STA3431H (Monte Carlo Methods), Winter 2009

Homework #2

Due: In class by 2:10 p.m. **sharp** on Monday March 2.

NOTE: All the same "Notes" from HW#1 still apply. In particular, **homeworks which are late, even by one minute, will be penalised.**

The assignment:

1. Suppose $\pi(x, y) = c\, g(x, y)$ where π and $c > 0$ are unknown, and $g(x, y) = x^2 y^3 \sin(y/5) \cos(x^4 y^2/6) e^{-(x-y)^2}$ for $0 < x < 1$ and $0 < y < 2$, otherwise 0. Let f be the density of the uniform distribution on $[0, 1] \times [0, 2]$. Compute $\mathbf{E}_\pi[X^2 Y]$ using each of the following two different Monte Carlo algorithms, each based on samples from this same density f. For each algorithm, you should explain and program and run it, and also estimate (with explanation) its standard error.

(a) An Importance Sampler.

(b) A Rejection Sampler.

2. Consider the standard variance components model described in lecture. Suppose $K = 2$ and $J_1 = J_2 = 4$, with data $Y_{11} = 1$, $Y_{12} = 2$, $Y_{13} = 3$, $Y_{14} = 4$, $Y_{21} = 2$, $Y_{22} = 4$, $Y_{23} = 6$, $Y_{24} = 8$. Use the prior values $a_1 = a_2 = b_1 = b_2 = 1$, $a_3 = 0$, and $b_3 = 4$. Program an appropriate Metropolis algorithm for the corresponding posterior distribution, and use it to estimate (as best as you can, together with standard errors) the posterior probability that $V > W$.

3. Repeat the previous question where now $K = 6$ and $J_i \equiv 5$, and the data Y_{ij} are the famous "dyestuff" data, available in the file "Rdye".

4. Consider the homerun baseball data in the file "Rhomerun", giving the number of homeruns H_i and number of attempts (at-bats) A_i for players $1 \leq i \leq 12$. Consider the A_i to be fixed, known, constants, and the H_i to be observed data. Assume that $H_i \sim \text{Binomial}(A_i, \theta_i)$ (cond. ind.), where $\theta_i \sim \text{Beta}(1001,\, 1 + 1000\, S)$ (cond. ind.) are unknown. Finally, put a prior $S \sim \text{Poisson}(4)$ on S.

(a) Specify (up to a normalising constant) the joint probabilities for $\theta_1, \ldots, \theta_{12}, S, H_1, \ldots, H_{12}$.

(b) By conditioning on the observed values of the H_i (from "Rhomerun"), specify (up to a normalising constant) the conditional (posterior) probabilities for $\theta_1, \ldots, \theta_{12}, S$.

(c) Run a Metropolis algorithm (or other MCMC algorithm) for this posterior distribution (with appropriate proposal scaling and run length and burn-in, as best as you can, together with standard errors), to estimate the posterior means of each of the 13 variables $\theta_1, \ldots, \theta_{12}, S$. [Note: this is not an easy simulation, and will probably require very long runs with very small proposal scalings to get it right.]