Homework 5

Steve Harms

STAT 520

November 15, 2018

1 Part 1

1.1

Let $f(y_i|\theta) = \exp[\phi\{\theta y_i - b(\theta)\} + c(y,\phi)]$ and $\pi(\theta) = \exp[\lambda\{m\theta - b(\theta)\} + d(\lambda, m)]$. We can take the derivative of both sides of the prior to get

$\frac{\partial \pi}{\partial \theta} = \lambda\{m - b'(\theta)\}\pi(\theta)$

then integrate both sides of the above to get

$\int_{\Re} \frac{\partial \pi}{\partial \theta} d\theta = \int_{\Re} [\lambda\{m - b'(\theta)\}\pi(\theta)] d\theta$

$\int_{\Re} \frac{\partial \pi}{\partial \theta} d\theta = \lambda m \int_{\Re} \pi(\theta) d\theta - \lambda \int_{\Re} b'(\theta)\pi(\theta) d\theta$

Then we note that $\int_{\Re} \pi(\theta) d\theta = 1$ and exchange differentiation and integration on the left side to get

$\frac{\partial}{\partial \theta} \int_{\Re} \pi(\theta) d\theta = \lambda m \int_{\Re} \pi(\theta) d\theta - \lambda \int_{\Re} b'(\theta)\pi(\theta) d\theta$

$\frac{\partial}{\partial \theta}(1) = \lambda m(1) - \lambda \int_{\Re} b'(\theta)\pi(\theta) d\theta$

$0 = \lambda m(1) - \lambda \int_{\Re} b'(\theta)\pi(\theta) d\theta$

$\lambda \int_{\Re} b'(\theta)\pi(\theta) d\theta = \lambda m$

$\boxed{\int_{\Re} b'(\theta)\pi(\theta) d\theta = m}$ $\square$

For the variance, we can take the second derivative of the prior to get

$\frac{\partial^2 \pi}{\partial \theta^2} = \lambda\{m - b'(\theta)\}\pi'(\theta) - \lambda b''(\theta)\pi(\theta)$

Then use the first derivative from above to rewrite as

$\frac{\partial^2 \pi}{\partial \theta^2} = \lambda^2\{m - b'(\theta)\}^2\pi(\theta) - \lambda b''(\theta)\pi(\theta)$

Then note that $V(\mu) = v_0 + v_1\mu + v_2\mu^2 = b''(\theta)$:

$\frac{\partial^2 \pi}{\partial \theta^2} = \lambda^2\{m - b'(\theta)\}^2\pi(\theta) - \lambda V(\mu)\pi(\theta)$

$\frac{\partial^2 \pi}{\partial \theta^2} = \lambda^2\{m - b'(\theta)\}^2\pi(\theta) - \lambda(v_0 + v_1\mu + v_2\mu^2)\pi(\theta)$

Then integrate both sides

$\int_{\Re} \frac{\partial^2 \pi}{\partial \theta^2} d\theta = \lambda^2 \int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta - \lambda \int_{\Re}(v_0 + v_1\mu + v_2\mu^2)\pi(\theta) d\theta$

$\frac{\partial \pi}{\partial \theta} \int_{\Re} \frac{\partial \pi}{\partial \theta} d\theta = \lambda^2 \int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta - \lambda \int_{\Re}(v_0 + v_1\mu + v_2\mu^2)\pi(\theta) d\theta$

$\frac{\partial \pi}{\partial \theta} \int_{\Re} \lambda\{m - b'(\theta)\}\pi(\theta) d\theta = \lambda^2 \int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta - \lambda \int_{\Re}(v_0 + v_1\mu + v_2\mu^2)\pi(\theta) d\theta$

$\frac{\partial \pi}{\partial \theta} \int_{\Re}\{m - b'(\theta)\}\pi(\theta) d\theta = \lambda \int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta - \int_{\Re}(v_0 + v_1\mu + v_2\mu^2)\pi(\theta) d\theta$

$\frac{\partial \pi}{\partial \theta} \int_{\Re}\{m - b'(\theta)\}\pi(\theta) d\theta = \lambda \int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta - \int_{\Re}(v_0 + v_1\mu + v_2\mu^2)\pi(\theta) d\theta$

$\int_{\Re} m\pi(\theta) d\theta \int_{\Re}\{b'(\theta)\}\pi(\theta) d\theta = \lambda \int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta - \int_{\Re}(v_0 + v_1\mu + v_2\mu^2)\pi(\theta) d\theta$

$\int_{\Re}\{m - b'(\theta)\}^2\pi(\theta) d\theta = \frac{V(m)}{\lambda - v_2}$

1.2

The posterior for θ is
$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}$
and the joint distribution for the data model is
$f(\mathbf{y}|\theta) = \Pi f(y_i|\theta)$
$= \exp[-\phi n b(\theta) + \phi\theta \sum y_i + \sum c(y_i, \phi)]$
Then we can write the posterior as
$f(\theta|\mathbf{y}) \propto \exp[\phi\theta \sum y_i - \phi n b(\theta) + \sum c(y_i, \phi)] \cdot \exp[\lambda\{m\theta - b(\theta)\} + d(\lambda, m)]$
$\propto \exp\{[\theta(\lambda m + \phi \sum y_i)] - b(\theta)[-\phi n - \lambda] + \sum c(y_i, \phi) + d(\lambda, m)\}$
$\propto \exp\{(-\phi n - \lambda)[\theta(m + \phi\bar{y}) - b(\theta)] + \sum c(y_i, \phi) + d(\lambda, m)\}$
$\propto \pi(\theta)$
Thus $\pi(\theta)$ is conjugate.

1.3

From the second to last line above, with $E[\mu_i] = b'(\theta) = b'(m - \bar{y})$,
$E[\mu_i] = \frac{\lambda\bar{y} + (\phi/n)m}{\lambda + \phi/n} = \frac{n\phi\bar{y} + \lambda m}{\lambda + n\phi}$
and this is just a weighted average of the sample mean and prior mean m.

1.4

For data model $Poisson(y_i|\lambda)$ with prior $\pi(\lambda) \sim Gamma(\alpha_0, \beta_0)$, we can write the posterior as in 1.2 as
$f(\lambda|y) \propto \exp\{1[n\bar{y}\log(\lambda) - (\exp(\log(\lambda)))] - \sum \log(y_i!)\} \cdot \exp\{(1/\alpha_0)[m \log(\beta_0/\alpha_0) - (-\log(-\beta_0/\alpha_0))] + (\alpha_0 - 1)\log(m) - \log(\Gamma(\beta_0))\}$
and using the form from 1.2, this corresponds to
$\propto \exp\{\lambda(\beta_0/\alpha_0)m + \lambda \log(-\beta_0/\alpha_0) + n\bar{y}\log\{\beta_0/\alpha_0\} + \cdots\}$
and from 1.3, this is proportional to a Gamma distribution with parameters $\alpha = \alpha_0 + n\bar{y}$ and $\beta = 1/(n + \beta_0)$. The dispersion parameter for a Gamma distribution is $\phi = 1/\alpha_0$ and for a Poisson it is 1. Then the expected value (as in 1.3) is
$E[\mu_i] = \frac{(1/\alpha_0)\bar{y} + (1/n)(\beta_0/\alpha_0)}{n + \beta_0}$

$$\boxed{E[\mu_i] = \frac{n}{n + \beta_0}\bar{y} + \frac{\beta_0}{n + \beta_0}\frac{\alpha_0}{\beta_0}}$$

For a normal distribution with normal prior, we can characterize the distributions in terms of precision rather than variance parameters and the posterior distribution is
$f(\mu|y) \propto \exp\{\frac{-1}{2}(n\sigma^2 + \tau_0)(\mu - \frac{n\sigma^2\bar{y} + \tau_0\mu_0}{n\sigma^2 + \tau_0})^2\}$
and using the form from 1.3, this corresponds to a normal posterior with mean (as given in class)

$$\boxed{E[\mu] = \frac{n\sigma^2\bar{y} + \tau_0\mu_0}{n\sigma^2 + \tau_0}}$$

2

2.1

For $Y \sim Poisson(\lambda)$ with uniform prior, the posterior is just the joint data model for Y times 1,
$f(\lambda|\mathbf{y}) \propto \frac{\lambda^{\sum y_i}\exp(n\lambda)}{\Pi y_i!} \cdot (1)$

which is just the kernel of a Gamma($\sum y_i + 1, n$) distribution.

2.2

To derive the Jeffreys prior, we need the Fisher information. The log-likelihood for the data model is

$\ell = (\sum y_i)\log(\lambda) - n\lambda - \sum log(y_i!)$

Then take the negative of the second derivative with respect to λ to get

$-\frac{\partial^2 \ell}{\partial \lambda^2} = \frac{\sum y_i}{\lambda^2}$

Thus the Fisher information is

$I_n(\lambda) = -E[\frac{\partial^2 \ell}{\partial \lambda^2}] = \frac{n\lambda}{\lambda^2} = n/\lambda$

Then Jeffreys prior is

$\pi(\lambda) = I_1(\lambda)^{0.5} = \sqrt{1/\lambda} \boxed{= \lambda^{-0.5}, \lambda \in (0,\infty)}$

Integrating over the sample space for λ we have

$$\boxed{\int_{\Re+} \pi(\lambda) d\lambda = \int_{\Re+} \lambda^{-0.5} d\lambda = 2\lambda^{1/2}]_0^\infty = \infty}$$

Thus the Jeffreys prior is <u>not</u> proper

The posterior distribution is given by

$f(\lambda|\mathbf{y}) \propto \frac{\lambda^{\sum y_i}\exp(n\lambda)}{\Pi y_i!} \cdot (\lambda^{-0.5})$

$$\boxed{f(\lambda|\mathbf{y}) \propto \frac{\lambda^{\sum y_i - 1/2}\exp(n\lambda)}{\Pi y_i!}}$$

which is the kernel of a Gamma($\sum y_i + 1/2, n$) distribution. Thus the posterior distribution with Jeffreys' prior is a proper distribution.

2.3

(a) The posterior distribution is

$f(\lambda|\mathbf{y}) \propto \frac{\lambda^{\sum y_i}\exp(n\lambda)}{\Pi y_i!} \cdot \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\lambda^{\alpha_0 - 1}\exp(-\beta_0\lambda)$

$\propto \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)\Pi y_i!}\lambda^{\sum y_i + \alpha_0 - 1}\exp(n\lambda - \beta_0\lambda)$

which is $\boxed{Gamma(\sum y_i + \alpha_0, \beta_0 + n)}$

(b) For $n = 5$, $\sum y_i = 15$, $\alpha_0 = 1$ and $\beta_0 = 0.1$, the posterior distribution is $f(\lambda|y) \sim Gamma(16, 5.1)$. Then an 95% equal tailed credible set is given by

(L_λ, U_λ)

$L_\lambda = G^{-1}_{0.025}(16, 5.1)$

$U_\lambda = G^{-1}_{0.975}(16, 5.1)$

where $G^{-1}(\cdot)$ is the inverse cdf of a Gamma(α_0, β_0) distribution.

(c) In general, the equal-tailed interval is not the same as the HPD interval, because the Gamma distribution is not symmetric.

3

(a) We compare the performance of the uniform and Jeffreys priors with a Monte Carlo study. The results are in the table below (R code is attached). We can see that the CI coverage is slightly too high for the Jeffreys prior, indicating that the equal-tailed interval is too wide. For the uniform prior, we can see that the coverage is less than 90%, indicating that the interval is too narrow. For a Monte Carlo approximation to the true λ parameter, the

Jeffreys prior gives a closer approximation to the true value of $\lambda = 9$ and also has a smaller Monte Carlo standard error, indicating stronger performance.

Monte Carlo Study Results			
Model	**MC Estimate of λ**	**95% CI for λ**	**90% CI coverage rate**
Jeffreys'	9.021	(8.991 , 9.0496)	90.25%
Improper Uniform	9.046	(9.016 , 9.076)	89.2%

(b) For the Bayesian confidence intervals, we are using concepts of epistemic probability, updating our belief about λ based on the observed data. For the Monte Carlo estimates, we are using hypothetical limiting relative frequency as we are using the likelihood of the data and repeating the experiment many times to approximate the probabilities.

4

1. Define our random variables as $Y_i : i = 1, ..., 26$ be the rainfall volumes from clouds in the seeded (treatment) group, and $X_i : i = 1, ..., 26$ be the rainfall volumes from clouds in the control group, measured in cm. From Assignment 2 (see solutions), we have a mean parameterization for the Gamma (joint) distribution of the Y_is and X_is as
   $f(Y_i|\mu, \alpha) = \frac{\alpha}{\Gamma(\alpha)\mu}[\frac{y_i\alpha}{\mu}]^{\alpha-1}exp\{(-1/\mu)\alpha y_i\}$ and
   $f(X_i|\mu_c, \alpha) = \frac{\alpha}{\Gamma(\alpha)\mu_c}[\frac{x_i\alpha}{\mu_c}]^{\alpha-1}exp\{(-1/\mu_c)\alpha x_i\}$
   where the two groups share a common shape parameter α. Since the observations are independent, the data model for each group is just the product of the individual likelihood functions, $f(\mathbf{Y}|\mu, \alpha) = \prod_1^{26} f(y_i|\mu, \alpha)$ and $f(\mathbf{X}|\mu_c, \alpha) = \prod_1^{26} f(x_i|\mu_c, \alpha)$, and these are both gamma distributions as well. The conjugate prior distribution for the mean parameters μ and μ_c is conveniently in the form of an inverse gamma distribution. I choose a diffuse prior, assuming that I have no prior information on the mean. For α, we can specify a non-informative but proper prior that enforces $\alpha > 0$. I choose a half-Cauchy distribution based on internet recommendations. Our model is then specified as:
   $Y_1, ..., Y_{26} \sim Gamma(\alpha, \frac{\mu}{\alpha})$
   $X_1, ..., X_{26} \sim Gamma(\alpha, \frac{\mu_c}{\alpha})$
   $\pi(\mu) \sim invGamma(\alpha_0, \beta_0)$
   $\pi(\mu_c) \sim invGamma(\alpha_0, \beta_0)$
   $\pi(\alpha) \sim halfCauchy(\mu = 0, \tau = 2.5)$
   All of the priors are proper in this case, so the posterior is also proper (although possibly not easy to write down).

2. The posterior distribution for the full model is difficult to write down, so we can use Gibbs sampling to simulate from the conditional posterior distributions instead. The joint posterior of α and μ, μ_c can be written as
   $p(\alpha, \mu, \mu_c|\mathbf{x}, \mathbf{y}) \propto \pi(\alpha)\pi(\mu)f(\mathbf{y}|\alpha, \mu)\pi(\mu_c)f(\mathbf{x}|\alpha, \mu_c)$
   which does not have a recognizable form. However, we have that
   $p(\mu|\mathbf{x}, \mathbf{y}, \mu_c, \alpha) \propto \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)\Gamma(\alpha)} \prod y_i^{\alpha-1}\mu^{-\alpha_0-n\alpha-1+n}exp\{(1/\mu)(-n\alpha\sum y_i - \beta_0)\}$
   which is recognizable as an $invGamma(\alpha_0+n(\alpha-1), n\alpha\sum y_i+\beta_0)$ distribution. The same holds for the conditional posterior for μ_c, with x_i instead of y_i. Since we recognize these conditional distributions (α considered fixed), we know how to simulate from these directly and can use them in our Gibbs sampling scheme. Then the full conditional posterior for α becomes
   $p(\alpha|\mu, \mu_c, \mathbf{x}, \mathbf{y}) \propto \pi(\alpha)\pi(\mu)f(\mathbf{y}|\alpha, \mu)\pi(\mu_c)f(\mathbf{x}|\alpha, \mu_c)$
   $\propto \pi(\alpha)p(\mu|\mathbf{x}, \mathbf{y}, \mu_c, \alpha)p(\mu_c|\mathbf{x}, \mathbf{y}, \mu, \alpha)$
   This is still not recognizable, but we know how to simulate from the final two distributions, and thus we can simulate

from a distribution proportional to the conditional posterior, using a step from the Metropolis-Hastings algorithm with the above as our target distribution. Thus our full MCMC simulation algorithm is a Metropolis-within-Gibbs:

- Simulate an observation from each prior distribution $\pi(\alpha)$, $\pi(\mu)$, $\pi(\mu_c)$ to initialize.
- Simulate from the conditional posteriors for the means, $p(\mu|\mathbf{x}, \mathbf{y}, \mu_c, \alpha)$ and $p(\mu_c|\mathbf{x}, \mathbf{y}, \mu, \alpha)$
- (M-H step) Sample from the conditional posterior for α, $p(\alpha|\mu, \mu_c, \mathbf{x}, \mathbf{y})$ using 1 step of the Metropolis-Hastings algorithm.
- (Gibbs) Iterate 1 and 2 until burn-in period is reached.
- Keep samples of μ, μ_c, α from (b) and (c) beyond burn-in period as samples from the full conditional posterior distribution.

3. Luckily for us, JAGS will choose the most efficient way to simulate from the posterior distribution, so we don't have to implement the algorithm from above. Using the clouds data, I implement the Bayesian model specified in (1) with diffuse inverse Gamma priors for the mean parameters and a non-informative half-Cauchy prior for α. The burn-in period for the MCMC is 50000, after which I keep 60,000 samples from the posterior. The `R` code is attached at the end.

I will provide analysis summaries first. The posterior distributions are summarized in Table 3, and all of the relevant graphics are at the end. Comparing to the original likelihood-based analysis from Assignment 2, our estimation results are almost exactly the same. The posterior means for the group mean parameters are quite different from each other, however both have quite large posterior standard deviations and thus there is a large amount of uncertainty in our estimates (perhaps due to relatively small sample size). We can see that the 95% HPD intervals slightly overlap, which leads us to conclude that, at a 5% significance level, the group means are not significantly different: there is not evidence that the silver iodide seeding treatment had an effect on the rainfall volume.

Cloud Seeding Analysis Posterior Summary				
Parameter	**Posterior Mean**	**Posterior S.D.**	**HPD Interval**	$\hat{R}$
α	0.598	0.098	(0.419 , 0.799)	1.001
μ	458.159	123.777	(255.927 , 694.605)	1.001
μ_c	174.305	47.050	(98.344 , 268.783)	1.001

Table 1: Results from Bayesian analysis of cloud seeding data

Next, I examine the MCMC convergence using output from JAGS. The traceplots indicate that all 3 chains converge similarly, so there is no issue with mixing visible here. The larger issue comes from examining plots of the scale reduction factor. We can see that even after a large burn-in period, the scale reduction factors still do not converge to 1 quickly, indicating that we do, in fact have some difficulty with mixing in the MCMC chains. This is likely to the diffuse priors. However, based on the trace plots and the autocorrelation plots, we can see that the Markov chain does converge well enough for reliable results. Finally, the effective sample size for α is only 3500, which is a small portion of the actual samples kept, indicating poor convergence as well. One possible reason for this is that we have 3 diffuse priors which all have a non-negligible influence on the posterior distributions because of the small sample size (n=26) in our data model.

4. If we know the mean rainfall volume for the control group is between 50 and 300, we can change the prior for μ_c to a distribution that is either bounded by 50 and 300 or contains most of its probability density between 50 and 300. An example would be to change the prior to $\pi(\mu_c) \sim Unif(50, 300)$ or possibly even a Gamma/invGamma that has mean and variance parameterized such that most of the density is between 50 and 300.

5. (EC) Changing the prior of the control group mean to $\pi(\mu_c) \sim Unif(50, 300)$ and running the analysis as before, the results are in Table 5. We can see that the posterior estimates of α and μ are largely the same, while for μ_c the posterior mean has increased by 9 units and the posterior standard deviation has decreased slightly as well, from the first analysis. This indicates that updating our prior has had an effect. However, the HPD intervals still overlap and thus our conclusion about the scientific question of interest has not changed. Now, the effective sample size is 18000, which is much improved over the first analysis.

Cloud Seeding Analysis Posterior Summary				
Parameter	**Posterior Mean**	**Posterior S.D.**	**HPD Interval**	$\hat{R}$
α	0.598	0.099	(0.408 , 0.787)	1.001
μ	457.375	120.876	(246.898 , 697.419)	1.001
μ_c	183.095	43.863	(108.133 , 273.232)	1.001

Table 2: Results from Bayesian analysis of cloud seeding data with updated prior

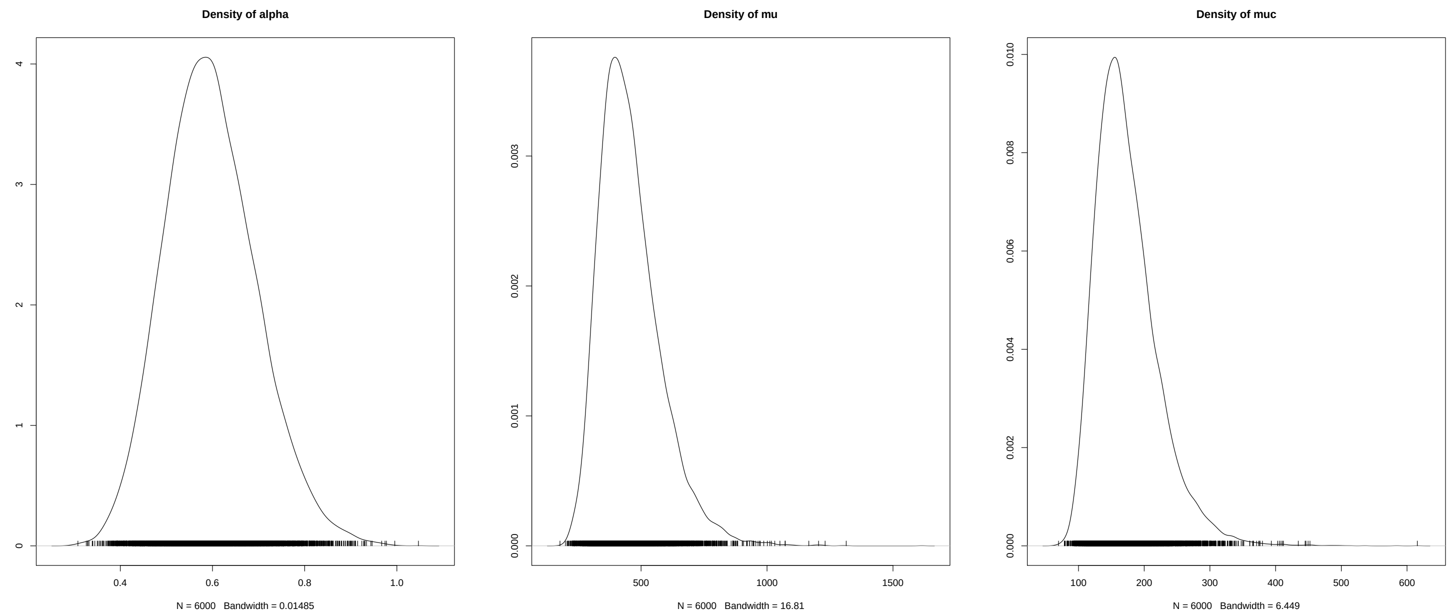

Figure 1: Posterior Density Plots

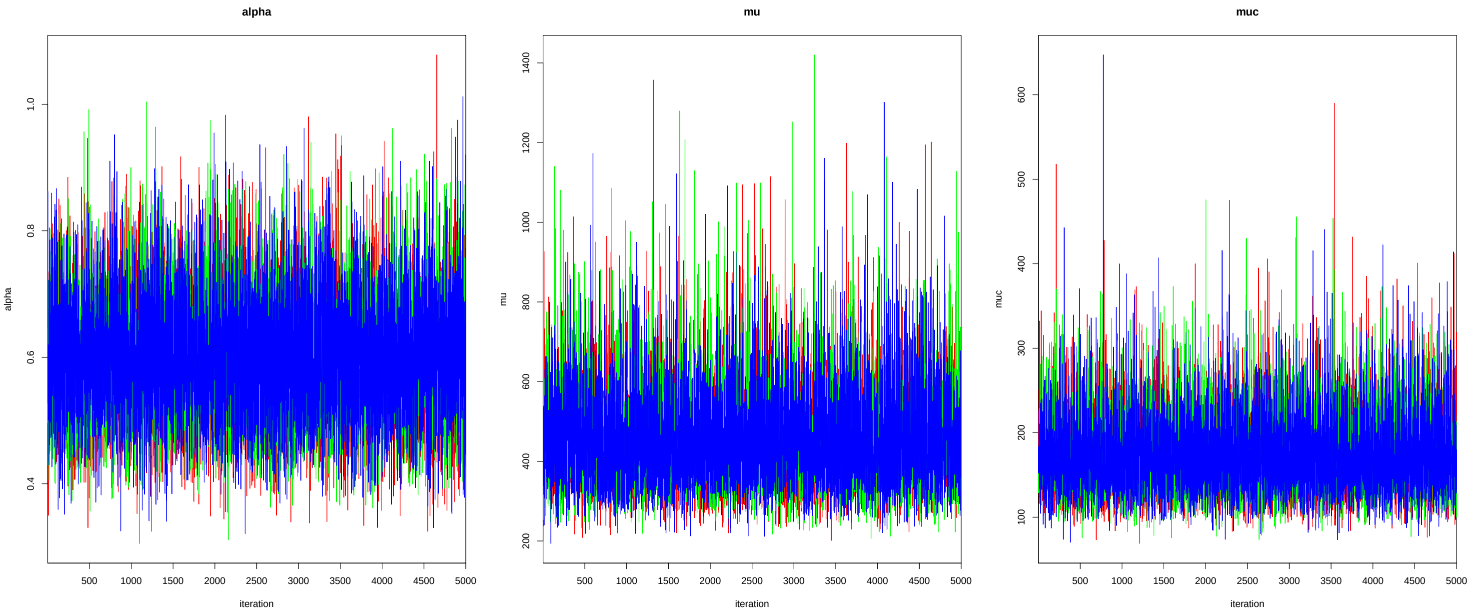

Figure 2: Trace Plots

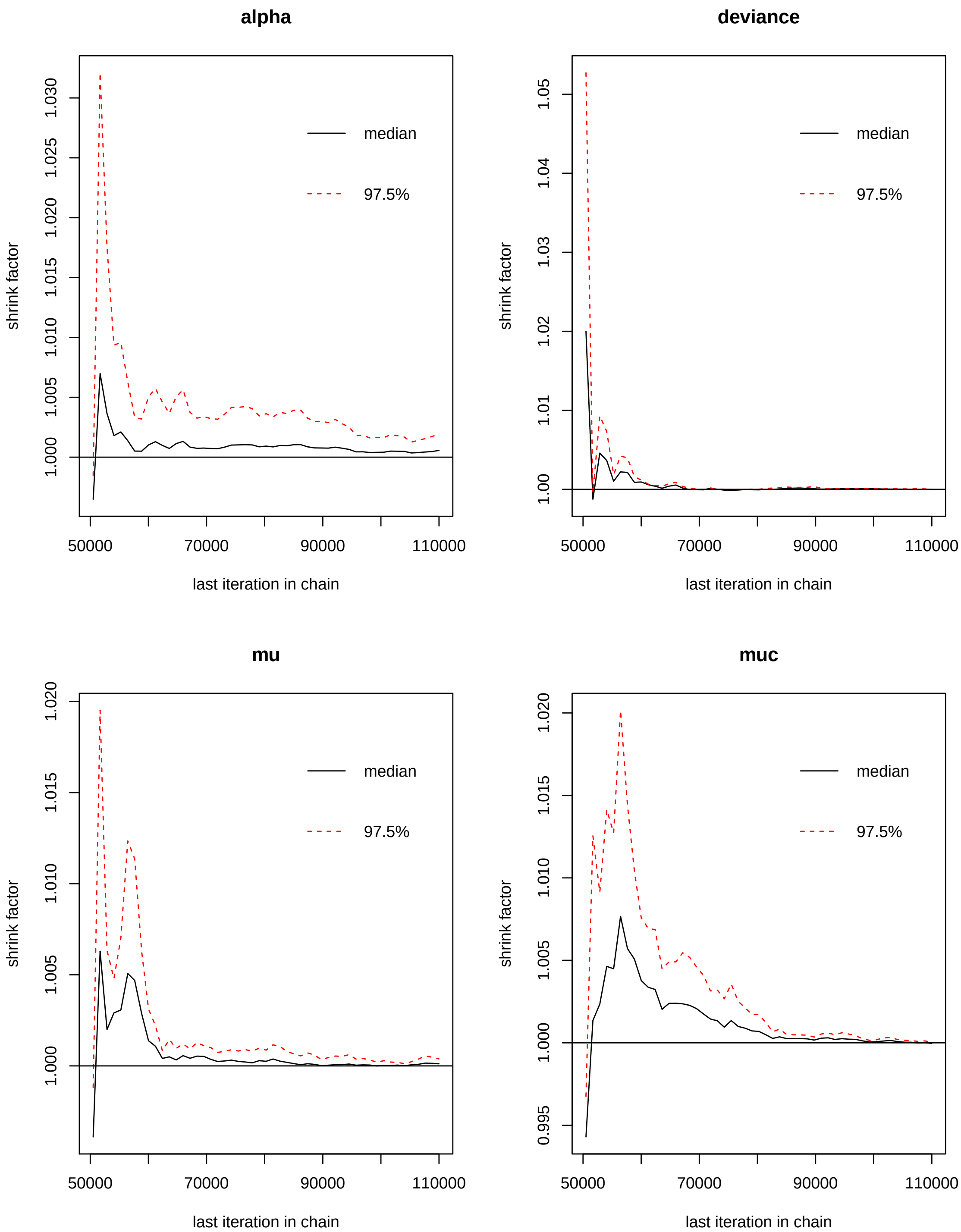

Figure 3: Gelman Plot

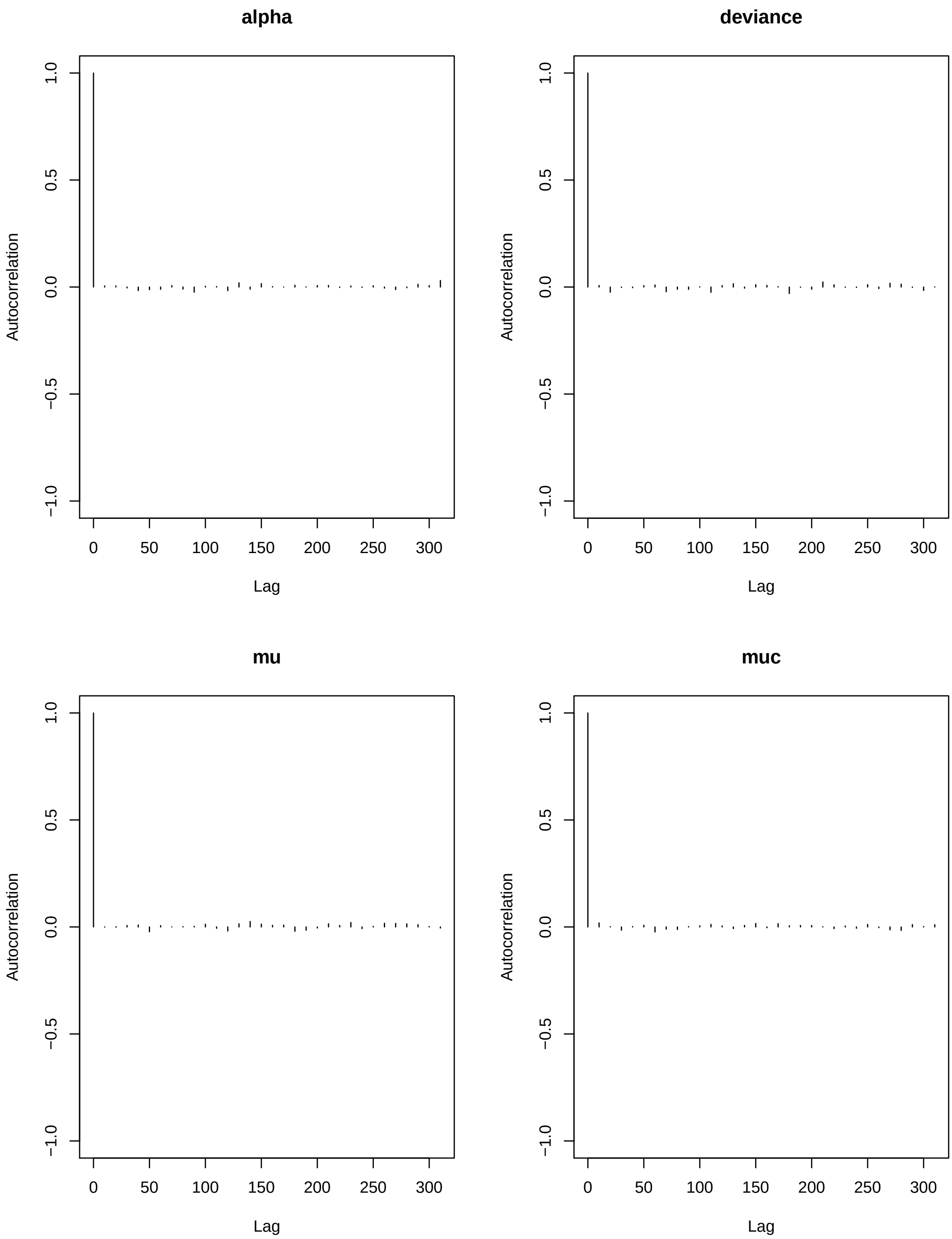

Figure 4: Autocorrelation Plots