REVUE DE STATISTIQUE APPLIQUÉE

E. MORICE

Étude des possibilités d'une machine à partir d'un échantillon de n observations

Revue de statistique appliquée, tome 12, nº 2 (1964), p. 71-77

<http://www.numdam.org/item?id=RSA_1964__12_2_71_0>

ÉTUDE DES POSSIBILITÉS D'UNE MACHINE A PARTIR D'UN ÉCHANTILLON DE n OBSERVATIONS

E. MORICE

Dans le cas particulier où la distribution de la cote envisagée dans la population produite par la machine est normale, la machine ne subissant pas de déréglage systématique, au cours des observations, la moyenne m et la variance σ^2 de cette distribution pourront être correctement estimées à partir d'un échantillon suffisamment important (par exemple : n >100), par la valeur :

$$x = \frac{1}{2} \sum_{i=1}^{n} x_i$$

$$s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n-1} \# \frac{\sum (x_i - \bar{x})^2}{n}$$

Les tables de la loi normale permettront alors d'estimer la proportion de la population produite qui se trouve dans un intervalle quelconque (a, b), soit :

$$P\ (a, b) = F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right),$$

$$F\ (u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

étant donnée par les tables.

P (a, b) est une variable aléatoire qui dépend des estimateurs utilisés x et s pour m et σ.

La recherche d'un intervalle de confiance pour P est un problème difficile, mais si n est suffisamment grand, sous réserve de l'hypothèse de normalité et aussi de la stabilité du réglage pendant la période d'étude, on pourra considérer P (a, b) comme pratiquement utilisable.

PROBLEME DE WILKS

Wilks a envisagé un problème plus général sous l'angle suivant :

Etant donné un échantillon d'effectif n préleve au hasard dans une population continue de loi quelconque non spécifiée, avec quel degré de confiance γ peut-on dire qu'au moins une proportion β de la population échantillonnée se trouve entre la plus petite valeur x_1 et la plus grande valeur x_n, observées dans l'échantillon.

Wilks a montré que les trois paramètres n, β, γ, étaient liés par la relation :

$$1 - \gamma = n\, \beta^{n-1} - (n-1)\, \beta^{n} \qquad (1)$$

La résolution de cette équation permet, en principe, de calculer l'une quelconque des trois inconnues n, β, γ, en fonction des deux autres.

De nombreuses formules d'approximation ont été proposées pour calculer β ou n, des tables ont été calculées et des abaques ont été calculées et des abaques ont été construits.

A - Formules approchées

OWEN [1]

$$\beta = \frac{n - 1}{n - 1 + 2\, F_{\gamma} \cdot (\nu_1 = 4, \quad \nu_2 = 8\ n-2)}$$

avec $\gamma = Pr\ (F < F_{\gamma})$, distribution de Fisher-Snedecor.

SCHEFFE et TUCKEY [2]

$$n = 4{,}74\ \chi^2_{\gamma}(\nu = 4) + 0{,}03 \qquad \text{pour } \beta = 0{,}90$$

$$n = 9{,}75\ \chi^2_{\gamma}(\nu = 4) + 0{,}01 \qquad \text{pour } \beta = 0{,}95$$

$$n - 49{,}75\ \chi^2_{\gamma}(\nu = 4) \qquad \text{pour } \beta = 0{,}99$$

avec $\gamma = Pr\ (\chi^2 < \chi^2_{\gamma})$

BOWKER ET LIEBERMAN [3]

$$n \ \# \ 0{,}25\ \frac{1 + \beta}{1 - \beta}\ \chi^2_{\gamma}(\nu + 4) + 0{,}5$$

B - Tables

SOMERVILLE [4]

Valeurs de γ pour n = 3...(1)...20, 25, 30...(10)...100

et β = 0,5 - 0,90 - 0,95 - 0,99

OWEN (1)

Valeurs de n pour β et γ = 0,5 - 0,7... (0,05)...0,95

0,975 - 0,95 - 0,99 - 0,995

DIXON et MASSEY [5]

Valeurs de n pour β = 0,5 - 0,9 - 0,95 - 0,99

et γ = 0,9 - 0,95 - 0,99

DION [6]

Valeurs de n pour β et γ = 0,01 - 0,02 - 0,05 - 0,10...

(0,10)...0,90 - 0,95 - 0,98 - 0,99

C - Abaques

MURPHY [7]

Valeurs de n pour $0,01 \leqslant \beta \leqslant 0,999$ et

γ = 0,9 - 0,95 - 0,99

BIRMBAUM et ZUCKERMAN [8]

Valeurs de n pour $0,8 \leqslant \beta \leqslant 0,999$ et

$0,8 \leqslant \gamma \leqslant 0,99$

NELSON [9]

Abaque à points alignés pour $0,75 < \gamma < 0,99$

$0,75 < \beta < 0,997$ et $10 < n < 1\ 000$

Exemple : **Déterminer l'effectif de l'échantillon nécessaire pour qu'au niveau de confiance 0,99 on puisse conclure qu'il y a au moins 95 % de la population compris à l'intérieur de l'étendue d'un tel échantillon.**

Pour γ = 0,99, β = 0,95, l'équation approchée de Bowker et Liebermann donne :

$$n = 0,25\ \frac{1,95}{0,05} \times 13,28 + 0,5 \ \# \ 130$$

On peut donc conclure que, quelle que soit la loi de distribution dans la population, il y a au moins 95 % de la population à l'intérieur de l'étendue observée, au risque 1 - γ = 1 % d'une conclusion erronée.

Valeurs entières de n vérifiant l'équation :

$$n\,\beta^{n-1} - (n-1)\,\beta^{n} = 1 - \gamma$$

(extrait de la table de Dion - MIT 1951.

$\gamma \backslash \beta$	0,99	0,98	0,95	0,90	0,80	0,70	0,50
0,99	661	330	130	64	31	20	11
0,98	581	290	115	56	27	17	9
0,95	473	236	93	46	22	14	8
0,90	388	191	77	38	18	12	7
0,80	299	149	59	29	14	9	5
0,70	244	122	49	24	12	8	5
0,50	168	84	34	17	9	6	3

Pour n = 130, on a une probabilité $\gamma = 0,99$ que 95 % au moins des valeurs de la population (quelconque) dont a été tiré l'échantillon, soient à l'intérieur de l'étendue de l'échantillon observé.

ANNEXE THEORIQUE

Le schéma général de n observations possibles se présente de la manière ci-après :

Nombre d'observation	0	1		n - 2	1	0	
Intervalles	$-\infty$	x_1	$x_1 + dx_1$		x_n	$x_n + dx_n$	$+\infty$

Compte tenu des permutations possibles de n valeurs, la loi du couple (x_1, x_n) sera :

$$p\,(x_1, x_n)\,dx_1 dx_n = \frac{n!}{(n-2)!}\left[\int_{x_1}^{x_n} f\,(x)\,dx\right]^{n-2} f\,(x_1)\,f\,(x_n)\,dx_1 dx_n$$

$$= n\,(n-1)\left[F\,(x_n) - F\,(x_1)\right]^{n-2} f\,(x_1)\,f\,(x_n)\,dx_1\,dx_n \qquad (1)$$

Si l'on pose :

$$u = \int_{-\infty}^{x_1} f\,(x)\,dx = F\,(x_1)$$

$$v = \int_{x_1}^{x_n} f\,(x)\,dx = F\,(x_n) - F\,(x_1)$$

on constatera que la loi de distribution de v, c'est à dire de la proportion de la population qui se trouve entre la plus petite et la plus grande valeur de l'échantillon, est indépendante de la fonction f (x).

En effet, la loi du couple (u, v) est définie par :

$$g\,(u,v) = p\,[x\,(u,v),\ x_n\,(u,v)]\ .\ \begin{vmatrix} \dfrac{\partial u}{\partial x_1} & \dfrac{\partial v}{\partial x_1} \\ \dfrac{\partial u}{\partial x_n} & \dfrac{\partial v}{\partial x_n} \end{vmatrix}^{-1}$$

avec :

$$\frac{\partial u}{\partial x_1} = f(x_1) \qquad \frac{\partial u}{\partial x_n} = 0$$

$$\frac{\partial v}{\partial x_1} = -f(x_1) \qquad \frac{\partial v}{\partial x_n} = f(x_n)$$

Le jacobien de la transformation est donc égal à $\frac{1}{f(x_1) f(x_n)}$ d'où la loi du couple (u, v) :

$$g(u,v)\, du\, dv = n(n-1)\, v^{n-2} du\, dv$$

En intégrant par rapport à u pour $0 < u < 1-v$, on obtient la loi de v :

$$h(v)\, dv = n(n-1)\, v^{n-2}(1-v)\, dv, \tag{2}$$

loi de la proportion de la population qui est comprise entre x_1 et x_n, loi qui dépend de n, mais non de f (x).

Si on se propose de déterminer n de façon que la probabilité soit γ (suffisamment élevée) pour qu'au moins une proportion β (suffisamment élevée) de la population se trouve à l'intérieur de l'étendue observée, il faudra calculer n satisfaisant à l'équation :

$$\gamma = \int_{\beta}^{1} n(n-1)\, v^{n-2}(1-v)\, dv$$

c'est-à-dire :

$$n\beta^{n-1} - (n-1)\beta^{n} = 1 - \gamma$$

Remarque : La loi de v (équation (2) est précisément celle de la distribution de l'étendue dans le cas d'échantillons provenant d'une population suivant une loi uniforme :

$$f(x) = 1 \qquad \text{pour} \qquad 0 < x < 1$$

L'équation (1) devient alors :

$$p(x_1, x_n)\, dx_1 dx_n = n(n-1)\left[\int_{x_1}^{x_n} dx\right]^{n-2} dx_1 dx_n$$

soit :

$$= n(n-1)(x_n - x_1)^{n-2} dx_1\, dx_n$$

Si l'on désigne l'étendue par :

$$W = x_n - x_1$$

on obtiendra la loi de W en intégrant par rapport à x_1 pour $0 < x_1 < 1-W$, soit encore :

$$h(W)\, dW = n(n-1)\, W^{n-2}(1-W)\, dW$$

probabilité pour que l'étendue soit comprise entre W et W + dW pour un échantillon provenant d'une population uniformément distribuée dans l'intervalle (0, 1), cas auquel on peut toujours se ramener, moyennant un changement d'origine et d'unité.

BIBLIOGRAPHIE

[1] OWEN - Distribution free tolerance limits for an additionnal finite sample obtained from Hypergeometric distribution (Wester Region Conf. ASQC - Anaheim - March 1961).

[2] SCHEFFE et TUKEY - Another Beta function approximation (Memorandum report N° 28 - Stat. Res. Group - Princeton University - 1969).

[3] BOWKER et LIEBERMANN - Engineering statistics (Prentice Hall - Englewood Cliffs - 1959 p. 229).

[4] SOMERVILLE - Tables for obtaining non-parametric tolerance limits (Ann. Math. Stat. 29 - 1958 - p. 599 - 601).

[5] DIXON et MASSEY - "Introduction to statistical analysis" (Mc Graw hill - New York 1957 - p. 294).

[6] DION - An approximate solution of Wilk's tolerance limits equation (M.I.T. These 1951).

[7] MURPHY - Non parametric tolerance limits (Ann. Math. Stat. 19 (1948) p. 581 - 589).

[8] BIRNBAUM et ZUCKERMANN - A graphical determination of sample size for Wilk's tolerance limits (Ann. Math. Stat. 20 (1949) p. 313-316).

[9] NELSON - A nomograph for two sided distribution-free tolerance intervals (Industr. Quality Control XIX N° 12 - 1963).

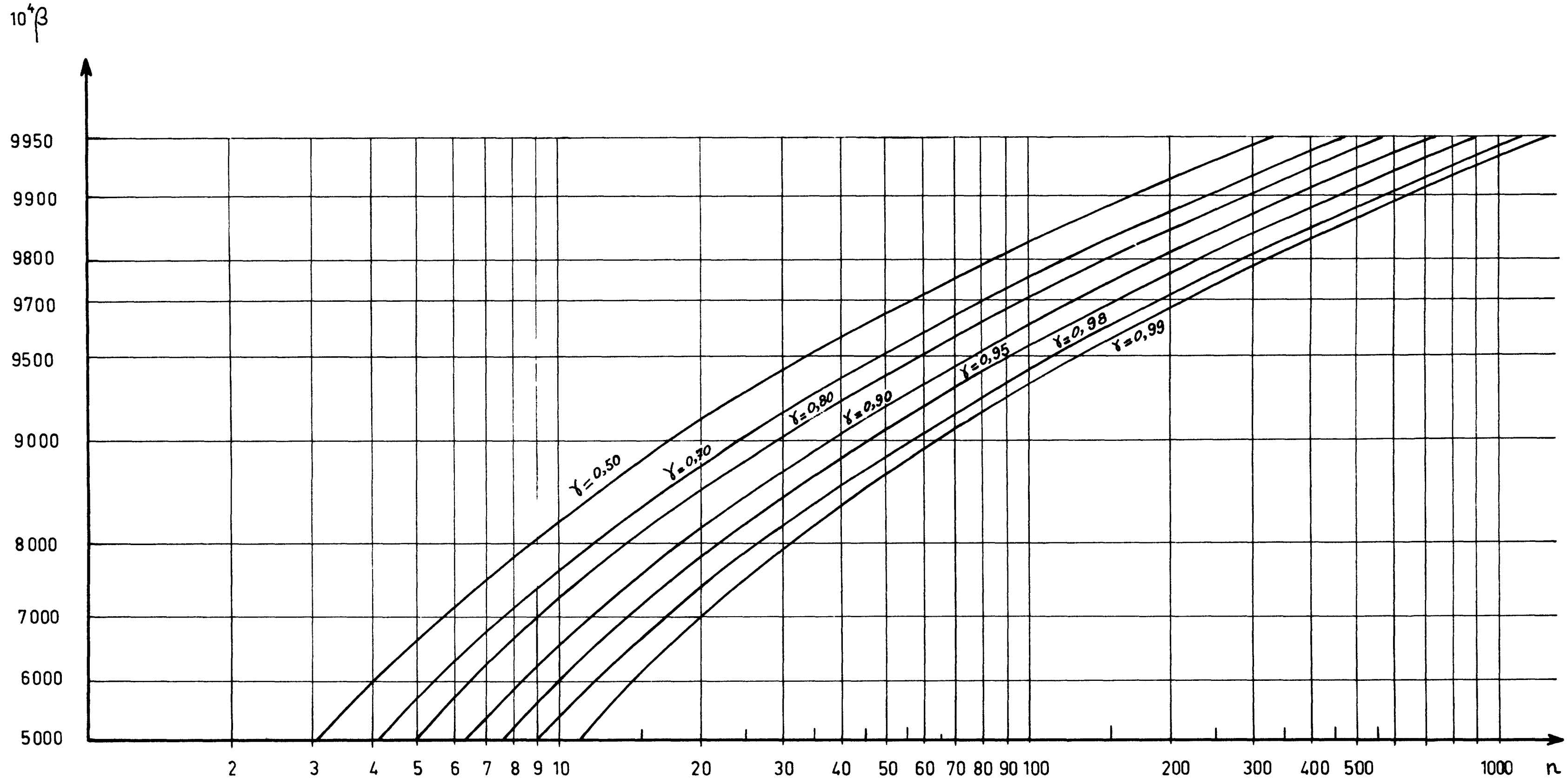
[10] BINGHAM - Tolerance limits and process capability studies (Ind. Quality Control XIX N° 1 - 1962).

[11] GRYNA - Determination of process capability (Ind. Quality Control XI N° 3 - 1954).

[12] WILDS - Statistical prediction with special reference to the problem of tolerance limits (Ann. Math. Stat. 13 N° 4 - 1942).

[13] MITRA - Tables for tolerance limits of a normal population based on sample mean and range (Journal of Amer. Statist. Assoc. 1952 - N° 277).

[14] PROSAHAM - Confidence and tolerance limits for the normal distribution (Journal of Amer. Statist. Assoc. 1948 - N° 263.



Equation de Wilks $1-\gamma = n\beta^{n-1} - (n-1)\beta^{n}$

γ = probabilité qu'une fraction au moins égale à β de la Population soit comprise dans l'étendue d'un échantillon de n observations

(Population continue ; loi non spécifiée)

20.096-64