IFT-3655: Modèles Stochastiques **Hiver 2024**

Prof. Pierre L'Ecuyer

DEVOIR 6

Devoir à remettre le *lundi 11 novembre 2024, avant 10h30*.

Les devoirs doivent être faits individuellement: un devoir par étudiant. Rendez votre devoir dans Studium sous forme d'un fichier .pdf. Il est très important de **bien expliquer** tout ce que vous faites. Dans la correction, on accordera davantage d'importance à la clarté des explications qu'à la réponse comme telle. Attention au plagiat: il est interdit de copier et/ou modifier des solutions venant de quelqu'un d'autre, ou prises sur Internet ou dans des livres.

Pour la question 3, si vous voulez faire vos simulations an Java, il y a la librairie SSJ disponible ici: ssj-github. Avant de l'utiliser, regardez bien la section "Documentation and tutorial". Ceci dit, vous pouvez utiliser le logiciel que vous voulez.

---

1. *(10 points)*

Dans l'exemple de la page 42 des diapos sur la simulation, on considère une méthode de rejet pour une loi beta en utilisant une fonction chapeau h constante par morceaux (en bleu). Cette fonction chapeau est proportionnelle à une densité g, qui a une cdf correspondante G.

(a) Trouvez une formule pour G^{-1} et un algorithme pour générer X par inversion selon la densité g.

(b) Donnez un algorithme qui utilise cela pour générer X selon la densité f par la méthode de rejet. Pas besoin de l'implanter.

2. *(20 points)*

Vous voulez générer n réalisations indépendantes d'une variable aléatoire qui suit la loi de Poisson de moyenne λ, par inversion. Vous pouvez implanter ce qui suit en Java ou Python ou C++.

(a) Pour générer les n variables aléatoires indépendantes $\mathcal{U}(0,1)$ dont vous avez besoin, vous allez implanter une fonction qui utilise le générateur décrit à la page 18 des diapos et retourne la prochaine valeur u_n dans la séquence. Vous pouvez l'implanter directement en utilisant des entiers de 64 bits. Comme état initial, prenez simplement 12345 pour chacun des 6 entiers $x_{-1}, x_{-2}, x_{-3}, y_{-1}, y_{-2}, y_{-3}$. La première valeur générée sera u_0. Pas besoin de mémoriser n, mais il faut bien sûr mémoriser l'état du générateur.

(b) Écrivez aussi une fonction qui prend λ et U (des `double`) en entrée et qui retourne $X = F_\lambda^{-1}(U)$ où F_λ est la cdf de la loi de Poisson de moyenne λ. Cette fonction doit fonctionner pour n'importe quel $\lambda > 0$ et $0 < U < 1$. Elle peut utiliser la méthode de recherche séquentielle de la diapo 32. L'avantage d'une telle fonction est que λ peut varier d'un appel à l'autre.

(c) Écrivez ensuite une fonction qui implante la méthode d'inversion qui utilise un index, décrite à la diapo 33, disons pour $\lambda = 25$, avec $c = 32$ intervalles. Vous devez d'abord précalculer et mémoriser dans des tableaux les valeurs de $F_{25}(x)$ pour $x = 0, 1, 2, 3, \ldots$ et de

i_s pour $s = 0, \ldots, c-1$. Pour simplifier, vous pouvez supposer que $F_{25}(50) = 1$. Ensuite écrire une fonction qui prend seulement U en entrée et utilise ces tableaux efficacement pour retourner $X = F_{25}^{-1}(U)$.

(d) Faites maintenant une expérience pour comparer les vitesses des fonctions en (b) et (c) pour générer $n = 10^6$ variables de Poisson avec $\lambda = 25$. Démarrez votre générateur dans le même état initial pour les deux cas, de manière à obtenir exactement la même séquence de valeurs dans les deux cas. Pour chacun des deux cas, calculez le temps total d'exécution ainsi que le nombre de fois que vous avez obtenu chacune des valeurs de X de 0 à 50. Assurez-vous que les fréquences observées correspondent à peu près aux probabilités de la loi de Poisson. Discutez vos résultats.

3. *(20 points)*

Pour la question 4 du devoir 5, implantez la méthode de Gibbs avec *balayage systématique* pour le graphe donné en (c), pour $k = 4$. Faites $n = 216,000$ itérations de la méthode et comptez le nombre de fois que chaque k-coloriage de $\mathcal{C}$ est visité. Vous avez vu au devoir 5 qu'il y a 216 coloriages possibles avec $k = 4$, et si la chaine est irréductible, sa loi stationnaire devrait être uniforme sur ces 216 coloriages, donc chacun devrait apparaître environ 1000 fois. Est-ce que vos résultats sont en accord avec cela?

Pour tester si les observations sont en accord avec cette hypothèse que tous les coloriages admissibles ont la même probabilité, on peut faire un test de chi-deux sur les résultats. Calculez le chi-deux, puis la p-valeur du test. Discutez vos résultats.

Rappel: S'il y a k possibilités, et si O_j est le nombre observé d'observations et e_j le nombre espéré d'observations pour la possibilité j, pour $j = 1, \ldots, 216$, alors le chi-deux est défini par

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - e_j)^2}{e_j}.$$

Sous l'hypothèse donnée ci-haut, et en supposant aussi que les observations sont indépendantes, cette variable aléatoire suit approximativement la loi de chi-deux avec $k-1$ degrés de liberté. Voir `https://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test`, par exemple. Si on observe $\chi^2 = x$, alors la p-valeur du test est $p = \mathbb{P}[X > x]$ où X est une variable aléatoire qui suit la loi du chi-deux à $k-1$ degrés de liberté. Si l'hypothèse est vraie, ce p devrait suivre à peu près la loi uniforme sur $(0, 1)$. Lorsque p est trop petit, on peut mettre en doûte l'hypothèse. Dans ce cas-ci, on a $k = 216$ et $e_j = n/k = 1000$ pour tout j.

Formellement, le test du chi-deux est valide sous l'hypothèse que les n observations sont indépendantes. Ici, elles ne le sont pas, car ce sont les états successifs visités par une chaine de Markov, mais la loi du chi-deux est quand même une bonne approximation quand n est grand, et le test peut détecter si la loi dévie beaucoup de l'uniformité, par exemple.