IUFM de Toulouse
Préparation au CAPES de Maths
2000-2001

Feuille d'exercices 2

1 Correction de l'exercice sur la loi géométrique

On admet que la probabilité qu'une famille ait n enfants est $p^{n-1}(1-p)$, $(n \geq 1,\ 0 \leq p \leq 1)$. On suppose que toutes les répartitions des sexes de n enfants sont équibrobables; montrer que la probabilité qu'une famille ait exactement k garçons est : $\dfrac{2p^{k-1}(1-p)}{(2-p)^{k+1}}$.

On se place dans le cas de familles possédant au moins un enfant. Pour $k \in \mathbb{N}$ et $n \in \mathbb{N}^*$ on pose

$$\begin{aligned} A_k &= \{\text{il y a } k \text{ garçons dans la famille}\} \\ B_n &= \{\text{il y a } n \text{ enfants dans la famille}\} \\ \alpha_k &= P(A_k). \end{aligned}$$

On définit ensuite pour s réel de module inférieur à 1 la fonction génératrice associée à (α_k) :

$$G(s) = \sum_{k \in \mathbb{N}} \alpha_k s^k.$$

On remarque que, puisque la série des (α_k) est absolument convergente, le rayon de convergence de la série entière définissant G est supérieur à 1. La formule des probabilités totales donne alors :

$$\alpha_k = \sum_{n \geq \max(k,1)} P(A_k | B_n) P(B_n).$$

Par ailleurs, on montre facilement que

$$P(A_k|B_n) = C_n^k \left(\frac{1}{2}\right)^n \quad (k \geq 0, n \geq \max(1,k)),$$

et l'énoncé donne $P(B_n)$. Donc

$$\alpha_k = \frac{1-p}{p} \sum_{n \geq \max(1,k)} C_n^k \left(\frac{p}{2}\right)^n.$$

On a alors pour $|s| < 1$

$$G(s) = \frac{1-p}{p} \sum_{k \in \mathbb{N}} s^k \sum_{n \geq \max(1,k)} C_n^k \left(\frac{p}{2}\right)^n = \frac{1-p}{p} \left[\sum_{k \in \mathbb{N}} s^k \sum_{n \geq k} C_n^k \left(\frac{p}{2}\right)^n - 1 \right].$$

On utilise maintenant le théorème de permutation de somme pour les séries doubles à termes positifs et on obtient

$$G(|s|) = \frac{1-p}{p} \left[\sum_{n \geq 0} \left(\frac{p}{2}\right)^n \sum_{k=0}^{n} C_n^k |s|^k - 1 \right] = \frac{1-p}{p} \left[\sum_{n \geq 0} \left(\frac{p(1+|s|)}{2}\right)^n - 1 \right] = \frac{1-p}{p} \left[\frac{1}{1 - \frac{p(1+|s|)}{2}} - 1 \right].$$

La série double est donc sommable et l'on a

$$G(s)=\frac{1-p}{p}\left[\frac{1}{1-\frac{p(1+s)}{2}}-1\right]=\frac{1-p}{p}\left[\frac{2}{(2-p)(1-\frac{sp}{2-p})}-1\right]=\frac{1-p}{p}\left[\frac{2}{2-p}\sum_{k\in\mathbb{N}}\left(\frac{sp}{2-p}\right)^k-1\right].$$

Par identification des coeficients de la série entière, on trouve pour $k>0$

$$\alpha_k=\frac{2(1-p)p^k}{p(2-p)(2-p)^k}=\frac{2(1-p)p^{k-1}}{(2-p)^{k+1}},$$

et

$$\alpha_0=\frac{1-p}{p}\left(\frac{2}{2-p}-1\right)=\frac{1-p}{2-p}.$$

2 Mon voisin- Simulation numérique

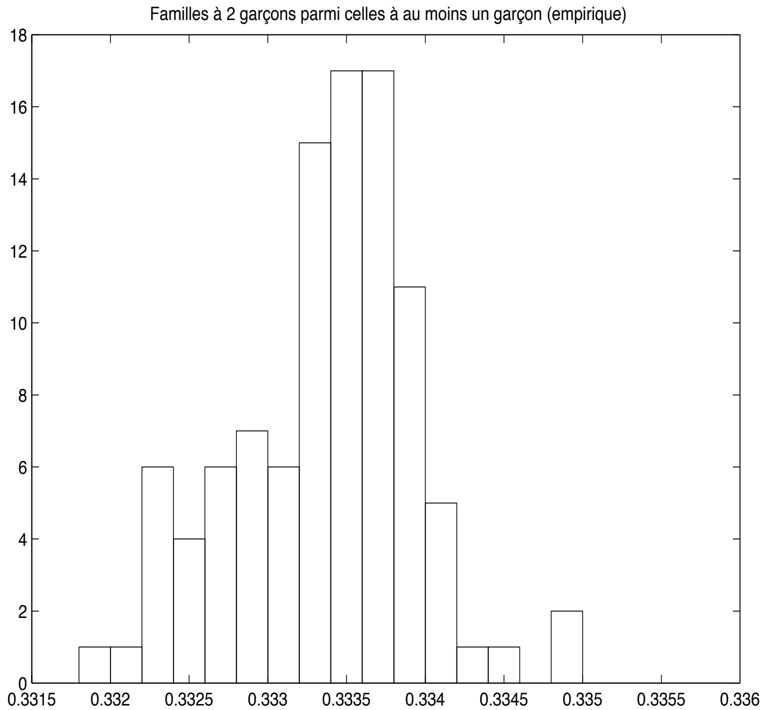
Le programme suivant (écrit en MATLAB) simule 100 fois 1000000 familles de 2 enfants. Parmi les familles possédant au moins un garçon on calcule 100 fois la fréquence observée de familles possédant 2 garçons. Ces résultats sont reportés dans l'histogramme ci-dessous. On retrouve évidemment (aux fluctations près), la valeur théorique 1/3.

Programme

```
%%%% Initialisation
clear;
n=1000000;
nmax=100;
teta=zeros(nmax,1);
%% Boucle
for k=1:nmax
%%% Génération de couples de loi de Bernoulli indépendantes
u=rand(n,2);
x=(u<0.5);
%% Recherche des réalisations possédant au moins un 1
y=sum(x');
y=y';
i=find(y>0);
xx=x(i,:);
nn=size(xx,1);
%% Calcul de la proportion de réalisations de type (1,1) parmi
%% les réalisation possédant au moins un 1
y=sum(xx');
y=y';
i=find(y==2);
teta(k,1)=size(i,1)/nn;
end
histo(teta);
```

```
title('Familles à 2 garçons parmi celles à au moins un garçon (empirique)');
%%
```

Histogramme



3 Variations sur le théorème de Bayes

1) Un examen comporte 15 questions, chacune admettant 3 réponses possibles. Les étudiants répondent à chaque question indépendamment; on suppose que les étudiants ayant préparé l'examen sont en proportion de 70% et répondent correctement à une question avec une probabilité de 0.8, les 30% d'étudiants restant choisissant les réponses au hasard. Il faut au moins 8 bonnes réponses pour réussir l'examen.

   a) Si un étudiant échoue, quelle est la probabilité qu'il ait préparé l'examen?

   b) Soit M le nombre moyen de bonnes réponses pour un étudiant ayant préparé l'examen. Si un étudiant obtient cette note M, quelle est la probabilité qu'il n'ait pas préparé l'examen?

2) Une mouche pond N œufs où N est distribué suivant une loi de Poisson $\mathcal{P}(\lambda)$ $(\lambda > 0)$. Chaque œuf éclot avec probabilité p $(p \in [0,1])$.

   a) À N donné, quelle est la loi de X, nombre d'œufs éclos? En déduire la loi marginale de X et sa fonction génératrice.

   b) Sachant que x œufs ont éclos, quelle est la loi de N?

3) Une population de N individus comporte un nombre inconnu N_A d'individus malades. On suppose que N_A peut prendre indifféremment les valeurs $0, \ldots, N$. On a selectionné n individus dans cette population et obtenu X_A individus malades.

   a) Donner la loi de X_A conditionnellement à N_A.

   b) Actualiser la distribution de probabilité de N_A en fonction de ces observations.

   c) Donner la probabilité que le prochain individu sélectionné soit malade.

4 Lois de Poisson

Soient X et Y deux variables aléatoire indépendantes suivant respectivement les lois de Poisson $\mathcal{P}(\alpha)$ et $\mathcal{P}(\beta)$, où α et β sont deux réels > 0, et soit $Z = X + Y$. Pour tout entier $n \geq 0$, déterminer la loi de probabilité conditionnelle de X sachant que $Z = n$ (c'est-à-dire les probabilités conditionnelles $P(X = k \mid Z = n)$ pour $k \geq 0$).

5 Bernoulli

Soient q et r deux réels strictement compris entre 0 et 1, et soit $(X_n)_{n\geq 1}$ une suite de variables de Bernoulli. On suppose que pour tout $n \geq 1$

$$P[X_{n+1} = 1 \mid X_n = 0] = q, \qquad P[X_{n+1} = 0 \mid X_n = 1] = r.$$

On note $p_n = P[X_n = 1]$ $(n \geq 1)$.

1) Écrire une équation de récurrence qui exprime p_{n+1} en fonction de p_n.

2) Montrer qu'il existe un $p \in [0, 1]$ tel que cette équation de récurrence admette la solution constante $p_n = p$ pout tout n, et montrer que dans le cas général p_n tend vers p lorsque n tend vers l'infini.

6 Décomposition dyadique sur $[0, 1]$

Soit l'espace de probabilité $([0, 1], \mathcal{B}([0, 1]), P)$ où $\mathcal{B}([0, 1])$ est la tribu borélienne de $[0, 1]$ (la plus petite tribu contenant tous les ouverts de $[0, 1]$) et P est la probabilité uniforme ($P([a, b]) = b - a$ pour tout intervalle $[a, b]$ de $[0, 1]$). Sur cet espace on définit les variables aléatoires dyadiques, d_n $(n \geq 1)$: pour $\omega \in [0, 1]$, $d_n(\omega)$ est le n-ième chiffre dans la décomposition binaire de ω. Pour les décompositions multiples, comme $1 = 1.0 = 0.1111...$ ou $1/2 = 0.1 = 0.0111..$, on choisit la décomposition infinie ($d_n(1) = 1$ pour tout n, $d_n(1/2) = 1$ pour tout $n \geq 2$).

1) Montrer que

$$P(\{\omega : \ d_j(\omega) = u_j\}) = \frac{1}{2} \text{ et } P(\{\omega : \ d_i(\omega) = u_i, \ i = 1, \ldots, n\}) = \frac{1}{2^n}$$

quelques soient $n, j \geq 1$ et $u_j, u_i \in \{0,1\}$. En déduire que

$$P\left(\left\{\omega : \sum_{i=1}^{n} d_i(\omega) = k\right\}\right) = C_n^k \frac{1}{2^n}, \quad 0 \leq k \leq n.$$

2) Soient, pour $n \geq 1$, les fonctions l_n définies sur $[0,1]$ par

$$l_n(\omega) = k \text{ si } d_n(\omega) = \cdots = d_{n+k-1}(\omega) \neq d_{n+k}(\omega) \ \ (k \in \mathbb{N}^*).$$

Montrer que

$$P(\{\omega : \ l_n(\omega) = k\}) = \frac{1}{2^k} \text{ et } P(\{\omega : \ l_n(\omega) \geq r\}) = \frac{1}{2^{r-1}} \ \ (k, r \in \mathbb{N}^*).$$

3) Soit une suite croissante de réels r_n. Montrer que, si $\sum_n 2^{-r_n}$ converge et si

$$A_n = \{\omega : \ l_n(\omega) \geq r_n\},$$

alors

$$P(\limsup_{n\to\infty} A_n) = 0.$$

Considérant le cas particulier $r_n = (1+\varepsilon)\log_2 n$, en déduire que

$$P\left(\left\{\omega : \ \limsup_{n\to\infty} \frac{l_n(\omega)}{\log_2 n} \leq 1\right\}\right) = 1.$$

4) On dira que si $d_n = 1$ on obtient un succès à l'instant $n \geq 2$ et sinon un échec.

   a) Montrer que l'événement *"on obtient qu'un nombre fini de succés"* a une probabilité nulle.

   b) Pour $n \geq 1$, soit T_n l'instant du n-ième succès, c'est-à-dire la variable aléatoire entière égale au plus petit indice k tel qu'on ait n succès parmi $d_1, \ldots, d_k$. Déterminer la loi de T_n $(n \geq 1)$.

   c) Montrer que les variables T_1 et $T_{n+1} - T_n$ pour $n \geq 1$ sont indépendantes et de même loi.

5) On dit qu'on obtient 01 à l'instant $n \geq 2$ si $d_{n-1} = 0$ et $d_n = 1$.

   a) Montrer que la probabilité de ne jamais obtenir 01 jusqu'à l'instant n compris est égale à $\dfrac{n+1}{2^n}$.

   b) En déduire que l'événement *"on obtient au moins une fois 01"* est de probabilité 1.

   c) Soit T la variable aléatoire entière égale au premier instant où l'on obtient 01. Déterminer la loi de probabilité de T.

7 Entropie

Soit Ω un ensemble dénombrable et P une probabilité sur Ω. Son **entropie** est définie par

$$H(P) = -\sum_{\omega \in \Omega} P(\omega) \log P(\omega).$$

On ne précise pas la base du logarithme et on convient que $u \log u$ vaut 0 pour $u = 0$ et $\frac{0}{0}$ vaut 0.

a) Montrer que $H(P)$ est positive; quand s'annule-t-elle?

b) Soit Q une autre probabilité. Prouver que $-\sum_{\omega \in \Omega} P(\omega) \log \frac{Q(\omega)}{P(\omega)}$ est positif, nul si et seulement si P et Q coïncident. Ce nombre est **l'information de Kullback** de P sur Q.

c) Soit $\Omega = \{1, 2, \ldots, n\}$ et Q la probabilité uniforme sur Ω. Calculer $H(Q)$. Prouver que, pour toute P différente de Q, $H(P)$ est strictement majorée par $H(Q)$.

d) Soit X une fonction définie sur Ω; on désigne par $H(X)$ l'entropie de sa loi. Prouver que $H(X)$ est majorée par $H(P)$.

e) Soit X et Y deux fonctions définies sur Ω; $P(Y = \cdot | X = x)$ est la loi de Y conditionnelle à $\{X = x\}$, pour $x \in \Omega' = \{\omega \in \Omega : \; P(\omega) > 0\}$. Prouver :

$$H(X, Y) = H(X) + \sum_{x \in X(\Omega')} P(X = x) H\left(P(Y = \cdot | X = x)\right).$$

L'expression $H(X, Y) - H(X)$ est donc une moyenne des entropies des lois de Y conditionnelles aux diverses valeurs de X. On la note $H(Y|X)$, et on l'appelle *entropie de Y conditionelle à X*. Montrer que $H(Y|X)$ s'annule si et seulement si Y est une fonction de X.

f) Un joueur cherche à déterminer le plus vite possible un nombre Y, que son adversaire a choisi au hasard compris entre 0 et $N-1$. Pour cela il peut poser les questions qu'il veut ("pair"? "égal à 47"? "supérieur à 50"?...) à condition que les réponses soient oui ou non : au bout de n questions il connaît la valeur d'une fonction X à valeur dans $\{0, 1\}^n$. Montrer que, même s'il est très astucieux, il ne peut être sur de déterminer Y avant d'avoir posé un nombre de questions supérieur à $\frac{\log N}{\log 2}$ (Comparer l'entropie de Y et l'entropie maximum de X).

g) Dans un tas de N pièces, toutes sont identiques sauf une fausse plus légère. Pour la déceler on a une balance à deux plateaux et on peut mettre le nombre de pièces que l'on veut sur chaque plateau : à chaque expérience correspondent 3 réponses possibles, "le plateau de gauche est plus lourd", "le plateau de droite est plus lourd", ou "il sont égaux". Prouver que la meilleure tactique ne nous assurera du succès qu'avec un nombre de pesées supérieur à $\frac{\log N}{\log 3}$.

8 Loi multinomiale

a) Soit X une v.a. de loi concentrée sur $\{1, \dots, k\}$;

$$P(X = j) = p_j, 1 \leq j \leq k.$$

Soit $Z^j = \mathbf{1}_{(X=j)}$ et $s = (s_1, \dots, s_k) \in \mathbb{R}^k$. Calculer:

$$E\left[s_1^{Z^1} s_2^{Z^2} \cdots s_k^{Z^k}\right].$$

b) Soit $X_1, \dots, X_n$ un n-échantillon de la loi précédente (n variables indépendantes de même loi), et

$$N^j = \sum_{i=1}^{n} \mathbf{1}_{(X_i=j)}.$$

Calculer:

$$E\left[s_1^{N^1} s_2^{N^2} \cdots s_k^{N^k}\right].$$

En déduire, pour $a_1, \dots, a_k$ entiers de somme n:

$$P(N^1 = a_1, \dots, N^k = a_k) = \frac{n!}{a_1! \dots a_k!} p_1^{a_1} \dots p_k^{a_k}.$$

C'est la loi multinomiale de paramètre $(p_1, \dots, p_k)$ et d'ordre n.

9 Loi de Poisson

Un auto stoppeur attend au péage de l'autoroute A6 à Avallon. Le nombre de véhicules passant par ce péage durant une heure est une variable aléatoire X. Pour chacun de ces véhicules il y a une probabilité $p \in]0, 1[$ qu'il vienne de la direction de Paris et donc $q = 1 - p$ qu'il vienne de la direction Lyon. On note Y et Z le nombre de véhicules venant de Paris (resp. Lyon), donc $Y + Z = X$.

a) On suppose que X suit une loi de Poisson de paramètre λ. Déterminer les lois de Y et Z et montrer que Y et Z sont indépendantes.

b) On suppose que Y et Z sont indépendantes. Quelle est la loi de X?