TP de modélisation stochastique - séance 1

Professeur: Olivier Wintenberger. Chargé de TP: Sébastien Farkas, `ISUP CS1`

Thème : Simulation de variables aléatoires. NB : Les questions (BONUS) ne sont pas obligatoires.

1. Simulation de variables aléatoires uniformes

(a) À partir de la méthode des générateurs congruentiels linéaires, créez une fonction permettant de simuler n réalisations d'une variable aléatoire uniforme sur $[0; 1]$. Par défaut, fixer les arguments de la fonction (n, seed, m, a, b) égaux à ceux du générateur *RANDU (IBM)*, que l'on rappelle ci-dessous :
— seed = 1
— m = $2^{31} - 1$
— a = 65539
— b = 0

(b) Testons l'efficacité du générateur précédemment défini. Les deux principales propriétés à vérifier sont les suivantes :
— les réalisations sont identiquement distribuées selon $\mathcal{U}([0; 1])$,
— les réalisations sont indépendantes.

1) Pour tester la première propriété, représentez par un histogramme les réalisations et superposez la densité d'une loi $\mathcal{U}([0; 1])$ puis commentez.

2) Pour tester la seconde, à partir des réalisations $(U_i)_{i\in 1,\ldots,n}$, créez un triplet (X, Y, Z) défini par $X = (U_i)_{i\in 1,\ldots,n-2}$, $Y = (U_i)_{i\in 2,\ldots,n-1}$ et $Z = (U_i)_{i\in 3,\ldots,n}$. Représentez ce triplet grâce à la fonction *plot3d* du package *rgl*. Que pouvez-vous conjecturer en observant le cube sous des angles différents ?

(c) Testons à présent l'efficacité du générateur par défaut de R stocké dans la fonction *runif* en répondant à nouveau aux deux précédentes sous-questions. Que peut-on conjecturer concernant les performences respectives du générateur RANDU et du générateur par défaut de R ?

2. Méthodes générales de simulation de variables aléatoires

À partir de maintenant, nous utilisons la fonction *runif* pour simuler une ou des variables uniformes.

(a) Simulation selon la méthode de transformation inverse.

1) Appliquons cette méthode à la simulation d'une variable suivant une loi exponentielle. Définissez une fonction permettant de simuler n réalisations indépendantes d'une loi exponentielle de paramètre λ.

2) Testez la fonction en simulant des réalisations suivant une loi $\mathcal{E}(\lambda)$ pour $\lambda \in \{1; 10\}$ et en superposant, pour chacun des paramètres :
— un histogramme des simulations,
— la densité théorique de $\mathcal{E}(\lambda)$,
— les moyennes théorique et empirique (par des droites verticales).

(b) Simulation selon la méthode des distributions mélanges.

1) Appliquons cette méthode à la simulation selon la loi de Laplace, de densité $f(x) = \frac{e^{-|x|}}{2}$. En remarquant que cette loi peut être définie à partir de la loi exponentielle, écrire une fonction permettant de simuler n réalisations indépendantes d'une loi de Laplace.

2) Testez la fonction en simulant des réalisations et en superposant un histogramme des simulations, la densité théorique d'une loi de Laplace, et les moyennes théorique et empirique (par des droites verticales).

(c) Simulation selon la méthode de rejet.

1) Écrivez une fonction permettant de simuler n réalisations d'une loi dont on connaît la densité f, en utilisant une constante M, une densité instrumentale g selon laquelle il est possible, grâce à une fonction rg, de simuler des réalisations. Affichez également, avant la fin de l'exécution, le ratio de rejet de la méthode sur la console.

2) Appliquons cette fonction à la simulation d'une variable suivant une loi normale centrée réduite. En utilisant comme densité instrumentale celle de le loi de Laplace et une constante associée qu'il conviendra de calculer, simulez n réalisations d'une loi normale centrée réduite.

3) Pour différentes valeurs de n, testez la performance de cette application en superposant un histogramme des simulations, la densité théorique d'une loi normale centrée réduite, et les moyennes théorique et empirique. Notez également le ratio de rejet de la méthode.

4) (BONUS) En modifiant la sortie de la précédente fonction, représentez la convergence lorsque n tend vers l'infini du ratio de rejet de la méthode. La valeur de convergence est égale à $\frac{1}{M}$.

(d) Simulation selon la méthode de ratio rejet :

1) Écrivez une fonction permettant de simuler n réalisations d'une loi dont on connaît la densité h à une constante près, à partir des bornes de l'espace $C_h = \{(u,v) \in \mathbb{R}^2 : 0 \leq u \leq \sqrt{h\left(\frac{v}{u}\right)}\}$. Affichez également, avant la fin de l'exécution, le ratio de rejet de la méthode sur la console.

2) Appliquons cette méthode à la simulation d'une variable suivant une loi normale centrée réduite en utilisant $h(x) = e^{-x^2}$. Simulez n réalisations d'une loi normale centrée réduite.

3) Pour différentes valeurs de n, testez la performance de cette application en superposant un histogramme des simulations, la densité théorique d'une loi normale centrée réduite, et les moyennes théorique et empirique. Notez également le ratio de rejet de la méthode.

4) En modifiant la sortie de la précédente fonction, représentez empiriquement l'espace C_h.

5) (BONUS) En modifiant la sortie de la précédente fonction, représentez la convergence lorsque n tend vers l'infini du ratio de rejet de la méthode. Quelle est la valeur de cette limite théorique ?

(e) Simulation selon la méthode de Box Muller.

1) Écrivez une fonction permettant de simuler un couple de n réalisations d'une loi normale centrée réduite en utilisant la méthode de Box Muller. Affichez également, avant la fin de l'exécution, le ratio de rejet de la méthode sur la console.

2) Pour différentes valeurs de n, testez la performance de cette application en superposant un histogramme des simulations, la densité théorique d'une loi normale centrée réduite, et les moyennes théorique et empirique. Notez également le ratio de rejet de la méthode.

3) (BONUS) En modifiant la sortie de la précédente fonction, représentez la convergence lorsque n tend vers l'infini du ratio de rejet de la méthode. Quelle est la valeur de cette limite théorique ?

(f) Comparaison des méthodes. Dans le cas particulier de la simulation de réalisations selon une loi normale centrée réduite de dimension 1 et selon vos observations des différents ratios de rejet, classez les méthodes de simulations précédemment considérées suivant leurs performances.

3. Méthodes particulières de simulation de variables aléatoires

(a) (BONUS) En utilisant la méthode Box Muller, simulez des réalisations d'une loi normale centrée réduite. À partir de cet échantillon simulé, construire le vecteur contenant les quantiles empiriques de cet échantillon. Créez ensuite un vecteur d'une même longueur contenant les quantiles théoriques aux mêmes niveaux d'une loi normale centrée réduite (vous pourrez utiliser la fonction R déjà programmée *qnorm*). Représentez le couple de vecteurs obtenu. Ce graphique porte le nom de *graphique qqplot*.

(b) (BONUS) Simulation selon la loi de Poisson.

1) Soit $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$. Rappelez la valeur de $\mathbb{P}(X = k)$, $\forall k \in \mathbb{N}^+$

2) Soit E_i des réalisations iid de loi $\mathcal{E}(\lambda)$ avec $\lambda > 0$. Soit $S_n \coloneqq \sum_{i=1}^n E_i$.

En admettant le résultat suivant déjà vu en cours et rappelé ci-dessous, proposez un algorithme de simulation d'une variable aléatoire suivant une loi de Poisson.

$$\mathbb{P}(S_n \leq 1 \leq S_{n+1}) = \frac{\lambda^n}{n!} \exp(-\lambda)$$

3) Créez une fonction *poisson* permettant de simuler n réalisations d'une variable aléatoire suivant une loi de Poisson.

4) Créez une fonction *poissoninverse* permettant de simuler selon une loi de poisson grâce à la méthode de transformation inverse.