Méthodes de Monte Carlo – Projet

stoehr@ceremade.dauphine.fr

- **À rendre avant** 23 : 59**, le 21 décembre 2022. Tout retard sera pénalisé : 1 point en moins toutes les 30 minutes à partir de** 00 : 01**.**
- R **est le seul langage autorisé.** Les questions nécessitant un code R sont indiquées par le symbole ♠. L'ensemble des générateurs aléatoires de la librairie de base de R sont autorisés.
- Le rapport (**nom du fichier :** `numero_groupe_noms`, *e.g.*, `001_Stoehr_Luciano`)
  - à rendre au format **.pdf** et doit contenir vos réponses et commentaires. Une rédaction soignée est attendue. Il est important de justifier/commenter les résultats théoriques et numériques.
  - Pour intégrer tout ou partie de votre code et des sorties dans votre rapport, vous pouvez utiliser les outils dédiés : Notebook, Rmarkdown ou LaTeX+ knitr.
  - Les graphiques doivent être soigneusement annotés et présentés (titre, couleur, légendes, ...).
- Une version du code pouvant être testée doit être fournie (**même nom de fichier**). Ce code doit
  - s'exécuter sans erreurs et permettre de reproduire l'intégralité des résultats présentés dans le rapport. Vous préciserez la graine utilisée pour les résultats obtenus.
  - être bien commenté. Il est possible qu'une explication orale vous soit demandée.
  - utiliser autant que possible les spécificités du language (bonus pour les codes les plus efficaces).

Exercice 1. Soient X et Y deux variables aléatoires réelles indépendantes telles que

- X suit la loi gamma de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+^*$

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\theta}\right), \quad \text{avec} \quad \alpha = 3 \quad \text{et} \quad \theta = 2.$$

- Y suit la loi de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$g(y) = \frac{y-\mu+2b}{4b^2} \mathbb{1}_{\{y\in[\mu-2b,\mu]\}} + \frac{1}{2b} \exp\left(-\frac{y-\mu}{b}\right) \mathbb{1}_{\{y>\mu\}}, \quad \text{avec} \quad \mu = -13 \quad \text{et} \quad b = 2.$$

L'objectif de cet exercice est d'estimer

$$\delta = \mathbb{P}[X+Y>t], \quad \text{avec} \quad t = 0.$$

Méthode Monte Carlo n°1.

1. *(♠)* Décrire une méthode pour simuler suivant la loi de (X, Y) et écrire le code R associé. Vous préciserez les différents calculs effectués.
2. *(♠)* Vérifier graphiquement que votre méthode et votre code sont corrects.
3. Écrire l'estimateur Monte Carlo $\widehat{p}_n$ de δ utilisant une suite de n variables aléatoires $(X_k, Y_k)_{k\geq 1}$ *i.i.d.* suivant la loi de (X, Y).

4. *(♠)* Implémenter un code R qui fournit, pour $n = 1000$, une estimation de δ, de l'erreur quadratique moyenne de $\widehat{p}_n$ et un intervalle de confiance asymptotique bilatère pour δ au niveau 95%.

5. *(♠)* Vérifier empiriquement si l'hypothèse du régime asymptotique $\widehat{p}_n$ est satisfaite.

6. On note F la fonction de répartition de X. Montrer que

$$\delta = \mathbb{E}[1 - F(t - Y)],$$

et en déduire un estimateur $\widehat{\delta}_n$ de δ utilisant une suite de n variables aléatoires $(U_k)_{k\geq 1}$ *i.i.d.* suivant la loi uniforme sur $[0,1]$.

7. *(♠)* Implémenter un code R qui fournit, pour $n = 1000$, l'estimation $\widehat{\delta}_n$, l'erreur quadratique moyenne de $\widehat{\delta}_n$ et un intervalle de confiance asymptotique bilatère pour δ au niveau 95%.

8. Selon ces résultats, l'estimateur $\widehat{\delta}_n$ est-il plus performant en terme de variance que l'estimateur $\widehat{p}_n$? Si oui, montrer qu'on a toujours $\mathbb{V}\mathrm{ar}\left[\widehat{\delta}_n\right] \leq \mathbb{V}\mathrm{ar}\left[\widehat{p}_n\right]$.

Méthode Monte Carlo n°2.

9. En utilisant la méthode de la variable antithétique, proposer un estimateur $\widehat{\delta}_n^{\mathrm{A}}$ de δ. Montrer que $\mathbb{V}\mathrm{ar}\left[\widehat{\delta}_n^{\mathrm{A}}\right] \leq \mathbb{V}\mathrm{ar}\left[\widehat{\delta}_n\right]$.

10. *(♠)* Implémenter un code R qui fournit, pour $n = 1000$, l'estimation $\widehat{\delta}_n^{\mathrm{A}}$, l'erreur quadratique moyenne de $\widehat{\delta}_n^{\mathrm{A}}$ et un intervalle de confiance asymptotique bilatère pour δ au niveau 95%.

Méthode Monte Carlo n°3. On note q_ε le quantile d'ordre ε de la loi de X et on définit les fonctions :

$$h_{0,1} \;:\; \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ y & \mapsto & y \end{array} \qquad \text{et} \qquad h_{0,2} \;:\; \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ y & \mapsto & \mathbb{1}_{\{y \geq t - q_\varepsilon\}}. \end{array}$$

Dans la suite on prend $\varepsilon = 0.6$.

11. *(♠)* Entre $h_{0,1}$ et $h_{0,2}$, quelle fonction utiliseriez-vous pour réduire la variance de $\widehat{\delta}_n$ à l'aide de la méthode de la variable de contrôle ? Écrire l'estimateur $\widehat{\delta}_n^{\mathrm{cont}}$ associé.

12. *(♠)* Implémenter un code R qui fournit, pour $n = 1000$, l'estimation $\widehat{\delta}_n^{\mathrm{cont}}$, l'erreur quadratique moyenne de $\widehat{\delta}_n^{\mathrm{cont}}$ et un intervalle de confiance asymptotique bilatère pour δ au niveau 95%.

Pour $(Y_k)_{k\geq 1}$, une suite de variables aléatoires *i.i.d.* suivant g, on veut construire le meilleur estimateur de la forme

$$\widehat{\delta}_n(\beta) = 1 - \frac{1}{n}\sum_{k=1}^{n} F(t - Y_k) - <\beta, \begin{pmatrix} Y_k - \mathbb{E}[Y] \\ \mathbb{1}_{\{Y_k \geq t - q_\varepsilon\}} - \mathbb{P}\left[Y \geq t - q_\varepsilon\right] \end{pmatrix}>, \quad \text{avec} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^2.$$

$<\cdot,\cdot>$ désigne le produit scalaire usuel de $\mathbb{R}^2$, *i.e.*, $<U,V> = U^\intercal V$.

13. Montrer que

$$\mathbb{V}\mathrm{ar}\left[\widehat{\delta}_n(\beta)\right] = \mathbb{V}\mathrm{ar}\left[\widehat{\delta}_n\right] - \frac{1}{n}\left(2<\beta, C> - <\beta, \Sigma\beta>\right), \quad \text{avec}$$

$$C = \begin{pmatrix} \mathrm{Cov}[F(t-Y), Y] \\ \mathrm{Cov}\left[F(t-Y), \mathbb{1}_{\{Y\geq t-q_\varepsilon\}}\right] \end{pmatrix} \quad \text{et} \quad \Sigma = \begin{pmatrix} \mathbb{V}\mathrm{ar}[Y] & \mathrm{Cov}\left[Y, \mathbb{1}_{\{Y\geq t-q_\varepsilon\}}\right] \\ \mathrm{Cov}\left[Y, \mathbb{1}_{\{Y\geq t-q_\varepsilon\}}\right] & \mathbb{V}\mathrm{ar}\left[\mathbb{1}_{\{Y\geq t-q_\varepsilon\}}\right] \end{pmatrix}.$$

14. Calculer explicitement Σ et vérifier qu'elle est inversible.

15. Montrer que

$$\beta^\star = \underset{\beta\in\mathbb{R}^2}{\arg\min}\, \mathbb{V}\mathrm{ar}\big[\widehat{\delta}_n(\beta)\big] = \Sigma^{-1}C \quad \text{et} \quad \mathbb{V}\mathrm{ar}\big[\widehat{\delta}_n(\beta^\star)\big] \leq \mathbb{V}\mathrm{ar}\big[\widehat{\delta}_n\big].$$

16. L'estimateur $\widehat{\delta}_n(\beta^\star)$ est-il préférable en terme de variance à un estimateur basé sur la méthode de la variable de contrôle utilisant $h_0 : y \mapsto h_{0,1}(y) + h_{0,2}(y)$? Justifier rigoureusement votre réponse.

17. *(♠)* Implémenter un code R qui fournit, pour $n = 1000$, l'estimation $\widehat{\delta}_n(\beta^\star)$, l'erreur quadratique moyenne de $\widehat{\delta}_n(\beta^\star)$ et un intervalle de confiance asymptotique bilatère pour δ au niveau 95%.

Méthode Monte Carlo n°4.

18. Soit $h \in L^2([0,1])$. Montrer que pour une suite $(U_k)_{k\geq 1}$ de variables aléatoires *i.i.d.* de loi uniforme sur $[0,1]$, on a

$$\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{k=1}^{n} h\left(\frac{k-1+U_k}{n}\right)\right] \leq \mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{k=1}^{n} h(U_k)\right].$$

19. En déduire un nouvel estimateur $\widehat{\Delta}_n$ de δ. Comment peut-on interpréter cet estimateur ?

20. *(♠)* Implémenter un code R qui fournit, pour $n = 1000$, l'estimation $\widehat{\Delta}_n$, l'erreur quadratique moyenne de $\widehat{\Delta}_n$ et un intervalle de confiance asymptotique bilatère pour δ au niveau 95%.

21. *(♠)* Comparer l'efficacité relative des différents estimateurs de δ que vous avez proposés.

Exercice 2 *(Bonus).* Soient $m \in \mathbb{N}^*$ et $k \in [\![2, m-1]\!]$. Avec la densité g définie à l'Exercice 1 et G la fonction de répartition associée à g, on définit

$$\Psi(x) = \frac{m!}{(k-1)!(m-k)!} G(x)^{k-1}[1-G(x)]^{m-k} g(x).$$

1. *(♠)* Implémenter un algorithme du rejet qui simule n réalisations suivant la densité Ψ.

2. *(♠)* Vérifier graphiquement que votre code est valide pour $m = 15$ et $k = 7$.