STAT2651: Bayesian Inference 4/9/2021

HW4

Name: Kamaru-Deen Lawal
Email: kal246@pitt.edu

10.3

We ran the simulation with three different Beta priors. The results are below

- Beta(1,1) - Flat Prior
- Beta(.5, .5) - Jeffrey's Prior
- Beta(20,20) - Informative Prior

a) 95% CI : $[-0.256, 0.408]$ (Flat Prior)
95% CI : $[-0.262, 0.432]$ (Jeffrey's Prior)
95% CI : $[-0.164, 0.205]$ (Informative Prior)

b) $p(p1 > p2)$: .687 (Flat Prior)
$p(p1 > p2)$: .696 (Jeffrey's Prior)
$p(p1 > p2)$: .586 (Informative Prior)

10.4

a) Let $f(\theta) = p(\theta|y)$ and $A|\theta \sim \text{Bernoulli}(\frac{f\theta}{M \cdot g(\theta|y)})$. $A|\theta$ is an indicator RV that takes the value 1 with probability equivalent to the probability a draw of θ is accepted. Let X be a RV representing the rejection sampling. Let Y_f be a RV corresponding to the true probability distribution. Let Y_g e a RV corresponding to samples from $M \cdot g(\theta|y)$. We need to show that for any event E, $p(X \in E) = p(Y_f \in E)$. Using the definition of conditional probability.

$$P(X \in E) = P(Y_g \in E|A = 1) = \frac{P(Y_g \in E, A = 1)}{P(A = 1})$$

. If we examine the denominator, we see:

$$P(A = 1) = \int_{-\infty}^{\infty} P(A = 1|\theta)\, d\theta = \frac{1}{M}$$

Next we examine the numerator

$$P(Y_g \in E, A = 1) = \int_E P(A = 1|\theta) g(\theta)\, d\theta \tag{1}$$
$$= \int_E g(\theta) \frac{p(\theta|y)}{M g(\theta|y)}\, d\theta \tag{2}$$
$$= \int_E \frac{p(\theta|y)}{M}\, d\theta \tag{3}$$
$$= \frac{P(Y_f \in E)}{M} \tag{4}$$

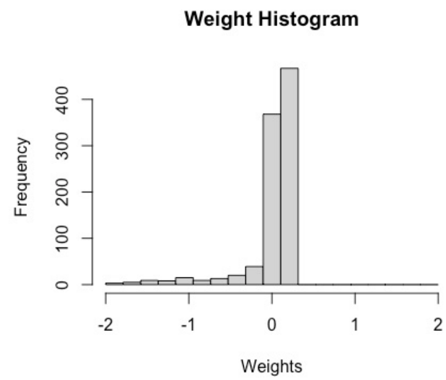
Plugging both of these values back into the numerator and denominator of the initial definition of $P(X \in E)$ we see that the sampling distribution is equal to the rejection sampling.

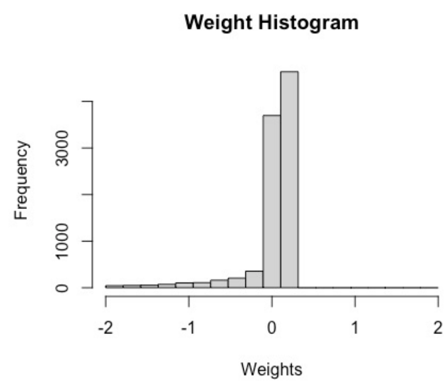
b) Let $b = \frac{p(\theta|y)}{q(\theta)}$ such that b is unbounded. $b \to \pm\infty$. b must go to either $-\infty$ or $+\infty$, and we know it can not go to $-\infty$ based on the definition of probability spaces. It is also clear that b can not approach $+\infty$ based on how normalizing constants work.

10.6

Below we simulate the true distribution $N(0,1)$ and the approximation with three D.o.F

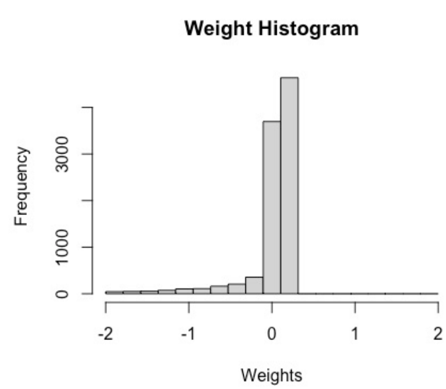
a) The following histogram includes the 95% highest density region, with outliers omitted for analysis of the higher density region.





b) Importance sampling tells us that $E[\theta|y] = -.04$ and $\text{var}(\theta|y) = .95$.

c) Below we see a histogram for the same with run with $S = 10,000$.

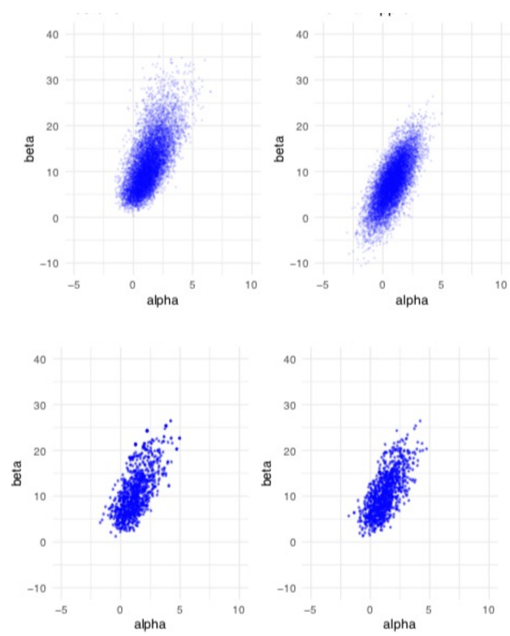


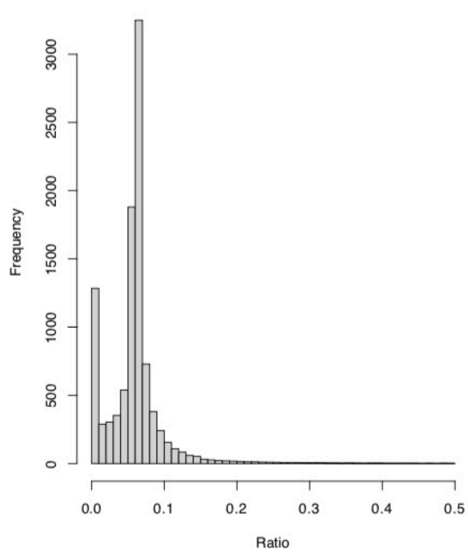
10.8

a) The samples from resampling with replacement approximate the highest density region of the distribution, but the variance is obviously much lower than that of the true posterior. We see the plots in the figure below our analysis of the parts.

b) Most of the calculated importances are either close to 0, or in the range [.03, .09]. A very small portion of the importance ratios were greater than .5. We see the plots in the figure below for our analysis of the parts.

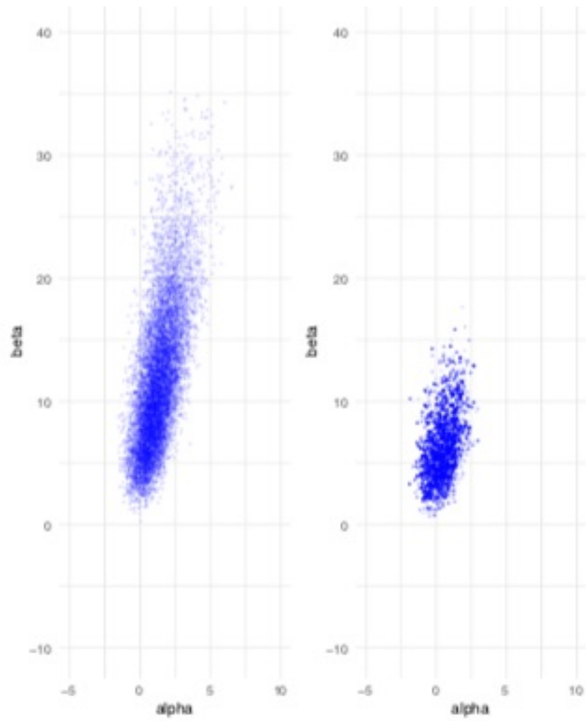
c) Sampling without replacement leads to a much smoother distribution that seems to have slightly higher variance. More or less the distributions look very similar.





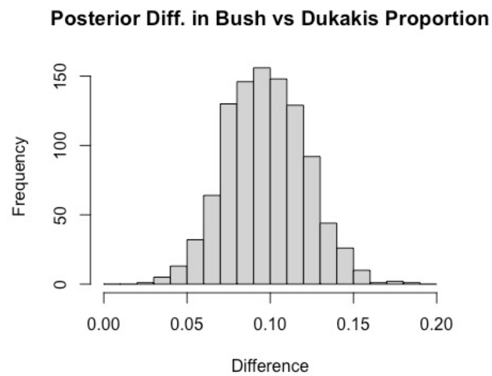
11.2

Below we see results for the simulation compared against the actual posterior. Simulation did a great job at capturing the highest density region of the posterior, but could not capture the lower density tails. After varying the parameter values it did not seem possible to capture this variance.



11.7

a) The results are very similar to the corresponding figure in the book.



b) The distributions for the hyper-parameters had stdev. .08, and the stdev. for the distribution of β was .05. Overall the resultant distribution that I produced using a jumping rule overlaps a lot with the the results in Figure 8.1b.

13.5

(a) $X_i \sim \text{Poisson}(\lambda_i)$; $\lambda_i \sim \text{Gamma}(\alpha, \beta)$, $p(\alpha, \beta) \propto 1$.

$$\Rightarrow p(x_i \mid \alpha, \beta) = \frac{p(x_i \mid \lambda_i; \alpha, \beta, y)\, p(\lambda_i \mid \alpha, \beta)}{p(\lambda_i \mid y, \alpha, \beta)} = \frac{\text{Poisson}(\cancel{x_i} \mid \lambda_i)\, \text{Gamma}(\lambda_i \mid \alpha, \beta)}{\text{Gamma}(\lambda_i \mid \alpha + y, \beta + 1)}$$

$$\propto \text{NegBin}(\alpha, \beta).$$

We also recall from the 2nd chapter that the negative binomial is a mixture of Poisson distributions that follow the gamma distribution, which is what we concluded from our analysis.

(b) X_i can be thought of as an independent trial that belongs to category y_i iff $X_i = i$. We also know that the prob. $X_i = i$ is equivalent to $\text{NegBin}(i \mid \alpha, \beta)$. Since we are given that we have exactly N species the # of independent trial is N. Therefore $y \sim \text{Mult}(N, p)$; $p = \{p_i \mid p_i \propto \text{NegBin}(i \mid \alpha, \beta), i \in [N]\}$.

See next page for 13.5 (c)

13.5

(c) $$p(Y \mid \alpha, \beta, N=24) = \frac{p(Y \mid X, \alpha, \beta, N)\, p(X \mid \alpha, \beta, N)}{p(X \mid Y, \alpha, \beta, N)}$$

$$= \frac{\mathrm{Mult}(24, p) \prod_i \mathrm{NegBin}(X_i \mid \alpha, \beta)}{p(X \mid Y, \alpha, \beta, N)}$$

$$p(X \mid Y, \alpha, \beta, N) = \frac{\prod \mathrm{NegBin}(X_i \mid \alpha, \beta)}{\mathrm{NegBin}\left(\sum X_i \mid N \alpha, \beta\right)} = \frac{\left(\frac{\beta}{1+\beta}\right)^{\sum_i \alpha} \left(\frac{1}{\beta+1}\right)^{\sum_i x_i} \prod_i \binom{\alpha + x_i - 1}{x_i}}{\left(\frac{\beta}{1+\beta}\right)^{N\alpha} \left(\frac{1}{\beta+1}\right)^{\sum_i x_i} \binom{N\alpha + \sum x_i - 1}{\sum_i x_i}}$$

$$= \frac{\prod_i \binom{\alpha + x_i - 1}{x_i}}{\binom{N\alpha + \sum_i x_i - 1}{\sum_i x_i}}$$

$$\propto \prod_i \frac{\Gamma(\alpha + x_i)}{\Gamma(1 + x_i)}$$

135

(c) continued.

We know the sum of the values of X_i, this is equivalent to a Dirichlet-Multinomial likelihood. w/ parameterization. $DM(x_1, \dots, x_n \mid \alpha_1, \dots, \alpha_n)$. Therefore.

we know the likelihood is given by $p(y \mid \alpha, \beta, N) = Mult(y \mid 24, p) \prod_i NegBin(x_i \mid \alpha, \beta)$

$$\overline{}$$

$$DM(\cancel{x} \mid \alpha_1, \dots, \alpha)$$

13.6.

If $E_{old} \log p(\gamma | \phi, y)$ is maximized at $\phi = \phi^{old}$, then we know that $E_{old} \log p(\gamma|\phi, y) \leq E_{old} \log p(\gamma | \phi^{old}, y)$ for all ϕ. This is all we need to show that $E_{old} \log p(\gamma|\phi, y) - E_{old} \log p(\gamma|\phi^{old}, y) \leq 0$

$$E_{old} \log p(\gamma|\phi, y) - E_{old} \log p(\gamma|\phi^{old}, y) = E_{old}\left[\log\left[\frac{p(\gamma|\phi, y)}{p(\gamma|\phi^{old}, y)}\right]\right]$$

by Jensen's Inequality we know that

$$E_{old}\left[\log \frac{p(\gamma|\phi, y)}{p(\gamma|\phi^{old}, y)}\right] \leq \log\left(E_{old}\left[\frac{p(\gamma|\phi, y)}{p(\gamma|\phi^{old}, y)}\right]\right)$$

$$RHS = \int p(\gamma|\phi^{old}, y) \frac{p(\gamma|\phi, y)}{p(\gamma|\phi^{old}, y)} d\gamma = \int p(\gamma|\phi, y) d\gamma.$$

Therefore

$$E_{old}\left[\log \frac{p(\gamma|\phi, y)}{p(\gamma|\phi^{old}, y)}\right] \leq \log\left[E_{old}\left[\frac{p(\gamma|\phi, y)}{p(\gamma|\phi^{old}, y)}\right]\right]$$

$$= \log\left(\int p(\gamma|\phi, y) d\gamma\right) \leq \log(1) = 0$$