JOURNAL
DE
MATHÉMATIQUES
PURES ET APPLIQUÉES
FONDÉ EN 1836 ET PUBLIÉ JUSQU'EN 1874
PAR JOSEPH LIOUVILLE

M. BRELOT

Sur l'influence des erreurs de mesure en statistiques

Journal de mathématiques pures et appliquées 9^e série, tome 15 (1936), p. 113-131.
<http://www.numdam.org/item?id=JMPA_1936_9_15__113_0>

Sur l'influence des erreurs de mesure en statistique;

PAR M. BRELOT.

I. — Introduction.

1. On sait comment, en statistique et dans les applications biologiques en particulier, on utilise et compare constamment certains *coefficients* (caractéristiques d'une espèce, comme les moyennes, moyennes d'écarts, de carrés d'écarts, etc.) pour des groupes de n grandeurs γ_i de même nature (par exemple les tailles ou certains rapports de dimensions chez n individus de même âge et d'une même espèce). On a bien étudié comment, dans un ensemble de telles grandeurs analogues, ces coefficients peuvent varier légèrement d'un groupe à l'autre et suivant la valeur de n; ce sont là des problèmes fondamentaux de statistique. Mais il est pour le moins peu connu (sauf pour la moyenne) qu'on ait cherché systématiquement l'ordre de grandeur des *erreurs* sur les coefficients d'un groupe *déterminé* résultant des erreurs expérimentales de détermination des γ_i (celles-ci provenant des *erreurs* sur les *mesures* directes qui servent à déterminer expérimentalement les γ_i) (1). Or, *a priori*, un coefficient expérimental peut différer du coefficient qui correspondrait à des mesures très précises d'une quantité peut-être relativement importante et dont on ignore l'ordre de grandeur; et cette incertitude peut être, comme l'ont signalé

(1) C'est sur la suggestion et la demande de MM. Dieuzeide et Murat, biologistes de la Station expérimentale d'aquiculture et de pêche de Castiglione (Algérie), que j'ai essayé ici de combler un peu cette lacune.

des biologistes, un obstacle ou une cause d'erreurs pour les conséquences. De même qu'en physique, l'utilisation d'une formule suppose un calcul d'erreurs, il convient, pour les formules propres à la statistique, de faire des évaluations d'erreurs analogues. Seulement, les formules de physique ne portent ordinairement que sur *quelques* mesures entachées d'erreurs, tandis que celles de statistique portent sur *beaucoup* de mesures *analogues* entachées d'erreurs; aussi y a-t-il lieu de s'occuper spécialement des calculs d'erreurs en statistique, et plus particulièrement dans le *cas biologique courant où n peut descendre à quelques dizaines*.

2. En ce qui concerne la *moyenne*, rappelons les résultats :

Une limitation absolue ε d'erreur sur chaque grandeur γ_i entraîne la même limitation absolue d'erreur sur la moyenne.

Introduisons séparément ou non, comme hypothèses possibles, des lois d'erreur de Gauss.

A. Hypothèse G_a (loi antérieure). Soit pour γ_i la valeur exacte y_i, la valeur expérimentale x_i, l'erreur $z_i = x_i - y_i$. On suppose, *avant* la détermination de tout γ_i, qu'il existe une probabilité élémentaire du résultat x_i égale à

$$\frac{k}{\sqrt{\pi}} e^{-k^2(x_i-\gamma_i)^2}\, dx_i$$

avec la précision k, *la même pour tous les* γ_i du groupe (et indépendance pour ces divers γ_i).

En conséquence $X = \frac{\Sigma x_i}{n}$ obéit à une loi de Gauss avec précision $k\sqrt{n}$; c'est-à-dire qu'étant données les grandeurs γ_i, non encore mesurées, la probabilité élémentaire relative au résultat futur X est

$$\frac{k\sqrt{n}}{\sqrt{\pi}} e^{-nk^2(X-Y)^2} dX \quad \text{où} \quad Y = \frac{\Sigma y_i}{n}.$$

B. Hypothèse G_p (loi postérieure) (1). En imaginant au besoin une

(1) L'interprétation concrète des probabilités postérieures et l'adoption de cette hypothèse présentent pour les applications pratiques des difficultés que j'expose dans un article moins strictement mathématique (Périodique de la

famille de grandeurs dont sont extraits les γ_i avec une certaine probabilité élémentaire à priori, on admet que, *la détermination expérimentale d'un x_i étant faite*, il existe une probabilité élémentaire (à posteriori) pour la valeur vraie inconnue correspondante y_i égale, pour chaque γ_i, à

$$\frac{k}{\sqrt{\pi}} e^{-k^2(y_i-x_i)^2}\, dy_i,$$

avec indépendance et une précision commune k.

On en déduit des conséquences analogues à celles de plus haut. On distinguera donc les deux cas de probabilités *antérieures* (relatives aux résultats expérimentaux futurs) et de probabilités *postérieures* (relatives, après les déterminations expérimentales supposées connues, aux valeurs exactes inconnues). Dans ces deux cas respectifs les hypothèses G_a et G_p se traduisent par l'expression de la probabilité élémentaire sur l'erreur $z_i = x_i - y_i$

$$\frac{k}{\sqrt{\pi}} e^{-k^2 z_i^2}\, dz_i,$$

et elles entraînent, pour l'erreur $Z = \frac{\Sigma x_i}{n} - \frac{\Sigma y_i}{n}$, la probabilité élémentaire

$$\frac{k\sqrt{n}}{\sqrt{\pi}} e^{-nk^2 Z^2}\, dZ;$$

d'où, à égalité de probabilités, la réduction dans le rapport $\sqrt{n}$, quand on passe d'une limite des $|z_i|$ à une limite de $|Z|$.

3. Le coefficient le plus important après la moyenne est la *dispersion*, ou écart quadratique moyen par rapport à la moyenne, ou encore, selon le vocable anglais très usité, la *standard deviation*. C'est

---

Station d'aquiculture et de pêche de Castiglione) (Algérie) (1935), où je reprends toutes ces recherches en donnant beaucoup plus de résultats numériques. D'ailleurs à cause de l'usage fréquent de ce genre de probabilités des causes, je ne pouvais passer sous silence les conséquences de G_p, d'autant moins qu'elles sont très voisines de celles de G_a.

pour le groupe des n grandeurs γ_i :

$$\text{la valeur exacte } \sigma_{y_i}=\sqrt{\frac{\sum\left(y_i-\frac{\sum y_i}{n}\right)^2}{n}},$$

$$\text{la valeur expérimentale } \sigma_{x_i}=\sqrt{\frac{\sum\left(x_i-\frac{\sum x_i}{n}\right)^2}{n}},$$

qu'on remplace d'ailleurs souvent par les quantités voisines

$$\sqrt{\frac{\sum\left(y_i-\frac{\sum y_i}{n}\right)^2}{n-1}} \quad \text{et} \quad \sqrt{\frac{\sum\left(x_i-\frac{\sum x_i}{n}\right)^2}{n-1}},$$

pour lesquelles une légère correction nous ramène aussitôt au cas précédent.

C'est ce coefficient σ que j'étudie dans cet article aux mêmes points de vue que je viens de rappeler pour la moyenne.

D'abord, *en toute généralité*, par voie purement algébrique, la limitation absolue d'erreur $|x_i-y_i|<\varepsilon$ pour tous les γ_i entraîne la même limitation d'erreur pour σ, soit $|\sigma_{x_i}-\sigma_{y_i}|<\varepsilon$.

Mais ensuite on obtient des limites *pratiques* d'erreurs plus serrées en remplaçant une certitude par une grande probabilité et utilisant *les hypothèses* G_a et G_p. On essaiera dans chacun des cas de probabilités antérieures ou postérieures de trouver des limites d'erreur sur σ correspondant à quelques probabilités comme $1-\frac{1}{20}$, $1-10^{-3}$, $1-10^{-6}$. Il sera même frappant de comparer la limite d'erreur sur les γ_i correspondant à une certaine probabilité et une limite d'erreur sur σ correspondant à une probabilité au moins égale. Ainsi, pour la probabilité $1-10^{-3}$, la limite correspondante d'erreur sur les γ est $\varepsilon_0 \# \frac{2,327}{k}$; on trouve, dans ces deux cas de probabilités, que l'erreur sur σ est moindre que $\frac{\varepsilon_0}{2}$ avec une probabilité supérieure à $1-10^{-3}$, dès que $n\geqq 12$. Autrement dit, dans ce cas toujours pratiquement réalisé $n\geqq 12$, si on néglige les risques de probabilités moindres que 10^{-3}, les limitations absolues d'erreur peuvent être réduites d'au moins moitié quand on passe de la détermination des γ à celle de σ.

On donnera bien d'autres formules plus avantageuses et plus ou moins générales selon les valeurs de ε, σ, n (1).

4. Ainsi un résultat essentiel est-il absolu, indépendant de toute loi de répartition des erreurs. Les autres seront souvent applicables, car beaucoup de mesures directes satisfont sensiblement à une loi de Gauss; il s'ensuit que la somme ou toute combinaison linéaire à coefficients constants de telles mesures *indépendantes* (comme la moyenne de p mesures d'une même grandeur) satisfont à des lois de Gauss; et si même les erreurs sur de telles mesures indépendantes $(x, y, z, \ldots)$ sont pratiquement assez petites pour qu'on puisse confondre une fonction $f(x, y, z, \ldots)$ avec une fonction linéaire de x, y, z dans le champ assez restreint où peut se trouver le point $(x, y, z, \ldots)$, les résultats indiqués seront encore valables lorsque les grandeurs γ_i seront les valeurs d'une telle fonction f pour les n éléments du groupe considéré.

II. — Limitations absolues.

5. Comparons le σ expérimental

$$\sigma_{x_i}=\sqrt{\frac{\sum\left(x_i-\frac{\sum x_i}{n}\right)^2}{n}}$$

et le σ qui correspondrait à des mesures sans erreurs

$$\sigma_{y_i}=\sqrt{\frac{\sum\left(y_i-\frac{\sum y_i}{n}\right)^2}{n}}.$$

Il vient immédiatement, en mettant en évidence les erreurs

(1) Un aperçu des résultats et des méthodes a été donné dans une conférence à la section d'Alger de la Société de Physique de France, le 9 mai 1935; une courte notice en a paru dans *le Bulletin de la Société de Physique.*

$z_i = x_i - y_i$,

$$n\sigma_{y_i}^2 = \sum\left[\left(x_i - \frac{\Sigma x_i}{n}\right) - \left(z_i - \frac{\Sigma z_i}{n}\right)\right]^2,$$

$$n(\sigma_{y_i}^2 - \sigma_{x_i}^2) = \sum\left(z_i - \frac{\Sigma z_i}{n}\right)^2 - 2\sum\left(x_i - \frac{\Sigma x_i}{n}\right)\left(z_i - \frac{\Sigma z_i}{n}\right).$$

Or

$$\sum\left(z_i - \frac{\Sigma z_i}{n}\right)^2 = \Sigma z_i^2 - \frac{2}{n}(\Sigma z_i)^2 + n\left(\frac{\Sigma z_i}{n}\right)^2$$

$$= \Sigma z_i^2 - \frac{1}{n}(\Sigma z_i)^2 \leqq \Sigma z_i^2$$

et

$$\sum\left(x_i - \frac{\Sigma x_i}{n}\right)\left(z_i - \frac{\Sigma z_i}{n}\right) = \Sigma z_i\left(x_i - \frac{\Sigma x_i}{n}\right),$$

d'où l'inégalité fondamentale

$$(1) \qquad |\sigma_{x_i}^2 - \sigma_{y_i}^2| \leqq \frac{\Sigma z_i^2}{n} + \frac{2}{n}\left|\Sigma z_i\left(x_i - \frac{\Sigma x_i}{n}\right)\right|.$$

De même, en permutant x_i et y_i et changeant z_i en $-z_i$, il vient

$$(1') \qquad |\sigma_{x_i}^2 - \sigma_{y_i}^2| \leqq \frac{\Sigma z_i^2}{n} + \frac{2}{n}\left|\Sigma z_i\left(y_i - \frac{\Sigma y_i}{n}\right)\right|.$$

6. Supposons les erreurs z_i en module au plus égales à ε, c'est-à-dire $|z_i| \leqq \varepsilon$. Alors

$$(2) \qquad |\sigma_{x_i}^2 - \sigma_{y_i}^2| \leqq \varepsilon^2 + 2\varepsilon\,\frac{\sum\left|x_i - \frac{\Sigma x_i}{n}\right|}{n},$$

$$(2') \qquad |\sigma_{x_i}^2 - \sigma_{y_i}^2| \leqq \varepsilon^2 + 2\varepsilon\,\frac{\sum\left|y_i - \frac{\Sigma y_i}{n}\right|}{n}.$$

On pourrait utiliser pratiquement la limitation (2). Mais cela exigerait le calcul de l'écart absolu moyen par rapport à la moyenne

$$l_{x_i} = \frac{\sum\left|x_i - \frac{\Sigma x_i}{n}\right|}{n},$$

qui n'est généralement pas utilisé en statistique.

Mais on sait que $l_{x_i} \leqq \sigma_{x_i}$ (1); utilisons cette majoration, d'ailleurs souvent pas très large, puisque, si la série des x_i est normale, $\frac{l_{x_i}}{\sigma_{x_i}}$ est voisin de $\sqrt{\frac{2}{\pi}} \neq 0,8$. D'où la formule générale

$$|\sigma_{x_i}^2 - \sigma_{y_i}^2| \leqq \varepsilon^2 + 2\varepsilon\sigma_{x_i}. \tag{3}$$

De même à partir de (2')

$$|\sigma_{x_i}^2 - \sigma_{y_i}^2| \leqq \varepsilon^2 + 2\varepsilon\sigma_{y_i}, \tag{3'}$$

d'où

$$-(\varepsilon^2 + 2\varepsilon\sigma_{x_i}) \leqq \sigma_{x_i}^2 - \sigma_{y_i}^2 \leqq \varepsilon^2 + 2\varepsilon\sigma_{y_i},$$
$$\sigma_{x_i}^2 \leqq (\sigma_{y_i} + \varepsilon)^2,$$
$$\sigma_{y_i}^2 \leqq (\sigma_{x_i} + \varepsilon)^2$$

et enfin

$$|\sigma_{xi} - \sigma_{yi}| \leqq \varepsilon, \tag{4}$$

résultat fondamental et tout à fait général, annoncé dans l'Introduction.

III. — Utilisation de lois d'erreur. Méthode directe.

7. Étudions directement $|\sigma_{x_i} - \sigma_{y_i}|$ dans chacun des cas de probabilités antérieures ou postérieures avec les hypothèses G_a ou G_p.

La probabilité *antérieure*, pour que $|\sigma_{x_i} - \sigma_{y_i}| < \lambda$ est, d'après G_a,

---

(1) Quels que soient α_i, β_i, on vérifie que

$$\sum_i \alpha_i^2 \sum_i \beta_i^2 = \left(\sum_i \alpha_i\beta_i\right)^2 + \sum_{i,j(i<j)} (\alpha_i\beta_j - \alpha_j\beta_j)^2,$$

d'où, en posant $\alpha_i = \sqrt{\frac{1}{n}}$, $\beta_i = \sqrt{\frac{1}{n}}\,|a_i|$, a_i réel quelconque,

$$\frac{\Sigma a_i^2}{n} \geqq \left(\frac{\Sigma\,|a_i|}{n}\right)^2,$$

quels que soient a_i et n.

l'intégrale multiple

$$(5)\qquad P(\lambda, n)=\left(\frac{k}{\sqrt{\pi}}\right)^n \int\cdots\int e^{-k^2\Sigma(x_i-y_i)^2}\,dx_1\,dx_2\ldots dx_n,$$

étendue au champ R suivant de l'espace euclidien à n dimensions et coordonnées courantes $(x_1, x_2, \ldots, x_n)$:

1° Si $\lambda < \sigma_{y_i}$, champ défini par

$$\sqrt{\frac{\Sigma\left(y_i-\frac{\Sigma y_i}{n}\right)^2}{n}}-\lambda<\sqrt{\frac{\Sigma\left(x_i-\frac{\Sigma x_i}{n}\right)^2}{n}}<\sqrt{\frac{\Sigma\left(y_i-\frac{\Sigma y_i}{n}\right)^2}{n}}+\lambda,$$

ou

$$\sqrt{\Sigma\left(y_i-\frac{\Sigma y_i}{n}\right)^2}-\lambda\sqrt{n}<\sqrt{\Sigma\left(x_i-\frac{\Sigma x_i}{n}\right)^2}<\sqrt{\Sigma\left(y_i-\frac{\Sigma y_i}{n}\right)^2}+\lambda\sqrt{n}.$$

2° Si $\lambda \geqq \sigma_{y_i}$, champ défini par

$$\sqrt{\Sigma\left(x_i-\frac{\Sigma x_i}{n}\right)^2}<\sqrt{\Sigma\left(y_i-\frac{\Sigma y_i}{n}\right)^2}+\lambda\sqrt{n}.$$

Or

$$\sqrt{\Sigma\left(x_i-\frac{\Sigma x_i}{n}\right)^2}=\sqrt{\Sigma x_i^2-\left(\frac{\Sigma x_i}{\sqrt{n}}\right)^2}$$

est la distance du point (x_i) à la droite $x_1 = x_2 = \ldots = x_n$ et le lieu

$$\sqrt{\Sigma\left(x_i-\frac{\Sigma x_i}{n}\right)^2}=\text{const.}$$

est un cylindre de révolution H, d'axe cette droite et de rayon cette constante.

On considérera donc le cylindre H_{y_i} de ce type et passant par le point (y_i), puis le cylindre H'_{y_i} de même axe qui s'en déduit en augmentant le rayon de $\lambda\sqrt{n}$, enfin si $\lambda < \sigma_{y_i}$ le cylindre H''_{y_i}, qui se déduit de H_{y_i} en diminuant le rayon de $\lambda\sqrt{n}$. Le champ R sera alors, dans le cas (1), l'espace compris entre H'_{y_i} et H''_{y_i} et dans le cas (2) tout l'intérieur de H'_{y_i}. L'intégration dans ce champ conduit à une expression compliquée peu propice aux calculs numériques; comme pour les recherches en vue on peut se contenter d'une limite inférieure de P, observons que R contient toujours le volume cylin-

drique de révolution h, d'axe parallèle à $x_1 = x_2 = \ldots = x_n$, passant par le point (y_i) et de rayon $\lambda\sqrt{n}$; alors, si $r^2 = \Sigma(x_i - y_i)^2$,

$$(6) \qquad P(\lambda, n) > \left(\frac{k}{\sqrt{\pi}}\right)^n \int_{(h)} e^{-k^2 r^2}\, dv \qquad (dv, \text{ élém. de vol. pour } n \text{ dim.}).$$

Si C est la section droite de h passant par (y_i) (sphère de rayon $\lambda\sqrt{n}$ dans un espace à $n-1$ dimensions), ρ la distance du point courant à l'axe de R, $d\sigma$ l'élément de volume dans l'espace à $n-1$ dimensions

$$\int_{(h)} e^{-k^2 r^2}\, dv = \int_{(C)} e^{-k^2\rho^2}\, d\sigma \cdot \int_{-\infty}^{+\infty} e^{-k^2 u^2}\, du.$$

Or

$$\int_{-\infty}^{+\infty} e^{-k^2 u^2}\, du = \frac{\sqrt{\pi}}{k},$$

$$\int_{C} e^{-k^2\rho^2}\, d\sigma = S_{n-1} \int_0^{\lambda\sqrt{n}} e^{-k^2\rho^2} \rho^{n-2}\, d\rho,$$

où $S_{n-1}\rho^{n-2}$ représente la surface de la sphère de rayon ρ dans l'espace à $n-1$ dimensions. Donc

$$P(\lambda, n) > \left(\frac{k}{\sqrt{\pi}}\right)^{n-1} S_{n-1} \int_0^{\lambda\sqrt{n}} e^{-k^2\rho^2} \rho^{n-2}\, d\rho,$$

$$(7) \qquad P(\lambda, n) > \frac{S_{n-1}}{\pi^{\frac{n-1}{2}}} \int_0^{k\lambda\sqrt{n}} e^{-u^2} u^{n-2}\, du.$$

Si maintenant on évalue la probabilité *postérieure* pour que $|\sigma_{x_i} - \sigma_{y_i}| < \lambda$ à partir de G_p, on est conduit, en raisonnant de même (les y_i devenant les coordonnées courantes), à la même inégalité finale.

Ainsi, *dans chacun des deux cas de probabilités* (et d'après G_a ou G_p), *la probabilité* $P(\lambda, n)$ *d'une erreur moindre que* λ *satisfait à l'inégalité* (7).

8. Rappelons (¹) que, si n est pair,

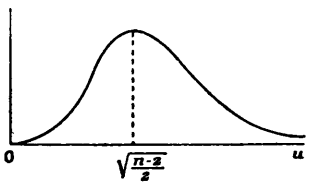
$$n = 2p, \qquad S_{2p} = \frac{2\pi^p}{(p-1)!};$$

(¹) *Voir* par exemple : BOREL, *Introduction géométrique à quelques théories physiques*, p. 62 (Gauthier-Villars); DELTHEIL, *Probabilités géométriques* (fascicule du *Traité des probabilités de Borel*, p. 109 (Gauthier-Villars).

si n est impair,

$$n = 2p+1, \qquad S_{2p+1} = \frac{2(2\pi)^p}{1.3.\ldots.(2p-1)}.$$

On démontre aussi que la courbe représentative de la fonction $\frac{S_{n-1}}{\pi^{\frac{n-1}{2}}} e^{-u^2} u^{n-2}$ (ci-contre) passe par un maximum pour $u = \sqrt{\frac{n-2}{n}}$ et limite avec Ou une aire égale à 1. A équidistance du maximum l'ordonnée de droite est supérieure à celle de gauche, de sorte que l'aire à droite du maximum est plus grande que l'aire à gauche.



Si donc l'on veut, *par la formule* (7), étudier des probabilités voisines de 1, donc $> \frac{1}{2}$, il faudra que

$$\lambda k\sqrt{n} > \sqrt{\frac{n-2}{n}} \qquad \text{ou} \qquad \lambda > \frac{1}{k\sqrt{2}}\sqrt{\frac{n-2}{n}},$$

quantité voisine de $\frac{1}{k\sqrt{2}}$.

On supposera donc dans la suite $\lambda > \frac{1}{k\sqrt{2}}$ et, grâce à (7), on étudiera numériquement $P(\lambda, n)$ quand n varie et pour diverses valeurs de λ.

D'*abord* $P(\lambda, n)$ tend vers 1 *quand* $n \to +\infty$.

Posons $\lambda k\sqrt{2} = \alpha > 1$ et

$$\frac{S_{n-1}}{\pi^{\frac{n-1}{2}}} \int_0^{k\lambda\sqrt{n}} e^{-u^2} u^{n-2}\, du = \frac{S_{n-1}}{\pi^{\frac{n-1}{2}}} \int_0^{\alpha\sqrt{\frac{n}{2}}} e^{-u^2} u^{n-2} = J(\alpha, n).$$

Si $n = 2p+1$, il vient aussitôt

$$(8) \quad J(\alpha, 2p+1) = \frac{1}{(p-1)!} \int_0^{\left(p+\frac{1}{2}\right)\alpha^2} e^{-v} v^{p-1}\, dv$$

$$= 1 - \frac{1}{(p-1)!} \int_{\left(p+\frac{1}{2}\right)\alpha^2}^{\infty} e^{-v} v^{p-1}\, dv.$$

Si $n = 2p$, il vient

$$J(\alpha, 2p) = \frac{2^p}{1.3.5.\ldots.(2p-3)\sqrt{\pi}} \int_0^{\alpha\sqrt{p}} e^{-u^2} u^{2p-2}\, du$$

$$> 1 - \frac{2^p}{1.3.5.\ldots.(2p-3)\sqrt{\pi}} \int_{\alpha\sqrt{p}}^{\infty} e^{-u^2} u^{2p-2} \frac{u}{\alpha\sqrt{p}}\, du,$$

d'où, en utilisant la propriété que

$$\frac{1}{\sqrt{\pi}} \frac{2^{p-1}}{1.3.5.\ldots.(2p-3)\sqrt{\pi}} < \frac{1}{(p-1)!},$$

$$J(\alpha, 2p) > 1 - \frac{1}{\alpha} \frac{1}{(p-1)!} \int_{p\alpha^2}^{\infty} e^{-v} v^{p-1}\, dv. \tag{9}$$

D'après (8) et (9),

$$\left.\begin{matrix} J(\alpha, 2p) \\ J(\alpha, 2p+1) \end{matrix}\right\} > \frac{1}{(p-1)!} \int_0^{(p-1)\alpha^2} e^{-v} v^{p-1}\, dv \tag{10}$$

$$= 1 - \frac{1}{(p-1)!} \int_{(p-1)\alpha^2}^{\infty} e^{-v} v^{p-1}\, dv.$$

Or, on démontre aisément que le second membre est une fonction *croissante* de p et, d'autre part, que

$$1 - \frac{1}{(p-1)!} \int_0^{(p-1)\alpha^2} e^{-v} v^{p-1}\, dv \tag{11}$$

$$= e^{-(p-1)\alpha^2} \left[1 + \frac{\alpha^2(p-1)}{1} + \frac{[\alpha^2(p-1)]^2}{2!} + \ldots + \frac{[\alpha^2(p-1)]^{p-1}}{(p-1)!} \right];$$

$$< e^{-(p-1)\alpha^2} \frac{[\alpha^2(p-1)]^{p-1}}{(p-1)!} \frac{1 - \left(\frac{1}{\alpha^2}\right)^p}{1 - \frac{1}{\alpha^2}} < \frac{[e^{-(\alpha^2-1)}\alpha^2]^{p-1}}{\sqrt{2\pi(p-1)}} \cdot \frac{1}{1 - \frac{1}{\alpha^2}}, \tag{12}$$

quantité qui tend vers o quand $p \to \infty$.

D'après cela, $n \geqq 2p_0$ *entraîne*

$$P(\lambda, n) > \frac{1}{(p_0 - 1)!} \int_0^{(p_0-1)2k^2\lambda^2} e^{-v} v^{(p_0-1)}\, dv, \tag{13}$$

quantité qui tend en croissant vers 1 *quand* $p_0 \to \infty$.

9. On aura aisément des résultats *numériques et pratiques* en utilisant (11), (12) et des tables de la fonction incomplète Γ (1).

Posons

$\varepsilon_q = \frac{1}{k\sqrt{2}} = \frac{0,707\ldots}{k}$, erreur quadratique moyenne théorique sur les x_i;

$\varepsilon_m = \frac{1}{k\sqrt{\pi}} = \frac{0,564\ldots}{k}$, erreur moyenne ou probable;

$\varepsilon_\mu = \frac{0,4769\ldots}{k}$, erreur médiane $\left(\text{probabilité } \frac{1}{2}\right)$;

$\varepsilon_0 = \frac{2,327\ldots}{k} = 3,29\ldots\,\varepsilon_q < 5\,\varepsilon_\mu$, erreur absolue à la probabilité $1 - 10^{-3}$ de ne pas être dépassée.

Exemples :

1° $\lambda = 1,1\,\varepsilon_q \neq \frac{\varepsilon_0}{3}$: $P(\lambda, n) > 1 - 10^{-3}$ dès que $n > 500$.

2° $\lambda = \frac{4}{3}\varepsilon_q \neq 2\varepsilon_\mu \neq 0,4\,\varepsilon_0$: $P(\lambda, n) > 0,95$ dès que $n \geqq 10$; $\boxed{P(\lambda, n) > 1 - 10^{-3}}$ dès que $n \geqq 43$.

3° $\lambda = \frac{\varepsilon_0}{2} = 1,64\ldots\varepsilon_q \neq 2\varepsilon_m$: $\boxed{P(\lambda, n) > 1 - 10^{-3}}$ dès que $n \geqq 12$; $P(\lambda, n) > 1 - 10^{-6}$ dès que $n \geqq 30$.

4° $\lambda = \varepsilon_0$: $P(\lambda, n) > 1 - 10^{-13}$ dès que $n \geqq 10$; $P(\lambda, n) > 1 - 10^{-40}$ dès que $n \geqq 25$.

Le point faible de cette méthode est l'approximation du champ d'intégration R par le champ h, qui conduit à une limitation inférieure (7) trop large de la probabilité.

On va donner une autre méthode qui, sans être aussi générale, conduit à des résultats *pratiques* bien meilleurs dans les cas les plus courants.

---

(1) *Tables of the incomplete* Γ*-function* de Karl Pearson.

Dans le champ des tables, on utilisera d'abord (13) puis, pour quelques valeurs de n inférieures à $2p_0$, les formules (8) et (9).

IV. — Utilisation de lois d'erreur. Autre méthode.

10. Partons des inégalités (1) et (1′) et cherchons, au point de vue des probabilités, des limitations des termes des seconds membres.

Étudions les probabilités *antérieures* avec l'hypothèse G_a. L'expression

$$\Sigma z_i\left(y_i-\frac{\Sigma y_i}{n}\right)=A_{y_i}$$

est linéaire par rapport à z_i, à coefficients indépendants des z_i. Ces z_i indépendants, satisfaisant à des lois de Gauss à même précision k, l'expression A_{y_i} satisfait donc à une loi de probabilité de Gauss avec précision H telle que

$$\frac{1}{H^2}=\frac{\sum\left(y_i-\frac{\Sigma y_i}{n}\right)^2}{k}=\frac{n\sigma_{y_i}^2}{k},\qquad \text{d'où}\qquad H=\frac{k}{\sigma_{y_i}\sqrt{n}}.$$

La probabilité (antérieure) pour que $|A_{y_i}|\leq\eta$ est alors

$$\frac{H}{\sqrt{\pi}}\int_{-\eta_1}^{+\eta}e^{-H^2u^2}\,du=\frac{1}{\sqrt{\pi}}\int_{-H\eta_1}^{+H\eta}e^{-u^2}\,du=\Theta(H\eta)\qquad (\text{fct classique }\Theta).$$

En posant $\eta_1=\rho\frac{\sigma_{y_i}\sqrt{n}}{k}$, il vient

$$(14)\qquad |A_{y_i}|<\rho\frac{\sigma_{y_i}\sqrt{n}}{k}\qquad [\textit{avec la probabilité (antérieure)}\ \Theta(\rho)].$$

En étudiant les probabilités *postérieures* avec l'hypothèse G_p, il vient de même, pour l'expression $\Sigma z_i\left(x_i-\frac{\Sigma z_i}{n}\right)=A_{x_i}$,

$$(14')\qquad |A_{x_i}|<\rho\frac{\sigma_{x_i}\sqrt{n}}{k}\qquad [\textit{avec la probabilité (postérieure)}\ \Theta(\rho)],$$

11. Pour utiliser (1) et (1′) dans chacun des cas de probabilités, il reste à étudier l'expression $\frac{\Sigma z_i^2}{n}$.

Dans les deux cas de probabilités, les z_i étant indépendants et obéis-

sant à G_a ou G_p, la probabilité pour que

$$\frac{\Sigma z_i^2}{n} < \frac{\beta}{2k^2} = \beta \varepsilon_q^2$$

est l'intégrale multiple

$$\Pi(\beta, n) = \left(\frac{k}{\sqrt{\pi}}\right)^n \int \cdots \int e^{-k^2 \Sigma z_i^2} dz_1 \ldots dz_n,$$

étendue dans l'espace euclidien à n dimensions et coordonnées courantes (z_i) à la sphère

$$\Sigma z_i^2 < \frac{n\beta}{2k^2},$$

c'est-à-dire (*voir* n° **8**)

$$\Pi(\beta, n) = \frac{S_n}{\pi^{\frac{n}{2}}} \int_0^{\sqrt{\frac{n\beta}{2}}} e^{-u^2} u^{n-1} du. \tag{15}$$

On désignera par $\beta(\Pi, n)$ la fonction β tirée de cette relation.

Il est évident que si n est fixé, $\Pi \to 1$ (en croissant) quand $\beta \to \infty$ et $\beta \to \infty$ (en croissant) quand $\Pi \to 1$. Il est d'autre part intuitif, d'après la signification de $\frac{1}{2k^2}$, que Π ne peut être voisin de 1, comme il importera, que si β est choisi convenablement > 1. D'ailleurs, d'après l'allure de la courbe $e^{-u^2} u^{n-1}$, $\Pi > \frac{1}{2}$ impose

$$\sqrt{\frac{n\beta}{2}} > \sqrt{\frac{n-1}{2}} \text{ (abscisse du maximum)} \quad \text{ou} \quad \beta > \frac{n-1}{n}.$$

Il est immédiat que chacune des deux propositions suivantes entraîne l'autre :

1° Si β est fixé (> 1), $\Pi(\beta, n) \to 1$ quand $n \to +\infty$;

2° Si Π (< 1) est fixé $> \frac{1}{2}$, $\beta(\Pi, n) \to 1$ quand $n \to +\infty$.

Supposons pour la suite $\beta > 1$ et démontrons directement la première proposition. En utilisant le n° **8**, ou raisonnant de façon très voisine, on trouve

$$\Pi(\beta, 2p) = \frac{1}{(p-1)!} \int_0^{\beta p} e^{-u} u^{p-1} du, \tag{16}$$

$$\Pi(\beta, 2p+1) > 1 - \frac{1}{\sqrt{\beta}} \frac{1}{p!} \int_{\beta\left(p+\frac{1}{2}\right)}^{\infty} e^{-u} u^p du. \tag{17}$$

Par suite $n \geq 2p_0 + 1$ *entraîne* (pour tout β fixé > 1)

$$\Pi(\beta, n) > \frac{1}{p_0!} \int_0^{\beta p_0} e^{-u} u^{p_0} du, \tag{18}$$

quantité qui tend en croissant vers 1 *quand* $n \to \infty$.

12. Remontons à (1) et (1′) pour conclure (¹).

Au point de vue des probabilités *antérieures*, avec G_a, il vient

$$|\sigma_{x_i}^2 - \sigma_{y_i}^2| < \beta\varepsilon_q^2 + \rho \frac{2\sqrt{2}}{\sqrt{n}} \sigma_{y_i} \varepsilon_q, \tag{19}$$

avec une probabilité $> \Pi(\beta, n) + \Theta(\rho) - 1$.

Au point de vue des probabilités *postérieures*, avec G_p, il vient de même

$$|\sigma_{x_i}^2 - \sigma_{y_i}^2| < \beta\varepsilon_q^2 + \rho \frac{2\sqrt{2}}{\sqrt{n}} \sigma_{x_i} \varepsilon_q, \tag{20}$$

avec une probabilité $> \Pi(\beta, n) + \Theta(\rho) - 1$.

Pratiquement, dans ce dernier cas, on tirera aussitôt, connaissant σ_{x_i}, ε_q et n, des limites pour σ_{y_i}, avec une limite inférieure de probabilité correspondante.

Si l'on veut n'admettre que G_a et se placer au point de vue des probabilités antérieures, il faudra partir de (19); on admettra avec un risque mesuré par sa probabilité, que, les mesures faites, le σ_{x_i} obtenu satisfait à (19); en résolvant en σ_{y_i}, on trouvera des limites pour ce coefficient théorique *en fait* inconnu.

Dans les deux cas, si l'on se donne une probabilité minima, il faudra pour obtenir les limites les plus serrées pour σ_{y_i}, choisir β et ρ au mieux — par exemple en tâtonnant — et cela dépend des ordres de grandeur de σ, ε_q et n.

(¹) Grâce à la remarque suivante : Soit relativement à une expérience les événements A et B (réalisations de certaines inégalités) de probabilités x_A et x_B.

Probabilité de A et B simultanément $\geq x_A + x_B - 1$.

13. A côté de ce procédé pénible, mais utilisant au mieux la méthode, indiquons comment l'on obtient, grâce à des approximations, certaines *limitations*, moins bonnes mais d'expression générale *simple*, portant *directement* sur $|\sigma_{x_i} - \sigma_{y_i}|$ et cela pour des cas (d'ailleurs très étendus) où les résultats sont nettement meilleurs que ceux du paragraphe III.

Puisque, couramment, ε_η est petit devant σ_{x_i}, faisons en même temps que

$$(21)\qquad \begin{cases} n \geqq 25, \\ \textit{l'hypothèse } \varepsilon_\eta < \dfrac{\sigma_{x_i}}{6} \left(\text{satisfaite si } \varepsilon_0 < \dfrac{\sigma_{x_i}}{2} \text{ ou si } \varepsilon_{\frac{1}{2}} < \dfrac{\sigma_{x_i}}{9} \right). \end{cases}$$

Cas des probabilités postérieures. — Prenons

$$\left.\begin{array}{l} \rho = 2{,}4 \quad \Theta(\rho) = 0{,}9993\ldots \\ \beta = 2{,}4 \\ n \geqq 25 \end{array}\right\} \Pi(\beta, n) > 0{,}9998\ldots \left.\vphantom{\begin{array}{l} \rho = 2{,}4 \quad \Theta(\rho) = 0{,}9993\ldots \\ \beta = 2{,}4 \\ n \geqq 25 \end{array}}\right\} \Pi(\beta, n) + \Theta(\rho) - 1 > 1 - 10^{-3}.$$

De (20) résulte

$$(22)\qquad \sigma_{x_i}\sqrt{1 - \left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}}} < \sigma_{y_i} < \sigma_{x_i}\sqrt{1 + \left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}}},$$

où

$$\left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}} < 0{,}3;$$

or

$$\sqrt{1 - \left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}}} > 1 - 0{,}55\left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}}$$
$$> 1 - 0{,}55\left(2{,}4\frac{\varepsilon_\eta}{\sigma_{x_i}} + \frac{6{,}8}{\sqrt{n}}\right)\frac{\varepsilon_\eta}{\sigma_{x_i}},$$

et d'autre part

$$\sqrt{1 + \left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}}} < 1 + \frac{1}{2}\left[\beta\frac{\varepsilon_\eta}{\sigma_{x_i}} + \rho\frac{2\sqrt{2}}{\sqrt{n}}\right]\frac{\varepsilon_\eta}{\sigma_{x_i}}$$
$$< 1 + 0{,}5\left(2{,}4\frac{\varepsilon_\eta}{\varepsilon_{x_i}} + \frac{6{,}8}{\sqrt{n}}\right)\frac{\varepsilon_\eta}{\sigma_{x_i}}.$$

Les inégalités (22) entraînent alors

$$|\sigma_{x_i} - \sigma_{y_i}| < 0{,}55\left(2{,}4\frac{\varepsilon_q}{\sigma_{x_i}} + \frac{6{,}8}{\sqrt{n}}\right)\varepsilon_q < \left(1{,}32\frac{\varepsilon_q}{\sigma_{x_i}} + \frac{3{,}75}{\sqrt{n}}\right)\varepsilon_q.$$

Donc, pour $n \geq 25$ et $\varepsilon_q < \frac{\sigma_{x_i}}{6}$, *on conclut qu'on a :*

a. Avec une probabilité postérieure $> 1 - 10^{-3}$,

$$|\sigma_{x_i} - \sigma_{y_i}| < \left(1{,}32\frac{\varepsilon_q}{\sigma_{x_i}} + \frac{3{,}75}{\sqrt{n}}\right)\varepsilon_q < \varepsilon_q < \frac{\varepsilon_0}{3}, \tag{23}$$

ce qui, pour $n \geq 43$, donnerait une limitation moindre que $0{,}8\,\varepsilon_q$ et ε_m.

On trouverait, par des calculs analogues :

b. Avec une probabilité postérieure $> 0{,}95$ (1),

$$|\sigma_{x_i} - \sigma_{y_i}| < \left(1{,}05\frac{\varepsilon_q}{\sigma_{x_i}} + \frac{2{,}1}{\sqrt{n}}\right)\varepsilon_q < 0{,}6\,\varepsilon_q < \varepsilon_\mu. \tag{24}$$

c. Avec une probabilité postérieure $> 1 - 10^{-6}$ (2),

$$|\sigma_{x_i} - \sigma_{y_i}| < \left(1{,}8\frac{\varepsilon_q}{\sigma_{x_i}} + \frac{5{,}65}{\sqrt{n}}\right)\varepsilon_q < \frac{3}{2}\varepsilon_q. \tag{25}$$

Cas des probabilités antérieures. — Remarquons d'abord comment on peut utiliser les calculs qui précèdent.

En remplaçant l'hypothèse $\varepsilon_q < \frac{\sigma_{x_i}}{6}$ par $\varepsilon_q < \frac{\sigma_{y_i}}{6}$, permutant σ_{x_i} et σ_{y_i} et changeant probabilités postérieures en probabilités antérieures, on obtient, par les mêmes calculs, les mêmes limitations (23), (24) et (25) (au changement près de σ_{x_i} en σ_{y_i}).

Dans l'application pratique il faudra seulement être sûr que $\varepsilon_q < \frac{\sigma_{y_i}}{6}$. Or supposons que les mesures aient donné $\sigma_{x_i} > 10\varepsilon_q$. La probabilité antérieure pour que $|\sigma_{y_i} - \sigma_{x_i}| > 3{,}3\ \varepsilon_q$ est, avons-nous vu, extrê-

(1) En prenant $\rho = 1{,}4$, $\beta = 2$.

(2) En prenant $\rho = 3{,}5$, $\beta = 3{,}2$.

mement faible, moindre que 10^{-13} ($n \geqq 10$), et que 10^{-40} ($n \geqq 25$). On pourra donc pratiquement admettre que $\sigma_{y_i} > 6\varepsilon_q$.

Donc on pourra adopter les limitations de droite de (23), (24), (25) avec des probabilités antérieures dépassant $1 - 10^{-3}$, 0,95 et $1 - 10^{-6}$ ($n \geqq 25$) sous l'hypothèse $\varepsilon_q < \frac{\sigma_{x_i}}{10}$.

Mais il est plus avantageux de partir de (19). On en tire toujours *dans les mêmes hypothèses* $\left[n \geqq 25; \frac{\varepsilon_\eta}{\sigma_{x_i}} < \frac{1}{6}\right]$ et en prenant encore $\rho = \beta = 2{,}4$:

$$\sigma_{y_i} - \sigma_{x_i} < \frac{\rho\sqrt{2}}{\sqrt{n}}\varepsilon_\eta + \sigma_{x_i}\left[\sqrt{1 + \left(\frac{2\rho^2}{n} + \beta\right)\left(\frac{\varepsilon_\eta}{\sigma_{x_i}}\right)^2} - 1\right],$$

$$\sigma_{x_i} - \sigma_{y_i} < \frac{\rho\sqrt{2}}{\sqrt{n}}\varepsilon_q + \sigma_{x_i}\left[1 - \sqrt{1 - \left(\beta - \frac{2\rho^2}{n}\right)\left(\frac{\varepsilon_\eta}{\sigma_{x_i}}\right)^2}\right],$$

d'où

$$(23') \qquad |\sigma_{x_i} - \sigma_{y_i}| < \left(1{,}32\,\frac{\varepsilon_\eta}{\sigma_{x_i}} + \frac{3{,}4}{\sqrt{n}}\right)\varepsilon_\eta \leqq 0{,}9\,\varepsilon_\eta.$$

Telle est une limitation d'erreur à laquelle on conclut *en négligeant le risque de probabilité antérieure* $< 10^{-3}$ que (19) ne soit pas vérifiée; et ce résultat est très voisin du résultat (23) avec probabilités postérieures.

V. — Conclusions pratiques.

On cherchera d'abord une limitation absolue ε d'erreur sur la détermination des x_i. Quels que soient les ordres de grandeur de ε, σ_{x_i} et n, ce nombre ε sera aussi une limitation absolue d'erreur sur le σ.

Si l'on désire une meilleure limitation, on cherchera si l'on peut admettre une loi de Gauss G_a ou G_ρ, au moins G_a. Rappelons quant à G_a que pour chaque grandeur γ_i, la répartition de m valeurs obtenues en répétant la détermination expérimentale fournit la précision k par la relation $\frac{1}{2k^2} \# \frac{\Sigma\eta_i}{m-1}$ (η_i écart par rapport à la moyenne des résultats). De plus, il sera immédiat d'évaluer grossièrement l'erreur médiane ε_μ, donc une limite supérieure qui pourra suffire. Quant à G_ρ, si elle est admissible, ce sera ordinairement si G_a l'est déjà et avec la même précision que pour G_a.

Dans bien des questions de statistique, il y a à la base assez de causes mal connues d'incertitude pour qu'on puisse se permettre, dans le champ strict du calcul, de confondre avec une certitude pratique une probabilité dépassant $1 - 10^{-3}$. *Alors* il sera légitime d'adopter comme limitation *pratique* d'erreur sur σ_{x_i}, les quantités $\frac{5}{3}\varepsilon_q$ ou $2{,}5\,\varepsilon_\mu\left(\text{ou } \frac{\varepsilon_0}{2}\right)$ $[n \geqq 12]$ (quels que soient les ordres de grandeur de ε_q et σ_{x_i}). Si même, comme il est courant, $\varepsilon_q < \frac{\sigma_{x_i}}{6}$, on pourra prendre la limitation ε_q ou $1{,}5\,\varepsilon_\mu\left(\text{ou } \frac{\varepsilon_0}{3}\right)$ $[n \geqq 25]$.

Si l'on voulait encore de meilleures limitations, il n'y aurait qu'à se reporter aux divers résultats numériques donnés plus haut (1), en se contentant au besoin d'une moins grande probabilité. Enfin, on pourra aisément, par les méthodes indiquées et les tables de la fonction incomplète Γ obtenir dans chaque cas particulier les meilleures limitations; c'est ce qu'il conviendra de faire pour de grandes valeurs de n. Je me suis en effet limité *numériquement* au cas le plus utile aux biologistes, où n est relativement petit, et dans ce Mémoire aux fins pratiques, j'ai laissé de côté certaines questions théoriques intéressantes comme l'étude du cas où n est infiniment grand (2).

---

(1) Voir les tableaux numériques plus riches du mémoire précité (Périodique de la Station expérimentale d'aquiculture et de pêche de Castiglione).

(2) Voir à ce sujet un exercice utile des *Aufgaben und Lehrsätze aus der Analysis*, de Polya et Szégö, vol. I, p. 80.