École des Mines de Nancy, année 2012–13

Méthodes probabilistes pour la simulation [SG241]
Examen final

lundi 25 mars 2013

ÉNONCÉ

PROBLÈME 1 — Dudo

Le but de cet exercice est de répondre à la question : « Quand on lance 25 dés, quelle est la probabilité, notée p, qu'un des chiffres au moins apparaisse 7 fois ou plus ? ». [1] *(Pour ceux qui ne souhaitent pas utiliser de paramètres explicites dans leurs formules, je suggère de noter* $25 =: Z$ *le nombre de dés,* $6 =: F$ *le nombre de faces d'un dé, et* $7 =: D$ *le nombre de dés identiques qui intervient dans la définition de p).*

5 pt **1.** Écrire une fonction appelée `Pb1_simulation` qui simule le lancer de 25 dés, et renvoie 1 si l'un des chiffres apparait 7 fois ou plus, et 0 sinon.
Indication : Ceux qui travaillent sous MATLAB pourront regarder avec profit la documentation des commandes suivantes : `eq`, `max`, `randi`, `sum`.

3 pt **2.** À l'aide de la fonction `Pb1_simulation` écrite à la question précédente, écrire un programme appelé `Pb1_MC_p` pour évaluer la probabilité recherchée à l'aide de la méthode de Monte-Carlo, en précisant l'intervalle de confiance.

Dans la suite du problème, on appelle chiffre record *le chiffre qui apparait en plus grand nombre au tirage des 25 dés, avec la règle suivante : quand le plus grand nombre d'occurrences est partagé par plusieurs chiffres, on tire au hasard* uniformément *parmi ces chiffres l'un d'eux qui sera décrété être le chiffre record. On note q la probabilité de l'évènement « Le chiffre record est '1', et celui-ci apparait au moins 7 fois ».*

4 pt **3.** Exprimer p en fonction de q, en justifiant le résultat. Si l'intervalle de confiance pour q (à une certaine précision) est $\hat{q}\pm r$, quel est l'intervalle de confiance correspondant pour p à la même précision ?

Dans la suite du problème, nous n'allons pas chercher à évaluer directement p, mais plutôt q, étant donné que d'après la question précédente cela donne du même coup un estimateur pour p.

6 pt **4.** Écrire une méthode de Monte-Carlo "naïve" pour évaluer q. (le programme sera appelé `Pb1_MC_q`).

5 pt **5.** À quelle technique de réduction de variance présentée dans le cours vous fait penser le lien entre p et q ? Pourquoi l'idée d'évaluer q plutôt que p est-elle *à priori* tout-à-fait stupide ? À votre avis, pourquoi avons-nous retenu cette idée malgré tout ?

Pour un lancer des 25 dés, on note A le nombre de '1' qui apparaissent.

4 pt **6.** Calculer exactement $\mathbb{E}(A)$, $\mathbb{E}(A^2)$, et [facultatif] $\mathbb{E}(A^3)$.

6 pt **7.** Écrire une méthode de Monte-Carlo utilisant les variables de contrôle A et A^2 (et éventuellement A^3) pour évaluer q. (le programme sera appelé `Pb1_MC_q_ctrl`).

---

[1]. La motivation pour calculer p est librement inspirée d'un jeu de dés chilien appelé *dudo* — mais il est inutile de connaitre ce jeu pour aborder le problème.

On définit maintenant la probabilité Q d'un tirage de 25 dés où les dés sont pipés : pour chaque dé, le chiffre '1' a 25 % de chances de sortir, tandis que les autres chiffres n'ont que 15 % de chances chacun. (Ceux qui ne souhaitent pas écrire les paramètres pourront poser $0{,}15 =: (1-\varepsilon)/6$*, resp.* $0{,}25 =: (1+5\varepsilon)/6$*).*

3 pt **8.** Pour un tirage $\omega = (\omega_1, \ldots, \omega_{25})$ des dés (les ω_i étant les chiffres affichés par les différents dés), exprimer la densité relative $\big(dP/dQ\big)(\omega)$ — où P désigne la loi du tirage de dés non pipés.

4 pt **9.** En déduire une technique d'échantillonnage préférentiel pour évaluer q, et l'implémenter. (le programme sera appelé `Pb1_MC_q_pref`).

PROBLÈME 2 — Croissance de fissures

Un ingénieur s'intéresse au risque d'éclatement d'un tube de générateur de vapeur dans une centrale nucléaire. Nous faisons ici la modélisation grossière suivante : il y a dans le tube une unique fissure dont la taille initiale est nulle ; et au cours du fonctionnement du générateur de vapeur, la fissure croît selon le processus markovien suivant :

- *La longueur de la fissure croît par des sauts brutaux, entre lesquels elle reste constante ;*
- *Les sauts se produisent de façon poissonnienne, c'est-à-dire qu'à tout instant, indépendamment du passé, il y a la même probabilité qu'un saut se produise, avec un taux de* $4{,}5 \cdot 10^{-4}$ *par heure de fonctionnement (taux noté* τ*) que ce saut se produise ;*
- *Quand un saut se produit, l'amplitude de ce saut (c'est-à-dire l'accroissement de la longueur de la fissure) suit une loi exponentielle de longueur catactéristique* $0{,}4$ *mm* $=: \lambda$*, c'est-à-dire que la loi de ce saut* X *est telle que* $\mathbb{P}(X \geqslant x) = e^{-x/\lambda}$.

Un tube est destiné à être utilisé pendant un cycle de fonctionnement de durée $10\ 000$ *h* $=: T$*, et éclate si la longueur de la fissure dépasse* 10 *mm* $=: L^*$*. On cherche à évaluer la probabilité* p *qu'un tube éclate avant la fin du cycle.*

5 pt **10.** Écrire une fonction (qu'on appellera `Pb2_evolution`) qui simule l'évolution de la longueur $l(t)$ de la fissure au cours du temps sur la durée T, plotte la courbe correspondante, et renvoie la longueur finale de la fissure (éventuellement supérieure à L^*, auquel cas le tube aura en réalité éclaté avant).
Indication : On pourra utiliser la fonction `stairs` de MATLAB pour plotter une fonction évoluant par sauts.

4 pt **11.** Pourquoi sera-t-il pratiquement indispensable d'utiliser une technique de réduction de variance pour ce problème ?

5 pt **12.** Quelle est la loi du nombre de fois que la fissure s'allonge au cours du cycle de fonctionnement ? Coder une fonction appelée `Pb2_probanbsauts` qui calcule *exactement*, pour $n \geqslant 1$, la probabilité que la fissure subisse au moins n allongements au cours du cycle.
Indication : Histoire que cette question ne risque pas de faire "barrage", je précise que la loi du nombre de fois que la fissure s'allonge appartient à une famille de lois classiques, ce qui vous devrait vous permettre de l'identifier facilement même si vous ne parvenez pas à le justifier.

3 pt **13.** Supposons qu'on simule l'allongement final de la fissure ainsi : *d'abord*, on simule la liste (infinie) $\delta_1, \delta_2, \ldots$ des longueurs des allongements successifs de la fissure — sachant que tous ces accroissements n'auront pas forcément lieu effectivement — ;

et *ensuite*, on simule le nombre d'allongements qui se sont produits au cours du cycle de fonctionnement pour en déduire la longueur totale atteinte. Conditionnellement à la première partie de la simulation, quelle est la probabilité que le tube éclate ?

5 pt **14.** En déduire une technique de conditionnement pour évaluer la probabilité d'éclatement d'un tube, et implémenter celle-ci (le programme s'appellera `Pb2_MC_cond`). Pourquoi n'est-ce en fait pas un problème que, dans la première partie de simulation, on doive théoriquement simuler une suite infinie de valeurs (ce qui serait évidemment impossible) ? [Facultatif] À quelle technique de réduction de variance peut se rattacher l'argument qui dit qu'on n'a en fait pas à simuler une suite infinie de valeurs ?
Indication : Dans la mesure où la fonction `Pb2_pbanbsauts` sera appelée très régulièrement par le programme `Pb2_MC_cond`, on pourra gagner un temps considérable en tabulant à l'avance les valeurs de retour les plus fréquentes de `Pb2_pbanbsauts`.

Un industriel propose à EDF d'utiliser un acier plus cher mais un peu plus résistant pour fabriquer les tubes, acier pour lequel le taux des sauts est de $\tau' = 4 \cdot 10^{-4}$ par heure de fonctionnement, avec une longueur caractéristique $\lambda' = 0{,}38$ mm pour les sauts (toujours supposés exponentiels). Les tubes éclatent toujours pour une longueur de fissure de 10 mm. EDF cherche à savoir si ce nouvel acier vaut le coup en évaluant avec précision la diminution de probabilité d'éclatement qu'il induit.

4 pt **15.** Toujours en utilisant la technique de conditionnement de la question précédente, utiliser la méthode des simulations communes pour évaluer la différence entre les probabilités d'éclatement selon qu'on utilise le premier ou le second acier (le programme s'appellera `Pb2_MC_cond_CRN`).

PROBLÈME 3 — Intervalles de confiance dans le cas non L^2

Soit X une variable aléatoire réelle non L^2, mais qu'on sait avoir un moment d'ordre $3/2$, i.e. $\mathbb{E}(X^{3/2}) < \infty$. Le but de cet exercice est de voir, à partir de simulations i.i.d. $X_1, X_2, \ldots$ de X, comment on peut obtenir rigoureusement un intervalle de confiance (asymptotique) pour $\mathbb{E}(X)$ dans un tel cadre.

3 pt **16.** On pose $\tilde{X} := X - \mathbb{E}(X)$ et $\sigma := \mathbb{E}\big(|\tilde{X}|^{3/2}\big)^{2/3}$. Démontrer que $\sigma < \infty$.

5 pt **17.** Montrer que $\sigma^{3/2} \leqslant \mathbb{E}\big(|X_1 - X_2|^{3/2}\big)$. En déduire un estimateur [ici, « estimateur » signifie simplement « quantité dépendant déterministiquement du tirage des X_i »] $\hat{\tau}$ (dépendant du nombre N de simulations effectuées) qui est *asymptotiquement majorant* de σ, càd. que pour tout $\varepsilon > 0$, $\mathbb{P}(\hat{\tau} \leqslant \sigma - \varepsilon) \overset{N\to\infty}{\to} 0$.
Indication : On pourra observer que $X_1 - X_2 = \tilde{X}_1 - \tilde{X}_2$ (avec des notations évidentes), et utiliser alors que $\mathbb{E}(|\tilde{X}_1|^{3/2}) = \sigma^{3/2}$ d'une part et que $\tilde{X}_2$ est centrée d'autre part.

3 pt **18.** Il y avait aussi un estimateur $\hat{\sigma}$ tout simple qui, pour le coup, convergeait *exactement* vers σ... Écrire cet estimateur. Quel est l'intérêt que $\hat{\tau}$ par rapport à $\hat{\sigma}$ qui justifie que nous ayons préféré considérer celui-ci ?

7 pt **19.** Soit $y \in \mathbb{R}$ et Z une variable aléatoire de classe $L^{3/2}$ *centrée*. Montrer qu'il existe un réel $\gamma < \infty$ ne dépendant pas de y ni de Z, tel que $\mathbb{E}(|y+Z|^{3/2}) \leqslant |y|^{3/2} + \gamma\, \mathbb{E}(|Z|^{3/2})$. Dans la suite du problème, nous admettrons que la valeur $\gamma = 1{,}31$ convient.
Indication : On pourra introduire la variable aléatoire $|y + Z|^{3/2} - aZ$, où a est un réel judicieusement choisi.
Indication : On pourra éventuellement remarquer qu'il suffit de démontrer cela pour $y = 1$. (pourquoi ?).

3 pt **20.** Déduire de la question précédente que, notant $N^{-1} \sum_{i=1}^{N} X_i =: \hat{m}_N$, on a

$\mathbb{E}\big(|\hat{m}_N - \mathbb{E}(X)|^{3/2}\big) \leqslant \gamma\sigma^{3/2}N^{-1/2}$.

4 pt **21.** En déduire un intervalle de confiance à 95 % (asymptotiquement s'entend) pour $\mathbb{E}(X)$. Commenter la façon dont la largeur de cet intervalle de confiance décroit. *Indication :* Penser à l'inégalité de Markov.

5 pt **22.** Application : évaluer $\mathbb{E}(\sin X \times X)$ avec un intervalle de confiance à 95 % pour X suivant la loi de Pareto de paramètre 1,8, càd. que X est à valeurs dans $[1, +\infty)$ avec

$$\mathbb{P}(X \geqslant x) = x^{-1,8} \qquad \forall x \geqslant 1.$$