Magistère
Probabilités
2008-2009

# Feuille d'exercices 2-Modèles discrets

## 1 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 $0, \dots, N$. 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.

## 2 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 $\alpha$ et $\beta$ 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 $k \geq 0$).

## 3 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.

# 4 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 $\omega$. 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) Montrer que, si $B_n = \{\omega :\ l_n(\omega) = 0\}$, $P(\limsup_{n\to\infty} B_n) = 0$.

5) Considérant une suite croissante $(r_n)$ de réels positifs telle que la série $\sum_n 2^{-r_n} r_n^{-1}$ diverge, montrer que $P(\limsup_{n\to\infty} A_n) = 1$. En prenant $r_n = \log_2 n$, déduire que

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

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 $T$.

# 5 Loi multinomiale

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

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

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

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

b) Soit $X_1, \ldots, X_n$ 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 $a_1, \ldots, a_k$ entiers de somme $n$ :

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

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

# 6 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 $\lambda$. 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 $X$ ?

# 7 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 $N_{\text{min}}$ 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 ?
   – Si on dispose de 40 échantillons, combien en moyenne d'intervalles de confiance (au risque 5%) comprendront la valeur exacte inconnue $p$ ?
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 $P(N_n^* \geq s)$.

# 8 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 $\rho$ la probabilité que la population s'éteigne et $\rho_t$ celle qu'elle ait disparu à la génération $t$. Exprimer $\rho_t$ en fonction de $f$ et déduire de la relation de récurrence entre $\rho_t$ et $\rho_{t-1}$ une équation définissant $\rho$.

5) Déduire de la convexité de $f$ que le fait que $\rho$ 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 $\rho$ 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 $g_t(s) = sf(g_{t-1}(s))$.

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

$$E\left[\left(\frac{q}{p}\right)^{S_\nu}\right] = 1, \qquad E[S_\nu] = (2p-1)E(\nu). \tag{1}$$

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 (1) 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 $E(\nu_{a,b})$.