Université Paris Dauphine
Département MIDO
Master 1 – Méthodes de Monte Carlo 2019–2020

Méthodes de Monte Carlo
– Projet –

stoehr@ceremade.dauphine.fr

Consignes.

À rendre avant le 08 janvier 2020

- Rapport contenant réponses/commentaires, codes utilisés et sorties (Notebook, Rmarkdown ou LaTeX+ knitr). À défaut, un rapport au format **.pdf** et un script contenant l'ensemble des codes utilisés. Dans ce cas, **il est interdit de copier-coller du code dans le corps du texte**. Une rédaction soignée est attendue. Il est notamment important de justifier/commenter les résultats théoriques et numérique et de préciser le choix des paramètres dans les différentes méthodes utilisées.
- **seul le language R est autorisé**.
- Les codes doivent :
  - être bien commentés. Il est possible qu'une explication orale vous soit demandée.
  - être optimisés (vectorisés) un minimum pour utiliser les spécificités du language.
  - s'exécuter sans erreurs et permettre de reproduire l'intégralité des résultats présentés dans le rapport.
- Les graphiques doivent être soigneusement annotés et présentés (titre, couleur, légendes, ...).
- **Chaque jour de retard sera pénalisé d'un point.**

Exercice 1.

On considère $X_1, \ldots, X_n$, des variables *i.i.d.* suivant la loi de densité donnée, pour tout $x \in \mathbb{R}$ par

$$f(x) = \frac{1}{\beta} \exp\left\{-\exp\left(-\frac{x-\mu}{\beta}\right)\right\} \exp\left(-\frac{x-\mu}{\beta}\right), \quad \text{avec} \quad \mu \in \mathbb{R} \quad \text{et} \quad \beta \in \mathbb{R}_+^*.$$

On note $X_{(1)} = \min(X_1, \ldots, X_n)$ et $X_{(n)} = \max(X_1, \ldots, X_n)$. Pour les implémentations pratiques, on prendra $n = 100$, $\mu = 1$ et $\beta = 2$.

Partie I – Simulations de variables aléatoires

1. **(a)** En utilisant un générateur aléatoire de la loi uniforme sur $[0, 1]$, décrire une méthode de simulation suivant la densité f et donner le code associé.

(b) La loi de densité f s'appelle la loi de Gumbel. Valider votre méthode de simulation à l'aide d'un ou plusieurs outils de comparaison.

2. **(a)** Montrer que la densité jointe de $(X_{(1)}, X_{(n)})$ s'écrit

$$f_{1,n}(x, y) = n(n-1)\{F(y) - F(x)\}^{n-2} f(x) f(y) \mathbb{1}_{\{x \leq y\}},$$

où F est la fonction de répartition associée à la densité f.

(b) Décrire une méthode de simulation suivant la densité $f_{1,n}$ et donner le code associé.

Partie II – Estimation de l'étendue par la méthode de Monte Carlo Classique

L'étendue de l'échantillon est définie par $\Delta = X_{(n)} - X_{(1)}$. Dans cette partie, on s'intéresse à l'estimation de $\delta = \mathbb{E}[\Delta]$.

1. Donner une estimation de δ au niveau de confiance 95% avec une précision $\epsilon = 10^{-2}$, en utilsant :

(a) la densité f ; **(b)** la densité $f_{1,n}$.

Quelle est l'efficacité relative de ces deux méthodes d'estimation ?

2. **(a)** Donner une estimation de δ par la méthode de la variable de contrôle. On donnera un intervalle de confiance au niveau 95%.

(b) Comparer les variances et l'efficacité relative entre la méthode de Monte Carlo Classique et la méthode de la variable de contrôle. Si la méthode de la variable de contrôle permet de réduire la variance, expliquer pourquoi.

Partie III – Estimation d'un événement rare

On considère $V_1, \dots, V_n$ des variables *i.i.d.* suivant la loi exponentielle $\varepsilon(2)$ et on note $V_{(n)} = \max(V_1, \dots, V_n)$. On s'intéresse à l'estimation de $p = \mathbb{P}\left[V_{(n)} \geq t\right]$. Dans les applications numériques, on prendra $t = 8$ et $n = 1000$.

1. Donner une estimation de p par la méthode de Monte Carlo classique avec un intervalle de confiance au niveau 95%. Que constatez vous ?

2. **(a)** Montrer que $V_{(n)} - 0.5 \log(n)$ converge en loi vers une loi de Gumbel dont on précisera les paramètres.

(b) En déduire une estimation de p par la méthode d'échantillonage préférentiel avec un intervalle de confiance au niveau 95%. Vous justifierez le choix de la loi d'importance.

3. Quelle est l'efficacité relative de ces deux méthodes d'estimation ?

4. **(a)** Partant de l'estimateur d'échantillonage préférentiel, construire un nouvel estimateur de p à l'aide de la méthode de la variable antithétique.

(b) Que pouvez-vous dire de l'efficacité relative de cet estimateur par rapport à l'estimateur d'échantillonage préférentiel ? Si cette modification de l'estimateur permet de réduire la variance de la méthode d'estimation, expliquer pourquoi.

Exercice 2. Le nombre de précipitation S sur un mois est modélisé par une loi de Poisson de paramètre $\lambda = 3.7$. La quantité d'eau Q_s tombant lors d'une précipitation s est modélisé par un loi de Weibul de paramètre de forme $k = 0.5$ et de paramètre d'échelle $\lambda = 2$ (on supposera que les précipitations sont indépendantes). La quantité de pluie tombant en 1 mois est donc

$$X = \begin{cases} 0 & \text{, si } S = 0, \\ \sum_{s=1}^{S} Q_s & \text{, sinon.} \end{cases}$$

On s'intéresse aux mois présentant de faibles précipitations et on cherche à estimer $p = \mathbb{P}[X < 3]$ (*i.e.*, il y a moins de 3 cm de pluie par mois).

1. Donner une estimation de p par la méthode de Monte Carlo classique. On donnera l'intervalle de confiance au niveau 95%.

2. **(a)** Donner une estimation de p par la méthode de stratification avec allocation proportionnelle, en précisant les strates utilisées. On donnera l'intervalle de confiance au niveau 95%.

(b) Comparer les variances et l'efficacité relative des deux méthodes d'estimations. Discuter de façon concise les résultats obtenus.

3. **(a)** Proposer une méthode d'estimation de p par la méthode de stratification avec allocation optimale. Quelles difficultés rencontrez-vous ?

(b) Comparer les variances et l'efficacité relative de ces trois méthodes d'estimations. Discuter de façon concise les résultats obtenus.