

**Date :** Mardi le 4 novembre 2003

**Heure :** 10 h 30 à 12 h 30

**Professeur :** Pierre L'Ecuyer

Aucune documentation permise. Le nombre de points accordés à chaque question est indiqué entre parenthèses.

---

— 1 —

(20 points)

Supposons que l'on dispose de  $n$  observations indépendantes  $x_1, \dots, x_n$ , provenant d'une loi de probabilité continue définie sur  $[0, \infty)$ , que l'on ne connaît pas.

- Comment définit-on la fonction de répartition empirique  $\hat{F}_n$  de ces observations?
- Donnez un algorithme, le plus simple possible, permettant de générer des variables aléatoires i.i.d. selon cette fonction de répartition empirique, par inversion.
- Proposez une fonction de répartition quasi-empirique raisonnable, pour ces observations, dont la densité est strictement positive dans tout l'intervalle  $(0, \infty)$ .
- Expliquez comment générer des variables aléatoires i.i.d. selon votre fonction de répartition quasi-empirique, par inversion.

— 2 —

(10 points)

Dans l'exemple du centre d'appels téléphonique vu en classe, une variable aléatoire importante dans le modèle est  $B$ , le facteur d'achalandage pour la journée, qui suit la loi gamma de paramètres  $(\alpha_0, \alpha_0)$ . Rappelons qu'une v.a.  $X$  qui suit la loi gamma de paramètres  $(\alpha, \lambda)$  a la densité

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{pour } x > 0,$$

la moyenne  $E[X] = \alpha/\lambda$  et la variance  $\text{Var}[X] = \alpha/\lambda^2$ .

Supposons que l'on veut estimer une mesure de performance  $\mu$  qui dépend d'événements qui ne se produisent que lorsqu'il y a beaucoup d'achalandage dans le centre (par exemple la probabilité d'un grand nombre d'abandons). Soit  $Y$  un estimateur sans biais de cette mesure  $\mu$ , obtenu en simulant une journée.

Dans ce contexte, supposons que l'on décide de générer les valeurs de  $B$  en utilisant une loi gamma de paramètres  $(\alpha_1, \lambda_1)$  au lieu de  $(\alpha_0, \alpha_0)$ , afin d'augmenter la moyenne de  $B$ . Alors la variable aléatoire  $Y$  ne sera plus un estimateur sans biais de  $\mu$ .

- (a) Expliquez comment récupérer un estimateur sans biais de  $\mu$  dans ce cas.
- (b) Quel avantage peut-il y avoir à modifier ainsi les paramètres de la loi de  $B$  afin d'augmenter sa moyenne?

— 3 —

(10 points)

Supposons que les réclamations d'assurance automobile chez un assureur arrivent selon un processus de Poisson non stationnaire dont la fonction de taux est  $\lambda(t)$ ,  $t \geq 0$ , et le taux cumulé est  $\Lambda(t) = \int_0^t \lambda(s) ds$ ,  $t \geq 0$ . Soit  $T_0 = 0$  et  $T_i$  l'instant de la  $i$ -ième arrivée. On dispose d'une suite de v.a. i.i.d.  $U(0, 1)$ ,  $U_1, U_2, \dots$ , et on veut générer les  $T_i$  à partir des  $U_i$ .

Définissons le processus stochastique  $\tilde{N}(t) = \sup\{i \mid \Lambda(T_i) \leq t\}$ ,  $t \geq 0$ . Quelle est la propriété principale de ce processus? Justifiez. Donnez un algorithme générant les  $T_i$  à partir des  $U_i$  en utilisant cette propriété de  $\tilde{N}$ . Sous quelle condition cet algorithme est-il facile à implanter?

— 4 —

(20 points)

Un modèle souvent utilisé pour l'évolution du prix d'un actif (en économie ou en finance) est le mouvement brownien géométrique. Rappelons qu'un processus  $\{S(t), t \geq 0\}$  est un *mouvement brownien géométrique* si

$$S(t) = S(0) \exp[\mu t + \sigma B(t)]$$

où  $\mu$  et  $\sigma > 0$  sont des constantes et  $\{B(t), t \geq 0\}$  est un mouvement brownien standard, avec  $B(0) = 0$ .

Rappelons aussi que pour un mouvement brownien *ordinaire* de paramètres  $\mu$  et  $\sigma^2 > 0$ , conditionnellement à  $B(0) = a$  et  $B(t) = b$ , pour  $0 < s < t$ , on a  $E[B(s)] = (sb + (t - s)a)/t$  et  $\text{Var}[B(s)] = \sigma^2 s(t - s)/t$ .

Supposons que l'on veut simuler la valeur d'un mouvement brownien géométrique aux instants  $t = 1, 2, \dots, 32$ , afin d'évaluer le revenu d'un contrat financier qui dépend des valeurs  $S(1), S(2), \dots, S(32)$ .

Donnez et expliquez *deux* algorithmes différents pour générer ces 32 valeurs, l'un qui opère de manière séquentielle et l'autre qui utilise le pont brownien. Quel pourrait être l'avantage de cette seconde manière dans le contexte où on veut utiliser une méthode de type quasi-Monte Carlo pour simuler le processus?

(20 points)

Soit  $\{C_i, i \geq 1\}$  un processus stochastique faiblement stationnaire pour lequel  $E[C_i] = \mu$ ,  $\text{Var}[C_i] = \sigma^2 > 0$ , et soit  $\rho_k$  l'autocorrelation de délai  $k$ , pour  $k \geq 0$  (avec  $\rho_0 = 1$ ).

On veut estimer  $\mu$  par la moyenne empirique

$$\bar{C}_n = \frac{1}{n} \sum_{i=1}^n C_i.$$

On veut aussi estimer la variance de cet estimateur,  $\text{Var}[\bar{C}_n]$ .

(a) Donnez une expression exacte pour cette variance en fonction de  $\sigma^2$  et des  $\rho_k$ , pour  $n$  fini. Donnez ensuite une expression pour  $n\text{Var}[\bar{C}_n]$  lorsque  $n \rightarrow \infty$ , en supposant que cette limite existe.

(b) Supposons maintenant que  $C_i = h(\mathcal{S}_i)$  où  $\mathcal{S}_i \in \mathbb{S}$  représente l'état d'un modèle de simulation lors du  $i$ -ième événement,  $\{\mathcal{S}_i, i \geq 1\}$  est une chaîne de Markov  $v$ -uniformément ergodique telle que  $|h(\mathcal{S})| \leq v(\mathcal{S})$  pour tout  $\mathcal{S} \in \mathbb{S}$ .

Sous cette condition, énoncez un théorème de limite centrale pour  $\bar{C}_n$ , qui permet de calculer un intervalle de confiance asymptotiquement valide pour  $\mu$  à condition de disposer d'un estimateur de variance approprié. Dire aussi quel est le lien entre ceci et la limite mentionnée en (a).

(c) L'une des possibilités pour estimer cette variance est d'utiliser la méthode des moyennes par lots ("batch means"). Expliquez en quoi consiste cette méthode et sous quelles conditions elle est asymptotiquement valide (lorsque  $n \rightarrow \infty$ ). Y a-t-il un lien entre ces conditions et la condition donnée en (b)?

(20 points)

Dans un modèle de simulation, supposons que le  $i$ -ième événement a lieu au temps  $t_i$  et qu'un coût  $C_i \geq 0$  est encouru lors de cet événement. Les  $t_i$  et  $C_i$  sont des variables aléatoires. On suppose aussi que les coûts sont actualisés au taux  $\rho > 0$ . On veut estimer l'espérance du coût total actualisé sur horizon infini, noté  $v_\rho^\infty$ , que l'on suppose fini.

(a) Si on dit que le processus  $\{C_i, i \geq 1\}$  est *régénératif*, expliquez ce que cela veut dire (donnez une définition précise).

(b) Dans le cas où le processus  $\{C_i, i \geq 1\}$  est régénératif, expliquez comment on peut écrire  $v_\rho^\infty$  comme un quotient de deux espérances, chacune étant l'espérance d'une variable aléatoire définie sur un cycle régénératif (horizon fini).

(c) Sous les hypothèses de (b), calculer un intervalle de confiance pour  $v_\rho^\infty$  se ramène à calculer un intervalle de confiance sur un quotient de deux espérances. Expliquer comment on peut calculer un tel intervalle de confiance pour un quotient.