Instructor: Victor F. Araman February 7, 2005

Stochastic Simulation

Homework 1

due on Thursday 17 February, 2005

Exercise 1 We consider here a statistical procedure that is useful to validate simulation models. Assume that $Y_1, Y_2, ..., Y_n$ are independent random variables and we are interested in testing the null hypothesis, H_0, that they have the common distribution F, where F is a given continuous distribution. One approach to testing H_0 is to discretize the random variables Y_j's and apply the so called "chi-square goodness of fit" test. In this problem we discuss another way to test continuous random variables which is more efficient than discretizing. After observing $Y_1, ..., Y_n$ we define the empirical distribution function F_e by

$$F_e(x) = \frac{\#i : Y_i \leq x}{n},$$

where $\#i$ counts the number of i in the set and F_e is the proportion of observed values that are less or equal to x.

(i) Explain in couple of lines why should we be interested in the following quantity

$$D = \max_{x \in \mathbb{R}} \{|F_e(x) - F(x)|\}$$

(ii) Now set $Y_j = y_j$ for $j = 1, .., n$. Let $y_{(i)}, y_{(2)}, ..., y_{(n)}$ denote the values of y_j in increasing order ($y_{(j)} = j^{th}$ smallest of $y_1, ..., y_n$). Write a formulation of F_e based on the position of x in the intervals $(y_{(k)}, y_{(k+1)})$, $k = 0, 1, ..., n$ and $y_{(0)} = 0, y_{(n+1)} = +\infty$.

(iii) Show that

$$D = \max\{j/n - F(y_{(j)}), F(y_{(j)}) - \frac{j-1}{n},\ j = 1, ..., n\}$$

Hint: start by noting that $\max_x\{F_e(x) - F(x)\} \geq 0$

(iv) Considering D as a random variable, show that the p-value$\equiv \mathbb{P}(D \geq d)$, is given by

$$p\text{-value} = \mathbb{P}(\max_{0\leq y\leq 1} |\frac{\#i : U_i \leq y}{n} - y| \geq d),$$

where $U_1, U_2, ..., U_n$ are independent uniform $(0, 1)$ random variables. Conclude that the distribution of D when H_0 is true does not depend on the actual distribution F

(v) Give an algorithm that will test whether the data observed are consistent with the null hypothesis H_0

This method is known as the Kolmogorov-Smirnov test for continuous data.

Exercise 2 Give an algorithm and explain its steps that will generate a Beta distribution. The Beta density on $[0, 1]$ with parameters $\alpha_1, \alpha_2 > 0$ is given by

$$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1}$$

for $0 \leq x \leq 1$, and

$$B(\alpha_1, \alpha_2 = \int_0^1 x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1} dx = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

Exercise 3 In simulating the waiting time process for the M/M/1 queue, the simulator must generate a sequence of independent, identically distributed random variables, $\{X_n : n \geq 1\}$. A typical X_i, say X_1, is given by $X_1 = V_0 - U_1$, where V_0 and U_1 are independent, exponentially distributed with parameters μ and λ respectively. Develop a method for generating an observation of X_1 that requires only ONE uniform $(0, 1)$ random number.