IUFM de Toulouse
Préparation au CAPES de Maths
2000-2001

Feuille d'exercices 3

1 Correction de l'exercice sur la 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

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

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 (1), 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

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

et

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

Après factorisation, on en déduit

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

- **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 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

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

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

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

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

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.

2 Modèles binomiaux et de Poisson

1) Pour estimer la taille d'une population animale, on installe successivement r trappes. Soit q la probabilité qu'un animal soit capturé par une trappe et N la taille de la population avant l'expérience.

   a) Supposant que les animaux capturés ne sont pas relachés, donner la probabilité que les effectifs de capture soient $n_1, \dots, n_r$.

   b) Une méthode alternative consiste à relacher les animaux après *marquage*. Donner la probabilité que les effectifs des captures soient $n_1, \dots, n_r$ et que les effectifs d'animaux recapturés soient $m_1, \dots, m_{r-1}$.

2) Soit une population à k modalités en proportions $p_1, \dots, p_k$. On en tire n individus (avec remise). Déterminer la loi de $(n_1, \dots, n_k)$, nombres d'individus prenant les modalités $1, \dots, k$. En déduire la loi marginale de n_1.

3) La ville de Londres étant divisée en 576 districts, on a relevé le nombre d'impacts de bombes subi par chaque district pendant la seconde guerre mondiale. Les résultats sont donnés dans le tableau suivant :

k	0	1	2	3	4	5 et plus
N_k	229	211	93	35	7	1

a) Comment tester l'adéquation de ces observations à une loi de Poisson? (On rappelle que, si $Z \sim \chi^2(4)$ alors $P(Z \leq 7.8) = 0.9$ et $P(Z \leq 9.5) = 0.95$)

b) Si l'adéquation à une loi de Poisson est acceptable, donner la probabilité que la dernière observation vaille 6.

3 Approximation de Poisson de De Moivre

1) Trouver la probabilité (approximative) que, sur 500 personnes, $k = 0, 1, 2, 3$ aient leur anniversaire le 20 janvier.

2) Une compagnie de téléphone déservant 2000 abonnés cherche à déterminer le nombre N de lignes nécessaires pour que le réseaux ne soit saturé qu'avec une probabilité inférieure à 1%. En moyenne, chaque utilisateur donne un appel de deux minutes toutes les heures. Déterminer $N_{\min}$ en utilisant l'approximation de De Moivre.

3) On effectue une enquête, durant une épidémie de grippe, dans le but de connaître la proportion p de personnes présentant ensuite des complications graves. On observe un échantillon représentatif de 400 personnes et pour un tel échantillon 40 personnes ont présenté des complications.

a) Donner un intervalle de confiance pour p au risque 5%.

b) On désire que la valeur estimée $\hat{p}$ diffère de la proportion inconnue exacte p de moins de $0,005$ avec une probabilité égale à 95%. Quel sera l'effectif d'un tel échantillon?

c) Quel devrait être le risque pour obtenir le même intervalle qu'à la question précédente en conservant l'effectif $n = 400$. Quelles conclusions peut-on en tirer?

4) Soit $N_n \sim \mathcal{B}(n, p)$. On note

$$N_n^* = \frac{N_n - np}{\sqrt{np(1-p)}}.$$

Construire la transformée de Laplace de N_n^*, $M_n(t) = \mathbb{E}[\exp(tN_n^*)]$ et déduire de l'inégalité de Markov,

$$P(N_n^* \geq s) \leq M_n(t)e^{-ts} \quad (s, t > 0),$$

une majoration de $P(N_n^* \geq s)$.

4 Introduction à l'estimation

On considère $X_1, X_2, \ldots, X_n$ des variables aléatoires i.i.d. de loi de Poisson de paramètre θ inconnu, $\theta \in [0, \infty[$.

1) On cherche à estimer θ.

   a) L'estimateur $\overline{X} = \frac{X_1+\ldots+X_n}{n}$ est-il d'espérance θ (sans biais)? Que vaut $E(\overline{X}-\theta)^2$ (risque quadratique)?

   b) Montrer que, pour tout estimateur sans biais $h(X_1, \ldots, X_n)$, il existe un estimateur $\hat{h}(X_1 + \ldots + X_n)$ de θ de risque quadratique inférieur : $\overline{X}$ est une statistique exhaustive.

   c) Quel est l'estimateur qui maximise la probabilité de la réalisation observée (maximum de vraisemblance)?

2) On cherche maintenant à estimer $e^{-l\theta}$, probabilité pour que sur l expériences futures, on observe toujours 0.

   a) On propose l'estimateur, $e^{-l\overline{X}}$? Est-il sans biais? Quel est son risque quadratique?

   b) Déterminer une fonction g de N dans R telle que $g(n\overline{X})$ soit un estimateur sans biais. Que se passe-t-il si on prend $n = 1, \quad l = 2$?

5 Écart d'une loi de Poisson à sa moyenne

Soit N une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. Prouver que, pour tout u réel, $\exp(uN)$ a une moyenne que l'on calculera. Soit h la fonction $x \to (1+x)\log(1+x) - x$; montrer que h est positive, strictement pour $x \neq 0$. Montrer que, pour $\xi > 0$:

$$\inf_{u>0} E\left[e^{u(N-(1+\xi)\lambda)}\right] = e^{-\lambda h(\xi)} \quad \text{et} \quad P(N \geq (1+\xi)\lambda) \leq e^{-\lambda h(\xi)}.$$

Prouver de même : $P(N \leq (1-\xi)\lambda) \leq e^{-\lambda h(-\xi)}$.

6 Extinction d'une population

Soient $X_1, \ldots, X_n$ des variables entières indépendantes de même fonction génératrice $f(s) = E(s^{X_1})$. Dans ce qui suit, on convient qu'une somme comprenant 0 terme est nulle.

1) Soit N une variable aléatoire de loi $\mathcal{P}(\lambda)$ indépendante de $X_1, \ldots, X_n$. On définit $S = X_1 + \cdots + X_N$. Montrer que S est *infiniment divisble*, c'est-à-dire que, pour tout $m \in \mathbb{N}$, elle peut se représenter comme la somme de m variables Y_i^m i.i.d.

2) On suppose que les X_i sont des variables aléatoires décrivant le nombre de descendants d'une paramécie. Les descendants directs des paramécies de la génération t forment la génération $t+1$ et on note Z^t la taille de la génération t. Écrire Z^{t+1} en fonction des descendants de chaque paramécie de la génération t, $X_1^t, \ldots, X_{Z_t}^t$.

3) Partant d'une paramécie à l'instant 0, $Z^0 = 1$, montrer que la fonction génératrice de Z^t est $f^t(s) = f(f^{t-1}(s))$ $(t > 1)$. Quelle est cette fonction si X_1 suit une loi de Bernoulli?

4) Une caractéristique importante concerne les chances de survie de la population. Soit ρ la probabilité que la population s'éteigne et ρ_t celle qu'elle ait disparu à la génération t. Exprimer ρ_t en fonction de f et déduire de la relation de récurrence entre ρ_t et ρ_{t-1} une équation définissant ρ.

5) Déduire de la convexité de f que le fait que ρ soit égale à 1 dépend de $f'(1)$. Interpréter en terme de $\mathbb{E}(X_1)$.

6) Montrer que, si $\rho < 1$, $f^t(s)$ tend vers ρ pour tout $0 < s < 1$ et interpréter.

7) Considérant Y^t, nombre de descendants de la paramécie jusqu'à la génération t, déduire de
$$Y^t = 1 + Z^1 + \cdots + Z^t$$
que la fonction génératrice g_t de Y^t est telle que $g_t(s) = sf(g_{t-1}(s))$.

7 La ruine du joueur

Les fluctuations d'un jeu sont souvent décrites par une suite (Y_n) de variables indépendantes égales à $+1$ ou à -1 ($+1$ pour un succès, -1 pour un échec). Le nombre de points gagnés après n jeux est donc $S_n = Y_1 + \cdots + Y_n$.
On suppose $P(Y_n = +1) = p$, $P(Y_n = -1) = q$, $(p + q = 1)$.

1) Montrer que $\frac{S_n}{n}$ converge presque sûrement vers $2p - 1$.

2) Calculer pour $n \geq 1$, $E\left[\left(\frac{q}{p}\right)^{S_n}\right]$ et $E(S_n)$.

3) Soit Γ un événement qui ne dépend que de $S_1, \ldots, S_k$ avec $1 \leq k < n$. Montrer que $E\left[\left(\frac{q}{p}\right)^{S_n} \mathbb{I}_\Gamma\right] = E\left[\left(\frac{q}{p}\right)^{S_k} \mathbb{I}_\Gamma\right]$ et $E[(S_n - n(2p-1))\mathbb{I}_\Gamma] = E[(S_k - k(2p-1))\mathbb{I}_\Gamma]$.

4) Soit ν une variable aléatoire, à valeurs dans $\mathbb{N}$, telle que, pour $n \geq 1$, l'événement $\{\nu = n\}$ ne dépend que de $S_1, \ldots, S_n$. Une telle variable aléatoire s'appelle un temps d'arrêt. On suppose de plus qu'il existe un entier M strictement positif tel que $P(\nu \leq M) = 1$. Prouver
$$E\left[\left(\frac{q}{p}\right)^{S_\nu}\right] = 1, \qquad E[S_\nu] = (2p-1)E(\nu). \tag{4}$$

5) On considère, pour $a, b \in \mathbb{N}$, la variable aléatoire $\nu_{a,b} = \inf\{n : \; S_n = -a \text{ ou } S_n = b\}$. On pose ensuite, pour $M \geq 1$, $\nu_{a,b}^M = \min(M, \nu_{a,b})$. Vérifier que $\nu_{a,b}^M$ est un temps

d'arrêt et en déduire que (4) vaut pour $\nu_{a,b}$. Supposons que le joueur possède une fortune initiale égale à a. Calculer la probabilité pour que le joueur s'arrête ruiné, $P(S_{\nu_{a,b}} = -a)$, et, pour $p \neq \frac{1}{2}$, la valeur moyenne de la durée du jeu $E(\nu_{a,b})$.

6) On suppose $p = \frac{1}{2}$, prouver : $E(S^2_{\nu_{a,b}}) = E(\nu_{a,b})E(Y_1^2)$. En déduire $E(\nu_{a,b})$.