EN.553.732 Homework 3

Joseph High Hopkins ID: 9E1FDC

October 17, 2017

Problem 1.
Here, we are implementing an importance sampler and let $g(x) \sim Normal(0,1)$.
We can then compute the expected value of the mixture of beta distributions by using the criteria

$$E(x) = \frac{E_g(x(f(x)/g(x))}{E_g(f(x)/g(x))}$$

From the corresponding R code (attached) it was found that the expected value is 0.3637159.
Moreover, the probability that the random variable is within the interval (0.35,0.55) was found to be 0.1205.
R code and results are attached

Problem 2. *Proof.*
W.T.S: $P(Y < y) = P(X < x|U < f(x))$
We first generate $X \sim g$ and $U|X = x \sim U_{[0,Mg(x)]}$
Then, the pdf of U is $P(U|X = x) = \dfrac{1}{Mg(x)}$
Then,

$$P(U < f(x)) = E(P(U < f(x)|x)) = E(\int_0^{f(x)} \frac{1}{Mg(x)} du) = E(\frac{f(x)}{Mg(x)}) = \int_{-\infty}^{\infty} \frac{f(x)}{Mg(x)} g(x)dx$$

$$= \frac{1}{M} \int_{-\infty}^{\infty} f(x)dx = \frac{1}{M} \qquad \text{(since } f(x) \text{ a pdf)}$$

Similarly,

$$P(U < f(x), X < x) = \int_{-\infty}^{x} P(U < f(x)g(x)dx = \int_{-\infty}^{x} \frac{f(x)}{Mg(x)} g(x)dx$$
$$= \frac{1}{M} \int_{-\infty}^{x} f(x)dx = \frac{F(x)}{M}.$$

Finally,

$$P(X < x|U < f(x)) = \frac{P(U < f(x), X < x)}{P(U < f(x))} = \frac{F(x)/M}{1/M} = F(x) = P(X < x) = P(Y < y)$$

Proving that the given algorithm is analogous to the standard Accept-Reject algorithm. □

Problem 3.
Under the assumption that μ and τ are independent, we can write their joint prior distribution as

$$p(\mu, \tau) \sim Beta(2, 2)Lognormal(1, 10) \ .$$

The data likelihood of X is

$$p(X|\mu, \tau) \sim \prod_{i=1}^{n} Normal(\mu, \tau)$$

The posterior distribution of μ and τ is proportional to the product of the likelihood and prior, so we have:

$$p(\mu,\tau|X) \propto Beta(\mu;2,2)Lognormal(\tau;1,10)\prod_{i=1}^{n} Normal(x_i;\mu,\tau)$$

Assuming that the proposal distribution is chosen to be symmetric, we have the Metropolis algorithm. With t iterations, there are two possibilities for $\mu^{(t)}$ and $\tau^{(t)}$: $\mu^{(t)} = \mu^*$ and $\tau^{(t)} = \tau^*$ with probability θ and $\mu^{(t)} = \mu^{(t-1)}$ and $\tau^{(t)} = \tau^{(t-1)}$ with probability $1-\theta$, where

$$\theta = \min(1, \frac{p(\mu^*,\tau^*|X)}{p(\mu^{(t)},\tau^{(t)}|X)})$$

This is a result of the symmetry of the proposal distribution, i.e. $q(y|z) = q(z|y)$
The posterior probability was found to be $P(\mu \le 0.5|X) = 0.82797$
The trace plots indicate convergence and the ACF plots drop precipitously over time, the desired result.
R code and results are attached

Problem 4.
Part a The respective R code, outputs, and graphics are attached. It can be seen that the empirical distribution skews to the right, and so it deviates from a normal distribution.

Part b
Solution Reference: Hoff, Peter D. (2009). *A First Course in Bayesian Statistical Methods.* New York, NY: Springer.

Denote $n_1 = \sum_{\{i:X_i=1\}} 1,\ n_2 = \sum_{\{i:X_i=2\}} 1,\ n = n_1 + n_2$
$\mathrm{Y}_1 = \sum_{\{i:X_i=1\}} y_i,\ \mathrm{Y}_2 = \sum_{\{i:X_i=2\}} y_i.$

we then have,
$p(X_i|p,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ Y) = \dfrac{p\times normal(\theta_1,\sigma_1^2)}{p\times normal(\theta_1,\sigma_1^2) + (1-p)\times normal(\theta_2,\sigma_2^2)},\ i = 1,\ \cdots\ n$
$p(X|p,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ Y) = \prod_{i=1}^n p(X_i|p,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ Y)$
$p(p|X,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ Y) \sim beta(a+n_1,\ b+n_2)$
$p(\theta_1|X,p,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ Y) \sim normal(\mu_n,\ \tau_n^2)$, where $\mu_n = \frac{\mu_0/\tau_0^2 + y_1/\sigma_1^2}{1/\tau_0^2 + n_1/\sigma_1^2}$ and $\tau_n^2 = \frac{1}{1/\tau_0^2 + n_1/\sigma_1^2}$
$p(\theta_2|X,p,\ \theta_1,\ \sigma_1^2,\ \sigma_2^2,\ Y) \sim normal(\mu_n,\ \tau_n^2)$, where $\mu_n = \frac{\mu_0/\tau_0^2 + y_2/\sigma_2^2}{1/\tau_0^2 + n_2/\sigma_2^2}$ and $\tau_n^2 = \frac{1}{1/\tau_0^2 + n_2/\sigma_2^2}$
$p(\sigma_1^2|X,p,\ \theta_1,\ \theta_2,\ \sigma_2^2,\ Y) \sim inverse-gamma(\nu_n/2,\ \nu_n\sigma_n^2/2)$, where $\nu_n = \nu_0 + n_1$ and $\sigma_n^2 = \frac{1}{\nu_n}(\nu_0\sigma_0^2 + \sum_{\{i:X_i=1\}}(y_i-\theta_1)^2)$
$p(\sigma_2^2|X,p,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ Y) \sim inverse-gamma(\nu_n/2,\ \nu_n\sigma_n^2/2)$, where $\nu_n = \nu_0 + n_2$ and $\sigma_n^2 = \frac{1}{\nu_n}(\nu_0\sigma_0^2 + \sum_{\{i:X_i=2\}}(y_i-\theta_2)^2)$

Part c Referring to the R code, plots and outputs, the ACF has a steep decline with respect to the increase in time - a desired result. The effective sample size for $\theta_{(1)}^{(s)}$ was found to be 418.4169 while the effective sample size for $\theta_{(2)}^{(s)}$ was found to be 230.2658.
R code and results are attached

Part d From the histogram (attached) and the density found in part a, we see that they are very similar. In particular, both are right skewed.

Problem 5.
Part 1 We first note that the likelihood of $\theta = (\theta_1,\theta_2,\theta_3,\theta_4)$ is

$$\mathcal{L}(\theta|x,n,y) = \prod_{j=1}^{4}(logit^{-1}(\alpha+\beta x_j))^{y_j}(1-logit^{-1}(\alpha+\beta x_j))^{n_j-y_j}$$

where $x, n,$ and y represent the respective data vectors. For parts 1-4, α and βhave $Normal(0,10^2)$ prior

distributions. However, for part 1, we let $\beta = 10$.
So for the posterior, we obtain:

$$p(\theta|x,n,y) \propto Normal(\alpha;0,10^2)\prod_{j=1}^{4}(logit^{-1}(\alpha+\beta x_j))^{y_j}(1-logit^{-1}(\alpha+\beta x_j))^{n_j-y_j}$$

Of course, α could change depending on the iteration. Note that we are picking symmetric normal jumps for α. For the entirety of this problem, let us pick symmetric normal jumps of $Norm(0,1)$. So then, $\alpha^* = \alpha^{(t-1)} + \varepsilon$, where $\varepsilon \sim Norm(0,1)$
In other words, we can say $\alpha^* \sim Normal(\alpha^{(t-1)}, 1)$
R code and results are attached.
Part 2 This part is similar to part 1; however, now β is not fixed.
Since α and β independent, the posterior is

$$p(\theta|x,n,y) = Norm(\beta;0,10^2)Norm(\alpha;0,10^2)\prod_{j=1}^{4}(logit^{-1}(\alpha+\beta x_j))^{y_j}(1-logit^{-1}(\alpha+\beta x_j))^{n_j-y_j}$$

Similar to part a, we are picking symmetric normal jumps of $Norm(0,1)$, but this time we are iterating for α and β.
R code and results are attached.
Part 3 While similar to parts 1 and 2, there is a difference in how we pick the jump. We let $\theta^* \sim Normal(\theta^{(t-1)}, I)$, so we jump α and β together instead of separately as in part b.
R code and results are attached.
Part 4 This part was similar to parts 1 - 3. The difference is in how we jump, which moves us in the direction of the mode. This would presumably give us faster convergence. In accordance with the definition of $\theta^*|\theta^t$ given in the problem, we let $\delta = 1$ and the covariance matrix $= I$.
R code and results are attached.
Part 5 While the efficiency of part 2 and part 3 are similar, there are differences. In particular, in part 2, we jump α and β separately, while in part 3 we jump them simultaneously using a bivariate normal distribution. Referring to the ACF and trace plots, there is no significant improvement in efficiency for α. On the other hand, for β, the method in part 2 is noticeably more efficient in terms of convergence compared to the method in part 3. Indeed, if we refer to the corresponding ACF, it decreases faster for β in part 2 compared to that of part 3. The trace plot for β in part 2 also shows stronger convergence when compared to the trace plot for β in part 3. This was to be expected since in part 2 we first jump α and then β with α already updated, subsequently resulting in more efficient convergence for β. However, in part 3, this was not the case since we jumped them simultaneously.
The algorithm in part 4 is significantly more efficient than part 2 and 3, as expected. This can be observed in the ACF, which vanishes after only a few lag times. Moreover, the corresponding trace plot displays stronger efficiency and strong convergence. This was also to be expected since the algorithm in part 4 moves in the direction of the mode, resulting in faster convergence of our parameters.