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 Survery 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 1

Note that I have modified Model 1 from my original takehome as you suggested. Given that we are dealing with count data, it is reasonable to consider the

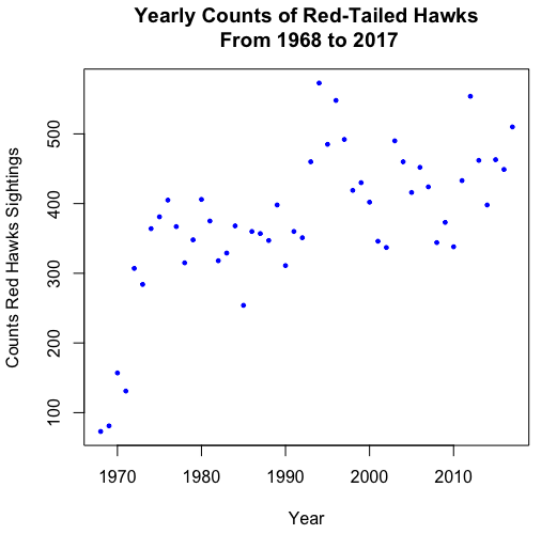

Figure 1: Scatter plot of the counts of Red Hawk Sightings by year in California

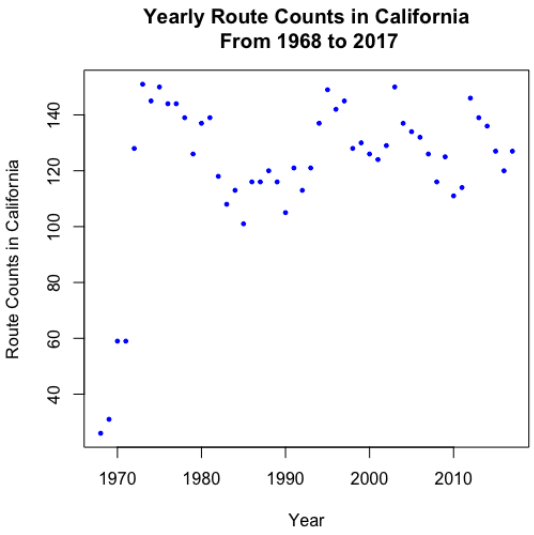

Figure 2: Scatter plot of the route counts for California over time

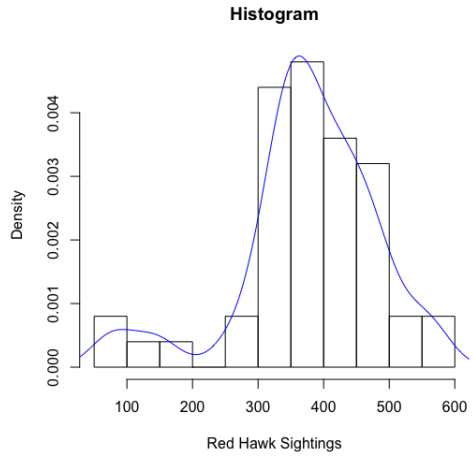

Figure 3: A histogram plot of the Red Hawk sightings in California with a density overlay

red hawk sightings (y_i) to be distributed from a Beta-Binomial model, or as Poisson-Gamma model.

2.1 M1: Specifications

Model 1 now incorporates the rate counts appropriately and I used a simpler hierarchical structure as follows

$$\begin{aligned} y_i &\sim \text{Pois}(\lambda_i c_i) \\ \lambda_i &\sim \text{Exp}(\theta) \\ \theta &\sim \text{Gamma}(\alpha, \beta) \end{aligned}$$

Where the values α and β are fixed. The joint posterior distribution for the above mentioned model then becomes

$$\begin{aligned} p(\lambda, \theta|y) &\propto \prod_{i=1}^{n} \left[f(y|\lambda_i, c_i)\pi(\lambda_i|\theta)\right] \pi(\theta) \\ &\propto \prod_{i=1}^{n} \left[\frac{(c_i\lambda_i)^{y_i} \exp\{-\lambda_i c_i\}}{y_i!} \theta \exp\{-\theta\lambda_i\}\right] \\ &\quad \times \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} \exp\{-\beta\theta\} \end{aligned}$$

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

$$\begin{aligned} \lambda_i\Big|\cdot &\sim \text{Gamma}\Big(\lambda_i\Big|y_i + 1, c_i + \theta\Big) \\ \theta\Big|\cdot &\sim \text{Gamma}\left(\theta\Big|n + \alpha, \beta + \sum_{i=1}^{n} \lambda_i\right) \end{aligned}$$

Since the full conditionals (under this model specification) are known distributions, I re-implemented model 1 using a straightforward Gibbs sampling approach.

3. Model 2

3.1 Specifications

Next, we will consider a modification of model 1 as follows

$$\begin{aligned} y_i &\sim \text{Bin}(N_i, \theta_i) \\ N_i &\sim \text{Pois}(\lambda_i c_i) \\ \lambda_i &\sim \text{Gamma}(\alpha_\lambda, \beta_\lambda) \\ \theta_i &\sim \text{Beta}(1, \beta_\lambda) \end{aligned}$$

where y_i corresponds to the observed count of hawks, out of a population of N_i hawks for the i^{th} year. Unfortunately, N_i is not known, so we assign a discrete prior.

The joint posterior is given by the following

$$\begin{aligned} p(N_i, \theta_i, \lambda_i|y) &\propto \prod_{i=1}^{n} \Big[f(y_i|N_i, \theta_i)\pi(N_i|\lambda_i, c_i)\Big] \\ &\quad \times \prod_{i=1}^{n} \Big[\pi(\lambda_i|\alpha_\lambda, \beta_\lambda)\pi(\theta_i|\beta_\theta)\Big] \\ &\propto \prod_{i=1}^{n} \left[\frac{\beta_\lambda^{\alpha_\lambda}}{\Gamma(\alpha_\lambda)} \frac{1}{Be(1, \beta_\lambda)} \frac{1}{y_i!(N_i - y_i)!}\right] \\ &\quad \times \prod_{i=1}^{n} \Big[\lambda_i^{N_i + \alpha_\lambda - 1} \exp\{-\lambda_i(c_i + \beta_\lambda)\}\Big] \\ &\quad \times \prod_{i=1}^{n} \Big[c_i^{N_i} \theta_i^{y_i} (1 - \theta_i)^{N_i - y_i + \beta_\theta - 1}\Big] \end{aligned}$$

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

$$\begin{aligned} \theta_i|\cdot &\sim \text{Beta}\Big(\theta_i\Big|y_i + 1, N_i - y_i + \beta_\theta\Big) \\ \lambda_i|\cdot &\sim \text{Gamma}\Big(\lambda_i\Big|N_i + \alpha_\lambda, c_i + \beta_\lambda\Big) \\ \pi_N(N_i|\cdot) &\propto \frac{1}{(N_i - y_i)!}(c_i\lambda_i)^{N_i}(1 - \theta_i)^{Ni} \end{aligned}$$

The hyper parameters of the priors $\alpha_\lambda, \beta_\lambda, \beta_\theta$ are fixed, and the choice of hyperparameter will be described in the next section.

3.2 M2: Sampling Algorithm

The full conditional for N_i is not a recognizable distribution. So, unlike model 1, a Gibbs sampling algorithm is not appropriate.

We consider a Metropolis-Hastings algorithm, in which we sample a proposed N_i^* from a proposal distribution at each iteration t. Since we know that N_i is the population of hawks, our *proposal distribution must be discrete*. Ultimately, I decide to use a Poisson distribution as my proposal distribution for N_i^*. We know that the proposal distribution should have a mean centered around the rate of the population of red-hawks, $\lambda_i c_i$, so we use our λ_i samples at each iteration as a parameter for our proposal distribution. The c_i, of course, are fixed. We are

unable to use a Metropolis-within-Gibbs sampling algorithm since that requires the proposal distribution to be symmetric. (I did also consider a Binomial proposal distribution, but in the short time frame I went with Poisson and didn't look back).

We use 20000 iterations with a burn-in of 2000. At each iteration we compute the following ratio

$$r = \frac{\pi_N(N_i^*|\lambda_i, \theta_i, \cdot)/J(N_i^*|N_i^{(t-1)})}{\pi_N(N_i^{(t-1)}|\lambda_i, \theta_i, \cdot)/J(N_i^{(t-1)}|N_i^*)}$$

Where $J(\cdot|\cdot)$ is our proposal distribution. We then set

$$N_i^{(t)} = \begin{cases} N_i^* & \text{with probability } min(r, 1) \\ N_i^{(t-1)} & \text{otherwise.} \end{cases}$$

which requires the generation of a uniform random number. Based on our hyperparameters and proposal density, we have an acceptance ratio that is approximately 80%. The ideal range is between 20-25%, but autotuning our acceptance rate to a target hasn't really been discussed so I am not going to do this.

3.3 M2: Hyperparameter Choice

For this model, we do have some prior knowledge of the population N_i of hawks in the sense that we know this number should be large (there's 1.9 million red hawks in North America alone according to a quick google search). I used this prior knowledge as a starting point. In other words, we need $\lambda_i c_i$ to be very high. Since c_i is fixed, we can only control the λ_i.

Since λ_i is just a latent variable with no interpretation, we can choose any α_λ and β_λ such that $E(\lambda_i)$ is large and $V(\lambda_i)$ is small. This was a trial and error assessment, similar to homework 2. Ultimately, I decide to set $\alpha_\lambda = 10000$ and $\beta_\lambda = 10$, which leads to $E(\lambda_i) = \frac{\alpha_\lambda}{\beta_\lambda} = 1000$ and $V(\lambda_i) = \frac{\alpha_\lambda}{\beta_\lambda^2} = 100$.

Next we look at the hyper parameter for θ_i, β_θ. We have to ensure that our prior mean for theta, $E(\theta) = \frac{1}{1+\beta_\theta}$, is proportional to the MLE, mean$(y_i)/N_i$. Thus, we select $\beta_\theta = 500$.

As one last sanity check, we can look at the mean of the data, $mean(y_i) = 376.1$ and compare it to the mean of the samples (i.e. $mean(N_i^{(1:iters)} \times \theta_i^{(1:iters)}) = 375.5692$). We see that these numbers are close, which means that our MCMC samples are converging to reasonable estimates.

3.4 M2: Parameter Inference

To check for mixing and convergence we examine the sampling traces for the parameters.

First we look at the population count N_i. Based on the trace plots, histograms, and autocorrelation plots for randomly selected years (Figures 4 - 6). We see that our M-H sampler is mixing and converging. This autocorrelation is clearly decreasing as the lag increases (i.e. samples can be considered as independent). We summarize a few years in table 2.

Year	Mean	2.5%	97.5%
1968	26002.57	25493.00	25659.00
1982	118043.31	115969.95	116633.00
1998	127982.51	125774.95	126487.00
2015	127026.29	124794.00	125512.00

Table 2: Summary of MCMC samples of population count by year.

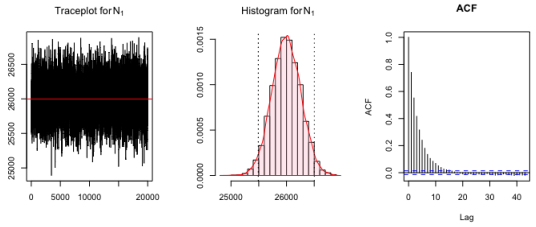

Figure 4: Posterior samples for model 2 for the year 1968. Left: Trace, Middle: Histogram with 95% credible intervals plotted vertically, Right: ACF for year 1968. For a better view and description see appendix. Just zoom it, man,

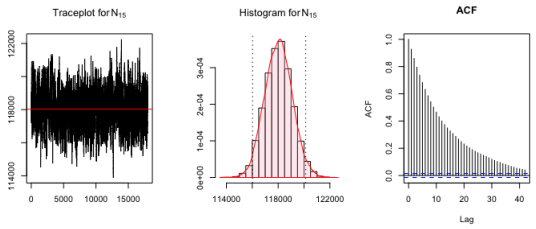

Figure 5: Trace and Histogram for population of Hawks at year 1982.

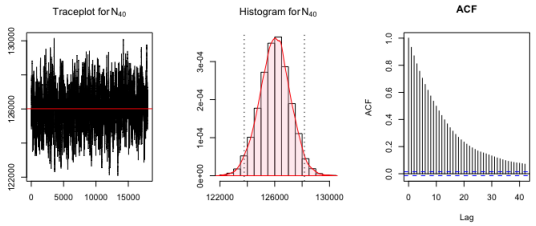

Figure 6: Trace and Histogram for population of Hawks at year 2015

Next we examine the posterior samples for θ_i. These are summarized in table 3. We see by the traceplots, histograms, and autocorrelation plots, that the conclusion is similar to N_i case.

Year	Mean	2.5%	97.5%
1981	0.0026	0.0024	0.0025
2004	0.0033	0.0030	0.0031
2012	0.0037	0.0034	0.0035
2015	0.0036	0.0033	0.0034

Table 3: Summary of posterior samples of θ_i

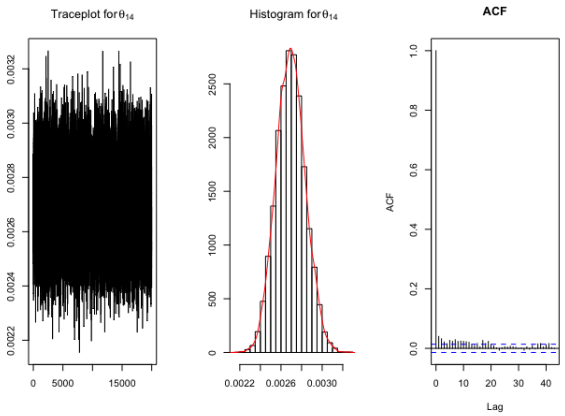

Figure 7: Probability of observing a hawk at year 1981.

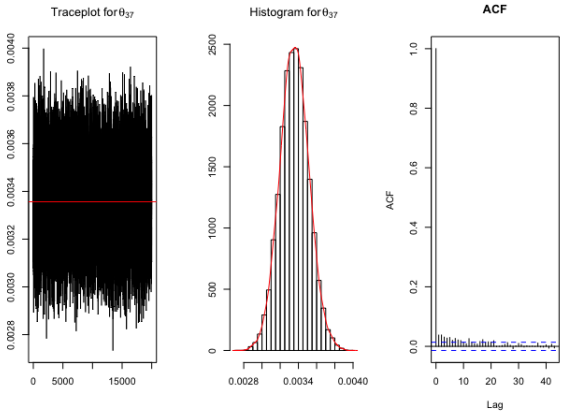

Figure 8: Probability of observing a hawk at year 2004.

Year	Mean	2.5%	97.5%
1984	999.93	982.83	988.26
1988	999.64	982.84	988.10
1993	1000.32	982.75	988.67
1994	999.95	982.52	988.52

Table 4: Summary of posterior samples of λ_i (these are chosen at random by the way, the same for θ_i and N_i).

Lastly, we examine the posterior samples for λ_i. Summary is in Table 4 and posterior traceplots, histograms and ACF plots for λ_i for randomly selected years are in figures 9 and 10.

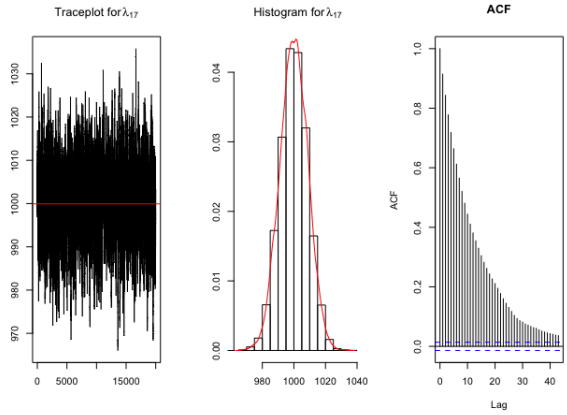

Figure 9: Posterior samples of λ_i for the year 1984.

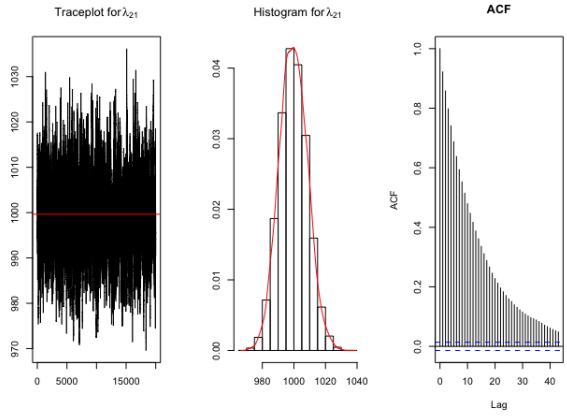

Figure 10: Posterior samples of λ_i for the year 1988.

4. Model Validation and Model Comparison

First, we should explain the hierarchical structure of model 2 a bit more. The random variable $y_i \sim Bin(N_i, \theta_i)$ has a distribution given by

$$P(Y_i = y_i) = \sum_{n_i=0}^{\infty} P(Y_i = y_i, N_i = n_i) \quad (1)$$

$$= \sum_{n_i=0}^{\infty} P(Y_i = y_i | N_i = n_i) P(Y_i = y_i) \quad (2)$$

$$= \sum_{n_i=y_i}^{\infty} \left[\left(\binom{n_i}{y_i} \right) \theta_i^{y_i} (1-\theta_i)^{n_i - y_i} \right] \quad (3)$$

$$\times \left[\frac{\exp\{-n_i\} (\lambda_i c_i)^{n_i}}{n_i!} \right] \quad (4)$$

$$= \frac{(\lambda_i c_i \theta_i)_i^y \exp\{-\lambda_i c_i\}}{n_i!} \quad (5)$$

$$\sum_{n_i=y_i}^{\infty} \frac{((1-\theta_i)\lambda_i c_i)^{n_i - y_i}}{(n_i - y_i)!} \quad (6)$$

$$= \frac{1}{y_i!} \left((\lambda_i c_i \theta_i)^{y_i} \exp\{-\lambda_i c_i \theta_i\} \right) \quad (7)$$

Thus, for model 2, $y_i \sim \text{Pois}(\lambda_i c_i \theta_i)$ and any inference on y_i is with respect to a $\text{Pois}(\lambda_i c_i \theta_i)$ distribution. Introducing N_i into the hierarchical model yields the following interpretation: on average, $Ey_i = \lambda_i c_i \theta_i$ red hawks are observed.

To obtain posterior predictive replicates for Model 2, we simulate 200 Poisson random variables using our route count data, c_i, and posterior samples of θ_i, λ_i. The comparison of the actual Red Hawk sightings to the posterior predictive replicates, y_i^{rep}, show that this model does a good job at fitting the data (Figure 12). Compared to our results using the prior specifications of model 1 (Figure 12), the difference is almost indistinguishable. Looking very, very closely there might be a slight improvement in predicting observations centered around the mean for model 2 (Figure 11).

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.*

Next, we use the Gelfand and Ghosh criterion to compare model 1 and model 2. For model 1, we get 1476.299 and for model 2 we get 1054.122. This value is smaller for model 2, which adds support to the earlier claim (Figure 11) that model 2 better describes the data.

Intuitively, model 2 seems like it should be higher because it is more complex since more latent variables are involved and GG penalizes complex models. However, I showed in equations (1-7) that any marginal inference on y_i is with respect to a $\text{Poisson}(\lambda_i c_i \theta_i)$ distribution, with N_i playing no part at all. Thus, the complexity of model 2 is not much different than model 1. So the biggest factor contributing to the GG criterion is simply good-ness of fit. I am confident that I have defended my argument that model 2 is better than model 1.

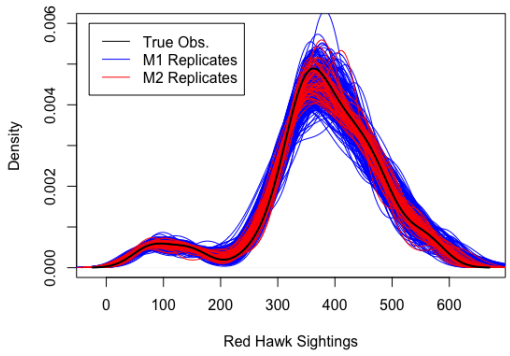

Figure 11: 200 samples from model 1 plotted against 200 samples from model 2. Here it is more obvious that model 2 *may* provide posterior predictive samples that are closer to the true observations.

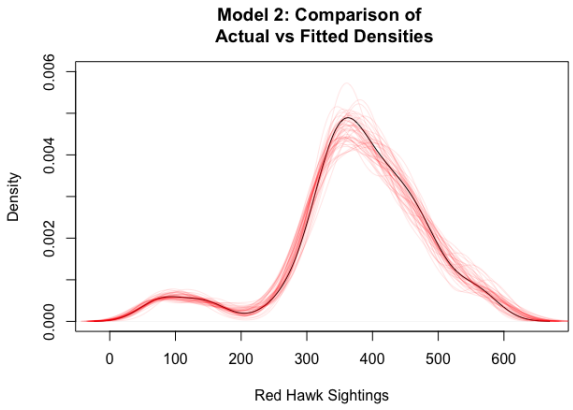

Figure 12: Distribution of the observed y_i and the distributions of some of the posterior replicates (y_i^{rep}) for Model 2. Looking very, very closely there might be a slight improvement in predicting observations centered around the mean for model 2.

Estimate the probability of observing more than 450 red hawks in California in a yea with a route count of 120:

Our model 2 (unlike my first attempt in exam 1) now includes the rate count. So what we are asked to predict is, given a completely new observation, for a new year with route count equal 120, what is the probability that the observed $y_i > 450$. To calculate this, I use 200 posterior predicted replicates. From those replicates, I count how many are greater than 450 and and divide by the total replicates. This ratio comes out to be 0.2273. Which

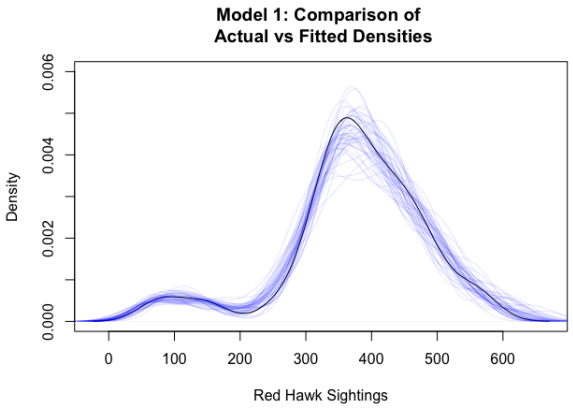

Figure 13: Distribution of the observed y_i and the distributions of some of the posterior replicates (y_i^{rep}) for Model 1.

is exactly what we discussed on exam 1.

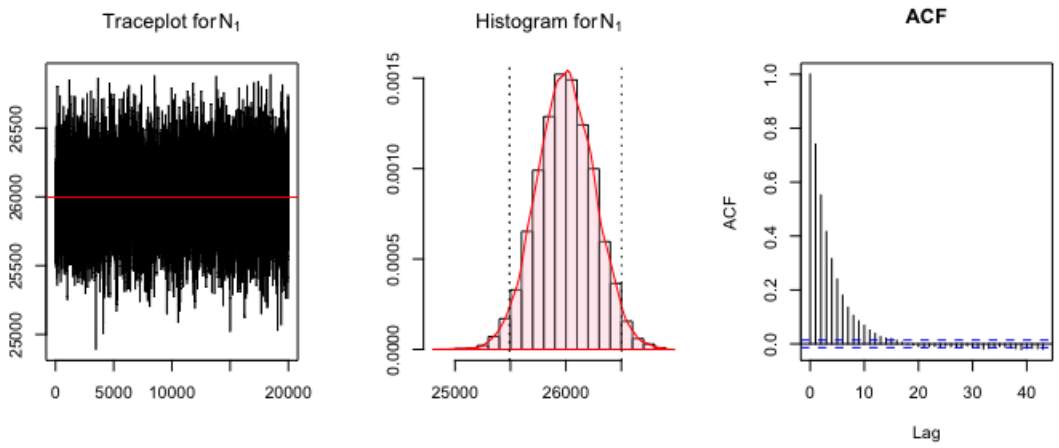

Figure 14: Left:Trace, Middle: Histogram with 95% credible intervals plotted vertically, Right: ACF for year 1968.