École des Mines de Nancy, année 2013 –14

# Méthode de Monte-Carlo & Application aux processus aléatoires [1]
# Examen final

Saint-Anicet [2] 2014, 14 h 00 – 17 h 00

## PROBLÈME 1 — Provision du casino

*Ce problème concerne un casino qui se demande quel montant il doit provisionner dans ses coffres pour une soirée. On modélise ainsi le déroulement d'une soirée dans le casino : au début de la soirée, le casino a préparé dans ses coffres un certain montant en euros, appelé* provision. *Au cours de la soirée, de nombreuses parties de jeux de hasard sont jouées. On supposera ici que le seul jeu de ce casino est la* roulette *: dans une partie de roulette, le joueur mise une certaine somme d'argent et parie sur un numéro ; avec une chance sur 37, le numéro parié par le joueur sort et celui-ci récupère alors 36 fois sa mise ; dans le cas contraire, le casino garde le mise du joueur et ne lui rend rien. (C'est le déséquilibre entre le nombre 37 de paris possibles et le facteur 36 du gain qui permet au casino de faire un bénéfice sur le long terme).*

*On suppose pour simplifier que les parties sont jouées de façon poissonnienne, autrement dit qu'à un instant donné le fait qu'une partie soit jouée ou pas (ainsi que le montant misé et le résultat de la partie) est totalement indépendant de ce qui se passe aux autres instants. Nous supposerons qu'il se joue en moyenne 10 parties par minute au cours de la soirée, la soirée durant 8 heures. Lorsqu'une partie est jouée, dans 50 % des cas la mise est de 5 €, qui correspond à la mise minimale autorisée ; dans 5 % des cas la mise est de 100 €, qui correspond à la mise maximale autorisée ; et dans les 45 % de cas restants, le montant misé $m$ (exprimé en euros) suit une loi telle que* $\log m$ *soit uniforme sur l'intervalle* $[\log 5, \log 100]$.

<u>Indication :</u> *Il pourra être judicieux d'introduire des notations littérales pour manipuler plus aisément les différentes quantités numériques introduites dans l'énoncé. Je propose que la probabilité de gain à la roulette soit notée* `pgain` *; que le facteur de gain en cas de succès (càd. le ratio entre le montant récupéré et le montant misé) soit noté* `fgain` *; que le nombre moyen de parties par unité de temps soit noté* `lambda` *; que la durée totale de la soirée soit notée* `Tsoiree` *; et que les montants minimal et maximal pour une mise soient notés respectivement* `misemin` *et* `misemax` *et que les probabilités qu'un joueur joue chacune de ces mises soient notées respectivement* `pmmin` *et* `pmmax`. *Je suggère également que les unités de base choisies soient l'euro pour les sommes d'argent, et la minute pour les durées.*

$\sim$ 5 pt **1.** Écrire une fonction `mise` qui simule la loi de la mise d'un joueur. Cette fonction ne prendra aucun argument et renverra la somme misée.

$\sim$ 3 pt **2.** Écrire une fonction `partie` qui simule le déroulement d'une partie d'un joueur (la mise étant choisie aléatoirement comme précisé dans l'énoncé). Cette fonction ne prendra aucun argument et renverra le gain net de la partie *pour le casino*.

$\sim$ 3 pt **3.** Écrire une fonction `attente` qui simule le temps qu'il faut attendre à un instant

[1]. Réf. SG241
[2]. Voir figure 1 ;-)

donné avant qu'un joueur ne joue une partie. La fonction ne prendra aucun argument et renverra le temps attendu.

∼ 9 pt **4.** Écrire une fonction `soiree` qui simule l'évolution de la réserve disponible dans les coffres du casino au cours d'une soirée, sans se soucier de la contrainte physique qui fait qu'en pratique, la réserve est obligée de rester positive. Cette fonction prendra en argument la provision initialement constituée par le casino et affichera une visualisation graphique de l'évolution de la réserve au cours du temps, sans retourner de valeur.

*Le casino s'intéresse à la probabilité qu'un joueur fasse "*sauter la banque*" : on dit que la banque saute lorsque, à un instant donné, le casino se retrouve engagé à verser à un joueur une quantité d'argent dont il ne dispose pas dans ses coffres.*

∼ 3 pt **5.** Par rapport à la simulation de la question précédente, à quoi correspond le fait que la banque saute au cours de la soirée ?

∼ 5 pt **6.** Écrire une fonction `sautage` qui simule l'évolution d'une soirée et nous dit si la banque a sauté. Cette fonction prendra en argument la provision initialement constituée par le casino, et renverra 1 si la banque saute et 0 sinon (vous pouvez aussi utiliser des booléens).

∼ 6 pt **7.** Écrire un programme `psautage` pour estimer par la méthode de Monte-Carlo la probabilité que la banque saute. Ce programme prendra en argument la provision initialement constituée par le casino et affichera la probabilité estimée que la banque saute, assortie d'un intervalle de confiance à 90 %, sans retourner de valeur.

∼ 5 pt **8.** Justifier que la probabilité (exacte) que la banque saute est une fonction décroissante de la provision initialement constituée par le casino.

*En réalité, ce qui préoccupe la casino n'est pas uniquement le risque que la banque saute. Un autre risque est celui de braquage... Or, en cas de braquage, si les voleurs parviennent à vider la caisse, le préjudice subi par le casino sera évidemment d'autant plus sévère que la provision était élevée : il n'est donc pas forcément judicieux de constituer une provision trop importante ! Le casino estime qu'un sautage de la banque lui coute en pratique 100 000 € (à cause de la perte de réputation que cela entraine) ; quant au risque de braquage, il est estimé à une malchance sur 10 000 par soirée, et le cas échéant on suppose pour simplifier que le montant perdu correspond exactement à celui de la provision. Ce que la banque cherche à minimiser en réalité, c'est l'espérance du préjudice total subi.*

<u>*Indication :*</u> *Je suggère d'appeler* `csautage` *le cout d'un sautage de la banque, et* `pbraquage` *la probabilité de braquage pour une soirée.*

∼ 5 pt **9.** Écrire un programme `Eprejudice` estimant l'espérance du préjudice subi (par soirée), avec un intervalle de confiance à 90 %. Ce programme prendra en argument la provision initialement constituée par le casino, et affichera l'estimation de l'espérance du préjudice subi, assortie d'un intervalle de confiance à 90 %, sans retourner de valeur.

∼ 9 pt **10.** Notre casino hésite entre deux montants possibles pour provisionner ses coffres, mais compte tenu de la puissance de calcul dont il dispose, le programme `Eprejudice` ne lui permet pas d'avoir des résultats suffisamment fins pour trancher sur l'identité de la meilleure option... Quelle technique de réduction de la variance peut-il utiliser pour pallier ce souci ? Implémenter cette technique dans un programme appelé `comparaison`, qui prendra en arguments les deux montants de provisions envisagés, et qui répondra si on peut affirmer que le premier montant est le meilleur choix au risque de 10 %, si on peut affirmer que le second montant est le meilleur choix au risque de 10 %, ou si on ne peut pas trancher au niveau de risque considéré (la fonction ne retournera par ailleurs aucune valeur).

*Pour rendre les calculs plus rapides, on peut envisager de modéliser certains aspects du processus de l'évolution de la réserve* $(R_t)_{t\in[0,T_{\text{soirée}}]}$ *par une équation différentielle stochastique. Par exemple, on pourrait souhaiter approximer la contribution des petites parties (parties dont les mises sont de montant* $\text{mise}_{\text{min}}$*) sous la forme d'un incrément de mouvement brownien avec dérive, de sorte que*

$$dR_t \simeq a dW_t + b dt + (\textit{contribution des moyennes et grosses mises}). \qquad (1)$$

$\sim 5$ pt **11.** Pourquoi l'équation (1) est-elle une approximation raisonnable pour la contribution des petites parties ?

$\sim 6$ pt **12.** Calculer théoriquement les valeurs de $a$ et $b$ adaptées à l'approximation de la contribution des moyennes et grosses mises.

<u>*Indication :*</u> Les valeurs de $a$ et $b$ doivent être choisies de sorte que le mouvement brownien avec dérive ait la même moyenne et la même variance que le véritable processus ; il faut donc calculer la moyenne et la variance de la contribution des petites mises pour déterminer $a$ et $b$.

## PROBLÈME 2 — L'or de la rivière

*Une compagnie minière s'intéresse à l'achat d'une vallée dans une région réputée pour ses cours d'eau aurifères, afin de prospecter de l'or dans la rivière qui y coule. On modélisera le cours de la rivière par le segment* $[0, L]$*, où* $L$ *est la longueur de la vallée (*$L = 50$ km*). Les ingénieurs géologues ont établi que dans cette région, quand on parcourt un segment, le profil de la densité d'or le long de ce segment (en masse d'or par unité de longueur) peut être modélisée de la façon suivante :*

- *L'or est présent sous la forme de "gisements" indépendants, dont les contributions se superposent les unes aux autres. (Un même point peut très bien appartenir à plusieurs gisements différents).*
- *Chaque gisement* $G$ *est caractérisé par l'emplacement de son centre (*$z_G$*), la quantité d'or qu'il contient (*$q_G$*) et l'étalement spatial du gisement autour de son centre (*$r_G$*). La contribution du gisement au profil de densité d'or le long de la rivière est*

  $$d_G(x) = \big(\alpha_G - \beta_G(x - z_G)^2\big)_+,$$

  *où* $d_G(x)$ *désigne la densité d'or en* $x$ *qui est due au gisement* $G$*, et où « $\cdot_+$ » est l'opération "partie positive" (càd.* $a_+ := \max(a, 0)$*). Les paramètres* $\alpha_G$ *et* $\beta_G$ *sont reliés à* $q_G$ *et* $r_G$ *de sorte que*

  $$\int_{\mathbb{R}} d_G(x)\, dx = q_G$$

  *et*

  $$\operatorname{supp} d_G = [z_G - r_G, z_G + r_G],$$

  *où 'supp' désigne le support d'une fonction (càd. [l'adhérence topologique de] l'ensemble des points où cette fonction est non nulle).*
- *À chaque point* $x$*, la probabilité qu'il y ait un gisement d'or centré sur ce point est indépendante, égale à* $\lambda dx$*, où* $\lambda$ *est un certain paramètre du modèle (*$\lambda =$ 0,04 km$^{-1}$*). En d'autres termes, l'ensemble des centres de gisements d'or est un processus ponctuel de poisson sur* $\mathbb{R}$*, d'intensité* $\lambda$*.*

– *Pour un gisement d'or donné, la loi des paramètres $q_G$ et $r_G$, qui est la même indépendamment d'un gisement à l'autre, est telle que*

$$d\mathbb{P}\big(q_G = q, r_G = r\big) = Z\mathbf{1}_{r<R}\frac{\exp\big(-q\,/\,\kappa(R-r)\big)}{R-r}dqdr, \qquad (2)$$

*où $R$ et $\kappa$ sont des paramètres du modèle ($R = 25$ km, $\kappa = 1$ t $\cdot$ km$^{-1}$) ('t' est le symbole de la tonne), et où $Z$ est la constante de renormalisation qui assure qu'on ait bien une densité de probabilité.*

*Indication : Dans tout ce problème, je suggère de prendre comme unités de base respectivement le kilomètre pour les distances et la tonne pour les quantités d'or.*

$\sim$ 8 pt **13.** Écrire une fonction `qetr` qui simule la loi décrite par (2). Ce programme ne prendra aucun argument et renverra les valeurs de $q$ et $r$ pour le gisement.

Indication : Pour arriver à simuler cette loi, la meilleure méthode en l'occurrence consiste à d'abord déterminer la loi (marginale) de $r$, puis la loi conditionnelle de $q$ sachant la valeur de $r$, de façon qu'on simulera d'abord $r$, puis $q$.

Indication : Si vous ne parvenez pas à résoudre cette question, pour la suite du problème, je vous suggère de prendre $q$ et $r$ indépendants, avec $q$ uniforme sur $[0, \kappa]$ et $r$ uniforme sur $[0, R]$.

$\sim$ 4 pt **14.** Déterminer $\alpha$ et $\beta$ en fonction de $q$ et $r$ (j'enlève ici les indices '$\cdot_G$'), et écrire une fonction `alphabeta` qui prend en argument $q$ et $r$ et renvoie $\alpha$ et $\beta$.

$\sim$ 4 pt **15.** Écrire une fonction `gisement` qui prend en argument les paramètres $(q_G, r_G, z_G)$ d'un gisement, et renvoie le profil de densité dû à ce gisement sur le cours d'eau de la rivière. Ce profil sera renvoyé sous la forme de deux vecteurs-lignes de même longueur : le vecteur des densités sur différents points de discrétisation de $[0, L]$ (on prendra un pas de discrétisation de 50 m au plus), et le vecteur des abscisses correspondantes.

*On note $\bar{d}(x)$ la densité d'or en un point $x$ de la rivière, qui est la somme des contributions $\bar{d}_G(x)$ des différents gisements.*

$\sim$ 5 pt **16.** Écrire une fonction `profil` qui simule le profil $(\bar{d}(x))_{x\in[0,L]}$ de densité de l'or sur le cours de la rivière et qui trace celui-ci. La fonction ne prendra pas d'argument et ne renverra rien.

Indication : Attention, il peut y avoir un effet de certains gisements dont le centre n'est pas sur le cours d'eau de la rivière, mais avant ou après celui-ci ! On remarquera toutefois qu'on dispose d'une borne à priori sur les valeurs possibles de $r$ (laquelle ?), ce qui nous permet de déterminer un segment sur lequel il suffit de regarder les centres de gisements à prendre en compte qui sont susceptibles d'avoir un impact sur le cours d'eau.

$\sim$ 2 pt **17.** Le processus $(\bar{d}(x))_{x\in[0,L]}$ est-il markovien ? Justifier.

$\sim$ 2 pt **18.** Justifier que le processus $(\bar{d}(x))_{x\in[0,L]}$ est *stationnaire*, càd. que la loi de (la restriction de) le profil $d$ sur un segment $[a_0, a_1]$ est indentique à sa loi sur le segment $[a_0 + b, a_1 + b]$ pour tous $a_0, a_1, b$ (sous réserve que ces segments soient contenus dans $[0, L]$, s'entend).

*Pour savoir si la vallée vaut le coup d'être achetée, la compagnie procède à un sondage en deux points de la rivière d'abscisses $x_1 = 15$ km et $x_2 = 35$ km, qui lui permet d'y déterminer les valeurs de la densité $\bar{d}$. On note $\bar{\delta} := \big(\bar{d}(x_1) + \bar{d}(x_2)\big)/2$. On note par ailleurs $\bar{q}$ la quantité totale d'or dans le lit de la rivière.*

$\sim$ 2 pt **19.** Écrire une fonction `deltaqbarre` qui ne prend aucun argument et, après simulation d'un profil $\bar{d}$, renvoie les valeurs de $\bar{\delta}$ et $\bar{q}$ associées.

$\sim 4$ pt **20.** Quelle est l'espérance de $\bar{\delta}$ conditionnellement à la valeur de $\bar{q}$ ? Justifier. En déduire un estimateur sans biais de $\bar{q}$ en fonction de $\bar{\delta}$.

*Indication :* Il y a un argument qui permet de trouver l'espérance conditionnelle de $\bar{\delta}$ sachant $\bar{q}$ sans avoir à faire de calculs compliqués...

$\sim 4$ pt **21.** L'ingénieur mathématicien de la compagnie suggère à celle-ci de calculer la régression linéaire entre $\bar{q}$ et $\bar{\delta}$, afin d'avoir une idée de ce que vaut l'espérance de $\bar{q}$ conditionnellement à $\bar{\delta}$. Autrement dit, on cherche une approximation de $\mathbb{E}(\bar{q}|\bar{\delta})$ sous la forme

$$\mathbb{E}(\bar{q}|\bar{\delta}) \approx A\bar{\delta} + B.$$

Comment l'estimateur des moindres carrés dit-il qu'il faut estimer $A$ et $B$ dans un tel cas ?

$\sim 3$ pt **22.** La question précédente nous amène à devoir estimer certaines espérances, disons les espérances de quantités que nous désignerons par $f_1, f_2, \ldots, f_k$. Implémenter une méthode de Monte-Carlo pour estimer ces espérances (en précisant un intervalle de confiance). **Attention, pour cette question je vous demanderai de ne pas travailler sous la loi $\mathbb{P}$, mais sous la loi $\mathbb{P}(\cdot|\mathcal{E})$, où $\mathcal{E}$ est l'évènement $\{\bar{\delta} \in [\bar{\delta}_-, \bar{\delta}_+]\}$. ($\bar{\delta}_- = 4\ \mathrm{t} \cdot \mathrm{km}^{-1}, \bar{\delta}_+ = 8\ \mathrm{t} \cdot \mathrm{km}^{-1}$).**

*Indication :* Pour simuler la loi $\mathbb{P}(\cdot|\mathcal{E})$, on procèdera simplement par rejet à partir de $\mathbb{P}$.

*Indication :* À défaut d'avoir trouvé la réponse à la question précédente, vous pouvez prendre $k = 1$ et $f_1 = \bar{q}\bar{\delta}$.

$\sim 5$ pt **23.** Implémenter une méthode d'échantillonnage préférentiel pour estimer $\mathbb{P}(\mathcal{E})$ et $\mathbb{E}(f_1 \mathbf{1}_{\mathcal{E}})$ ($\mathcal{E}$ étant l'évènement défini à la question précédente), où la loi d'échantillonnage diffère de la loi vraie par sa valeur de $\gamma$ (on pourra par exemple multiplier $\gamma$ par deux). Commenter les résultats obtenus.

$\sim 4$ pt **24.** On observe le paradoxe suivant : pour les grandes valeurs $\bar{q}_1$ de $\bar{q}$, notant $\bar{\delta}_1 := \mathbb{E}(\bar{\delta}|\bar{q} = \bar{q}_1)$, on a $\mathbb{E}(\bar{q}|\bar{\delta} = \bar{\delta}_1) \ll \bar{q}_1$ ! Commenter : En quoi est-ce paradoxal ? Comment expliquer le phénomène ? Notre compagnie doit-elle plutôt s'intéresser à trouver un estimateur sans biais de $\bar{q}$ ou à connaitre l'espérance de $\bar{q}$ conditionnellement à $\bar{\delta}$ ? Quelles seraient les conséquences possibles d'un mauvais choix ?

## PROBLÈME 3 — Dans l'esprit du compositeur

$\sim 5$ pt **25.** À votre avis, à la question 22, pourquoi vous ai-je demandé de travailler sous la loi $\mathbb{P}(\cdot|\mathcal{E})$ plutôt que sous la loi $\mathbb{P}$ ?

*Indication :* Il y a deux réponses possibles : l'une est une raison strictement mathématique ; l'autre concerne les implications pratiques de se restreindre à l'évènement $\mathcal{E}$ pour la situation d'entreprise présentée par l'énoncé.

$\sim 7$ pt **26.** À partir des résultats de la question 23, on peut obtenir un estimateur de $\mathbb{E}(f_1|\mathcal{E})$ (lequel ?). Comment calculeriez-vous un intervalle de confiance associé à cet estimateur, aussi finement que possible ? Dans le cas où on échantillonne sous la véritable loi, votre estimateur de $\mathbb{E}(f_1|\mathcal{E})$ donne-t-il (essentiellement) les mêmes résultats que l'estimateur "naïf" en échantillonnant sous $\mathbb{P}(\cdot|\mathcal{E})$, ou perd-on quelque chose à utiliser la nouvelle méthode ?


FIGURE 1 – Illustration d'*Anicet Le Pingouin*, par Totor & Coco