Examen LM346, 1ère session de l'année 2018–2019, sans document, ni calculatrice.

(1) On considère une variable aléatoire X de densité $f_X(x) = Cx\sin(x)1_{[0,\pi]}(x)$.

Donner la constante C et la fonction de répartition $F_X(x)$ (On pourrait penser à une intégration par parties).

(2) Décrire une méthode de simulation de X possible à partir du résultat de (1). Donne-t-elle des résultats satisfaisants dans ce cas ? Pourquoi ?

(3) Soit U une variable aléatoire de loi uniforme sur $[0,1]$. Comment simuler une variable aléatoire de loi uniforme sur $[0,\pi]$? *Justifier* votre réponse (sinon elle ne sera pas prise en compte).

(4) On simule $W_1, Z_1, W_2, Z_2, \ldots$ une suite de variables aléatoires independantes, $W_1, W_2, \ldots$ étant de loi uniforme sur $[0,\pi]$, $Z_1, Z_2, \ldots$ étant de loi unforme sur $[0, D]$ où D est une constante, choisie telle que $D \geq \max_{0\leq x\leq\pi} Cx\sin x$.

On pose $\nu(\omega) = \min\left\{n \geq 1 : \frac{Z_n}{W_n} \leq C\sin(W_n)\right\}$. Calculer $P(\nu(\omega) = k)$. Quelle est la loi de ν ?

(5) Rédiger une méthode alternative à celle proposée dans la question (2) pour simuler une variable aléatoire de même loi que X. (On demande de rédiger et non seulement nommer).

(6) Calculer $P(0 \leq W_\nu \leq \pi/3)$ (ceci ne demande pas beaucoup de calculs si vous avez fait les questions précédentes).

(7) Calculer $\mathbf{E}\nu$ en fonction de D. Comment devrait-on choisir D pour minimiser $E\nu$?

(8) Formuler (proprement!) le Théorème de la limite Centrale vectoriel.

Soient $X_1, X_2, \ldots, X_n, \ldots$ des variables aléatoires indépendantes, de même loi, de second moment fini, $\mathbf{E}X_1 = 0$. Calculer

$$\lim_{n\to\infty} E\exp\Big(-(3/2)\frac{(X_1+\cdots+X_n)^2}{n\mathrm{Var}X_1}\Big).$$

(Ce calcul est très rapide).

(9) Soient $X_1, X_2, \ldots, X_n, \ldots$ comme dans la question précédente. Pour $\gamma \in]0,1/2[$, on note $q(\gamma)$ tel que

$(\sqrt{2\pi})^{-1}\int_{q(\gamma)}^{\infty}\exp(-t^2/2)dt = \gamma$. Donner la valeur de la limite suivante :

$$\lim_{n\to\infty} \mathbf{P}\Big(-q(\gamma) < \frac{X_1+\cdots+X_n}{\sqrt{n\mathrm{Var}\,(X_1)}} < q(\gamma)\Big).$$

(10) Soit $\gamma \in]0,1/2[$. Construire un intervalle de confiance asymptotique $[I_n(\gamma), \infty[$ de niveau de fiabilité $1-2\gamma$ pour la variance $\mathrm{Var}\,(X_1)$, c'est-à-dire tel que

$$\lim_{n\to\infty} \mathbf{P}\big(\mathrm{Var}\,(X_1) \in [I_n(\gamma),\infty[\big) \geq 1-2\gamma.$$

(11) Le deuxième tour de l'élection présidentielle au sufrage universel entre deux candidats A et B s'annonce. Une proportion p de votants est déjà determinée dans son choix : exactement la moitié compte voter pour A et l'autre moitié pour B. La proportion $(1-p)$ de la population n'est pas encore determinée dans son choix. On veut estimer p. Pour cela on fait un sondage de 10000 individus : 4600 se disent indécis, 2800 vont voter pour A et 2600 pour B. Construire un intervalle de confiance asymptotique pour le paramètre p de niveau $1-2\gamma$.

TOURNEZ LA PAGE SVP

(12) **Partie 3.** On considère une chaîne de Markov $(X_n)_{n\geq 0}$ de matrice de transition

$$\begin{pmatrix} 1/4 & 3/4 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 2/3 & 0 & 0 \\ 1/4 & 0 & 1/8 & 1/4 & 1/4 & 1/8 \\ 3/5 & 0 & 0 & 2/5 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/4 & 1/8 & 1/8 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Donner les classes d'états de cette chaine de Markov, dire si elles sont récurrentes ou transientes.

(13) Soit $h_i^2 = P(T^2 < \infty \mid X_0 = i)$ où $T^2 = \inf\{n > 0 : X_n = 2\}$. Calculer h_i^2 pour $i = 1, 2, 3, 4, 5, 6$.

(14) On note $N_i = \sum_{n=0}^{\infty} 1_{\{i\}}(X_n)$ le nombre de visites dans l'état i. Donner $P(N_1 = k \mid X_0 = 2)$ $k = 0$, pour $k = 2$ et pour $k = \infty$. Donner $P(N_1 = k \mid X_0 = 3)$ $k = 0$, pour $k = 2$ et pour $k = \infty$.

(15) On note $h_i^6 = P(T^6 < \infty \mid X_0 = i)$ avec $T^6 = \inf\{n > 0 : X_n = i\}$. Calculer h_i^6, pour $i = 1, 2, 3, 4, 5, 6$.

(16) Donner toutes les mesures invariantes de cette chaine de Markov.

(17) Donner $\lim_{n\to\infty} P(X_n = 2 \mid X_0 = i)$ pour $i = 1, 2, 3, 4, 5, 6$.

(18) Calculer $P(X_3 = 1 \mid X_0 = 4)$.

(19) Soit $A = \{1, 2, 4, 6\}$. Soit $\tau_A = \min\{n \geq 0 : X_n \in A\}$.
Calculer $E(\tau_A \mid X_0 = 3)$.

(20) Soit $T^2 = \min\{n > 0 : X_n = 2\}$. Donner $E(T^2 \mid X_0 = 2)$. (Ceci ne demande pas beaucoup de calculs si les calculs précédents sont corrects.)

Correction

- $\int_0^x t\sin t dx = -\int_0^x t d\cos t = -x\cos x + \int_0^x cosx dx = \sin x - x\cos x$ Donc $const = 1/\pi$, la fonction de répartition est 0 pour $x<0$, $(\sin x - x\cos x)/\pi$ pour $0 \le x \le 1$, et 1 pour $x \ge 1$.

- Pour réaliser cette méthode, pour tout $u \in]0,1[$, il faut chercher $X(u) \in]0,\pi[$ tel que $\sin X(u) - X(u)\cos X(u) = \pi u$. Alors la v.a. $X(U)$ est de même loi que X. En absence d'expréssion explicite pour $X(u)$, la méthode n'est pas très satisfaisante.

- C'est πU. Pour justifier, on prend une fonction mesurable bornée, $Ef(\pi U) = \int_0^1 f(\pi u)du = \int_0^\pi f(t)(1/\pi)dt$, donc la densité de πU est $(1/\pi)1_{[0,\pi]}(t)$.

- $P(\nu = k) = (1-s)^{k-1}s$ où $s = P(\frac{Z_n}{W_n} \le C\sin(W_n)) = P(Z_n < CW_n \sin W_n) = (1/\pi)(1/D)\int_0^\pi \int_0^{Cw_n \cos w_n} dzdw = 1/(\pi D)$. C'est la loi géomètrique du paramètre s.

- Il faut simuler W_ν (un Thm du cours).

- Comme W_ν et X sont de même loi, on calcule $P(0 \le X \le \pi/3) = F_X(\pi/3) = (1/\pi)(\sqrt{3}/2 - (\pi/3)\times 1/2)$.

- $E\nu = \sum_{k\ge 1} k(1-s)^{k-1}s = s\Big(-\sum_{k\ge 0}(1-s)^k\Big)' = -s\big((1-(1-s))^{-1}\big)' = -(-1)s \times s^{-2} = s^{-1}$, donc $E\nu = \pi D$. Il faut choisir D minimal possible.

8 Soient $X_1, X_2, \ldots$ une suite de v.a. indépendantes de même loi, de second moment fini, alors $(X_1 + \cdots + X_n - n\mathbf{E}X_1)/\sqrt{n\mathrm{Var}\,(X_1)}$ converge en loi(!) vers la loi Normale d'espérance 0 et de Variance 1 quand $n \to \infty$.

La convergence en loi d'une suite ξ_n vers ξ signifie que pour toute fonction continue bornée (ce qui est la cas de $\exp(-(3/2)t^2)$) on a $Ef(\xi_n) \to Ef(\xi)$. On a donc par le Thm de la limite Centrale à la limite $(1/\sqrt{2\pi})\int_R \exp(-(3/2)t^2/2)\exp(-t^2/2)dt$ ce qui après le changement de variable $t = s/2$ est bien 1/2.

9 La limite vaut $(\sqrt{2\pi})^{-1}\int_{-q(\gamma)}^{q(\gamma)} \exp(-t^2/2)dt = 1 - 2\gamma$ par le Thm de la limite centrale.

10 L'inégalité $-q(\gamma) < \frac{X_1+\cdots+X_n}{\sqrt{n\mathrm{Var}\,(X_1)}} < q(\gamma)$ est équivalente à $\sqrt{\mathrm{Var}(X_1)} > \max\{(X_1 + \cdots X_n)/(\sqrt{n}q(\gamma)),\ -(X_1 + \cdots X_n)/(\sqrt{n}q(\gamma))\} = |(X_1 + \cdots X_n)/(\sqrt{n}q(\gamma))|$.
Alors $I_n(\gamma) = (X_1 + \cdots + X_n)^2/(n(q(\gamma))^2)$.

11 On pose $X_i = 0$ si le $i$ème sondé n'est pas determiné dans son choix et $X_i = 1$ (resp. $X_i = -1$) si le $i$ème sondé est pour le candidat A (resp. B). Alors $X_1, \ldots, X_{10000}$ sont des v.a. indép., de même loi, à valeurs dans 0, 1, -1 avec probabilités $1-p$, $p/2$, $p/2$ respectivement. On remarque que $\mathbf{E}X_1 = 0$, $\mathrm{Var}X_1 = p$. Le résultat du sondage nous donne : $X_1 + \cdots + X_{10000} = 2800 - 2600 = 200$. Alors $I_{10000}(\gamma) = (X_1 + \cdots + X_{10000})^2/(10000(q(\gamma))^2) = 4/(q(\gamma))^2$. L'intervalle approximative pour p est $[4/(q(\gamma))^2, 1]$.

12 Les classes récurrentes sont $\{6\}$ et $\{1,2,4\}$. La classe transiente est $\{3,5\}$.

13 $h_2^2 = h_4^2 = h_1^2 = 1$ car dans une classe récurrente chaque état est visité p.s. $h_6^2 = 0$ car 6 est un état absorbant. On cherche h_3^2 et h_5^2 a partir du système $h_3^2 = (1/8)h_3^2 + (1/4)h_5^2 + 1/4 \times 1 + 1/4 \times 1 + 1/8 \times 0 + 0 \times 1$, $h_5^2 = 1/8h_5^2 + 1/4h_3 + 1/4 \times 1 + 1/4 \times 1 + 1/8 \times 0 + 0 \times 1$, $h_3^2 = h_5^2 = 4/5$.

14 Comme les états 1 et 2 appartiennent à la même classe réccurente, $\{1\}$ est visité au départ de 2 p.s., et il est visité ensuite une infinité de fois. Pour $k = \infty$, $P(N_1 = k \mid X_0 = 2)$ vaut 1, et donc pour $k = 0, 2$, elle vaut zéro.

A partir de l'état 3, la classe $\{1,2,4\}$ est visité avec probabilité 4/5. Si cette classe est visitée, chaque état est visitée une infinité de fois p.s. Alors $P(N_1 = k \mid X_0 = 3)$ vaut 4/5 pour $k = \infty$, elle vaut 0 pour $k = 2$ et elle vaut 1/5 pour $k = 0$.

15 On déduit $h_3^6 = 1 - h_3^2 = 1 - 4/5 = 1/5$, de même $h_5^6 = 1/5$, $h_2^6 = h_4^6 = h_1^6 = 0$, $h_6^6 = 1$.

16 On pose $\pi_3 = \pi_5 = 0$, $\pi_6 = 1$, on résout alors $\pi = \pi P$ qui se réduit à un système de trois équations avec les inconnues π_2, π_4, π_1. On obtient les mesures invariantes $(c_1(8/27), c_1(1/3), 0, c_1(10/27), 0, c_2)$ $c_1, c_2 \ge 0$, $c_1 + c_2 = 1$.

17 Comme $\{1,2,4\}$ est apériodique ($p_{2,2} > 0$), cette limite vaut 1/3 pour $i = 1,2,4$, elle vaut $(4/5) \times 1/3$ pour $i = 3,5$, elle vaut 0 pour $i = 6$.

18 On calcule $(P^3)_{4,1} = 51/125$.

19 On note $k_i = E(\tau \mid X_0 = i)$. On doit résoudre $k_3 = 1 + 1/8k_3 + 1/4k_5$, $k_5 = 1 + 1/4k_3 + 1/8k_5$, alors $k_3 = 8/5$.

20 On a $E(\tau_2 \mid X_0 = 2) = 1/(1/3) = 3$, car la mesure de probabilité invariante de la classe fermée $\{1, 2, 4\}$ est $(8/27, 1/3, 10/27)$.