

Correction du TD 3. Intégration et Probabilités

1 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, $N \rightarrow \infty$ et $\frac{N_1}{N} \rightarrow p$ ($0 < p < 1$).

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

Correction

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

$$P(X = k) = \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n}. \quad (1)$$

Calculons maintenant les deux premiers moments de X .

– **Première méthode**

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

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

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

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

Donc

$$E(X) = E\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n E(Y_i) = \frac{nN_1}{N}, \quad (2)$$

et

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

Après factorisation, on en déduit

$$\begin{aligned}\text{Var } X = E(X^2) - (E(X))^2 &= \frac{nN_1}{N} + \frac{n(n-1)N_1(N_1-1)}{N(N-1)} - \left(\frac{nN_1}{N}\right)^2 \\ &= \frac{nN_1N_2(N-n)}{N^2(N-1)}.\end{aligned}\quad (3)$$

– **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 raisonnement 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

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

On retrouve donc bien, en utilisant cette méthode, les résultats (2) et (3).

Etudions maintenant les propriétés asymptotiques de la loi hypergéométrique.

- i) n est fixe, $N \rightarrow \infty$ et $\frac{N_1}{N} \rightarrow p$ ($0 < p < 1$). 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 $k \in \{0, \dots, n\}$

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

On trouve finalement

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

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

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

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

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 $\lim_{N \rightarrow \infty} P(X = 0) = \exp(-\lambda)$.
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

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

Par suite

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

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