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 23, 2025

Contents

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[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. 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.

[1] To manage risk exposure, insurers frequently employ reinsurance strategies, which help limit 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. This approach helps insurers mitigate the impact of high-variance claims.