Finger exercises for Week 13 — To be discussed Monday Nov 25, 2002

1. We continue with the data Y

   ```
   1.2 2.4 1.3 1.3 0.0 1.0 1.8 0.8 4.6 1.4 ,
   ```

   modeled as iid observations y from a Cauchy likelihood

   $$f(y|\theta, s) = \frac{1}{\pi s} \frac{1}{1 + \left(\frac{y-\theta}{s}\right)^2}$$

   with s fixed. Assume now that $s = 0.1$ and that θ has a prior distribution, $\theta \sim N(1, 1)$.

   (a) Graph the log-posterior density $\log p(\theta|Y)$ (or something log-proportional to it) for the Cauchy data above.

   (b) Compute an approximate 95% credible interval for θ, using the posterior mode $\tilde{\theta}$ and $-\frac{d^2}{d\theta^2} \log p(\tilde{\theta}|Y)$, using whatever tools are convenient (optimization functions in R or Splus, numerical or symbolic differentiation, etc.)

   (c) Use the R/Splus function `integrate` to calculate the coverage probability for the interval calculated in part 1b. In other words, calculate $P(\theta \in [a, b]|Y)$, where $[a, b]$ is the interval you calculated in part 1b.

   (d) Try calculating $E[\theta|Y]$ and $\sqrt{\text{Var}(\theta|Y)}$ using SPLUS's `integrate` function. Is the approximate 95% credible interval based on $E[\theta|Y]$ and $\sqrt{\text{Var}(\theta|Y)}$ much different from the interval in part 1b? Is the coverage probability much different?

   (e) To explore what happens to numerical integrals when the likelihood becomes more and more peaked, recalculate $E[\theta|Y]$ and $\sqrt{\text{Var}(\theta|Y)}$, when the dataset Y is replaced with

   - The same dataset, with each observation repeated 2 times.
   - The same dataset, with each observation repeated 3 times.
   - The same dataset, with each observation repeated 5 times.
   - The same dataset, with each observation repeated 10 times.
   - The same dataset, with each observation repeated 20 times.
   - The same dataset, with each observation repeated 30 times.

   and describe/explain your findings.

2. The inverse-CDF method says that if $U \sim Unif[0, 1]$ then $X = F^{-1}(U)$ will have CDF $P[X \leq x] = F(x)$. If X is a discrete random variable taking on the values $0, 1, 2, \ldots$ then $F(x)$ is a step function: $F(x) \equiv F(k)$ for all $x \in [k, k+1)$. This suggests a "search" method of generating X: First generate $U \sim Unif[0, 1]$; then find the k such that $F(k) \leq U < F(k + 1)$, and return $X = k$.

   (a) Show that this "search" method produces a random draw from $X \sim F(x)$.

(b) Implement the method in an Splus or R function to generate random draws from a Poisson(λ) random variable. Show that it works.

(c) Discuss at least two drawbacks of the method.

3. In the rejection methods, on average, how many pairs (Y, U) have to be drawn, before Y is accepted as an observation from the "target" density $f(x)$? What does this say about what the relationship between the "target" density $f(x)$ and the "convenient" density $g(x)$ should be?

4. Show that you can generate $X \sim N(0,1)$ by applying the rejection method to Cauchy observations $Y = \tan(\pi(U - 1/2))$. What is (the ***best***) a? Show (analytically) that it does ***not*** work to generate $X \sim Cauchy$ by applying the rejection method to $N(0,1)$ observations Y. *What goes wrong???*

5. Can you develop a rejection method for discrete random variables (the method for continuous rv's was discussed in class). Does the "convenient" density $g(x)$ have to be for a discrete random variable?

6. In the notes and in class we discussed two versions of importance sampling, one involving

$$\tilde{I}_n = \frac{1}{n}\sum_1^n h(Y_i)f(Y_i)/g(Y_i) \to E_f[h(X)]\,, \quad \text{as } n \to \infty$$

and the other involving $\hat{I}_n = \tilde{I}_n/\tilde{O}_n$ where

$$\tilde{O}_n = \frac{1}{n}\sum_1^n 1 \cdot f(Y_i)/g(Y_i) \to 1\,, \quad \text{as } n \to \infty$$

Apply the delta method to the function $g(x, y) = x/y$, where $x = \frac{1}{n}\tilde{I}_n$ and $y = \frac{1}{n}\tilde{O}_n$, to show that

$$\text{Var}(\hat{I}_n) \approx \text{Var}(\tilde{I}_n) - 2E_f[h(X)]\text{Cov}(\tilde{I}_n, \tilde{O}_n) + (E_f[h(X)])^2\text{Var}(\tilde{O}_n)$$

Under what conditions could we expect $\text{Var}(\hat{I}_n) \leq \text{Var}(\tilde{I}_n)$?