MATH0062-1 - Éléments du Calcul des Probabilités
Projet 1

Généralités

Le premier travail portera sur le problème des anniversaires, connu aussi sous le nom de *paradoxe des anniversaires*, qui fut formulé pour la première fois par l'ingénieur autrichien Richard von Mises. Ce travail a pour but de familiariser les étudiants avec la notion d'arbre de probabilités et avec la méthode de Monte Carlo. Cette dernière, inventée dans les années 1940 par John von Neumann, Stanislaw Ulam et Nicholas Metropolis, permet d'évaluer une intégrale, en particulier l'espérance d'une variable aléatoire, là où une approche analytique ou déterministe serait difficile, voire impossible.

Ce travail devra être réalisé individuellement. Chaque étudiant devra rendre d'une part un rapport dont le nombre de pages conseillé (hormis annexes) est de 4 et d'autre part ses codes sources MATLAB. Ce rapport devra respecter les indications du guideline `http://www.montefiore.ulg.ac.be/~lduchesne/proba/guidelines.pdf`.
Toutes vos questions pour ce projet doivent être envoyées à l.duchesne@ulg.ac.be.

Le travail devra être rendu sous forme d'archive `.zip` au plus tard pour le lundi 27/03/2017 23h59 via la plateforme `http://submit.montefiore.ulg.ac.be/`.

Notez qu'au delà de la deadline, il ne sera plus possible de rendre le projet.

Présentation du problème

Le problème des anniversaires consiste à déterminer la probabilité qu'au moins deux personnes, dans un ensemble de N, aient leur anniversaire le même jour. On supposera, pour des raisons de simplicité, que toutes les années sont non-bissextiles et donc que personne ne soit né le 29 février. Le travail se décompose en deux parties :

1. Recherche d'une solution exacte : l'objectif de cette partie est de démontrer votre compréhension de la méthode de Lehman & Leighton et votre capacité à l'adapter à un problème particulier.
2. Recherche d'une solution approchée par simulation : l'objectif de cette partie est d'utiliser la méthode de Monte Carlo et d'en comprendre les avantages et limitations.

Questions

1. **Méthode de Lehman & Leighton**
   (a) Utiliser la fonction `naiveBirthdaySol` du fichier `naiveBirthdaySol.m` pour calculer la probabilité que deux personnes aient leur anniversaire le même jour. La réponse vous semble-t-elle correcte ? Utilisez les fonctions `tic` et `toc` pour déterminer le temps que prend MATLAB pour effectuer ce calcul sur votre machine.

   (b) Que se passe-t-il lorsque vous essayez d'utiliser `naiveBirthdaySol` pour calculer la probabilité qu'au moins deux personnes parmi trois aient leur anniversaire le même jour ? En faisant l'hypothèse que le temps d'exécution est proportionnel au nombre de branches de l'arbre naïf construit par la fonction `createBirthdayTree` du fichier `createBirthdayTree.m`, estimez le temps nécessaire à la résolution des cas $N = 3$, $N = 4$ et $N = 5$. Que pouvez-vous en déduire quant à l'utilisation pratique de cette approche ?

   (c) Imaginez une simplification de l'arbre permettant de calculer efficacement la probabilité qu'au moins deux personnes dans un ensemble de N aient leur anniversaire le même jour. Détaillez le raisonnement qui vous a conduit à construire l'arbre simplifié et donnez un exemple, sous forme de schéma, de cet arbre pour $N = 3$.

   (d) Implémentez dans Matlab une fonction se basant sur votre arbre simplifié. Déterminez la probabilité recherchée pour $N = 2, 3, 4, 5, 20, 30, 40, 50, 60, 80$. Mesurez le temps nécessaire à l'exécution de votre fonction dans chacun de ces cas et comparez-le le cas échéant aux temps estimés à la question 1b. Comment pouvez-vous expliquer intuitivement les probabilités obtenues ?

2. **Méthode de Monte Carlo**
   (a) Soit la variable aléatoire binaire $\mathcal{X}$ = *1 si au moins deux personnes parmi quarante ont leur anniversaire le même jour, 0 sinon.* Quelle loi de probabilité cette variable aléatoire suit-elle ? Calculez l'espérance et la variance de cette variable aléatoire.

   (b) Estimez l'espérance et la variance de $\mathcal{X}$ à l'aide de la méthode de Monte Carlo. Pour ce faire, générez des vecteurs binaires dont les éléments sont obtenus en utilisant la fonction `birthday40` du fichier `birthday40.m` et dont la taille est de respectivement 10, 10^2, 10^3 et 10^4. Les moyennes (variances) respectives de ces vecteurs servent dans les quatre cas d'*estimateur* de l'espérance (la variance) demandée. Comparez vos résultats avec les valeurs théoriques calculées précédemment. Qu'observez-vous lorsque la taille des vecteurs augmente ?

   (c) Remarquez que réaliser le point (b) peut conduire à des résultats fort différents d'une fois à l'autre. Pour éviter ce problème, répétez désormais 1000 fois l'expérience décrite au point (b) [1]. Calculez dans les quatre cas

      i. la moyenne et la variance des 1000 estimateurs de l'espérance ;

      ii. la moyenne et la variance des 1000 estimateurs de la variance.

   Présentez vos résultats sous forme d'un tableau. Qu'observez-vous lorsque la taille des vecteurs augmente ? Serait-il intéressant de réaliser l'expérience avec une taille de vecteur de 10^5 ?

1. Utilisez à nouveau des vecteurs dont la taille est respectivement de 10, 10^2, 10^3 et 10^4

(d) Une personne fête son anniversaire à la maison et invite tous ses amis pour une certaine heure. Supposons que le retard d'un ami (exprimé en minutes) soit une variable aléatoire normale $\mathcal{N}(5, 5)$. Remarquez que ce nombre peut être négatif [2]. Soit la variable aléatoire $\mathcal{Y}$ = *rapport entre les retards de deux amis arrivant indépendamment les uns des autres*. Répétez 1000 fois l'expérience suivante : générer des vecteurs dont les éléments suivent la loi de probabilité étudiée et dont la taille est de respectivement 10^2, 10^3, 10^4 et 10^5. Calculez dans les quatre cas la moyenne et la variance des 1000 estimateurs de l'espérance de $\mathcal{Y}$. Présentez vos résultats sous forme d'un tableau. Que constatez-vous ? Comment expliquez-vous ce phénomène ?

Cauchy !

(pas utiliser formule toute faite wikipedia mais créer arbre)

2. ce qui correspond à un ami arrivant en avance.