Problem 1

Lemma 2.4.1 *If*

$$f(x) = \frac{f_1(x) - \varepsilon f_2(x)}{1 - \varepsilon}$$

where f_1 and f_2 are probability densities such that $f_1(x) \geq \varepsilon f_2(x)$, the algorithm

Generate

$$(X, U) \sim f_1(x)\, \mathbf{I}_{[0,1]}(u)$$

until $U \geq \varepsilon f_2(X)/f_1(X)$

produces a variable X distributed according to f.

(a) Show that the distribution of X satisfies

$$P(X \leq x_0) = \int_{-\infty}^{x_0} \left(1 - \frac{\varepsilon f_2(x)}{f_1(x)} \right) f_1(x)\, dx \sum_{i=0}^{\infty} \varepsilon^i \ .$$

(b) Evaluate the integral in (a) to complete the proof.

Problem 2

For the Accept–Reject algorithm [A.5], with f and g properly normalized,

(a) Show that the probability of accepting a random variable is

$$P\left(U < \frac{f(X)}{Mg(X)} \right) = \frac{1}{M}.$$

(b) Show that $M \geq 1$.

(c) Show that the bound M does not have to be tight; that is, there may be $M' < M$ such that $f(x) \leq M'g(x)$. Give an example where it makes sense to use M instead of M'.

(d) When the bound M is too tight (i.e., when $f(x) > Mg(x)$ on a non-negligible part of the support of f), show that the algorithm [A.5] does not produce a generation from f. Give the resulting distribution.

(e) When the bound is not tight, show that there is a way, using Lemma 2.4.1, to recycle part of the rejected random variables (*Note:* See Casella and Robert 1998 for details.)

Problem 3

Generate 100 standard normal random variables and use these values to perform rejection sampling to calculate the integral of the normal probability density function from 2 to infinity. Repeat this experiment 100 times and calculate the empirical mean and variance of your estimate. Now generate 100 Cauchy values and use the importance sampling to obtain an estimate of the same integral. Repeat this second experiment 100 times and again calculate the empirical mean and variance. Which approach is more accurate?

Problem 4

For the genetic linkage model: Use the EM algorithm to obtain the MLE, θ^*, of θ. Then use this MLE θ^* as the true parameter to simulate the original multinomial distribution with four cells and the total sample size n = 125+18+20+34=197 to obtain a new sample (y_1, y_2, y_3, y_4). Use the EM algorithm to obtain the new MLE based on this set of data. Repeat this simulation experiment 400 times. Each time use θ^* as the true parameter to simulate the original multinomial distribution with n = 197. Calculate the sample standard deviation of these 400 new MLE's of θ.

Problem 5

For the genetic linkage model:

(a) Use importance sampling to obtain the posterior mean and posterior standard deviation for the data Y = (125, 18, 20, 34). Use the matching normal distribution as the importance function. Compare your importance sampling estimates of the posterior mean and posterior standard deviation to those obtained via Laplace's method. Draw the histogram of the weights and compute their standard deviation.

(b) Repeat (a) for the data Y = (14, 0, 1, 5). Compare the histograms of the weights for both data sets. Discuss the adequacy of the importance sampling for each data set.

(c) Repeat (a) and (b) using the matching beta distribution as the importance function.