1 Loi géométrique

Soit X une variable aléatoire de loi de géométrique $\mathcal{G}(p)$, ($p \in]0,1[$).

1) Rappelons que si Z est une variable aléatoire à valeurs dans les entiers positifs, on a

$$\mathbb{E}(Z) = \sum_{k\geq 1} \mathbb{P}(Z \geq k). \tag{1}$$

On a, pour $k \geq 1$

$$\mathbb{P}(X \geq k) = \sum_{j\geq k}(1-p)p^{j-1} = (1-p)\sum_{j\geq k} p^{j-1} = (1-p)p^{k-1}\sum_{j\geq 0} p^j = p^{k-1}\frac{1-p}{1-p} = p^{k-1}.$$

Donc en utilisant (1), on obtient :

$$\mathbb{E}(X) = \sum_{k\geq 1} \mathbb{P}(X \geq k) = \sum_{k\geq 1} p^{k-1} = \frac{1}{1-p}. \tag{2}$$

2) Calculons maintenant la fonction génératrice de X. Pour $|t| < 1/p$, on a

$$G_X(t) = \mathbb{E}\left(t^X\right) = \sum_{k\geq 1} \mathbb{P}(X=k)t^k = (1-p)t\sum_{k\geq 1}(pt)^{k-1} = \frac{(1-p)t}{1-pt}. \tag{3}$$

Pour, $|t| < 1/p$, on a

$$\begin{aligned} G'_X(t) &= \frac{(1-p)(1-pt)+pt(1-p)}{(1-pt)^2} = \frac{1-p}{(1-pt)^2} \\ G''_X(t) &= \frac{2(1-p)p}{(1-pt)^3}. \end{aligned}$$

Donc on obtient,

$$\mathbb{E}(X) = G'_X(1) = \frac{1-p}{(1-p)^2} = \frac{1}{1-p}, \tag{4}$$

et

$$\begin{aligned} \mathrm{Var}(X) &= G''_X(1) + G'_X(1) - \left[G'_X(1)\right]^2 \\ &= \frac{2p}{(1-p)^2} + \frac{1}{1-p} - \frac{1}{(1-p)^2} \\ &= \frac{p}{(1-p)^2}. \end{aligned} \tag{5}$$

2 Loi hypergéométrique

On interroge au hasard n individus différents dans une population de N individus dont N_1 fument et $N_2 = N - N_1$ ne fument pas. Soit X le nombre de fumeurs parmi les n interrogés.

a) Déterminer $P(X = k)$ $(0 \leq k \leq n)$; calculer l'espérance et la variance de X.

b) Que se passe t'il à la limite lorsque

i) n est fixe, $N \to \infty$ et $\frac{N_1}{N} \to p$ $(0 < p < 1)$.

ii) $n,N,N_1 \to \infty$ et $\frac{nN_1}{N} \to \lambda$ $(\lambda > 0)$.

Pour simplifier on suppose que $n \leq \min(N_1,N_2)$. Il s'agit d'un sondage sans remise. Commençons par dénombrer le nombre d'échantillons possibles. Il s'agit de compter le nombre de sous-ensembles à n éléments à choisir parmi $N = N_1 + N_2$. Il y a donc C_N^n possibilités. Soit k un entier naturel inférieur à n. Le nombre d'échantillons qui contiennent exactement k fumeurs est $C_{N_1}^k C_{N_2}^{n-k}$. En effet, il faut choisir k fumeurs parmi N_1 puis $n-k$ non fumeurs parmi N_2. La loi de X est donc donnée par

$$P(X = k) = \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n}. \tag{6}$$

Calculons maintenant les deux premiers moments de X.

- **Première méthode**

  Pour un ordre donné de $\{1,2,\ldots,n\}$, on utilise la décomposition de X

  $$X = \sum_{i=1}^n Y_i \quad \text{où } Y_i = 1 \quad \text{si le } i\text{-ème individu fume} \quad Y_i = 0 \quad \text{sinon} \quad (i = 1,2,\ldots,n).$$

  Les variables aléatoires $(Y_i)_{i=1,\ldots,n}$ ont toutes la même loi mais ne sont pas indépendantes. Par les mêmes arguments que ceux utilisés pour établir (6), on montre que pour $i,j \in \{1,\ldots,n\}$ avec $i \neq j$ on a

  $$\begin{aligned} E(Y_i^2) &= E(Y_i) = P(Y_i = 1) = \frac{N_1}{N}, \\ E(Y_iY_j) &= P(Y_i = 1, Y_j = 1) = \frac{C_{N_1}^2 C_{N_2}^0}{C_N^2} = \frac{N_1(N_1-1)}{N(N-1)}. \end{aligned}$$

  Donc

  $$E(X) = E(\sum_{i=1}^n Y_i) = \sum_{i=1}^n E(Y_i) = \frac{nN_1}{N}, \tag{7}$$

  et

  $$E(X^2) = E\left[\left(\sum_{i=1}^n Y_i\right)^2\right] = \sum_{i=1}^n E(Y_i) + 2\sum_{1\leq i<j\leq n} E(Y_iY_j) = \frac{nN_1}{N} + \frac{n(n-1)N_1(N_1-1)}{N(N-1)}.$$

  Après factorisation, on en déduit

  $$\begin{aligned} \text{Var } X = E(X^2) - (E(X))^2 &= \frac{nN_1}{N} + \frac{n(n-1)N_1(N_1-1)}{N(N-1)} - \left(\frac{nN_1}{N}\right)^2 \\ &= \frac{nN_1N_2(N-n)}{N^2(N-1)}. \end{aligned} \tag{8}$$

- **Deuxième méthode**

  On a pour $n,N_1 \geq 1$,

  $$E(X) = \sum_{k=0}^n kP(X = k) = \sum_{k=0}^n k\frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n} = \frac{nN_1}{N}\sum_{j=0}^{n-1}\frac{C_{N_1-1}^j C_{N_2}^{n-1-j}}{C_{N-1}^{n-1}} = \frac{nN_1}{N}.$$

  En effet, l'égalité $\sum_{j=0}^{n-1}\frac{C_{N_1-1}^j C_{N_2}^{n-1-j}}{C_{N-1}^{n-1}} = 1$ s'obtient en considérant l'équation de normalisation pour la loi hypergéométrique de paramètres $N_1-1,N_2,n-1$. Par le même raisonement on obtient d'autre part pour $n,N_1 \geq 2$

  $$\begin{aligned} E[X(X-1)] &= \sum_{k=0}^n k(k-1)P(X = k) = \sum_{k=0}^n k(k-1)\frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n} \\ &= \frac{nN_1(n-1)(N_1-1)}{N(N-1)}\sum_{j=0}^{n-2}\frac{C_{N_1-2}^j C_{N_2}^{n-2-j}}{C_{N-2}^{n-2}} = \frac{nN_1(n-1)(N_1-1)}{N(N-1)}. \end{aligned}$$

D'où

$$\begin{aligned}\operatorname{Var} X = E[X(X-1)] + E(X)[1-E(X)] &= \frac{nN_1(n-1)(N_1-1)}{N(N-1)} + \frac{nN_1}{N}\left(1-\frac{nN_1}{N}\right)\\ &= \frac{nN_1N_2(N-n)}{N^2(N-1)}.\end{aligned}$$

On retrouve donc bien, en utilisant cette méthode, les résutats (7) et (8).

Etudions maintenant les propriétés asymptotiques de la loi hypergéométrique.

i) n est fixe, $N \to \infty$ et $\frac{N_1}{N} \to p$ $(0 < p < 1)$. On s'attend intuitivement à trouver la loi d'échantillonnage pour un tirage avec remise. C'est-à-dire la loi binomiale. On a pour $k \in \{0,\ldots,n\}$

$$\begin{aligned}P(X=k) &= \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n}\\ &= C_n^k \frac{N_1(N_1-1)\cdots(N_1-k+1)N_2(N_2-1)\cdots(N_2-n+k+1)}{N(N-1)\cdots(N-n+1)}\\ &= C_n^k\left(\frac{N_1}{N}\right)^k\left(\frac{N_2}{N}\right)^{n-k}\frac{(1-\frac{1}{N_1})\cdots(1-\frac{k-1}{N_1})(1-\frac{1}{N_2})\cdots(1-\frac{n-k-1}{N_2})}{(1-\frac{1}{N})\cdots(1-\frac{n-1}{N})}.\end{aligned}$$

On trouve finalement

$$\lim_{N\to\infty} P(X=k) = C_n^k p^k (1-p)^{n-k}.$$

C'est la loi binomiale de paramètres n,p.

ii) $n,N,N_1 \to \infty$ et $\frac{nN_1}{N} \to \lambda$ $(\lambda > 0)$. On écrit ici

$$\begin{aligned}P(X=0) = \frac{C_{N_1}^0 C_{N_2}^n}{C_N^n} &= \frac{(N-N_1)(N-1-N_1)\cdots(N-n+1-N_1)}{N(N-1)\ldots(N-n+1)}\\ &= \left(1-\frac{N_1}{N}\right)\left(1-\frac{N_1}{N-1}\right)\cdots\left(1-\frac{N_1}{N-n+1}\right).\end{aligned}$$

D'où

$$\left(1-\frac{N_1}{N}\right)^n \geq P(X=0) \geq \left(1-\frac{N_1}{N-n+1}\right)^n.$$

En passant à la limite dans ces inégalités on obtient $\lim_{N\to\infty} P(X=0) = \exp(-\lambda)$. D'autre part pour $k \in \mathbb{N}$

$$\frac{P(X=k+1)}{P(X=k)} = \frac{(n-k)(N_1-k)}{(k+1)(N_2-n+k+1)} = \left(\frac{nN_1}{N}\right)\left(\frac{1}{k+1}\right)\frac{\left(1-\frac{k}{N_1}\right)\left(1-\frac{k}{n}\right)}{1-\frac{N_1+n-k-1}{N}}.$$

On en déduit que

$$\lim_{N\to\infty}\frac{P(X=k+1)}{P(X=k)} = \frac{\lambda}{k+1}.$$

Par suite

$$\lim_{N\to\infty} P(X=k) = \exp(-\lambda)\frac{\lambda^k}{k!},$$

il s'agit de la loi de Poisson de paramètre λ.