Bayesian models for Kilauea volcano interevent times

Wyara Vanesa Moura e Silva
AMS 207 MidTerm 1

Abstract

Kilauea is a currently active shield volcano in the Hawaiian Islands. The high level of activity of Klauea has a great impact on the ecology of its mountains. In this paper, we present two different models to analyze the interevent time days of a historical data from 1923 to 1983. The results found favored the simpler model in one criterion, which this model use the Normal distribution for the data. And in two criterion the complex model was favored, in which this model assign a Student-t distribution with a random variable in the mean. However, both models gets close predictions for the observed data.

KEY WORDS: volcano, model, predictive, comparison

Figure 1: Histogram of the Kilawea volcano dataset from interevent times in days. Red line correspond to the density.

1. Introduction

In this work, it will be applied two models in a dataset related to volcanic eruptions, such data have already been analyzed by Passarelli et al. (2010). We will analyze specifically, a data of interevent time in days from a volcano called Kilauea, which can be found in Passarelli et al. (2010). Kilauea volcano is the youngest volcano on the Big Island of Hawaii. The data used in this analysis from the eruption history it was limit to the 42 events after 1923 until 1983. Look at the paper (Passarelli et al., 2010) for more details about the data and the Kilauea volcano.

Figure 1 display the histogram for the log transformation of the interevent time data, and we can notice a asymmetric behavior in the distribution.

Some descriptive statistics about the original data would be that, $n = 41$, for the mean $\overline{T}_i = 528.805$ and for the standard deviation $\mathrm{sd}(T_i) = 1021.569$, and for the transformed data, the mean $\overline{\log T}_i = 5.513$ and for the standard deviation $\mathrm{sd}(\log T_i) = 1.246$.
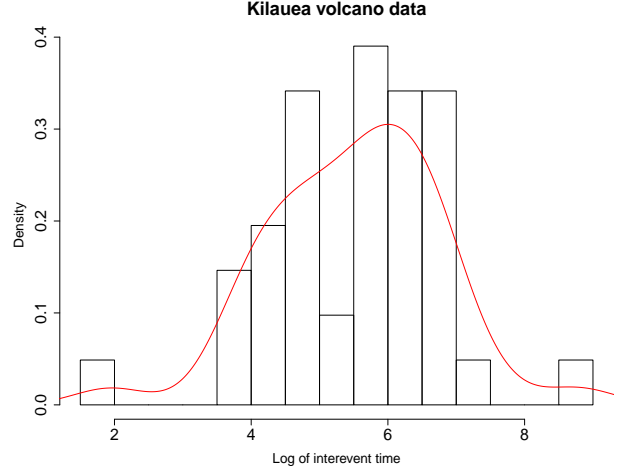
2. Methods

Denoting T_i the i-th observation of the interevent times in days, and given that this observations are too skew to be considering normal distributed (see Figure 1). Then, we are going to use a logarithm transformation in T_i, and the data will be modeling using a Student-t distribution and after that, we will apply another model for the data, now using a Normal distribution.

2.1 Model 1: Student t distribution

The first model using a Student-t distribution is setting by the following:

$$
\begin{aligned}
\log T_i &= \mu_i + \epsilon_i, & \epsilon_i &\sim \mathcal{St}_\nu(0, \sigma^2) \\
\mu_i &= \mu + v_i, & v_i &\sim \mathcal{N}(0, \tau^2)
\end{aligned}
$$

In which, it is fixed the degrees of freedom with $\nu = 5$. Now, the distribution for the random variables in function of the parameters that will be estimated are given by:

$$p(\log T_i | \mu_i, \sigma^2) \quad \sim \quad \mathcal{S}t_\nu(\mu_i, \sigma^2) \quad i = 1, ..., n$$

$$p(\mu_i | \mu) \quad \sim \quad \mathcal{N}(\mu, \tau^2) \quad i = 1, ..., n$$

In order to make easy the computation of the posterior distribution for the parameters, we are going to use a data augmentation technique, in which we can write the distribution for the $\log(T_i)$ as:

$$p(\log T_i | \mu_i, \sigma^2) \quad = \quad \int_0^\infty \mathcal{N}\left(\log T_i | \mu_i, \frac{\sigma^2}{\lambda_i}\right) \times$$

$$\mathcal{G}\left(\lambda_i | \frac{\nu}{2}, \frac{\nu}{2}\right) d\lambda_i$$

Now the model has the following form:

$$p(\log T_i | \lambda_i, \mu_i, \sigma^2) \quad \sim \quad \mathcal{N}\left(\mu_i, \frac{\sigma^2}{\lambda_i}\right)$$

$$p(\lambda_i | \nu) \quad \sim \quad \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

$$p(\mu_i | \mu, \tau^2) \quad \sim \quad \mathcal{N}\left(\mu, \tau^2\right)$$

We can complete the model by setting suitable priors for the parameters. The priors distribution chose are given below.

$$p(\mu | m, s^2) \quad \sim \quad \mathcal{N}\left(m, s^2\right)$$

$$p(\sigma^2 | a, b) \quad \sim \quad \mathcal{IG}\left(a, b\right)$$

$$p(\tau^2 | c, d) \quad \sim \quad \mathcal{IG}\left(c, d\right)$$

Then, we can find the full conditionals distributions for the parameters $\boldsymbol{\theta} = (\lambda_i, \mu_i, \mu, \sigma, \tau)$, for Model 1, which are given by:

$$\pi(\lambda_i | \cdot) \quad \sim \quad \mathcal{G}\left(\frac{\nu + 1}{2}, \frac{\nu}{2} + \frac{(\log T_i - \mu_i)^2}{2\sigma^2}\right)$$

$$\pi(\mu_i | \cdot) \quad \sim \quad \mathcal{N}\left(\frac{\lambda_i \tau^2 \log T_i + \mu \sigma^2}{\lambda_i \tau^2 + \sigma^2}, \frac{\tau^2 \sigma^2}{\lambda_i \tau^2 + \sigma^2}\right)$$

$$\pi(\mu | \cdot) \quad \sim \quad \mathcal{N}\left(\frac{s^2 \sum_{i=1}^n \mu_i + m \tau^2}{ns^2 + \tau^2}, \frac{\tau^2 s^2}{ns^2 + \tau^2}\right)$$

$$\pi(\sigma^2 | \cdot) \quad \sim \quad \mathcal{IG}\left(\frac{n}{2} + a, \frac{\sum_{i=1}^n (\log T_i - \mu_i)^2 \lambda_i}{2} + b\right)$$

$$\pi(\tau^2 | \cdot) \quad \sim \quad \mathcal{IG}\left(\frac{n}{2} + c, \frac{\sum_{i=1}^n (\mu_i - \mu)^2}{2} + d\right)$$

in which the $(\cdot)$ represents the data and all the other variables. And about the other notations: $\mathcal{S}t_\nu$ the Student-t density, $\mathcal{N}$ the normal density, $\mathcal{G}$ the gamma density, and $\mathcal{IG}$ the inverse gamma density.

2.2 Model 2: Normal distribution

The second model is a simplified version of the first model, in which now we use the Normal distribution, more formal, the model would be given by:

$$\log(T_i) \quad = \quad \mu + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Then, the distribution for the log transformation of the interevent would be:

$$p(\log(T_i) | \mu, \sigma^2) \quad \sim \quad \mathcal{N}(\mu, \sigma^2) \quad i = 1, ..., n$$

Setting suitable priors for the parameters, we can complete the model. The priors distribution chose are given below.

$$p(\mu | m, s^2) \quad \sim \quad \mathcal{N}\left(m, s^2\right)$$

$$p(\sigma^2 | a, b) \quad \sim \quad \mathcal{IG}\left(a, b\right)$$

Then, we can find the full conditionals distributions, for the parameters $\boldsymbol{\theta} = (\mu, \sigma^2)$, for the model, which are given by:

$$\pi(\mu | \cdot) \quad \sim \quad \mathcal{N}\left(\frac{s^2 \sum_{i=1}^n \log(T_i) + m\sigma^2}{ns^2 + \sigma^2}, \frac{s^2 \sigma^2}{ns^2 + \sigma^2}\right)$$

$$\pi(\sigma^2 | \cdot) \quad \sim \quad \mathcal{IG}\left(\frac{n}{2} + a, \frac{\sum_{i=1}^n (\log(T_i) - \mu)^2}{2} + b\right)$$

in which the $(\cdot)$ represents the data and all the other variables. And about the other notations: $\mathcal{N}$ the normal density and $\mathcal{IG}$ the inverse gamma density.

The choice for the distributions of the priors was done using the fact that the normal distribution seems to fit good for a distribution of the mean, and Inverse Gamma distribution for the variances of the errors, because is a very flexible distribution.

Now we can derive a Gibbs sampling to explore the posterior distribution for all the parameters. The idea is to generate posterior samples by sweeping through each variable to sample from its conditional distribution with the remaining variables fixed to their current values (Gelman et al., 2014). Inside the Gibbs sampling, we updated the parameters in the following order: λ_i, μ_i, μ, τ^2 and σ^2, for Model 1 and μ and σ^2 for Model 2. The steps used to find those full conditionals are given in the Appendix.

2

2.3 Model Checking and Model Comparison

2.3.1 Model Checking

In order to perform the model checking we are going to use first the basic technique which consist of draw simulated values from the joint posterior predictive distribution of replicated data and compare these samples to the observed data. Then, we are going to choose some test quantity $T(y)$ to compare the observed data to the posterior predictive distribution. And then we calculate The Bayesian p-value which is defined by:

$$p_B = \Pr(T(y^{rep}) > T(y))$$

in which the p-value correspond to the proportion of the sampled replicate statistics above the observed value from the data.

The second approach to model checking was done using leave one out cross-validation predictive distributions, for each future data point comparing to the inference given all the other data:

$$p_i = \Pr(y_i^{rep} \leq y_i | y_{-i})$$

where y_{-i} contains all other data except y_i. And in the case of our data, the cross-validation predictive p-values should have uniform distribution if the model is acceptable (Gelman et al., 2014). As the data is not too large, than it should not be a big problem to compute.

2.3.2 Model Comparison

For the model comparison it was used three different criterion. First it was perform the deviance information criterion (DIC) which is a criterion that use the information from the posterior distribution of the parameters. This criterion used in the analysis was setting by:

$$DIC = \overline{D}(\theta) + 2p_D$$

in which $D(\theta) = -2\log(f(\mathbf{y}|\theta))$ is the deviance and the $\overline{D}$ is the mean of the posterior over all the θ sampled. And the $p_D = \mathbb{V}ar\,(D(\theta))/2$ which is the sample variance of the $D(\theta)$ divided by 2 (Gelman et al., 2014).

The second will be the Gelfand and Gosh criterion, which is known as posterior predictive loss approach, This method provides an index of the predictive capacity of a model comparing observed data with replicated data. This criterion is defined by:

$$D = G + P = \sum_{l=1}^{n}(y_l - \mu_l)^2 + \sum_{l=1}^{n}\sigma_l^2$$

which $\mu_l = \mathbb{E}(y_l^{rep}|\mathbf{y})$ and $\sigma_l^2 = \mathbb{V}ar(y_l^{rep}|\mathbf{y})$, in which y_l^{rep} are the replicas of the posterior values for each data point (Gelfand et al. 1998).

The Bayes Factor was the last approach to compare the models, this method is defined by:

$$BF_{12} = \frac{\int_{\Theta_1} f_1(\mathbf{y}|\theta)\pi_1(\theta)d\theta}{\int_{\Theta_2} f_2(\mathbf{y}|\theta)\pi_2(\theta)d\theta} = \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}$$

A good way to perfom the Bayes Factor would be using the method by simulations of the marginals $(f_j(\mathbf{y}))$. This method is known as the naive Monte Carlo estimator of marginal likelihood (Gronau et al. 2017).

Basically we will sample θ_i^k, $k = 1, ..., s$, from the prior density $\pi_i(\theta_i)$ and we use the approximation:

$$f_i(\mathbf{y}) = \frac{1}{s}\sum_{k=1}^{s} f_i(\mathbf{y}|\theta_i^k)$$

in which this s has to be really large to the approximation to make sense. In here, we chose $s = 1$ million. By doing this we get an approximation for the Bayes Factor (Christense et al., 2011).

3. Results

In the following results, it was obtained $M = 10000$ posterior samples after the burn-in period, for the two models. Some information about the convergence will be given in Subsection 3.1. Figures 2 and 3 display the posterior densities of the parameters of interesting for Model 1 and Model 2, in those densities it was draw in red line the credibility intervals and the posterior mean.

It was assumed independence between the priors distribution for the parameters assigned. And the choices for the hyperparameters, m, s^2, a, b, c and d were assigned in order to be as non-informative as possible. For the $m = 5$ which is around the mean of the data, the $s^2 = 8$ which is a very vague parameter for the variance. And the $a = c = b = d = 3$, which result in a mean and variance for the prior distribution of σ^2 and τ^2 around 1.5 and 2.25.
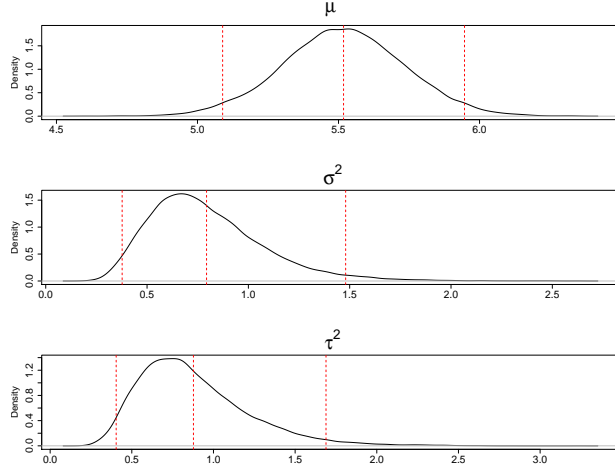
Figure 2: Posterior densities for the parameters of Model 1. Red dashed line correspond to the credibility interval (2.5%;97.5%) and mean.
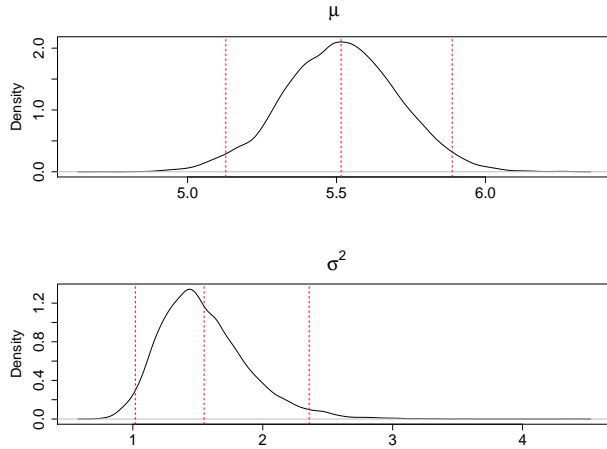

Figure 3: Posterior densities for the parameters of Model 2. Red dashed line correspond to the credibility interval (2.5%;97.5%) and mean.

The densities for the λ_i and μ_i were not shown here because there was not any issue, those densities appear to have the same behavior, the only thing different observed was that, if we marginalize each λ_i and μ_i the densities for the observations 9-*th* and 15-*th* behaved differently from other observations. These two observations are the maximum and the minimum of the data, respectively. Therefore, the model correctly captures the extremes.

3.1 Convergence Checking

Figure 4 shows the plots for the Markov chain for the three parameters of interesting and a plot of the autocorrelation function. As well as Figure 5 shows the same plots for Model 2. It was sample a $M = 41000$ and the burn-in chose was $B = 1000$ and it was thinning by 4. We can notice that for each model the chain converge and the autocorrelation is not too high.
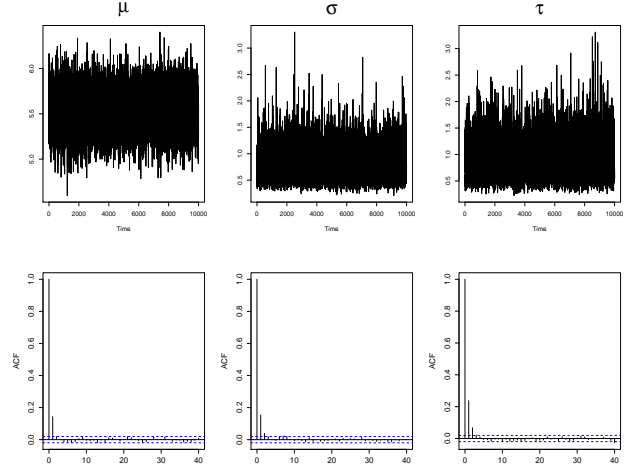

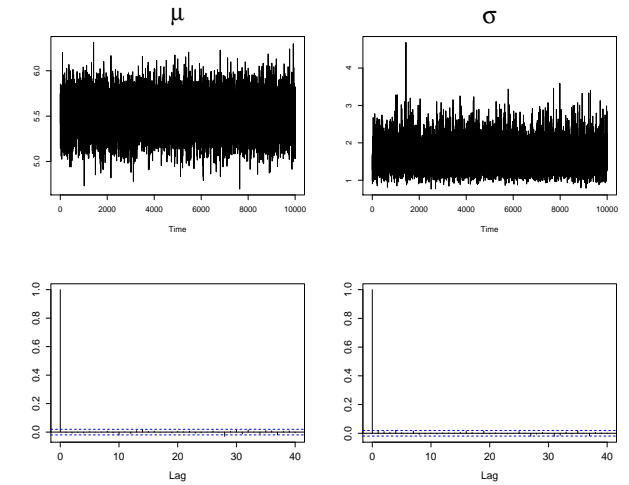Figure 4: Markov Chain and Autocorrelation plots for Model 1.

Figure 5: Markov Chain and Autocorrelation plots for Model 2.

For the others, parameters corresponding to λ_i and

4

μ_i were not observed any issue with concerning to the convergence. Then, it was chosen not to show the chains plots for those parameters.

3.2 Model Checking

We simulate 10.000 samples from the posterior predictive distribution, which in this case we can do the prediction for the $\log T_i$ observation using the posterior distribution for the parameters. For Model 1, the posterior distribution for λ_i, μ_i and σ^2. And for Model 2, just the posterior of μ and σ^2. We then calculated for both models, statistics from predictions samples and statistics from the observed data, which were: the maximum, the minimum, the mean and the standard deviation.
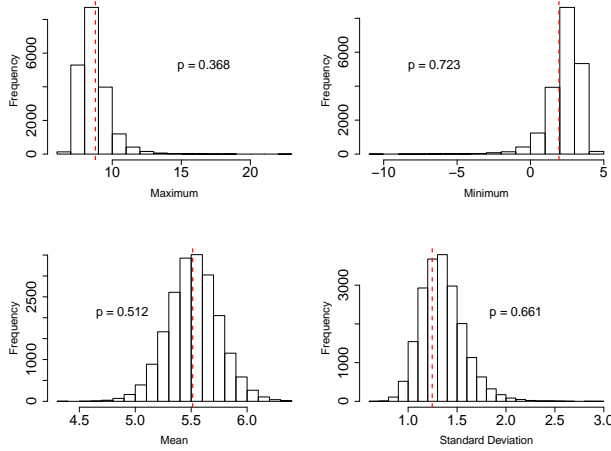

Figure 6: Histogram of samples from the posterior predictive distributions of some summaries of the interevent times for the Kilauea volcano, for Model 1. Red dashed lines corresponds to the observed values. Also the p-values from the comparison.

Figures 6 and 7 shows the comparisons between the histograms of the statistics from predictive posterior distributions and the corresponding observed values from data. We can see the plots and the p-value that the predictions are in good agreement with observed values for all four statistics.

From the graphs in Figures 6 and 7 we can notice that the maximum has the smallest p-value in the two models, and specifically in Model 2 the p-value is about 0.181 which can indicate a slight disagreement with the data, but not too strong.

Passarelli et al. (2010) point out that the maximum corresponds to the 18 years of quiescence of the

Kilauea volcano (i.e. 19341952 AD). And this could be the reason for this slight discrepancy of the models. In which the minimum on the interevent time is the 15-th observations corresponding to $T_{15} = 7$ ($\log T_{15} = 1.945$). And the maximum is the 9-th observations and is equal to $T_9 = 6504$ ($\log T_9 = 8.780$).
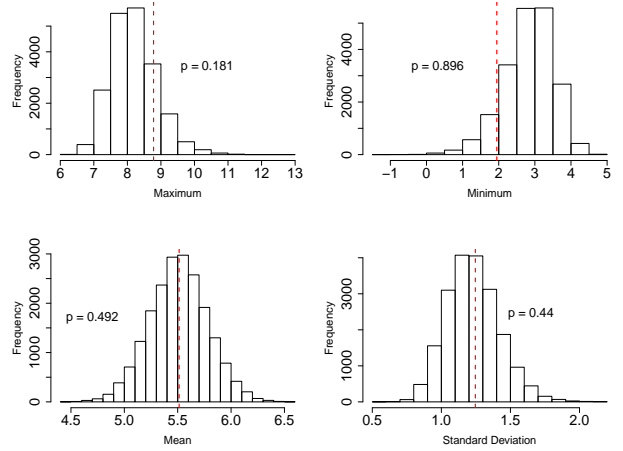

Figure 7: Histogram of samples from the posterior predictive distributions of some summaries of the interevent times for the Kilauea volcano, for Model 2. Red dashed lines corresponds to the observed values. Also the p-values from the comparison

Figures 8 and 9 show the posterior predictive densities for each observation from a amount of 50 replicas for Model 1 and 2. What we can point out looking these densities is that in Model 1 the posterior predictive has heavier tails than Model 2, which is something expected as we modeled Model 1 with a Student-t distribution that has thicker tails than a Normal distribution.

The second approach using leave-one-out cross validation we achieve larger p-values ($p_{M1} = 0.8802$ and $p_{M2} = 0.8086$) when we test the probabilities from the approach using a Kolmogorov-Smirnov test (it was used the function `ks.test` in R), in order to test if the results are uniformly distributed. In Figure 10 is shown the densities for the probabilities obtained by the method for both methods, which have very similar behavior. Therefore, with this outcomes we can conclude by this approach that the models fits well the data.
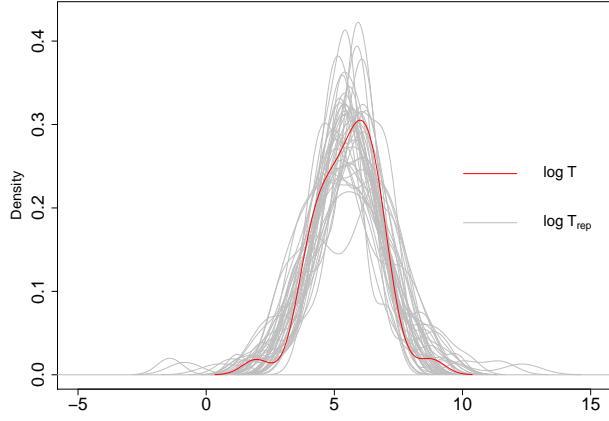
Figure 8: Densities corresponding to 50 replicas from the posterior predictive distribution of each observation for Model 1. Red line the density of the observed data. Grey lines the posterior predictive density for replicas.
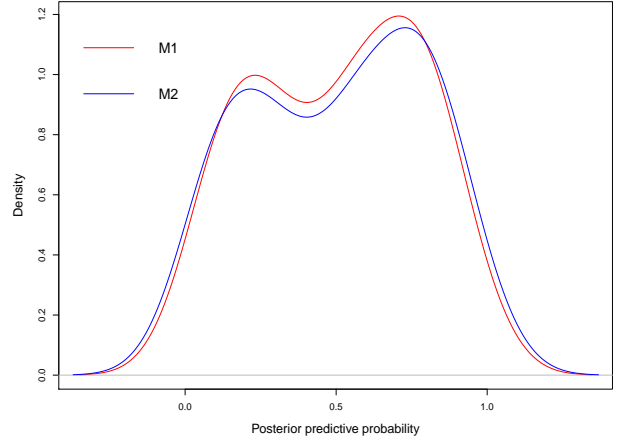

Figure 10: Densities from the leave-on-out cross validated method for both models.

Now, in order to predict future observations for the models, first for Model 1 we must sample a μ_{ir} from the $\mathcal{N}(\mu, \tau^2)$ using our posterior distribution of the μ and τ^2, and in this work we sample λ_{ir} from $\mathcal{G}(\nu/2, \nu/2)$ and then sample $\log T_{ir}$ from the $\mathcal{N}(\mu_{ir}, \sigma^2/\lambda_{ir})$. Moreover, for Model 2, we sample the Normal distribution using the posterior distribution for the parameters, μ and σ^2. Figure 11 shows the posterior predictive densities for the future observations for the two models.
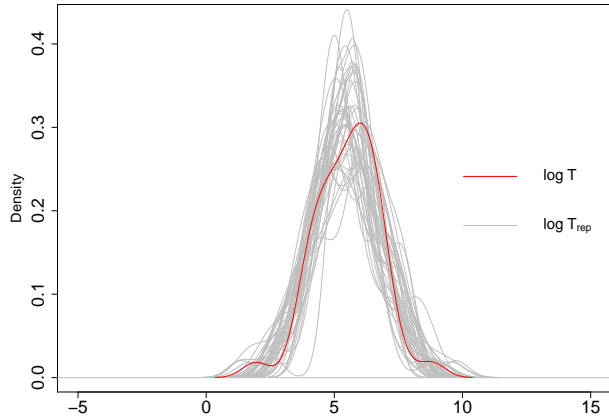

Figure 9: Densities corresponding to 50 replicas from the posterior predictive distribution of each observation for Model 2. Red line the density of the observed data. Grey lines the posterior predictive density for replicas.
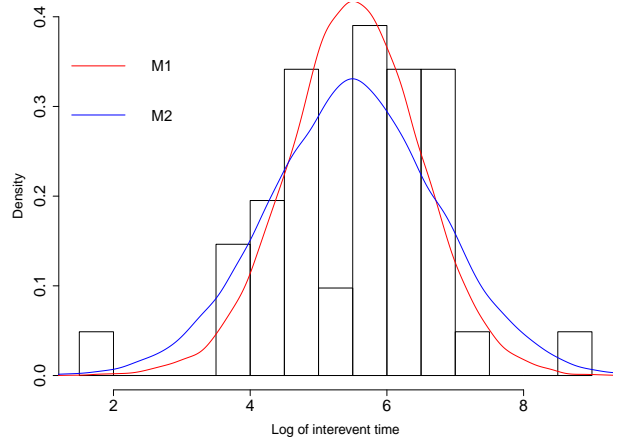

Figure 11: Histogram of interevent time. Red line the posterior predictive for future observation of Model 1. Blue line the posterior predictive density for future observation of Model 2.

3.3 Model Comparison

Table 1 has the values for the criteria results of the DIC, Gelfand and Gosh and the Bayes Factor. In summary, looking at the results from the DIC method we can notice that this criteria favor Model 2, but if we evaluate the goodness of fit, it can be seen that it favor Model 1, and for the Effective size (penalty) Model 2 has the lower value.

For the Gelfand and Gosh criterion, it put more beliefs in favor of Model 1, as can be seen in Table 1. For the goodness of fit, the Model 1 is a little smaller than Model 2, favoring Model 1 then. But, the penalty for both models have close values, that is the reason that this criterion the belief of favored are more strong in Model 1 (see Table 1).

Perhaps the reason for this is because this criterion uses posterior predictive distribution, and as we saw in Figures 8 and 9, Model 2 has better predictions for extremes values in the tails of the distribution.

Now evaluating the Bayes Factor, we can conclude that this approach puts more beliefs in favor of Model 1, as the value for BF_{12} is less than 1. However, Bayes Factor is not very reliable given that we are only making an approximation. And also the bayes factor is very affected by the choice of the prior.

And even for the other criteria, after making a quickly prior sensitivity analysis, we could notice that the results can be affected by the choice of the parameters of the distributions, but as said before, we tried to make the priors less informative possible.

Table 1: Criterion values for the model comparison.

Criterion		Model 1	Model 2
DIC		193.882	137.258
	GF	107.261	135.338
	ES	86.621	1.919
GG		85.681	124.690
	GF	20.061	64.619
	P	60.958	63.763
BF (Beliefs in)		1.313	0.761

Note: GG: Gelfand and Gosh; GF: Goodness of Fit; ES: Effective Size; P: Penalty

The results from the Gelfand and Gosh and the Bayes Factor seems reasonable given that the Model 1 is the one that reflect better extreme values in the distribution of the data.

4. Remarks

In this work, we perform two different models in order to evaluate a data of interevent of time from Kilauea volcano. The first model was a more elaborate model, using the Student-t distribution to model the data to try to fit better the thicker tails in the observed data. The second model uses the Normal distribution for modeling and it is very simple compared to Model 1. It was done a model checking which the results do not show any significant behavior that could assign a bad fit for both models. Then, after a model comparison, we found that Model 2 was better using DIC. Although the Gelfand and Gosh criterion and Bayes Factor favored Model 1, it was not a considered big difference between the models. The choice for Model 1 seems reasonable given that the model gets better inferences and predictions.

5. Appendix

5.1 Full Conditionals Distribution

In this section, it will be showing some steps to find the full conditional distribution of the parameters for $\boldsymbol{\theta} = \lambda_i, \mu_i, \mu, \sigma^2$ and τ^2, in this case for Model 1.

First, we need the joint posterior distribution, which is given by:

$$\pi(\boldsymbol{\theta}|\log \mathbf{T}) \quad \propto \quad \pi(\lambda_i, \mu_i, \mu, \sigma^2, \tau^2)\pi(\mathbf{y}|\boldsymbol{\theta})$$

$$\begin{aligned}
\pi(\boldsymbol{\theta}|\log \mathbf{T}) \quad \propto \quad & \prod_{i=1}^{n} \left(\frac{\lambda_i}{\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda_i(\log T_i - \mu_i)^2}{2\sigma^2}\right\} \\
& \times \prod_{i=1}^{n} (\lambda)^{\frac{\nu}{2}-1} \exp\left\{-\lambda_i\frac{\nu}{2}\right\} \\
& \times \prod_{i=1}^{n} \left(\tau^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{(\mu_i - \mu)^2}{2\tau^2}\right\} \\
& \times \exp\left\{-\frac{(\mu - m)^2}{2s^2}\right\}(\sigma^2)^{-a-1}\exp\left\{-\frac{b}{\sigma^2}\right\} \\
& \times \left(\tau^2\right)^{-c-1}\exp\left\{-\frac{d}{\tau^2}\right\}
\end{aligned}$$

Then, we can find all the full conditionals distribution, in which the $(\cdot)$ represents the data and all the other variables, $\mathcal{N}$ the normal density, $\mathcal{G}$ the gamma density, and $\mathcal{IG}$ the inverse gamma density, the results are following:

- For λ_i

$$\pi(\lambda_i|\cdot) \quad \propto \quad \prod_{i=1}^{n}(\lambda_i)^{\frac{1}{2}}\exp\left\{-\frac{\lambda_i(\log T_i - \mu_i)^2}{2\sigma^2}\right\}$$
$$\times \quad \prod_{i=1}^{n}(\lambda)^{\frac{\nu}{2}-1}\exp\left\{-\lambda_i\frac{\nu}{2}\right\}$$

$$\propto \quad \prod_{i=1}^{n}\left(\frac{\lambda_i}{\sigma^2}\right)^{\frac{\nu+1}{2}-1}$$
$$\times \quad \exp\left\{-\lambda_i\left(\frac{\nu}{2}+\frac{(\log T_i - \mu_i)^2}{2\sigma^2}\right)\right\}$$

$$\sim \quad \mathcal{G}\left(\frac{\nu+1}{2},\frac{\nu}{2}+\frac{(\log T_i - \mu_i)^2}{2\sigma^2}\right)$$

$$\pi(\sigma^2|\cdot) \quad \propto \quad \prod_{i=1}^{n}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}\exp\left\{-\frac{\lambda_i(\log T_i - \mu_i)^2}{2\sigma^2}\right\}$$
$$\times \quad (\sigma^2)^{-a-1}\exp\left\{-\frac{b}{\sigma^2}\right\}$$

$$\propto \quad (\sigma^2)^{-(n/2+a)-1}$$
$$\times \quad \exp\left\{-\frac{\sum_{i=1}^{n}(\log T_i - \mu_i)^2\lambda_i}{2\sigma^2}-\frac{b}{\sigma^2}\right\}$$

$$\sim \quad \mathcal{IG}\left(\frac{n}{2}+a,\frac{\sum_{i=1}^{n}(\log T_i - \mu_i)^2\lambda_i}{2}+b\right)$$

- For μ_i

- For τ^2

$$\pi(\mu_i|\cdot) \quad \propto \quad \prod_{i=1}^{n}\exp\left\{-\frac{\lambda_i(\log T_i - \mu_i)^2}{2\sigma^2}\right\}$$
$$\times \quad \prod_{i=1}^{n}\exp\left\{-\frac{(\mu_i - \mu)^2}{2\tau^2}\right\}$$

$$\propto \quad \prod_{i=1}^{n}\exp\left\{-\left[\frac{\lambda_i\tau^2+\sigma^2}{2\tau^2\sigma^2}(\mu_i^2)\right]\right\}$$
$$\times \quad \exp\left\{-\left[\frac{\lambda_i\tau^2+\sigma^2}{2\tau^2\sigma^2}\left(-2\mu_i\frac{\lambda_i\log T_i\tau^2+\mu\sigma^2}{\lambda_i\tau^2+\sigma^2}\right)\right]\right\}$$

$$\sim \quad \mathcal{N}\left(\frac{\lambda_i\tau^2\log T_i+\mu\sigma^2}{\lambda_i\tau^2+\sigma^2},\frac{\tau^2\sigma^2}{\lambda_i\tau^2+\sigma^2}\right)$$

$$\pi(\tau^2|\cdot) \quad \propto \quad \prod_{i=1}^{n}(\tau^2)^{-\frac{1}{2}}\exp\left\{-\frac{(\mu_i - \mu)^2}{2\tau^2}\right\}$$
$$\times \quad (\tau^2)^{-c-1}\exp\left\{-\frac{d}{\tau^2}\right\}$$

$$\propto \quad (\tau^2)^{-(\frac{n}{2}+c)-1}\exp\left\{-\frac{\sum_{i=1}^{n}(\mu_i - \mu)^2}{2\tau^2}-\frac{d}{\tau^2}\right\}$$

$$\sim \quad \mathcal{IG}\left(\frac{n}{2}+c,\frac{\sum_{i=1}^{n}(\mu_i - \mu)^2}{2}+d\right)$$

- For μ

Now, for $\boldsymbol{\theta} = \mu$ and σ^2, corresponding to Model 1. First, we need the joint posterior distribution, which is given by:

$$\pi(\mu|\cdot) \quad \propto \quad \prod_{i=1}^{n}\exp\left\{-\frac{(\mu_i - \mu)^2}{2\tau^2}\right\}$$
$$\times \quad \exp\left\{-\frac{(\mu - m)^2}{2s^2}\right\}$$

$$\propto \quad \exp\left\{-\left[\frac{\tau^2+ns^2}{2\tau^2s^2}(\mu^2)\right]\right\}$$
$$\times \quad \exp\left\{-\left[\frac{\tau^2+ns^2}{2\tau^2s^2}\left(-2\mu\frac{s^2\sum_{i-1}^{n}\mu_i+m\tau^2}{2\tau^2+ns^2}\right)\right]\right\}$$

$$\sim \quad \mathcal{N}\left(\frac{s^2\sum_{i=1}^{n}\mu_i+m\tau^2}{ns^2+\tau^2},\frac{\tau^2s^2}{ns^2+\tau^2}\right)$$

$$\pi(\boldsymbol{\theta}|\log\mathbf{T}) \quad \propto \quad \pi(\mu,\sigma^2)\pi(\mathbf{y}|\boldsymbol{\theta})$$

$$\pi(\boldsymbol{\theta}|\log\mathbf{T}) \quad \propto \quad \prod_{i=1}^{n}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}\exp\left\{-\frac{(\log T_i - \mu)^2}{2\sigma^2}\right\}$$
$$\times \quad \exp\left\{-\frac{(\mu - m)^2}{2s^2}\right\}(\sigma^2)^{-a-1}\exp\left\{-\frac{b}{\sigma^2}\right\}$$

Then, we can find all the full conditionals distribution, in which the $(\cdot)$ represents the data and all the other variables, $\mathcal{N}$ the normal density and $\mathcal{IG}$ the inverse gamma density, the results are following:

- For σ^2

- For μ

$$
\begin{aligned}
\pi(\mu_i|\cdot) &\propto \prod_{i=1}^{n} \exp\left\{-\frac{(\log T_i - \mu)^2}{2\sigma^2}\right\} \\
&\times \prod_{i=1}^{n} \exp\left\{-\frac{(\mu - m)^2}{2s^2}\right\} \\[2mm]
&\propto \exp\left\{-\left[\frac{ns^2 + \sigma^2}{2s^2\sigma^2}(\mu^2)\right]\right\} \\
&\times \exp\left\{-\left[\frac{ns^2 + \sigma^2}{2s^2\sigma^2}\left(-2\mu\frac{\sum_{i=1}^{n}(\log T_i)s^2 + m\sigma^2}{ns^2 + \sigma^2}\right)\right]\right\} \\[2mm]
&\sim \mathcal{N}\left(\frac{s^2\sum_{i=1}^{n}\log(T_i) + m\sigma^2}{ns^2 + \sigma^2}, \frac{s^2\sigma^2}{ns^2 + \sigma^2}\right)
\end{aligned}
$$

- For σ^2

$$
\begin{aligned}
\pi(\sigma^2|\cdot) &\propto \prod_{i=1}^{n}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}\exp\left\{-\frac{(\log T_i - \mu)^2}{2\sigma^2}\right\} \\
&\times (\sigma^2)^{-a-1}\exp\left\{-\frac{b}{\sigma^2}\right\} \\[2mm]
&\propto (\sigma^2)^{-(n/2+a)-1} \\
&\times \exp\left\{-\frac{\sum_{i=1}^{n}(\log T_i - \mu_i)^2\lambda_i}{2\sigma^2} - \frac{b}{\sigma^2}\right\} \\[2mm]
&\sim \mathcal{IG}\left(\frac{n}{2} + a, \frac{\sum_{i=1}^{n}(\log(T_i) - \mu)^2}{2} + b\right)
\end{aligned}
$$

REFERENCES

Christensen, R., Johnson, W., Branscum, A., & Hanson, T. E. (2011). Bayesian ideas and data analysis: an introduction for scientists and statisticians. CRC Press.

Gelfand, A. E., & Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. Biometrika, 85(1), 1-11.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian data analysis (Vol. 3). Boca Raton, FL: CRC press.

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D.S., Forster, J.J., Wagenmakers, E.J. & Steingroever, H. (2017). A tutorial on bridge sampling. Journal of mathematical psychology, 81, 80-97.

Passarelli, L., Sans, B., Sandri, L., & Marzocchi, W. (2010). Testing forecasts of a new Bayesian time-predictable model of eruption occurrence. Journal of Volcanology and Geothermal Research, 198(1-2), 57-75.