose de combler dans une certaine mesure, en me bornant au cas simple de $p=\frac{1}{2}$, le cas général pouvant être traité d'une façon analogue, mais un peu plus compliquée. Pour fixer les idées je supposerai le nombre d'expériences n impair

$$n=2\mu-1\,.$$

Il est clair que d'une façon générale la probabilité de l'égalité

$$m-\frac{n}{2}=E,$$

est la même que la probabilité de l'égalité

$$m-\frac{n}{2}=-E.$$

On peut donc se borner à envisager les E positifs.

La plus petite valeur positive de E correspondera à $m=\mu$; on a alors

$$\mu-\frac{2\mu-1}{2}=\frac{1}{2};$$

et la probabilité pour que le nombre d'apparitions de l'événement soit exactement égal à μ, a pour valeur

$$I_{\mu}=\frac{1}{2^n}C_n^{\mu}=\frac{1.3\ldots(2\mu-1)}{2.4\ldots 2\mu}$$

D'une façon générale la probabilité que

$$m-\frac{n}{2}=\frac{1}{2}+k$$

a pour valeur

$$I_{\mu+k}=\frac{1}{2^n}C_n^{\mu+k}=I_{\mu}\frac{(\mu-1)(\mu-2)\ldots(\mu-k)}{(\mu+1)(\mu+2)\ldots(\mu+k)}$$

Donc,

$$\log I_{\mu+k}=\log I_{\mu}+[\log(\mu-1)-\log(\mu+1)]+$$
$$+[\log(\mu-2)-\log(\mu+2)]+\ldots+[\log(\mu-k)-\log(\mu+k)]=$$
$$=\log I_{\mu}+\left[\log\left(1-\frac{1}{\mu}\right)-\log\left(1+\frac{1}{\mu}\right)\right]+\left[\log\left(1-\frac{2}{\mu}\right)-\log\left(1+\frac{2}{\mu}\right)\right]+\ldots$$
$$\ldots+\left[\log\left(1-\frac{k}{\mu}\right)-\log\left(1+\frac{k}{\mu}\right)\right].$$

D'ailleurs, k est toujours inférieur à μ; de sorte que

$$\log\left(1-\frac{k}{\mu}\right)<-\frac{k}{\mu}-\frac{k^2}{2\mu^2},\quad \log\left(1+\frac{k}{\mu}\right)>\frac{k}{\mu}-\frac{k^2}{2\mu^2}$$

Par conséquent,

$$\log I_{\mu+k}<\log I_{\mu}-\frac{2}{\mu}-\frac{4}{\mu}-\frac{6}{\mu}-\ldots-\frac{2k}{\mu}$$

ou

$$\log I_{\mu+k}<\log I_{\mu}-\frac{k(k+1)}{\mu}$$

et

$$I_{\mu+k}<I_{\mu}e^{-\frac{k(k+1)}{\mu}}$$

Prenons ensuite la somme des termes $I_{\mu+k}$ qui correspondent à toutes les valeurs de k supérieures à un nombre donné k_0

$$\sum_{\mu+k_0+1}^{2\mu-1} I_m<I_{\mu}\sum_{k_0+1}^{\mu-1}e^{-\frac{k(k+1)}{\mu}}=I_{\mu}e^{\frac{1}{4\mu}}\sum_{k_0+1}^{\mu-1}e^{-\frac{\left(k+\frac{1}{2}\right)^2}{\mu}} \qquad (1)$$

Mais, il est clair que l'on a

$$e^{\frac{1}{4\mu}} e^{-\frac{\left(k+\frac{1}{2}\right)^2}{\mu}} < \frac{1}{2}\left(e^{-\frac{k^2}{\mu}} + e^{-\frac{\left(k+\frac{1}{2}\right)^2}{\mu}}\right)$$

pour toute valeur de $k \geqq 1$, et d'autre part, en vertu de la décroissance de la fonction $e^{-\frac{k^2}{\mu}}$, on a

$$\frac{1}{2}\left(e^{-\frac{k^2}{4\mu}} + e^{-\frac{\left(k+\frac{1}{2}\right)^2}{\mu}}\right) < \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} e^{-\frac{k^2}{\mu}} dk$$

Donc,

$$I_\mu \sum_{k_0+1}^{\mu-1} e^{\frac{1}{4\mu}} - e^{-\frac{\left(k+\frac{1}{2}\right)^2}{\mu}} < I_\mu \int_{k_0+\frac{1}{2}}^{\infty} e^{-\frac{k^2}{\mu}} dk,$$

et finalement,

$$\sum_{m=\mu+k_0+1}^{m=2\mu-1} I_m < I_\mu \int_{k_0+\frac{1}{2}}^{\infty} e^{-\frac{k^2}{\mu}} dk, \tag{2}$$

En faisant sous le signe d'intégrale le changement de variables $\lambda = \frac{k}{\sqrt{\mu}}$, on donne à l'inégalité (2) la forme

$$\sum_{\mu+k_0+1}^{2\mu-1} I_m < I_\mu \sqrt{\mu} \int_z^{\infty} e^{-\lambda^2} d\lambda, \tag{3}$$

où

$$z = \frac{k_0 + \frac{1}{2}}{\sqrt{\mu}}$$

On sait que le produit

$$P_\mu = I_\mu \,.\, \sqrt{\mu} = \frac{1 . 3 . \ldots . 2\mu-1}{2 . 4 . \ldots . 2\mu} \sqrt{\mu}$$

a pour limite $\frac{1}{\sqrt{\pi}}$. De plus en tendant vers cette limite P_μ lui reste toujours inférieur.

On a ainsi
$$P_\mu < \frac{1}{\sqrt{\pi}}.$$

Il est possible donc de transformer l'inégalité (3) dans la suivante

$$\sum_{\mu+k_0+1}^{2\mu-1} I_m < \frac{1}{\sqrt{\pi}} \int_z^{\infty} e^{-\lambda^2} d\lambda. \tag{4}$$

On en tire que

$$\sum_{\mu}^{\mu+k_0} I_m > \frac{1}{2} - \frac{1}{\sqrt{\pi}}\int_z^{\infty} e^{-\lambda^2}\,d\lambda = \frac{1}{\sqrt{\pi}}\int_z^{\infty} e^{-\lambda^2}\,d\lambda - \frac{1}{\sqrt{\pi}}\int_z^{\infty} e^{-\lambda^2}\,d\lambda =$$

$$= \frac{1}{\sqrt{\pi}}\int_0^{z} e^{-\lambda^2}\,d\lambda = \frac{1}{2}\,\Phi(z)$$

Et, à cause de la symétrie que nous avons remarqué au debut, on a aussi

$$\sum_{\mu-1}^{\mu-k_0-1} I_m > \frac{1}{2}\,\Phi(z).$$

Donc, finalement

$$\sum_{\mu-k_0-1}^{\mu+k_0} I_m = \sum_{\frac{n}{2}-\left(k_0+\frac{1}{2}\right)}^{\frac{n}{2}+\left(k_0+\frac{1}{2}\right)} I_m > \Phi(z) \tag{5}$$

où

$$z = \frac{k_0+\frac{1}{2}}{\sqrt{\mu}} = \frac{k_0+\frac{1}{2}}{\sqrt{\frac{n+1}{2}}}$$

En d'autres termes, *la probabilité pour que*

$$\left| m - \frac{n}{2} \right| \leqq z\sqrt{\frac{n+1}{2}}$$

est (pour toute valeur impaire de n) supérieure [1] *à*

$$\Phi(z) = \frac{2}{\sqrt{\pi}}\int_0^{z} e^{-\lambda^2}\,d\lambda$$

Telle est l'inégalité exacte par laquelle il faut remplacer la formule limite de Laplace et qui montre en même temps le haut degré de précision de cette formule classique. Ainsi, par exemple, pour $n = 199$, en

---

[1]) Il ne faut pas oublier d'ailleurs que

$$\frac{1}{2} + z\sqrt{\frac{n+1}{2}}$$

est supposé être un nombre entier.

prenant $z = 2,25$ l'inégalité obtenue prouve que la probabilité pour que le nombre m d'apparitions de l'événement, dont la probabilité est $p = \frac{1}{2}$, soit compris entre les limites

$$122 \geqq m \geqq 77$$

est supérieure à $\Phi(2,25) = 0,9985373$, tandis que la formule de Laplace nous conduirait à attribuer la même probabilité 0,9985373 à l'inégalité

$$121,9 > m > 77,1$$

ou, ce qui revient au même, à l'inégalité

$$121 \geqq m \geqq 78$$

La comparaison de notre inégalité et de la formule de Laplace montre que c'est là un fait général:

L'écart maximum qu'on peut effectivement affirmer $\left(\textit{pour } p = \frac{1}{2}\right)$ *avec la probabilité* $\Phi(z)$ *dépasse, en général, au plus d'une unité l'écart fourni par l'application de la formule classique de Laplace.*