University of Milan

Faculty of Political, Economic and Social Sciences

A Bayesian Approach to Aggregate Insurance Claim Modeling

Final Project in the Subject Bayesian Analysis

Julia Maria Wdowinska (43288A)
Edoardo Zanone (33927A)

Data Science for Economics
II Year
Master's Degree

April 29, 2025

Contents

1 Introduction

2 Theoretical Considerations

The goal is to apply Bayesian statistics to our models. The Bayesian approach differs from the classical frequentist one, where, usually, the sample size is large and the parameter θ is assumed to be fixed.

In contrast, according to the Bayesian methodology, θ is treated as a random variable with prior distribution $\pi(\theta)$. The information available about the possible values of θ, prior to observing the data $\vec{y}$, is summarized by $\pi(\theta)$.

Both approaches to statistics have their own strengths and weaknesses, and, in practice, they often complement each other. A common critique of the Bayesian method is the subjectivity involved in choosing a prior distribution. There are various different methods for selecting the right priors. However, as the sample size increases, the influence of the prior diminishes, and the outcomes of the two methods tend to converge. Although the classical approach is often computationally easier, it can sometimes produce inconsistent results, especially in the context of hypothesis testing. In contrast, the Bayesian method offers coherent and flexible framework by incorporating prior knowledge and providing full posterior distributions. Even though the choice of the prior may arise potential inconsistencies, Bayesian inference tends to be more advantageous in situations with limited or complex data. However, in addition to the subjectivity of the priors, Bayesian methods can be computationally intense and complex, moreover, when dealing with complex models or non-conjugate priors.

Through the use of Bayes' Theorem, the ultimate aim is to find the posterior distribution of our parameter θ:

$$\pi(\theta|y) = \frac{f(\vec{y}|\theta)\pi(\theta)}{\int f(\vec{y}|\theta)\pi(\theta)\,d\theta},$$

where $f(\vec{y}|\vec{\theta})$ is the likelihood function. Given that the denominator is constant with respect to θ it is possible to simplify the expression to :

$$\pi(\theta|\vec{y}) \propto \pi(\theta)f(\vec{y}|\theta).$$

Usually, to simplify computations, it is common to use conjugate distributions. A family $\mathcal{F}$ is said to be conjugate to the likelihood if, for every prior that belongs to $\mathcal{F}$, the posterior distribution of θ also belongs to the same family.

When the posterior distributions are analytically intractable, the use of *Markov Chain Monte Carlo* (MCMC) methods becomes essential. The main goal of MCMC is to draw samples from a probability distribution from which direct sampling is difficult.

Among the different algorithms available for implementing MCMC, two are considered here: the *Gibbs sampler* and the *Metropolis-Hastings algorithm.*

The first algorithm mentioned aims to find the full conditional posterior distributions starting from the joint or marginal posterior distributions. It is particularly efficient when these conditional distributions have a closed form. Hence, conjugate priors are often preferred. The algorithm proceeds through a clear and structured series of steps:

- Fix the initial value of the parameter $\vec{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$ and initialize the iteration counter as $m = 1$;
- Obtain $\theta^{(0)}$ such that $\pi(\theta^{(m)}1|\theta_2^{(m-1)}, \ldots, \theta_p^{(m-1)}, \vec{y}) \propto f(\vec{y}|\vec{\theta}^{(m-1)})\pi(\vec{\theta}^{(m-1)}$, now draw $\theta_2^{(m)}$ from the full conditional posterior distribution $\pi(\theta_2^{(m)}|\theta_1^{(m)}, \theta_3^{(m-1)} \ldots, \theta_p^{(m-1)}, \vec{y})$, and so on until $\pi(\theta_p^{(m)}|\theta_1^{(m)}, \ldots, \theta^{(m)}p - 1, \vec{y})$;
- Increment the counter from m to $m + 1$ and return to the first step.

The Metropolis-Hastings algorithm, on the other hand, is used when the full conditional posterior distribution is not in closed form. Suppose a realization from the full conditional distribution is desired: a proposal distribution with density $q(\theta^*|\theta)$ is used to generate candidate values θ^* from the current state θ. This algorithm also follows a clear and structured procedure:

- Initialize the chain θ_0 and the iteration counter;

- Generate a candidate value θ^* from a proposal distribution $q(\theta^*|\theta)$;
- Evaluate the acceptance probability $p(\theta \cdot \theta^{(j-1)})$ of the proposed draw $p(\theta \cdot \theta^{(j-1)}) = min(1, \frac{\pi(\theta|\vec{y})}{\pi(\theta^{(j-1)}|\vec{y})} \frac{q(\theta^{(j-1)}|\theta)}{p(\theta^*|\theta^{(j-1)})}$;
- Draw a uniform value $U \sim U(0,1)$, set $\theta^{(j)} = \theta^*$ if $U < p(\theta^*, \theta^{(j-1)})$, otherwise set $\theta^{(j)} = \theta^{(j-1)}$;
- Update the iteration counter and return to the first step.

After running the chain, the successive samples are likely to be correlated. Therefore, to obtain a random sample, a thinning process is applied, which consists of keeping one iteration and discarding the following m.

Finally, the chains must run long enough to reach equilibrium, a phase known as *burn-in*.

There are several methods to determine whether a chain is converging adequately. Two general approaches exist: one involves running a single chain for a larger number of iterations, while the other involves running multiple chains for a shorter period of time. Various diagnostics to check if the chain actually reached convergence have been developed, including visual inspection, the Gelman-Rubin test and the Geweke diagnostic.

Visually inspection consist in assessing how well the chain mixes or moves through the parameter space. If the chain moves slowly then the convergence will also be slow. It's important to inspect all the model parameters, since some may converge while others do not. The traceplot, which shows sampled parameter values over iterations, helps detect poor mixing. Thanks to this visualization it is possible to see if a chain get stuck in a certain area of the parameter space, which is the result of a bad mixing. Another useful tool is the running mean plot, which plots the cumulative average of the samples across iterations and helps evaluate stability over time.

Looking at the autocorrelation function (A.C.F) can help assessing the convergence. It represents the correlation among the draws of the Markov Chain. High autocorrelation implies strong dependence and slow mixing. In such cases, thinning can help reducing the correlation.

The Gelman-Rublin Test, which works with multiple chains, follows predetermined steps:

- Run the m different chains of length 2N, where N is equal to the number of iterations of the MCMC algorithm.
- Discard the number of the burn-in iterations (the first N draws?)
- Compute the within-chain variance and the between-chain variance, which are, respectively, $W = \frac{1}{m} \sum_{j=1}^{m} (s_j^2)$ and $B = \frac{N}{m-1} \sum_{j=1}^{m} (\overline{\theta_j} - \overline{\overline{\theta}})^2$, where $s_j^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\theta_{ij} - \overline{\theta_j})^2, \overline{\theta_j} = \frac{1}{N} \sum_{i=1}^{N} (\theta_{ij})$ and $\overline{\overline{\theta}} = \frac{1}{N} \sum_{j=1}^{m} \overline{\theta_j}$
- Compute the estimated variance of the parameter as the weighted average between the sum of between-variance anche within-variance $\hat{var}(\theta) = (1 - \frac{1}{N})W + \frac{1}{N}B$
- Calculate the Potential Scale Reduction factor $\hat{R} = \sqrt{\frac{\hat{var}(\theta)}{W}}$.

When $\hat{R}$ is larger than 1.2 it is better to run the chain for a larger number of iterations in order to improve the convergence of the algorithm.

Differently from the Gelman-Rubin, the Geweke Diagnostic test works with a single chain. It takes two non-overlapping parts of the chain of each parameter. The two parts, by default, are the first 0.1 and the last 0.5 of the Markov chain. The Geweke test compares the means of both parts, using a difference of means test to see if the two part of the chain care from the same distributions. The test statistic is a $N(0,1)$. If the absolute value of the statistic is less than 1.96 then there is no problem with the chain convergence.

3 Replication of Dudley 2006

The first objective of this project was to replicate the analysis conducted by *dudley2006bayesian*. The dataset used comprises insurance claim amounts exceeding 1.5 million over a period of five years from an automobile insurance portfolio. The data, originally presented in Rytgaard (1990), is shown in Table 1.

To model this dataset within a Bayesian framework, assumptions about the distributions of both the number of claims in year t (N_t) and the amount of the i-th claim in year t ($Y_{i,t}$) were necessary. Claims were assumed to occur randomly and independently at a constant rate over time, so N_t was modeled using a Poisson distribution.

[1] To manage risk exposure, insurers frequently employ reinsurance strategies, which help reduce their financial liability on large claims. Under such arrangements, if a claim amount y exceeds a predetermined threshold d (the retention), the insurer is responsible only for paying up to d, while any excess $y - d$ is covered by the reinsurer.

Table 1: Insurance Claim Amounts Exceeding 1.5 Million (Data from Rytgaard, 1990)

Year	Claim Amounts (in millions)				
1	2.495	2.120	2.095	1.700	1.650
2	1.985	1.810	1.625	–	–
3	3.215	2.105	1.765	1.715	–
4	–	–	–	–	–
5	19.180	1.915	1.790	1.755	–

The threshold of 1.5 million corresponds to the retention level of an excess-of-loss insurance policy[1].

A Pareto distribution was chosen for $Y_{i,t}$, as a heavy-tailed loss distribution was needed to account for the fact that individual claim amounts are positive and may include large outliers. That is,

$$N_t \sim \text{Poisson}(\theta), \quad 0 < \theta < \infty,$$
$$Y_{i,t} \sim \text{Pareto}(\alpha, \beta), \quad \alpha > 0, \quad 0 < \beta < y.$$

The Pareto(α, β) distribution with support $[\beta, \infty)$ was particularly suitable in this context, as we were modeling claim amounts exceeding a certain threshold.

In addition, the following assumptions were made:

- N_t are independently and identically distributed (i.i.d.) across t,
- $Y_{i,t}$ are i.i.d. across both i and t,
- N_t and $Y_{i,t}$ are independent for all i and t.

Under these assumptions, the aggregate claim amount in year t, denoted by

$$S_t = Y_{1,t} + Y_{2,t} + \cdots + Y_{N_t,t},$$

follows a compound Poisson distribution, since it represents the sum of independent Pareto-distributed random variables. [This is wrong?]

Next, prior distributions for the parameters α, β, and θ were specified. Due to limited prior knowledge about their true values—beyond the assumption that they are strictly positive—vague Gamma priors were chosen:

$$\alpha \sim \text{Gamma}(1, 0.0001), \quad \beta \sim \text{Gamma}(1, 0.0001), \quad \theta \sim \text{Gamma}(1, 0.0001),$$

with the constraint $0 < \beta < \min\{y_{i,t}\}$ to ensure validity of the Pareto distribution. Each of these Gamma priors has a variance of 10^8, implying minimal prior influence so that most of the information about the parameters is derived from the dataset. Additionally, the Gamma distribution is conjugate to both the Poisson and Pareto likelihoods, facilitating analytical tractability in Bayesian inference.

Finally, the posterior distributions were derived. First, the joint posterior distribution of (α, β) was obtained via Bayes' theorem[2]:

$$\begin{aligned}
\pi(\alpha, \beta \mid \mathbf{y}) &\propto \pi(\alpha) \cdot \pi(\beta) \cdot f(\mathbf{y} \mid \alpha, \beta) \\
&\propto 0.0001 \cdot \exp(-0.0001\alpha) \cdot 0.0001 \cdot \exp(-0.0001\beta) \cdot \prod_{i=1}^{n} \frac{\alpha \beta^{\alpha}}{y_i^{\alpha+1}} \\
&\propto \exp(-0.0001\alpha) \cdot \exp(-0.0001\beta) \cdot \alpha^n \cdot \beta^{n\alpha} \left(\prod_{i=1}^{n} y_i \right)^{-(\alpha+1)} \\
&\propto \alpha^n \cdot \exp(-0.0001\alpha) \cdot \left(\prod_{i=1}^{n} y_i \right)^{-\alpha} \cdot \beta^{n\alpha} \cdot \exp(-0.0001\beta) \\
&\propto \alpha^n \cdot \exp\left(- \left(0.0001 + \sum_{i=1}^{n} \ln(y_i) \right) \alpha \right) \cdot \beta^{n\alpha} \cdot \exp(-0.0001\beta)
\end{aligned}$$

As a result, the full conditional posterior distributions of α and β were as follows:

$$\pi(\alpha \mid \beta, \mathbf{y}) \propto \alpha^n \cdot \exp\left(- \left(0.0001 - n \ln(\beta) + \sum_{i=1}^{n} \ln(y_i) \right) \alpha \right)$$
$$\pi(\beta \mid \alpha, \mathbf{y}) \propto \beta^{n\alpha} \cdot \exp(-0.0001\beta)$$

[2]Here, assuming that α and β are independent, the joint prior $\pi(\alpha, \beta)$ was computed as $\pi(\alpha) \cdot \pi(\beta)$.

which implied that:

$$\alpha \mid \beta, \mathbf{y} \sim \text{Gamma}\left(n+1, \sum_{i=1}^{n} \ln(y_i) - n\ln(\beta) + 0.0001\right),$$
$$\beta \mid \alpha, \mathbf{y} \sim \text{Gamma}(n\alpha + 1, 0.0001)$$

Similarly, the posterior distribution of θ was obtained via Bayes' theorem:

$$\begin{aligned}
\pi(\theta \mid \mathbf{n}) &\propto \pi(\theta) \cdot f(\mathbf{n} \mid \theta) \\
&\propto \exp(-0.0001\theta) \cdot \prod_{t=1}^{T} (\theta^{n_t} \cdot \exp(-\theta)) \\
&\propto \exp(-0.0001\theta) \cdot \theta^{\sum_{t=1}^{T} n_t} \cdot \exp(-5\theta) \\
&\propto \exp(-5.0001\theta) \cdot \theta^{\sum_{t=1}^{T} n_t}
\end{aligned}$$

which implied that:

$$\theta \mid \mathbf{n} \sim \text{Gamma}\left(\sum_{t=1}^{T} n_t + 1, 5.0001\right)$$

Since all three posterior distributions were standard distributions, the Gibbs sampling method was employed to draw realizations from them. This was implemented using the `JAGS` program, which was called from within `R`. Three Markov chains were run in parallel. The initial values of α, β, and θ were chosen to be well-dispersed and are presented in Table 2.

Table 2: Initial Parameter Values

Chain	α	β	θ
1	1×10^{-5}	1×10^{-5}	1×10^{-5}
2	1×10^{5}	1	1×10^{5}
3	3.076	1.625	3.200

The burn-in period was set to 20,000 iterations. The statistics computed over the results of the subsequent 30,000 iterations are presented in Table 3. A comparison with the statistics reported by Dudley shows a close match, indicating that the model was properly specified and the Gibbs sampler was executed correctly.

Table 3: Posterior Statistics

	Mean	**Standard Deviation**	**95% Bayesian Credible Interval**
α	3.079	0.763	(1.771, 4.741)
β	1.592	0.035	(1.498, 1.624)
θ	3.396	0.821	(1.982, 5.192)
$E[Y]$	2.499	0.621	(2.024, 3.621)

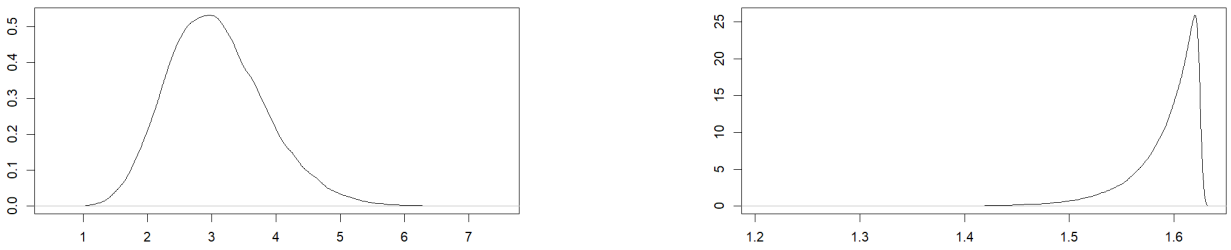
Note: $E[Y]$ was calculated for each simulated set of parameters α and β, and from these values, the mean, standard deviation, and 95% Bayesian credible interval were subsequently computed.

In addition, density plots were generated for each of the parameters and for $E[Y]$, as presented in Figure 1. The resulting densities for the parameters resemble Gamma distributions, with the density of β appropriately truncated at $\min\{y_{i,t}\} = 1.625$. The density plot for $E[Y]$ displays a right-skewed distribution that permits very large values, albeit with very low probability—consistent with expectations.

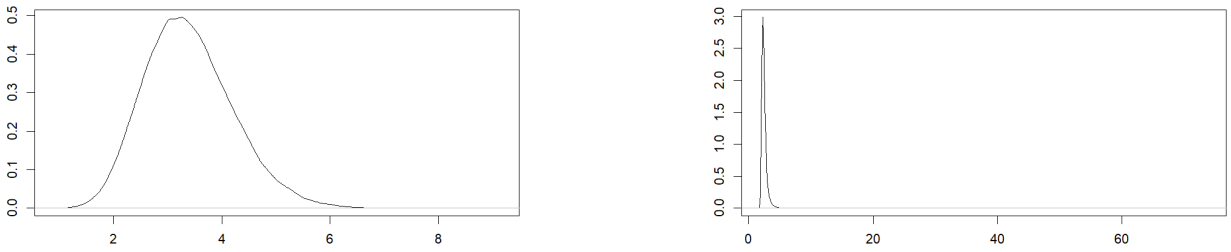
The posterior means of α and β were used as parameters of the Pareto distribution, and the corresponding cumulative distribution function (CDF) was plotted against the empirical cumulative data ($y_{i,t}$). Similarly, the posterior mean of θ was used as the parameter of the Poisson distribution, and its CDF was plotted against the empirical cumulative data (n_t).

The Pareto(3.079, 1.592) distribution provides a close fit to the empirical data. The Poisson(3.396) distribution also fits the observed frequencies quite well.

Throughout all computations, a burn-in period of 20,000 iterations was applied, with only samples from iterations 20,001 to 50,000 retained for analysis. This approach followed the assumptions made by Dudley. However, it is essential to verify that the chains have indeed converged to a stationary distribution after discarding the initial samples. The first method of assessment involves visual inspection. Figure 4 presents trace plots for all three parameters, showing the sampled values across iterations. These plots indicate good mixing, suggesting that the chains have converged.



(a) α (b) β



(c) θ (d) $E[y]$

Figure 1: Posterior Densities

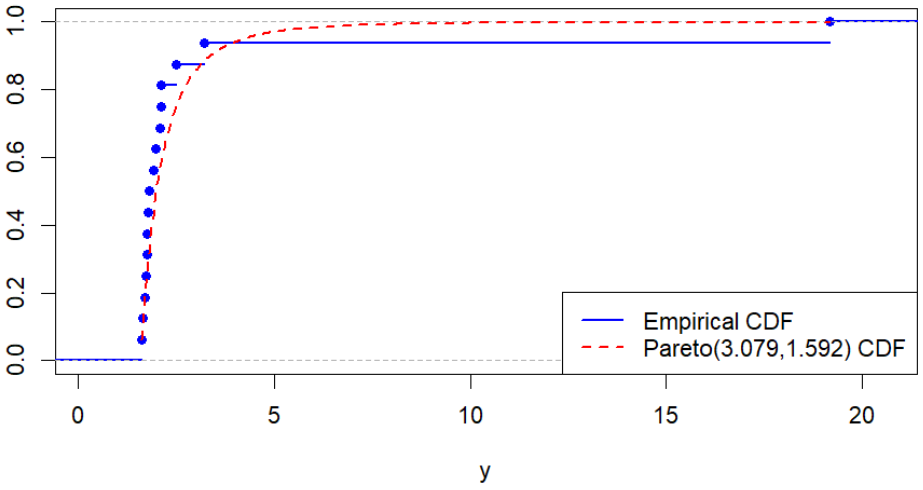


Figure 2: Empirical vs. Fitted Pareto CDF

Table 4: Potential Scale Reduction Factors (Gelman–Rubin Diagnostic)

Parameter	**Point Estimate**	**Upper C.I.**
α	1.00	1.00
β	1.00	1.00
θ	1.00	1.00
Multivariate PSRF	1.00	

The convergence of the chains was also assessed using the Gelman–Rubin diagnostic. The diagnostic was applied to the post-burn-in iterations (20,001–50,000), and the results are summarized in Table 4.
All univariate potential scale reduction factors (PSRFs) have point estimates and upper confidence bounds at

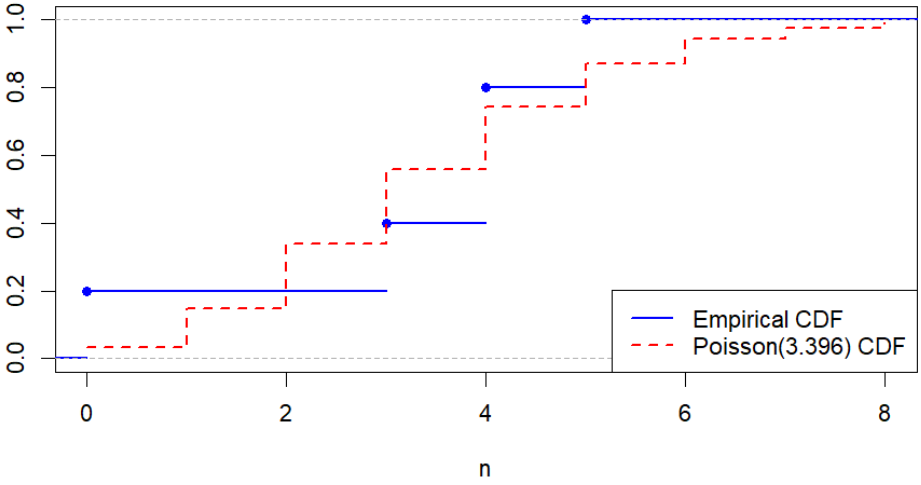
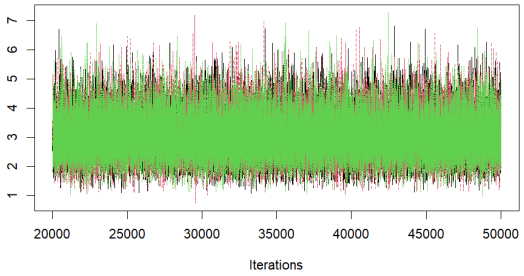
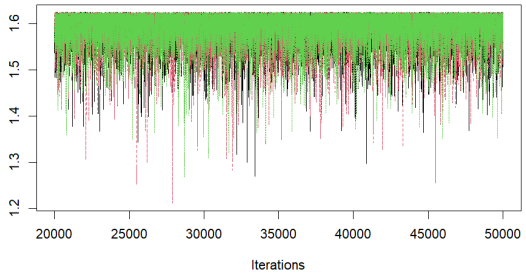


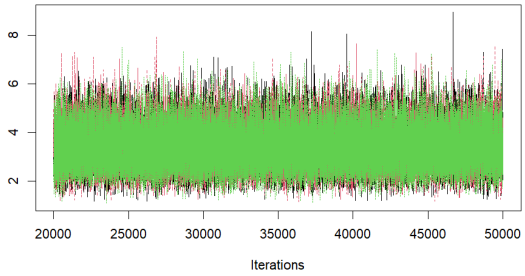
Figure 3: Empirical vs. Fitted Poisson CDF



(a) α



(b) β



(c) θ

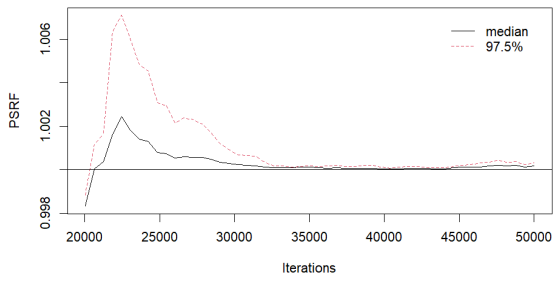
Figure 4: Trace Plots

1.00. The multivariate PSRF is also equal to 1.00. These values suggest that the Markov chains have likely converged, both for individual parameters and jointly.

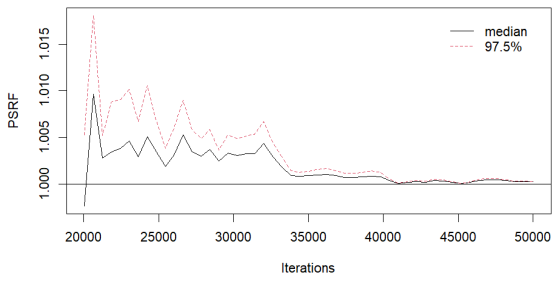
Figure 5 shows how the univariate PSRF point estimates evolve with increasing iterations. Throughout all iterations, all estimates remain below 1.1, which is commonly considered an acceptable threshold for convergence. This further confirms that the chains have likely reached a stable distribution.

In addition, autocorrelation plots were generated for all three parameters (see Figure 6), and values at lags 1 through 10 are reported in Table 5. Several high autocorrelations were observed, particularly for β, which motivated the use of a thinning interval of 10 iterations, as suggested by Dudley.

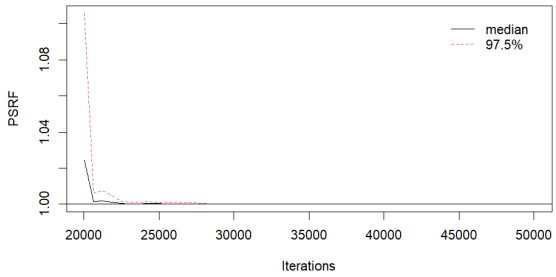
Consequently, the chains were rerun with this thinning. Figure 7 presents the corresponding trace plots, and Figure 8 shows the updated autocorrelation plots. The trace plots indicate that the chains have mixed well,



(a) α

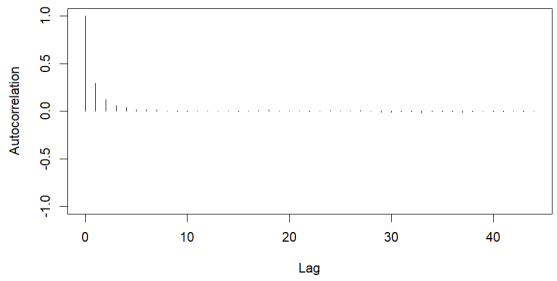


(b) β

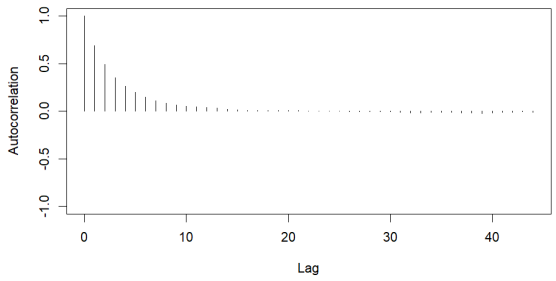


(c) θ

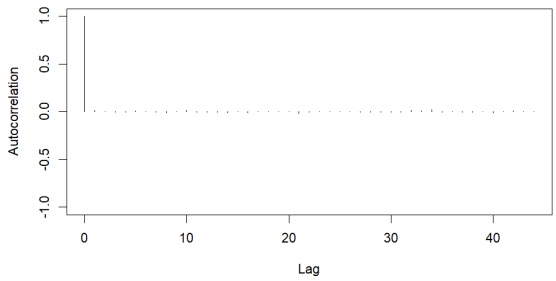
Figure 5: PSRF Values (Gelman–Rubin Diagnostic)



(a) α



(b) β



(c) θ

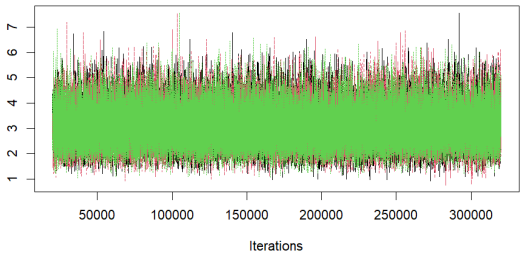
Figure 6: Autocorrelation Plots

and all autocorrelation values at lags 2, 3, and beyond have become negligible.

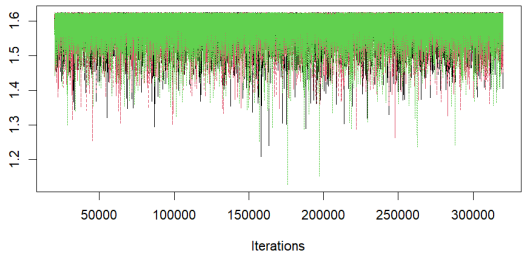
Since the ultimate goal of the analysis was to predict the values of S_t, the posterior predictive distribution was employed. For a variable z, this distribution is defined as:

Table 5: Autocorrelations at Lags 1–10

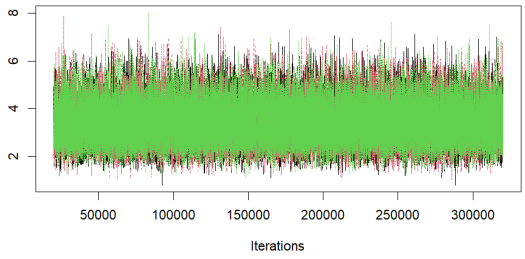
Lag	α	β	θ
1	0.297	0.705	0.003
2	0.119	0.509	0.005
3	0.061	0.373	−0.005
4	0.036	0.278	−0.004
5	0.023	0.210	0.001
6	0.016	0.161	−0.002
7	0.010	0.126	−0.004
8	0.009	0.102	−0.004
9	0.002	0.081	−0.001
10	0.007	0.065	0.004



(a) α



(b) β



(c) θ

Figure 7: Trace Plots After Thinning

$$\pi(z \mid \mathbf{y}) = \int_{\Theta} f(z \mid \phi)\, \pi(\phi \mid \mathbf{y})\, d\phi$$

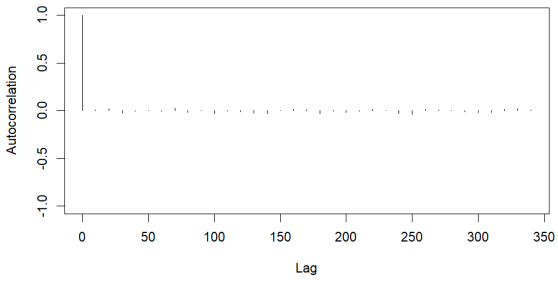
where $\phi = (\alpha, \beta, \theta)$, $\pi(\phi \mid \mathbf{y})$ is the posterior distribution of ϕ, and $f(z \mid \phi)$ is the likelihood of z given ϕ. This approach accounts for the uncertainty in ϕ by integrating over its possible values, weighted by their posterior probabilities.

Let $\mathbf{n}$ denote the observed data and N_f represent a future observation. Then the posterior predictive distribution for N_f is:

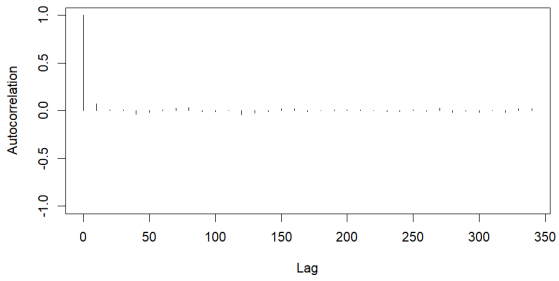
$$\begin{aligned} p(N_f = n \mid \mathbf{n}) &= \int_0^{\infty} p(N_f = n \mid \theta)\pi(\theta \mid \mathbf{n})\, d\theta \\ &= \mathbb{E}_{\theta|\mathbf{n}}\left[p(N_f = n \mid \theta)\right] \\ &= \mathbb{E}_{\theta|\mathbf{n}}\left[\frac{\theta^n e^{-\theta}}{n!}\right] \end{aligned}$$

Since the integral could not be solved analytically, it was approximated using samples from the posterior distribution obtained via MCMC. Specifically, the expectation was approximated as:

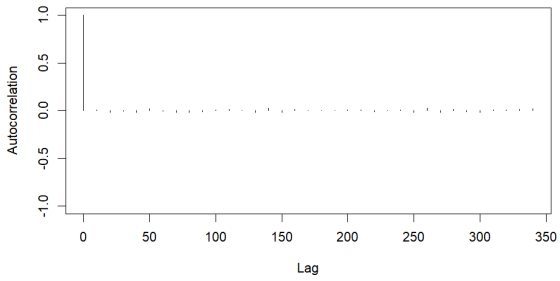
$$p(N_f = n \mid \mathbf{n}) \approx \frac{1}{m}\sum_{i=1}^{m} \frac{(\theta^{(i)})^n e^{-\theta^{(i)}}}{n!} \tag{1}$$



(a) α



(b) β



(c) θ

Figure 8: Autocorrelation Plots After Thinning

where $\theta^{(i)}$ is the i-th sample from the MCMC chain and m is the number of iterations after burn-in and thinning.

Similarly, let $\mathbf{y}$ denote the observed data, and let Y_f represent a future observation. The posterior predictive cumulative distribution function (CDF) of Y_f is given by:

$$\begin{aligned} p(Y_f \leq y \mid \mathbf{y}) &= \int_{\mathbf{u}} p(Y_f \leq y \mid \mathbf{u})\, \pi(\mathbf{u} \mid \mathbf{y})\, d\mathbf{u} \\ &= \mathbb{E}_{\mathbf{u}|\mathbf{y}} \left[p(Y_f \leq y \mid \alpha, \beta) \right] \end{aligned}$$

Again,

$$p(Y_f \leq y \mid \mathbf{y}) \approx \frac{1}{m} \sum_{i=1}^{m} \left(1 - \left(\frac{\beta^{(i)}}{y} \right)^{\alpha^{(i)}} \right) \tag{2}$$

where $\alpha^{(i)}$ and $\beta^{(i)}$ denote the i-th samples from the MCMC chain, and m is the number of post-burn-in, thinned iterations.

Table 6 contains the estimated probabilities of $N_f = n$ for $n = 0, \ldots, 14$.

These results are consistent with those obtained by Dudley. As Y_f is a continuous variable, the probability density function (PDF) was estimated rather than discrete probabilities. The estimation employed the inverse cumulative distribution function (CDF) method.

A total of 1000 values $U \sim \text{Uniform}(0, 1)$ were generated. For each value, the transformation

$$y^{(i)} = \frac{\beta^{(i)}}{(1 - U)^{1/\alpha^{(i)}}}$$

was applied. For each i, the mean of the resulting values was computed. The resulting values were then used to approximate the PDF of Y_f via kernel density estimation (KDE). The estimated predictive density is illustrated in Figure 9.

The inverse CDF method was also used to estimate the predictive distribution of S_f, representing a future observation of S_t. The procedure was as follows: 1,000 values were drawn from the posterior predictive distribution of N_f. For each uniformly drawn U, a Poisson sample was generated using each posterior value of $\theta^{(i)}$, and the average of these samples was computed and rounded to obtain N_f. Then, for each simulated value of N_f,

Table 6: Estimates of $p(N_f = n \mid \mathbf{n})$

n	Probability
0	0.0452
1	0.1279
2	0.1918
3	0.2023
4	0.1685
5	0.1178
6	0.0719
7	0.0393
8	0.0196
9	0.0091
10	0.0039
11	0.0016
12	0.0006
13	0.0002
14	0.0001

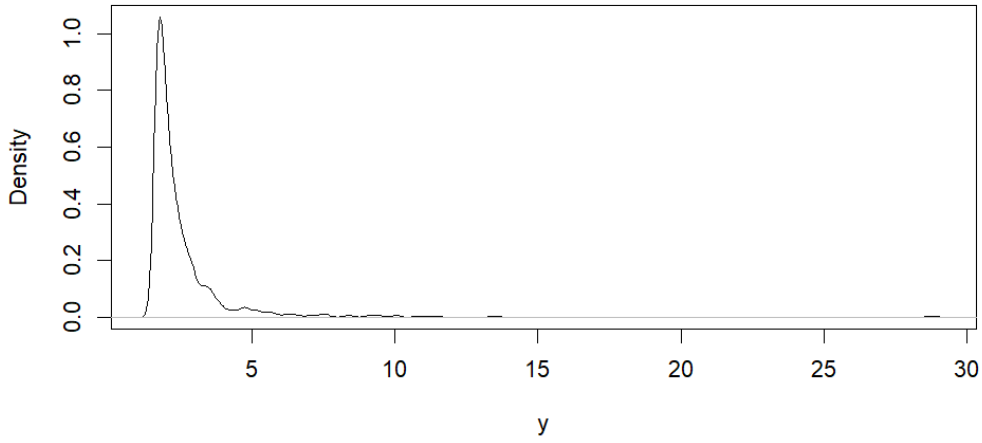


Figure 9: Estimated Predictive PDF of Y_f

that many samples were drawn from the predictive distribution of Y_f (as described earlier), and the resulting values were summed to obtain a draw of S_f.

Figure 10 presents the histogram of the resulting S_f samples, along with the estimated density and three fitted distributions using moment matching. As observed by Dudley, the Gamma distribution provides the best fit. The fitted Gamma distribution has parameters $\alpha = 2.435$ and $\beta = 0.292$.

Table 7 shows various percentiles of the simulated S_f values. The distribution exhibits a suitably long tail, which aligns with expectations for a heavy-tailed claim size distribution. This indicates that the simulation method used was effective in generating large Y values, thereby capturing the tail behavior of the predictive distribution of S more accurately. Proper representation of the tail is important, as most aggregate claims are moderate, but extreme values can occasionally occur.

Table 7: Percentiles of Simulated S_f Values

Percentile	**Value**
Median	7.630
90th Percentile	15.170
95th Percentile	18.103
99th Percentile	26.410
Maximum	35.851

References

Dudley, C. (2006). Bayesian Analysis of an Aggregate Claim Model Using Various Loss Distributions. Master's dissertation, Heriot-Watt University, School of Mathematical and Computer Sciences, Actuarial Mathematics

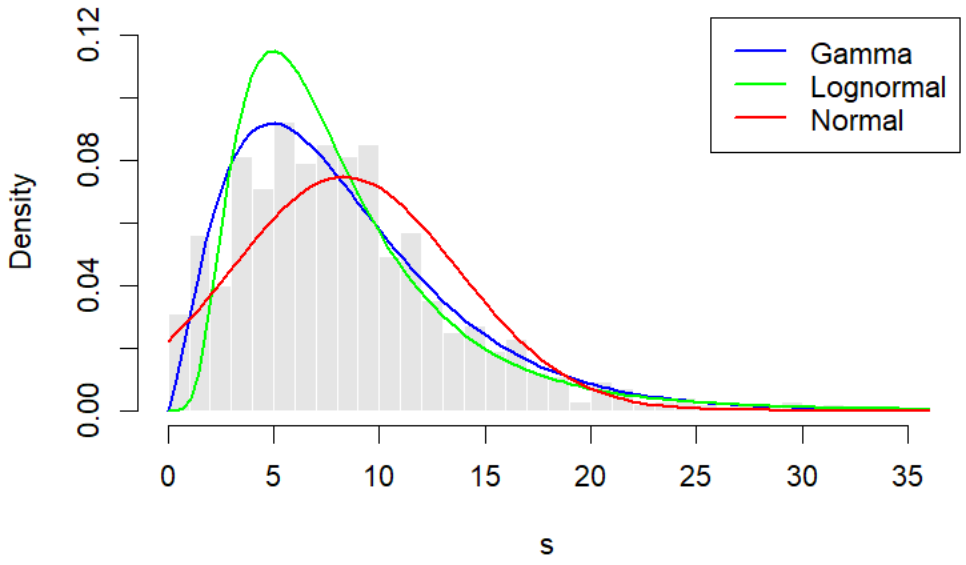


Figure 10: Histogram and Fitted Distributions for Predictive S_f

& Statistics.

Rytgaard, M. (1990). Estimation in the Pareto Distribution. *ASTIN Bulletin*, 20(2):201–216.