Echantillonnage préférentiel

Travaux dirigés

Pierre Gloaguen

Contents

1 Évènement rare

On veut estimer la probabilité p^* qu'une loi normale centrée réduite dépasse la valeur 3.

1. Pour un effort de Monte Carlo de taille M, proposer un estimateur de Monte Carlo pour p^* se basant sur la loi $\mathcal{N}(0,1)$.
2. Implémenter cet estimateur sur `R` pour un effort de Monte Carlo de taille $M = 10000$
3. On se propose d'utiliser un échantillonnage préférentiel pour estimer cette probabilité. On utilisera comme loi d'échantillonnage une loi exponentielle translatée de 3, de paramètre λ notée $Y \sim t\mathcal{E}(3, \lambda)$, i.e., la variable aléatoire Y telle que $Y - 3 \sim \mathcal{E}(\lambda)$ Calculer les poids d'importance associés, et proposer un estimateur de p^*
4. Ecrire la variance de l'estimateur comme fonction de λ. Comment doit on choisir λ?
5. Donner un estimateur empirique de cette variance, et donner l'expression de l'intervalle de confiance à 95% pour l'estimateur de p^*.
6. Implémenter cet estimateur sur `R` avec $\lambda = 3.5$ et le comparer à celui de Monte Carlo.

2 Cas problématique

2.1 Premier cas jouet

Afin de montrer qu'un mauvais choix de loi d'échantillonnage peut augmenter drastiquement la variance de l'estimateur, on peut prendre un exemple très simple.

Supposons qu'on veuille estimer par méthode de Monte Carlo $\mathbb{E}[X]$ où $X \sim \mathcal{N}(0, 1)$. On se propose de le faire par méthode de Monte Carlo standard et par échantillonnage préférentiel en tirant dans une loi $Y \sim \mathcal{N}(\mu, 1)$ où μ est choisi par l'utilisateur.

1. Quelle est la variance de l'estimateur de Monte Carlo standard?
2. Quelle est la variance de l'estimateur par échantillonnage préférentiel?

2.2 Encore pire!

Soit X une variable aléatoire de densité $f_X(x) = 3x^{-4}\mathbf{1}_{x\geq 1}$ (on dit que X suit une loi de Pareto de paramètres (1, 3)).

1. Calculer $p = \mathbb{P}(X > 10)$
2. Supposons qu'on veuille estimer p par méthode de Monte Carlo. Proposer une méthode de simulation de X par méthode d'inversion. En déduire un estimateur de Monte Carlo standard.
3. Représenter graphiquement les performances empiriques de cet estimateur pour un effot Monte Carlo de $M = 10^4$.
4. On propose maintenant un estimateur par échantillonnage préférentiel comme dans l'exercice précédent. On choisit comme densité d'échantillonnage celle d'une loi exponentielle translatée de 10, de paramètre $\lambda = 1$. Donnez l'estimateur associé et mettez le en place sous R. Que constatez vous? Comment l'expliquez vous.
5. Que pouvez vous dire de la variance de cet estimateur?

3 Marche aléatoire du joueur optimiste

Le problème suivant peut être vu par le prisme d'un joueur dans un jeu à espérance négative, qui décide *a priori* de s'arrêter, soit quand il est ruiné, soit quand ses gains ont dépassé le jackpot.

Soit $X_0 = 0$ et $(X_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. de loi $\mathcal{N}(-\mu, 1)$ où μ est une constante strictement positive (gain pour une partie). On note $S_n = \sum_0^n X_n$ (somme des gains jusqu'à l'instant n). On se donne un réel $r < 0$ (représentant la ruine) et un réel $j > 0$ (représentant le jackpot). On considère le temps d'arrêt $T = \min\{n, S_n \leq r \text{ ou } S_n \geq j\}$. On souhaite estimer la probabilité de sortir vainqueur, soit $p^* = \mathbb{P}(S_T \geq j)$.

1. Proposer un algortihme de simulation de pour une trajectoire $(S_n)_{1\leq n\leq T}$.
2. Pour un effort de Monte Carlo de taille M, proposer un estimateur pour p^*.
3. Pour $\mu = 1, r = -50, j = 5$, implémenter cet estimateur, quelle est la valeur de $\hat{p}^*$ obtenue pour $M = 1000$? Représentez une trajectoire d'une estimation (i.e., la valeur de l'estimation en fonction de M)?
4. Pour une séquence observée $s_1, \dots, s_T$ simulée grâce à la méthode de Monte Carlo ci dessus, donner la densité de l'échantillon, notée $f(s_{1:T})$.
5. On considère maintenant la marche aléatoire où les X_n sont tirés selon la loi $\mathcal{N}(\mu, 1)$ (n a toujours $X_0 = 0$ et $S_n = \sum_0^n X_n$). Pour une séquence $s_1, \dots, s_T$ simulée ainsi on note $g(s_1, \dots, s_T)$, la densité de l'échantillon, écrire le rapport $\frac{f(s_{1:T})}{g(s_{1:T})}$.
6. Proposer un estimateur par échantillonnage préférentiel pour estimer p^*, basé sur la simulation selon la densité g. L'estimateur sera noté $\hat{p}_M^{IS}$. De cet estimateur, vous déduirez que $p^* \leq \mathrm{e}^{-2\mu j}$.
7. Implémenter cet algorithme, tracez la valeur de l'estimation ainsi que son intervalle de confiance en fonction de M.