Examen final du 10 juin 2015 (2ème session)

Durée : 2 heures. Tous documents interdits. Dans tout l'examen on travaille sur un espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$. La qualité et la rigueur de la rédaction seront prises en compte.

Exercice 1. Soit $\alpha \in]0,1[$ et (X,Y) un couple de variables aléatoires entières dont la loi est donnée par

$$\mathbb{P}(X=i, Y=j) = \alpha^{i+j}(1-\alpha)^2 \qquad i,j \geq 0.$$

a) Donner la loi de X et la loi de Y.

b) Les v.a. X et Y sont-elles indépendantes ?

c) Déterminer la loi de $X+Y$.

d) Pour $n \geq 0$, déterminer la loi de X conditionnelle à $\{X+Y=n\}$.

Solution de l'exercice 1.

a) Le problème est symétrique en X et en Y. Par exemple pour X : pour tout $i \geq 0$,

$$\mathbb{P}(X=i) = \sum_{j\geq 0} \mathbb{P}(X=i, Y=j) = (1-\alpha)^2\alpha^i \sum_{j\geq 0} \alpha^j = (1-\alpha)\alpha^i.$$

Donc X et Y sont toutes deux géométriques de paramètre α.

b) Les v.a. X et Y sont indépendantes car pour tous $i,j \geq 0$

$$\mathbb{P}(X=i, Y=j) = (1-\alpha)^2\alpha^{i+j} = (1-\alpha)\alpha^i(1-\alpha)\alpha^j = \mathbb{P}(X=i)\mathbb{P}(Y=j).$$

c) Pour tout $n \geq 0$,

$$\mathbb{P}(X+Y=n) = \sum_{i=0}^{n} \mathbb{P}(X=i, Y=n-i) = \sum_{i=0}^{n} (1-\alpha)^2\alpha^n = (n+1)(1-\alpha)^2\alpha^n.$$

d) Pour tous $0 \leq i \leq n$,

$$\mathbb{P}(X=i|X+Y=n) = \frac{\mathbb{P}(X=i, X+Y=n)}{\mathbb{P}(X+Y=n)} = \frac{\mathbb{P}(X=i, Y=n-i)}{\mathbb{P}(X+Y=n)} = \frac{(1-\alpha)^2\alpha^n}{(n+1)(1-\alpha)^2\alpha^n} = \frac{1}{n+1}.$$

En conclusion, la loi conditionnelle de X sachant $X+Y=n$ est la loi uniforme sur $\{0,1,\ldots,n\}$.

Exercice 2. Considérons un tiroir contenant n paires de chaussettes assorties (soit $2n$ chaussettes) bien mélangées, dont les n couleurs sont supposées distinctes 2 à 2 et notées symboliquement couleur 1, couleur 2,..., couleur n.
On sort les chaussettes une par une de ce tiroir, uniformément au hasard (et sans remise) à chaque tirage successif.
Soit la variable aléatoire $X_k \in \{1, \ldots, n\}$ définie comme la couleur de la chaussette sortie du tiroir au $(k+1)-$ième tirage. Pour tout entier $1 \leq k \leq 2n-1$, on définit l'événement A_k :

$$A_k := \bigcap_{1 \leq i < j \leq k+1} \{X_i \neq X_j\},$$

qui se réalise lorsque les $k+1$ premières chaussettes tirées sont toutes dépareillées.
L'objectif de ce problème est d'étudier la variable aléatoire T_n égale au nombre minimal de tirages nécessaires pour avoir deux chaussettes de la même couleur, c'est-à-dire

$$T_n := \min\{j \geq 2 : \exists i < j, X_i = X_j\}.$$

a) Question préliminaire. Soit $a > 0$ et $F : \mathbb{R}_+ \longrightarrow [0,1]$ définie par $F_a(x) = 1 - \exp(-x^2/a)$. Soit R_a une variable aléatoire de fonction de répartition F_a.
   i) Déterminer la densité f_a de R_a.
   ii) En utilisant vos connaissances sur la variance des variables normales, déterminer $\mathbb{E}(R_a)$.

b) Dans cette question, on cherche à caractériser la loi de T_n.
   i) Que vaut $\mathbb{P}(A_k)$ pour $k > n$?
   ii) Déterminer $\mathbb{P}(A_1)$.
   iii) Sans utiliser la loi de Bayes, montrer que $\mathbb{P}(A_k|A_{k-1}) = u_k$, où l'on a noté

$$u_k := \frac{2(n-k)}{2n-k}$$

   iv) Montrer que la suite (A_k) est décroissante et en déduire

$$\mathbb{P}(A_k) = \prod_{j=1}^{k} u_j.$$

   v) Montrer que $\{T_n > k\} = A_{k-1}$.
   vi) Montrer que la suite (T_n) converge en probabilité vers $+\infty$ lorsque $n \to \infty$.

c) Dans cette dernière question on cherche à connaître l'ordre de grandeur de T_n lorsque $n \to \infty$. Soit $f : [0, 1/2] \longrightarrow [0,1]$ définie par

$$f(x) = \frac{1-2x}{1-x}$$

   i) Montrer que $u_k = f(k/(2n))$.
   ii) Soit $g : [0, 1/2[\longrightarrow \mathbb{R}$ définie par $g(x) = x + \ln(f(x))$. Montrer que g est négative, décroissante et vérifie $\lim_{x \to 0+} g(x)/x = 0$.

iii) Montrer que

$$\ln(\mathbb{P}(A_k)) = -\sum_{j=1}^{k} \frac{j}{2n} + \sum_{j=1}^{k} g\left(\frac{j}{2n}\right).$$

iv) En déduire

$$\left|\ln(\mathbb{P}(A_k)) + \frac{k(k+1)}{4n}\right| \leq k\,|g(k/2n)|\,.$$

v) Soit $x > 0$. Montrer que

$$\lim_{n\to\infty} \ln(\mathbb{P}(T_n > x\sqrt{n})) = -\frac{x^2}{4}.$$

vi) Montrer que $T_n/\sqrt{n}$ converge en loi vers une v.a. R dont on précisera la loi et l'espérance.

vii) Si l'on tire au hasard 20 chaussettes dans un tiroir contenant 100 paires de chaussettes de 100 couleurs différentes, que vaut approximativement la probabilité d'avoir tiré au moins une paire ?

Solution de l'exercice 2.

a) i) La densité f_a est la dérivée de la fonction de répartition, soit $f_a(x) = (2x/a)e^{-x^2/a}$. Cette loi est appelée loi de Rayleigh.

ii) En posant $a = 2\sigma^2$, l'espérance de R_a est donnée par

$$\mathbb{E}(R_a) = 2a^{-1}\int_0^\infty x^2 e^{-x^2/a}\,dx = (2\sigma^2)^{-1}\int_{-\infty}^{\infty} x^2 e^{-x^2/(2\sigma^2)}\,dx = (2\sigma^2)^{-1}\sigma\sqrt{2\pi}\,\mathbb{E}(N^2),$$

où N est une v.a. normale centrée de variance σ^2, donc $\mathbb{E}(N^2) = \sigma^2$ et $\mathbb{E}(R_a) = \sigma\sqrt{2\pi}/2 = \sqrt{a\pi}/2$.

b) L'événement A_k est réalisé ssi les $k+1$ premières chaussettes tirées sont toutes de couleurs distinctes 2 à 2.

i) $\mathbb{P}(A_k) = 0$ dès que $k > n$, car si les $k+1$ chaussettes tirées étaient de couleurs distinctes 2 à 2, alors il y aurait plus de couleurs dans l'échantillon que dans le tiroir d'origine !

ii) A_1 n'est pas réalisé ssi au deuxième tirage on tire la seule chaussette qui est de la même couleur que la première, parmi les $2n-1$ chaussettes restantes, autrement dit $\mathbb{P}(A_1) = 1 - 1/(2n-1) = (2n-2)/(2n-1)$.

iii) En faisant le même raisonnement conditionnellement à A_{k-1}, c'est-à-dire conditionnellement à avoir tiré k chaussettes dépareillées en k tirages, A_k n'est pas réalisé ssi au $(k+1)-$ème tirage, on tire une des k chaussettes qui font la paire avec une des k chaussettes dépareillées déjà tirées, parmi les $2n-k$ chaussettes restantes, autrement dit $\mathbb{P}(A_k|A_{k-1}) = 1 - k/(2n-k) = u_k$.

iv) Si les k premières chaussettes sont dépareillées alors les $k-1$ premières le sont aussi, autrement dit $A_k \subset A_{k-1}$. Il en résulte que

$$\mathbb{P}(A_k) = \mathbb{P}(A_k \cap A_{k-1}) = \mathbb{P}(A_k|A_{k-1})\mathbb{P}(A_{k-1}) = u_k\mathbb{P}(A_{k-1}),$$

d'où le résultat, par une récurrence immédiate.

v) évident.

vi) En examinant l'expression de u_j, pour j fixé, on voit que la valeur qu'elle prend tend vers 1 lorsque $n \to \infty$, aussi pour tout k fixé, $\mathbb{P}(T_n > k) = \mathbb{P}(A_{k-1}) \to 1$ lorsque $n \to \infty$, ce qui ne signifie rien d'autre que $T_n \to +\infty$ en probabilité.

c) i) $u_k = 2(n-k)/(2n-k) = (1-(k/n))/(1-(k/(2n)) = f(k/2n)$.

ii) Il s'agit de calculs élémentaires :

$$g(x) = x + \ln(1-2x) - \ln(1-x),$$

donc

$$g'(x) = 1 - \frac{2}{1-2x} + \frac{1}{1-x} = \frac{1-3x+2x^2-2(1-x)+1-2x}{(1-2x)(1-x)} = \frac{x(2x-3)}{(1-2x)(1-x)},$$

qui ne prend que des valeurs négatives sur $]0, 1/2[$. De plus, $g(0) = g'(0) = 0$, ce qui donne la conclusion.

iii) Il suffit de passer au logarithme l'expression de $\mathbb{P}(A_k)$.

iv) Comme $\sum_{j=1}^{k} j = k(k+1)/2$, on obtient

$$\left|\ln(\mathbb{P}(A_k)) + \frac{k(k+1)}{4n}\right| = \left|\sum_{j=1}^{k} g\left(j/2n\right)\right| \leq \sum_{j=1}^{k} \left|g\left(j/2n\right)\right| \leq \sum_{j=1}^{k} \left|g\left(k/2n\right)\right| = k\left|g\left(k/2n\right)\right|.$$

v) En remplaçant k par $x\sqrt{n}$ (plus rigoureusement, sa partie entière $[x\sqrt{n}]$), on obtient

$$\left|\ln(\mathbb{P}(A_{[x\sqrt{n}]})) + \frac{[x\sqrt{n}]([x\sqrt{n}]+1)}{4n}\right| \leq [x\sqrt{n}]\left|g\left([x\sqrt{n}]/2n\right)\right|.$$

Mais comme $\mathbb{P}(A_{[x\sqrt{n}]}) = \mathbb{P}(T_n > [x\sqrt{n}]+1)$ et que $g(x) = o(x)$, on obtient

$$\left|\ln(\mathbb{P}(T_n > [x\sqrt{n}]+1)) + \frac{[x\sqrt{n}]([x\sqrt{n}]+1)}{4n}\right| \leq [x\sqrt{n}]\left|g\left([x\sqrt{n}]/2n\right)\right| \longrightarrow 0,$$

ce qui donne le résultat demandé.

vi) Une autre façon de l'écrire est

$$\lim_{n\to\infty} \mathbb{P}(T_n/\sqrt{n} > x) = \exp(-\frac{x^2}{4}) = 1 - F_a(x),$$

avec $a = 4$. Autrement dit, $T_n/\sqrt{n}$ converge vers R_4, dont l'espérance vaut (question a) $\sqrt{\pi}$.

vii) Si l'on tire $20 = 2\sqrt{100}$ chaussettes dans un tiroir de 100 paires de chaussettes, la probabilité d'avoir (au moins) une paire peut être approchée par $\mathbb{P}(R_4 < 2) = 1 - e^{-1} \approx 0.63$.