REVUE DE STATISTIQUE APPLIQUÉE

HARRY R. LARSON

Un nomogramme de la distribution cumulative binomiale

Revue de statistique appliquée, tome 15, nº 2 (1967), p. 39-58
<http://www.numdam.org/item?id=RSA_1967__15_2_39_0>

© Société française de statistique, 1967, 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/

UN NOMOGRAMME
DE LA DISTRIBUTION CUMULATIVE BINOMIALE (1)

Harry R. LARSON

Dans un récent article, publié dans Industrial Quality Control, Larson a présenté un nomogramme à points alignés permettant de résoudre de nombreux problèmes liés à la distribution cumulative de la loi binomiale. On en trouvera ci-après une analyse détaillée.

La fonction cumulative de la loi binomiale $\mathcal{B}(n\ ,\ p)$:

$$\Pr(x \leqslant c) = \sum_{x=0}^{c} \frac{n!}{x!(n-x)!}\ p^x(1-p)^{n-x} = P \qquad (1)$$

est la probabilité d'observer un nombre, inférieur ou égal à c, d'éléments possédant un certain caractère, dans un échantillon de n éléments prélevés au hasard, avec remise, dans une population dans laquelle la proportion de tels éléments est égale à p.

Pratiquement, cette relation est valable dans le cas beaucoup plus fréquent d'un tirage sans remise, à condition que le rapport n/N soit petit $\left(\text{par exemple } \frac{n}{N} < \frac{1}{10}\right)$, N étant l'effectif de la population dans laquelle est fait le tirage de l'échantillon.

Dans la relation ci-dessus figurent 4 variables n, c, p et P : le nomogramme permet de déterminer l'un quelconque de ces 4 éléments lorsque les 3 autres sont donnés, il permet de plus de résoudre divers problèmes annexes.

Tous les problèmes seront résolus à l'aide de la droite joignant les points cotés p et P sur les deux échelles linéaires et passant par le point d'intersection des deux courbes n et c (Fig. 1).

On en trouvera divers exemples ci-après.

1/ <u>Probabilité P que dans un échantillon de n éléments provenant d'une population à deux catégories de fréquences respectives p, q = 1 − p, on en trouve un nombre x inférieur ou égal à c, appartenant à la catégorie p.</u>

Exemple : $n = 100$, $p = 0{,}10$, $c = 5$

---

(1) Traduction et adaptation par E. Morice de "A nomograph of the Cumulative binomial distribution" par H. R. Larson. Industrial Quality Control - Décembre 1966 pp. 270-278.

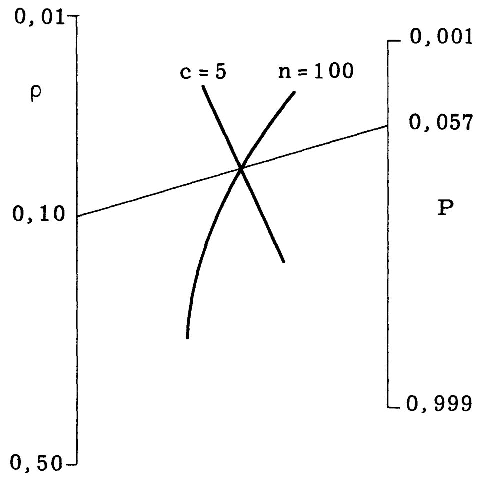


Figure 1

Le nomogramme donne :

$$P = 0,057 = \Pr[x \leqslant 5 \mid n = 100,\ p = 0,10]$$

on aurait de la même manière la solution des problèmes ayant comme inconnue n ou c ou p, satisfaisant à l'équation

$$P = \sum_{x=o}^{c} C_n^x\, p^x (1 - p^{\,n-x}$$

2/ Même problème pour $p < 0,01$. (Ces valeurs ne figurent pas sur l'échelle des valeurs de p).

Dans ce cas on peut utiliser la loi de Poisson comme approximation de la loi binomiale (approximation valable pour $p < 0,10$).

L'équation (1) s'écrit alors

$$P = \sum_{o}^{c} e^{-np} \frac{(np)^x}{x!}$$

Pour une valeur fixée de c, les valeurs associées de n et de p, sont liées par

$$np = n_o p_o$$

n_o et p_o pouvant être choisis arbitrairement de manière que leur produit soit égal à np, avec $p_o < 0\ 10$.

Exemple : $n = 100$, $p = 0,004$, $c = 1$

En prenant

$$n = 10 \quad , \quad p = 0,04 \quad , \quad c = 1$$

le nomogramme donne

$$P = 0,94$$

La table de Molina donne

$$P = 0,9384.$$

3/ Même problème pour $p > 0,50$ (ces valeurs ne figurent pas sur l'échelle des valeurs de p).

On sait que pour la loi binomiale on a la relation

$$\sum_{0}^{c} (n\ ,\ p) = 1 - \sum_{0}^{n-c-1} (n\ ,\ 1-p)$$

qui permettra d'utiliser le nomogramme en remplaçant p par $1 - p$ et c par $n - c - 1$.

Exemple : $n = 50 \quad , \quad p = 0,90 \quad , \quad c = 40$

d'où

$$n = 50 \quad , \quad p_1 = 1 - p = 0,10 \quad , \quad c_1 = n - c - 1 = 9$$

d'où

$$P = 1 - Pr[x \leqslant 40] = 1 - \sum_{0}^{9} (50 - 0,10) = 1 - 0,975 = 0,025$$

4/ Probabilité, pour n et p donnés, d'avoir exactement c éléments

On a

$$Pr[x = c] = Pr[x \leqslant c] - Pr[x \leqslant c - 1]$$

Pour $n = 100$, $p = 0,10$, on aura :

$$Pr[x = 4] = Pr[x \leqslant 5] - Pr[x \leqslant 4]$$
$$= 0,057 - 0,024 = 0,033$$

5/ Test de l'hypothèse $p = p_o$ à partir d'un échantillon (n , c)

Pour un test *bilatéral* , au risque $\alpha, \left(\frac{\alpha}{2} + \frac{\alpha}{2}\right)$, de rejeter l'hypothèse $p = p_o$ si elle est vraie, cette hypothèse sera acceptée si :

$$c' \leqslant c \leqslant c''$$

c' et c'' étant définis par :

$$Pr\left[\sum(n\ ,\ p_o\ ,\ c \leqslant c' - 1\right] \leqslant \frac{\alpha}{2}$$

$$Pr\left[\sum(n\ ,\ p_o\ ,\ c \leqslant c''\right] \geqslant 1 - \frac{\alpha}{2}$$

(les inégalités correspondant au fait que c' et c'' devant être entiers, on devra se contenter de solutions c', c'' donnant lieu à un risque total simplement voisin de α).

Exemple : Test bilatéral de l'hypothèse $p_o = 0,05$, au risque $\alpha = 0,05$, à partir d'un échantillon de $n = 80$.

Le nomogramme donne les solutions $c' = 0$ et $c'' = 8$ qui correspondent respectivement à des risques 0,018 et 0,02 (le risque réel est voisin de 0,04 au lieu de 0,05).

Pour un test *unilatéral* au risque α, l'unique région d'acceptation sera définie par une des conditions ci-dessus, en y remplaçant $\alpha/2$ par α.

6/ Puissance du test

La région de rejet de l'hypothèse $p = p_o$ étant constituée par les deux intervalles

$$0 < x < c' \qquad \text{et} \qquad c'' < x < n \qquad \text{(test bilatéral)}$$

ou par un seul de ces intervalles s'il s'agit d'un test unilatéral, la puissance du test, probabilité de rejet de l'hypothèse p_o, si la vraie valeur est p, est

$$\Pi = \sum_{x=o}^{c'-1} C_n^x p^x(1-p)^{n-x} + \sum_{x=c''+1}^{n} C_n^x p^x(1-p^{n-x}$$

$$= \sum_{o}^{c'-1} C_n^x p^x(1-p)^{n-x} + \left[1 - \sum_{o}^{c''} C_n^x p^x(1-p)^{n-x}\right]$$

$$= P' + [1 - P'']$$

dans le cas d'un test bilatéral (ou un seul des deux termes dans le cas d'un test unilatéral).

Les sommes P' et P'' figurant dans l'expression de Π seront données par le nomogramme avec, respectivement $c = c' - 1$ et $c = c''$.

7/ Intervalle de confiance à $1 - \alpha$ pour p après observation de c éléments, appartenant à la catégorie de fréquence p dans la population et observés dans un échantillon d'effectif n.

a) Intervalle unilatéral à $1 - \alpha$

S'il s'agit d'une limite supérieure p'' telle que l'on ait

$$Pr[x \leqslant c \mid p''] = P = \alpha$$

l'équation du problème est, avec cette définition généralement admise

$$\sum_{x=o}^{c} C_n^x p''^x(1-p'')^{n-x} = \alpha$$

De même dans le cas d'une limite inférieure p' telle que l'on ait

$$\Pr[x \geqslant c \mid p'] = \alpha$$

l'équation du problème est

$$\sum_{x=o}^{c-1} C_n^x p'^x (1 - p')^{n-x} = P = 1 - \alpha$$

Avec cette définition de l'intervalle de confiance, on notera que l'une des sommations est faite jusqu'à c − 1, l'autre jusqu'à c.

b) Intervalle bilatéral symétrique à $1 - \alpha$ $\left(\alpha = \frac{\alpha}{2} + \frac{\alpha}{2}\right)$

Les deux équations du problème sont alors

$$\sum_{o}^{c} (n , p'') = P = \frac{\alpha}{2}$$

$$\sum_{o}^{c-1} (n , p') = P = 1 - \frac{\alpha}{2}$$

Exemple : Intervalle de confiance bilatéral symétrique à 0,95 pour p, si n = 50, c = 5.

Pour $\frac{\alpha}{2} = 0,025$ on lit sur l'abaque :

$$p'' = 0,218 \qquad p' = 0,043$$

$$0,95 = \Pr[0,043 < p < 0,218]$$

8/ Détermination de n (effectif de l'échantillon) et c (critère d'acceptation pour un plan simple (α, p_1 , β , p_2).

Les valeurs de n et c sont solutions du système

$$\left\{\begin{array}{l} \sum_{x=o}^{c} C_n^x p_1^x (1 - p_1)^{n-x} = 1 - \alpha \\ \sum_{x=o}^{o} C_n^x p_2^x (1 - p_2)^{n-x} = \beta \end{array}\right.$$

L'emploi du nomogramme donnera directement n et c à l'intersection des droites [$p = p_1$, $P = 1 - \alpha$ et $p = p_2$, $P = \beta$], (Fig. 2).

Exemple : $\alpha = 0,05$, $p_1 = 0,02$ $\beta = 0,10$, $p_2 = 0,08$

Le nomogramme donne (Fig. 2) par interpolation graphique c = 4 (valeur entière par excès) et n ∼ 98 correspondant à des risques réels α' # 0,04, β' # 0,11.

Le même mécanisme s'applique évidemment à la détermination d'un plan d'échantillonnage pour acceptation d'un lot d'éléments appartenant à une population dont la durée de vie est distribuée exponentiellement (essais sans remplacement des défaillants, de durée préfixée T, (essai tronqué)).

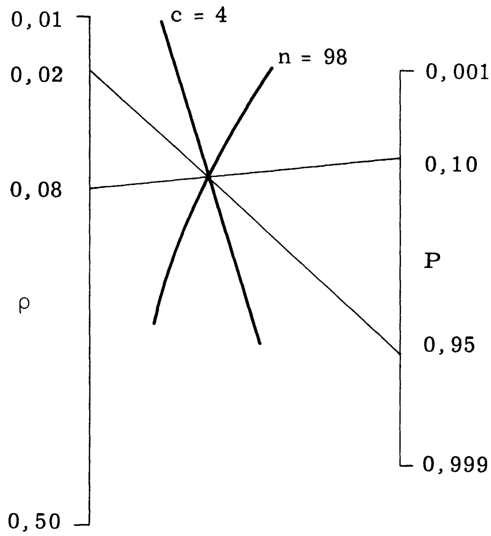


Figure 2

Si θ est la durée moyenne de vie, caractérisant la loi exponentielle et si β et 1 − α sont les probabilités d'acceptation correspondant respectivement à des valeurs θ_2 et θ_1 ($\theta_2 < \theta_1$), l'effectif n de l'échantillon et le critère d'acceptation c seront définis par

$$P_a(\theta_2) = \sum_{x=0}^{c} C_n^x [1 - e^{-T/\theta_2}] \ (e^{-T/\theta_2})^{n-x} = \beta$$

$$P_a(\theta_1) = \sum_{x=0}^{c} C_n^x [1 - e^{-T/\theta_1}] \ (e^{-T/\theta_1})^{n-x} = 1 - \alpha$$

de la même forme que les précédentes avec $p_1 = 1 - e^{-T/\theta_1}$, $p_2 = 1 - e^{-T/\theta_2}$, valeurs données en fonction de T/θ par les tables de la fonction e^{-x}.

Exemple

$$\alpha = 0,05 \qquad T/\theta_1 = \frac{1}{20} \quad , \quad \beta = 0,10 \quad , \quad T/\theta_2 = \frac{3}{20}$$

d'où

$$p_1 = 0,0487 \quad , \quad p_2 = 0,1393$$

$$P_1 = 0,95 \quad , \quad P_2 = 0,10$$

Le nomogramme donne

$$c = 8 \qquad n \sim 85$$

(les tableaux publiés par H. 108 [11], pour un nombre limité de valeurs α, β, T/θ_1, T/θ_2, donnent c = 8, n = 83 pour l'exemple ci-dessus.

9/ Courbe d'efficacité d'un plan d'échantillonnage simple

Pour un échantillon d'effectif n et un critère d'acceptation c, la probabilité d'acceptation P_a d'un lot contenant une fraction p de défectueux est

$$P_a = \sum_{x=c}^{c} C_n^x p^x (1 - p)^{n-x}$$

n et c étant fixés, les valeurs correspondantes de p et P se liront sur une droite pivotant autour du point (n, c) du réseau de courbes du nomogramme.

10/ Carte de contrôle du pourcentage de défectueux

Si p_0 est le pourcentage de défectueux correspondant à un processus sous contrôle, les limites inférieure et supérieure au risque bilatéral 0,002 = 0,001 + 0,001 pour des échantillons d'effectif constant n seront les valeurs $p' = \frac{c'}{n}$ et $p'' = \frac{c''}{n}$ telles que

$$\sum_{0}^{c'} C_n^x p_0^x (1 - p_0)^{n-x} = 0,001$$

$$\sum_{0}^{c''} C_n^x p_0^x (1 - P_0)^{n-x} = 0,999$$

Le nomogramme donnera c' et c'' d'où p' et p''.

Exemple

$$n = 100 \qquad p_0 = 0,12$$

Le nomogramme donne c' = 2,6 et c'' = 22,5, d'où p' = 0,026 et p'' = 0,225 (ces résultats sont d'ailleurs très voisins de ceux que donne l'approximation normale

$$p_0 \pm 3,09 \sqrt{\frac{p_0(1 - p_0)}{n}} = (0,02, \ 0,22))$$

11/ Test des signes

C'est un cas particulier d'application de la loi binomiale dans le cas particulier p_0 = 0,50, ce test est utilisé dans la comparaison par paires :

Chaque paire d'observations d'un même phénomène sur deux unités , dont on veut tester qu'il n'y a pas lieu de supposer qu'il existe entre elles de différence systématique, donnera lieu à deux observations qualitatives ou quantitatives dont la différence pourra être affectée du signe + ou du signe –.

Si n est le nombre des paires observées, c le nombre des différences positives, par exemple, la probabilité d'obtenir un nombre inférieur ou égal à c est, dans l'hypothèse p_0 = 1/2

$$P = \sum_{0}^{c} C_n^x \left(\frac{1}{2}\right)^n$$

Exemple $n = 14 \quad , \quad c = 3$

Pour $p_o = 0,5$ le nomogramme donne $P = 0,026$ qui conduit à rejeter l'hypothèse d'homogénéité dans un test bilatéral au risque $\alpha = 0,05$.

12/ Comparaison de deux proportions observées

Si c_1 et c_2 sont les nombres d'éléments observés sur des échantillons d'effectifs n_1 et n_2, d'où $f = \frac{c_1}{n_1}$, $f_2 = \frac{c_2}{n_2}$, le test de l'hypothèse $p_1 = p_2 = p$ dans les deux populations peut généralement, si p n'est pas trop petit et si n_1, n_2 sont assez grands (n_1p, $n_2p > 10$), être fait à l'aide de la variable normale réduite

$$u = \frac{f_2 - f_1}{\sqrt{p'(1 - p')\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

avec

$$p' = \frac{c_1 + c_2}{n_1 + n_2}$$

c_1	c_2
$n_1 - c_1$	$n_2 - c_2$
n_1	n_2

Si c_1 par exemple, est inférieur à la valeur c_1' qui correspondrait à l'homogénéité des deux échantillons

$$\frac{c_1'}{n} = \frac{c_1 + c_2}{n_1 + n_2},$$

le test "exact" de Fisher-Yates conduit à refuser l'hypothèse au risque unilatéral α , si

$$\sum_{x=0}^{c_1} \frac{C(n_1, x)\, C(n_2, c_1 + c_2 - x)}{C(n_1 + n_2, c_1 + c\ )} < \alpha$$

avec

$$C(n_1, x) = \frac{n_1!}{x!(n_1 - x)}$$

$$C(n_2, c_1 + c_2 - x) = \frac{n_2!}{(c_1 + c_2 - x)!\,(n_2 - c_1 - c_2 + x)!}$$

$$C(n_1 + n_2, c_1 + c_2) = \frac{(n_1 + n_2)!}{(c_1 + c_2)!\,(n_1 + n_2 - c_1 - c_2)!}$$

Ce calcul relativement facile si c_1, c_2, n_1, n_2 sont petits devient rapidement très laborieux, mais on peut remarquer que l'approximation binomiale de la loi hupergéométrique permet, si $\frac{c_1 + c_2}{n_1 + n_2} < 0,10$ de remplacer la condition ci-dessus par la condition :

$$\sum_{x=0}^{c_1} C^x_{c_1+c_2}\left(\frac{n_1}{n_1 + n_2}\right)^x \left(\frac{n_2}{n_1 + n_2}\right)^{c_1+c_2-x} < \alpha$$

relative à la loi binomiale

$$n = c_1 + c_2 \qquad p_o = \frac{n_1}{n_1 + n_2}$$

dont la solution est donnée, comme ci-dessus, par le nomogramme.

Exemple $n_1 = n_2 = 50$ $\quad c_1 = 2$ $\quad c_2 = 6$, $\left(\frac{2 + 6}{100} = 0,08 < 0,10\right)$

$n = 8$ $\quad p_o = 0,50$ $\quad c = 2$

Le nomogramme donne $P = \sum_0^2 C^x_8(0,5)^8 = 0,13$

Si on s'est fixé $\alpha = 0,05$, on acceptera l'hypothèse $p_1 = p_2$.

L'approximation normale - non valable dans ce cas où $n_1p' = n_2p' = 4$, appliquée à la comparaison de p_1 et p_2 à partir de $f_1 = \frac{2}{50} = 0,04$ et $f_2 = \frac{6}{50} = 0,12$ d'où $p' = 0,08$, donnerait

$$u = \frac{0,12 - 0,04}{\sqrt{0,08 \times 0,92\left(\frac{1}{50} + \frac{1}{50}\right)}} = 1,47,$$

valeur qui a une probabilité 0,07 d'être dépassée.

Le test de Fisher-Yates donne 0,13.

(On notera que, si dans ce cas $\alpha = 0,05$, l'approximation normale conduit à la même conclusion (différence non significative) que le test de Fisher-Yates ou le nomogramme, il pourrait ne pas en être ainsi avec d'autres valeurs de n_1, c_1, n_2, c_2 ou avec une autre valeur de α, par exemple $\alpha = 0,10$).

13/ Probabilités cumulées de la variable F de Fisher-Snedecor.

La relation :

$$\Pr[x \leqslant c] = \sum_{x=0}^{c} C^x_n\, p^x(1 - p)^{n-x} = \alpha$$

est équivalente à

$$\Pr\left[F > \frac{n - c}{c + 1}\,\frac{p}{1 - p}\right] = \alpha$$

soit

$$F_{1-\alpha} = \frac{n - c}{c + 1}\,\frac{p}{1 - p} \qquad \begin{cases} \nu_1 = 2(c + 1) \\ \nu_2 = 2(n - c) \end{cases}$$

ou

$$p = \frac{(c + 1)F_{1-\alpha}}{n - c + (c + 1)\, F_{1-\alpha}}$$

Pour une valeur observée de F, les éléments de la loi binomiale correspondante sont donc la valeur de p ci-dessus et :

$$c = \frac{\nu_1}{2} - 1$$

$$n = \frac{\nu_1 + \nu_2}{2} - 1 \qquad \text{(pour } \nu_1 \text{ et } \nu_2 \text{ pairs)}$$

Le nomogramme permet alors de déterminer la valeur de α telle que

$$\alpha = \Pr\left[F > \frac{n - c}{c + 1}\,\frac{p}{1 - p}\right]$$

Exemple - On a calculé, dans une comparaison de deux variances (échantillons provenant de deux populations normales), une valeur F = 2,08 avec $\nu_1 = 10$ et $\nu_2 = 40$ degrés de liberté.

On a :

$$c = 4, \quad n = 24, \quad p = \frac{5 \times 2,08}{20 + 5 \times 2,08} = \frac{2,08}{6,08} = 0,342.$$

Le nomogramme donne

$$\alpha = P \# 0,05 = \Pr[F_{(10,\ 40)} > 2,08]$$

Le nomogramme permet donc de compléter les tables généralement assez limitées dont on dispose pour la variable F.

Pour une valeur $F_{(\nu_1 . \nu_2)} < 1$, il n'y a aucun changement :

Pour	$\nu_1 = 10$,	$\nu_2 = 40$	$F = 0,375$
	$c = 4$	$n = 24$	$p = 0,0855$

Le nomogramme donne

$$\alpha = P \# 0,95 = \Pr[F_{(10,\ 40)} > 0,375]$$

14/ Limites de tolérance d'un processus

Wilks a démontré [6] que, pour une distribution continue, quelle que soit sa forme, l'effectif minimum n d'échantillon nécessaire pour que l'étendue de cet échantillon contienne au moins une large fraction β de la population, avec un coefficient de confiance élevé α, était défini par :

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

Cette équation peut s'écrire

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

soit

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

avec

$$p = 1-\beta$$

dont la solution en n est donnée par le nomogramme pour

$$c = 1 \qquad p = 1-\beta \qquad P = 1-\alpha$$

Exemple

$$\beta = 0,95 \qquad \alpha = 0,95$$

soit

$$c = 1 \qquad p = 0,05 \qquad P = 0,05$$

Le nomogramme donne $n = 92$, effectif tel que l'on peut dire au risque 0,05 que 95 % au moins des unités de la population sont contenues dans l'étendue de cet échantillon au hasard.

(Diverses solutions approchées (1) de l'équation transcendante ci-dessus ont été proposées, par exemple

$$n \sim \frac{1}{4}\ \chi^2_\alpha(\nu = 4)\ \frac{1+\beta}{1-\beta}$$

soit

$$n \sim \frac{1}{4} \times 9,49 \times \frac{1,95}{0,05} = 92.$$

Inversement, le nomogramme permet de déterminer $\beta = 1 - p$, lorsque n est donné ainsi que α.

Exemple - Déterminer avec un coefficient de confiance $\alpha = 0,90$, soit $P = 0,01$ la fraction minimale β de la population que l'on peut espérer dans l'étendue d'un échantillon de $n = 100$.

Avec $n = 100$ $P = 0,01$ $c = 1$, le nomogramme donne

$p = 0,065$ soit $\beta = 0,935$

Le calcul (cf N° 13) donne

$$p = \frac{2F_{0,99}}{99 + 2F_{0,99}}, \qquad \begin{cases} \nu_1 = 4 \\ \nu_2 = 198 \end{cases}$$

$$p = \frac{6,82}{105,82} = 0,0644$$

15/ Précision du nomogramme

Dessiné sur une feuille 21 × 27, le nomogramme permet de lire p avec une erreur relative maximale variant de $\frac{1}{100}$ (pour p = 0,50) à $\frac{1}{20}$ (pour p = 0,01) et de lire P avec une erreur relative variant de $\frac{1}{100}$ (pour P = 0,50) à $\frac{1}{10}$ (pour P = 0,01 ou 0,99).

L'imprécision dans l'emploi provient surtout de la lecture des valeurs de n et de c dans le réseau de courbes. Cependant une expérience portant sur 110 utilisations du nomogramme pour déterminer P en fonction de (n, c, p) à montré que l'erreur médiane sur P varie de 0,0007 $\left(\text{pour } P = 0,01, \text{ soit } \frac{7}{1000}\right)$ à 0,01 pour $\left(P = 0,50, \text{ soit } \frac{1}{50}\right)$.

La précision d'ensemble du nomogramme paraît suffisante en pratique pour les problèmes faisant intervenir la loi binomiale.

ANNEXE

APPLICATION A LA LOI DE POISSON

Dans le cas d'une loi de Poisson correspondant alors à l'équation :

$$\Pr[x \leqslant c] = \sum_{x=o}^{c} e^{-m} \frac{m^x}{x!} = P \qquad (2)$$

certains des problèmes envisagés ci-dessus peuvent être résolus à l'aide du nomogramme établi par M. Cavé [9] pour la détermination d'un plan d'échantillonnage simple (α, p_1, β, p_2) en utilisant l'approximation de la loi binomiale par la loi de Poisson, soit :

$$\sum_{x=o}^{c} C_n^x\, p^x (1-p)^{n-x} \sim \sum_{x=o}^{o} e^{-np} \frac{(np)^x}{x!} = P$$

pratiquement valable si $p < 0,10$.

Pour l'exemple traité au § 8, l'abaque Cavé donne c = 4, n = 100.

Mais le nomogramme de M. Cavé est limité à $0 \leqslant c \leqslant 24$, (il serait évidemment facile de l'étendre), d'autre part l'échelle de P est remplacée par un réseau discontinu de courbes P = Cte, correspondant à quelques valeurs usuelles de $1 - \alpha$ et de β, réseau qui n'offre pas les mêmes facilités d'interpolation qu'une échelle linéaire continue.

Le nomogramme de Larson, offre une gamme plus étendue de solutions suffisamment approchées de l'équation (2), en posant $np_o = m$, p_o étant choisi arbitrairement petit, inférieur à 0,1, par exemple $p_o = 0,01$ d'où $n = \frac{m}{p_o} = 100\ m$ (on pourrait tout aussi bien choisir $p_o = 0,05$, d'où $n = 20\ m$ ou toute autre valeur $p < 0,10$, associée à la courbe $n = \frac{m}{p_o}$).

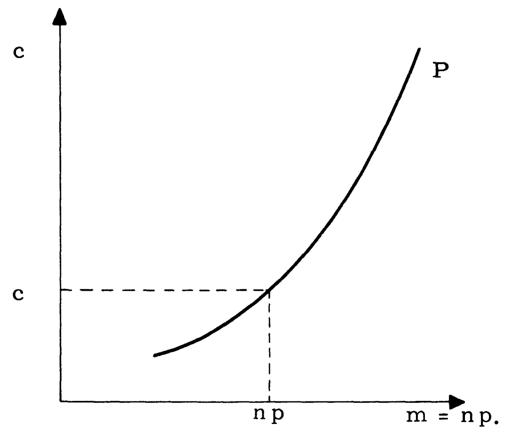


Figure 3

Dans ces conditions l'échelle p du nomogramme de Larson se réduit à un point p_o choisi pour un problème particulier, aussi petit que possible de manière à être compatible avec le tracé du réseau de courbes.

De ce point rayonneront toutes les droites utilisées dans la solution graphique des problèmes liés directement à la seule équation (2).

Exemples

1/ Calcul de $P = \sum_{o}^{10} e^{-17} \frac{(17)^x}{x!}$

En prenant $p_o = 0,05$, d'où $n = \frac{17}{0,05} = 340$, le nomogramme pour $c = 10$, donne

$$P \sim 0,05$$

(La table de la loi de Poisson donne P = 0,0491).

2/ Intervalle de confiance bilatéral à 0,90 pour le paramètre m si on a observé c = 6.

L'intervalle $m' < m < m''$ est défini par

$$\begin{cases} \sum_{x=o}^{5} e^{-m'} \frac{(m')^x}{x!} = 0,95 \\ \sum_{x=o}^{6} e^{-m''} \frac{(m'')^x}{x!} = 0,05 \end{cases}$$

En prenant comme pivot arbitraire $p_o = 0,05$, on trouve (fig. 4)

pour	c = 5	et	P = 0,95	n = 51	$m' = 51 \times 0,05 = 2,55$
	c = 6		P = 0,05	n = 230	$m'' = 230 \times 0,05 = 11,5$

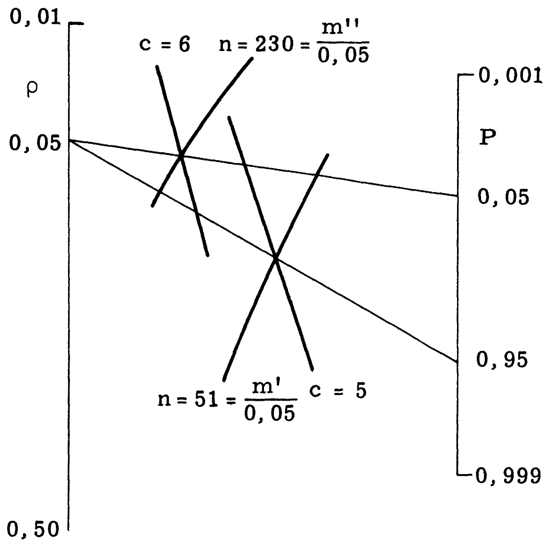


Figure 4

La table de Molina [5] donne

$$\sum_{0}^{5} (m' = 2,55) = 0,955$$

$$\sum_{0}^{6} (m'' = 11,5) = 0,06$$

3/ <u>Déterminer la valeur de c, telle que</u>

$$\sum_{x=0}^{c} e^{-8} \frac{8^x}{x!} = 0,10$$

En utilisant le même pivot $p_0 = 0,05$, ce qui correspond à $n = \frac{m}{p_0} = \frac{8}{0,05} = 160$ le nomogramme donne c = 4.

La table de la loi de Poisson donne

$$\sum_{0}^{4} e^{-8} \frac{8^x}{x!} = 0,0996$$

4/ <u>Test bilatéral au risque 0, 10 = 0, 05 + 0, 05 de l'hypothèse m = 15.</u>

L'intervalle d'acceptation $c' \leqslant c \leqslant c''$ est défini par les conditions :

$$\sum_{0}^{c'-1} e^{-15} \frac{15^x}{x!} \leqslant 0,05$$

$$\sum_{0}^{c''} e^{-15} \frac{15^x}{x!} \geqslant 0,95$$

d'où, avec $p_0 = 0,05$, $n = \frac{15}{0,05} = 300$. on trouve les valeurs (Fig. 6)

$$c' = 9 \qquad c'' = 22$$

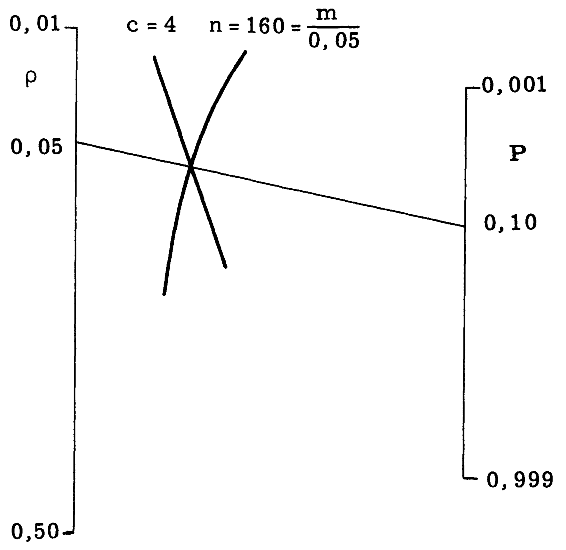


Figure 5

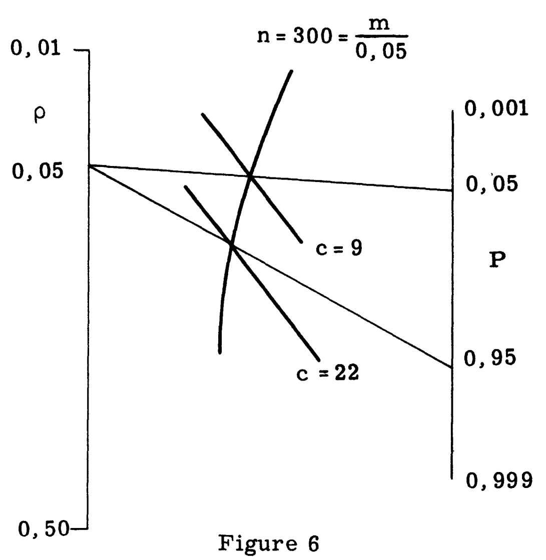


Figure 6

Pour m = 15, la table de la loi de Poisson donne :

$$\sum_{0}^{8} = 0,037 \qquad \sum_{0}^{22} = 0,967$$

5/ <u>Déterminer n et c (p_1, p_2, P_1, P_2 étant fixés) pour que l'on ait simultanément</u>

$$\sum_{x=0}^{c} e^{-np_1} \frac{(np_1)^x}{x!} = P_1$$

$$\sum_{x=0}^{c} e^{-np_2} \frac{(np_2)^x}{x!} = P_2$$

Pour $p_o = 0,05$ par exemple, ce système est pratiquement équivalent à :

$$\sum_{x=0}^{c} C_{n_1}^x p_o^x(1 - p_o)^{n_1-x} = P_1$$

$$\sum_{x=0}^{c} C_{n_2}^x p_o^x(1 - p_o)^{n_2-x} = P_2$$

avec

$$\left.\begin{array}{l} np_1 = n_1p_o \\ np_2 = n_2p_o \end{array}\right\} \frac{n_1}{n_2} = \frac{p_1}{p_2}$$

Les solutions en n_1, n_2 et c doivent être telles que les deux droites p_oP_1 et p_oP_2 coupent une même courbe c en deux points A et B situés sur deux courbes n_1, n_2, telles que $\frac{n_1}{n_2} = \frac{p_1}{p_2}$.

L'emploi d'un transparent sur lequel on aura tracé les deux droites p_oP_1 et p_oP_2 permettra, assez aisément, de déterminer c, n_1, n_2 d'où deux approximations de n (Fig. 7).

$$n \sim n_1 \frac{p_o}{p_1} \qquad n = n_2 \frac{p_o}{p_2}$$

dont on pourra, par exemple, prendre la moyenne.

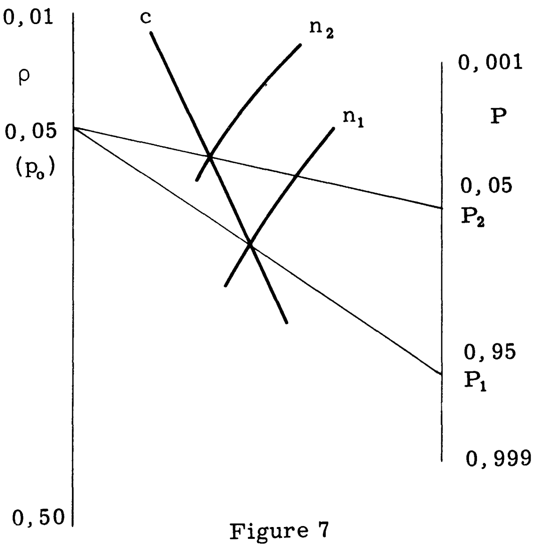


Figure 7

Application - Détermination d'un plan d'échantillonnage pour acceptation d'un lot d'éléments appartenant à une population dont la durée de vie est distribuée exponentiellement (essai avec remplacement des défaillants, de durée préfixée T ; essai tronqué).

$1 - \alpha$ et β étant respectivement les probabilités d'acceptation du lot correspondant aux valeurs θ_1 et θ_2, $(\theta_2 < \theta_1)$, de la durée de vie moyenne θ, l'effectif n, maintenu constant par remplacement des défaillants pendant la durée de l'essai, et le critère d'acceptation c seront définis par

$$P_a(\theta_1) = \sum_{x=0}^{c} e^{-nT/\theta_1} \frac{(nT/\theta_1)^x}{x!} = P_1 = 1 - \alpha$$

$$P_a(\theta_2) = \sum_{x=0}^{o} e^{-nT/\theta_2} \frac{(nT/\theta_2)^x}{x!} = P_2 = \beta$$

de la forme indiquée ci-dessus avec :

$$p_1 = \frac{T}{\theta_1} \quad , \quad p_2 = \frac{T}{\theta_2}$$

En posant

$$\begin{aligned} np_1 &= n_1 p_o \\ np_2 &= n_2 p_o \end{aligned} \qquad \text{d'où} \qquad \frac{n_1}{n_2} = \frac{p_1}{p_2} = \frac{T/\theta_1}{T/\theta_2} \ ,$$

on est ramené au problème précédent.

<u>Exemple</u> : Pour $\alpha = 0,10$, $T/\theta_1 = \frac{1}{5}$, $\beta = 0,05$, $T/\theta_2 = \frac{2}{5}$, en prenant $p_o = 0,05$, les deux droites $(p = 0,05 \ P = 0,05)$ et $(p = 0,05 \ P = 0,90)$, coupent la droite $c = 18$, respectivement sur les courbes $n_1 \sim 275$ et $n_2 \sim 520$ telles que $\frac{n_1}{n_2} \sim \frac{1}{2}$, d'où les deux valeurs estimées de n :

$$\left.\begin{aligned} n &= \frac{n_1 p_o}{p_1} = 275 \times 0,05 \times 5 = 69 \\ n &= \frac{n_2 p_o}{p_2} = 520 \times 0,05 \times \frac{5}{2} = 65 \end{aligned}\right\} \text{soit } n \sim 67$$

(résultat voisin de celui donné par H. 108 [11] soit n = 64).

Le nomogramme permet de déterminer c à une unité près, par contre l'erreur sur n peut être beaucoup plus importante.

Le recours à une table de Molina [5] permettra de préciser c et n avec un minimum de tâtonnements.

6/ <u>Probabilités cumulées de la variable $\chi^2(\nu)$</u>

On a la relation

$$\alpha = \sum_{o}^{c} e^{-m} \frac{m^x}{x!} = \Pr[\chi^2 > 2m, \ \nu = 2(e + 1)]$$

d'où

$$2m = \chi^2_{1-\alpha} \quad , \quad \nu = 2(c + 1)$$

Pour une valeur de χ^2 résultant de l'observation, les éléments de la loi binomiale correspondante sont $c = \nu/2 - 1$ et, en prenant $p_o = 0,05$ comme pivot, $m = \frac{1}{2}\ \chi^2$, soit $n = \frac{\chi^2}{2p_o} = 10\ \chi^2$, d'où la valeur de α donnée par le nomogramme pour $p = 0,05$, $n = 10\ \chi^2$, $c = \frac{\nu}{2} - 1$.

Exemple - On a calculé $\chi^2 = 25$ avec $\nu = 16$ degrés de liberté.

Pour $p = 0,05$, $c = 7$, $n = 250$, le nomogramme donne (Fig. 8)

$$= P = 0,08 = \Pr[\ ^2 > \ 25 \ \ , \ \ \nu = 16]$$

(Pour $\nu = 16$, la table de χ^2 donne

$$0,10 = \Pr[\chi^2 > 23,5]$$
$$0,05 = \Pr[\chi^2 > 26,3])$$

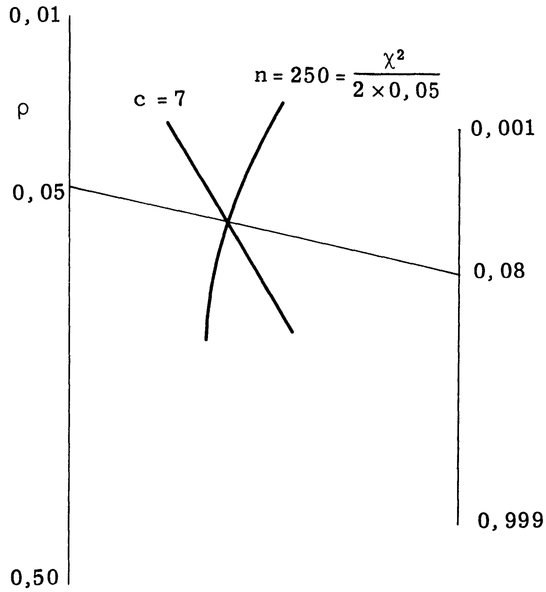


Figure 8

BIBLIOGRAPHIE

[1] ALICOCK, JONES and MICHEL - The nomogram. Pitman Pub. Corp. New-York, 1963, Ch. VI.

[2] NELSON - Nomograph for two sided distribution free tolerance intervals. Industrial Quality Control, Juin 1963.

[3] Tables of the cumulative binomial probability distribution. Harvard University Press 1955.

[4] ROMIG - 50-100 Binomial Tables. Wiley.

[5] MOLINA - Poisson's exponential binomial limits. Van Nostrand, New-York.

[6] EISENHART, HASTAY and WALLIS - Selected techniques in Statistical Analysis. Mac-Graw Hill Book C°, New-York 1947, pp. 249/253.

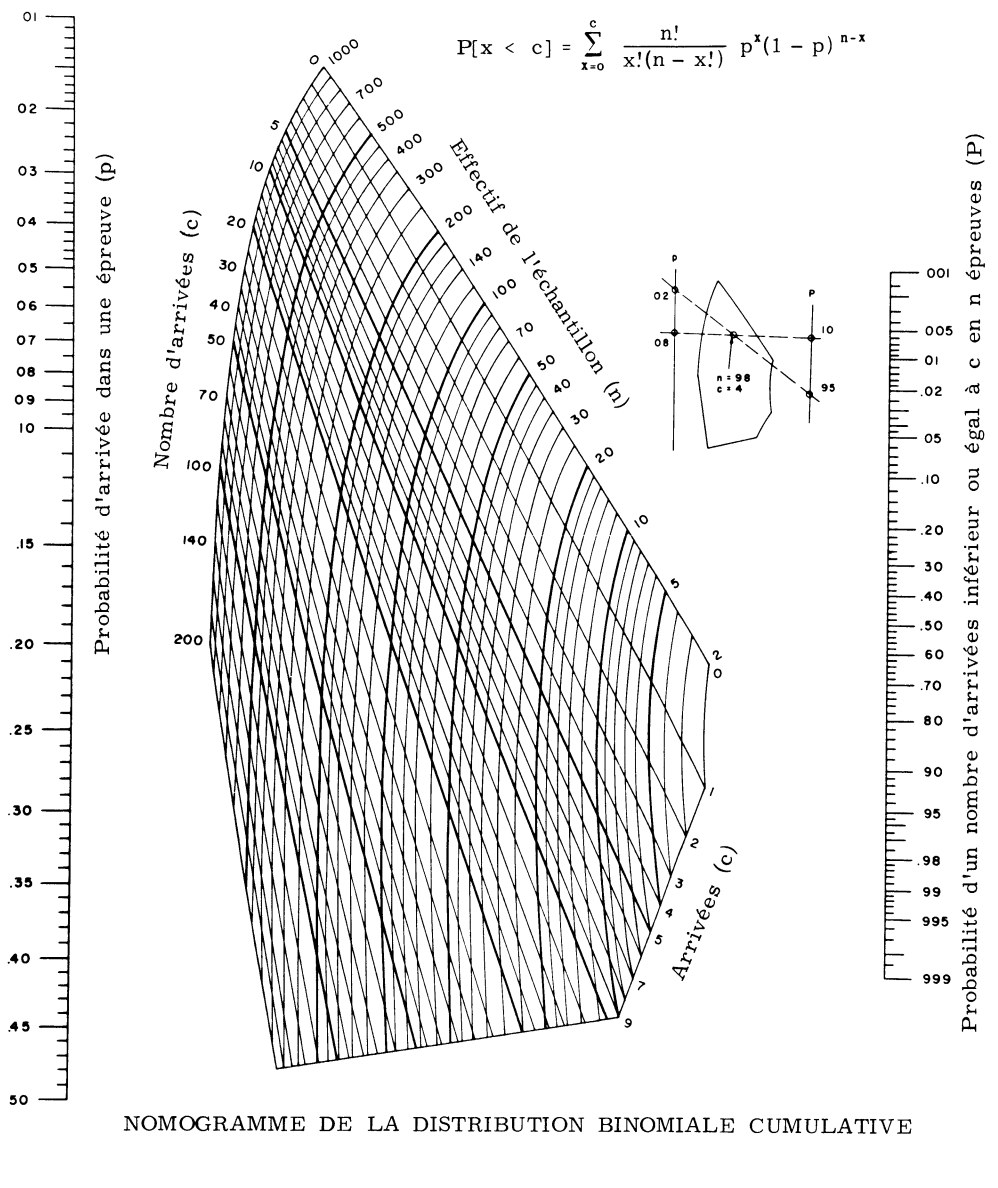
[7] HALD - Statistical theory with engineering applications. John Wiley, New-York, 1955.

[8] WILKS - Mathematical statistics. Princeton University Press, 1943.

[9] CAVE - Le contrôle statistique des fabrications. Eyrolles, 1961.

[10] MORICE - Les graphiques à échelles fonctionnelles du statisticien Revue de Statistique Appliquée, 1964 Vol XII N° 3, p. 73.

[11] H. 108 - Quality Control and reliability Handbook. Secretary of Defense, Washington 25 D. S.



NOMOGRAMME DE LA DISTRIBUTION BINOMIALE CUMULATIVE