Red Hawk Sightings in California

Mary Silva

AMS 207 - Take Home Exam 1

Abstract

The North American Breeding survey provides information about the abundance of different species of birds in North America. We look to examine the counts of Red Hawk sightings in California using a Bayesian Hierarchical model.

KEY WORDS: Hierarchical model, rejection sampling, Poisson-Gamma model, North American Breeding Survey.

Table 1: A portion of the data from North American Breeding Survey for the years 1968 - 1977 showing the counts of Red-tailed Hawks in California and the Route Counts for each year.

	years	Route Count	Red-tailed Hawk
1	1968	26	73
2	1969	31	81
3	1970	59	157
4	1971	59	131
5	1972	128	307
6	1973	151	284
7	1974	145	364
8	1975	150	381
9	1976	144	405
10	1977	144	367

1. Introduction

The data is obtained from the North American Breeding Survey for Red Hawks in California (Table 1). The full data set contains the years 1977 through 2017. The variables of interest are the bird route counts and the counts of Red Hawk sightings.

Initial exploratory analysis shows an increasing trend of Red Hawk sightings over time (Figure 1) as well as an increasing trend in bird flight route counts (Figure 2). We see that the density for red hawk sightings in California 3 is multimodal and skewed, which should be taken into consideration when building a model.

2. Model Methodology

Given that we are dealing with count data, it is reasonable to consider the red hawk sightings (y_i) to be
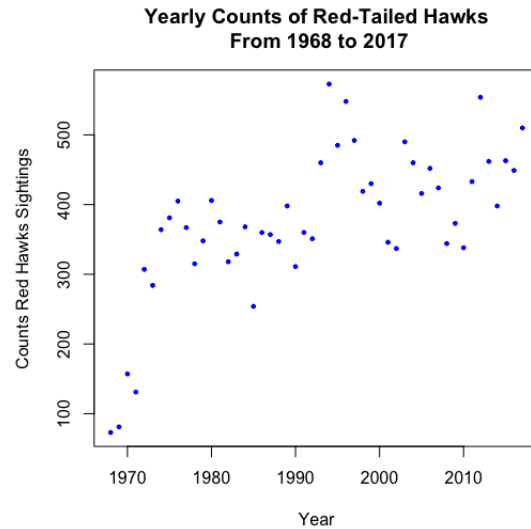

Figure 1: Scatter plot of the counts of Red Hawk Sightings by year in California
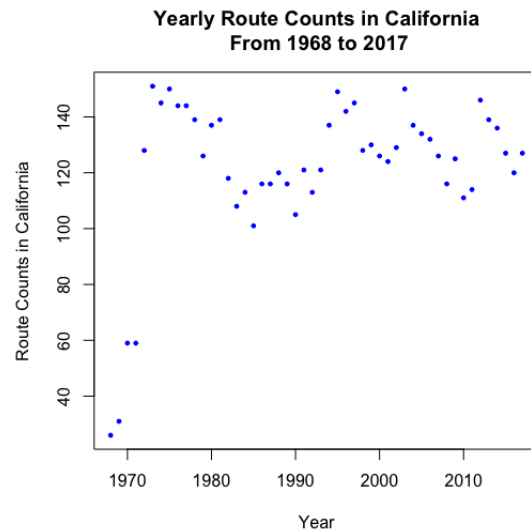

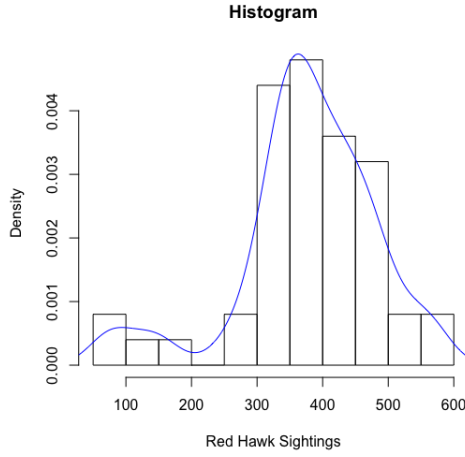Figure 2: Scatter plot of the route counts for California over time

Figure 3: A histogram plot of the Red Hawk sightings in California with a density overlay

distributed from a Beta-Binomial model, or as Poisson-Gamma model.

2.1 Model 1: Methodology

For this data, we'll first attempt the following hierarchical structure:

$$y_i \sim \text{Pois}(\mu_i)$$
$$\mu_i \sim \text{Gamma}(\alpha, \beta)$$
$$\beta \sim \text{Gamma}(a, b)$$

where $\mu_i = \lambda_i/c_i$ and c_i represents the route count for the i-th year. Furthermore, we will assume α_b, β_b and α are fixed. Later, we will assign a prior to alpha as well.

The joint posterior distribution for the above mentioned model then becomes

$$p(\mu, \beta|y) \propto \prod_{i=1}^{n} [f(y|\mu_i)p(\mu_i|\alpha, \beta)] \, p(\beta)$$

$$\propto \prod_{i=1}^{n} \mu_i^{y_i+\alpha-1} exp\left\{-(\beta+1)\sum_{i=1}^{n} \mu_i\right\}$$

$$\times \frac{b^a}{\Gamma(a)} \beta^{a-1} \exp\{-b\beta\}$$

The full conditionals (up to proportionality) under this model specification are as follows

$$\mu_i|\cdot \sim \text{Gamma}\left(y_i + \alpha, \beta + 1\right)$$

$$\beta|\cdot \sim \text{Gamma}\left(\alpha n + a, b + \sum_{i=1}^{n} \mu_i\right)$$

Since the full conditionals (under this model specification) are known distributions, we can use R to implement Gibbs sampling algorithms to simulate from the joint posterior of all parameters.

We use 2000 MCMC iterations and discard a burn-in of 1000 iterations so sample from the full conditionals. Since this is a Gibbs algorithm, our hyper parameters can be specified as vague. In this case we set $\alpha = 10, a = 1$, and $b = 1$. We summarize our posterior samples in (Table 2.1). A brief analysis of the traceplots and histograms shows that our Gibbs sampler is mixing and converging properly (Appendix Figure 8). We also can verify that there are no issues with autocorrelation (Figure 9) which further prove that the Gibbs sampler is working.

Table 2: Summary of posterior means of select posterior samples of μ and β. Keep in mind that $\mu_i = \lambda_i/c_i$

	Mean	Sd	2.5%	97.5%
μ_1	80.36	8.78	63.57	98.03
μ_{10}	366.88	19.44	329.15	406.17
μ_{30}	488.38	21.52	446.05	533.37
μ_{49}	447.20	21.28	408.08	488.53
β	0.03	0.00	0.02	0.03

Figure 4: Distribution of the observed y_i and the distributions of some of the posterior replicates (y_i^{rep}) from model 1.
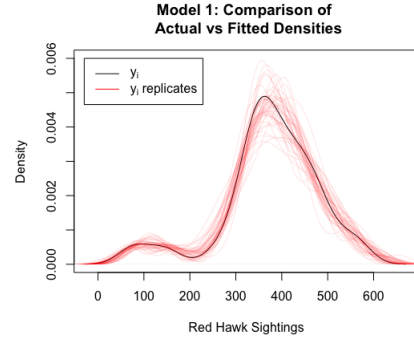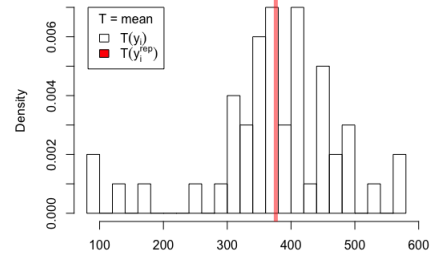

Figure 5: Posterior predictive distribution, $T(y) = mean(y)$. We compare the mean of the observed to the means of the y_i^{rep}

2.2 Model 1: Model Assessment

Next, we present predictive model checks and provide an assessment of the model fit to the observed records. Under the model 1 specifications, we simulate 200 posterior replicates of y_i (Figure 4). We see that under this model specification, the posterior density fits very well (maybe too well).

We may also compare the means of the posterior predictive distribution to the mean of the observed results (Figure 5). Based on the model 1 specifications, we can conclude this model accurately fits the data.

3. Model 2: Methodology

Next, we propose a hierarchical model similar to model 1, but also adds a prior on α as follows

$$y_i \sim \text{Pois}(\mu_i)$$
$$\mu_i \sim \text{Gamma}(\alpha, \beta)$$
$$\beta \sim \text{Gamma}(a, b)$$
$$\alpha \sim \text{Gamma}(c, d)$$

Where, here, the fixed parameters are a, b, c and d. The joint posterior distribution for model 2 then becomes

$$p(\mu, \beta, \alpha | y) \propto \prod_{i=1}^{n} \left[f(y_i | \mu_i) p(\mu_i | \alpha, \beta) \right] p(\alpha) p(\beta)$$

$$\propto \left(\frac{\beta^{\alpha n}}{\Gamma(\alpha)^n} \right) \prod_{i=1}^{n} \left(\mu_i^{y_i + \alpha - 1} \right)$$

$$\times \exp \left\{ - \sum_{i=1}^{n} \mu_i (1 + \beta) \right\}$$

$$\times \frac{b^a}{\Gamma(a)} \alpha^{a-1} \exp \{ -b\alpha \}$$

$$\times \frac{d^c}{\Gamma(c)} \beta^{c-1} \exp \{ -d\beta \}$$

From this, we can obtain the full conditionals (up to proportionality).

$$p(\mu_i | \cdot) \propto \mu_i^{y_i + \alpha - 1} \exp \left\{ - \mu_i (\beta + 1) \right\}$$

$$p(\beta | \cdot) \propto \beta^{\alpha n + c - 1} \exp \left\{ -\beta \left(d + \sum_{i=1}^{n} \mu_i \right) \right\}$$

$$p(\alpha | \cdot) \propto \left(\prod_{i=1}^{n} \mu_i^\alpha \right) \frac{\beta^{\alpha n}}{\Gamma(\alpha)^n} \alpha^{a-1} \exp \{ -b\alpha \}$$

We recognize that the full conditionals for μ_i and β are known distributions, thus

$$\mu_i | \cdot \sim \text{Gamma}(y_i + \alpha, \beta + 1)$$

$$\beta | \cdot \sim \text{Gamma}(\alpha n + c, d + \sum_{i=1}^{n} \mu_i)$$

The full conditional for α, however, is not a known standard distribution. Thus, we must apply either a Metropolis-within-Gibbs step or rejection sampling method to sample the α's. Using both methods, almost identical results were obtained, but for this paper only a Sampling Importance Resampling (SIR) method is used to jointly sample (α, β).

The computational procedure is as follows

- Draw from the joint posterior $p(\alpha, \beta) \propto p(\alpha) p(\beta)$

- Given α, β, draw μ_i from the corresponding Gamma distribution

A contour plot of α and β on the log scale for model 2, together with simulated draws from the rejection algorithm, are shown in Figure 6. A summary of the posterior samples can be seen in Table 3. We can note that these values are similar to model 1.
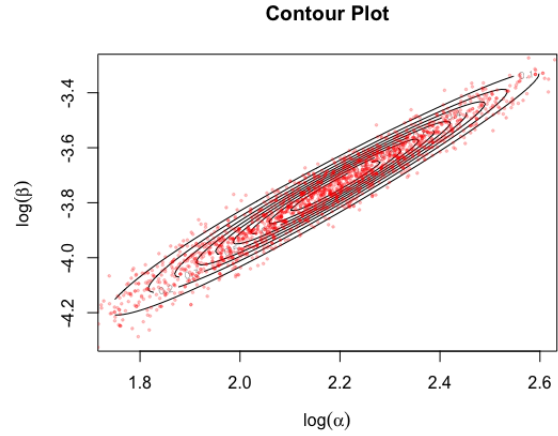

Contour Plot

Figure 6: Contour plot with 2000 samples of α and β on the log scale.

Table 3: Cap

	Mean	SD	2.5%	97.5%
α	8.79	1.72	5.75	12.39
β	0.02	0.00	0.02	0.03
μ_1	79.81	8.92	63.40	98.54
μ_2	87.86	9.48	69.65	108.25
μ_3	161.88	12.74	138.64	187.76
μ_4	136.32	11.80	113.68	159.85
μ_5	309.41	17.67	275.85	345.99

Again, to check for mixing and convergence we examine the sampling traces for the parameters (Appendix Figures 11 - 12). By examining the traceplots of μ_1 for both models, we see that the parameter converges almost identically to the same mean (It's not the same image, I triple checked!). The autocorrelation plots also show no problems (10).

3.1 Model 2: Model Assessment

We repeat the assesment steps as we did with model 1. The comparison of the actual Red Hawk sightings to the posterior predictive replicates, y_i^{rep}, (Figure 7) are expected to look identical for both models and we can see that this is true. The reason for this is that with model 1, we set $\alpha = 10$ in order to reflect the distributional assumptions allowing us to obtain the good estimates for $\mu_i = \lambda_i/c_i$. This choice of α in model 1 happens to be very close to the posterior mean for α in model 2 which explains the similar fitted values.
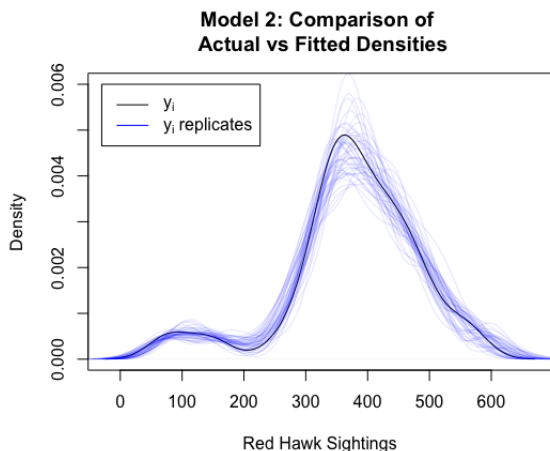
Figure 7: Distribution of the observed y_i and the distributions of some of the posterior replicates (y_i^{rep}) for Model 2.

4. Conclusion

As we expected, both model 1 and model 2 yield similar results. We can use a model comparison method to verify this.

Definition 1 (Gelfand & Ghosh) *Let $\boldsymbol{y}$ denote the observed data. Let $\mu_l = E(z_l|\boldsymbol{y})$ and $\sigma_l^2 = Var(z_l|\boldsymbol{y})$. Then let $G = \sum_l(\mu_l - y_l)^2$, and $P = \sum_l \sigma_l^2$. The Gelfand and Ghosh criterion is defined as $D = G + P$, where D seeks to reward goodness of fit penalizing complexity. So the smaller the D, the better the model.*

The Gelfand & Ghosh criterion for model 1 is 2449.3 and for model 2 is 2678.6. Since the Gelfand and Ghosh Criterion for model 1 is smaller this, by definition, indicates that model 1 fits the data better. The smaller value for model 1 is expected since model 2 is more complex and GG penalizes complexity. Of course, there are other methods to compare model fit, such as Deviance Information Criterion (DIC), and Bayes Factor, but those methods will be applied to future assignments once they are covered in the course.

To estimate the probability of observing more than 450 red hawks in california in a year with a route count of 120, I sampled 2500 iterations from the route counts from gamma distribution and kept the results where the rate was greater than 120, and set all others equal to 0. Then I used the vector of route counts where this statement was true to simulate 2500 y posterior predictive replicates. From this, I calculated the percentage of values where the result was greater than 450. This percentage came out to be around 22.4%. This compares well to the 26% in the original data.
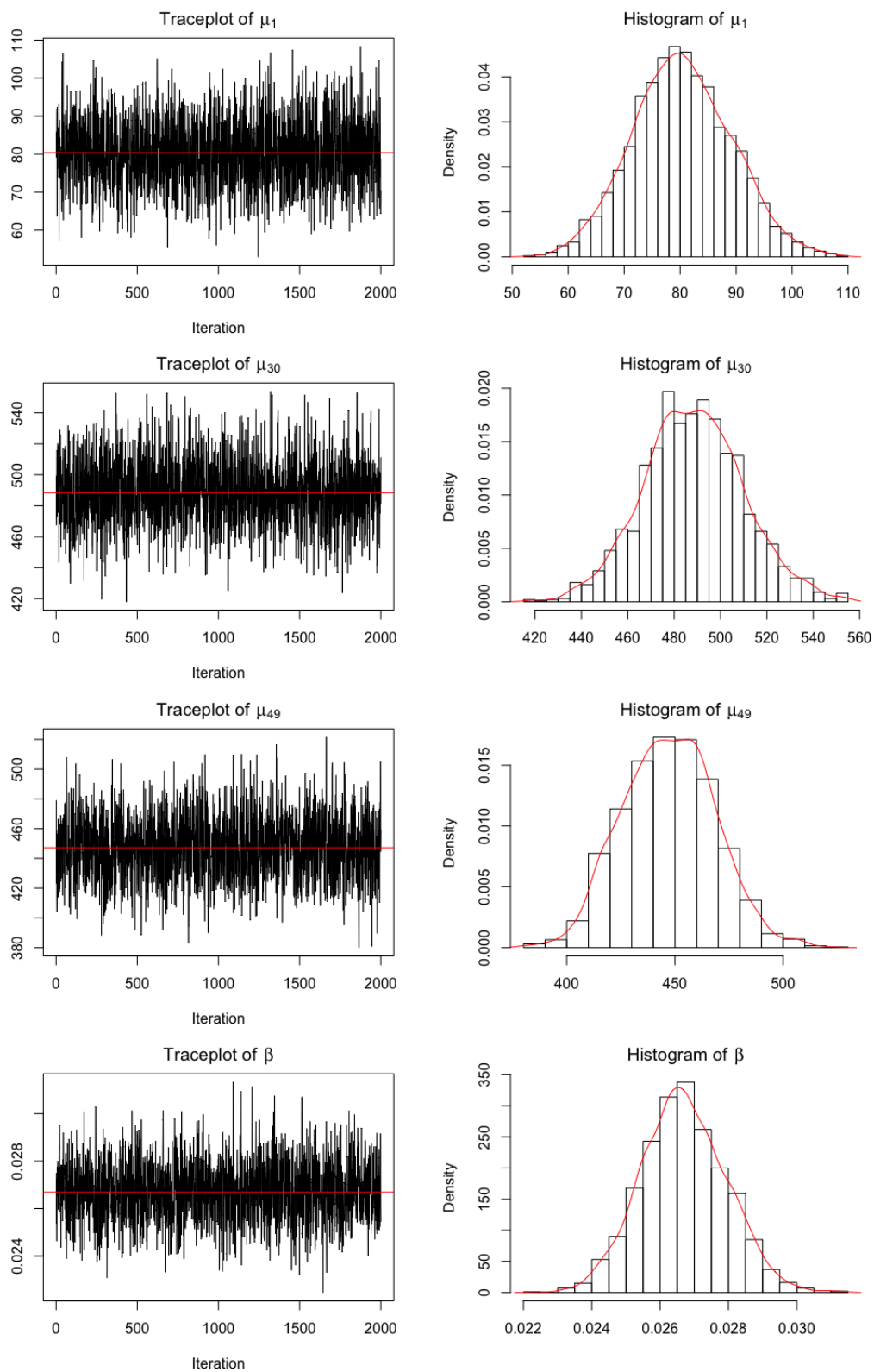
Figure 8: Traceplots for $\mu_1, \mu_{30}, \mu_{49}$, and β.
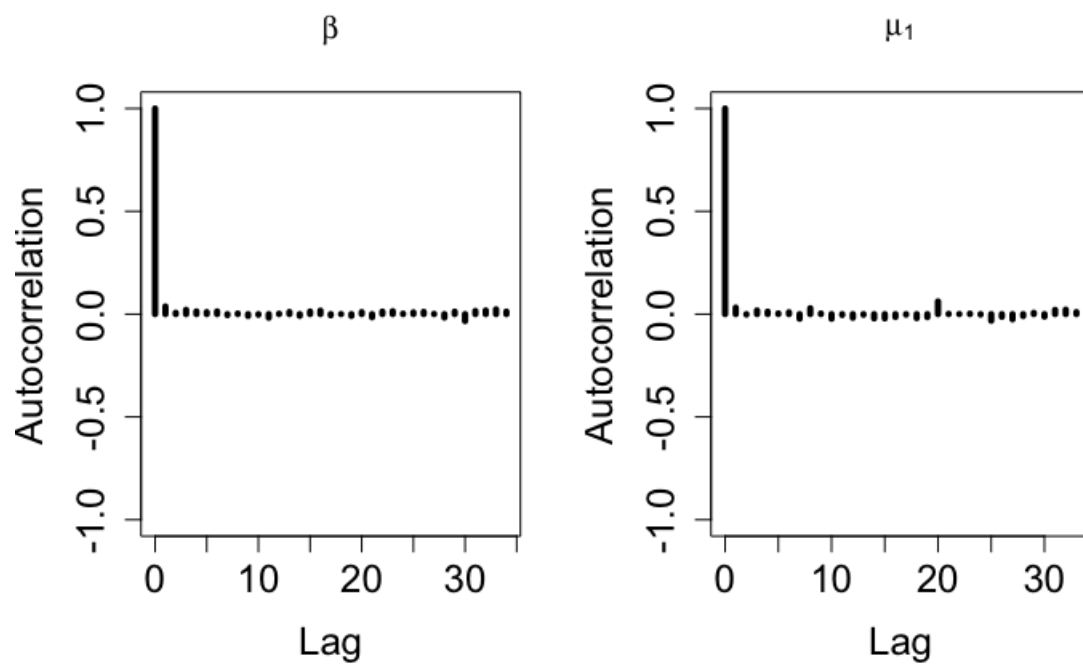
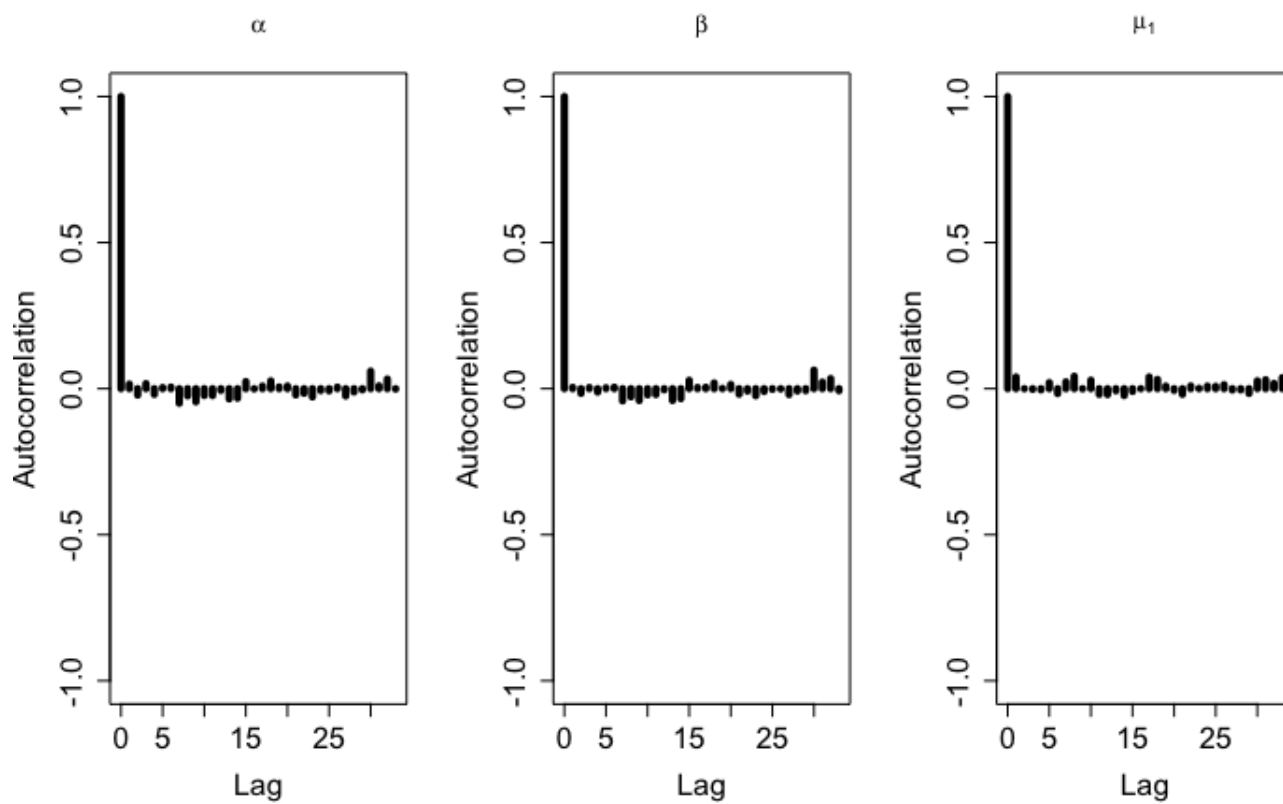Figure 9: Autocorrelation for model 1 parameters β, μ_1

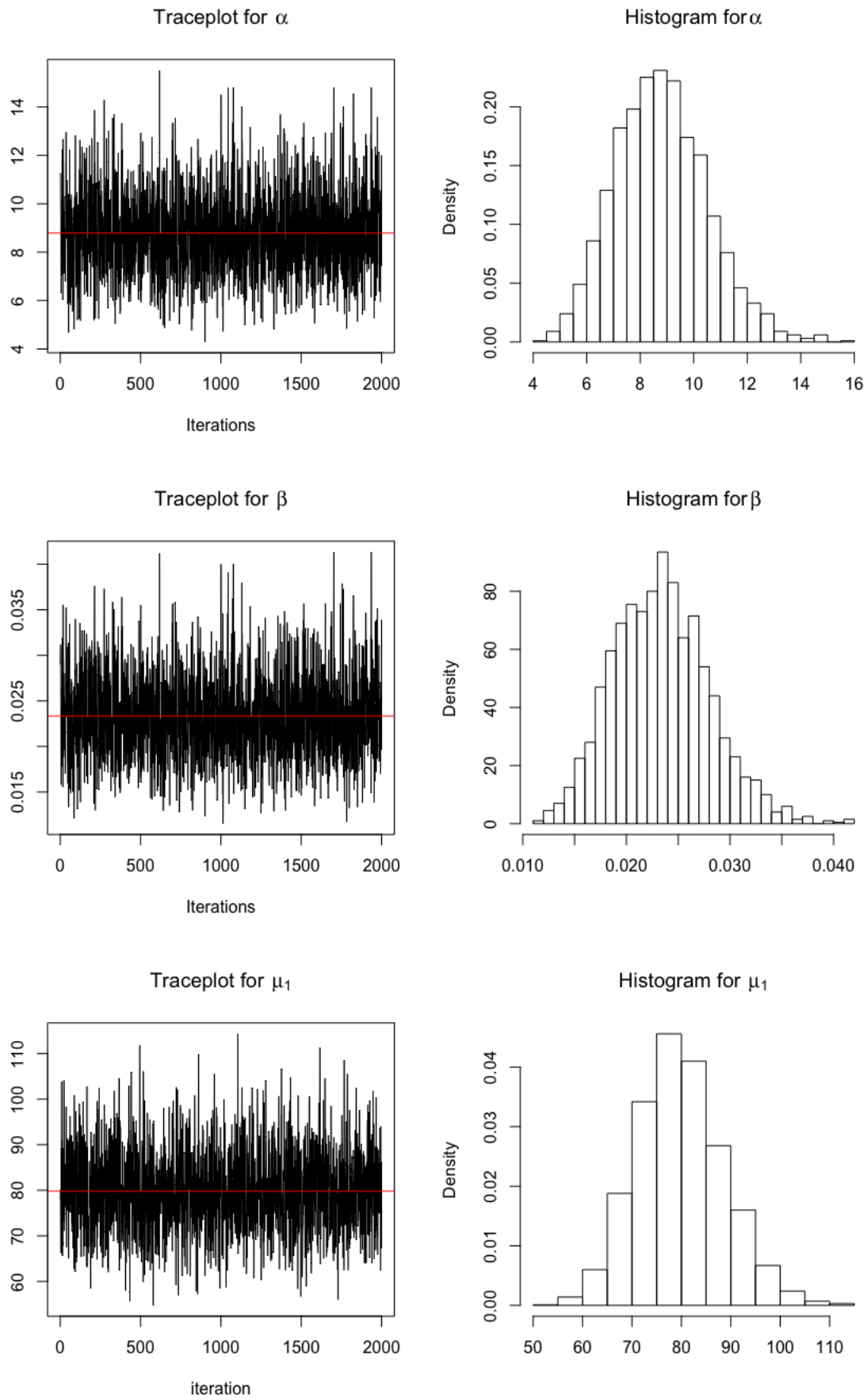Figure 10: Autocorrelation for model 2 parameters α, β, μ_1

Figure 11: Traceplots and histograms for posterior samples of α, β and μ_1 corresponding to model 2.
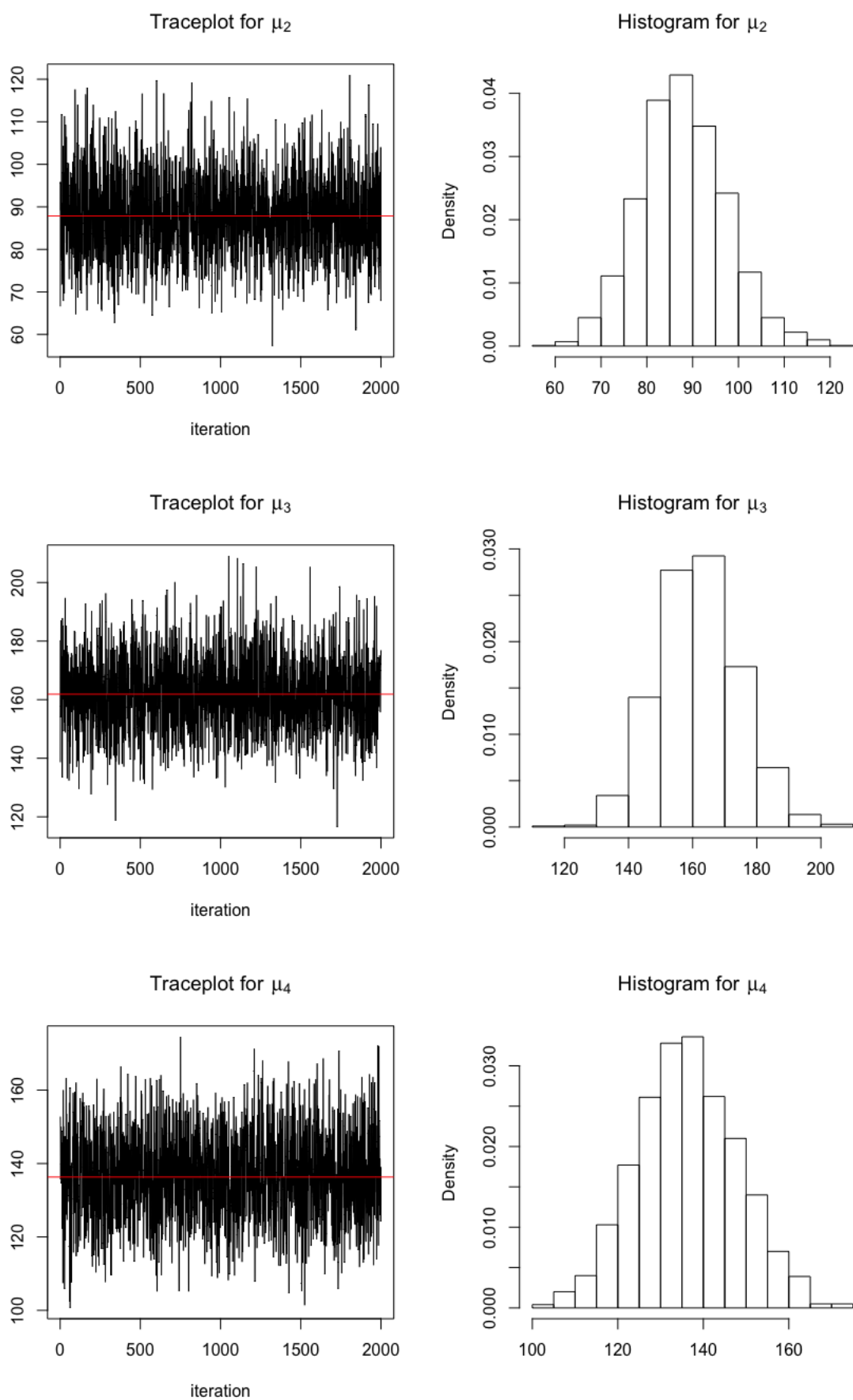
Figure 12: Traceplots and histograms for posterior samples of μ_i for $i = 2, ..., 4$. Clearly, convergence is achieved.