Année : 2008-2009 1er semestre
Niveau : MASTER IS 1ère année
Cours : Logiciel R
Enseignant : A. Illig

TP6 : Simulations en R, méthodes de Monte Carlo

Table des matières

1 Intégration par la méthode de Monte Carlo

La méthode de Monte Carlo consiste à calculer ou plus exactement à approximer une intégrale de la forme suivante :

$$\mathcal{J} = \int h(x)f(x)\,dx \tag{1}$$

où f est une densité. Si X est une variable aléatoire réelle de densité f, on remarque que l'intégrale $\mathcal{J}$ peut s'écrire sous forme d'une espérance mathématique :

$$\mathcal{J} = \mathbb{E}(h(X)). \tag{2}$$

Cette remarque est à la base de la méthode de Monte Carlo.

1.1 Description de la méthode

Considérons $X_1, \ldots, X_n, \ldots$ des copies indépendantes d'une variable aléatoire réelle X de densité f vérifiant $\mathbb{E}(h(X)) < \infty$ avec h la fonction de (1). Posons $Y = h(X)$ et $Y_i = h(X_i)$, puis appliquons la loi forte des grands nombres à $Y_1, \ldots, Y_n, \ldots$ ($\mathbb{E}(Y) < \infty$ par hypothèse) :

$$\hat{\mathcal{J}}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{p.s.} \mathbb{E}(Y) = \mathcal{J}.$$

Par conséquent, lorsque n est grand,

$$(\hat{\mathcal{J}}_n)_{obs} = \frac{1}{n}\sum_{i=1}^{n}(h(X_i))_{obs}$$

fournit une approximation de $\mathcal{J}$.

Remarque : Pour appliquer la méthode de Monte Carlo, il est nécessaire de savoir simuler des réalisation de la loi de densité f !

On peut s'interroger sur la précision de cette méthode. Une réponse à cette question est donnée par le théorème de la limite centrale (TLC) qui permet d'évaluer la précision de la méthode. Supposons que $\sigma^2 = \mathrm{Var}(Y)$ est finie et appliquons le TLC aux variables $Y_1, \ldots, Y_n, \ldots$:

$$\frac{\sqrt{n}(\hat{\mathcal{J}}_n - \mathcal{J})}{\sigma} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Par ailleurs, σ^2 étant inconnu, on introduit l'estimateur consistant de la variance :

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mathcal{J}}_n)^2$$

et le lemme de Slutsky permet de conserver l'approximation normale donnée par le TLC en remplaçant σ^2 par S_n^2 :

$$\frac{\sqrt{n}(\hat{\mathcal{J}}_n - \mathcal{J})}{S_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1). \tag{3}$$

Ainsi,

$$\hat{\mathcal{J}}_n \sim_{approx} \mathcal{N}(\mathcal{J}, \frac{S_n^2}{n})$$

et l'erreur d'approximation est évaluée par $\frac{(S_n^2)_{obs}}{n}$.

Remarques : L'erreur de précision est d'autant plus petite que $\sigma^2 = \mathrm{Var}(Y)$ est petite car

$$S_n^2 \xrightarrow[n\to\infty]{p.s.} \sigma^2.$$

La convergence (3) permet également d'obtenir un intervalle de confiance de niveau donné pour $\mathcal{J}$.

Exemple : Pour approximer le nombre π à l'aide de la méthode de Monte Carlo, remarquons que

$$\mathcal{I} = 4\int_0^1 \sqrt{x^2-1}\,dx = \pi.$$

En conservant les notations précédentes, on observe que $\mathcal{I} = \mathbb{E}(h(X))$ avec X de densité $f(x) = 1_{[0,1]}(x)$ et $h(x) = 4\sqrt{x^2-1}$. Voici un petit programme R qui donne une approximation de π et l'erreur de précision :

```
> n=100
> U=runif(n)
> h=function(x){
+ x=4*sqrt(1-x^2)
+ ;x}
> Y=matrix(0,ncol=1,nrow=n)
> for (i in 1:n){
+ Y[i]=h(U[i])}
> I=mean(Y)
> V=var(Y)/n
```

1.2 Amélioration de la méthode

On considère que l'intégrale $\mathcal{J}$ définie par (1) peut se réécrire sous la forme suivante :

$$\mathcal{J} = \int h(x)\frac{f(x)}{g(x)}g(x)\,dx \tag{4}$$

où g est une densité que l'on appelle alors fonction d'importante. La méthode de Monte Carlo s'applique maintenant à des variables $\tilde{X}_1, \ldots, \tilde{X}_n$ de densité g

et une fonction $H(x) = h(x)\frac{f(x)}{g(x)}$ pour approximer

$$\mathcal{J} = \mathbb{E}(H(\tilde{X}))$$

où $\tilde{X}$ une variable aléatoire de densité g. Cette alternative présente plusieurs intérêts :

– Elle est applicable à toute densité g telle que $supp(g) \supset supp(f)$.
– Elle permet d'utiliser une densité g correspondant à une loi plus simple.
– Pour un bon choix de g, la variance de la méthode peut être réduite.

Le choix de g minimisant la variance de $\hat{\mathcal{J}}_n$ est une fonction dépendant de la densité f et de la fonction h :

$$g^*(x) = \frac{|h(x)|f(x)}{\mathcal{J}}.$$

Dans ce cas, la variance $\sigma^2 = \mathrm{Var}(\tilde{Y})$ où $\tilde{Y} = H(\tilde{X})$ est nulle lorsque h est positive mais en pratique la fonction g^* est inconnue car elle dépend de $\mathcal{J}$.

2 Exercices

2.1 Exercice 1

1. Considérons une variable aléatoire X de densité proportionnelle à la fonction suivante :

$$\tilde{f}(x) = (2 + sin^2(x)) \exp(-(2 + cos^3(3x) + sin(2x))x) 1_{[0,+\infty}(x).$$

   Déterminer une approximation par une méthode de simulation de l'espérance de la variable X.

2. Soit $\tilde{X}$ une variable aléatoire de densité g proportionnelle à la fonction suivante :

$$\tilde{g}(x) = (2 + sin^2(x)) \exp(-(3 + cos(3x))x) 1_{[0,+\infty}(x).$$

   Déterminer une approximation par une méthode de simulation de la constante C telle que $g = C\tilde{g}$.

2.2 Exercice 2

1. Calculer par la méthode de Monte Carlo une approximation de

$$\mathcal{I} = \int_2^\infty \frac{1}{\pi(1 + x^2)} \, dx.$$

2. Ouvrir un dispositif graphique à deux fenêtres. Représenter dans la première fenêtre, l'évolution de l'approximation en fonction de la taille de l'échantillon utilisé. Dans la deuxième fenêtre, représenter l'évolution de la précision en fonction de la taille de l'échantillon.
3. Remarquer que

$$\mathcal{I} = \frac{1}{2} - \int_0^2 \frac{1}{\pi(1 + x^2)} \, dx$$

   et proposer une nouvelle approximation de $\mathcal{I}$.
4. Reprendre la question 2. pour la deuxième méthode d'approximation.