Assignment #2 STA410H1F/2102H1F

due Wednesday October 18, 2017

Instructions: Solutions to problems 1 and 2 are to be submitted on Blackboard (PDF files strongly preferred). You are strongly encouraged to do problems 3–5 but these are not to be submitted for grading.

1. An interesting variation of rejection sampling is the ratio of uniforms method. We start by taking a bounded function h with $h(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} h(x)\,dx < \infty$. We then define the region

$$\mathcal{C}_h = \left\{(u,v) : 0 \leq u \leq \sqrt{h(v/u)}\right\}$$

and generate (U,V) uniformly distributed on $\mathcal{C}_h$. We then define the random variable $X = V/U$.

(a) The joint density of (U,V) is

$$f(u,v) = \frac{1}{|\mathcal{C}_h|} \quad \text{for } (u,v) \in \mathcal{C}_h$$

where $|\mathcal{C}_h|$ is the area of $\mathcal{C}_h$. Show that the joint density of (U,X) is

$$g(u,x) = \frac{u}{|\mathcal{C}_h|} \quad \text{for } 0 \leq u \leq \sqrt{h(x)}$$

and that the density of X is $\gamma\, h(x)$ for some $\gamma > 0$.

(b) The implementation of this method is somewhat complicated by the fact that it is typically difficult to sample (U,V) from a uniform distribution on $\mathcal{C}_h$. However, it is usually possible to find a rectangle of the form $\mathcal{D}_h = \{(u,v) : 0 \leq u \leq u_+, v_- \leq v \leq v_+\}$ such that $\mathcal{C}_h$ is contained within $\mathcal{D}_h$. Thus to draw (U,V) from a uniform distribution on $\mathcal{C}_h$, we can use rejection sampling where we draw proposals (U^*,V^*) from a uniform distribution on the rectangle $\mathcal{D}_h$; note that the proposals U^* and V^* are independent random variables with $\text{Unif}(0,u_+)$ and $\text{Unif}(v_-,v_+)$ distributions, respectively. Show that we can define u_+, v_- and v_+ as follows:

$$u_+ = \max_x \sqrt{h(x)} \quad v_- = \min_x x\sqrt{h(x)} \quad v_+ = \max_x x\sqrt{h(x)}.$$

(Hint: It suffices to show that if $(u,v) \in \mathcal{C}_h$ then $(u,v) \in \mathcal{D}_h$ where $\mathcal{D}_h$ is defined using u_+, v_-, and v_+ above.)

(c) Implement (in R) the method above for the standard normal distribution taking $h(x) = \exp(-x^2/2)$. In this case, $u_+ = 1$, $v_- = -\sqrt{2/e} = -0.8577639$, and $v_+ = \sqrt{2/e} = 0.8577639$. What is the probability that proposals are accepted?

2. Suppose we observe $y_1, \cdots, y_n$ where

$$y_i = \theta_i + \varepsilon_i \quad (i = 1, \cdots, n)$$

where $\{\varepsilon_i\}$ is a sequence of random variables with mean 0 and finite variance representing noise. We will assume that $\theta_1, \cdots, \theta_n$ are dependent in the sense that $|\theta_i - \theta_{i-1}|$ is small for most values of i. The least squares estimates of $\theta_1, \cdots, \theta_n$ are trivial — $\widehat{\theta}_i = y_i$ for all i — but we can modify least squares in a number of ways to accommodate the "smoothness" assumption on $\{\theta_i\}$. In this problem, we will consider estimating $\{\theta_i\}$ by minimizing

$$\sum_{i=1}^{n}(y_i - \theta_i)^2 + \lambda \sum_{i=2}^{n}(\theta_i - \theta_{i-1})^2$$

where $\lambda > 0$ is a tuning parameter. (The term $\lambda \sum_{i=2}^{n}(\theta_i - \theta_{i-1})^2$ is a "roughness" penalty. In practice, it is more common to use $\lambda \sum_{i=2}^{n}|\theta_i - \theta_{i-1}|$ as it better estimates jumps in $\{\theta_i\}$.)

(a) Show the estimates $\widehat{\theta}_1, \cdots, \widehat{\theta}_n$ satisfy the equations

$$\begin{aligned}
y_1 &= (1+\lambda)\widehat{\theta}_1 - \lambda\widehat{\theta}_2 \\
y_j &= -\lambda\widehat{\theta}_{j-1} + (1+2\lambda)\widehat{\theta}_j - \lambda\widehat{\theta}_{j+1} \quad (j = 2, \cdots, n-1) \\
y_n &= -\lambda\widehat{\theta}_{n-1} + (1+\lambda)\widehat{\theta}_n.
\end{aligned}$$

(Note that if we write this in matrix form $\mathbf{y} = A_\lambda\widehat{\boldsymbol{\theta}}$, the matrix A is sparse.)

(b) Show (using results from class) that both the Jacobi and Gauss-Seidel algorithms can be used to compute the estimates defined in part (a).

(c) Write a function in R to implement the Gauss-Seidel algorithm above. The inputs for this function are a vector of responses `y` and the tuning parameter `lambda`. Test your function (for various tuning parameters) on data generated by the following command:

```
> y <- c(rep(0,250),rep(1,250),rep(0,50),rep(1,450)) + rnorm(1000,0,0.1)
```

How does varying λ change the estimates, particularly the estimates of the transitions from 0 to 1 in the step function?

Supplemental problems (not to hand in):

3. To generate random variables from some distributions, we need to generate two independent two independent random variables Y and V where Y has a uniform distribution over some finite set and V has a uniform distribution on $[0, 1]$. It turns out that Y and V can be generated from a single Unif$(0, 1)$ random variable U.

(a) Suppose for simplicity that the finite set is $\{0, 1, \cdots, n-1\}$ for some integer $n \geq 2$. For $U \sim \text{Unif}(0, 1)$, define

$$Y = \lfloor nU \rfloor \quad \text{and} \quad V = nU - Y$$

where $\lfloor x \rfloor$ is the integer part of x. Show that Y has a uniform distribution on the set $\{0, 1, \cdots, n-1\}$, V has a uniform distribution on $[0, 1]$, and Y and V are independent.
(b) What happens to the precision of V defined in part (a) as n increases? (For example, if U has 16 decimal digits and $n = 10^6$, how many decimal digits will V have?) Is the method in part (a) particularly feasible if n is very large?

4. One issue with rejection sampling is a lack of efficiency due to the rejection of random variables generated from the proposal density. An alternative is the acceptance-complement (A-C) method described below.
Suppose we want to generate a continuous random variable from a density $f(x)$ and that $f(x) = f_1(x) + f_2(x)$ (where both f_1 and f_2 are non-negative) where $f_1(x) \leq g(x)$ for some density function g. Then the A-C method works as follows:

1. Generate two independent random variables $U \sim \text{Unif}(0, 1)$ and X with density g.
2. If $U \leq f_1(X)/g(X)$ then return X.
3. Otherwise (that is, if $U > f_1(X)/g(X)$) generate X from the density

$$f_2^*(x) = \frac{f_2(x)}{\displaystyle\int_{-\infty}^{\infty} f_2(t)\, dt}.$$

Note that we must be able to easily sample from g and f_2^* in order for the A-C method to be efficient; in some cases, they can both be taken to be uniform distributions.
(a) Show that the A-C method generates a random variable with a density f. What is the probability that the X generated in step 1 (from g) is "rejected" in step 2?
(b) Suppose we want to sample from the truncated Cauchy density

$$f(x) = \frac{2}{\pi(1 + x^2)} \quad (-1 \leq x \leq 1)$$

using the A-C method with $f_2(x) = k$, a constant, for $-1 \leq x \leq 1$ (so that $f_2^*(x) = 1/2$ is a uniform density on $[-1, 1]$) with

$$f_1(x) = f(x) - f_2(x) = f(x) - k \quad (-1 \leq x \leq 1).$$

If $g(x)$ is also a uniform density on $[-1, 1]$ for what range of values of k can the A-C method be applied?

(c) Defining f_1, f_2, and g as in part (b), what value of k minimizes the probability that X generated in step 1 of the A-C algorithm is rejected?

5. Suppose we want to generate a random variable X from the tail of a standard normal distribution, that is, a normal distribution conditioned to be greater than some constant b. The density in question is

$$f(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}(1 - \Phi(b))} \quad \text{for } x \geq b$$

with $f(x) = 0$ for $x < b$ where $\Phi(x)$ is the standard normal distribution function. Consider rejection sampling using the shifted exponential proposal density

$$g(x) = b\exp(-b(x - b)) \quad \text{for } x \geq b.$$

(a) Define Y be an exponential random variable with mean 1 and U be a uniform random variable on $[0, 1]$ independent of Y. Show that the rejection sampling scheme defines $X = b + Y/b$ if

$$-2\ln(U) \geq \frac{Y^2}{b^2}.$$

(Hint: Note that $b + Y/b$ has density g.)
(b) Show the probability of acceptance is given by

$$\frac{\sqrt{2\pi}b(1 - \Phi(b))}{\exp(-b^2/2)}.$$

What happens to this probability for large values of b? (Hint: You need to evaluate $M = \max f(x)/g(x)$.)
(c) Suppose we replace the proposal density g defined above by

$$g_\lambda(x) = \lambda\exp(-\lambda(x - b)) \quad \text{for } x \geq b.$$

(Note that g_λ is also a shifted exponential density.) What value of λ maximizes the probability of acceptance? (Hint: Note that you are trying to solve the problem

$$\min_{\lambda>0} \max_{x \geq b} \frac{f(x)}{g_\lambda(x)}$$

for λ. Because the density $g_\lambda(x)$ has heavier tails, the minimax problem above will have the same solution as the maximin problem

$$\max_{x \geq b} \min_{\lambda>0} \frac{f(x)}{g_\lambda(x)}$$

which may be easier to solve.)