Devoir libre 13 (facultatif)

À rendre lundi 2 mars 2020

Toutes les variables aléatoires étudiées dans ce problème sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs réelles.
On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires et on pose :

$$\forall n \geqslant 1, \quad S_n = \sum_{k=1}^{n} X_k.$$

Préliminaires

1. Soit (X_n) une suite de variables aléatoires réelles de même loi, admettant une espérance m. Énoncer *avec précision* la loi faible des grands nombres pour la suite (X_n).

2. Soient $\delta > 0$ et A un borélien de $\mathbb{R}$ tel que $]m-\delta, m+\delta[\subset \mathbb{R} \setminus A$ (où $\mathbb{R} \setminus A$ désigne le complémentaire de A dans $\mathbb{R}$). Déterminer

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{S_n}{n} \in A\Big).$$

L'objet du problème est de préciser de manière quantitative les résultats ci-dessus.

Première partie

Un premier exemple : le cas gaussien

Dans cette partie, $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes suivant toutes une loi normale centrée réduite $\mathcal{N}(0,1)$.

3. Quelle est, pour $n \geqslant 1$, la loi de la variable aléatoire $\frac{S_n}{n}$?

4. Soit $\delta > 0$.

a. Montrer que :

$$\mathbb{P}\Big(\Big|\frac{S_n}{n}\Big| \geqslant \delta\Big) = 2\,\mathbb{P}\Big(\frac{S_n}{n} \geqslant \delta\Big) = \sqrt{\frac{2n}{\pi}} \int_{\delta}^{+\infty} \exp\Big(-\frac{nt^2}{2}\Big)\,\mathrm{d}t.$$

b. En posant $u = n(t-\delta)$, montrer que :

$$\mathbb{P}\Big(\Big|\frac{S_n}{n}\Big| \geqslant \delta\Big) = \sqrt{\frac{2}{n\pi}} \exp\Big(-\frac{n\delta^2}{2}\Big) \int_{0}^{+\infty} \exp\Big(-\frac{u^2}{2n} - u\delta\Big)\,\mathrm{d}u.$$

5. **a.** Montrer que pour tout $x \geqslant 0$, $0 \leqslant 1 - \mathbf{e}^{-x} \leqslant x$.

b. Montrer que l'intégrale $\int_0^{+\infty} \mathbf{e}^{-u\delta}\,\mathrm{d}u$ converge et la calculer.

c. Déterminer

$$\lim_{n\to\infty} \Big(\int_0^{+\infty} \mathbf{e}^{-u\delta}\,\mathrm{d}u - \int_0^{+\infty} \exp\Big(-\frac{u^2}{2n} - u\delta\Big)\,\mathrm{d}u\Big).$$

d. En déduire, lorsque $n \to \infty$, un équivalent de

$$\mathbb{P}\Big(\Big|\frac{S_n}{n}\Big| \geqslant \delta\Big).$$

Deuxième partie

Quelques résultats généraux

À l'instar des variables aléatoires discrètes, on admettra que si X et Y sont deux variables aléatoires à densité, indépendantes et admettant une espérance, alors XY admet une espérance et $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.
Étant donné un événement A, on note $\mathbb{1}_A$ la variable aléatoire indicatrice de l'événement A :

$$\mathbb{1}_A : \omega \in \Omega \longmapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}.$$

Soit X une variable aléatoire réelle (discrète ou à densité). Pour tout $s \in \mathbb{R}$ tel que $\mathbf{e}^{sX}$ admet une espérance, on pose $\varphi(s) = \mathbb{E}(\mathbf{e}^{sX})$.
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes suivant toutes la même loi que X.

6. **a.** Soit $s > 0$. Montrer que si $\varphi(s)$ existe, alors $\varphi(t)$ existe pour tout $t \in [0, s]$.
b. Justifier que le résultat précédent est encore valable pour $s < 0$ (en considérant que $[0, s]$ désigne le segment $[s, 0]$).
c. Montrer que pour tout $n \geqslant 1$, pour tout $s \in \mathbb{R}$ tel que $\varphi(s)$ existe,

$$\mathbb{E}\Big[\exp\Big(s\frac{S_n}{n}\Big)\Big] = \Big(\varphi\big(\frac{s}{n}\big)\Big)^n.$$

7. **a.** Soient Y une variable aléatoire réelle et $s > 0$ tel que $\mathbb{E}(\mathbf{e}^{sY})$ existe. Justifier que :

$$\forall a \in \mathbb{R}, \quad \mathbb{P}(Y \geqslant a) \leqslant \mathbf{e}^{-as}\,\mathbb{E}(\mathbf{e}^{sY}).$$

b. En déduire que pour $s > 0$ tel que $\varphi(s)$ existe, on a :

$$\forall n \geqslant 1, \quad \forall a \in \mathbb{R}, \quad \mathbb{P}\Big(\frac{S_n}{n} \geqslant a\Big) \leqslant \mathbf{e}^{-as}\Big(\varphi\big(\frac{s}{n}\big)\Big)^n.$$

8. **a.** Soient Y une variable aléatoire réelle et $s < 0$ tel que $\mathbb{E}(\mathbf{e}^{sY})$ existe. Pour $a \in \mathbb{R}$, établir l'inégalité $\mathbb{1}_{[Y \leqslant a]} \leqslant \mathbf{e}^{s(Y-a)}$ et en déduire que $\mathbb{P}(Y \leqslant a) \leqslant \mathbf{e}^{-as}\,\mathbb{E}(\mathbf{e}^{sY})$.
b. Montrer que pour $s < 0$ tel que $\varphi(s)$ existe :

$$\forall n \geqslant 1, \quad \forall a \in \mathbb{R}, \quad \mathbb{P}\Big(\frac{S_n}{n} \leqslant a\Big) \leqslant \mathbf{e}^{-as}\Big(\varphi\big(\frac{s}{n}\big)\Big)^n.$$

Troisième partie
Un deuxième exemple : le cas binomial

Soient X une variable aléatoire de loi de Bernoulli de paramètre $p \in]0, 1[$ et $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires indépendantes suivant toutes la même loi que X.

9. Expliciter le domaine de définition et l'expression de la fonction $\varphi : s \longmapsto \mathbb{E}(\mathbf{e}^{sX})$.

On considère un réel $a \in]0, 1[$.

10. On suppose dans cette question que $a > p$.
a. Étudier les variations de la fonction $\ell_a : s \in \mathbb{R}_+ \longmapsto as - \ln \varphi(s)$.
b. Montrer que la fonction ℓ_a atteint sur $\mathbb{R}_+$ un maximum strictement positif $h(a, p)$ qu'on exprimera en fonction de a et p.
c. Montrer que :

$$\forall n \geqslant 1, \quad \mathbb{P}\Big(\frac{S_n}{n} \geqslant a\Big) \leqslant \mathbf{e}^{-nh(a,p)}.$$

11. On suppose dans cette question que $a < p$ et on pose $h(a, p) = h(1 - a, 1 - p)$.
a. Quelle est, pour $n \geqslant 1$, la loi de la variable $n - S_n$?
b. Montrer que :

$$\forall n \geqslant 1, \quad \mathbb{P}\Big(\frac{S_n}{n} \leqslant a\Big) \leqslant \mathbf{e}^{-nh(a,p)}.$$

12. Déduire des questions précédentes que pour $0 < \varepsilon < \min\{p, 1 - p\}$,

$$\forall n \geqslant 1, \quad \mathbb{P}\Big(\Big|\frac{S_n}{n} - p\Big| \geqslant \varepsilon\Big) \leqslant 2\mathbf{e}^{-n\min\{h(p+\varepsilon,p),h(p-\varepsilon,p)\}}.$$

13. Soit $\alpha \in]0, 1[$. Montrer que, pour n assez grand, il est toujours possible de trouver deux réels a_1 et a_2 tels que $0 < a_1 < p < a_2 < 1$ vérifiant les inégalités :

$$\mathbb{P}\Big(\frac{S_n}{n} \leqslant a_1\Big) \leqslant \frac{\alpha}{2} \quad \text{et} \quad \mathbb{P}\Big(\frac{S_n}{n} \geqslant a_2\Big) \leqslant \frac{\alpha}{2}.$$

Indication. On pourra étudier les variations de la fonction $a \longmapsto h(a, p)$.

On considère une machine fabriquant un certain type d'objets qui, en fonctionnement normal, produit une proportion $p \in]0, 1[$ d'objets défectueux. Afin de connaître la valeur de p, on teste la machine en prélevant un échantillon de $n \geqslant 1$ objets qu'on analyse. Pour tout $i \in [\![1, n]\!]$, soit X_i la variable aléatoire de Bernoulli définie par

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ième objet prélevé est défectueux} \\ 0 & \text{sinon} \end{cases}.$$

On suppose que dans les conditions de prélèvement, les variables aléatoires $X_1, \dots, X_n$ sont indépendantes.

14. À partir de l'échantillon $(X_1, \dots, X_n)$, on décide d'estimer p à l'aide de la variable aléatoire $F_n = \frac{S_n}{n}$.

a. Justifier que $\mathbb{E}(F_n) = p$: on dit que F_n est un *estimateur sans biais* de p.

b. Calculer le *risque quadratique* $r_n = \mathbb{E}\big((F_n - p)^2\big)$ ainsi que sa limite lorsque $n \to \infty$.

c. En déduire que la suite $(F_n)_{n\geqslant 1}$ converge en probabilité vers p : on dit que F_n est un *estimateur convergent* de p.

15. Afin d'affiner l'estimation précédente, on décide de construire un intervalle $[U_n, V_n]$, où U_n et V_n sont deux variables aléatoires fonctions de $X_1, \dots, X_n$ ne dépendant pas de p, tel que :

$$\lim_{n\to\infty} \mathbb{P}\big(U_n \leqslant p \leqslant V_n\big) \geqslant 1 - \alpha.$$

Un tel intervalle est appelé [1] *intervalle de confiance asymptotique* pour p au niveau de confiance $1 - \alpha$.

a. Quelle est la limite en loi de la suite de terme général

$$\sqrt{n}\frac{F_n - p}{\sqrt{F_n(1 - F_n)}} = \sqrt{\frac{p(1-p)}{F_n(1-F_n)}} \times \sqrt{n}\frac{F_n - p}{\sqrt{p(1-p)}}, \qquad n \geqslant 1\,?$$

b. Soit T une variable aléatoire de loi normale centrée réduite. On note t_α le réel caractérisé par la condition $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$, où Φ désigne la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$. Que vaut $\mathbb{P}\big(|T| \leqslant t_\alpha\big)$?

c. Déterminer en fonction de F_n et t_α un intervalle de confiance asymptotique de p au niveau de confiance $1 - \alpha$.

Quatrième partie

Le cas général

Soit X une variable aléatoire de densité f. On considère la fonction

$$L_X : t \longmapsto \mathbb{E}(\mathbf{e}^{tX}) = \int_{-\infty}^{+\infty} \mathbf{e}^{tu} f(u)\,\mathrm{d}u.$$

On suppose que L_X est définie sur un intervalle $]\alpha, \beta[$ contenant 0.

16. Soient $t \in\,]\alpha, \beta[$ et $\delta > 0$ tel que $[t - \delta, t + \delta] \subset\,]\alpha, \beta[$.

a. Montrer que :

$$\forall u \in \mathbb{R}, \quad \left|\mathbf{e}^{\delta u} - 1 - \delta u\right| \leqslant \sum_{n=2}^{\infty} \frac{\delta^n |u^n|}{n!}.$$

b. Montrer que :

$$\forall u \in \mathbb{R}, \quad \mathbf{e}^{tu}\left|\mathbf{e}^{\delta u} - 1 - \delta u\right| f(u) \leqslant \big(\mathbf{e}^{(t-\delta)u} + \mathbf{e}^{(t+\delta)u}\big) f(u).$$

c. En déduire que l'intégrale $\int_{-\infty}^{+\infty} u\mathbf{e}^{tu} f(u)\,\mathrm{d}u$ converge puis que X admet une espérance m.

17. Soient $t \in\,]\alpha, \beta[$ et $\delta > 0$ tel que $[t - \delta, t + \delta] \subset\,]\alpha, \beta[$. Soit $h \in \mathbb{R}^*$ tel que $|h| < \delta$.

a. Montrer que :

$$\forall u \in \mathbb{R}, \quad \left|\frac{\mathbf{e}^{(t+h)u} - \mathbf{e}^{tu}}{h} - u\mathbf{e}^{tu}\right| f(u) \leqslant |h|\, \mathbf{e}^{tu} \sum_{n=2}^{\infty} \frac{\delta^{n-2} |u|^n}{n!} f(u)$$

puis que :

$$\forall u \in \mathbb{R}, \quad \delta^2 \left|\frac{\mathbf{e}^{(t+h)u} - \mathbf{e}^{tu}}{h} - u\mathbf{e}^{tu}\right| f(u) \leqslant |h|\, \mathbf{e}^{tu} \mathbf{e}^{\delta|u|} f(u).$$

b. Montrer que L_X est dérivable en t et que

$$L_X'(t) = \int_{-\infty}^{+\infty} u\mathbf{e}^{tu} f(u)\,\mathrm{d}u.$$

On admettra (et on démontrerait de manière analogue) que la fonction L_X est de classe $\mathscr{C}^2$ sur $]\alpha, \beta[$ et que :

$$\forall t \in\,]\alpha, \beta[, \quad L_X''(t) = \int_{-\infty}^{+\infty} u^2 \mathbf{e}^{tu} f(u)\,\mathrm{d}u.$$

1. La condition ci-dessus sera satisfaite sur le présent exemple, mais est trop restrictive pour être adoptée comme définition dans le cas général.

18. On considère la fonction $\psi : t \longmapsto \ln \mathrm{L_X}(t)$.

a. Donner le domaine de définition de la fonction ψ.

b. Calculer ψ' et ψ''.

c. En utilisant l'inégalité de Cauchy-Schwarz, montrer que ψ' est strictement croissante sur $]\alpha, \beta[$. En déduire que ψ' admet en α (resp. β) une limite (finie ou infinie) qu'on notera $\mathrm{L}_1(\alpha)$ (resp. $\mathrm{L}_1(\beta)$). Quelle est la valeur de $\psi'(0)$?

d. Montrer que ψ admet en α (resp. β) une limite (finie ou infinie) qu'on notera $\mathrm{L}_0(\alpha)$ (resp. $\mathrm{L}_0(\beta)$).

19. Pour $a \in \mathbb{R}$ donné, dresser le tableau de variations de la fonction $\ell_a : t \in]\alpha, \beta[\longmapsto at - \psi(t)$. On distinguera les trois cas $a \geqslant \mathrm{L}_1(\beta)$, $a \leqslant \mathrm{L}_1(\alpha)$ et $\mathrm{L}_1(\alpha) < a < \mathrm{L}_1(\beta)$.

On suppose dorénavant que $\mathrm{L}_1(\alpha) < a < \mathrm{L}_1(\beta)$ et on pose $h(a) = \sup_{\alpha<t<\beta} \ell_a(t)$.

20. Montrer que :

$$h(a) = a(\psi')^{-1}(a) - \psi\big((\psi')^{-1}(a)\big).$$

21. Montrer que si $a > m$,

$$h(a) = \sup_{0<t<\beta} \big(at - \psi(t)\big)$$

puis que si $a < m$,

$$h(a) = \sup_{\alpha<t<0} \big(at - \psi(t)\big).$$

22. Soit $(\mathrm{X}_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi que X. Montrer que si $a > m$,

$$\forall n \geqslant 1, \quad \mathbb{P}\Big(\frac{\mathrm{S}_n}{n} \geqslant a\Big) \leqslant \mathbf{e}^{-nh(a)}$$

puis que si $a < m$,

$$\forall n \geqslant 1, \quad \mathbb{P}\Big(\frac{\mathrm{S}_n}{n} \leqslant a\Big) \leqslant \mathbf{e}^{-nh(a)}.$$

23. Pour $0 < \varepsilon < \min\{\mathrm{L}_1(\beta) - m, m - \mathrm{L}_1(\alpha)\}$, montrer que :

$$\forall n \geqslant 1, \quad \mathbb{P}\Big(\Big|\frac{\mathrm{S}_n}{n} - m\Big| \geqslant \varepsilon\Big) \leqslant 2\mathbf{e}^{-n\min\{h(m-\varepsilon), h(m+\varepsilon)\}}.$$