

**Date :** Mardi le 6 novembre 2007

**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 —

(10 points)

Une variable aléatoire suit une loi de Weibull si sa fonction de répartition a la forme  $F(x) = 1 - \exp[-(\lambda x)^\alpha]$  pour  $x \geq 0$ . Développez et énoncez un algorithme pour générer des variables aléatoires selon cette loi, par inversion.

— 2 —

(10 points)

Supposons que l'on veut générer une paire de variables aléatoires  $(X, Y)$  telles que  $X \sim N(0, \sigma^2)$ ,  $Y \sim \text{Weibull}(\alpha, \beta)$ , et avec la plus grande corrélation possible entre  $X$  et  $Y$ . Expliquez comment faire cela. Et si on veut la plus petite corrélation possible? Dans les deux cas, dire quelle sera la corrélation des rangs (ou de Spearman) entre  $X$  et  $Y$ .

— 3 —

(15 points)

(a) Expliquer une méthode simple et efficace pour générer un point au hasard (uniformément) sur la surface d'une sphère de rayon 1, en  $d$  dimensions.

(b) L'ensemble des points  $\mathbf{x} \in \mathbb{R}^d$  qui satisfont à l'inégalité  $\mathbf{x}^t \mathbf{A} \mathbf{x} \leq 1$ , où  $\mathbf{A}$  est une matrice diagonale  $d \times d$  dont les éléments  $a_{11}, \dots, a_{dd}$  sont tous positifs, forme un ellipsoïde en  $d$  dimensions, centré à l'origine, aligné avec les axes des coordonnées cartésiennes. Quelle est la plus petite boîte rectangulaire dans lequel s'inscrit cet ellipsoïde? Expliquez en détail comment générer un point au hasard (uniformément) dans cet ellipsoïde, par une méthode de rejet. Discutez de l'efficacité de votre méthode lorsque  $d$  augmente (par exemple,  $d = 2, 3, 10, 50$ ).

(20 points)

Soit  $X$  une variable aléatoire définie par

$$X = \begin{cases} \max(0, Y_1 + Y_2 + Y_3 - K) & \text{si } \{Y_1 < a \text{ et } Y_2 < b\}, \\ 0 & \text{sinon,} \end{cases}$$

où  $a, b$  et  $K$  sont des constantes, et  $Y_1, Y_2, Y_3$  sont des variables aléatoires indépendantes  $N(1, 1)$ . On veut estimer  $\mu = \mathbb{E}[X]$ .

(a) Expliquez en trois lignes comment obtenir un intervalle de confiance pour  $\mu$  par la méthode Monte Carlo standard.

(b) Dans le cas où  $p = \mathbb{P}[X > 0]$  est très petit, par exemple inférieur à  $10^{-6}$ , quels sont les problèmes rencontrés avec cette méthode standard?

(c) Une solution dans ce cas est d'appliquer l'échantillonnage stratégique (importance sampling, ou IS). Expliquez clairement une façon simple de faire cela dans ce cas-ci. Expliquez en particulier comment seront générées les variables aléatoires et quel sera l'estimateur IS.

(20 points)

On considère une option financière dont le revenu net  $Y(s)$  est une variable aléatoire qui est fonction de la valeur d'un mouvement Brownien géométrique (MBG)  $\{S(t), t \geq 0\}$  aux instants  $0 < t_1 < \dots < t_c = T$ , dont la valeur initiale est  $S(0) = s$ . La valeur de l'option est  $v(s) = \mathbb{E}[Y(s)]$ , en supposant que l'espérance est par rapport à la mesure de probabilité de risque neutre. Rappelons que pour un MBG, on peut écrire

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

où  $r$  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 standard, conditionnellement à  $B(0) = a$  et  $B(t) = b$ , pour  $0 < s < t$ , on a  $E[B(s)] = (sb + (t - s)a)/t$  et  $\text{Var}[B(s)] = s(t - s)/t$ . On peut alors écrire

$$Y(s) = g(s, S(t_1), \dots, S(t_c)) = h(s, B(t_1), \dots, B(t_c))$$

pour certaines fonctions  $g$  et  $h$ , où  $\{B(t), t \geq 0\}$  est un mouvement Brownien standard.

(a) Expliquez comment simuler des copies indépendantes de  $Y(s)$  par une méthode séquentielle (en avançant le temps).

(b) Expliquez comment simuler des copies indépendantes de  $Y(s)$  par une méthode de pont Brownien, en supposant que  $c = 64$ .

(c) On veut estimer  $v'(s) = \partial v(s)/\partial s$  en utilisant comme estimateur la différence finie

$$X(s, \delta) = \frac{Y(s + \delta) - Y(s)}{\delta},$$

pour un petit  $\delta > 0$ , où  $Y(s + \delta)$  et  $Y(s)$  sont indépendantes. En supposant que  $v(s)$  possède un développement en série de Taylor:  $v(s + \delta) = v(s) + v'(s)\delta + v''(s)\delta^2/2 + O(\delta^3)$ , donnez des expressions pour le biais, la variance, et l'erreur quadratique moyenne (MSE) de  $X(s, \delta)$  en fonction de  $\delta$ .

(d) Supposons que l'on estime  $v'(s)$  par la moyenne de  $n$  copies indépendantes de  $X(s, \delta)$  défini en (c). Supposons aussi que chaque calcul de  $Y(s)$  ou  $Y(s + \delta)$  prend un temps constant  $c_0$ . Quelle est la valeur de  $\delta$  qui maximise l'efficacité de l'estimateur, en fonction de  $n$ ? (Une approximation asymptotiquement valide lorsque  $n \rightarrow \infty$  suffit. Il faut bien sûr prouver votre résultat.)

(e) Expliquez en quoi consiste l'utilisation de valeurs aléatoires communes (VAC) dans ce contexte, et comment cela peut améliorer l'efficacité.

— 6 —

(25 points)

Un générateur à congruence linéaire multiple (en anglais, “multiple recursive generator” ou MRG) est défini par une récurrence linéaire de la forme

$$x_n = (a_1x_{n-1} + \dots + a_kx_{n-k}) \bmod m,$$

où  $0 < a_k < m$ , et la sortie est  $u_n = x_n/m$ . Vous pouvez répondre à chacune des sous-questions suivantes en *moins de 10 lignes*.

(a) Est-ce que la suite  $\{u_n, n \geq 0\}$  produite par cette récurrence est nécessairement périodique? Si oui, quelle est la longueur maximale de sa période? En pratique, comment vérifie-t-on si la période est maximale ou pas?

(c) Supposons que l'on connaît  $\mathbf{x}_n = (x_n, \dots, x_{n-k+1})^t$  et que l'on veut calculer  $\mathbf{x}_{n+\nu} = (x_{n+\nu}, \dots, x_{n+\nu-k+1})^t$  pour  $\nu = 2^{100}$ . On veut répéter cela des milliers de fois pour le même  $\nu$ , mais pour des états  $\mathbf{x}_n$  différents et que l'on ne connaît pas à l'avance. Expliquez comment faire cela efficacement et à quoi cela peut servir.

(d) Pour rendre l'implantation plus rapide, est-ce une bonne idée de ne prendre que deux coefficients  $a_j$  non nuls dans la récurrence du MRG, et de les prendre tous les deux égaux à  $\pm 1$ ? Pourquoi?

(e) L'ensemble  $\Psi_t$  de tous les vecteurs de  $t$  valeurs successives que ce générateur peut produire, à partir de tous les états initiaux possibles, est l'intersection d'un réseau (“lattice”) avec l'hypercube unitaire  $[0, 1]^t$ . Expliquez ce que cela veut dire comment on utilise ce réseau pour mesurer l'uniformité de  $\Psi_t$

(f) Certains générateurs (par exemple MRG32k3a dans SSJ) sont définis en combinant 2 ou 3 MRGs ayant des moduli  $m_j$  qui sont des nombres premiers distincts. La combinaison peut se faire, par exemple, en additionnant les  $u_{j,n}$  modulo 1 (pour  $n$  fixé), où  $u_{j,n}$  est l'output de la composante  $j$  à l'étape  $n$ . Expliquez quels sont les principaux avantages de ces MRGs combinés, quelle est la structure de leur ensemble de points  $\Psi_t$ , et comment on fait pour mesurer leur uniformité.