

# REVUE DE STATISTIQUE APPLIQUÉE

WŁODZIMIERZ KRYSICKI

## **La variance empirique peut-elle être égale à zéro ?**

*Revue de statistique appliquée*, tome 13, n° 2 (1965), p. 83-94

[http://www.numdam.org/item?id=RSA\\_1965\\_\\_13\\_2\\_83\\_0](http://www.numdam.org/item?id=RSA_1965__13_2_83_0)

© Société française de statistique, 1965, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# LA VARIANCE EMPIRIQUE PEUT-ELLE ÊTRE ÉGALE A ZÉRO ?

Włodzimierz KRYSICKI

École Polytechnique de Lodz

Dans tous les problèmes concernant des mensurations, les grandeurs mesurées peuvent prendre toutes les valeurs d'un certain intervalle. Cependant, étant donné que, pour les diverses mensurations, on adopte des unités données ou, souvent, à la suite de l'exactitude limitée de chaque mensuration, nous attribuons aux grandeurs mesurées des valeurs arrondies ; par ex. si nous mesurons certaines longueurs avec une précision de 1/2 mm en "arrondissant" les résultats jusqu'aux millimètres entiers, nous n'obtiendrons que des nombres entiers.

Il peut donc arriver qu'en effectuant  $n$  mensurations indépendantes avec la même exactitude de moitié d'une certaine unité et en arrondissant le résultat obtenu jusqu'à un nombre entier de ces unités, nous obtenions  $n$  résultats identiques :  $L$  unités, tandis qu'en réalité nous pouvons être sûrs seulement que les véritables grandeurs inconnues  $l_1, l_2, \dots, l_n$ , satisfont l'inégalité

$$L - \frac{1}{2} \leq l_i < L + \frac{1}{2} \quad . \quad i = 1, 2, \dots, n.$$

S'il en est ainsi,  $l_i = L$  quel que soit  $i$ , la variance empirique calculée  $s_e^2$  sera égale à zéro, la variance empirique exacte définie par  $s^2 = \frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^2$  demeurant inconnue.

Cette étude a pour objet de calculer la probabilité qu'en adoptant une certaine exactitude constante pour chacune des  $n$  mensurations toutes ces mensurations donneront le même résultat, c. à. d. que la variance empirique sera égale à zéro.

Nous étudierons ce problème séparément pour les distributions de probabilité suivantes :

- 1/ uniforme (rectangulaire),
- 2/ normale
- 3/ normale tronquée unilatéralement et bilatéralement,
- 4/ triangulaire.

## § 1. CAS DE LA DISTRIBUTION RECTANGULAIRE.

Admettons qu'étant donné l'unité de mesure adoptée les résultats des mensurations ne peuvent donner que les nombres

$$a, a + h, a + 2h, \dots, a + kh,$$

c. à. d. des nombres de forme  $a + ih$ , où  $h > 0$ ,  $i = 0, 1, \dots, k$ , la limite supérieure de la valeur absolue de l'erreur dans chaque mensuration étant égale à  $\frac{h}{2}$ ; alors nous savons seulement que la grandeur réelle qui a donné comme résultat de la mensuration a unités est contenue entre  $a - \frac{1}{2}h$  et  $a + \frac{1}{2}h$ , et que la grandeur réelle dont la mensuration a donné le résultat  $a + kh$  est contenue entre  $a + kh - \frac{h}{2}$  et  $a + kh + \frac{h}{2}$ . Donc la différence entre la plus grande et la plus petite grandeur réelle peut être de  $a + kh + \frac{h}{2} - (a - \frac{1}{2}h) = (k + 1)h$ , c. à. d.  $k + 1$  intervalles de longueur  $h$ .

La probabilité que  $n$  grandeurs indépendantes, variant de façon continue dans l'intervalle

$$\left[ a - \frac{1}{2}h, a + (2k + 1) \frac{h}{2}, \right)^*$$

prendront, dans un des  $(k + 1)$  intervalles disjoints, les valeurs

$$\left[ a + (2i - 1) \frac{h}{2}, a + (2i + 1) \frac{h}{2} \right) \quad i = 0, 1, 2, \dots, k$$

en admettant une distribution de probabilité rectangulaire, (uniforme) est égale à :

$$P_{\text{rest.}} \left[ s_e^2 = 0 \mid n, h \right] = (k + 1) \frac{1}{(k + 1)^n} = \frac{1}{(k + 1)^{n-1}}. \quad (I)$$

C'est donc avec une telle probabilité que, dans les conditions considérées la variance empirique  $s^2$  sera égale à zéro.

Dans le tableau suivant on a présenté la probabilité  $p$  exprimée en % pour  $n = 2, 3, \dots, 10$  et  $k = 1, 2, \dots, 9$ .

n-k	1	2	3	4	5	6	7	8	9
2	50,00	33,33	25,00	20,00	16,67	14,29	12,50	11,11	10,00
3	25,00	11,11	6,250	4,000	2,778	2,041	1,563	1,389	1,000
4	12,50	3,703	1,563	0,800	0,462	0,292	0,195	0,155	0,100
5	6,250	1,234	0,391	0,160	0,077	0,042	0,024	0,017	0,010
6	3,125	0,411	0,098	0,032	0,013	0,006	0,003	0,002	0,001
7	1,563	0,137	0,025	0,006	0,002	0,001	0,000	0,000	0,000
8	0,781	0,046	0,006	0,002	0,000	0,000			
9	0,391	0,015	0,002	0,000					
10	0,191	0,005	0,000						

\* L'intervalle  $[c, d)$  désigne un intervalle fermé à gauche et ouvert à droite c. à. d. un ensemble de nombres satisfaisant l'inégalité  $c \leq x < d$ .

§ 2. DISTRIBUTION NORMALE.

Si la variable aléatoire suit la loi de Laplace-Gauss de moyenne  $m$  et de variance  $\sigma^2$ , la probabilité de satisfaire l'inégalité :

$$x_0 - \frac{h}{2} < x < x_0 + \frac{h}{2}$$

est égale à :

$$\frac{1}{\sigma \sqrt{2\pi}} \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} \exp \left[ -\frac{(t - m)^2}{2\sigma^2} \right] dt$$

En posant  $u = \frac{t - m}{\sigma}$  nous obtenons

$$P \left[ x_0 - \frac{h}{2} \leq x < x_0 + \frac{h}{2} \right] = \frac{1}{\sqrt{2\pi}} \int_{\frac{x_0 - m - \frac{h}{2}}{\sigma}}^{\frac{x_0 - m + \frac{h}{2}}{\sigma}} \exp \left( -\frac{u^2}{2} \right) du.$$

La probabilité que  $n$  variables aléatoires indépendantes  $x_i, i = 1, 2, \dots, n$ , dont chacune suit la loi de Laplace-Gauss de moyenne  $m$  de variance  $\sigma^2$ , prendront des valeurs dans le même intervalle  $x_0 - \frac{h}{2} \leq x_i < x_0 + \frac{h}{2}$ ,  $i = 1, 2, \dots, n$  est égale à

$$\left[ \frac{1}{\sqrt{2\pi}} \int_{\frac{x_0 - m - \frac{h}{2}}{\sigma}}^{\frac{x_0 - m + \frac{h}{2}}{\sigma}} \exp \left( -\frac{u^2}{2} \right) du \right]^n$$

La probabilité que les  $n$  grandeurs citées soient contenues dans un quelconque des intervalles disjoints

$$a_l = x_0 + (2l - 1) \frac{h}{2} \leq x_i < x_0 + (2l + 1) \frac{h}{2} = b_l, \quad i = 1, 2, \dots, n \quad (2)$$

où  $l$  peut varier de  $-\infty$  à  $+\infty$  en passant par tous les nombres entiers, s'exprimera par la formule

$$P_{L-G} [s_0^2 = 0 | n, h] = \sum_{l=-\infty}^{+\infty} \left[ \frac{1}{\sqrt{2\pi}} \int_{\frac{a_l}{\sigma}}^{\frac{b_l}{\sigma}} \exp \left( -\frac{u^2}{2} \right) du \right]^n. \quad (II)$$

Churchill Eisenhart dans "Effects of Rounding or Grouping Data" (1) présente le tableau suivant des valeurs de la probabilité que la variance empirique de  $n$  mensurations indépendantes "arrondies" d'une distribution normale soit égale à zéro.

Nombre des observations N	Longueur de l'intervalle exprimée en écarts-types			
	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$
2	0,271	0,1396	0,0703	0,0302
3	0,085	0,0225	0,0057	0,0014
4	0,028	0,0038	0,0005	0,0001
5	0,009	0,0007		
6	0,003	0,0001		
7	0,001			

(A)

Si nous posons  $\kappa = \frac{h}{\sigma}$ , les conditions suffisantes de possibilité d'appliquer ce tableau sont :

$$\begin{aligned} 0 < \kappa &\leq \frac{1}{200}, & n &\geq 2 ; \\ \frac{1}{200} < \kappa &\leq \frac{1}{10}, & n &\geq 3 ; \\ \frac{1}{10} < \kappa &\leq \frac{1}{3}, & n &\geq 4 ; \\ \frac{1}{3} < \kappa &\leq \frac{1}{2}, & n &\geq 5 ; \\ \frac{1}{2} < \kappa &\leq \frac{4}{5}, & n &\geq 6 ; \\ \frac{4}{5} < \kappa &\leq 1, & n &\geq 7 . \end{aligned}$$

Eisenhart suggère comme indication générale que l'unité dont on se sert dans les mensurations soit inférieure à  $\frac{\sigma}{3}$  ou même à  $\frac{\sigma}{4}$ .

Le calcul de la valeur exacte de l'expression (II) présente des difficultés. Nous pouvons obtenir une approximation tout à fait suffisante pour des fins pratiques de la façon suivante. Nous remplaçons d'abord l'intégrale entre parenthèses par le produit de la longueur de l'intervalle d'intégration  $\frac{b_i}{\sigma} - \frac{a_i}{\sigma} = \frac{h}{\sigma}$  par la valeur de la fonction à intégrer en un certain point  $x_e$  intérieur à cet intervalle et nous obtenons :

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{h}{\sigma}\right)^n \sum_{l=-\infty}^{+\infty} \exp\left(-\frac{nx_e^2}{2}\right)$$

En écrivant cette expression sous la forme :

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \left(\frac{h}{\sigma}\right)^{n-1} \sum_{l=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \frac{h}{\sigma} \exp\left(-\frac{nx_e^2}{2}\right)$$

et en la remplaçant par la valeur approximative de la somme des intégrales, on obtient :

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \left(\frac{h}{\sigma}\right)^{n-1} \sum_{l=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \int_{\frac{lh}{\sigma}}^{\frac{(l+1)h}{\sigma}} \exp\left(-\frac{nt^2}{2}\right) dt,$$

qui à son tour est égale à :

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \left(\frac{h}{\sigma}\right)^{n-1} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{nt^2}{2}\right) dt.$$

En calculant la dernière intégrale nous obtenons enfin comme valeur approximative

$$P_{L-\sigma} [s_e^2 = 0 | n, h] = \left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \left(\frac{h}{\sigma}\right)^{n-1} \frac{1}{\sqrt{n}} \quad (\text{II}^*)$$

Pour  $n = 1$  cette formule approchée, donne  $p = 1$ , ce qui est évident.

Voici les résultats obtenus d'après la formule approchée (II\*) pour quelques valeurs consécutives de n

pour n = 2	p = 28,2094	$\left(\frac{h}{\sigma}\right) \%$
pour n = 3	p = 9,18881	$\left(\frac{h}{\sigma}\right) \%$
pour n = 4	p = 3,17468	$\left(\frac{h}{\sigma}\right) \%$
pour n = 5	p = 1,13279	$\left(\frac{h}{\sigma}\right) \%$
pour n = 6	p = 0,412548	$\left(\frac{h}{\sigma}\right) \%$
pour n = 7	p = 0,152974	$\left(\frac{h}{\sigma}\right) \%$
pour n = 8	p = 0,057521	$\left(\frac{h}{\sigma}\right) \%$
pour n = 9	p = 0,0213874	$\left(\frac{h}{\sigma}\right) \%$
pour n = 10	p = 0,00809451	$\left(\frac{h}{\sigma}\right) \%$

En comparant les valeurs approximatives calculées selon les formules données pour  $p_2, p_3, \dots$  (des résultats voisins avec une exactitude de 4 chiffres significatifs, pour  $n = 2, 3, 4, 5, 6$  sont donnés par U. Graf et H. J. Henning [2] avec les valeurs présentées dans le tableau (A) nous constatons que les différences entre les valeurs correspondantes rarement atteignent 1 %. Dans la majorité des cas elles sont notablement plus petites.

Les représentations graphiques de la variabilité de ces fonctions dans le réseau double logarithmique sont des lignes droites. La figure 1 présente ces droites pour les valeurs du rapport  $\kappa = \frac{h}{\sigma}$  allant de  $\kappa = 0,025$  à  $\kappa = 2,5$  exprimant p ( $i = 2, \dots, 10$ ) en % ; nous remarquerons en même temps que la partie de la figure 1 depuis la valeur de  $\kappa$  égale à peu près à 1 à  $\kappa = 2,5$  n'a pas d'importance pratique et n'a été présentée que pour rendre compte du cours des lignes droites.

Remarquons enfin que dans l'intervalle limité aux valeurs  $m - 3\sigma$  et  $m + 3\sigma$  se placent  $\frac{6\sigma}{h} = \frac{6}{\kappa}$  unités de mesure.

### § 3. DISTRIBUTION NORMALE, UNILATERALEMENT ET BILATERALEMENT TRONQUÉE.

Lorsque la variable prend les valeurs de l'intervalle  $a \leq x \leq b$ , dans une distribution normale bilatéralement tronquée de moyenne m et de variance  $\sigma^2$ , sa densité de probabilité est :

$$\varphi(x) = \frac{\Phi\left(\frac{x-m}{\sigma}\right) \cdot \frac{1}{\sigma}}{\Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)} \quad (3)$$

pour  $a \leq x \leq b$ ,  $\varphi(x) = 0$  pour  $x < a$  et pour  $x > b$ ,

où

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{u^2}{2}\right) du. \quad (4)$$

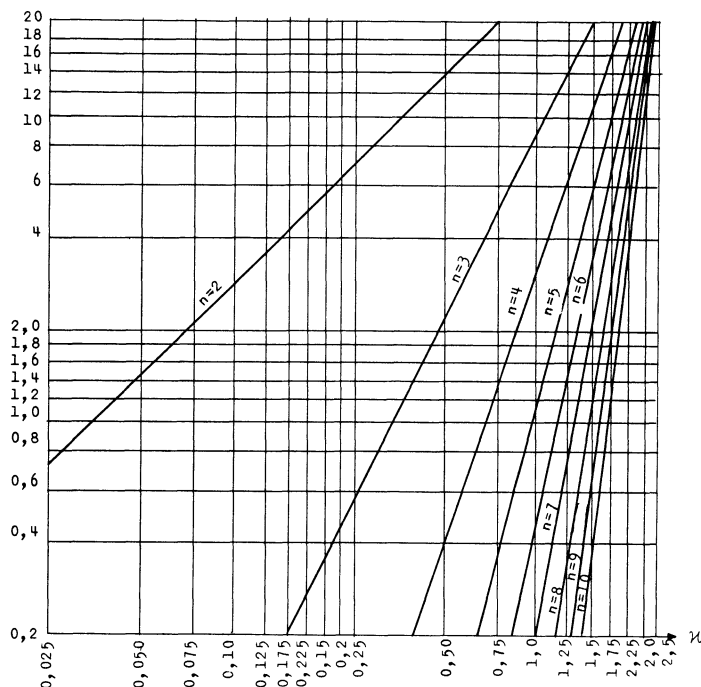


Figure 1 - Rapport de l'unité de mesure à l'écart type.

La forme des courbes  $\varphi(x)$  varie en fonction des paramètres  $m$ ,  $\sigma$  et des limites de l'intervalle  $a, b$  ; cependant indépendamment de cela ce sont des arcs de la courbe de Laplace-Gauss dont les ordonnées ont été multipliées par

$$\left[ \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) \right]^{-1}$$

tandis que les abscisses des extrémités de l'arc sont les valeurs  $a$  et  $b$ . Ces arcs n'auront un axe de symétrie que lorsque l'on a  $2m = a + b$ . Les écarts par rapport à la forme de la courbe de Laplace-Gauss seront importants si la longueur de l'intervalle  $b - a$  est considérablement plus petite que l'intervalle de longueur  $6\sigma^*$ .

Si, en mesurant certaines grandeurs suivant la distribution normale bilatéralement tronquée nous ne pouvons obtenir dans les conditions décrites au début du § 1 que les nombres suivants

$$a, a + h, a + 2h, \dots, a + kh,$$

ce qui veut dire que les grandeurs mesurées  $x$ , étant donné la précision des mensurations, peuvent être effectivement contenues dans l'intervalle

\* Le travail exhaustif de N. A. Borodaczew [3] présente de nombreux détails à ce sujet ainsi que de nombreux dessins des courbes  $\varphi(x)$  définies par la formule (3).

$$a_1 = a - \frac{h}{2} \leq x < a + (2k + 1) \frac{h}{2} = b_1,$$

la probabilité  $P_{L-G}^{tr,2}$ , que les  $n$  résultats des mensurations  $x$  se trouveront dans un même intervalle (quelconque)

$$a + (2i - 1) \frac{h}{2} \leq x < a + (2i + 1) \frac{h}{2}, \quad i = 0, 1, 2, \dots, k,$$

étant donné le raisonnement précédent (§ 2) et la formule (3), sera égale à :

$$P_{L-G}^{tr,2}[s_0^2 = 0 | n, h] = \frac{1}{\left[\Phi\left(\frac{b_1 - m}{\sigma}\right) - \Phi\left(\frac{a_1 - m}{\sigma}\right)\right]^n} \cdot \sum_{i=0}^k \left[ \frac{1}{\sqrt{2\pi}} \int_{\alpha_i}^{\beta_i} \exp\left(-\frac{u^2}{2}\right) du \right]^n \quad (III)$$

où

$$\alpha_i = \frac{a - m + (2i - 1) \frac{h}{2}}{\sigma},$$

$$\beta_i = \frac{a - m + (2i + 1) \frac{h}{2}}{\sigma}.$$

Pour une valeur finie de  $n$  avec  $a_1 = -\infty$ ,  $b_1 = +\infty$ , et par la même avec l'indice  $i$  variant de  $-\infty$  à  $+\infty$  nous obtenons de la formule (III) la formule (II).

$P_{L-G}^{tr,2}$  est la probabilité pour que la variance empirique  $s_0^2$  de  $n$  mensurations indépendantes "arrondies" soit égale à zéro dans ce cas.

La densité de probabilité d'une distribution normale unilatéralement tronquée, de moyenne  $m$  et de variance  $\sigma^2$  pour la variable prenant les valeurs de l'intervalle

$$a \leq x < +\infty$$

sera obtenue comme cas particulier de la formule (3) en y faisant  $b = +\infty$  ; on obtient alors

$$\varphi(x) = \begin{cases} \frac{\Phi\left(\frac{x-m}{\sigma}\right) \cdot \frac{1}{\sigma}}{1 - \Phi\left(\frac{a-m}{\sigma}\right)}, & a \leq x < +\infty \\ 0 & -\infty < x < a \end{cases} \quad (5)$$

Pour une distribution normale unilatéralement tronquée dont la densité de probabilité est donnée par la formule (5), la probabilité que  $n$  mensurations indépendantes seront contenues dans un quelconque des intervalles

$$a + (2i - 1) \frac{h}{2} \leq x < a + (2i + 1) \frac{h}{2},$$

pour  $i$  prenant toutes les valeurs entières non négatives s'exprimera par la formule



$$P_{L-G}^{tr.1} [s_e^2 = 0 | n, h] = \left[ 1 - \Phi \left( \frac{a_1 - m}{\sigma} \right) \right]^{-n} \sum_{i=0}^{\infty} \left[ \frac{1}{\sqrt{2\pi}} \int_{\alpha_i}^{\beta_i} \exp \left( -\frac{u^2}{2} \right) du \right]^n \quad (IV)$$

où

$$a_1 = a - \frac{h}{2}, \quad i = 0, 1, 2, \dots,$$

$$\alpha_i = \frac{a - m + (2i - 1) \frac{h}{2}}{\sigma}, \quad \beta_i = \frac{a - m + (2i + 1) \frac{h}{2}}{\sigma}.$$

Dans ce cas nous pouvons encore remplacer la formule exacte (IV), difficile à appliquer, par une formule approchée. En appliquant à la somme des  $n^{\text{èmes}}$  puissances du second membre de la formule (IV) les transformations déjà utilisées pour obtenir la formule (II\*) nous obtenons :

$$\left( \frac{1}{\sqrt{2\pi}} \right)^{n-1} \left( \frac{h}{\sigma} \right)^{n-1} \int_{\alpha_0}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{nt^2}{2} \right) dt.$$

Puis nous procédons au changement de variable  $\sqrt{nt} = u$ , nous obtenons

$$\left( \frac{1}{\sqrt{2\pi}} \right)^{n-1} \left( \frac{h}{\sigma} \right)^{n-1} \frac{1}{\sqrt{n}} \int_{\alpha_0/\sqrt{n}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{u^2}{2} \right) du.$$

Exprimant la dernière intégrale par l'intégrale de Laplace définie par la formule (4), nous obtenons au lieu de la formule (IV) la formule approchée suivante :

$$P_{L-G}^{tr.1} [s_e^2 = 0 | n, h] = \frac{1 - \Phi(\alpha_0 \sqrt{n})}{\left[ 1 - \Phi \left( \frac{a_1 - m}{\sigma} \right) \right]^n} \left( \frac{1}{\sqrt{2\pi}} \right)^{n-1} \left( \frac{h}{\sigma} \right)^{n-1} \frac{1}{\sqrt{n}} \quad (IV^*)$$

#### § 4. DISTRIBUTION TRIANGULAIRE.

Supposons que la variable aléatoire suit la loi de la distribution triangulaire (selon le triangle isocèle) dans l'intervalle

$$a - \frac{h}{2}, \quad a + kh + \frac{h}{2},$$

donc dans l'intervalle de longueur  $(k + 1)h$ . La densité de probabilité sera alors donnée par la formule

$$\varphi(x) = \left( x - a + \frac{h}{2} \right) \cdot \frac{4}{(k + 1)^2 h^2}$$

pour

$$a - \frac{1}{2} h \leq x < a + \frac{k}{2} h,$$

$$\varphi(x) = - \left[ x - a - (2k + 1) \frac{h}{2} \right] \cdot \frac{4}{(k + 1)^2 h^2} \quad (6)$$

pour

$$a + \frac{k}{2} h \leq x < a + (2k + 1) \frac{h}{2}.$$

Remarquons qu'il faut ici considérer 2 cas selon que le nombre des intervalles  $k + 1$  est pair ou impair.

Cas 1. Le nombre  $k$  des valeurs possibles  $a + (u - 1)h$ ,  $u = 1, 2, \dots, k$  des mensurations obtenues est impair, c. à. d. le nombre des intervalles  $k + 1$  est pair soit :

$$k = 2r - 1, \quad r = 1, 2, \dots, \frac{k + 1}{2}.$$

Nous avons  $BD = rh$ ,  $AD = \frac{1}{rh}$  (figure 2).

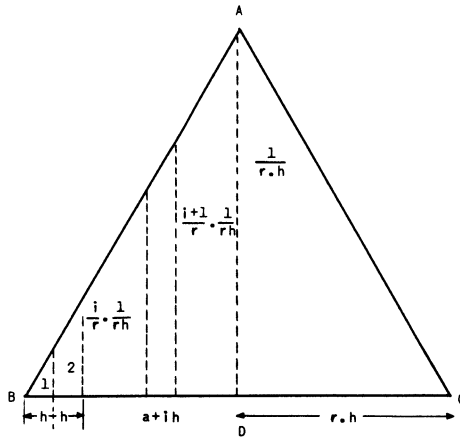


Fig. 2

La probabilité que la variable aléatoire soumise à la distribution envisagée soit contenue dans un intervalle de longueur  $h$

$$a + (2i - 1) \frac{h}{2} \leq x < a + (2i + 1) \frac{h}{2} \quad i = 0, 1, 2, \dots, r-1 \quad (7)$$

est égale à la surface du trapèze de rang  $(i + 1)$ . Cette surface est égale à  $\frac{2i + 1}{2r^2}$ , et la probabilité que  $n$  variables aléatoires indépendantes prennent des valeurs arbitraires de l'intervalle (7) est

$$\left( \frac{2i + 1}{2r^2} \right)^n. \quad (8)$$

Donc pour obtenir la probabilité que  $n$  variables aléatoires indépendantes prennent des valeurs arbitraires dans un quelconque des intervalles disjoints (7) pour  $i = 0, 1, \dots, 2r - 1$ , il faut additionner les valeurs (8) pour  $i = 0, 1, \dots, r-1$ , et ensuite multiplier par 2.

Nous obtiendrons alors

$$P_{\text{triang.}} [s_e^2 = 0 | n, h] = \frac{2}{(2r^2)} \sum_{i=0}^{r-1} (2i + 1)^n. \quad (V)$$

Pour une valeur entière arbitraire  $r > 0$  avec  $n = 1$  nous obtenons de cette formule  $P_{\text{triang.}} = 1$  (en effet la somme des nombres impairs consécutifs allant de 1 à  $2r - 1$  est égale à  $r^2$ ),

Pour  $n = 2$  et pour  $n = 3$ , en vue d'un calcul plus rapide, nous pouvons appliquer les formules suivantes

$$\sum_{i=0}^{r-1} (2i + 1)^2 = \frac{1}{3} r (4r^2 - 1) \quad (9)$$

$$\sum_{i=0}^{r-1} (2i + 1)^3 = r^2 (2r^2 - 1) \quad (10)$$

La formule générale pour la somme des  $n^{\text{èmes}}$  puissances des nombres impairs consécutifs à partir de 1 est la suivante :

$$\begin{aligned} \sum_{i=1}^r (2i - 1)^n &= \frac{2^n}{n + 1} r^{n+1} - \frac{1}{2} C_n^1 2^{n-1} (2^1 - 1) B_2 r^{n-1} \\ &- \frac{1}{4} C_n^3 2^{n-3} (2^3 - 1) B_4 r^{n-3} - \frac{1}{6} C_n^5 2^{n-5} (2^5 - 1) B_6 r^{n-5} - \dots \end{aligned} \quad (11)$$

le dernier terme contient  $r$  ou  $r^2$ , selon que  $n$  est pair ou impair ; les nombres  $B_2, B_4, B_6, \dots$  sont les nombres de Bernoulli apparaissant comme coefficients dans le développement de la fonction

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}.$$

Les quelques premiers nombres de Bernoulli ont les valeurs

$$B_2 = \frac{1}{6} ; \quad B_4 = -\frac{1}{30} ; \quad B_6 = \frac{1}{42} ; \quad B_8 = -\frac{1}{30} ; \quad B_{10} = \frac{5}{66}.$$

Le tableau suivant donne quelques valeurs de  $P_{\text{triang}}$ , pour  $r = 1, 2, \dots, 10$  ;  $n = 2, 3, \dots, 8$  (probabilités en %)

$n \backslash r$	1	2	3	4	5	6	7	8	9	10
2	50,00	31,25	21,6	16,4	13,2	11,0	9,5	8,30	7,4	6,7
3	25,00	11,11	5,25	3,03	1,96	1,4	1,01	0,77	0,61	0,50
4	12,50	4,00	1,35	0,60						
5	6,250	1,49								
6	3,125									
7	0,781									
8										

Cas 2. Le nombre  $k$  des valeurs possibles  $a + (u - 1)h$ ,  $u = 1, 2, \dots, k$  obtenues des mensurations est pair, c. à. d. le nombre des intervalles,  $k + 1$  est impair.

Soit  $k = 2s$ . Alors  $BC = (2s + 1)h$ , donc  $BD = (s + \frac{1}{2})h$ , (figure 3)

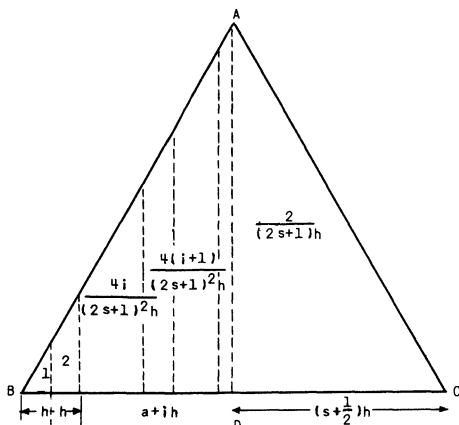


Fig. 3

La probabilité que  $n$  mensurations indépendantes soient alors contenues dans un même intervalle défini par l'inégalité

$$a + (2i - 1) \frac{h}{2} \leq x < a + (2i + 1) \frac{h}{2}, \quad i = 0, 1, \dots, s - 1 \quad (12)$$

est égale à

$$\left[ \frac{2(2i + 1)}{(2s + 1)^2} \right]^n \quad (13)$$

Pour calculer la probabilité que  $n$  mensurations indépendantes soient contenues dans un quelconque des intervalles disjoints (12), il faut additionner les valeurs (13) pour  $i = 0, 1, \dots, s-1$  ; nous obtiendrons alors

$$\frac{2^n}{(2s + 1)^{2n}} \sum_{i=0}^{s-1} (2i + 1)^n. \quad (14)$$

Il faut encore calculer la probabilité que  $n$  mensurations indépendantes soient contenues dans l'intervalle médian EG, c. à d. calculer la surface du pentagone EFAHG et ensuite calculer sa  $n^{\text{ème}}$  puissance. Cette surface calculée comme surface de deux trapèzes égaux est égale à :

$$\frac{4s + 1}{(2s + 1)^2}. \quad (15)$$

Donc en multipliant (14) par 2 et en additionnant (15) nous obtenons la probabilité cherchée.

$$P_{\text{triang.}} [s_e^2 = 0 | n, h] = \frac{1}{(2s + 1)^{2n}} \left[ 2^{n+1} \sum_{i=0}^{s-1} (2i + 1)^n + (4s + 1)^n \right] \quad (VI)$$

Pour une valeur entière arbitraire  $s$  avec  $n = 1$ , de la formule (VI) nous obtenons  $P_{\text{triang.}} = 1$ , ce qui est évident. La manière de calculer la somme des  $n^{\text{èmes}}$  puissances des nombres impairs consécutifs à partir de 1 est donnée par les formules (9), (10) et (11).

Le tableau suivant donne quelques valeurs de la probabilité  $P_{\text{tr.}}$  pour  $n = 2, 3, \dots, 8$  et pour  $s = 1, 2, \dots, s$ , exprimée en %

$\frac{n}{s}$	1	2	3	4	5	6	7	8
2	40,74	25,76	18,7	14,6	12,0	10,2	8,85	7,82
3	19,35	7,55	3,95	2,12	1,57	1,57	0,87	0,67
4	10,01	2,35	0,88	0,53	0,24	0,14		
5	5,401	0,76						
6	2,96							
7	1,64							
8	0,91							

#### REFERENCES

- [1] EISENHARTCHURCHILL - Effects of Rounding or Grouping Data. Selected Techniques of Statistical Analysis. Statistical Research Group. Columbia University 1947.
- [2] GRAF U., HENNING H.J. - Der Einfluss der Messgenauigkeit auf Strennung Spannweite bei kleinen Stichproben, Misseilungs für mathematische Statistik 1951, 2.
- [3] BORODATCHEV N.A. - Les problèmes principaux de l'exactitude de la production Académie de l'U.R.S.S. Moscou-Leningrand 1950 (en russe).