Stat 260/CS 294 Homework Assignment 5 (due May 5)

1. If $K(x, x')$ is a Markov transition kernel with stationary distribution g, show that the Metropolis-Hastings algorithm where, at iteration t, the proposed value $y_t \sim K(x^{(t)}, y)$ is accepted with probability

$$\min\left(1, \frac{f(y_t)}{f(x^{(t)})}\frac{g(x^{(t)})}{g(y_t)}\right)$$

provides a valid MCMC algorithm for the stationary distribution f.

2. Consider the probit model, with $x_i \in \Re$ and $Y_i \in \{0, 1\}$:

$$\begin{aligned} P(Y_i = 1 \mid Z_i) &= P(Z_i \geq 0) \\ Z_i &\sim \mathrm{N}(x_i\beta, \sigma^2) \\ \beta_i &\sim \mathrm{N}(0, 10^2) \\ \sigma^{-2} &\sim \mathrm{Ga}(1.5, 1.5), \end{aligned}$$

where (x_i, y_i), $i = 1, \ldots, n$ are the observed data.

   (a) Given the data in `www.cs.berkeley.edu/~jordan/courses/260-spring10/data/probit.dat`, implement a Gibbs sampler for approximating the posterior of (β, σ). Initialize the sampler at (25, 5) and run it for 20,000 iterations (with no burn-in). Plot the samples on top of the contours of the true posterior.

   (b) Now consider an alternative sampler that inserts a Metropolis-Hastings step after each Gibbs cycle which scales the current state of the Markov chain, $(\beta^{(t)}, \sigma^{(t)})$, by a factor s which is drawn from an exponential distribution, Exp(1). The rescaled state is accepted or rejected according to the usual Metropolis-Hastings procedure. Implement this sampler and redo the simulation from part (a), again plotting the samples on top of the contours of the true posterior. Which sampler mixes more quickly?

3. Let $(X^{(t)})$ be a Markov chain with transition kernel q. Derive the following:

$$\mathrm{var}\left(\sum_{t=1}^{T} h(X^{(t)})\frac{f(X^{(t)})}{q(X^{(t)} \mid X^{(t-1)})}\right) = \sum_{t=1}^{T} \mathrm{var}\left(h(X^{(t)})\frac{f(X^{(t)})}{q(X^{(t)} \mid X^{(t-1)})}\right).$$

We see that for importance sampling estimators, the importance-weighted terms are uncorrelated even if the samples are correlated.

4. The following two derivations will exercise your skills with the Dirichlet distribution.

   (a) Define independent random variables $U \sim \mathrm{Dir}(\alpha)$ and $V \sim \mathrm{Dir}(\gamma)$. Let $W \sim \mathrm{Beta}(\sum_i \alpha_i, \sum_i \gamma_i)$, independently of U and V. Prove that

$$WU + (1 - W)V \sim \mathrm{Dir}(\alpha + \gamma).$$

   (b) Let e_j denote a unit basis vector. Let $\beta_j = \gamma_j / \sum_i \gamma_i$. Prove that

$$\sum_j \beta_j \mathrm{Dir}(\gamma + e_j) = \mathrm{Dir}(\gamma).$$

5. Show that the expectation of the total number of occupied tables in the Chinese restaurant scales as $O(\alpha n^d)$ under the Pitman-Yor process, $\mathrm{PY}(d, \alpha, G_0)$.

6. Consider the following DP mixture-of-Gaussians model:

$$\begin{aligned} y_i \,|\, \theta_i, \phi &\sim \mathrm{N}(y_i; \theta_i, \phi), \qquad i = 1, \ldots, n \\ \theta_i \,|\, G &\sim G, \qquad i = 1, \ldots, n \\ G &\sim \mathrm{DP}(\alpha, \mathrm{N}(\mu, \tau^2)) \\ \phi &\sim \mathrm{IG}(a_\phi, b_\phi) \\ \mu &\sim \mathrm{N}(a_\mu, b_\mu) \\ \tau^2 &\sim \mathrm{IG}(a_{\tau^2}, b_{\tau^2}). \end{aligned}$$

Let $a_\phi = 1, b_\phi = 1, a_\mu = 0, b_\mu = 10, a_{\tau^2} = 1, b_{\tau^2} = 1$. The data in
`www.cs.berkeley.edu/~jordan/courses/260-spring10/data/mixture.dat`
were generated by sampling from this model for a particular choice of α.

(a) Implement a Gibbs sampler for this model, where the Gibbs state includes indicator variables that encode the allocation of data points to tables in the Chinese restaurant, as well as the θ parameters associated with the tables. Fix $\alpha = 1$. Run the sampler until it appears to have converged, and use the subsequent samples to plot a histogram of the posterior distribution of the number of occupied tables.

(b) Implement a Rao-Blackwellized Gibbs sampler for this model in which the θ parameters have been integrated out. Fix $\alpha = 1$. Run the sampler and plot a histogram of the posterior distribution of the number of occupied tables. Does this sampler mix more quickly than the basic sampler?

(c) Implement a sampler for a model that places a vague gamma prior on α using the augmentation discussed in class. Again plot a histogram of the posterior distribution of the number of occupied tables. Also plot a histogram of the posterior distribution of α. Explore the sensitivity of your results to the choice of parameters for the gamma distribution.

(d) Interpret your results.