TP de modélisation stochastique - séance 2

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

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

1. Illustration des théorèmes de convergence asymptotique

(a) La loi des grands nombres

1) Rappelez les hypothèses de la loi des grands nombres.
2) Rappelez le résultat de la loi des grands nombres.
3) Soit $X \sim \mathcal{E}(2)$. Peut-on appliquer la loi des grands nombres à des réalisations *iid* de X ?

Lors de ce TP, nous allons chercher à observer la variabilité d'estimations. Cette connaissance est essentielle pour prendre du recul sur un résultat et le nuancer. Ici, chacune des estimations sera issue d'une méthode de type Monte Carlo. Pour ce faire, nous procéderons de la manière suivante : nous réaliserons un certain nombre d'expériences, chacune permettant d'obtenir une estimation à partir d'un nombre de simulations donné.

4) L'objectif est d'étudier la convergence de l'estimateur de la moyenne. Plus précisément, pour chaque expérience, il s'agit de simuler 10000 réalisations de X grâce à la fonction *rexp*, puis de stocker la suite des moyennes empiriques obtenues après i simulations. Réalisez 10 expériences et représentez sur un même graphique les suites des moyennes empiriques cumulées des 10 expériences. Vous pourrez utiliser la fonction *plot* pour représenter la première trajectoire puis la fonction *lines* pour superposer sur le même graphique les autres trajectoires. Astuce : un chiffre peut être associé à l'argument *col* afin de varier simplement et dans une boucle, les couleurs des trajectoires.

(b) Le théorème centrale limite

1) Rappelez les hypothèses du théorème centrale limite.
2) Rappelez le résultat du théorème centrale limite.
3) Soit X la VA définie dans la question précédente. Peut-on appliquer le théorème centrale limite à des réalisations *iid* de X ?
4) Grâce à la fonction *rexp*, simulez respectivement 100, 1000 et 10000 réalisations de X. Réalisez un premier graphique permettant de visualiser l'adéquation en loi des moyennes empiriques obtenues lors de chaque expérience.
5) À présent, étudions la variabilité de l'estimateur de la moyenne. Pour les trois nombres de simulations précédents, réalisez 100 expériences puis représentez 3 boxplots des moyennes estimées pour 100, 1000 et 10000 simulations de X.

(c) Contre-exemple

1) Soit Y suivant une loi de Cauchy de paramètres 0 (localisation) et 10 (échelle).
2) Répondez à la question a) 4) dans le cas de cette loi et conclure.

2. Réduction de la variance

(a) Méthode Monte Carlo. Écrivez une fonction *MC* prenant en argument une fonction à intégrer f, une borne inférieure a, une borne supérieure b et un nombre de simulations n et retournant l'estimation via Monte Carlo après n simulations de $\int_a^b f(x)\,\mathrm{d}x$.

(b) Variables antithétiques

Les deux premières questions ne sont pas notées.

1) Rappelez le principe de l'utilisation de variables antithétiques dans l'optimisation de la performance d'une estimation de Monte Carlo. Dans quel cadre ce principe peut-il s'appliquer ?

2) Soit f définie par $f(x) = \frac{1}{1+x}$. Calculez (exactement) $\int_0^1 f(x)\,\mathrm{d}x$ (cf TD)

3) Écrivez une fonction *MCantithetique* prenant en argument une fonction f et un nombre de simulations n et retournant l'estimation après n simulations de $\int_0^1 f(x)\,\mathrm{d}x$.

4) Pour un nombre de simulation n fixé, par exemple $n = 100$, réalisez 1000 expériences d'application des fonctions *MC* et *MCantithetique* à l'estimation de $\int_0^1 f(x)\,\mathrm{d}x$. Représentez deux boxplot côte à côte, pour comparer la variabilité des estimations de chacune des méthodes. Commentez sur la performance relative des méthodes.

(c) Variable de contrôle

Les questions 1 et 3 ne sont pas notées.

1) Rappelez le principe de l'utilisation d'une variable de contrôle dans l'optimisation de la performance d'une estimation de Monte Carlo.

2) Écrivez une fonction *MCcontrole* prenant en argument une fonction f à intégrer entre a et b, une fonction g de contrôle, son intégrale entre a et b et un nombre de simulations n et retournant l'estimation après n simulations de $\int_0^1 f(x)\,\mathrm{d}x$.

3) Proposez une variable de contrôle pour optimiser l'estimation de $\int_0^1 f(x)\,\mathrm{d}x$ (cf TD) Nous considérons la fonction de contrôle $g(x) = c(1 + x)$.

4) Pour $c \in [0; 1]$ avec un pas de 0.05, réalisez 100 expériences avec un nombre de simulation égal à 100 en appliquant la fonction *MCcontrole* à l'estimation de $\int_0^1 f(x)\,\mathrm{d}x$. Représentez des boxplots montrant l'évolution de la variabilité de l'estimation en fonction de c. Conjecturez un choix de c optimal.

5) Comparez les méthodes méthodes *MCcontrole*, *MCantithetique* et *MC* en représentant des boxplots montrant les différentes variabilités des estimations. Pour la méthode de contrôle, nous fixons c égal à la valeur optimale conjecturée précédemment.

6) Commentez sur la performance relative des méthodes.

(d) Échantillonage préférentiel

La question 1 n'est pas notée.

1) Rappelez le principe de l'échantillonnage préférentiel dans l'optimisation de la performance d'une estimation de Monte Carlo.

2) Écrivez une fonction *MCpreferentiel* prenant en argument une fonction f à intégrer, une fonction rapport de vraisemblance rd, une fonction $rsimu$ permettant de simuler suivant la loi choisie et un nombre de simulations n et retournant l'estimation après n simulations de $\int_0^1 f(x)\,\mathrm{d}x$.

3) Soit f la fonction définie par $h(x) = 10.e^{-2.|x-5|}$. Appliquez respectivement les fonctions *MC* et *MCpreferentiel* en simulant suivant une loi normale de paramètres 5 et 1 pour estimer $\int_0^{10} h(x)\,\mathrm{d}x$. Pour comparer les résultats obtenus, nous prendrons $n = 100$ et nous réaliserons 1000 expériences.

4) Pour $\sigma \in [1; 5]$ avec un pas de 0.1, appliquez la fonction *MCpreferentiel* en simulant suivant une loi normale de paramètres 5 et σ pour estimer $\int_0^{10} h(x)\,\mathrm{d}x$. Représentez un graphique montrant l'évolution de l'estimation en fonction de σ. Vous pourrez superposer sur le graphique la valeur théorique de l'intégrale. Représentez de plus un graphique mettant en évidence la réduction de la variance en fonction de σ. Conjecturez un choix de σ optimal.