Homework 9, Stat 515, Spring 2015

Due Wednesday, April 22, 2015 <u>**beginning of class**</u>

Make sure you submit R code to the Angel drop folder, using the same naming conventions as in the last homework, e.g. your name should appear in the name of the R program. Please pay attention to good programming style in order to receive full credit.

1. **Importance Sampling**
   Return to the univariate Poisson kernel density function (Yang, 2004; or see "wrapped Cauchy" in Levy (1939) and Wintner (1947)), which is defined as follows:

$$f(\theta;\mu,\rho) = \frac{1}{2\pi}\frac{1-\rho^2}{1-2\rho\cos(\theta-\mu)+\rho^2}, \quad \mu-\pi \le \theta \le \mu+\pi \tag{1}$$

   (a) Approximate the expectation $E(\theta^2)$ for $\mu = 3, \rho = 0.7$ using importance sampling. State the importance function you used and report associated Monte Carlo standard errors.

   (b) Approximate the expectation $E(\theta)$ as a function of $\rho \in (-1, 1)$, that is, find $E_\rho(\theta)$, with $\mu = 3$ using importance sampling. (a) Evaluate the expectation on a grid of ρ values ranging from -0.95 to +0.95 in increments of 0.02. You can use the `R` function `seq` to generate the grid of values. For example, `rhovals=seq(-.95,0.95,by=0.02)`. Estimate these expectations by *using only one set of samples from the importance function.* (a) State the importance function you used and how it satisfies requisite conditions for validity of the algorithm, (b) Plot the estimated expectation as a function of ρ, and (c) report Monte Carlo standard errors along with the estimate of the expectation at the 1st, 24th, 48th, 72nd and 96th value of ρ. It may be best to report this information in the form of a table.

2. **Importance Sampling for Bayesian inference**
   Suppose $Y_1, \ldots, Y_n|\alpha \sim$ Poisson($\exp(\alpha)$) (they are conditionally independent given α), and the prior distribution for α is $\alpha \sim N(0, 100)$. The data $Y_1, \ldots, Y_n$ are available at `http://www.stat.psu.edu/~mharan/515/hwdir/hw09.dat`. You can read the data by using the command: `ys = scan("http://www.stat.psu.edu/~mharan/515/hwdir/hw09.dat")` (directly from the website) OR `ys = scan("hw09.dat")`
   We are interested in the posterior distribution of α given the data, i.e., the conditional distribution of $\alpha|Y$.

   (a) Estimate the posterior expectation of α using importance sampling.

   (b) Estimate the posterior probability that α is greater than 2 using importance sampling.

   (c) Use importance sampling to approximate the posterior expectation of α if the prior distribution for α is changed to $\alpha \sim N(0, 5)$, using the same set of samples as above. What is a potential problem if you switched the order in which you did this, that is, you found an importance function that worked for $\alpha \sim N(0, 5)$ and then tried to use the same for $\alpha \sim N(0, 100)$?

   You need to report: (i) importance function you used (you may use more than one and report your results for all of them if you like), (ii) plot the estimate versus the Monte Carlo sample size to see how your estimate converges as a function of the Monte Carlo sample size. Do this for at least $n = 100$ to 5,000, in increments of 100, and (iii) plot the Monte Carlo standard error as a function of n in similar fashion.

3. **Univariate Metropolis-Hastings**
   Return to the Bayesian inference problem above, but now change the prior so $\alpha \sim t_3(0, 100)$, a t-density with $\nu = 3$ degrees of freedom and $\sigma^2 = 100$ (so variance=$\nu\sigma^2/(\nu - 2)$).

(a) Simulate draws from $\alpha \mid Y$ by using a random walk Metropolis-Hastings algorithm with a normal proposal. Try at least two different values for the variance of the random walk proposal and compare your results for the different algorithms. In particular:

   i. Observe how the estimate of the expected value of $\alpha \mid Y$ and $\alpha^2 \mid Y$ change over time by plotting the estimate versus number of iterations (similar to what you did for importance sampling). You can plot both the estimates in the same figure by using the "lines" command in `R`.
   You may choose to use the function `estvssamp` in `http://www.stat.psu.edu/~mharan/batchmeans.R`.

   ii. Check the autocorrelation in your samples (not your estimates). For this you can use the `acf` function in `R`. How is the autocorrelation affected by different values of the variance of the random walk proposal?

   iii. Report the acceptance rates of the algorithm for different values of σ^2.

(b) Draw a smoothed estimated density plot based on the draws.

(c) Using these draws estimate the posterior expectation of α and α^2. Report MCMC standard errors for your estimates. Draw enough samples so you feel comfortable that your estimates appear to be stabilizing, and that your Monte Carlo standard errors are reasonable. For MCMC, you can use the **`consistent batchmeans`** procedure to estimate standard errors. The corresponding R code is available here: `http://www.stat.psu.edu/~mharan/batchmeans.R`.