Estimation de l'incertitude d'un écart-type

Un problème pratique nous a récemment obligés à chercher un moyen pour déterminer numériquement l'erreur statistique d'une variance empirique. Par ailleurs, la nécessité de connaître "l'écart-type d'un écart-type" se manifeste inévitablement dans tous les cas où l'écart-type ne sert pas seulement à donner une mesure pour la précision d'une valeur moyenne, mais où c'est une quantité qui présente en elle-même un intérêt, ou qui est fonction de l'information recherchée. Comme c'est toujours le cas, la fiabilité de cette information dépend alors essentiellement de l'erreur correspondante. C'est pourquoi une méthode générale - pourvu qu'elle ne soit pas trop compliquée - permettant d'évaluer l'écart-type d'un écart-type à partir des données disponibles ne nous a pas paru complètement inutile.

Après quelques efforts infructueux, nous avons réussi à trouver une méthode qui a l'air de correspondre assez bien à la plupart des besoins pratiques.

Si les résultats des n mesures effectuées pour déterminer une certaine quantité x sont désignés par x_i, leur valeur moyenne $\bar{x}$ et l'écart-type s_x (ou $s_{\bar{x}}$ pour $\bar{x}$) sont couramment déterminés à l'aide des formules

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{et} \qquad s_{\bar{x}}^2 = \frac{s_x^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n\,(n-1)} .$$

Après quelques calculs, on obtient comme estimation pour l'écart-type de $s_{\bar{x}}$ l'expression

$$s\,(s_{\bar{x}}) = \frac{s\,(s_x)}{\sqrt{n}} \cong \sqrt{\frac{M_4\,(x) - M_2^2\,(x)}{4n\,(n-1)\,M_2\,(x)}} . \qquad (1)$$

Pour évaluer cet écart-type on n'a donc besoin que du deuxième et du quatrième des moments centraux, qui sont définis par les espérances mathématiques

$$M_k\,(x) = E\left\{(x-\mu)^k\right\} \cong E\left\{(x-\bar{x})^k\right\} ,$$

avec $\mu = E\left\{x\right\} \cong \bar{x}$.

Pour les distributions les plus courantes, ces moments dépendent les uns des autres et sont dans un rapport bien défini et facile à déterminer. En particulier

- pour une densité de Gauss

$$\text{avec} \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2}\left[\frac{x_i - \bar{x}}{\sigma}\right]^2}$$

$$\text{on trouve} \quad M_2(x) = \sigma^2 \,, \qquad M_4(x) = 3\,\sigma^4 \qquad (2)$$

- pour une distribution de Poisson

$$\text{avec} \quad P(x) = e^{-\mu} \frac{\mu^x}{x!}$$

$$\text{on trouve} \quad M_2(x) = \mu \,, \qquad M_4(x) = \mu + 3\,\mu^2 . \qquad (3)$$

Si l'on exprime l'écart-type et son erreur relative sous la forme

$s_{\bar{x}}\,(1 \pm r)$ ou bien aussi $s_x\,(1 \pm r)$,

la relation (1) se réduit facilement, à l'aide de (2) et (3), à

$$r_G \cong \frac{1}{\sqrt{2\,(n-1)}} \,, \qquad \text{pour Gauss,} \qquad (4)$$

$$r_P \cong \frac{1}{2}\sqrt{\frac{2\mu + 1}{\mu(n-1)}} \,, \qquad \text{pour Poisson.} \qquad (5)$$

On vérifie aisément que pour $\mu \gg 1$ l'erreur relative r_P s'approche de plus en plus de r_G , comme on doit s'y attendre. Le simple résultat (4) pour une distribution normale se trouve parfois dans la littérature (même dans le petit "Topping" [1] , par exemple !). Il peut être confirmé en utilisant les moments de la distribution $\chi^2_n(x)$.

Un rapport plus détaillé sur cette étude est prévu.

(Mü.)

Avril 1969

[1] J. Topping: "Errors of observation and their treatment" (Chapman, London, 1963^3)