Introduction aux probabilités
Examen du 6 janvier 2011. Durée : 3 heures.
La consultation des notes de cours est autorisée.

On suppose que toutes les variables aléatoires qui apparaissent sont définies sur un espace de probabilité discret.

Exercice 1

Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d. à valeurs strictement positives. Pour tout $1 \leqslant m \leqslant n$, on pose $S_m = X_1 + X_2 + \cdots + X_m$. Montrer que pour tout $1 \leqslant m \leqslant n$, on a

$$\mathbf{E}\left(\frac{S_m}{S_n}\right) = \frac{m}{n}.$$

Exercice 2

On considère une urne qui contient à l'instant 0 une boule rouge et une boule verte. A l'étape n, on tire au sort une boule de l'urne, on la remet dedans et on ajoute à l'urne une boule de la même couleur que la boule tirée. Ainsi, à l'étape n, l'urne contient $n+2$ boules. On suppose que les tirages sont effectués de manière indépendante et selon la loi uniforme.

1. Soit X_n le nombre de boules rouges dans l'urne à l'étape n. Montrer que X_n suit la loi uniforme sur $\{1, \dots, n+1\}$.
2. On pose $Y_n = 1$ si la boule tirée à l'étape n est rouge, et $Y_n = 0$ si la boule tirée à l'étape n est verte. Est-ce que les variables aléatoires $(Y_n)_{n \geqslant 1}$ sont indépendantes ? Est-ce qu'elles ont même loi ?
3. Montrer que pour toute permutation $\sigma \in \mathfrak{S}_n$, les vecteurs aléatoires $(Y_1, \dots, Y_n)$ et $(Y_{\sigma(1)}, \dots, Y_{\sigma(n)})$ ont même loi.

Exercice 3

Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires indépendantes de loi uniforme sur $\{-1, 1\}$. On pose $S_n = X_1 + X_2 + \cdots + X_n$.

Dans les questions 1 à 4, on ne considère qu'une suite finie $(X_1, \dots, X_n)$ définie sur un espace de probabilité disrcet. Dans la question 5, on considère une suite infinie, définie sur l'espace $\{-1,1\}^{\mathbf{N}}$, et on affecte des probabilités à certains événements de la manière indiquée dans le cours.

1. Montrer que pour tous réels positifs a, x, on a
$$\mathbf{P}(S_n \geqslant a) \leqslant \exp(-ax)\mathbf{E}\exp(xS_n).$$
2. Donner une expression plus simple de $\mathbf{E}\exp(xS_n)$.
3. Montrer que l'inégalité $\operatorname{ch} x \leqslant \exp(x^2/2)$ est vraie pour tout nombre réel x (vous pouvez l'admettre, c'est une question d'analyse qui aura un poids très faible dans le barème).
4. En déduire que pour tout réel positif a, on a
$$\mathbf{P}(S_n \geqslant a) \leqslant \exp\left(-\frac{a^2}{2n}\right).$$
5. Montrer que pour $c > \sqrt{2}$, on a
$$\mathbf{P}\left(\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{n \log n}} > c\right) = 0.$$
Indication : *On pourra inclure cet événement dans la réunion d'une suite d'événements d'envergure finie dont la somme des probabilités est arbitrairement petite.*

Exercice 4

Soit X une variable aléatoire. On dit qu'un nombre réel μ est une médiane de X si on a $\mathbf{P}(X \geqslant \mu) \geqslant 1/2$ et $\mathbf{P}(X \leqslant \mu) \geqslant 1/2$.

1. Montrer que toute variable aléatoire admet une médiane. Est-elle unique ?
2. Soit X une variable aléatoire telle que $\mathbf{E}X^2 < +\infty$. On note $E = \mathbf{E}X$ l'espérance de X, μ une médiane de X et $\sigma = \sqrt{\mathbf{Var}\, X}$ l'écart-type de X. Montrer l'inégalité
$$|E - \mu| \leqslant \sigma.$$

Exercice 5

Soient X et Y des variables aléatoires positives. On dit que Y majore stochastiquement X (on écrit $X \leqslant_{st} Y$) si, pour tout réel t, on a
$$\mathbf{P}(X \geqslant t) \leqslant \mathbf{P}(Y \geqslant t).$$

1. Montrer que $X \leqslant_{st} Y$ si et seulement si, pour toute fonction croissante $h : \mathbf{R}^+ \to \mathbf{R}^+$, on a
$$\mathbf{E}h(X) \leqslant \mathbf{E}h(Y).$$
2. Soient X et Y des variables aléatoires indépendantes vérifiant $X \leqslant_{st} Y$. Montrer que $\mathbf{P}(X \leqslant Y) \geqslant 1/2$.
3. Soient X et Y deux variables aléatoires de loi de Poisson, de paramètres respectifs λ et μ. On suppose que $\lambda \leqslant \mu$. Montrer que $X \leqslant_{st} Y$.

Exercice 6

1. Soient $1 \leqslant p \leqslant q$ des nombres entiers, et A un sous-ensemble aléatoire de $\{1, \dots, q\}$ de cardinal p choisi selon la loi uniforme parmi les $\binom{q}{p}$ sous-ensembles possibles. On note $A = \{(x_i)_{1 \leqslant i \leqslant p}\}$ avec $x_1 < x_2 < \cdots < x_p$. Montrer que les variables aléatoires
$$x_1, x_2 - x_1, x_3 - x_2, \dots, x_p - x_{p-1}, q + 1 - x_p$$
suivent toutes la même loi.
2. Les 2^n premiers joueurs du classement mondial participent à un tournoi de tennis. Chaque jour, tous les joueurs non encore éliminés jouent un match contre un adversaire ; le perdant du match est éliminé du tournoi. On suppose que dans chaque match, le joueur le mieux classé l'emporte.

   Le $n$ème jour, le joueur classé premier remporte le tournoi. Soit X le classement du finaliste, c'est-à-dire du dernier joueur à avoir été eliminé.

   On suppose que, chaque jour, les matchs sont tirés au sort selon la loi uniforme sur l'ensemble des matchs possibles, de manière indépendante. Calculer l'espérance de X, et donner sa limite lorsque n tend vers $+\infty$.

CORRIGÉ

Exercice 1

Pour toute permutation σ, le vecteur aléatoire $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ a même loi que $(X_1, \dots, X_n)$, et on en déduit que la quantité $\mathbf{E}(X_i/S_n)$ de dépend pas de i. Puisque ces n quantités ont pour somme 1, chacune vaut $1/n$. Enfin, $\mathbf{E}(S_m/S_n) = m/n$ par linéarité de l'espérance.

Exercice 2

1. On procède par récurrence grâce à la formule

$$\mathbf{P}(X_n = k) = \mathbf{P}(X_n = k|X_{n-1} = k)\mathbf{P}(X_{n-1} = k) + \mathbf{P}(X_n = k) = \mathbf{P}(X_n = k|X_{n-1} = k-1)\mathbf{P}(X_{n-1} = k-1)$$

2. On vérifie facilement que $\mathbf{P}(Y_1 = 1, Y_2 = 1) \neq \mathbf{P}(Y_1 = 1)\mathbf{P}(Y_2 = 1)$ donc ces variables aléatoires ne sont pas indépendantes. Elles ont toute même loi, qui est une loi de Bernoulli de paramètre $1/2$. Pour le vérifier, on peut calculer $\mathbf{P}(Y_n = 1) = \sum_k \mathbf{P}(Y_n = 1|X_{n-1} = k)\mathbf{P}(X_{n-1} = k)$ grâce à la question 1, ou invoquer un argument de syémtrie : puisque X_n suit la loi uniforme sur $\{1, \dots, n+1\}$, elle a même loi que $n + 2 - X_n$ qui est le nombre de boules vertes dans l'urne, donc $\mathbf{P}(Y_n = 1) = \mathbf{P}(Y_n = 0)$.

3. Il suffit de le vérifier lorsque σ est la transposition qui échange i et $i+1$, puisque ces permutations engendrent le groupe $\mathfrak{S}_n$. On est ramené à vérifier que pour toute suite $(y_k) \in \{0,1\}^n$, on a

$$\mathbf{P}(Y_1 = y_1, \dots, Y_i = y_i, Y_{i+1} = y_{i+1}, \dots, Y_n = y_n) = \mathbf{P}(Y_1 = y_1, \dots, Y_i = y_{i+1}, Y_{i+1} = y_i, \dots, Y_n = y_n)$$

En écrivant

$$\mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n) = \mathbf{P}(Y_1 = y_1) \prod_{k=2}^{n} \mathbf{P}(Y_k = y_k | Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}),$$

et similairement pour l'autre membre, on est ramené à vérifier que

$$\mathbf{P}(Y_i = y_i | Y_1 = y_1, \dots, Y_{i-1} = y_{i-1})\mathbf{P}(Y_{i+1} = y_{i+1} | Y_1 = y_1, \dots, Y_i = y_i) =$$

$$\mathbf{P}(Y_i = y_{i+1} | Y_1 = y_1, \dots, Y_{i-1} = y_{i-1})\mathbf{P}(Y_{i+1} = y_i | Y_1 = y_1, \dots, Y_{i-1} = y_{i-1}, Y_i = y_{i+1})$$

Cette égalité se vérifie facilement : $S = X_{i-1}$ est déterminé par $(Y_1, \dots, Y_{i-1})$ et

$$\frac{S}{i+1} \frac{i+2-(S+1)}{i+2} = \frac{i+1-S}{i+1} \frac{S}{i+2}.$$

Exercice 3

1. $\mathbf{P}(S_n \geqslant a) = \mathbf{P}(\exp(xS_n) \geqslant \exp(xa)) \leqslant \exp(-xa)\mathbf{E}\exp(xS_n)$ par l'inégalité de Markov.

2. Puisque (X_n) sont i.i.d., on a $\mathbf{E}\exp(xS_n) = (\mathbf{E}\exp(xX_1))^n = \cosh(x)^n$.

3. Développer les deux membres en série entière et comparer les coefficients.

4. On a $\mathbf{P}(S_n \geqslant a) \leqslant \exp(-xa + nx^2/2)$ et le choix opptimal $x = a/n$ donne le résultat.

5. D'abord, S_n a la même loi que $-S_n$, et donc $\mathbf{P}(|S_n| \geqslant a) = \mathbf{P}(S_n \geqslant a) + \mathbf{P}(-S_n \geqslant a) \leqslant 2\exp(-a^2/2n)$. Soit A_n l'événement d'envergure finie "$|S_n| \geqslant c\sqrt{n \log n}$". On a $\mathbf{P}(A_n) \leqslant 2\exp(-c^2 n \log n/2n) = n^{-c^2/2}$. Pour tout N, l'événement considéré est inclus dans $\bigcup_{n \geqslant N} \mathbf{P}(A_n)$, et comme la série $\sum n^{-c^2/2}$ converge, la somme des probabilités de ces événements peut être rendue arbitrairement petite en choisissant N suffisemment grand.

Exercice 4

1. La médiane n'est pas unique (prendre une v.a. de loi de Bernoulli). Pour montrer son existence, on introduit par exemple

$$\mu = \inf\{t \text{ t.q. } \mathbf{P}(X \geqslant t) \geqslant 1/2\}.$$

La nombre μ ainsi défini est un réel (l'infimum est pris sur une partie non vide minorée de $\mathbf{R}$). La continuité à droite de la fonction de répartition implique que l'infimum est atteint et que $\mathbf{P}(X \geqslant \mu) \geqslant 1/2$. De plus pour tout $\varepsilon > 0$, on a $\mathbf{P}(X \leqslant t - \varepsilon) < 1/2$, donc $\mathbf{P}(X \geqslant t - \varepsilon) \geqslant 1/2$, et par continuité $\mathbf{P}(X \geqslant t) \geqslant 1/2$.

2. On peut (quitte a remplacer X par $(X-m)/\sigma$) supposer que $m=0$ et $\sigma=1$. Il faut alors montrer que $|\mu| \leqslant 1$, c'est-à-dire que $\mathbf{P}(X \geqslant -1) \geqslant 1/2$ et $\mathbf{P}(X \leqslant 1) \geqslant 1/2$. Par symétrie il suffit de montrer la deuxième inégalité. On écrit pour $t>0$

$$\mathbf{P}(X>1) = \mathbf{P}(X+t>1+t) \leqslant \mathbf{P}((X+t)^2 \geqslant (1+t)^2) \leqslant \frac{1}{(1+t)^2}\mathbf{E}(X+t)^2 = \frac{1+t^2}{(1+t)^2},$$

par l'inégalité de Markov ; le choix optimal $t=1$ donne le résultat.

Exercice 5

1. On commence remarquer que si $X \leqslant_{st} Y$ et $h: \mathbf{R}^+ \to \mathbf{R}^+$ est une fonction croissante, alors $h(X) \leqslant_{st} h(Y)$. Ensuite on a
$$\mathbf{E}h(X) = \int_0^\infty \mathbf{P}(h(X) \geqslant t)dt \leqslant \int_0^\infty \mathbf{P}(h(X) \geqslant t)dt = \mathbf{E}h(Y).$$
Réciproquement, on prend pour h la fonction indicatrice de $[t,+\infty[$.

2. On remarque que si X' est une copie indépendante de X, alors les événements $X \geqslant X'$ et $X' \geqslant X$ ont même probabilité (par symétrie), et leur réunion est tout l'espace de probabilités, donc chacun a probabilité au moins $1/2$. Ensuite,
$$\mathbf{P}(X \leqslant Y) = \sum_{x\in\mathbf{R}} \mathbf{P}(X=x)\mathbf{P}(Y \geqslant x) \geqslant \sum_{x\in\mathbf{R}} \mathbf{P}(X=x)\mathbf{P}(X' \geqslant x) = \mathbf{P}(X \leqslant X') \geqslant 1/2.$$

3. Soit Z une loi de Poisson de paramètre $\mu-\lambda$, indépendante de X. Alors $X+Z$ a même loi que Y (résultat facile à montrer avec les fonctions génératrices). Comme Z est positive,
$$\mathbf{P}(X \geqslant t) \leqslant \mathbf{P}(X+Z \geqslant t) = \mathbf{P}(Y \geqslant t).$$

Exercice 6

1. On peut invoquer un argument de symétrie, ou bien calculer directement les lois de ces variables aléatoires: pour tout $k \in \{1,\dots,q\}$, on dénombre (dans la deuxième ligne, j est la valeur de x_i)
$$\binom{q}{p}\mathbf{P}(x_1=k) = \binom{q-k}{p-1},$$
$$\binom{q}{p}\mathbf{P}(x_{i+1}-x_i=k) = \sum_j \binom{j-1}{i-1}\binom{q-(j+k)}{p-(i+1)} = \binom{q-k}{p-1}$$
$$\binom{q}{p}\mathbf{P}(x_p=k) = \binom{k-1}{p-1}.$$
Une remarque utile pour la quite : comme la somme des variables aléatoires $x_1, x_2-x_1, x_3-x_2, \dots, x_p-x_{p-1}, q+1-x_p$ est constante égale à $q+1$, une conséquence du fait qu'elles ont même loi est que chacune a pour espérance $\frac{q+1}{p+1}$.

2. L'ensemble des joueurs qui ne sont pas dans la moitié de tableau du joueur classé numéro 1 est un sous-ensemble I à 2^{n-1} éléments de $\{2,\dots,2^n\}$. On vérifie que I est distribué selon la loi uniforme, et que X le classement du finaliste est le minimum de I. Ainsi $X-1$ la loi du minimum d'un sous-ensemble aléatoire uniforme de 2^{n-1} éléments parmi $\{1,\dots,2^n-1\}$. Par la question précédente ($p=2^{n-1}, q=2^n-1$), on a
$$\mathbf{E}X = 1+\frac{2^n}{2^{n-1}+1},$$
qui tend vers 3 quand n tend vers $+\infty$.