Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2007

This paper is also taken for the relevant examination for the Associateship.

M3S9/M4SS9

STOCHASTIC SIMULATION

Date: Monday, 14th May 2007 Time: 10 am – 11:30 am

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

1. Suppose that we wish to sample from a probability density function f, and have a simple way to sample from a probability density function g, with $f(x)/g(x) \leq M < \infty$ for all $-\infty < x < \infty$. For a given auxiliary function $h(x)$, $0 \leq h(x) \leq 1$ for all $-\infty < x < \infty$, the following simulation algorithm is suggested:

   (1) Generate Y from pdf g.

   (2) With probability $h(Y)$ accept $X = Y$, else go to (1).

   (a) Find the distribution of accepted values X and show that by suitable choice of function h the accepted values will have pdf f.

   (b) For the choice of h found in part (a), determine

      (i) the probability that a generated Y is accepted, and

      (ii) the distribution of the number of trials before a Y is accepted.

   (c) Comment briefly upon the need to know the normalising constant of f when applying the above method and upon the efficiency of such a method.

   (d) Suppose we can find easily evaluated functions l and u with

$$0 \leq l(x) \leq \frac{f(x)}{g(x)} \leq u(x) \quad \text{, for all } \infty < x < \infty.$$

   Assume that $f(x)$ and $g(x)$ are time-consuming to calculate, suggest how we might use this knowledge to speed up the simulation algorithm.

2. (a) Consider the problem of estimating the integral $\theta = \int_0^1 e^x \, dx$.

      (i) Construct the Monte Carlo estimator $\widehat{\theta}_1$ of θ using a sequence, $\{X_1, \ldots, X_n\}$, of independent random variates with probability density function $f(x) = 1$, $0 \leq x \leq 1$. Calculate the variance of $\widehat{\theta}_1$.

      (ii) Construct the control variate Monte Carlo estimator $\widehat{\theta}_2$ of θ using the uniform random variable X as the control variate and calculate the smallest achievable variance of $\widehat{\theta}_2$.

   (b) Now, consider the problem of estimating the integral $\theta' = \int_{-1}^1 e^x \, dx$. Use stratification sampling to construct an estimator $\widetilde{\theta}'$ of θ, with two sequences $\{Y_1, \ldots, Y_{n/2}\} \sim U(-1, 0)$ and $\{Z_1, \ldots, Z_{n/2}\} \sim U(0, 1)$. Calculate the variance of $\widetilde{\theta}'$.

3. (a) Describe the 'Frequency Test of Digits' method for testing randomness of the digit in the first decimal place of a sequence of n uniform $(0,1)$ random numbers.

(b) Consider the following change point model:

$$X_i \sim N(\theta_1, \sigma^2) \qquad 1 \le i < \tau$$

$$X_i \sim N(\theta_2, \sigma^2) \qquad \tau \le i \le n,$$

with unknown parameters $\theta_1, \theta_2, \sigma^2$ and τ ($\tau \neq 1$). Write $\delta = \sigma^{-2}$. Suppose that prior distributions for the parameters are specified by

$$\begin{aligned} \theta_i &\sim N(0,1) \qquad i = 1,2 \\ \delta &\sim \text{Gamma}(\alpha_0, \beta_0) \\ \tau &\sim \text{Uniform distribution on}\{1, 2, \ldots, n\}, \end{aligned}$$

where $\theta_1, \theta_2, \delta$ and τ are considered to be independent a priori. Assume that α_0 and β_0 are known. Describe how to apply the Gibbs sampler to simulate from the posterior distribution of $(\theta_1, \theta_2, \delta, \tau)$.
(The probability density function of $X \sim$ gamma(α_0, β_0) is proportional to $x^{\alpha_0 - 1} e^{-\beta_0 x}$).

4. Let $X_1, \ldots, X_n$ be independent, identically distributed from a population F. Let $\widehat{\theta}(X_1, \ldots, X_n)$ be an estimator of a population parameter $\theta \equiv \theta(F)$.

(a) Explain what is meant by the *jackknife* and *bootstrap* estimators of the bias of $\widehat{\theta}$, assuming a non-parametric problem.

(b) Let μ and σ^2 be the mean and variance of F respectively. Let $\theta = \mu^2$. Consider the estimator $\widehat{\theta} = \bar{X}^2$, where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. Obtain the form of the jackknife bias estimator b_{jack} and bootstrap bias estimator b_{boot}.

For (c) and (d), refer to the setting of (b).

(c) Describe how the bootstrap bias estimator b_{boot} may be used to construct a refined estimator $\widehat{\theta}_1$, whose bias is of smaller order in the sample size n. Describe also how the jackknife bias estimator b_{jack} may be used to construct another refined estimator $\widetilde{\theta}$.

(d) Suppose the bootstrap procedure is iterated. Show that the bootstrap bias-corrected estimator $\widehat{\theta}_1$ after j iterations is

$$\widehat{\theta}_j = \bar{X}^2 - (n-1)^{-1}(1 - n^{-j})\widehat{\sigma}^2,$$

where $\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Based on this form, describe how to produce an estimator whose bias is arbitrarily small as a function of n.
Compare the estimator obtained as $j \to \infty$ with the estimator $\widetilde{\theta}$.