Examen final du 7 janvier 2016 (1ère session)

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

Exercice 1. Soit $g : [0,1] \to \mathbb{R}$ une fonction bornée. On cherche à donner une approximation de l'intégrale $I = \int_0^1 g(x)\,\mathrm{d}x$ à l'aide d'un algorithme aléatoire. Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes uniformes sur $[0,1]$. On pose pour tout $n \geq 1$

$$I_n = \frac{1}{n}\sum_{k=1}^{n} g(U_k).$$

a) i) Exprimer en fonction de g les quantités suivantes : $E(g(U_k))$, $E(I_n)$, $\mathrm{Var}(I_n)$.

ii) Que dire de la convergence de la suite de variables aléatoires (I_n) ?

b) Soit $M > 0$ tel que pour tout $x \in [0,1]$, $|g(x)| \leq M$.

i) En majorant $\mathrm{Var}(g(U_k))$ par $E(g(U_k)^2)$, donner une majoration de $\mathrm{Var}(I_n)$ en fonction de M et n.

ii) En utilisant une inégalité classique du cours, montrer que pour tout $n \geq 1$ et tout $\varepsilon > 0$,

$$P(|I_n - I| \geq \varepsilon) \leq \frac{M^2}{n\varepsilon^2}.$$

c) On suppose maintenant que $M = 1$.

i) Utiliser l'inégalité précédente pour donner le nombre minimal de variables U_k à utiliser pour obtenir un résultat approché de I à 0,1 près avec une confiance de 99%.

ii) Alternativement, comment peut-on utiliser le théorème central limite pour obtenir un contrôle de l'erreur entre I_n et I ? (On rappelle que si Z suit une loi gaussienne centrée réduite, $P(|Z| \leq 1,96) \approx 0,99$).

Solution de l'exercice 1.

a) i) g est bornée donc $g(U_k)$ a une espérance et un moment d'ordre 2 finis. L'espérance vaut :

$$m = E(g(U_k)) = \int_0^1 g(x)\mathrm{d}x = I,$$

et

$$\sigma^2 = \mathrm{Var}(g(U_k)) = \int_0^1 g^2(x)\mathrm{d}x - I^2.$$

Par linéarité de l'espérance, on a

$$E(I_n) = \frac{1}{n}\sum_{k=1}^{n} E(g(U_k)) = I.$$

La dernière égalité provient du fait que les (U_k) sont identiquement distribuées. Pour calculer la variance, on utilise aussi le fait qu'elles sont identiquement distribuées mais aussi indépendantes :

$$\mathrm{Var}(I_n) = \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}(g(U_k)) = \frac{\sigma^2}{n}.$$

ii) Les $(g(U_k))$ sont iid et admettent un moment d'ordre 1 fini. D'après la loi forte des grands nombres, la suite (I_n) converge p.s. vers I.

b) i)

$$\mathrm{Var}(I_n) \leq E(I_n^2) \leq \frac{1}{n}\int_0^1 M^2 \mathrm{d}x = \frac{M^2}{n}.$$

ii) On applique l'inégalité de Bienaymé-Chebyshev à I_n qui a un moment d'ordre 2 fini.

$$P(|I_n - I| \geq \varepsilon) \leq \frac{\mathrm{Var}(I_n)}{\varepsilon^2} \leq \frac{M^2}{n\varepsilon^2}.$$

c) i) On prend $M = 1$, $\varepsilon = 0,1$. On veut que $\frac{M^2}{n\varepsilon^2} = 0,01$. Ce qui donne $n = 10000$.

ii) Comme les $g(U_k)$ sont iid et ont un moment d'ordre 2, d'après le théorème central limite, $\frac{\sqrt{n}}{\sigma}(I_n - I)$ converge en loi vers une variable gaussienne centrée réduite. En particulier,

$$\lim_{n\to\infty} P(|I_n - I| \leq \frac{1,96\sigma}{\sqrt{n}}) = \int_{-1,96}^{1,96} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x \simeq 0,99$$

En majorant encore σ par $M^2 = 1$, et $1,96$ par 2, on augmente l'événement dans la probabilité. On trouve donc que si on est déjà dans le régime asymptotique où on peut faire l'approximation du théorème central limite pour n fini, on a la bonne précision pour $\frac{2}{\sqrt{n}} = 0,1$, soit $n = 400$.

Exercice 2. Soit $\theta > 0$. Des individus numérotés $1, 2, \ldots, n$, arrivent successivement dans une salle de restaurant contenant une infinité de tables infiniment longues. Le premier individu s'assied à une table au hasard. Pour tout entier $k \geq 1$, lorsque l'individu $k+1$ arrive, avec la probabilité $k/(k+\theta)$ il choisit uniformément au hasard un des k convives déjà attablés et s'assied à la même table, ou avec la probabilité $\theta/(k+\theta)$ occupe une table inoccupée.

On désigne par K_n le nombre de tables occupées lorsque n convives se sont installés, et par f_n la fonction génératrice de K_n.

a) i) Montrer que

$$K_n = \sum_{k=0}^{n-1} \varepsilon_k,$$

où les $(\varepsilon_k)_{k=0,\ldots,n-1}$ sont des variables de Bernoulli indépendantes dont on précisera les probabilités de succès respectives.

ii) Montrer les formules suivantes

$$E(K_n) = \sum_{k=0}^{n-1} \frac{\theta}{k+\theta} \quad \text{et} \quad \mathrm{Var}(K_n) = \sum_{k=0}^{n-1} \frac{k\ \theta}{(k+\theta)^2}\ .$$

iii) Soit Q_n le polynôme de degré n défini par $Q_n(s) = \prod_{k=0}^{n-1}(s+k)$. Montrer que

$$f_n(s) = \frac{Q_n(\theta s)}{Q_n(\theta)}.$$

b) i) À l'aide du théorème de Borel–Cantelli, calculer $P(\lim_{n\to\infty} K_n = +\infty)$.

ii) Donner un équivalent de $E(K_n)$ et de $\mathrm{Var}(K_n)$ lorsque $n \to \infty$.
Indication. On pourra calculer $E(K_n) - \theta H_n$ et $\mathrm{Var}(K_n) - \theta H_n$, où $H_n := \sum_{k=1}^{n-1} \frac{1}{k}$.

iii) Montrer que $K_n/\ln(n)$ converge vers θ en un sens que l'on précisera.

Solution de l'exercice 2.

a) i) La v.a. K_n peut s'écrire sous la forme

$$K_n = \sum_{k=0}^{n-1} \varepsilon_k,$$

où ε_k est l'indicatrice de l'évènement 'le convive $k+1$ choisit une table inoccupée', donc les (ε_k) sont des variables de Bernoulli indépendantes d'espérance $\theta/(k+\theta)$, pour $k = 0, \ldots, n-1$.

ii) On trouve immédiatement les formules demandées, en précisant comme il se doit que l'espérance est linéaire, et que la variance d'une somme de variables indépendantes égale la somme de leurs variances.

iii) En utilisant l'indépendance des ε_k,

$$f_n(s) = \prod_{k=0}^{n-1} E\left(s^{\varepsilon_k}\right) = \prod_{k=0}^{n-1}\left(\frac{k}{k+\theta} + \frac{\theta s}{k+\theta}\right) = \prod_{k=0}^{n-1}\left(\frac{k+\theta s}{k+\theta}\right)$$

b) i) Les variables ε_k sont indépendantes et

$$\sum_{k\geq 0} P(\varepsilon_k = 1) = \sum_{k\geq 0} \frac{\theta}{k+\theta} = \infty,$$

Donc avec probabilité 1, la suites des indices k tels que $\varepsilon_k = 1$ est infinie, et donc la probabilité que leur somme est infinie vaut 1, qui est la probabilité demandée.

ii) D'abord

$$E(K_n) - \theta H_n = 1 + \sum_{k=1}^{n-1}\left(\frac{\theta}{k+\theta} - \frac{\theta}{k}\right) = 1 - \sum_{k=1}^{n-1} \frac{\theta^2}{k(k+\theta)},$$

qui converge vers une quantité finie lorsque $n \to \infty$. Or $\lim_n H_n = \infty$, donc $E(K_n) \sim \theta H_n$. De même,

$$\mathrm{Var}(K_n) - \theta H_n = \sum_{k=1}^{n-1}\left(\frac{k\theta}{(k+\theta)^2} - \frac{\theta}{k}\right) = \sum_{k=1}^{n-1} \frac{k^2\theta - \theta(k+\theta)^2}{k(k+\theta)^2} = -\sum_{k=1}^{n-1} \frac{\theta(2k\theta+\theta^2)}{k(k+\theta)^2},$$

qui converge également vers une quantité finie lorsque $n \to \infty$, ce qui donne $\mathrm{Var}(K_n) \sim \theta H_n$. Comme il est connu que $H_n \sim \ln(n)$, on a

$$E(K_n) \sim \theta \ln(n) \quad \text{et} \quad \mathrm{Var}(K_n) \sim \theta \ln(n),$$

lorsque n tend vers $+\infty$.

iii) Soit $X_n := K_n / \ln(n)$. D'abord $\lim_n E(X_n) = \theta$. Ensuite $\mathrm{Var}(X_n) = \mathrm{Var}(K_n)/\ln(n)^2 \sim \theta/\ln(n)$, donc $\mathrm{Var}(X_n) \to 0$ et X_n converge vers θ dans L^2. Par conséquent, X_n converge vers θ en probabilité, ce que l'on peut aussi prouver directement par l'inégalité de Bienaymé-Chebyshev. On peut aussi remarquer par l'équivalent $\mathrm{Var}(X_n) \sim \theta/\ln(n)$ que la série de terme général $\mathrm{Var}(X_n)$ ne converge pas, si bien qu'on ne peut pas conclure quant à la convergence p.s.