REVUE DE STATISTIQUE APPLIQUÉE

B. SORIN

Sur l'approximation gaussienne des queues de distributions binomiales

Revue de statistique appliquée, tome 20, nº 3 (1972), p. 89-96
<http://www.numdam.org/item?id=RSA_1972__20_3_89_0>

SUR L'APPROXIMATION GAUSSIENNE DES QUEUES DE DISTRIBUTIONS BINOMIALES

B. SORIN

Faculté des Sciences – 37 - TOURS

En statistique ou dans les applications numériques de problèmes de probabilités, il est fréquent d'avoir à calculer les queues de distributions binomiales du type $P(X \geqslant a)$ où X suit la loi binomiale $\mathcal{B}(n, p)$, a entier donné $0 \leqslant a \leqslant n$; nous noterons $B(a ; n, p)$ cette valeur. Le calcul de $P(X \leqslant a')$ se ramène au précédent puisque :

$$P(X \leqslant a') = 1 - P(X \geqslant a' + 1) = 1 - B(a' + 1 ; n, p)$$

Or, on ne possède pas toujours sous la main le tableau des distributions binomiales $\mathcal{B}(n, p)$ correspondant aux deux paramètres n et p désirés.

Toutefois si n est grand et p non voisin de 0 ou 1*, il est très classique, en vertu du théorème central-limite, d'approcher $P(X \geqslant a)$ par $P(Y \geqslant a)$ où Y est une variable aléatoire qui suit la loi de Laplace-Gauss de même moyenne np et de même variance npq $(q = 1 - p)$ ou mieux (correction dite de continuité) par : $P\left(Y \geqslant a - \frac{1}{2}\right)$. Cette approximation conduit à prendre pour évaluation de $B(a ; n, p)$ la valeur

$$\Phi\left(\left(np - a + \frac{1}{2}\right) \Big/ \sqrt{npq}\right)$$

où $\Phi(x)$ est la fonction de répartition de la loi de Laplace-Gauss réduite, tabulée dans nombre de manuels.

Notre but est ici d'étudier :

- le comportement de l'erreur exacte commise :

$$\Delta(a ; n, p) = \Phi\left(\left(np - a + \frac{1}{2}\right) \Big/ \sqrt{npq}\right) - B(a ; n, p)$$

en fonction de a pour des paramètres n et p donnés,

- l'évolution avec n et p du maximum sur a de la valeur absolue de l'erreur commise :

$$\delta(n, p) = \max_{0 \leqslant a \leqslant n} |\Delta(a ; n, p)|$$

----

* Auquel cas on utiliserait l'approximation poissonnienne ; on se reportera alors pour l'étude de l'erreur commise au travail de MM. Morice et Thionet [1].

Pour diverses études théoriques conduisant à des majorations de l'erreur, on pourra se reporter à [2], [3], [4].

I - EVOLUTION AVEC a de Δ (a ; n, p)

I-1 Symétrie : $\Delta(a\ ;\ n,\ p) = -\Delta(n+1-a,\ n,\ q)$

Soit $\frac{1}{2} < p \leqslant 1$; on a :

$$\Delta(a\ ;\ n,\ p) = \Phi\left(\left(np - a + \frac{1}{2}\right)\Big/\sqrt{npq}\right) - B(a\ ;\ n,\ p) = 1 - \Phi\left(\left(-np + a - \frac{1}{2}\right)\Big/\sqrt{npq}\right) =$$

$$-1 + B(n+1-a\ ;\ n,\ q) = E(n+1-a\ ;\ n,\ q) - \Phi\left(\left(nq - (n+1) - a) + \frac{1}{2}\right)\Big/\sqrt{npq}\right) =$$

$$-\Delta(n+1-a\ ;\ n,\ q).$$

Ainsi, pour des valeurs p complémentaires, les graphes de :

$$a \longrightarrow .\Delta$$

sont symétriques par rapport au point I $\left(\frac{n+1}{2}\ ,\ 0\right)$; on n'envisagera donc les $\Delta(a\ ;\ n,\ p)$ que pour des paramètres p compris entre 0 et $\frac{1}{2}$.

I-2 Nature de l'erreur commise

Pour des valeurs de n et p fixées, quelconques (dans le cadre susprécisé de l'approximation gaussienne) nous donnons, en figure 1, l'allure générale des graphes de :

1° $a \longrightarrow P(X \geqslant a)$ a entier

c'est-à-dire des probabilités exactes ; les valeurs $P(X \geqslant a)$ étant, par définition, rigoureusement égales à 1 lorsque $a < 0$ et rigoureusement nulles lorsque $a > n$

2° $a \longrightarrow \Phi\left(\left(np - a + \frac{1}{2}\right)\Big/\sqrt{npq}\right)$

c'est-à-dire des approximations gaussiennes des probabilités ci-dessus. Les chevauchements des deux graphes délimitent 4 zones, respectivement :

(1) et (3) de sous-estimation $(\Delta < 0)$

(2) et (4) de sur-estimation $(\Delta > 0)$.

Les graphes $a \longrightarrow \Delta$ vont donc présenter 4 oscillations ; à chaque zone correspondra un extremum relatif.

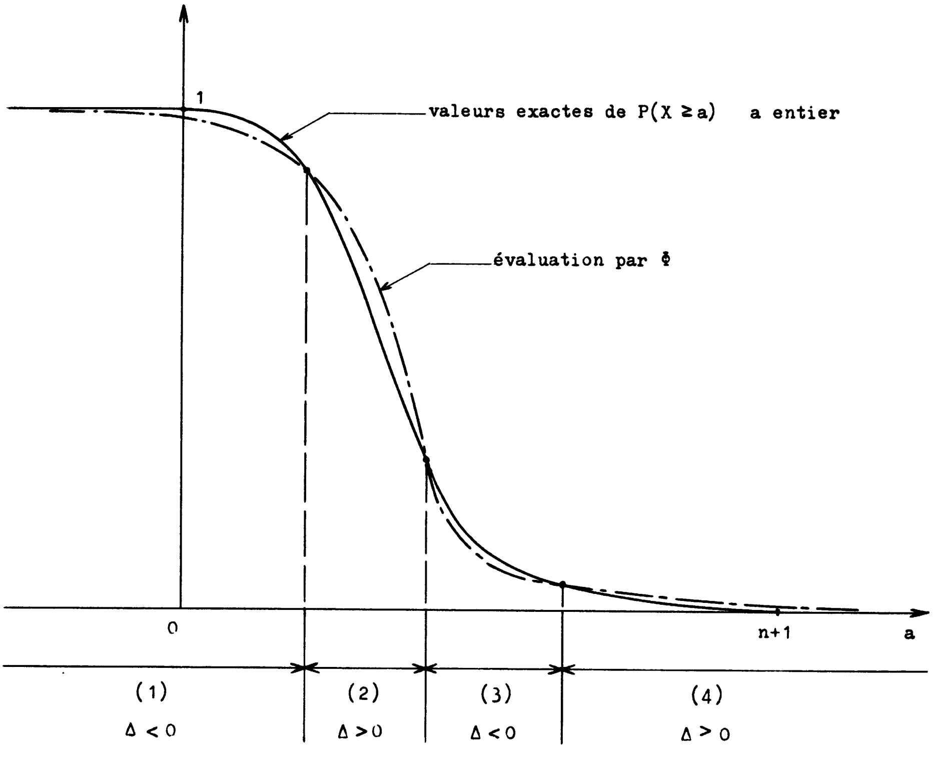

Figure 1 - Nature de l'erreur Δ commise en fonction de a.

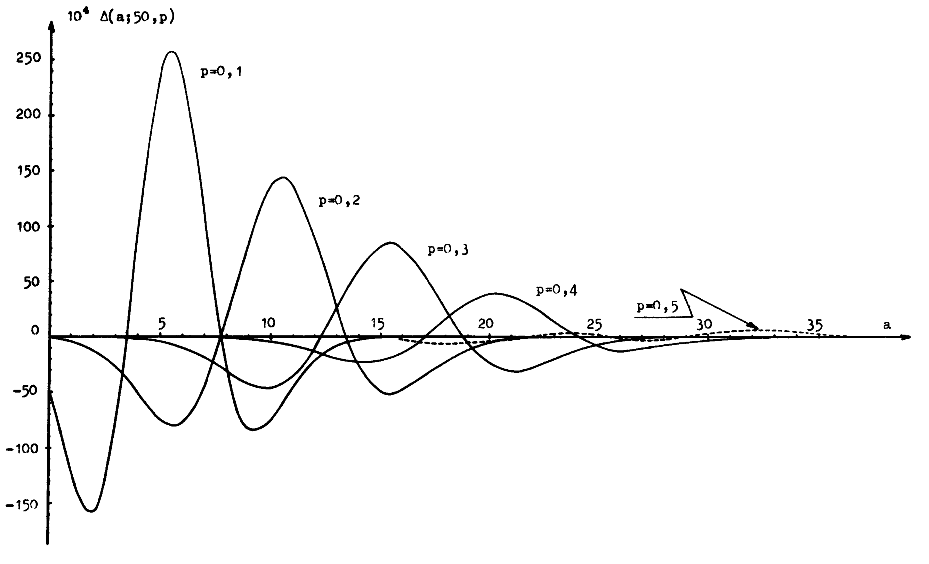

Figure 2 - Erreurs Δ (a ; 50, p) commises par l'approximation.

I-3 Graphes de Δ (a ; n, p), $p < \frac{1}{2}$.

En figure 2, pour n = 50 et p = 0,1 ; 0,2 ; 0,3 ; 0,4 ; 0,5 et à partir des tabulations exactes (à 10^{-5}) des lois binomiales (cf. [5]) et de la fonction Φ (cf. [6]), nous avons construit les graphes de $a \longrightarrow \Delta$ (a ; 50, p).

On peut en tirer les remarques suivantes :

A chaque zone correspond un maximum de $|\Delta|$ et, vers la moyenne np, dans la zone (2) [resp. (3)], quand $0 < p < \frac{1}{2}$ $\left[\text{resp. } \frac{1}{2} < p < 1\right]$ a lieu le maximum absolu. Seules exceptions, pour des valeurs de p très voisines de $\frac{1}{2}$ (en fait, quand n = 50, pour $\left|\frac{1}{2} - p\right| < 0{,}02$) où il a alors lieu en (1) [resp. (4)]. Hormis ces mêmes situations, dans la zone (4) [resp. (1)] les valeurs exactes et approchées sont négligeables à la précision utilisée ; il s'ensuit que Δ est pratiquement nulle (oscillation invisible).

Pour le cas limite $p = \frac{1}{2}$, le graphe de $a \longrightarrow \Delta\left(a ; n, \frac{1}{2}\right)$ présente des variations de très faibles amplitudes (approximation excellente) ; il est, en vertu de I-1, symétrique par rapport à $I\left(\frac{n+1}{2}, 0\right)$; il possède alors deux (faibles) extremums, respectivement en (1) et (4), de valeurs opposées.

Ces remarques sont générales, c'est-à-dire que, pour $0 < p < \frac{1}{2}$, un maximum absolu de $a \longrightarrow \Delta$ (a ; n, p) se présente au voisinage de la valeur moyenne np, excepté pour des valeurs de p très proches de $\frac{1}{2}$ (auquel cas l'approximation gaussienne est excellente) ; l'approximation, relativement au critère $\underset{a}{\text{Max}} |\Delta (a ; n, p)|$, est d'autant meilleure que n est grand et p voisin de $\frac{1}{2}$; elle se dégrade franchement lorsque p prend des valeurs très petites ou très proches de 1 (cf. note ci-dessus). Nous allons préciser quantitativement ceci au II.

II - EVOLUTION DE δ (n, p)

II-1 Décroissance de δ (n, p) à n donné et p variant de 0 à $\frac{1}{2}$

Posons :

$$\delta (n, p) = \underset{0 \leqslant a \leqslant n}{\text{Max}} |\Delta (a ; n, p)|$$

D'après I-1, on a :

$$\delta (n, p) = \delta (n, 1 - p).$$

δ (n, p) est une fonction décroissante de p pour $0 < p \leqslant \frac{1}{2}$ (croissante, symétriquement par rapport à $p = \frac{1}{2}$, pour $\frac{1}{2} \leqslant p < 1$) ; pour n = 50, nous avons, en figure 3, tracé son graphe dans un repère semi-logarithmique.

II-2 Décroissance de $\delta(n, p)$ à p donné pour n croissant

$\delta(n, p)$ est une fonction décroissante de n ; pour différentes valeurs de p : 0,1 (0,1) 0,9 -puisque $\delta(n, p) = \delta(n, 1-p)$- et $10 \leqslant n < 1000$, nous en avons tracé les graphes sur l'abaque figure 4.

Exemple d'utilisation de l'abaque 4

Soit p, $0 < p < 1$ et α, $0 < \alpha < 1$. A partir de quelle valeur N peut-on affirmer que l'erreur commise par évaluation gaussienne d'une queue de distribution binomiale $\mathcal{B}(n, p)$, $n \geqslant N$, est inférieure à α ?

Numériquement :

α = 0,015 ; 0,010 ; 0,005 et p = 0,1 ; 0,2 ; ... ; 0,9

Réponses :

Les valeurs de N sont sensiblement les suivantes :

α \ p	0,5	0,6 0,4	0,7 0,3	0,8 0,2	0,1 0,1
0,015	< 10	< 10	16	45	140
0,010	< 10	< 10	33	100	320
0,005	< 10	30	140	400	1 300

Valeurs N telles que $\delta(n, p) < \alpha$ pour $n \geqslant N$.

II-3 Valeurs N telles que $\delta\left(n, \frac{1}{m}\right) < \alpha$ pour $n \geqslant N$

Il est assez fréquent d'utiliser les distributions binomiales du type $\mathcal{B}\left(n, \frac{1}{m}\right)$ m entier modéré. Aussi avons-nous déterminé indépendamment de la figure 4 pour m = 2 ; 3 ; 4 ; 5 et α = 0,015 ; 0,010 ; 0,005 les valeurs N telles que $\delta\left(n, \frac{1}{m}\right) < \alpha$ pour $n \geqslant N$.

α \ m	2	3	4	5
0,015	3	10	28	45
0,010	4	24	$\simeq$ 60	$\simeq$ 100
0,005	6	$\simeq$ 88	$\simeq$ 240	$\simeq$ 400

Valeurs N telles que $\delta\left(n, \frac{1}{m}\right) < \alpha$ pour $n \geqslant N$.

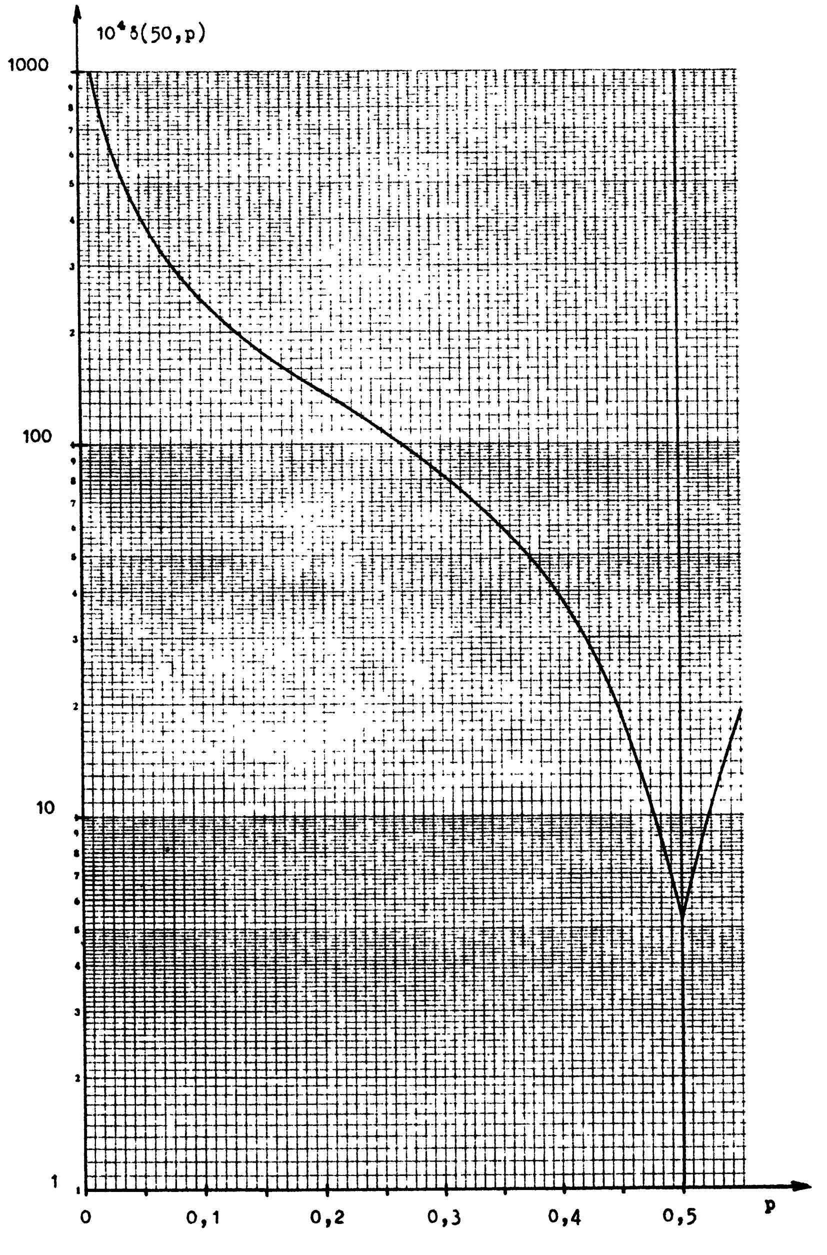

Figure 3 - Evolution de l'erreur maximum commise δ (50. p) en fonction de p.

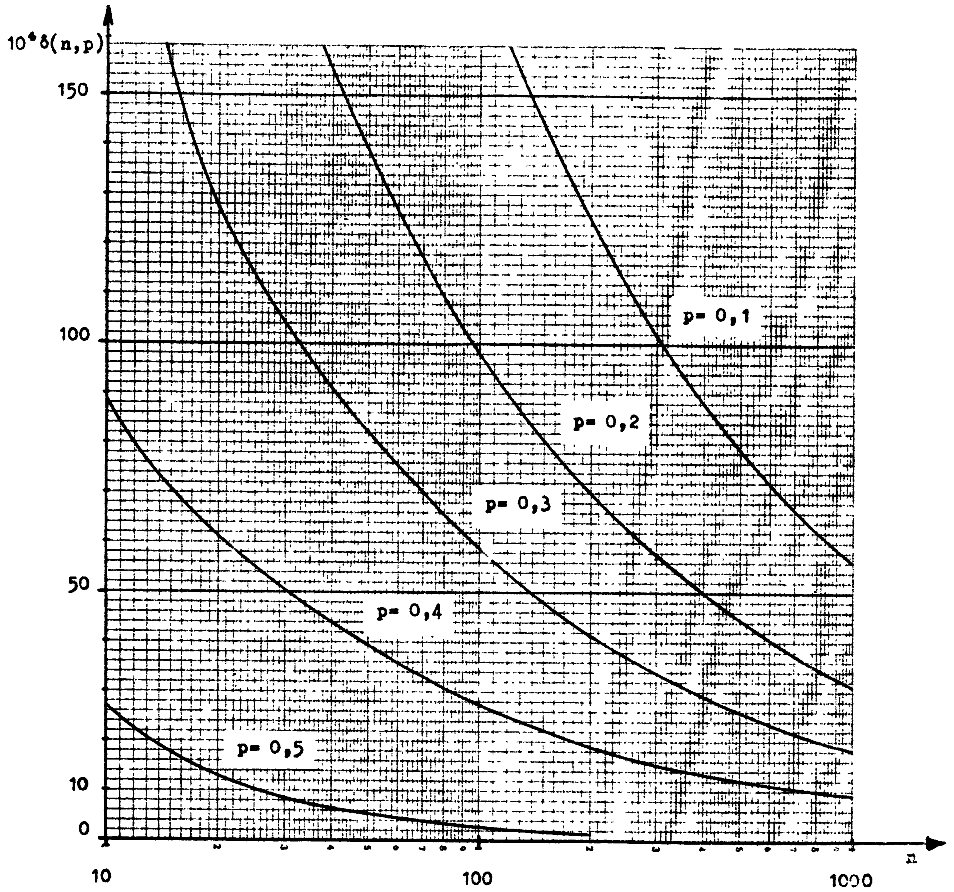

Figure 4 - Evolution de l'erreur maximum commise selon n pour diverses valeurs de p.

RESUME

Nous précisons ici la nature de l'erreur commise dans l'approximation gaussienne des queues de distributions binomiales $\mathcal{B}(n, p)$. Nous avons construit l'abaque qui, pour différentes valeurs de p, donne la décroissance de l'erreur maximum commise δ (n, p) en fonction de n.

REFERENCES

[1] MORICE E. et THIONET - Loi Binomiale et loi de Poisson , Rev. Stat. Appl., 1969, Vol. XVII, n° 3, p. 75-89.

[2] USPENSKY J.V. - Introduction to Mathematical Probability ; Mc Graw-Hill, 1937, p. 129.

[3] FELLER W. - On the Normal Approximation to the Binomial Distribution ; Ann. Math. Stat., 1945, Vol. 16, p. 319-329.

[4] SMITH E.S. - Binomial, Normal and Poisson Probabilities ; Ed. S. Smith, Bel Air, Maryland, 1953.

[5] XXXX - Tables of the Cumulative Binomial Probability Distribution ; Harvard University Press, 1955.

[6] BOLL M. - Tables Numériques Universelles ; Dunod, 1957.