

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



We declare that this material, which we now submit for assessment, is entirely our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of our work. We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by us or any other person for assessment on this or any other course of study.

April 23, 2025

# Contents

1	Introduction	1
2		1

## 1 Introduction

### 2

The first objective of this project was to replicate the analysis conducted by Dudley. 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.

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<sup>1</sup>.

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. 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  $\text{Pareto}(\alpha, \beta)$  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  $\alpha$ ,  $\beta$ , and  $\theta$  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.

<sup>1</sup>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.

Finally, the posterior distributions were derived. First, the joint posterior distribution of  $(\alpha, \beta)$  was obtained via Bayes' theorem<sup>2</sup>:

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

As a result, the full conditional posterior distributions of  $\alpha$  and  $\beta$  were as follows:

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

which implied that:

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

Similarly, the posterior distribution of  $\theta$  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  $\alpha$ ,  $\beta$ , and  $\theta$  were chosen to be well-dispersed and are presented in Table 2.

Table 2: Initial Parameter Values

Chain	$\alpha$		$\beta$		$\theta$	
1	1	$\times 10^{-5}$	1	$\times 10^{-5}$	1	$\times 10^{-5}$
2	1	$\times 10^5$	1		1	$\times 10^5$
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.

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  $\beta$  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.

<sup>2</sup>Here, assuming that  $\alpha$  and  $\beta$  are independent, the joint prior  $\pi(\alpha, \beta)$  was computed as  $\pi(\alpha) \cdot \pi(\beta)$ .

Table 3: Posterior Statistics

	Mean	Standard Deviation	95% Bayesian Credible Interval
$\alpha$	3.079	0.771	(1.756, 4.761)
$\beta$	1.591	0.037	(1.495, 1.624)
$\theta$	3.400	0.826	(1.978, 5.193)
$E[Y]$	2.503	0.724	(2.022, 3.653)

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

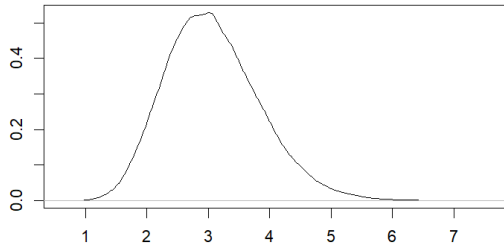
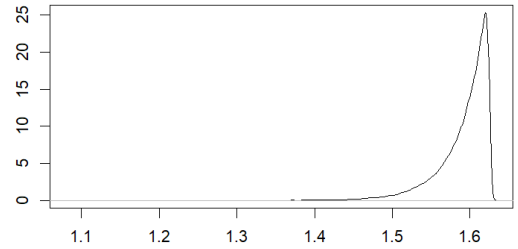
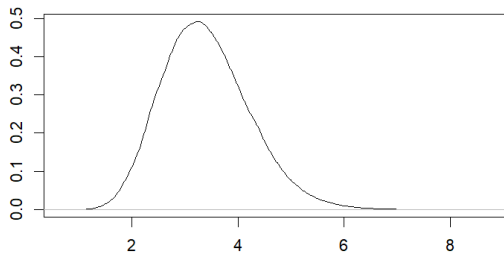
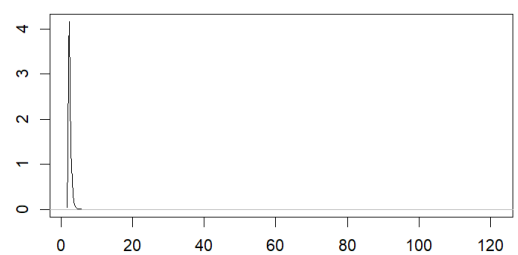
(a)  $\alpha$ (b)  $\beta$ (c)  $\theta$ (d)  $E[y]$ 

Figure 1: Posterior Densities

The posterior means of  $\alpha$  and  $\beta$  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  $\theta$  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, 1.591) distribution provides a close fit to the empirical data. The Poisson(3.4) 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 follows 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.

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.

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

Parameter	Point Estimate	Upper C.I.
$\alpha$	1.00	1.00
$\beta$	1.00	1.01
$\theta$	1.00	1.00
<b>Multivariate PSRF</b>	1.00	

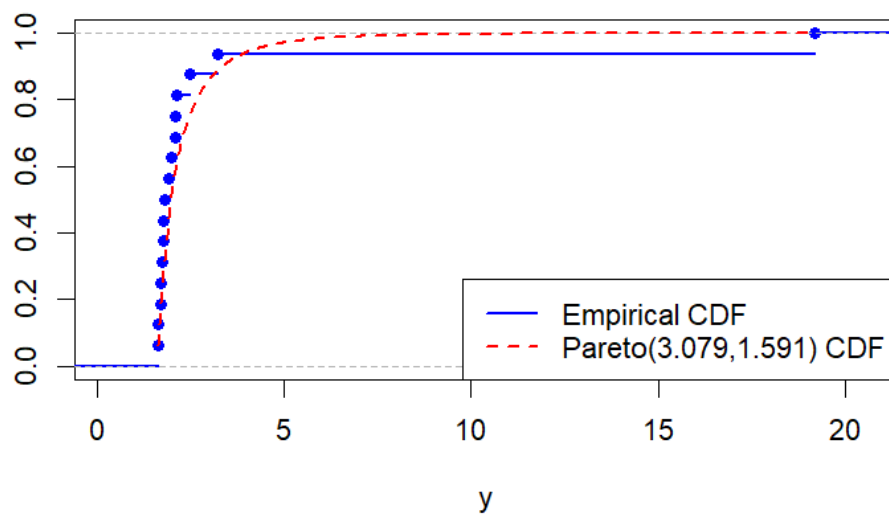


Figure 2: Empirical vs. Fitted Pareto CDF

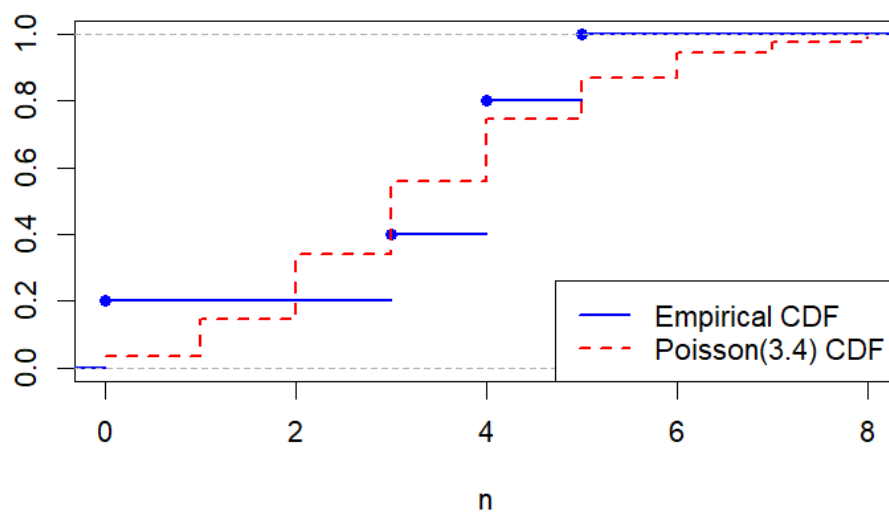
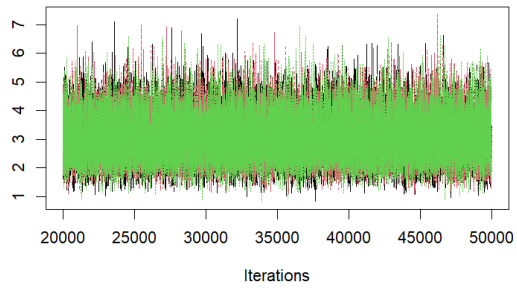
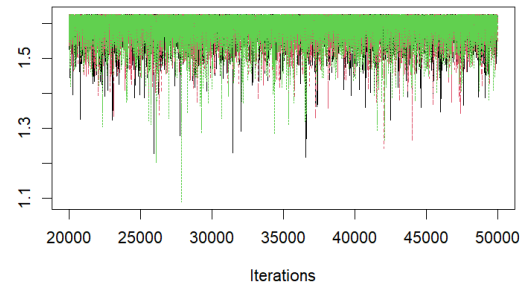


Figure 3: Empirical vs. Fitted Poisson CDF

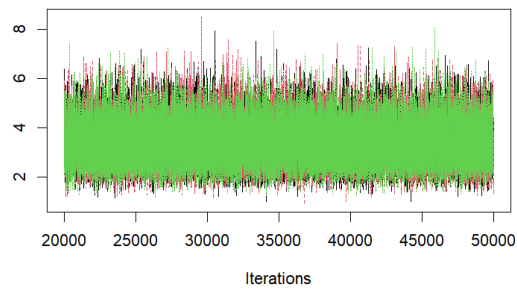
The results indicate that the chains have likely converged for all parameters, both individually and jointly.



(a) Trace plot for  $\alpha$



(b) Trace plot for  $\beta$



(c) Trace plot for  $\theta$

Figure 4: Trace Plots of Model Parameters