Penn State STAT 540
Homework #2, due Wednesday, October 7, 2015

Submit: (i) `R` code in a file titled PSUemailidHW1.R (e.g. muh10HW1.R), (ii) pdf file that contains a clear writeup for the questions below named PSUemailidHW1.pdf Note that your code should be readily usable. It should also be well commented.

1. Use the ratio of uniforms approach to generate random variates according to a GEV($\mu = 3, \sigma = 0.8, \xi = 0.4$) distribution, with pdf given by

$$f(x) = \exp\left(- \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right) \times \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1-1/\xi}, \ \mu, \xi \in \mathcal{R}, \sigma > 0.$$

   (a) Provide pseudocode for the algorithm.
   (b) Your `R` code for the algorithm.
   (c) Plot the ratio of uniforms region.
   (d) Approximate the expected value. Use a Monte Carlo sample size such that the Monte Carlo standard error is less than 5% of the estimated expectation. Report your final approximation and the Monte Carlo sample size. Plot how the estimate converges as the Monte Carlo sample size increases. Also plot how the Monte Carlo standard error decreases with an increase in sample size.
   (e) Report the number of samples per second.

2. MCMC for GEV

   (a) Construct a Metropolis-Hastings algorithm for the above distribution. Provide your pseudocode and submit your `R` code.
   (b) Approximate the expected value for a random variable. Use a Monte Carlo sample size such that the MCMC standard error is less than 5% of the estimated expectation.
   (c) Report anything else you used to confirm that your approximation is reliable.
   (d) Compare the effective samples per second versus the iid sampler.

3. The m-dimensional multivariate truncated normal random with $\boldsymbol{x} = (x_1, \ldots, x_m)$, has pdf $f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma, \boldsymbol{a}, \boldsymbol{b})$

$$= \frac{\exp\left(-0.5(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)}{C(\boldsymbol{x}, \boldsymbol{\mu}, \Sigma)} I(x_i \in (a_i, b_i), \ i = 1, \ldots, n),$$

   with $C(\boldsymbol{x}, \boldsymbol{\mu}, \Sigma)$ the normalizing function of this pdf, Σ positive definite $m \times m$ covariance matrix, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ is the mean vector, the truncation constants are $\boldsymbol{a} = (a_1, \ldots, a_m)$, $\boldsymbol{b} = (b_1, \ldots, b_m)$.

(a) Construct an iid sampler for the 3-dimensional truncated normal random variable with $\boldsymbol{\mu} = (0, 1, 0)$, covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.8 & 0.3 \\ 0.8 & 2 & 0.4 \\ 0.3 & 0.4 & 3 \end{bmatrix},$$

and $\boldsymbol{a} = (-1, 2, 3), \boldsymbol{b} = (1, 4, 5)$. Provide pseudocode in your writeup and submit your `R` code.

(b) Approximate the expected value of this random variable using the above algorithm. Provide the Monte Carlo standard error of your approximation. Also provide anything else you used to verify your results along with a brief explanation. Any plots you include should be well labeled.

(c) What is the number of samples generated per second by your algorithm?

4. Importance sampling

   (a) Construct an importance sampler to approximate the expected value of the random variable from the previous problem. Provide pseudocode in your writeup and submit your `R` code.

   (b) Approximate the expected value of this random variable using the above algorithm. Provide the Monte Carlo standard error of your approximation. Also provide anything else you used to verify your results along with a brief explanation. Any plots you include should be well labeled.

   (c) What is the number of effective samples obtained for every 10,000 samples generated? What is the number of effective samples generated per second by your algorithm?

5. MCMC

   (a) Construct a Metropolis-Hastings algorithm for this distribution. Provide pseudocode in your writeup and submit your `R` code.

   (b) Approximate the expected value of this random variable using the above algorithm. Provide the Markov chain Monte Carlo standard error of your approximation. Also provide anything else you used to verify your results along with a brief explanation. Any plots you include should be well labeled.

   (c) What is the number of effective samples obtained for every 10,000 samples generated? What is the number of effective samples generated per second by your algorithm?

   (d) Summarize in a single table: your mean estimates, the corresponding Monte Carlo (or MCMC) standard errors, the effective sample size for 10,000 samples and the effective samples per second for each algorithm in problems 3, 4, and 5.