Examen final du 8 janvier 2015 (1ère 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 X une v.a. de densité f sur $\mathbb{R}_+$, où

$$f(x) = \frac{e^{-2\sqrt{x}}}{\sqrt{x}} \qquad x \geq 0.$$

a) Calculer $\mathbb{E}(X)$.
b) Déterminer la loi de $\sqrt{X}$.

Solution de l'exercice 1.

a) Par définition, $\mathbb{E}(X) = \int_0^\infty x f(x)\,dx$ et par le changement de variable $z = \sqrt{x}$, on obtient

$$\mathbb{E}(X) = \int_0^\infty 2z^3 f(z^2)\,dz = \int_0^\infty 2z^2 e^{-2z}\,dz.$$

Par deux intégrations par parties successives, on obtient alors

$$\mathbb{E}(X) = \int_0^\infty 2z e^{-2z}\,dz = \int_0^\infty e^{-2z}\,dz = \frac{1}{2}.$$

b) En fait, si h est une fonction arbitraire positive ou bornée, on obtient par le même changement de variable

$$\mathbb{E}(h(\sqrt{X})) = \int_0^\infty h(\sqrt{x}) f(x)\,dx = \int_0^\infty 2z h(z) f(z^2)\,dz = \int_0^\infty 2h(z) e^{-2z}\,dz.$$

La densité de $\sqrt{X}$ est donc $2e^{-2z}$, c'est-à-dire que $\sqrt{X}$ est une variable exponentielle de paramètre 2.

Exercice 2. On cherche à prévoir l'achat d'un nombre suffisamment grand de galettes des rois pour que chacun des n convives d'une fête gourmande ait au moins une fois la fève. On suppose qu'il n'y a qu'une seule fève par galette et que pour chaque galette partagée, chaque convive a la même probabilité (égale à $1/n$, donc) de l'avoir. On suppose également que chaque galette est mangée intégralement avant de passer à la suivante.
On numérote les convives de 1 à n du plus jeune au plus vieux.
On définit la variable aléatoire $X_k \in \{1, \ldots, n\}$ comme le numéro du convive qui obtient la fève de la $k-$ème galette. Par exemple, l'événement $\{X_3 = 1\}$ est réalisé ssi c'est le plus jeune convive qui obtient la fève de la 3ème galette.

a) Question préliminaire. Soit G une v.a. géométrique de paramètre a, c'est-à-dire telle que $\mathbb{P}(G = n) = (1 - a)\, a^{n-1}$, $n \geq 1$.
   i) Donner une expression de $\mathbb{P}(G \geq n)$.
   ii) Donner la fonction génératrice f de G.
   iii) Donner (ou en déduire) $\mathbb{E}(G)$ et $\mathrm{Var}(G)$.
   iv) Soit Z une variable exponentielle de paramètre 1. Rappeler la densité de Z et sa fonction de répartition. Calculer également $\mathbb{P}(Z > x)$ pour $x \geq 0$.

b) Dans cette question, on cherche seulement à savoir combien il faut prévoir de galettes pour que le plus jeune convive ait au moins une fois la fève. Soit donc S_n la première galette où le convive 1 a la fève :

$$S_n := \min\{k \geq 1 : X_k = 1\}.$$

   i) Que peut-on dire de la suite (X_k) ? En déduire la loi de S_n.
   ii) Soit $x > 0$. Que vaut $\mathbb{P}(S_n > nx)$?
   iii) Montrer que S_n/n converge en loi vers une variable aléatoire que l'on précisera.
   iv) Avec n galettes, où n est suffisamment grand, quelle est approximativement la probabilité que le plus jeune convive ait au moins une fois la fève ?

c) On revient maintenant au problème de départ, c'est-à-dire l'étude de la loi du nombre T_n de galettes nécessaires pour que chaque convive ait au moins une fois la fève.
   On pose N_k le nombre de convives qui, juste après avoir fini la $k-$ème galette, ont déjà eu la fève au moins une fois, de sorte que

$$T_n := \min\{k \geq 1 : N_k = n\}.$$

   Pour tout $1 \leq j \leq n$, on définit $T_j := \min\{k \geq 1 : N_k = j\}$ et $G_j := T_j - T_{j-1}$, avec la convention $T_0 = 0$.
   i) Montrer que $T_n = \sum_{j=1}^{n} G_j$.
   ii) Montrer que les variables aléatoires (G_j) sont indépendantes, et que G_j est une v.a. géométrique de paramètre $(j - 1)/n$.
   iii) Montrer que $\mathbb{E}(G_{j+1}) = n/(n - j)$ et que $\mathrm{Var}(G_{j+1}) = (nj)/(n - j)^2$.
   iv) Montrer que

$$\mathbb{E}(T_n) = n \sum_{k=1}^{n} \frac{1}{k} \quad \text{et} \quad \mathrm{Var}(T_n) = n \sum_{k=1}^{n} \frac{n-k}{k^2}.$$

   v) Montrer que $\mathbb{E}(T_n) \sim n \ln(n)$ lorsque $n \to \infty$ et que $\mathrm{Var}(T_n) \leq cn^2$, pour une certaine constante c que l'on pourra préciser.
   vi) Montrer que pour tout $\varepsilon > 0$,

$$\mathbb{P}(|T_n - \mathbb{E}(T_n)| > \varepsilon n \ln(n)) \leq \frac{c}{\varepsilon^2 \ln(n)^2}.$$

   vii) Montrer que $T_n/n\ln(n)$ converge vers 1 lorsque $n \to \infty$, dans un sens que l'on précisera. Comparer avec le résultat obtenu à la question b.

Solution de l'exercice 2.

a) i) $\mathbb{P}(G \geq n) = \sum_{k\geq n}(1-a)a^{k-1} = (1-a)a^{n-1}\sum_{k\geq n} a^{k-n} = a^{n-1}$.

ii) Pour tout $s \in [0,1]$, $f(s) = \mathbb{E}(s^G) = \sum_{n\geq 1}(1-a)a^{n-1}s^n = (1-a)s/(1-as)$, que l'on peut aussi écrire

$$f(s) = \frac{(1-a)s}{1-as} = \frac{1-a}{a}\left(-1+\frac{1}{1-as}\right)$$

iii) Un calcul élémentaire donne alors

$$f'(s) = \frac{1-a}{(1-as)^2} \quad \text{et} \quad f''(s) = \frac{2a(1-a)}{(1-a)^3},$$

de sorte que

$$f'(1) = \frac{1}{1-a} \quad \text{et} \quad f''(1) = \frac{2a}{(1-a)^2}.$$

Or $\mathbb{E}(G) = f'(1)$ et $f''(1) = \mathbb{E}(G(G-1))$, de sorte que $\mathbb{E}(G) = (1-a)^{-1}$ et

$$\mathrm{Var}(G) = f''(1) + f'(1) - f'(1)^2 = \frac{2a}{(1-a)^2} + \frac{1}{1-a} - \frac{1}{(1-a)^2} = \frac{a}{(1-a)^2}.$$

iv) $f(x) = e^{-x}$, $F(x) = 1 - e^{-x}$, $\mathbb{P}(Z > x) = 1 - F(x) = e^{-x}$.

b) i) On remarque que les (X_k) sont i.i.d. uniformes donc les événements $\{X_k = 1\}$ sont indépendants et ont la même probabilité $1/n$. Donc S_n suit la loi géométrique de paramètre $1 - 1/n$.

ii) Si $[y]$ désigne la partie entière de y, alors

$$\mathbb{P}(S_n > nx) = \mathbb{P}(S_n \geq [nx]+1) = \left(1-\frac{1}{n}\right)^{[nx]} = \exp\left[[nx]\ln\left(1-\frac{1}{n}\right)\right]$$

iii) En utilisant $\ln(1-(1/n)) = -(1/n) + o(1/n)$, on obtient la convergence de $\mathbb{P}(S_n/n > x)$ vers e^{-x}, autrement dit, S_n/n converge en loi vers une variable exponentielle de paramètre 1.

iv) La probabilité demandée est $P(S_n \leq n) = \mathbb{P}(S_n/n \leq 1)$, qui vaut approximativement $\mathbb{P}(Z \leq 1)$, où Z est une variable exponentielle de paramètre 1. La probabilité cherchée peut donc être approchée par $1 - e^{-1} \approx 0.63$. On aurait pu obtenir ce résultat en comptant le nombre de fois où le convive 1 a la fève en n galettes, qui suit la loi binomiale de paramètres n et $1/n$, et qui converge donc vers une v.a. de Poisson de paramètre 1.

c) i) Série télescopique : $\sum_{j=1}^n G_j = \sum_{j=1}^n (T_j - T_{j-1}) = T_n - T_0 = T_n$.

ii) Quitte à sommer sur tous les ordres dans lequel les convives ont la fève pour la première fois, on peut supposer que cet ordre est l'ordre croissant des âges. On peut alors écrire pour tout $n-$uplet d'entiers $(a_1, \ldots, a_n)$ l'égalité entre événements

$$\bigcap_{j=1}^{n}\{G_j = a_j\} = \bigcap_{j=1}^{n}\{X_{t_j} \geq j\} \cap \bigcap_{h=1}^{a_j-1}\{X_{t_{j-1}+h} < j\}$$

où $t_j := a_1 + \cdots + a_j$. Or les (X_i) sont i.i.d. uniformes donc

$$\mathbb{P}(\forall 1 \leq j \leq n, G_j = a_j) = \prod_{j=1}^{n} \mathbb{P}(X_{t_j} \geq j) \prod_{h=1}^{a_j-1} \mathbb{P}(X_{t_{j-1}+h} < j) = \prod_{j=1}^{n} ((n-j+1)/n)\,((j-1)/n)^{a_j-1},$$

ce qui prouve le résultat.

iii) La v.a. G_{j+1} est géométrique de paramètre j/n donc d'après la question préliminaire,

$$\mathbb{E}(G_{j+1}) = \frac{1}{1-(j/n)} = \frac{n}{n-j},$$

et

$$\mathrm{Var}(G_{j+1}) = \frac{j/n}{(1-(j/n))^2} = \frac{nj}{(n-j)^2}.$$

iv) L'espérance étant linéaire,

$$\mathbb{E}(T_n) = \sum_{j=0}^{n-1} \mathbb{E}(G_{j+1}) = \sum_{j=0}^{n-1} \frac{n}{n-j} = n \sum_{k=1}^{n} \frac{1}{k},$$

avec le changement d'indice $k = n-j$. De même, comme les (G_j) sont indépendantes, la variance de leur somme égale la somme de leurs variances, ainsi

$$\mathrm{Var}(T_n) = \sum_{j=0}^{n-1} \mathrm{Var}(G_{j+1}) = \sum_{j=0}^{n-1} \frac{nj}{(n-j)^2} = n \sum_{k=1}^{n} \frac{n-k}{k^2},$$

avec le même changement d'indice.

v) Il est bien connu que $\sum_{k=1}^{n} \frac{1}{k} \sim \ln(n)$, ce qui donne $\mathbb{E}(T_n) \sim n\ln(n)$. D'autre part $\sum_{k=1}^{n}(1/k^2) \leq c$, où $c = \sum_{k=1}^{\infty}(1/k^2) = \pi^2/6$, donc $\mathrm{Var}(T_n) \leq cn^2$.

vi) Par l'inégalité de Bienaymé-Chebyshev,

$$\mathbb{P}(|T_n - \mathbb{E}(T_n)| > \varepsilon n \ln(n)) \leq \frac{\mathrm{Var}(T_n)}{\varepsilon^2 n^2 \ln(n)^2} \leq \frac{c}{\varepsilon^2 \ln(n)^2}.$$

vii) Comme $\mathbb{E}(T_n)/n\ln(n) \to 1$ et que d'après la question précédente

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{T_n}{n\ln(n)} - \frac{\mathbb{E}(T_n)}{n\ln(n)}\right| > \varepsilon\right) = 0,$$

on en conclut que $T_n/(n\ln(n))$ converge vers 1 en probabilité.