Statistical Computing
Homework 2: Markov Chain and Simulation
Due Date: March 24, 2003

1. Use R's runif function to generate (pseudo) random numbers. In order to make sure these random numbers are reasonably random, consider the "Runs Test" (see Knuth, *The Art of Programming Volume 2*). Suppose we take a random sequence U_j's and keep counting until $U_j > U_{j+1}$. For example, the following illustrates a run of length three: $U_0 = 0.1, U_1 = 0.2, U_2 = 0.9, U_3 = 0.8$. If we throw away the element that immediately follows a run, then the run lengths are independent. Thus when $U_j > U_{j+1}$ we discard U_{j+1} and start counting again at U_{j+2}.

   a. Let T denote the run length. Derive the probability mass function of T.

   b. If $(n_1, \ldots, n_k)$ is multinomial with success probabilities $\mathbf{p} = (\mathbf{p_1}, \ldots, \mathbf{p_k})$ and N trials then

   $$X = \sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i}$$

   follows a Chi-squared distribution with $k-1$ degrees of freedom as $N \to \infty$. Perform the run test for a very large sample. (Note there is no bound on the largest possible run length. Therefore the last multinomial category should be runs of length x or larger, for some value of x.)

2. Let $Y_i \sim$ Exponential(β) for $i = 1, \ldots, n$ (with mean $1/\beta$). Let β be a half normal (σ) random variable. That is the density of β is

   $$\sqrt{\frac{2}{\pi\sigma^2}} \exp(-\beta^2/2\sigma^2)$$

   for $\beta > 0$. Use rejection sampling to simulate from the posterior of $\beta \mid y_1, \ldots, y_n, \sigma$ for $\sigma = 0.5$ and the following y's

   $$20.100306, 2.272066, 3.796734, 2.265275, 3.480183.$$

   Conduct a study to evaluate the influence of the value of the hyperparameter σ. (You can plot the posterior mean as a function of the hyperparameter σ.)

3. In the precipitation monitor of the US Weather Service at Snoqualmie Falls, the record on January rainfall from 1948 through 1983 is summarized as following

	Today dry	Today wet	Total
Yesterday dry	186	123	309
Yesterday wet	128	643	771
Total	314	766	1080

   a. Since we only consider January, the assumption of stationary transition probabilities seems reasonable. Write down the Markov chain to describe the above data.

b. Assume that the 36 starting values at January 1st of each year as fixed numbers. Write down the likelihood function of this Markov chain and use MLE to derive estimate of transition probabilities.

c. Assume that the weather condition of those 36 starting values at January 1st of each year as determined by the process of flipping a coin with probability of getting head 314/766. Use Monte Carlo method to simulate the Markov Chain a large number of times. How will you use those simulation results to check whether this chain can be a good description of the observed data?

d. Under the above assumed Markov chain, find the probability of rain at Snoqualmie Falls on January 6, given rain on January 1.

e. Determine the correlation $Cov(X_n, X_{n+k})$ for $k = 1, \ldots, 6$ by analytic method. Verify your derivation with your simulation result in 34.

4. Write a function, partReg, that takes three real valued vector arguments, y, x and z and returns a plot of the residuals from the two regression equations: $y = \beta z + Error$ and $x = \alpha z + Error$. Such a plot is called a partial regression plot. It is often described as a plot of y against x having removed z. (Obviously we are simplifying things by making z a univariate vector. Partial regression plots are also used when z is multivariate.) Your homework is to write an R package for your function. You must include:

   a. Appropriate program comments within the R function

   b. Error checking in your function

   c. A sample data set with documentation

   d. A help file for partReg with an example based the sample data set

5. Suppose we want to estimate θ, where

$$\theta = \int_0^1 \exp(x^2)dx.$$

Implement a Monte Carlo estimator by sampling from the uniform and averaging $\exp(U_i^2)$'s. Also, utilize the idea of variance reduction using antithetic variables to propose another algorithm with less variance. Plot the convergence (estimates versus the sample size) of both the algorithms on the same plot.

6. Consider a 10-state random walk with the following transition matrix:

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0.5 & 0 & 0.5 & 0 & \cdots \\ 0 & 0.5 & 0 & 0.5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}.$$

Define

$$d(k) = E[\text{time to reach states } a_1 \text{ or } a_{10} \mid \text{initial state is } a_k].$$

Estimate $d(k)$ via simulation and plot the estimated values $d(k)$ versus k.