REVUE DE STATISTIQUE APPLIQUÉE

F. CHARTIER
E. MORICE

Relations entre quelques lois de probabilités

Revue de statistique appliquée, tome 15, nº 4 (1967), p. 17-33
<http://www.numdam.org/item?id=RSA_1967__15_4_17_0>

RELATIONS ENTRE QUELQUES LOIS DE PROBABILITÉS

F. CHARTIER et E. MORICE

Dans un précédent article (E. Morice "Quelques modèles mathématiques de durée de vie"), on a signalé brièvement la liaison entre la loi binomiale et la loi binomiale négative. On en trouvera ci-après la démonstration basée sur les relations qui lient leurs fonctions de répartition à la fonction de répartition de la loi Beta et par conséquent à la fonction de répartition de la variable F de Fisher-Snedecor.

Ce même procédé de démonstration permet aussi d'établir très simplement les relations qui existent entre la fonction de répartition d'une loi de Poisson et celles de la loi Gamma et de la loi de χ^2

En raison de l'intérêt de ces transformations dans divers problèmes d'estimation ou de tests, il a paru utile de les regrouper avec quelques exemples d'application et avec les diverses approximations qui peuvent, dans certains cas, être utilisées valablement.

I - LOIS BINOMIALE, BINOMIALE NEGATIVE, BETA ET F

1/ Introduction.

Considérons des épreuves aléatoires indépendantes, le résultat de chacune ne pouvant revêtir que deux formes, que nous appellerons le succès et l'insuccès, avec des probabilités complémentaires p et 1 - p respectivement.

On sait que la *loi binomiale* est celle du *nombre aléatoire de succès* sur un *nombre donné d'épreuves,* alors que la *loi binomiale négative* est celle du *nombre aléatoire d'épreuves* à réaliser pour obtenir un *nombre donné de* succès.

On va montrer que, k et n étant deux entiers positifs avec $k \leqslant n$, on a :

loi binomiale		loi binomiale négative
Pr (k succès ou plus de k sur n épreuves, n donné)	=	Pr (n épreuves ou moins de n pour k succès, k donné)

L'intérêt pratique de ce résultat est de permettre l'obtention de la fonction de répartition (ou des probabilités cumulées) de la loi binomiale négative à l'aide des tables de la loi binomiale, voire des approximations normale ou de Poisson de cette dernière loi.

Pour établir le résultat annoncé, on exprimera les deux probabilités sous forme d'une même intégrale eulérienne β incomplète.

2/ Expression de la somme des derniers termes d'une loi binomiale par une fonction β incomplète.

La loi binomiale $\mathcal{B}(n, p)$ est celle de la variable aléatoire X, nombre de "succès" sur n épreuves indépendantes, à chacune desquelles la probabilité du succès est p.

Pour $x \in \{0, 1, 2, \ldots, n\}$, on a

$$\Pr(X = x) = C_n^x p^x (1 - p)^{n-x} , \tag{2.1}$$

conformément au schéma :

épreuve n°	$\underbrace{1 \quad 2 \quad 3 \quad \ldots\ldots\ldots\ldots \quad n}$
résultat	x succès et (n − x) insuccès dans un ordre quelconque

La somme des derniers termes de la loi binomiale, *au delà de l'entier* k est

avec $k \in \{0, 1, \ldots, n - 1\}$,

$$\boxed{\Pr[X > k] = Q_B(k) = \sum_{x=k+1}^{n} C_n^x p^x (1 - p)^{n-x} = I_p(k + 1, n - k)} \tag{2.2}$$

où

$$I_p(a, b) = \frac{\int_0^p u^{a-1}(1 - u)^{b-1} \, du}{\int_0^1 u^{a-1}(1 - u)^{b-1} \, du}$$

Remarque - La sommation définissant $Q_B(k)$ partant de k + 1, $Q_B(k)$ est le complément à l'unité de la fonction de répartition de la loi binomiale :

$$Q_B(k) = 1 - P_B(k) \qquad \text{avec} \qquad P_B(k) = \Pr(X \leqslant k) .$$

On aurait de même

$$\Pr[X \geqslant k] = Q_B(k - 1) = I_p(k, n - k + 1) \tag{2.3}$$

Démonstration - Considérons $Q_B(k)$ comme une fonction de p que nous noterons plus simplement S(p).

$$S(p) = \sum_{k+1}^{n} C_n^x p^x (1 - p)^{n-x}$$

Sa dérivée par rapport à p est :

$$\frac{dS}{dp} = \sum_{k+1}^{n} x\, C_n^x\, p^{x-1}\, (1 - p)^{n-x} - \sum_{k+1}^{n-1} (n - x)\, C_n^x\, p^x (1 - p)^{n-x-1}$$

La première somme seule contient un terme en p^k, soit : $(k + 1)\, C_n^{k+1}\, p^k\, (1 - p)^{n-k-1}$, les autres termes de même exposant en p, (k + 1, k + 2, ..., n - 1), regroupés, donnent

$$\sum_{k+1}^{n} [x\, C_n^x - (n - x + 1)\, C_n^{x-1}]\, p^{x-1}(1 - p)^{n-x} = 0$$

car

$$x\, C_n^x - (n - x + 1)\, C_n^{x-1} = \frac{x.\, n!}{x!\,(n - x)!} - \frac{(n - x + 1)\, n!}{(x - 1)!\,(n - x + 1)!} = 0$$

On obtient finalement

$$dS = (k + 1)\, C_n^{k+1}\, p^k (1 - p)^{n-k-1}\, dp$$

Intégrons les deux membres de 0 à p :

$$\int_0^p dS = (k + 1)\, C_n^{k+1} \int_0^p p^k (1 - p)^{n-k-1}\, dp\,,$$

soit

$$S(p) - S(o) = \frac{n!}{k!\,(n - k - 1)!}\, B_p(k + 1\,,\ n - k)$$

Comme S(o) = 0, il reste :

$$S(p) = \frac{B_p(k + 1\,,\ n - k)}{B(k + 1\,,\ n - k)} = I_p(k + 1\,,\ n - k)$$

On rappelle que $B(a,\, b) = \frac{\Gamma(a)\, \Gamma(b)}{\Gamma(a + b)}$ et que si a est entier, positif, on a $\Gamma(a) = (a - 1)!$

3/ Expression de la somme des premiers termes d'une loi binomiale négative par une fonction β incomplète.

La loi binomiale négative BN(k, p) est celle de la variable aléatoire Y, nombre des épreuves indépendantes nécessaires et suffisantes pour obtenir k "succès", la probabilité du succès étant p à chaque épreuve.

Pour $y \in \{k, k+1, \ldots, \infty\}$, on a

$$\Pr(Y = y) = C_{y-1}^{k-1}\, p^k (1-p)^{y-k}, \qquad (3.1)$$

conformément au schéma :

épreuve n° :	$\underbrace{1 \quad 2 \quad \ldots \quad k \quad \ldots \quad y-1}$	$y \quad \ldots$
résultat	k − 1 succès et y − k insuccès dans un ordre quelconque	k^e succès

La somme des premiers termes de cette loi binomiale négative, *jusqu'à l'entier* n, avec $n \in \{k, k+1, \ldots, \infty\}$ vaut

$$\boxed{\Pr[Y \leqslant n] = P_{BN}(n) = \sum_{y=k}^{n} C_{y-1}^{k-1}\, p^k (1-p)^{y-k} = I_p(k, n-k+1)} \qquad (3.2)$$

Remarque - $P_{BN}(b) = \Pr(Y \leqslant n)$ est la fonction de répartition de la loi binomiale négative $BN(k, p)$.

Démonstration - Considérons $P_{BN}(n)$ comme une fonction de p que nous noterons $T(p)$ dans cette démonstration :

$$T(p) = p^k \sum_{k}^{n} C_{y-1}^{k-1} (1-p)^{y-k}$$

Sa dérivée par rapport à p est

$$\frac{dT}{dp} = k\, p^{k-1} \sum_{k}^{n} C_{y-1}^{k-1} (1-p)^{y-k} - p^k \sum_{k+1}^{n} (y-k)\, C_{y-1}^{k-1} (1-p)^{y-k-1}$$

Au second terme, remplaçons p^k par $p^{k-1}[1-(1-p)]$:

$$\frac{dT}{dp} = k\, p^{k-1} \sum_{k}^{n} C_{y-1}^{k-1} (1-p)^{y-k} - p^{k-1} \sum_{k+1}^{n} (y-k)\, C_{y-1}^{k-1} (1-p)^{y-k-1}$$

$$+ p^{k-1} \sum_{k+1}^{n} (y-k)\, C_{y-1}^{k-1} (1-p)^{y-k}$$

Regroupons les termes de même puissance en $(1-p)$ et écrivons la 2e somme privée de son premier terme :

$$\sum_{k+2}^{n} (y-k)\, C_{y-1}^{k-1} (1-p)^{y-k-1} = \sum_{k+1}^{n-1} (y+1-k)\, C_{y}^{k-1} (1-p)^{y-k}.$$

Il vient :

$$\frac{dT}{dp} = p^{k-1} \left[k\, C_{k-1}^{k-1} - 1\, C_{k}^{k-1}\right] (1-p)^{\circ}$$

$$+ p^{k-1} \sum_{k+1}^{n-1} \left[k\, C_{y-1}^{k-1} - (y+1-k)\, C_{y}^{k-1} + (y-k)\, C_{y-1}^{k-1}\right] (1-p)^{y-k}$$

$$+ p^{k-1} \left[k\, C_{n-1}^{k-1} + (n-k)\, C_{n-1}^{k-1}\right] (1-p)^{n-k}.$$

Le premier terme est nul. Dans la somme qui figure au second, le coefficient de $(1 - p)^{y-k}$ est

$$y \frac{(y-1)!}{(k-1)!\,(y-k)!} - (y+1-k) \frac{y!}{(k-1)!\,(y+1-k)!} = 0 ,$$

et le troisième terme se réduit à

$$p^{k-1}\, n\, C_{n-1}^{k-1} (1-p)^{n-k} = \frac{n!}{(k-1)!\,(n-k)!}\, p^{k-1} (1-p)^{n-k} .$$

On a donc

$$dT = \frac{n!}{(k-1)!\,(n-k)!}\, p^{k-1} (1-p)^{n-k}\, dp$$

Intégrons les deux membres de 0 à p :

$$T(p) - T(0) = \frac{n!}{(k-1)!\,(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k}\, dt .$$

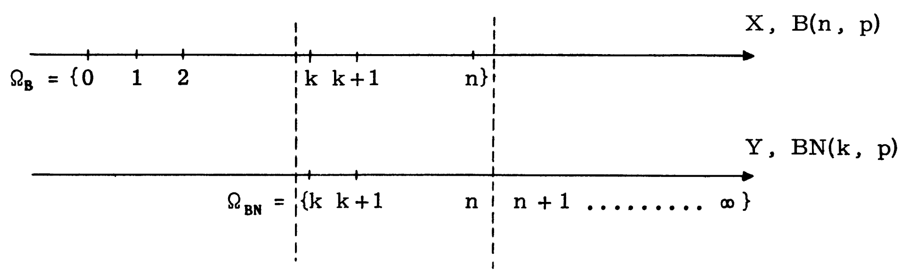
Comme T(0) = 0, on a bien :

$$P_{BN}(n) = Pr(Y \leqslant n) = T(p) = \frac{B_p(k,\ n-k+1)}{B(k,\ n-k+1)} = I_p(k,\ n-k+1)$$

4/ <u>Application des résultats précédents</u> : relation entre la loi binomiale et la loi binomiale négative.

$$0 < k \leqslant n .$$

Soit X la variable aléatoire binomiale B(n, p) dont l'ensemble des valeurs possibles est $\Omega_B = \{0, 1, 2, \ldots, n\}$ et Y la variable aléatoire binomiale négative BN(k, p) pour laquelle l'ensemble des valeurs possibles est $\Omega_{BN} = \{k, k+1, \ldots, \infty\}$, suivant le schéma



On a

$$Pr(k \leqslant X \leqslant n) = Pr(k \leqslant Y \leqslant n) , \qquad (4.1)$$

la valeur commune de ces deux probabilités étant $I_p(k, n - k + 1)$ d'après (2.3) et (3.2).

Ainsi la fonction de répartition de la variable aléatoire Y, BN(k, p), peut s'écrire :

$$
\left.\begin{aligned}
P_{BN}(Y \leqslant n \; ; \; k, \; p) &= \sum_{n=k}^{n} C_n^x \, p^x \, (1 - p)^{n-x} \\
&= 1 - \sum_{x=o}^{k-1} C_n^x \, p^x \, (1 - p)^{n-x} \\
&= 1 - P_B \, (X < k \; ; \; n, \; p)
\end{aligned}\right\} \qquad (4.2)
$$

ce qui permet d'utiliser les tables usuelles de la loi binomiale.

Remarque - Les tables de la loi $\mathcal{B}(n, p)$ sont établies pour $p \in]0, 1/2]$. Si $p \in]1/2, 1[$, on utilise les relations :

$$
\left.\begin{aligned}
\sum_{x=k}^{n} C_n^x \, p^x \, (1 - p)^{n-x} &= \sum_{x=o}^{n-k} C_n^x \, p_1^x \, (1 - p_1)^{n-x} \\
&= 1 - \sum_{x=n-k+1}^{n} C_n^x \, p_1^x \, (1 - p_1)^{n-x}
\end{aligned}\right.
$$

ou

$$
\sum_{x=o}^{k} C_n^x \, p^x \, (1 - p)^{n-x} = 1 - \sum_{x=o}^{n-k-1} C_n^x \, p_1^x \, (1 - p_1)^{n-x}
$$

avec $p_1 = 1 - p$.

Exemple - On lance un dé à jouer jusqu'à ce que l'on obtienne $k = 5$ faces autres que l'as $\left(p = \frac{5}{6} \text{ à chaque jet}\right)$. Quelle est la probabilité d'atteindre ce résultat en $n = 8$ jets ou moins de 8 jets du dé ?

$$
\begin{aligned}
P_{BN} \left(5 \leqslant Y \leqslant 8 \; ; \; k = 5, \; p = \frac{5}{6}\right) &= \sum_{x=5}^{8} C_8^x \left(\frac{5}{6}\right)^x \left(\frac{1}{6}\right)^{8-x} \\
&= 1 - \sum_{x=4}^{8} C_8^x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{8-x}
\end{aligned}
$$

Cette dernière expression est celle de la loi $\mathcal{B}\left(n = 8, \; p_1 = \frac{1}{6}\right)$. On trouve dans la table [1] :

$$
\sum_{x=4}^{8} C_8^x \, p^x (1 - p)^{8-x} = \begin{cases} 0,02667 & \text{pour} \quad p = 0,16 \\ 0,03279 & \text{pour} \quad p = 0,17 \end{cases}
$$

soit, par interpolation, $\approx 0,03075$ " $p = 0,166 \ldots$

d'où :

$$
P_{BN} \left(5 \leqslant Y \leqslant 8 \; ; \; k = 5, \; p = \frac{5}{6}\right) \approx 0,96925
$$

5/ Relation entre la loi binomiale et la loi de F.

La loi de la variable

$$
F = \frac{(1/\nu_1) \, \chi^2 \, (\nu_1)}{1/\nu_2 \, \chi^2 \, (\nu_2)}
$$

a pour densité de probabilité

$$f(F\ ;\ \nu_1, \nu_2) = \frac{1}{B(\nu_1/2, \nu_2/2)} \frac{\nu_1^{\nu_1/2}\ \nu_2^{\nu_2/2}\ F^{\frac{\nu_1}{2}-1}}{(\nu_2 + \nu_1 F)^{\frac{\nu_1+\nu_2}{2}}} \quad , \quad 0 < F < \infty$$

Si l'on fait le changement de variable défini par

$$F = \frac{\nu_1}{\nu_2} \frac{Z}{1 - Z} \quad , \quad Z = \frac{\nu_1 F}{\nu_2 + \nu_1 F} ,$$

la loi de la variable Z est la loi Beta définie par

$$g(z) = \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} z^{\frac{\nu_1}{2}-1} (1 - z)^{\frac{\nu_2}{2}-1} , \quad 0 < Z < 1 ,$$

d'où

$$P(z) = \Pr\ [Z < z] = \frac{B_z\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} = I_z\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \tag{5.1}$$

On a vu précédemment, à propos de la loi binomiale, que

$$P_B(k) = \sum_{x=0}^{k} C_n^x p^x (1 - p)^{n-x} = 1 - I_p(k + 1, n - k) \tag{5.2}$$

Le rapprochement des relations (5.1) et (5.2) montre que

$$\text{si} \quad \begin{cases} \nu_1 = 2(k + 1) \\ \nu_2 = 2(n - k) \end{cases} \quad \text{et} \quad z = p = \frac{\nu_1 F}{\nu_2 + \nu_1 F} ,$$

alors

$$P_B(k) = 1 - P(z) ,$$

soit

$$\begin{aligned} \sum_{x=0}^{k} C_n^x\ p^x (1 - p)^{n-x} &= 1 - \Pr\left(\frac{(k + 1)\ F}{(n - k) + (k + 1)\ F} < p\right) \\ &= 1 - \Pr\left(F < \frac{n - k}{k + 1} \cdot \frac{p}{1 - p}\right) \end{aligned} \tag{5.3}$$

Pour n, p et k donnés, posons α la probabilité $P_B(k)$:

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

D'après (5.3) :

$$\Pr\left(F < \frac{n - k}{k + 1} \cdot \frac{p}{1 - p}\right) = 1 - \alpha ,$$

de sorte que, en notant $F_{1-\alpha}$ le fractile d'ordre $1 - \alpha$ de la variable aléatoire F :

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

Cette relation se résout en p :

$$p = \frac{(k+1)\, F_{1-\alpha}}{(n-k) + (k+1)\, F_{1-\alpha}} \qquad \begin{cases} \nu_1 = 2(k+1) \\ \nu_2 = 2(n-k) \end{cases} \tag{5.4}$$

ou encore

$$p = \frac{k+1}{(n-k)\, F_\alpha + (k+1)} \qquad \begin{cases} \nu_1 = 2(n-k) \\ \nu_2 = 2(k+1) \end{cases} \tag{5.5}$$

l'une ou l'autre des formules (5.4) et (5.5) permettant d'utiliser les tables usuelles des fractiles de F suivant que α est inférieur ou supérieur à 0,5.

6/ Application - Intervalle de confiance symétrique en probabilité, à $1 - \alpha$, pour le paramètre p d'une loi binomiale, d'après les résultats d'un échantillon d'effectif n ayant donné k éléments de la catégorie de probabilité p

La limite supérieure p_s étant définie par $\sum_{x=0}^{k} C_n^x\, p_s^x (1 - p)^{n-x} = \frac{\alpha}{2}$, on a

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

De même la limite inférieure p_i étant définie par

$$\sum_{x=k}^{n} C_n^x\, p_i^x\, (1 - p_i)^{n-x} = \frac{\alpha}{2}$$

soit

$$\sum_{x=0}^{k-1} C_n^x\, p_i^x\, (1 - p_i)^{n-x} = 1 - \frac{\alpha}{2}$$

On a :

$$p_i = \frac{k\, F_{\alpha/2}}{(n-k+1)\quad k\, F_{\alpha/2}} \qquad \begin{cases} \nu_1 = 2k \\ \nu_2 = 2(n-k+1), \end{cases}$$

soit pour tenir compte de la présentation habituelle des tables

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

Exemple - $n = 10$; $k = 1$; $\alpha = 0{,}10$;

$$p_s = \frac{2 \times 2,93}{9 + 2 \times 2,93} = \frac{5,86}{14,86} = 0,39$$

$$p_i = \frac{1}{10 \times 19,4 + 1} = \frac{1}{195} = 0,005$$

7/ Fractiles de la variable aléatoire $F(x_i)$, x_i étant la $i^{ème}$ valeur d'un échantillon ordonné de n observations indépendantes d'une variable continue suivant la loi $F(x)$.

Soit une variable aléatoire X de fonction de répartition $F(x)$ et de densité de probabilité $f(x)$. On en réalise n observations indépendantes que l'on classe par valeurs croissantes, obtenant ainsi la série *ordonnée* :

$$x_1 < x_2 \ldots\ldots\ldots\ldots < x_i < \ldots\ldots\ldots\ldots < x_n$$

La loi de la variable aléatoire $X_{i,n}$, $i^{ème}$ observation, a pour densité de probabilité :

$$f_{in}(x) = \frac{n!}{(i-1)!\,(n-i)!}\,[F(x)]^{i-1}\,[1 - F(x)]^{n-i}\,f(x) \tag{7.1}$$

conformément au schéma :

Nombre d'observations	$i - 1$	1	$n - i$
		x … $x+dx$	
Probabilité	$F(x)$	$f(x)dx$	$1 - F(x)$

Si, au lieu de la $i^{ème}$ observation $X_{i,n}$, on considère la valeur $F(x_i)$, il s'agit encore d'une variable aléatoire, que nous noterons $Y(i, n)$ et qui est distribuée sur le segment $(0, 1)$ avec la densité de probabilité

$$g_{i,n}(y) = \frac{n!}{(i-1)!\,(n-i)!}\, y^{i-1}\,(1 - y)^{n-i}$$

C'est donc une variable aléatoire de loi $\beta(i, n - i + 1)$ dont le fractile d'ordre α, $y_\alpha(i, n)$, défini par

$$\alpha = \int_0^{y_\alpha(i,n)} g_{i,n}(y)\, dy\ ,$$

s'exprime en fonction du fractile F_α de la variable F, de degrés de liberté $\nu_1 = 2i$, $\nu_2 = 2(n - i + 1)$

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

A défaut d'une table de la loi β, ce fractile pourra donc, pour les valeurs usuelles de α, être calculé à l'aide d'une table de la loi de F.

Si l'on se souvient de ce que la relation (7.1) peut s'écrire :

$$\alpha = \int_0^{y_\alpha(i,n)} g_{i,n}(y)\, dy = I_{y_\alpha(i,n)}(i, n - i + 1)$$

$$= 1 - \sum_{x=0}^{i-1} C_n^x [y_\alpha(i, n)]^x [1 - y_\alpha(i, n)]^{n-x} ,$$

on voit que, pour :

$$i = 1 \qquad [1 - y_\alpha(1, n)]^n = 1 - \alpha \longrightarrow y_\alpha(1, n) = 1 - \sqrt[n]{1 - \alpha}$$

$$i = n \qquad [y_\alpha(n, n)]^n = \alpha \longrightarrow y_\alpha(n, n) = \sqrt[n]{\alpha}$$

8/ Estimation du $F(x_i)$

La méthode généralement employée pour préciser la densité de probabilité $f(x)$, de forme connue, représentant la distribution de la population d'où provient l'échantillon de n observations, consiste à estimer ses paramètres à l'aide d'une méthode adéquate (méthode des moments, maximum de vraisemblance, ...).

Mais on peut aussi envisager une estimation directe de la fonction de répartition $F(x)$. Cette méthode est d'un emploi particulièrement commode lorsqu'une anamorphose convenablement choisie permet de transformer la courbe des probabilités cumulées en droite. C'est le cas, en particulier, pour la loi normale, la loi exponentielle et la loi de Weibull.

Il s'agit alors de faire correspondre à chacune des n observations $x_1, x_2, \ldots, x_i, \ldots, x_n$ une estimation $y^*(i, n)$ de la valeur $F(x_i)$ de la fonction de répartition $F(x)$.

L'estimation généralement utilisée est l'espérance mathématique

$$E[y(i, n)] = \int_0^1 \frac{n!}{(i-1)!\,(n-i)!} y^i (1 - y)^{n-i}\, dy$$

$$= \frac{i}{n + 1}$$

Certains auteurs préfèrent utiliser la valeur médiane de $y(i, n)$ ("median rank"), soit :

$$y_{0,50}(i, n) = \frac{i\, F_{0,50}}{n - i + 1 + i\, F_{0,50}} \qquad \begin{cases} \nu_1 = 2i \\ \nu_2 = 2(n - i + 1) \end{cases}$$

Ces deux valeurs sont très peu différentes lorsque i est voisin de $\frac{n}{2}$. En particulier, pour n impair et $i = \frac{n+1}{2}$, on a $\nu_1 = \nu_2 = 2i$ d'où $F_{0,50} = 1$ et

$$y_{0,50}\left(\frac{n+1}{2}, n\right) = \frac{i}{2i} = 0,50$$

de même que

$$E\left[y\left(\frac{n+1}{2}, n\right)\right] = 0,50$$

D'une manière plus générale on a, en désignant par $M_o = \frac{i - 1}{n - 1}$ le mode de la variable $y(i, n)$:

$$M_o < y_{0,50} < E(y) \qquad \text{pour} \qquad n > 2i - 1$$

$$M_o > y_{0,50} > E(y) \qquad \text{pour} \qquad n < 2i - 1$$

L'utilisation, soit de $y_{0,50}$, soit de $E(y)$, tend donc à donner des indications différentes sur la pente de la droite qui sera utilisée comme approximation graphique de la courbe transformée de la fonction de répartition $F(x)$.

D'autre part on constate que quel que soit i ($1 \leqslant i \leqslant n$), on a pour tout petit intervalle $\pm \Delta y$:

$$Pr\,[y_{0,50} - \Delta y < y < y_{0,50} + \Delta y] > Pr\,[E(y) - \Delta y < y < E(y) + \Delta y]\ ,$$

ce qui conduit à préférer l'estimation $y_{0,50}$ à l'estimation $E(y)$ (par exemple, pour $n = 20$, $i = 1$, d'où $y_{0,50} = 0,035$ et $E(y) = 0,048$, le rapport des deux probabilités envisagées ci-dessus est voisin de 1,4).

Mais, ce qui est pratiquement plus intéressant du point de vue graphique, c'est la possibilité d'avoir une information sur les risques associés au tracé d'une droite, sur le graphique à échelles fonctionnelles à partir des points (x_i, y_i^*) quelle que soit celle des estimations y_i^* qui a été choisie.

On peut en effet associer à chaque valeur inconnue y_i un intervalle de confiance à $1 - \alpha$ défini par :

$$1 - \alpha = Pr\left[\frac{i\,F_{\alpha/2}}{n - i + 1 + i\,F_{\alpha/2}} < y_i < \frac{i\,F_{1-\alpha/2}}{n - i + 1 + i\,F_{1-\alpha/2}}\right]$$

$$[\nu_1 = 2i\ ,\ \nu_2 = 2(n - i - 1)]$$

Par exemple pour $n = 20$, on a

$$i = 1 \quad \begin{cases} 0,90 = Pr\,[0,00256 < y_1 < 0,141] \\ 0,40 = Pr\,[0,018 < y_1 < 0,058] \end{cases}$$

$$i = 10 \quad \begin{cases} 0,90 = Pr\,[0,35 < y_{10} < 0,60] \\ 0,40 = Pr\,[0,419 < y_{10} < 0,533] \end{cases}$$

$$i = 20 \quad \begin{cases} 0,90 = Pr\,[0,861 < y_{20} < 0,997] \\ 0,40 = Pr\,[0,942 < y_{20} < 0,982] \end{cases}$$

Pour une valeur fixée de α et pour un effectif donné n, l'amplitude de ces intervalles de confiance est maximale pour i voisin de $n/2$, mais ce qu'il convient de remarquer c'est l'amplitude graphique de ces intervalles lorsqu'on utilise un papier dont l'échelle des ordonnées graduée en $F(x)$ est une échelle fonctionnelle. Cette amplitude graphique, caractéristique de l'incertitude sur la mise en place des points (x_i, y_i^*) est particulièrement importante dans le cas de la loi de Weibull où l'on utilise un graphique dans lequel l'axe des ordonnées gradué en F(x), correspond à une échelle métrique en $\text{Log}_e\left(\text{Log}_e \frac{1}{i - F(x)}\right)$.

Ainsi, par exemple, dans un papier 21 × 27 avec un axe des ordonnées gradué de 0,001 à 0,999 sur une longueur de 18 cm, les intervalles (0,0025 - 0,14) et (0,018 - 0,058) ont des longueurs respectivement égales à 85 mm et 24 mm.

De même les intervalles (0,35 - 0,65) et (0,42 - 0,53) ont des longueurs respectivement égales à 22 mm et 8 mm et les intervalles (0,861 - 0,997) et (0,942 - 0,982) ont pour longueurs 24 mm et 7 mm.

Ceci montre les difficultés rencontrées dans l'estimation graphique d'une loi de répartition lorsque l'on utilise de petits échantillons, et plus particulièrement lorsque, pour les mêmes raisons d'économie, on se contente d'un petit nombre de premières défaillances.

II - LOIS DE POISSON, GAMMA ET χ^2

1/ Expression de la somme des derniers termes d'une loi de Poisson par une fonction Gamma incomplète.

La loi de Poisson $\mathcal{P}(m)$, de paramètre m est celle de la variable aléatoire X, définie par

$$\Pr[X = x] = e^{-m} \frac{m^x}{x!} \qquad \text{pour} \qquad x \in \{0, 1, \dots \infty\}$$

La somme des derniers termes au delà de l'entier k de la loi de Poisson de paramètre m est

$$\Pr[X > k] = Q_p(k) = \sum_{x=k+1}^{\infty} e^{-m} \frac{m^x}{x!}$$

Considérons $Q_p(k)$ comme une fonction de m que nous noterons S(m)

$$S(m) = \sum_{k+1}^{\infty} e^{-m} \frac{m^x}{x!}$$

d'où

$$\frac{dS}{dm} = \sum_{k+1}^{\infty} \left[- e^{-m} \frac{m^x}{x!} + e^{-m} \frac{m^{x-1}}{(x-1)!} \right] = e^{-m} \frac{m^k}{k!}$$

En intégrant de 0 à m, on a :

$$\int_o^m dS = S(m) - S(o) = \frac{1}{k!} \int_o^m e^{-m} m^k \, dm$$

soit, compte tenu de S(0) = 0 :

$$Q_p(k) = \sum_{k+1}^{\infty} e^{-m} \frac{m^x}{x!} = \frac{\Gamma_m(k+1)}{\Gamma(k+1)}, \qquad (1)$$

fonction de répartition d'une variable aléatoire continue m, distribuée suivant la loi Gamma de paramètre k + 1.

2/ Relation entre la loi de Poisson et la loi de χ^2.

La loi de variable $\chi^2(\nu)$ a pour densité de probabilité

$$f(\chi^2) = \frac{1}{2^{\nu/2}\,\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{\nu}{2}-1} \qquad 0 < \chi^2 < \infty$$

Si on fait le changement de variable défini par

$$\frac{\chi^2}{2} = Z$$

la loi de la variable Z est la loi Gamma de densité de probabilité

$$g(z) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} e^{-z} Z^{\frac{\nu}{2}-1} \quad ;$$

d'où

$$P\left(\frac{\chi^2}{2} < m\right) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \int_0^m e^{-z} Z^{\frac{\nu}{2}-1}\, dZ = \frac{\Gamma_m\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \tag{2}$$

La comparaison des relations (1) et (2) donne

$$\sum_{x=k+1}^{\infty} e^{-m} \frac{m^x}{x!} = \Pr\left[\chi^2 < 2m,\ \nu = 2(k+1)\right] \tag{3}$$

Pour une valeur fixée α telle que

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

on aura

$$1 - \alpha = \Pr\left[\chi^2 < 2m,\ \nu = 2(k+1)\right]$$

d'où

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

3/ Application - Intervalle de confiance, symétrique en probabilité, à $1-\alpha$ pour le paramètre m d'une loi de Poisson, d'après les résultats d'une observation ayant donné k éléments possédant la propriété considérée.

La limite supérieure m_s étant définie par

$$\sum_{x=0}^{k} e^{-m_s} \frac{(m_s)^x}{x!} = \frac{\alpha}{2} = 1 - \Pr\left[\chi^2 < 2m_s,\ \nu = 2(k+1)\right]$$

on a

$$m_s = \frac{1}{2}\chi^2_{1-\alpha/2}, \qquad \nu = 2(k+1)$$

De même pour la limite inférieure m_i définie par

$$\sum_{x=k}^{\infty} e^{-m_i} \frac{(m_i)^x}{x!} = \frac{\alpha}{2} = \Pr[\chi^2 < 2m_i \ , \ \nu = 2k]$$

on a

$$m_i = \frac{1}{2} \chi^2_{\alpha/2} \ , \ \nu = 2\,k$$

<u>Exemple</u> - $k = 24 \qquad \alpha = 0{,}10$

$$m_i = \frac{1}{2} \chi^2_{0{,}05}(48) = 16{,}55$$

$$m_s = \frac{1}{2} \chi^2_{0{,}95}(50) = 33{,}75$$

RESUME

1 - LOIS BINOMIALE, BETA ET F

Nombre aléatoire X de succès dans n épreuves indépendantes, avec p, probabilité de succès ($X = 0, 1, \ldots, n$).

$$\alpha = \Pr[X \leqslant k] = \sum_{x=0}^{k} C_n^x p^x (1-p)^{n-x} = \begin{cases} 1 - \dfrac{\int_0^p u^k (1-u)^{n-k-1} du}{\int_0^1 u^k (1-u)^{n-k-1} du} = 1 - I_p(k+1\,;\,n-k) \\[2ex] 1 - \Pr\left[F < \dfrac{n-k}{k+1} \dfrac{p}{1-p}\right] \end{cases}$$

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

Ces résultats sont pratiquement valables pour la loi hypergéométrique (tirages sans remise) si $p = n/N < 0{,}10$.

2 - LOIS BINOMIALE ET BINOMIALE NEGATIVE

a) Nombre aléatoire Y d'épreuves indépendantes pour obtenir k succès ($Y = k, k+1, \ldots \infty$)

$$\Pr[Y \leqslant n] = \sum_{y=k}^{n} C_{y-1}^{k-1} p^k (1-p)^{y-k} = 1 - \sum_{x=0}^{k-1} C_n^x p^x (1-p)^{n-x} = I_p(k,\ n-k+1)$$

$$= 1 - \Pr\left[F < \frac{n-k+1}{k} \frac{p}{1-p}\right] , \quad \nu_1 = 2k \ , \quad \nu_2 = 2(n-k+1)$$

b) Nombre aléatoire Z d'épreuves indépendantes, nécessaires, au-delà des k premières pour obtenir k succès ($Z = Y - k = 0, 1, \ldots, \infty$)

$$\Pr[Z \leqslant s] = \sum_{z=0}^{s} C_{k-1+z}^{k-1} p^k (1-p)^{z-k} = 1 - \sum_{x=0}^{k-1} C_{k+s}^{x} p^x (1-p)^{k+s-x} = I_p(k, s+1)$$

$$= 1 - \Pr\left[F < \frac{s+1}{k}\,\frac{p}{1-p}\right], \quad \nu_1 = 2k, \quad \nu_2 = 2(s+1)$$

Ces résultats sont pratiquement valables pour la loi hypergéométrique négative (tirages sans remise) si $p = n/N < 0,10$.

3 - LOIS DE POISSON, GAMMA ET χ^2

X variable aléatoire distribuée suivant une loi de Poisson de paramètre m ($X = 0, 1, \ldots, \infty$)

$$\alpha = \Pr[X \leqslant k] = \sum_{x=0}^{k} e^{-m} \frac{m^x}{x!} = \begin{cases} 1 - \dfrac{\Gamma_m(k+1)}{\Gamma(k+1)} \\ 1 - \Pr[\chi^2 < 2m, \ \nu = 2(k+1)] \end{cases}$$

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

4 - LOI BINOMIALE ET LOI DE POISSON

Pour $p < 0,10$

$$\alpha = \Pr[X \leqslant k] = \sum_{x=0}^{k} C_n^x p^x (1-p)^{n-x} \approx \sum_{x=0}^{k} e^{-np} \frac{(np)^x}{x!}$$

5 - LOI BINOMIALE ET LOI NORMALE

Pour $np > 15$ (p étant la plus petite des quantités p et $1 - p$)

$$\alpha = \Pr[X \leqslant k] = \sum_{x=0}^{k} C_n^x p^x (1-p)^{n-x} \sim \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

$$u_\alpha = \frac{k + 0,5 - np}{\sqrt{np(1-p)}}$$

6 - LOI DE POISSON ET LOI NORMALE

Pour $m > 15$

$$\alpha = \Pr[X \leqslant k] = \sum_{x=0}^{k} e^{-m} \frac{m^x}{x!} \sim \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

$$u_\alpha = \frac{k + 0,5 - m}{\sqrt{m}}$$

7 - LOI GAMMA ET LOI NORMALE

$$\alpha = \Pr[\gamma(m)] < a] = \frac{1}{\Gamma(m)} \int_0^a e^{-x} x^{m-1}\, dx = \frac{\Gamma_a(m)}{\Gamma(m)}$$

1/ pour $m > 50$

$$\alpha \sim \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$a = \gamma_\alpha(m) \sim m + u_\alpha \sqrt{m} \quad , \quad u_\alpha = \frac{a - m}{\sqrt{m}}$$

2/ pour $m > 30$

$$\alpha \sim \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$a = \gamma_\alpha(m) \sim \frac{1}{4} [u_\alpha + \sqrt{4m - 1}]^2 \quad , \quad u_\alpha = \sqrt{4a} - \sqrt{4m - 1}$$

8 - LOI DE χ^2 ET LOI NORMALE

$$\alpha = \Pr [\chi^2(\nu) < a] = \int_0^a \frac{1}{2^{\nu/2}\, \Gamma\left(\frac{\nu}{2}\right)} e^{-\chi^2/2} (\chi^2)^{\frac{\nu}{2}-1} d(\chi^2)$$

1/ pour $\nu > 100$

$$\alpha \sim \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$a = \chi^2_\alpha(\nu) \sim \nu + u_\alpha \sqrt{2\nu} \quad , \quad u_\alpha = \frac{a - \nu}{\sqrt{2\nu}}$$

2/ pour $\nu > 30$

$$\alpha \sim \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$a = \chi^2_\alpha(\nu) \sim \frac{1}{2} [\sqrt{2\nu - 1} + u_\alpha]^2 \quad , \quad u_\alpha = \sqrt{2a} - \sqrt{2\nu - 1}$$

BIBLIOGRAPHIE

[1] N.B.S. - Tables of the binomial probability distribution. Applied Math. Series 6, National Bureau of Standards

$$C_n^s p^s (1 - p)^{n-s} \text{ et } \sum_{s=r}^{n} C_n^s p^s (1 - p)^{n-s} \quad \text{pour} \quad \begin{cases} n = 2(1)\dots\dots\dots\dots 49 \\ p = 0{,}01(0{,}01)\dots\dots\dots\dots 0{,}50 \\ s = 0(1)\dots\dots\dots\dots n - 1 \\ r = 0(1)\dots\dots\dots\dots n \end{cases}$$

[2] H.G. ROMIG - 50-100 binomial tables, Wiley (New-York)

$$C_n^s p^s (1-p)^{n-s} \quad \text{et} \quad \sum_{s=o}^{r} C_n^s p^s (1-p)^{n-s} \quad \text{pour} \quad \begin{cases} n = 50(5)\ldots\ldots\ldots\ldots 100 \\ p = 0{,}01(0{,}01)\ldots\ldots\ldots\ldots 0{,}50 \\ s = 0(1)\ldots\ldots\ldots\ldots n-1 \\ r = 0(1)\ldots\ldots\ldots\ldots n \end{cases}$$

[3] The Staff of the Harvard University Computation Laboratory, Tables of the Cumulative Binomial Probability Distributions. Harvard University Press, Cambridge, Mass. (1955).

[4] E.C. MOLINA - Poisson's exponential binomial limits, Van Nostrand, New-York (1942).

$$e^{-a}\frac{a^x}{x!} \quad \text{et} \quad \sum_{s=o}^{r} e^{-a}\frac{a^x}{x!} \quad \text{pour} \quad \begin{cases} a = 0{,}001(0{,}001)\ldots\ldots\ldots\ldots 0{,}01 \\ \quad 0{,}01(0{,}01)\ldots\ldots\ldots\ldots 0{,}40 \\ \quad 0{,}40(0{,}1)\ldots\ldots\ldots\ldots 15 \\ \quad 15(1)\ldots\ldots\ldots\ldots 100 \end{cases}$$

[5] WILLIAMSON and BRETHARTON - Tables of the negative binomial distribution Wiley (1963)

$$C_{x+k-1}^{k-1} p^k q^x \quad \text{et} \quad \sum_{x=o}^{r} C_{x+k-1}\, p^k q^x \quad \text{pour} \quad \begin{cases} p = 0{,}05\ldots\ldots\ldots\ldots 0{,}95 \\ k = 0\ldots\ldots\ldots\ldots 200 \\ \text{(par intervalles divers)} \end{cases}$$

[6] BARTKO - Approximating the negative binomial.Technometrics (1966) 8, 345-350.

[7] HALD et SINKBAEK - A table of percentage points of the χ^2 distribution Skandinavisk Actuarietidskrift (1950).

$\nu = 1(1)\ldots\ldots\ldots\ldots 100$

α et $1 - \alpha$: 0,05 - 0,1 - 0,5 - 1 - 2,5 - 5 - 10 - 20 - 30 - 40 - 50 (en %)

(partiellement reproduites dans "HALD" Statistical Tables and Formulas" Wiley (1952), qui donne aussi les tables de F pour :

ν_1 = 1..(1)..20..(2)..30..(5)..50..60..80..100..200..500..∞

ν_2 = 1..(1)..30..(2)..50..(5)..70..(10)..100..125..150..200..500.. 1000..∞

α = 50 - 70 - 90 - 95 - 97,5 - 99 - 99,5 - 99,95 (en %)

[8] K. PEARSON - Tables of the Incomplete Gamma Function Biometrika (1922). The University Press.

[9] K. PEARSON - Tables of the Incomplete Beta Function. The University Press, Cambridge (1934).

[10] D.B. OWEN - Handbook of Statistical Tables. Addison-Wesley Publishing Company, Reading, Mass. (1962).

[11] G.P. PATIL - On the evaluation of the negative binomial distribution with examples. Technometrics (1960) 2 p. 501-505.

[12] D.B. OWEN - "Letter to the Editor". Technometrics Vol. 8, n° 4 (Nov. 1966).