MÉTHODES DE MONTE-CARLO EXERCICES DU 24 FÉVRIER 2014

par Rémi Peyre

EXERCICE 1 — La grande suite

Un joueur de yahtzee, à son dernier coup, cherche à obtenir la grande suite. Pour ceux qui ne connaissent pas le jeu du yahtzee, rappelons que ce jeu se joue avec 5 dés que le joueur a le droit de lancer jusqu'à trois fois : plus précisément, il lance une fois les cinq dés, puis relance les dés de son choix (éventuellement tous au aucun), puis relance à nouveau les dés de son choix (y compris, éventuellement, des dés qu'il n'avait pas encore relancés). La « grande suite » consiste à ce que, à l'issue des trois lancers, les dés affichent exactement une fois chaque chiffre de '2' à '6'.

Nous admettons que la meilleure stratégie possible pour le joueur est la stratégie naturelle consistant à garder chaque chiffre entre '2' et '6' qui apparaît, mais au plus un dé de chaque chiffre. Notre question consiste à savoir quelle est la probabilité que, en suivant cette stratégie, le joueur réussisse à obtenir une grande suite.

1. Conditionnellement aux deux premiers lancers et à leurs résultats, quelle est la probabilité que le joueur réussisse ?

2. En déduire un algorithme évaluant la probabilité de succès du joueur à l'aide d'une méthode de conditionnement.

3. Comparer à l'algorithme sans conditionnement ; évaluer le gain d'efficacité.

4. (★) Imaginer et implémenter une méthode d'échantillonnage préférentiel dans laquelle un des dés a une plus grande probabilité de tomber sur '2', un autre sur '3', ..., et le dernier sur '6'.

Indication : Implémentée trop naïvement, cette méthode va conduire à un problème de sous-échantillonnage : voyez-vous pourquoi ? Pour contourner le sous-échantillonnage, on choisit de ne considérer comme « succès » que l'événement dans lequel chaque dé finit par tomber sur le chiffre qu'il favorisait. Quel est le lien entre la probabilité de cet événement et la probabilité qui nous intéresse ?

Indication : Notez qu'en l'absence d'échantillonnage préférentiel, les cinq dés jouaient des rôles symétriques, et que donc le choix du joueur de garder tel ou tel dé quand un chiffre sortait plusieurs fois importait peu. Sous l'échantillonnage préférentiel en revanche, les dés ne jouent plus le même rôle... Y a-t-il alors une stratégie particulière que doit suivre le joueur de notre simulation ; et si oui, pourquoi ?

EXERCICE 2 — Une question de chance

On considère à nouveau le jeu du yahtzee. La « chance » consiste tout simplement à marquer autant de points que les dés affichent à la fin des trois lancers.

Nous admettons que la meilleure stratégie possible pour le joueur consiste à ne garder, à l'issue du premier lancer, que les dés qui affichent 5 ou plus ; et au second lancer, que les dés qui affichent 4 ou plus. La question est de savoir l'espérance du nombre de points marqués.

1. Implémenter un programme utilisant une méthode de variables antithétiques pour résoudre ce problème ; comparer l'efficacité à celle de la méthode de Monte-Carlo "naïve".

2. (★) Plutôt que les variables antithétiques, proposer une méthode de conditionnement aussi judicieuse que possible.

Indication : Si vous traitez bien cette question, vous devriez pouvoir en déduire que celui qui l'a posée est un petit comique...

EXERCICE 3 — Les imperfections du TLC

On considère la loi $\chi^2(8)$, qui peut s'obtenir comme la somme de 4 lois exponentielles $\text{Expon}(1/2)$ indépendantes. On note par X une variable aléatoire suivant cette loi.

1. (♣) Déterminer l'espérance μ et l'écart-type σ de la loi de X .

2. À l'aide de la fonction `erf` ou `erfinv` de MATLAB, écrire une routine qui simule deux variables aléatoires Y et Z telles que :

- Y suit une loi exponentielle ;
- Z suit une loi normale centrée réduite ;
- Y et Z sont des fonctions déterministes (strictement) croissantes l'une de l'autre.

3. Écrire une méthode de Monte-Carlo qui évalue la probabilité que X soit plus grande que $\mu + \lambda\sigma$, où λ sera un paramètre au choix de l'opérateur. Cette méthode devra reposer sur un couplage entre X et une variable gaussienne $\mathcal{N}(\mu, \sigma^2)$. Comparer à la méthode de Monte-Carlo sans couplage.

Indication : Commencer par remarquer que si X était gaussienne, on saurait calculer exactement les probabilités en question, à l'aide de la fonction `erf`.

EXERCICE 4 — Le coup de la panne

Pour augmenter la durée de vie d'un système électronique embarqué, les ingénieurs qui l'ont conçu ont décidé de tripler le circuit : c'est-à-dire qu'en plus du circuit principal (que nous appellerons n° 1), il y a deux circuits de secours (n° 2 et n° 3), le système fonctionnant tant qu'au moins un des trois circuits est en service. On suppose qu'en l'absence de mauvais coup, les durées de vie des différents circuits sont indépendantes et exponentielles, le circuit 1 ayant une durée de vie typique de 10 ans (càd. de loi $\text{Expon}(10^{-1})$), et les circuits 2 et 3 resp. de 5 et de 4 ans. Indépendamment de ces problèmes d'usure, le système peut également prendre un "mauvais coup", suivant une loi $\text{Expon}(20^{-1})$, qui met d'un seul coup les trois circuits hors service. On appelle T la durée de vie globale de la machine (incluant la possibilité d'un mauvais coup) ; on s'intéresse à l'évaluation de $\mathbb{E}(T)$.

1. (♣) Rappeler comment on simule une loi exponentielle. Écrire une routine simulant T et l'intégrer à une méthode de Monte-Carlo pour le calcul de $\mathbb{E}(T)$.

2. (♣) Si X et Y sont deux v.a. indépendantes de lois respectives $\text{Expon}(\lambda)$ et $\text{Expon}(\mu)$, montrer que $\min\{X, Y\}$ suit la loi $\text{Expon}(\lambda + \mu)$.

3. On appelle T_1, T_2, T_3 les durées de vie respective des circuits 1, 2 et 3. Calculer exactement $\mathbb{E}(T_1), \mathbb{E}(T_2), \mathbb{E}(T_3)$.

4. Utiliser la variable de contrôle T_1 (version rudimentaire) pour améliorer l'efficacité du calcul de $\mathbb{E}(T)$. Implémenter et commenter.

5. Même question avec la version paramétrée de la variable de contrôle.

6. Même question en utilisant les trois variables de contrôle T_1, T_2, T_3 .

EXERCICE 5 — Que d'eau, que d'eau... !

Un agronome souhaite connaître la quantité de pluie moyenne qui tombe à Nancy en un an. Il se rend donc au centre local de Météo France ; mais c'était à l'époque des débuts de la météo : on avait certes déjà enregistré 1 siècle de météo, mais on n'avait encore aucun traitement informatisé... Tout ce que Météo France savait, c'est que sur 36 524 jours il y avait eu 12 449 jours de pluie. Notre agronome n'ayant pas de temps d'éplucher trente-six-mille fiches, il décide d'en piocher mille au hasard pour se faire une idée, et d'extrapoler à partir de celles-ci.

1. (✿) Formaliser le problème ; et écrire l'estimateur correspondant à la démarche de l'agronome, ainsi que l'intervalle de confiance associé à cet estimateur. (On supposera que le nombre de fiches tirées est suffisamment petit en proportion pour qu'on puisse faire comme s'il y avait indépendance).

2. Quelle technique de réduction de variance l'agronome peut-il utiliser ? Écrire formellement l'estimateur et l'intervalle correspondants.

3. Supposons maintenant que les fiches de Météo France sont rangées dans deux bacs : « jours de pluie » et « jours sans pluie ». Quelle technique de réduction de variance cela suggère-t-il ? Décrire la méthode correspondante, et écrire formellement l'estimateur et l'intervalle de confiance.

4. Faire une simulation des différentes expériences et donner les valeurs numériques associées. (On supposera pour l'application que, les jours où il pleut, la quantité Q de pluie tombée, exprimée en mm, suit une loi telle que $\ln Q \sim 1,33 + \mathcal{N}(1)$, Q étant ensuite arrondi à l'entier le plus proche).

EXERCICE 6 — Téléphonomanie

Un adolescent téléphonomane passe, en moyenne, 32 coups de fil par mois sur son portable. Le nombre de coups de fil qu'il passe dans le mois est modélisé par une loi de Poisson. La durée moyenne de chaque coup de fil est de 8 minutes ; elle est modélisée par une loi exponentielle, chaque appel ayant une durée indépendante des autres. L'adolescent hésite entre deux forfaits pour son portable : un forfait "4 h" qui lui offre 4 h d'appels pour 16 €, puis 17 c€/min au-delà, et un forfait "5 h" qui lui offre 5 h d'appels pour 20 €, avec toujours 17 c€/min au-delà.

1. Écrire un programme évaluant par la méthode de Monte-Carlo le cout moyen par mois pour notre adolescent avec le forfait "4 h", resp. avec le forfait "5 h".

Indication : On rappelle que la loi de Poisson de paramètre λ est de nombre de variables indépendantes de loi $Expon(1)$ qu'on peut sommer avant de dépasser λ : si $X_1, X_2, \dots$ sont des v.a. indépendantes $Expon(1)$,

$$\max \left\{ n \in \mathbb{N} : \sum_{i=1}^n X_i \leq \lambda \right\}$$

suit une loi $Poisson(\lambda)$.

2. Quel est le forfait le plus avantageux pour l'adolescent ? On utilisera une technique de couplage pour répondre à cette question.

3. Évaluer le gain d'efficacité apporté par la technique de couplage.

Enfin, l'opérateur de téléphonie mobile a retiré le forfait "5 h" de son offre, ce qui règle la question du choix du forfait... Mais maintenant, la grande sœur de notre adolescent lui conseille de téléphoner un peu plus souvent mais un peu moins longtemps, disons 37 fois par mois pour une durée moyenne de 7 minutes par appel.

4. Le conseil de la grande sœur est-il judicieux ? On utilisera une technique

de couplage pour répondre à cette question. Évaluer ici aussi le gain d'efficacité apporté par la technique de couplage.