Simulation et R : Examen

Exercice 1.
On veut simuler suivant une loi dite "normale à support positif" dont la densité est donnée par :

$$f_\mu(x) = \frac{1}{C(\mu)} \exp\big(-\frac{(x-\mu)^2}{2}\big) \mathbb{1}_{\mathbb{R}_+}(x),$$

où $\mu \in \mathbb{R}$.

1. Ecrire $C(\mu)$ en fonction de la fonction de répartition F_μ de la loi $\mathcal{N}(\mu, 1)$. Montrer que $C(\mu) = \sqrt{2\pi}\Phi(\mu)$, où Φ est la fonction de répartition de la loi $\mathcal{N}(0, 1)$.
2. Calculer la fonction de répartition de cette loi "normale à support positif" en fonction de Φ et expliquer pourquoi la méthode d'inversion ne peut pas s'appliquer pour simuler suivant cette loi.

   Etude de l'algorithme du rejet On va, dans la suite, étudier l'algorithme du rejet pour simuler suivant cette loi de densité f_μ. On cherchera donc une densité g et une constante M telle que

   $$\forall x \in \mathbb{R}_+ \ : \ M(\mu) \geq \frac{f_\mu(x)}{g(x)}.$$

   On définit par ailleurs

   $$\tau = \inf\{n \geq 1, f_\mu(X_n) \geq M(\mu) U_n g(X_n)\},$$

   avec $X_1, \ldots, X_n, \ldots$ i.i.d. de densité g et $U_1, \ldots, U_n, \ldots$ i.i.d. de loi $\mathcal{U}([0, 1])$.
3. (a) Que représente τ ? Calculer son espérance en fonction de $M(\mu)$.
   (b) Justifier que $M(\mu) \geq 1$.
4. **Proposition gaussienne.** On étudie le cas où $g = g_\mu$ et g_μ est la densité de la loi $\mathcal{N}(\mu, 1)$.
   (a) Etudier le rapport $f_\mu(x)/g_\mu(x)$ pour $x \in \mathbb{R}_+$. En déduire la valeur optimale de la constante $M(\mu)$, notée $M_1(\mu)$ dans ce cas.
   (b) En étudiant la limite de $M_1(\mu)$ quand $\mu \to -\infty$ et puis $\mu \to +\infty$, commenter l'efficacité de cet algorithme dans ces deux cas.
5. **Proposition exponentielle.** On étudie le cas où $g = g_\alpha$ et g_α est la densité de la loi $\mathcal{E}(\alpha)$.
   (a) Montrer que, dans ce cas, la constante optimale $M(\mu)$ vaut $M_\alpha(\mu)$ avec

   $$M_\alpha(\mu) = \frac{1}{\alpha C(\mu)} \exp\big(\frac{\alpha}{2}(2\mu + \alpha)\big).$$

   (b) En déduire que le meilleur choix, noté α^*, de α pour μ donné est

   $$\alpha^* = \frac{\sqrt{\mu^2 + 4} - \mu}{2}.$$

   Quelle est la valeur de la constante $M_{\alpha^*}(\mu)$ quand $g = g_{\alpha^*}$?
   (c) En utilisant l'approximation :

   $$\Phi(-x) \underset{+\infty}{\sim} \frac{\exp(-x^2/2)}{\sqrt{2\pi}x},$$

   montrer que cet algorithme est efficace quand $\mu \to -\infty$.
   (d) Montrer qu'il est inefficace quand $\mu \to +\infty$.

   Indication. *On utilisera les équivalents :*

   $$\frac{\sqrt{\mu^2 + 4} - \mu}{2} \underset{+\infty}{\sim} \frac{1}{\mu} \quad et \quad \frac{\sqrt{\mu^2 + 4} - \mu}{2} \underset{-\infty}{\sim} -\mu$$

6. **Recommandation.** Pour les valeurs $\mu = -10$, $\mu = 0$ et $\mu = +10$, quel algorithme (proposition gaussienne/proposition exponentielle) recommandez-vous ?

Indication. $\frac{2}{\sqrt{2\pi}} \exp(1/2) \simeq 1.32$

Exercice 2.
On veut calculer l'intégrale

$$I = \int_{\mathbb{R}} \exp(-u^2) f_\nu(u) du,$$

où f_ν est la densité de la loi de Student $\mathcal{S}t(\nu)$ à $\nu > 1$ degrés de liberté définie par

$$f_\nu(x) = C_\nu \big(1 + x^2/\nu\big)^{-\frac{\nu+1}{2}} \quad \text{avec} \quad C_\nu = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}.$$

1. **Approximation directe.** Soit $T_1, \dots, T_N$ N v.a. i.i.d. de loi $\mathcal{S}t(\nu)$.
   (a) Proposer un estimateur $\hat{I}_1$ de l'intégrale I en fonction de $T_1, \dots, T_N$.
   (b) Montrer que la variance de cet estimateur s'écrit

$$\mathbb{V}(\hat{I}_1) = \frac{1}{N}\big(\mathbb{E}(\exp(-2T^2)) - I^2\big)$$

2. **Approximation via la loi de Cauchy.** Soit $C_1, \dots, C_N$ N v.a. i.i.d. de loi de Cauchy de densité

$$g_\nu(x) = \frac{1}{\pi\nu} \frac{1}{1 + (x/\nu)^2}$$

   (a) Proposer un nouvel estimateur $\hat{I}_2$ de l'intégrale I calculé à partir de $C_1, \dots, C_N$.
   (b) Montrer que la variance de $\hat{I}_2$ peut s'écrire

$$\mathbb{V}(\hat{I}_2) = \frac{1}{N}\Big(\mathbb{E}\big(\exp(-2T^2)\frac{f_\nu(T)}{g_\nu(T)}\big) - I^2\Big),$$

   où T suit la loi du Student $\mathcal{S}t(\nu)$.
   (c) Dans la figure "Rapport de densités", la courbe de $x \to f_\nu(x)/g_\nu(x)$ pour $\nu = 10$ et $x \in [-5, 5]$.
   Au vu de cette courbe, peut-on déduire que cette seconde méthode va améliorer l'estimation de I ? Expliquer pourquoi.

3. **Approximation par "Rao-Blackwellisation".** On rappelle que T suit une loi de Student $\mathcal{S}t(\nu)$ si et seulement si

$$T \overset{loi}{=} \frac{X}{\sqrt{Z/\nu}},$$

   où $X \sim \mathcal{N}(0, 1)$ est indépendant de $Z \sim \chi^2_\nu$.
   (a) Montrer successivement que

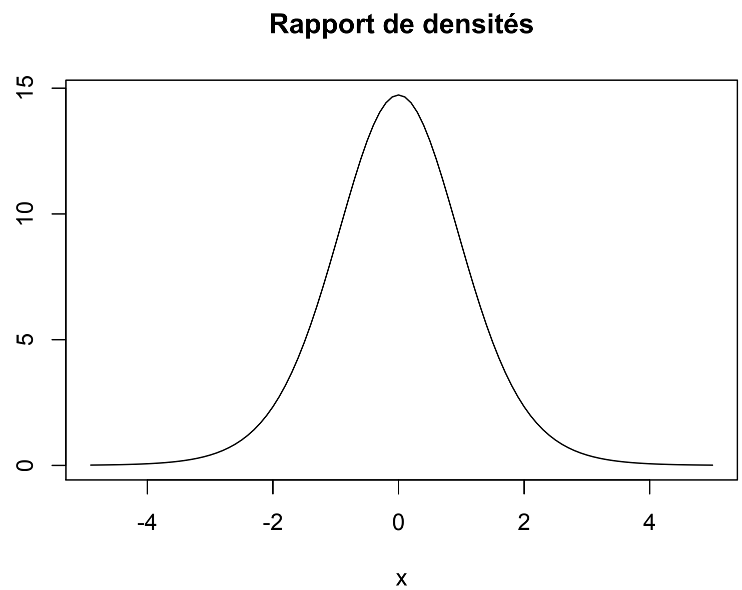
$$\begin{aligned} I &= \int_{\mathbb{R}} \int_{\mathbb{R}+} \exp(-(x/\sqrt{z/\nu})^2) f_{(X,Z)}(x, z) dx dz = \int_{\mathbb{R}} \int_{\mathbb{R}+} \exp(-t^2) f_{(T,Z)}(t, z) dx dz \\ &= \mathbb{E}(\mathbb{E}(\exp(-T^2)|Z)), \end{aligned}$$

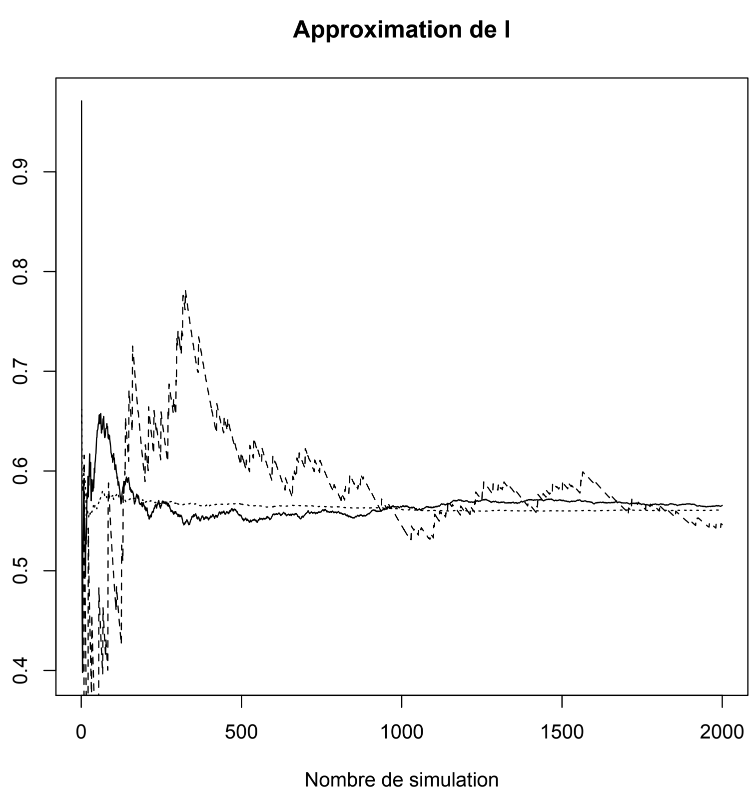
   en explicitant les densités $f_{(X,Z)}$ du couple (X, Z) et $f_{(T,Z)}$ du couple (T, Z) en fonction des densités φ de la loi normale $\mathcal{N}(0, 1)$ et f_Z de Z (de la loi χ^2_ν).
   (b) En déduire que la v.a. $T|Z = z$ suit la loi $\mathcal{N}(0, \nu/z)$. En utilisant l'indication ci-dessous, donner une nouvelle expression de I en fonction de ν et de Z v.a. qui suit une loi χ^2_ν .
   Indication. *Si X est une v.a. gaussienne de loi $\mathcal{N}(0, \sigma^2)$ alors*

$$\mathbb{E}(\exp(-X^2)) = \frac{1}{\sqrt{2\sigma^2 + 1}}$$

   (c) A partir de la question précédente, proposer un nouvel estimateur $\hat{I}_3$ de I construit à partir de N v.a. i.i.d. $Z_1, \dots, Z_N$ de loi du χ^2_ν. Donner l'expression de sa variance.
   (d) Grâce à l'inégalite de Jensen, montrer que $\hat{I}_3$ a une variance toujours plus petite que celle de $\hat{I}_1$ (indication : on approxime par "Rao-Blackwellisation"...)

4. On a représenté dans la figure "Approximation de I", les trajectoires des estimations $N \to \hat{I}_1(N)$, $N \to \hat{I}_2(N)$ et $N \to \hat{I}_3(N)$. Associer à chaque trajectoire une des trois méthodes étudiées dans les questions précédentes.


Rapport de densités
15
10
5
0
-4
-2
0
2
4
x


Approximation de I
0.9
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Nombre de simulation