

REVUE DE STATISTIQUE APPLIQUÉE

J. CHAHINE

Une généralisation de la loi binomiale négative

Revue de statistique appliquée, tome 13, n° 4 (1965), p. 33-43

http://www.numdam.org/item?id=RSA_1965__13_4_33_0

© Société française de statistique, 1965, 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/>

UNE GÉNÉRALISATION DE LA LOI BINOMIALE NÉGATIVE

J. CHAHINE

Responsable du Département Statistique à l'A.C. NIELSEN C°

I - INTRODUCTION

Le problème qui sera traité dans cette note a été soulevé par une question pratique de sondage.

On disposait d'un fichier exhaustif d'une population donnée, en l'occurrence un fichier des établissements.

La population avait été stratifiée selon deux tailles : grands et petits établissements, en fonction d'un seuil de chiffre d'affaires.

Il n'était pas possible au moyen du fichier disponible, de connaître la taille de l'établissement.

On se proposait de tirer un échantillon de n individus, composés de n_1 grands établissements et n_2 petits établissements. La taille des établissements était par la suite vérifiée sur le terrain.

Le nombre de tirages nécessaires pour obtenir cet échantillon était évidemment aléatoire et il y avait lieu de déterminer le nombre moyen de tirages nécessaires, ainsi qu'une valeur critique ayant une probabilité donnée d'être dépassée.

Les nombres n_1 et n_2 sont supposés ajustés en fonction des taux de refus connus dans les deux strates. Si nous désignons par r_1 et r_2 ces taux de refus, on procèdera au tirage de $\frac{n_1}{1-r_1}$ et $\frac{n_2}{1-r_2}$ individus des strates 1 et 2.

Dans le cas où l'un des deux nombres n_1 ou n_2 est nul, le problème se ramène à une distribution binomiale négative.

Plus généralement, ayant une population stratifiée en k strates dont les importances relatives sont p_1, p_2, \dots, p_k ($p_1 + p_2 + \dots + p_k = 1$), quel est le nombre moyen de tirages nécessaires pour avoir un échantillon composé de n_1, n_2, \dots, n_k unités des strates 1, 2, ..., k .

Le problème peut être traité en supposant les tirages effectués avec remise ou sans remise.

Dans la présente note, nous traiterons le problème d'une urne à deux couleurs, $p_1 + p_2 = 1$ en supposant les tirages effectués avec remise et sans remise.

Un problème analogue a été étudié par F. CHARTIER⁽¹⁾: le problème traité était celui d'une urne à k couleurs équiprobables avec $n_1 = n_2 = \dots = n_k = 1$.

II - DISTRIBUTIONS BINOMIALE ET HYPERGEOMETRIQUE NEGATIVES

II.1 - Distribution Binomiale Négative.

Rappelons tout d'abord quelques caractéristiques de la distribution binomiale négative : X est la variable aléatoire, nombre de tirages nécessaires pour obtenir n boules blanches, les tirages étant faits avec remise, dans une urne où la proportion de boules blanches est $p(p+q = 1)$.

La loi de probabilité de x est

$$P(X = x) = p C_{x-1}^{n-1} p^{n-1} q^{x-1} \longrightarrow x \geq n$$

L'espérance mathématique de x est $\frac{n}{p}$ et sa variance $\frac{nq}{p^2}$.

Une propriété intéressante de cette distribution concerne sa fonction de répartition. Nous allons montrer que cette dernière se ramène à la répartition d'une distribution binomiale.

Soient :

$\bar{B}_n^p(x)$ la probabilité de faire moins de x tirages pour obtenir n arrivées (x variable aléatoire, n fixé)

$$\bar{B}_n^p(x) = \sum_{j=n}^{x-1} C_{j-1}^{n-1} p^n q^{j-n} \quad (\text{Binomiale négative})$$

$B_x^p(n)$ la probabilité d'obtenir moins de n arrivées dans une série de x tirages (n variable aléatoire, x fixé)

$$B_x^p(n) = \sum_{j=0}^{n-1} C_x^j p^j q^{x-j} \quad (\text{Binomiale})$$

Pour obtenir n arrivées en moins de x tirages, il est nécessaire et suffisant que dans les x - 1 tirages éventuellement permis, on observe n arrivées ou plus, cette dernière probabilité est égale à

$$1 - B_{x-1}^p(n)$$

Par conséquent nous avons :

$$\bar{B}_n^p(x) = 1 - B_{x-1}^p(n) \quad (1)$$

$$\sum_{j=1}^{x-1} C_{j-1}^{n-1} p^n q^{j-n} = \sum_{j=n}^{x-1} C_{x-1}^j p^j q^{x-1-j}$$

Une démonstration algébrique de cette propriété sera donnée un peu plus loin.

(1) F. CHARTIER - Le problème du Collectionneur - Revue Statistiques Appliquées Vol. VII, n° 1 1958.

Signalons encore que l'estimateur sans biais de p est :

$$\hat{p} = \frac{n-1}{x-1} \quad n > 1$$

En effet

$$\begin{aligned} E(\hat{p}) &= \sum_{x=n}^{\infty} \frac{n-1}{x-1} C_{x-1}^{n-1} p^n q^{x-n} \\ &= p \sum_{x=n}^{\infty} C_{x-2}^{n-2} p^{n-1} q^{x-n} \\ &= p \sum_{u=(n-1)}^{\infty} C_{u-1}^{(n-1)-1} p^{(n-1)} q^{u-(n-1)} \end{aligned}$$

en posant

$$u = x - 1$$

La somme écrite n'est autre que la somme des probabilités d'une loi binomiale négative de paramètre $(n-1)$.

Par conséquent, nous aurons

$$E(\hat{p}) = p.$$

II.2 - Distribution hypergéométrique négative.

Définissons de même quelques propriétés de la distribution que nous appellerons "hypergéométrique négative" par analogie avec la distribution binomiale négative quand les tirages sont faits sans remise.

Nous prendrons une urne de composition $N_1, N_2, N_1 + N_2 = N$.

La loi de probabilité de X , nombre de tirages nécessaires pour avoir n boules blanches provenant des N_1 boules blanches de l'urne est :

$$P(X = x) = P(x) = \frac{N_1 - n + 1}{N - x + 1} \frac{C_{N_1}^{n-1} C_{N_2}^{x-n}}{C_N^{x-1}} \quad n \leq x \leq N_2 + n$$

Le x^e tirage est obligatoirement une boule blanche dans une urne contenant $(N_1 - n + 1)$ boules blanches sur $(N - x + 1)$ boules, les $(x-1)$ tirages précédents comportant $(n-1)$ boules blanches et $(x-n)$ boules noires ($x-1$ tirages hypergéométriques).

$P(x)$ peut encore s'écrire :

$$P(x) = \alpha(n, N_2, N)(x-1)(x-2)\dots(x-n+1)(N-x)(N-x-1)\dots(N_2-x+n+1),$$

$$\text{avec } \alpha(n, N_2, N) = \frac{N_1! N_2!}{N! (N_1 - n)! (n-1)!}$$

Nous avons par ailleurs :

$$\alpha(n, N_2, N) \sum_{x=n}^{N_2+n} (x-1)(x-2)\dots(x-n+1)(N-x)(N-x-1)\dots(N_2-x+n+1) = 1 \quad (2)$$

Calculons dans ce cas $E(x)$

$$E(x) = \alpha(n, N_2, N) \sum_{x=n}^{N_2+n} x(x-1)\dots(x-n+1)(N-x)(N-x-1)\dots(N_2-x+n+1)$$

$$E(x) = \alpha(n, N_2, N) \sum_{u=(n+1)}^{N_2+(n+1)} (u-1)(u-2)\dots[u-(n+1)+1][(N+1)-u]\dots[N_2-u+(n+1)+1]$$

en posant $u = x + 1$.

La somme écrite est analogue à la somme calculée en (2) avec les paramètres $(n + 1)$, N_2 , $N + 1$.

Par conséquent

$$E(x) = \frac{\alpha(n, N_2, N)}{\alpha(n + 1, N_2, N + 1)}$$

On trouve ainsi

$$E(x) = n \frac{N + 1}{N_1 + 1}$$

Cette espérance est peu différente de $\frac{n}{p}$, espérance trouvée dans le cas avec remise, pour N_1 assez grand.

Nous allons illustrer cette formule par un petit exemple : ayant un trousseau de N clés, quel est le nombre moyen d'essais nécessaires pour ouvrir une porte en faisant *attention* aux clés déjà essayées et en ne faisant *pas attention*.

Dans ce cas :

$$p = \frac{1}{N} \quad N_1 = 1 \quad n = 1$$

En ne faisant pas attention :

$$E(x) = \frac{1}{N} = N$$

En faisant attention :

$$E(x) = 1 \frac{N + 1}{1 + 1} = \frac{N + 1}{2}$$

Le nombre d'essais est pratiquement réduit de moitié quand on n'essaie pas une deuxième fois la même clé.

De la même manière, on calcule $V(x)$ au moyen du moment factoriel d'ordre : deux $E[x(x + 1)]$.

$$V(x) = E[x(x + 1)] - E(x) - E(x)^2$$

On trouve :

$$E[x(x + 1)] = \frac{\alpha(n, N_2, N)}{\alpha(n + 2, N_2, N + 2)}$$

et

$$V(x) = \frac{n(N+1)N_2(N_1-n+1)}{(N_1+1)^2(N_1+2)}$$

Cette variance est évidemment inférieure à celle du tirage avec remise :

$$\frac{n N N_2}{N_1^2}$$

De même, un estimateur sans biais de p c'est-à-dire de $\frac{N_1}{N}$ est $\hat{p} = \frac{n-1}{x-1}$, $n > 1$

En effet,

$$E(\hat{p}) = (n-1) \alpha(n, N_2, N) \sum_{x=n}^{N_2+n} (x-2)(x-3)\dots(x-n+1)(N-x)(N-x-1)\dots(N_2-x+n+1)$$

On trouve :

$$E(\hat{p}) = (n-1) \frac{\alpha(n, N_2, N)}{\alpha(n-1, N_2, N-1)}$$

Soit :

$$E(\hat{p}) = (n-1) \frac{N_1}{N(n-1)} = \frac{N_1}{N} = p$$

Signalons aussi une propriété analogue à (1) pour la distribution hypergéométrique négative.

Soient :

$\bar{H}_n^{N_1}(x)$ la probabilité de faire moins de x tirages pour obtenir n arrivées (x variable aléatoire, n fixé, loi hypergéométrique négative) ;

$H_x^{N_1}(n)$ la probabilité d'obtenir moins de n arrivées dans une série de x tirages (n variable aléatoire, x fixé, loi hypergéométrique).

Pour obtenir n arrivées en moins de x tirages, il est nécessaire et suffisant que dans les $(x-1)$ tirages éventuellement permis, on observe n arrivées ou plus, cette dernière probabilité est égale à :

$$1 - H_{x-1}^{N_1}(n)$$

Par conséquent, nous avons :

$$\bar{H}_n^{N_1}(x) = 1 - H_{x-1}^{N_1}(n)$$

III - GENERALISATION DU PROBLEME A DEUX COULEURS

III.1 - Tirage avec remise.

Soit X le nombre de tirages nécessaires pour obtenir un échantillon de n_1 et n_2 individus des strates 1 et 2.

La loi de probabilité de X s'écrit en distinguant les cas où la dernière boule tirée est de couleur 1 ou 2 :

$$P(X = x) = P(x) = p_1 C_{x-1}^{n_1-1} p_1^{n_1-1} p_2^{x-n_1} + p_2 C_{x-1}^{n_2-1} p_2^{n_2-1} p_1^{x-n_2}$$

$$x \geq n_1 + n_2 = n$$

$P(x)$ est la somme de deux distributions binomiales négatives dont le domaine de variation est restreint de n à l'infini.

Démontrons d'abord que $\sum_{x=n}^{\infty} p(x) = 1$.

$$\sum_{x=n}^{\infty} p(x) = \frac{p_1^{n_1}}{(n_1-1)!} \sum_{x=n}^{\infty} (x-1)(x-2)\dots(x-n_1+1) p_2^{x-n_1} + \frac{p_2^{n_2}}{(n_2-1)!} \sum_{x=n}^{\infty} (x-1)(x-2)\dots(x-n_2+1) p_1^{x-n_2} \quad (3)$$

Posons :

$$\delta(p) = \sum_{x=n}^{\infty} p^{x-1} = p^{n-1} \frac{1}{1-p} \quad (4)$$

Nous pouvons alors écrire :

$$\sum_{x=n}^{\infty} p(x) = \frac{p_1^{n_1}}{(n_1-1)!} \delta^{(n_1-1)}(p_2) + \frac{p_2^{n_2}}{(n_2-1)!} \delta^{(n_2-1)}(p_1) \quad (5)$$

où $\delta^{(n_1-1)}(p_2)$ est la dérivée d'ordre (n_1-1) .

En prenant la dérivée d'ordre (n_1-1) dans (4) au moyen de la formule de Leibniz nous trouvons :

$$\delta^{(n_1-1)}(p_2) = \sum_{i=0}^{n_1-1} C_{n_1-1}^i \frac{(n-1)! i!}{(n_2+i)!} \frac{p_2^{n_2+i}}{p_1^{i+1}} = \sum_{i=0}^{n_1-1} \frac{(n-1)! (n_1-1)!}{(n_1-i-1)! (n_2+i)!} \frac{p_2^{n_2+i}}{p_1^{i+1}}$$

$$\delta^{(n_2-1)}(p_1) = \sum_{i=0}^{n_2-1} C_{n_2-1}^i \frac{(n-1)! i!}{(n_1+i)!} \frac{p_1^{n_1+i}}{p_2^{i+1}} = \sum_{i=0}^{n_2-1} \frac{(n-1)! (n_2-1)!}{(n_2-i-1)! (n_1+i)!} \frac{p_1^{n_1+i}}{p_2^{i+1}}$$

En remplaçant les valeurs de $\delta^{(n_1-1)}(p_2)$ et $\delta^{(n_2-1)}(p_1)$ dans (5), nous trouvons :

$$\sum_{x=n}^{\infty} p(x) = \sum_{i=0}^{n_1-1} \frac{(n-1)!}{(n_1-i-1)! (n_2+i)!} p_2^{n_2+i} p_1^{n_1-i-1} + \sum_{i=0}^{n_2-1} \frac{(n-1)!}{(n_2-i-1)! (n_1+i)!} p_1^{n_1+i} p_2^{n_2-i-1}$$

ou encore :

$$\sum_{x=n}^{\infty} p(x) = \sum_{i=0}^{n_1-1} C_{n_1+n_2-1}^{n_2+i} p_2^{n_2+i} p_1^{n_1-i-1} + \sum_{i=0}^{n_2-1} C_{n_1+n_2-1}^{n_1+i} p_1^{n_1+i} p_2^{n_2-i-1} \quad (6)$$

Ceci n'est autre,

$$(p_1 + p_2)^{n_1+n_2-1}$$

D'où :

$$\sum_{x=n}^{\infty} p(x) = (p_1 + p_2)^{n_1+n_2-1} = 1$$

Ce calcul est utile dans la mesure où il permettra de calculer par une méthode directe l'espérance mathématique et la variance de x .

Nous avons par ailleurs démontré en comparant (3) et (6) que

$$\sum_{x=n}^{\infty} C_{x-1}^{n_1-1} p_1^{n_1} p_2^{x-n_1} = \sum_{i=0}^{n_1-1} C_{n-1}^{n_2+i} p_2^{n_2+i} p_1^{n_1-i-1}$$

or

$$\sum_{x=n}^{\infty} C_{x-1}^{n_1-1} p_1^{n_1} p_2^{x-n_1} = 1 - \sum_{x=n_1}^{n-1} C_{x-1}^{n_1-1} p_1^{n_1} p_2^{x-n_1} = 1 - \bar{B}_{n_1}^{p_1}(n)$$

Par ailleurs :

$$\sum_{i=0}^{n_1-1} C_{n-1}^{n_2+i} p_2^{n_2+i} p_1^{n_1-i-1} = \sum_{j=0}^{n_1-1} C_{n-1}^j p_1^j p_2^{n-1-j} = B_{n-1}^{p_1}(n_1)$$

en posant $j = n_1 - i - 1$.

D'où :

$$1 - \bar{B}_{n_1}^{p_1}(n) = B_{n-1}^{p_1}(n_1)$$

Ceci n'est autre que la relation (1) dans laquelle on aurait remplacé

$$\begin{array}{l} p \text{ par } p_1 \\ n \text{ par } n_1 \\ x \text{ par } n \end{array}$$

Espérance mathématique de x .

Nous avons :

$$E(x) = \frac{p_1^{n_1}}{(n_1-1)!} \sum_{x=n}^{\infty} x(x-1)\dots(x-n_1+1) p_2^{x-n_1} + \frac{p_2^{n_2}}{(n_2-1)!} \sum_{x=n}^{\infty} x(x-1)\dots(x-n_2+1) p_1^{x-n_2} \quad (7)$$

$$\sum_{x=n}^{\infty} x(x-1)\dots(x-n_1+1) p_2^{x-n_1} = \sum_{u=(n_1+1)+n_2}^{\infty} (u-1)(u-2)\dots[u-(n_1+1)+1] p_2^{u-(n_1+1)}$$

Ceci n'est autre que $\delta^{n_1}(p_2)$ dans laquelle on aurait remplacé n_1 par (n_1+1) .

D'où :

$$E(x) = \frac{p_1^{n_1}}{(n_1-1)!} \sum_{\alpha=0}^{n_1} C_{n_1}^{\alpha} \frac{n!}{(n_2+\alpha)!} \frac{p_2^{n_2+\alpha}}{p_1^{\alpha+1}} + \frac{p_2^{n_2}}{(n_2-1)!} \sum_{\alpha=0}^{n_2} C_{n_2}^{\alpha} \frac{n!}{(n_2+\alpha)!} \frac{p_1^{n_1+\alpha}}{p_2^{\alpha+1}}$$

$$E(x) = \frac{n_1}{p_1} \sum_{\alpha=0}^{n_1} C_{n_1+n_2}^{n_1-\alpha} p_1^{n_1-\alpha} p_2^{n_2+\alpha} + \frac{n_2}{p_2} \sum_{\alpha=0}^{n_2} C_{n_1+n_2}^{n_2-\alpha} p_2^{n_2-\alpha} p_1^{n_1+\alpha}$$

$$\sum_{\alpha=0}^{n_1} C_{n_1+n_2}^{n_1-\alpha} p_1^{n_1-\alpha} p_2^{n_2+\alpha}$$

n'est autre que la probabilité dans un tirage de Bernoulli d'obtenir moins de $n + 1$ boules de couleur (1), en n tirages. Cette probabilité est égale à :

$$B_n^{p_1} (n_1 + 1)$$

Nous obtenons ainsi :

$$E(x) = \frac{n_1}{p_1} B_n^{p_1} (n_1 + 1) + \frac{n_2}{p_2} B_n^{p_2} (n_2 + 1) \quad (8)$$

Cette formule est à rapprocher de celle donnant l'espérance de la loi binomiale négative.

D'autre part $0 \leq B \leq 1$ entraîne

$$E(x) \leq \frac{n_1}{p_1} + \frac{n_2}{p_2}$$

Cette inégalité était évidente a priori, on a besoin de moins de tirages pour avoir n_1 et n_2 indépendamment.

Nous allons maintenant calculer $E(x)$ en utilisant la relation (1). Nous avons :

$$\frac{p_1^{n_1}}{(n_1 - 1)!} \sum_{x=n_1}^{\infty} (x - 1) (x - 2) \dots (x - n_1 + 1) p_2^{x-n_1} = 1$$

en considérant une distribution binomiale négative de paramètre n et p .

D'après (7) nous pouvons écrire :

$$E(x) = \frac{n_1}{p_1} \left\{ \frac{p_1^{n_1+1}}{n_1!} \sum_{u=(n_1+1)+n_2}^{\infty} (u - 1) (u - 2) \dots [u - (n_1 + 1) + 1] p_2^{u-n_1} \right\} +$$

$$\frac{n_2}{p_2} \left\{ \frac{p_2^{n_2+1}}{n_2!} \sum_{u=(n_2+1)+n_1}^{\infty} (u - 1) (u - 2) \dots [u - (n_2 + 1) + 1] p_1^{u-n_2} \right\}$$

en posant $x = u - 1$.

L'expression entre accolades n'est autre que la probabilité d'avoir $(n_1 + 1)$ boules blanches en $n + 1$ tirages ou plus, cette probabilité est la même que celle d'obtenir moins de $n_1 + 1$ boules blanches en n tirages, or cette probabilité n'est autre que :

$$B_n^{p_1} (n_1 + 1) = \sum_{x=0}^{n_1} C_n^x p_1^x (1 - p_1)^{n-x} \quad \text{Relation (1)}$$

On retrouve donc :

$$E(x) = \frac{n_1}{p_1} B_n^{p_1} (n_1 + 1) + \frac{n_2}{p_2} B_n^{p_2} (n_2 + 1)$$

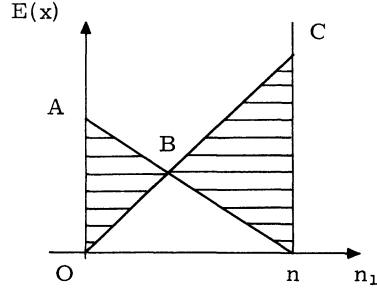
Cas particulier. Trouver le minimum de $E(x)$ lié par la condition $n_1 + n_2 = n$.

Nous avons :

$$E(x) \geq \frac{n_1}{p_1}$$

$$E(x) \geq \frac{n - n_1}{p_2}$$

Le domaine de variation de $E(x)$ est le domaine hachuré et $E(x)$ est respectivement égale à $\frac{n_1}{p_1}$, $\frac{n - n_1}{p_2}$ pour $n_1 = n$, $n_1 = 0$. Le point B définit donc le minimum de $E(x)$ lié par la condition $n_1 + n_2 = n$. En ce point nous



$$\frac{n_1}{p_1} = \frac{n - n_1}{p_2} = \frac{n_2}{p_2} = n.$$

Ceci signifie que la composition de l'échantillon tiré est proportionnelle à celle de l'urne. Le minimum de $E(x)$ est égal dans ce cas à :

$$\text{Min} [E(x)/(n_1 + n_2 = n)] = \frac{n_1}{p_1} [B_n^{p_1} (n_1 + 1) + B_n^{p_2} (n_2 + 1)]$$

$$B_n^{p_1} (n_1 + 1) = \sum_{x=0}^{n_1} C_n^x p_1^x p_2^{n-x}$$

$$B_n^{p_2} (n_2 + 1) = \sum_{x=0}^{n_2} C_n^x p_2^x p_1^{n-x} = \sum_{x=n_1}^n C_n^x p_1^x p_2^{n-x}$$

en remplaçant x par $(n - x)$.

D'où :

$$B_n^{p_1} (n_1 + 1) + B_n^{p_2} (n_2 + 1) = \sum_{x=0}^n C_n^x p_1^x p_2^{n-x} + C_n^{n_1} p_1^{n_1} p_2^{n_2} =$$

$$(p_1 + p_2)^n + C_n^{n_1} p_1^{n_1} p_2^{n_2} = 1 + C_n^{n_1} p_1^{n_1} p_2^{n_2}$$

Par conséquent :

$$\text{Min} [E(x)/(n_1 + n_2 = n)] = \frac{n_1}{p_1} (1 + C_n^{n_1} p_1^{n_1} p_2^{n_2}) = \frac{n_2}{p_2} (1 + C_n^{n_2} p_2^{n_2} p_1^{n_1})$$

Variance de x .

Nous calculerons d'abord le moment factoriel d'ordre deux $E[x(x+1)]$.

Nous avons :

$$E[x(x+1)] = \frac{p_1^{n_1}}{(n_1 - 1)!} \sum_{x=n_1}^{\infty} (x+1) x \dots (x - n_1 + 1) p_2^{x-n_1} +$$

$$\frac{p_2^{n_2}}{(n_2 - 1)!} \sum_{x=n}^{\infty} (x + 1) x \dots (x - n_2 + 1) p_1^{x-n_2}$$

Cette expression se ramène à :

$$E(x(x + 1)) = \frac{n_1(n_1 + 1)}{p_1^2} B_{n+1}^{p_1} (n_1 + 2) + \frac{n_2(n_2 + 1)}{p_2^2} B_{n+1}^{p_2} (n_2 + 2)$$

$$V(x) = E(x(x + 1)) - E(x) - E(x)^2$$

Il est facile de voir que :

$$B_{n+1}^{p_1} (n_1 + 2) = B_n^{p_1} (n_1 + 1) + p_2 C_n^{n_1+1} p_1^{n_1+1} p_2^{n_2-1}$$

V(x) s'écrit alors :

$$V(x) = \frac{n_1 p_2}{p_1^2} [B_n^{p_1} (n_1 + 1) + (n_1 - 1) C_n^{n_1+1} p_1^{n_1+1} p_2^{n_2-1}] +$$

$$\frac{n_2 p_1}{p_2^2} [B_n^{p_2} (n_2 + 1) + (n_2 - 1) C_n^{n_2} p_2^{n_2+1} p_1^{n_1-1}]$$

$$- 2 \frac{n_1 n_2}{p_1 p_2} B_n^{p_1} (n_1 + 1) B_n^{p_2} (n_2 + 1)$$

III.2 - Tirage sans remise.

La loi de probabilité de x s'écrit dans ce cas :

$$P(x) = \frac{N_1 - n_1 + 1}{N - x + 1} \frac{C_{N_1}^{n_1-1} C_{N_2}^{x-n_1}}{C_N^{x-1}} + \frac{N_2 - n_2 + 1}{N - x + 1} \frac{C_{N_2}^{x-n_2} C_{N_1}^{x-n_2}}{C_N^{x-1}}$$

en se plaçant dans le tirage d'ordre (x - 1), selon que la dernière boule tirée est de couleur 1 ou 2.

Dans la première expression x varie de n à N₂ + n₁ et dans la deuxième expression x varie de n à N₁ + n₂.

P(x) peut encore s'écrire sous la forme :

$$P(x) = \frac{N_1! N_2!}{N! (N_1 - n_1)! (n_1 - 1)!} \frac{(x - 1)! (N - x)!}{(x - n_1)! (N_2 - x + n_1)!} + \frac{N_1! N_2!}{N! (N_2 - n_2)! (n_2 - 1)!} \frac{(x - 1)! (N - x)!}{(x - n_2)! (N_1 - x + n_2)!} \quad (9)$$

Nous avons ainsi :

$$E(x) = \frac{N_1! N_2!}{N! (N_1 - n_1)! (n_1 - 1)!} \sum_{x=n}^{N_2+n_1} \frac{x! (N - x)!}{(x - n_1)! (N_2 - x + n_1)!} + \frac{N_1! N_2!}{N! (N_2 - n_2)! (n_2 - 1)!} \sum_{x=n}^{N_1+n_2} \frac{x! (N - x)!}{(x - n_2)! (N_1 - x + n_2)!}$$

ou encore, en posant x = u - 1 :

$$E(x) = \frac{(N + 1)}{(N_1 + 1)} n_1 \left[\frac{(N_1 + 1)! N_2!}{(N + 1)! (N_1 - n_1)! n_1!} \sum_{u=(n_1+1)+n_2}^{N_2+n_1+1} \frac{(u - 1)! [(N + 1) - u + 1]!}{[u - (n_1 + 1)]! [N_2 - u + (n_1 + 1)]!} \right]$$

$$+ \frac{N+1}{(N_2+1)} n_2 \left[\frac{(N_2+1)! N_1!}{(N+1)! (N_2-n_2)! n_2!} \sum_{u=(n_2+1)+n_1}^{n_1+n_2+1} \frac{(u-1)! [(N+1)-u]!}{[u-(n_2+1)]! [N_1-u+(n_2+1)]!} \right]$$

L'expression entre crochets est une somme de probabilités identiques à (9) dont les paramètres sont respectivement

$$(N_1+1), N_2, (n_1+1), n_2$$

Elle représente la probabilité d'avoir n_1+1 boules blanches en plus de $n+1$ tirages. Comme plus haut, cette probabilité est celle d'obtenir moins de n_1+1 boules blanches en n tirages.

Cette dernière probabilité est donnée par la fonction de répartition hypergéométrique.

$$H_{n+1}^{N+1}(x) = \text{Prob.}\{X < x\}$$

avec n tirages dans une urne comportant $N+1$ boules dont N_1+1 blanches.

$E(x)$ s'écrit donc :

$$E(x) = \frac{N+1}{N_1+1} n_1 H_{n+1}^{N+1}(n_1+1) + \frac{N+1}{N_2+1} n_2 H_{n+1}^{N+1}(n_2+1) \quad (10)$$

$\frac{N+1}{N_1+1} n_1$ est comme nous l'avons vu, l'espérance mathématique, dans une loi hypergéométrique négative.

Par conséquent, le résultat trouvé en (10) est absolument analogue à celui trouvé en (8). La variance de x se calcule de la même manière que dans le cas avec remise, en calculant $E[x(x+1)]$.