# STA3431H (Monte Carlo Methods), Winter 2011, Homework #2

**Due:** In class by 2:10 p.m. **sharp** on Monday March 28.

**NOTES:**

- **Late homeworks, even by one minute, will be penalised!**
- Homework assignments are to be solved by each student individually; you may discuss assignments in general terms with other students, but you must solve it on your own, including doing all of your own computing and writing.
- For full points you should provide very complete solutions, including explaining all of your reasoning clearly and neatly, performing detailed Monte Carlo investigations including multiple runs as appropriate, justifying all of the choices you make, etc.
- You may use results from lecture, but clearly state when you are doing so.
- When writing computer programs for homework assignments:
  - R is the "default" computer programming language and should normally be used for homework; under some circumstances it may be permitted to use other standard computer languages, but only with prior permission from the instructor.
  - You should include both the complete source code and the program output.
  - Programs should be clearly explained, with comments, so they are easy to follow.
- Reminder: your final project is due April 4 at 2:10pm.

**THE ACTUAL ASSIGNMENT:**

**1.** Consider the Markov chain with state space $\mathcal{X} = \{1, 2, \ldots, 100\}$, and transition probabilities given by $P(j, j) = 0.8$ for all $j \in \mathcal{X}$, and $P(1, 2) = P(100, 99) = 0.2$, and $P(j, j+1) = P(j, j-1) = 0.1$ for $2 \le j \le 99$, with $P(i, j) = 0$ otherwise.

**(a)** Compute $\lim_{n\to\infty} \mathbf{P}(X_n = 3)$ analytically, including a complete proof of your answer. [Hint: is this Markov chain reversible with respect to some probability distribution?]

**(b)** Write and run a computer program to simulate this Markov chain, and relate (with discussion) your program's output to the answer from part (a).

**2.** Consider an independence sampler algorithm on $\mathcal{X} = (1, \infty)$, where $\pi(x) = 5\, x^{-6}$ and $q(x) = r\, x^{-r-1}$ for some choice of $r > 0$, with identity functional $h(x) = x$.

**(a)** For what value of $r$ will the algorithm provide i.i.d. samples?

**(b)** For what values of $r$ will the sampler be geometrically ergodic?

**(c)** For $r = 1/20$, find a number $n$ such that $D(x, n) < 0.01$ for all $x \in \mathcal{X}$.

**(d)** Write and run a computer program to estimate $\mathbf{E}_\pi(h)$ with this algorithm in the two cases $r = 1/20$ and $r = 10$, each with $M = 10^5$ and $B = 10^4$. Estimate the corresponding standard errors by two different methods: (i) using "varfact", and (ii) from repeated independent runs.

**(e)** Discuss and compare the standard errors estimated by each of the two methods in each of the two cases, including discussion of which method is "better" for assessing uncertainty, and which case is a "better" sampling algorithm.

**3.** Let $\mathcal{X} = \mathbf{R}$, and let $\pi(x) = c\, g(x)$, where $g(x) = e^{-|x|/10}(1 + \cos(x)\sin(x^3))$, and let $h(x) = x + x^2$. With appropriate choice of $M$ and $B$ and $\sigma$ and starting distribution $\mathcal{L}(X_0)$, estimate $\mathbf{E}_\pi(h)$ in each of two different ways:

**(a)** With a usual random-walk Metropolis algorithm for $\pi$, with the usual proposal distributions $Y_n \sim N(X_{n-1}, \sigma^2)$.

**(b)** With a Langevin (Metropolis-Hastings) algorithm with proposals $Y_n \sim N(X_{n-1} + \frac{1}{2}\sigma^2\, g'(X_{n-1}) \,/\, g(X_{n-1}),\, \sigma^2)$. [Note: Here $g'(x)$ is the usual derivative of $g$, and should be computed analytically by you in advance and entered into your program.]

**(c)** Compare the two algorithms and discuss which one is "better".

**4.** Consider the standard variance components model described in lecture, with $K = 6$ and $J_i \equiv 5$, and $\{Y_{ij}\}$ the famous "dyestuff" data (from the file "Rdye"), with prior values $a_1 = a_2 = a_3 = b_1 = b_2 = b_3 = 100$. Estimate (as best as you can, together with a discussion of accuracy etc.) the posterior mean of $W/V$, in each of three ways:

**(a)** With a random-walk Metropolis algorithm.

**(b)** With a Metropolis-within-Gibbs algorithm.

**(c)** With a Gibbs sampler. [Note: first <u>derive</u> from scratch all of the conditional distributions, whether or not they were already described in lecture.]

**(d)** Finally, discuss the relative merits of all three algorithms for this example.

**5.** 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 $1 \le i \le 12$. 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 $S \sim \text{Poisson}(5)$ on $S$ (thus, $S$ is integer-valued).

**(a)** Specify (up to a normalising constant) the joint posterior distribution of $S, \theta_1, \ldots, \theta_{12}$.

**(b)** Run at least one MCMC algorithm of your choice for this posterior distribution, to estimate (as best as you can, together with standard errors), the posterior means of each of the 3 variables $S, \theta_1, \theta_2$.

**(c)** For $i = 1, 2$, compare the estimated posterior mean of $\theta_i$ to the value $H_i/A_i$. Are they different, and if so, how and why? Discuss.