

REVUE DE STATISTIQUE APPLIQUÉE

F. CHARTIER

E. MORICE

Validité de l'approximation de Poisson pour les bornes d'un intervalle de confiance relatif à une proportion

Revue de statistique appliquée, tome 25, n° 2 (1977), p. 27-32

http://www.numdam.org/item?id=RSA_1977__25_2_27_0

© Société française de statistique, 1977, 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/>

VALIDITÉ DE L'APPROXIMATION DE POISSON POUR LES BORNES D'UN INTERVALLE DE CONFIANCE RELATIF A UNE PROPORTION

F. CHARTIER et E. MORICE

1 — L'EQUATION BINOMIALE

Soit à former l'intervalle de confiance relatif à la proportion p des unités d'une population qui possèdent un caractère donné à l'aide d'un échantillon aléatoire de cette population, d'effectif n et dont k unités présentent le caractère considéré.

La borne supérieure de cet intervalle, p_s , aléatoire comme k dont elle est fonction, est déterminée par la condition :

$$\text{Prob}(p_s < p) = \alpha \dots \dots (1)$$

pour un intervalle de confiance unilatéral correspondant au niveau de confiance $1 - \alpha$ ou pour un intervalle de confiance bilatéral, (p_i, p_s) , symétrique en probabilité, de niveau de confiance $1 - 2\alpha$.

Si les n unités constituant l'échantillon sont indépendantes en probabilité, k est une réalisation d'une variable aléatoire de la loi binomiale $\mathcal{B}(n, p)$ et la condition (1) conduit à rechercher la valeur de p telle que :

$$\sum_{x=0}^k C_n^x p^x (1-p)^{n-x} = \alpha \dots \dots (2)$$

Cette équation en p comporte une racine unique, p_{sB} , qui peut être interprétée comme la valeur maximale de p telle que la probabilité d'observer k unités ou moins de k possédant le caractère considéré ne soit pas inférieure à α .

2 — L'APPROXIMATION DE POISSON

La détermination de la racine exacte, p_{sB} , de l'équation (1) suppose l'utilisation des tables de la loi binomiale, peu courantes à cause de leur volume, ou des tables de la loi de F (voir [1]).

Il est bien connu que si n tend vers l'infini et p vers zéro, de façon que le produit np tende vers un nombre fini m , la loi binomiale $\mathcal{B}(n, p)$ tend vers la

loi de Poisson de paramètre m , $\mathcal{P}(m)$, de sorte que l'équation (2) est remplacée par :

$$\sum_{x=0}^k e^{-m} \frac{m^x}{x!} = \alpha \dots \dots \quad (3)$$

Cette équation en m peut être résolue par consultation des tables de la loi de Poisson, ce qui nécessitera généralement une interpolation pour obtenir la racine m_s , d'où l'on déduira l'approximation poissonnienne de p_{sB} , soit $p_{sP} = m_s/n$.

Il est équivalent, mais plus commode et plus précis, d'obtenir p_{sP} à l'aide du fractile d'ordre $1 - \alpha$ de la variable χ^2 à $2(k + 1)$ degrés de liberté :

$$p_{sP} = \frac{1}{2n} \chi_{1-\alpha}^2(2k + 2) \dots \dots \quad (4)$$

(Pour la relation entre loi de Poisson et loi de χ^2 , on pourra se reporter à [1]).

La question qui se pose alors au praticien est la suivante : pour une valeur de n qui n'est pas infinie, parfois de l'ordre de quelques dizaines seulement, et une valeur de p qui n'est pas nulle et dont on ne sait même pas a priori ce qu'elle sera, sa valeur p_{sP} n'étant connue qu'après l'approximation, *quelle est l'importance de l'écart $e = p_{sP} - p_{sB}$?*

3 – LA VALIDITE DE L'APPROXIMATION DE POISSON

Les deux abaques qui suivent permettent, pour cinq valeurs usuelles de α (0,1 – 0,05 – 0,025 – 0,01 et 0,005), de voir pour tout couple (n, k) où $n \leq 100$, si cet écart est inférieur ou supérieur à 0,01 (abaque n° 1), inférieur ou supérieur à 0,02 (abaque n° 2), en situant le point représentatif du couple (n, k) par rapport à la courbe correspondant à la valeur choisie de α .

On constate que plus α est faible, plus les conditions théoriques de l'approximation de Poisson (n grand, p faible, soit ici k petit) doivent être fortement respectées.

Par exemple, d'après l'abaque n° 1, pour $n = 70$ et $k = 8$, l'approximation de Poisson donnera un écart $e = p_{sP} - p_{sB} < 0,01$ sur la borne supérieure d'un intervalle unilatéral $(0, p_s)$ de niveau de confiance $1 - \alpha = 0,95$ (soit $\alpha = 0,05$). En revanche, l'écart sur la borne supérieure d'un intervalle bilatéral symétrique en probabilité, (p_i, p_s) , de même niveau de confiance 0,95, serait supérieur à 0,01, la valeur de α à considérer n'étant plus que de 0,025.

Ces constatations sont conformes à ce qui a été établi dans un article antérieur [2].

On notera aussi que les conditions pratiques de validité de l'approximation de la loi binomiale par la loi de Poisson, qui sont souvent indiquées dans les cours de statistique, soit la seule condition $p < 0,10$, soit les conditions combinées $p < 0,10$ et $n > 50$, peuvent être – en ce qui concerne le calcul des bornes de

l'intervalle de confiance — trop ou trop peu exigeantes, si l'utilisateur est tenté d'y remplacer p inconnu par k/n .

Ainsi, par exemple, pour $n = 100$, $k = 15$ et $\alpha = 0,05$, la loi de Poisson donne $p_{sP} = 0,23$ pour $p_{sB} = 0,22$, bien que $k/n = 0,15$ soit supérieur à $0,10$.

Par contre, pour $n = 12$, $k = 1$, $\alpha = 0,05$, la loi de Poisson donne $p_{sP} = 0,40$ pour $p_{sB} = 0,34$ bien que $k/n = 0,083$ soit inférieur à $0,10$.

ANNEXE

MODE DE CONSTRUCTION DES ABAQUES

Il s'agissait de déterminer un certain nombre de couples (n, k) tels que $e = p_{sP} - p_{sB} = 0,01$ (abaque n° 1) ou $0,02$ (abaque n° 2), successivement pour $\alpha = 0,1 ; 0,05 ; \dots$.

On a procédé à k donné : $k = 0, 2, 5, 9, 14, 19, \dots$, ce qui permettait de lire le fractile $\chi^2_{1-\alpha}(2k+2)$ dans les Biometrika Tables, puis d'en déduire p_{sP} par division par $2n$, pour différentes valeurs de n .

Parallèlement, on a recherché dans les tables de la loi binomiale cumulée [3] — [4], pour les mêmes couples (n, k) , la valeur p_1 (en nombre entier de centièmes) telle que :

$$\sum_0^k C_n^x p^x (1 - p)^{n-x} \begin{cases} < \alpha & \text{pour } p = p_1 \\ > \alpha & \text{pour } p = p_1 + 0,01. \end{cases}$$

On en a tiré la valeur p_{sB} telle que la somme soit égale à α par interpolation linéaire sur p entre p_1 et $p_1 + 0,01$, avec arrondissage au 10^{-4} le plus proche, ce qui ne justifie sans doute pas une méthode d'interpolation plus raffinée.

De l'examen des valeurs p_{sB} et p_{sP} correspondant aux différents couples (n, k) , on a déduit les deux nombres n_1 et n_2 ($n_2 > n_1$) tels que :

$$e = p_{sP} - p_{sB} \begin{cases} = e_1 > 0,01 \text{ (ou } 0,02) & \text{pour } (n_1, k) \\ = e_2 < 0,01 \text{ (ou } 0,02) & \text{pour } (n_2, k) \end{cases}$$

Par interpolation linéaire sur n entre n_1 et n_2 , on a finalement obtenu la valeur n_0 (au dixième le plus proche) à associer à k , telle que $e = 0,01$ (ou $0,02$).

Remarque : en ce qui concerne la borne inférieure p_i d'un intervalle unilatéral $(p_i, 1)$ à $1 - \alpha$ ou bilatéral (p_i, p_s) symétrique, à $1 - 2\alpha$, l'équation binomiale, de racine p_{iB} , s'écrit :

$$\sum_{x=k}^n C_n^x p^x (1 - p)^{n-x} = \alpha \dots \dots \quad (2)'$$

et l'équation de Poisson, de racine m_i :

$$\sum_{x=k}^{\infty} e^{-m} \frac{m^x}{x!} = \alpha \dots \dots (3)'$$

d'où la valeur approchée $p_{iP} = m_i/n$, valeur que l'on obtient plus commodément par :

$$p_{iP} = \frac{1}{2n} \chi^2_{\alpha}(2k) \dots \dots (4)'$$

Il résulte de [2] que si n et k satisfont aux conditions nécessaires pour que $p_{sP} - p_{sB} < 0,01$ (ou $0,02$), l'écart entre p_{iP} et p_{iB} sera très inférieur à $0,01$ (ou $0,02$).

A titre indicatif, pour $n = 70$ et $k = 8$, on a les valeurs suivantes pour trois intervalles de même niveau de confiance $1 - \alpha = 0,95$:

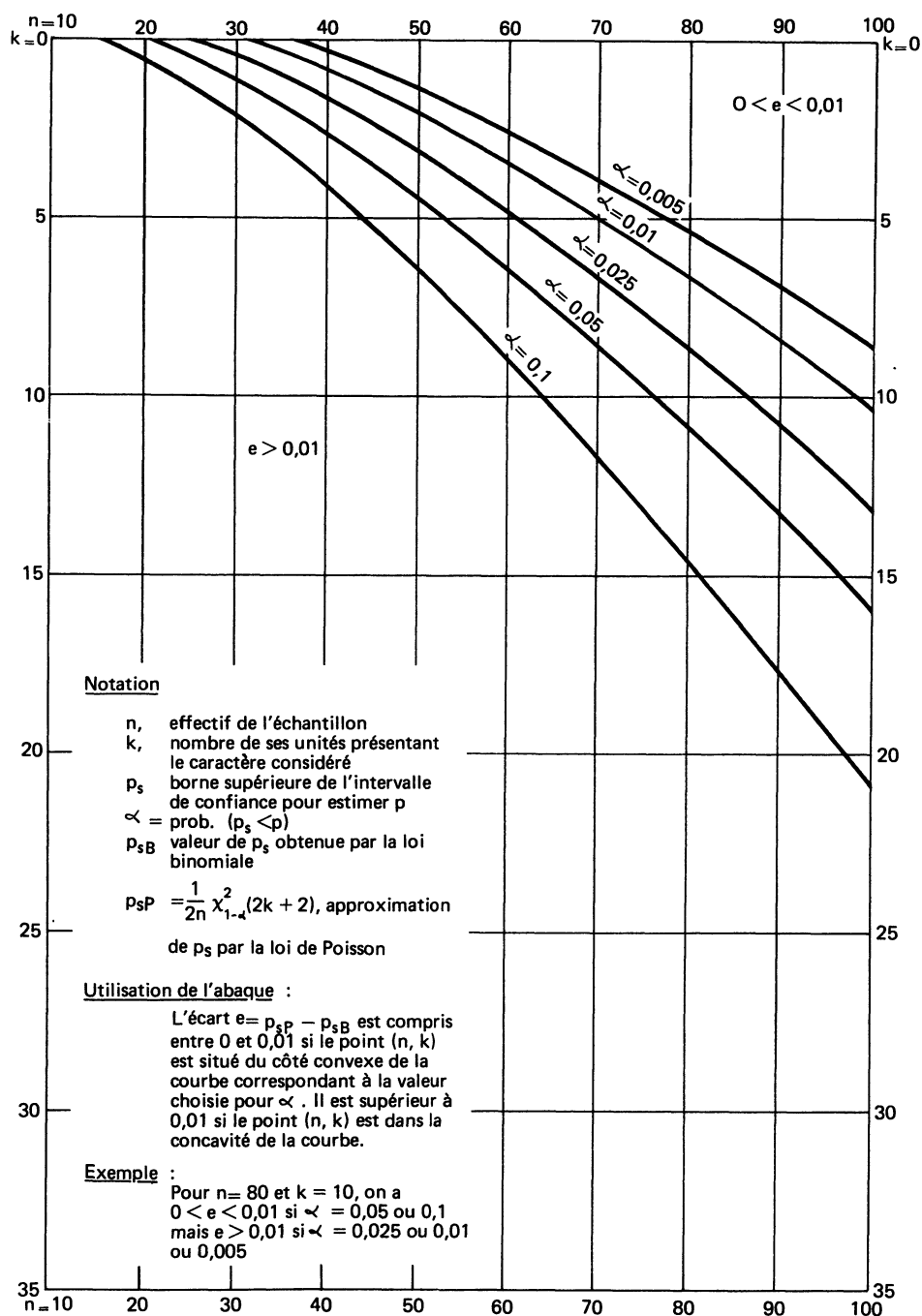
	Intervalle unilatéral (0 , p_s)	Int. bilatéral symétrique (p_i , p_s)	Intervalle unilatéral (p_i , 1)
p_{sB}^*	0,1971	0,2132	
p_{sP}	0,2062	0,2252	
$p_{sP} - p_{sB}$	0,0091	0,0120	
p_{iB}^*		0,0505	0,0577
p_{iP}		0,0493	0,0569
$p_{iP} - p_{iB}$		- 0,0012	- 0,0008

* Valeur obtenue par interpolation linéaire.

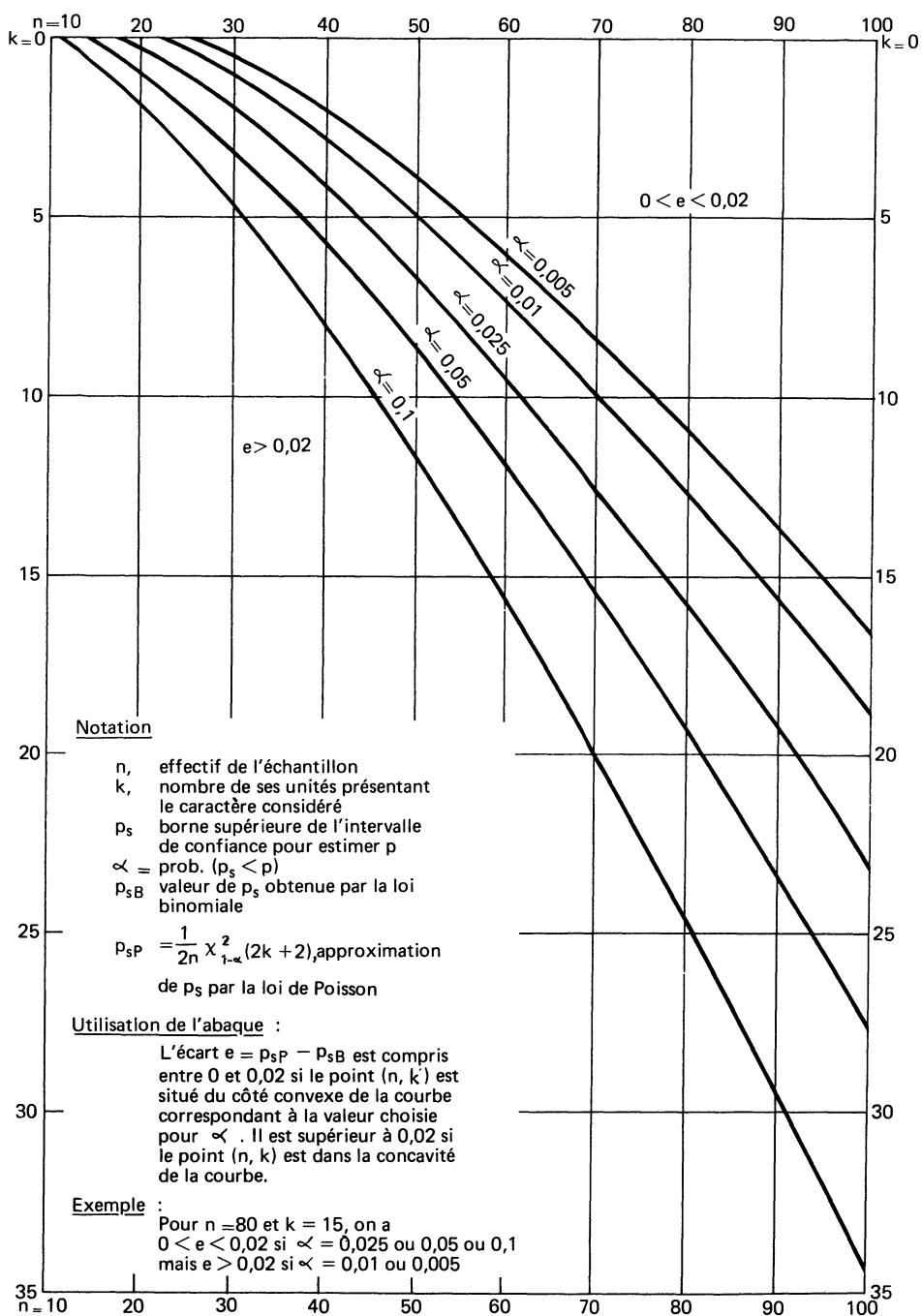
Les écarts entre p_{sP} et p_{sB} sont conformes aux conclusions que l'on peut déduire de l'abaque n° 1. Les écarts en p_{iP} et p_{iB} sont tout à fait négligeables.

BIBLIOGRAPHIE

- [1] CHARTIER et MORICE. — Relations entre quelques lois de probabilité. *Revue de Statistique Appliquée*, 1967, n° 4, p. 17/33.
- [2] MORICE et THIONET. — Loi binomiale et loi de Poisson. *Revue de Statistique Appliquée*, 1969, n° 3, p. 75/89.
- [3] Tables of the Binomial Probability Distribution. *Dept. of Commerce. Nat. Bureau of Standards*, pour $n = 2(1)49$.
- [4] H.G. ROMIG. — 50-100 Binomial Tables. *J. Wiley*, pour $n = 50(5)100$.



Abaque 1



Abaque 2