Simulation Methods Project : Group 10

LORENZ Lucien, LORENZ Arno, MONETA Marc

May 2023

1 Executive Summary

The analysis of the policy portfolio reveals that the collective risk model accurately captures the behaviour of claims, with a Poisson distribution governing claim occurrences and a symmetric Gamma distribution characterising claim severity. On average, each policyholder is expected to experience one claim, and the claim amounts are centred around the mean value. Notably, in 95% of instances where claims occur, the amounts are anticipated to fall within the range of CHF 614.94 and CHF 784.46.

Furthermore, our assessment indicates that the premiums charged adequately safeguard the solvency capital, considering a ruin probability of 0.5% over a one-year period. Precisely, the total premiums collected surpass the required amount by 7.067%. This surplus serves as an additional buffer, enhancing the company's financial resilience and ability to cover potential losses.

2 Data Exploration

The dataset encompasses comprehensive information on claim amounts, claim severity, and various policy-related attributes. Upon examining the shape of the claim severity distribution, we observed characteristics indicative of a light-tailed distribution. Consequently, we implemented two data restrictions to refine the dataset :

- Claims exceeding 5000 were identified as potential misinputs. As this only affected a single claim, it was deemed negligible for model estimation purposes.
- Negative claims are also considered erroneous.

Both of those restrictions are replaced by non-values (NaN) in our dataset instead.
Following the initial cleanup process, we opted to aggregate all claims into a unified vector. This decision was supported by the clear indication that the portfolio predominantly comprises a single type of claim, suggesting the assumption of a singular distribution governing all claims.

With the data organised accordingly, we proceeded to visualise the claim frequency and severity distributions. The observed shape of the frequency distribution exhibited

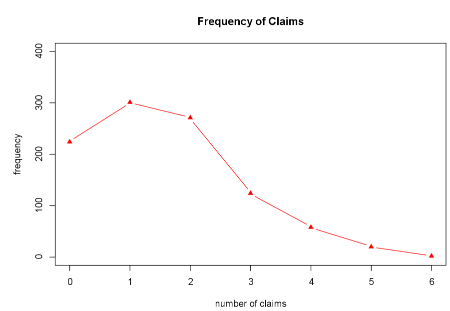

Figure 1: Plot of claim amount

characteristics consistent with various discrete distributions, with the exception of the geometric and hypergeometric distributions.

As for claim severity, its distribution displayed a nearly symmetrical shape, as evidenced by a computed skewness value of 0.093. This finding allows us to exclude highly skewed distributions from consideration.
Moreover, despite the presence of a few outliers among the 1559 claims, the severity values remained within the quartiles, indicating that the data largely adheres to its expected range. This observation supports the inclusion of symmetric distributions, specifically the Gamma and Normal distributions, for further consideration.

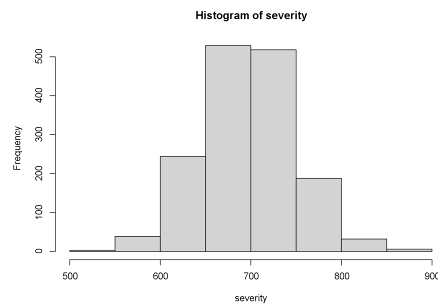

Figure 2: Histogram of claim severity

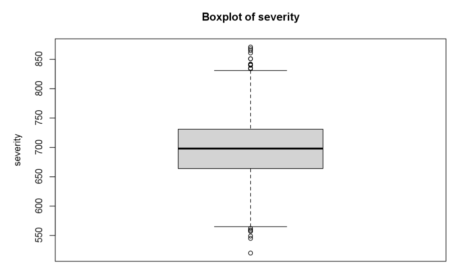

Figure 3: Boxplot of claim severity

3 Model Selection and Validation

Claim Amount

When examining the nature ans shape of claims, the data suggests a Poisson distribution for claim occurrences N. However, we further compare it with the negative binomial distribution to evaluate the best fit.

$$\begin{aligned} N &\sim Poi(\lambda) \\ \mathbb{E}(N) &= \lambda \\ \mathrm{Var}(N) &= \lambda \end{aligned}$$

To estimate the expectation and variance using Monte Carlo methods, we substitute the sample mean and variance as estimators.
It is worth noting that the issue with this model is that it restricts the variance to be

”close enough” to the mean. When computed, we find :

$$\frac{1}{n}\sum N_i = 1.559$$
$$\frac{1}{n-1}\sum (N_i - \bar{N})^2 = 1.572091$$

They are not perfectly aligned, but we judge them close enough to use this model. This gives us then the estimator $\hat{\lambda} = 1.559$.
We can determine if the sample comes from a sample distribution using Pearson's χ^2 test of independence, with following hypothesis :

$$H_0 = \text{Sample doesn't come from a Poisson}(\lambda)$$
$$H_1 = \text{Samples comes from a Poisson}$$

the test provides a p-value of 0.227, indicating that we cannot reject the null hypothesis. We can then plot the theoretical density on our empirical distribution

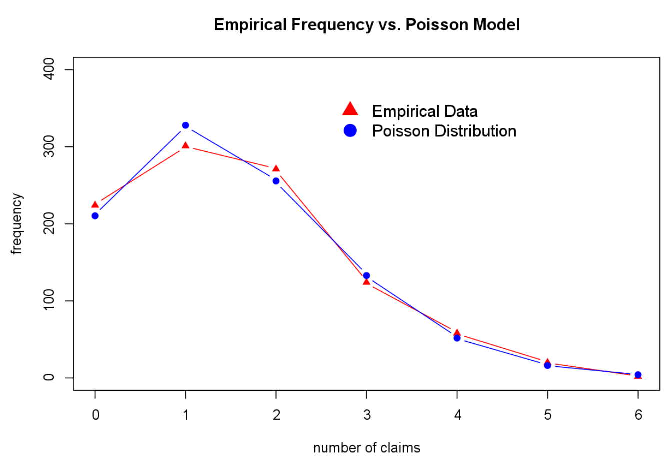

Figure 4: Theoretical Poisson against empirical

Attempting to fit a negative binomial distribution proves more challenging. We know that, in this case :

$$N_i \sim Nbin(r, p)$$
$$\mathbb{E}(N_i) = \frac{r(1-p)}{p}$$
$$\text{Var}(N_i) = \frac{r(1-p)}{p^2}$$

Based on the available data, we estimate the parameters $\hat{r}$ to be 185.6592 and $\hat{p}$ to be 0.9918. The Pearson's χ^2 test of independence yields a p-value of 0.227, indicating that we do not reject the null hypothesis of independence.

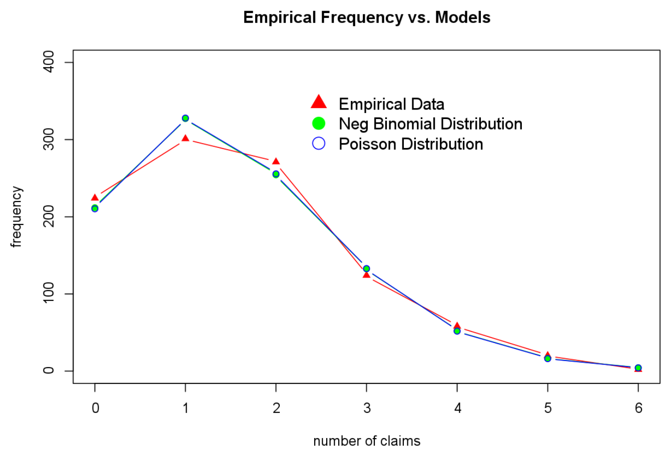

Figure 5: Empirical vs. Models

Comparing the two models, we find very similar results, which is not surprising : Both models are viable options, but the decision was made to choose the Poisson distribution over the negative binomial distribution. This choice is driven by a careful consideration of the trade-off between simplicity and accuracy. Furthermore in this specific case the assumption of equidispersion holds reasonably well.

By assuming equidispersion, we can effectively capture the variability in claim occurrences over a one-year time frame using the Poisson model. This assumption simplifies the modelling process, as it requires only a single parameter estimation (the mean). Furthermore, it allows for easier interpretation of the results, as the Poisson distribution provides a clear and intuitive representation of the probability distribution for the number of claims within the specified time period, given the average claim rate.

The choice of the Poisson model also takes into account computational efficiency. Estimating parameters and conducting statistical inference with the Poisson distribution is generally faster and more computationally efficient compared to the negative binomial model. This efficiency becomes particularly advantageous when dealing with large datasets or complex modelling tasks.

Considering the balance between simplicity, accuracy, and computational efficiency, the Poisson distribution is deemed suitable for modelling claim occurrences in this context. It captures the essence of the data adequately, given the sample mean and variance, and provides a practical and efficient framework for analysing claim frequencies over a one-year period.

Claim Severity

The shape of claim severity suggests a somewhat symmetrical distribution, eliminating the exponential, uniform and Pareto distributions. The Gamma distribution is our initial guess, we will first compute its parameters.

$$\begin{aligned} Y_i &\sim \Gamma(\alpha, \beta) \\ \mathbb{E}(Y_i) &= \frac{\alpha}{\beta} \\ \mathrm{Var}(Y_i) &= \frac{\alpha}{\beta^2} \end{aligned}$$

Giving us estimated parameters $\hat{\alpha} = 183.339$ and $\hat{\beta} = 0.2628$.
We can check if the sample truly comes from a Gamma distribution, with hypothesis

$$\begin{aligned} H_0 &= \text{Sample doesn't come from a } \Gamma(\alpha, \beta) \\ H_1 &= \text{Sample comes from a Gamma} \end{aligned}$$

Using a Kolmogorov-Smirnov test of distribution, the p-value returns 0.6054, not rejecting the null-hypothesis. Comparing the theoretical density to empirical frequency :

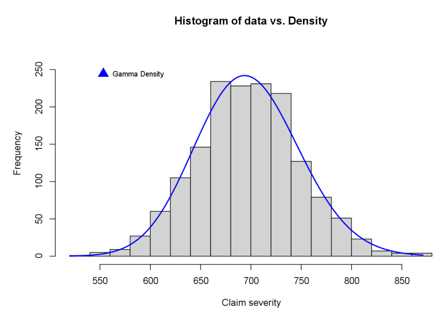

Figure 6: Fitted Gamma against Severity

We can also check if the quantiles match up :
Our second choice was to use a Normal approximation. In this case :

$$\begin{aligned} Y_i &\sim \mathcal{N}(\mu, \sigma^2) \\ \mathbb{E}(Y_i) &= \mu \\ \mathrm{Var}(Y_i) &= \sigma^2 \end{aligned}$$

the estimated parameters are $\hat{\mu} = 697.569$, $\hat{\sigma}^2 = 2654.188$. In the same fashion, the KS-test returns a p-value of 0.6443, not rejecting the null-hypothesis that the data doesn't come from a gamma. Plotting the quantile plot :
We see that the quantiles start to deviate on the extremes.

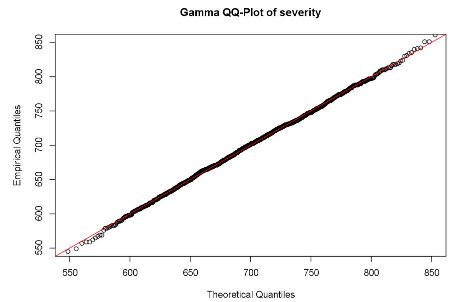

Figure 7: Gamma QQ-plot

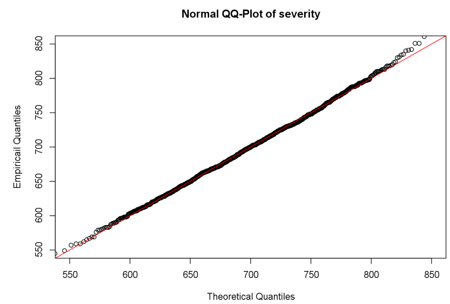

Figure 8: Normal QQ-plot

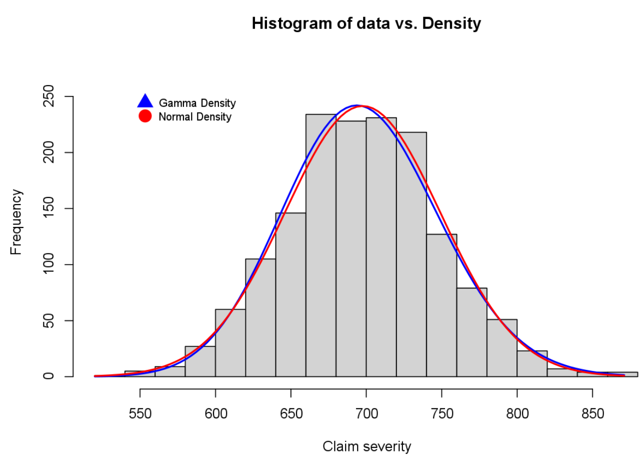

Figure 9: Models against Empirical Severity

Plotting both distributions gives us Figure9. The only issue with the Normal distribution is, even though negative values are very unlikely (given the 99% confidence interval is nowhere close to 0), it can support negative values.

The choice to employ the gamma distribution for modelling claim severity is justified by two compelling factors: the presence of positive skewness (0.093) in the data and the requirement of non-negative values.

The gamma distribution excels at accommodating such positively skewed distributions, rendering it a fitting choice to accurately capture the underlying characteristics of the data. Its inherent ability to flexibly represent distributions with varying degrees of skewness ensures a faithful portrayal of the claim severity distribution.

Additionally, the gamma distribution satisfies the fundamental constraint of non-negative claim amounts, which is essential in the context of modelling claim severity. By design, the gamma distribution guarantees that modelled claim severity values remain within a realistic range, effectively excluding the possibility of negative values. This alignment with the inherent nature of claim severity data enhances the credibility and reliability of the modelling results.

Variance Reduction

The main problem with the crude Monte-Carlo estimator (the sample mean), is that it converges in $O(\sigma/\sqrt{n})$. We implemented the antithetic estimator as a proof of concept that we can reduce the variance of the mean.
The antithetic estimator is useful as it allows to only generate half the values needed. We divided the dataset in two, and used the first half as our estimator $f(U)$, and the second half as $f(1-U)$, because they come from the same distribution. Implementing this technique for claims frequency and severity shows us that the standard deviation is smaller than the Monte-Carlo estimator.
Note : the mean for the antithetic estimator is slightly different than the one in the Monte-Carlo estimation. This is due to the fact that we have an odd number of non-NaN severities ; one observation is not taken into account.

	Frequency		Severity	
	Monte Carlo	Antithetic	Monte Carlo	Antithetic
Mean	1.559	1.559	697.569	697.5802
Standard Deviation	0.039	0.027	1.305	0.927

Table 1: Comparison between Monte Carlo and Antithetic Estimator

4 Application

To determine the Value at Risk, we can compute it in two ways : using the collective risk model, or simulating it. We will do both and check if they are similar.

Simulation

Using a Poisson and Gamma distribution allows us to make the simulation simpler. These distributions exhibit the following properties :

$$\begin{aligned} N_1, N_2 &\sim Poi(\lambda) \\ Y_1, Y_2 &\sim \Gamma(\alpha, \beta) \end{aligned}$$

$$\begin{aligned} N_1 + N_2 &\sim Poi(2\lambda) \\ Y_1 + Y_2 &\sim \Gamma(2\alpha, \beta) \end{aligned}$$

With a dataset that is 1000 individuals long, we can therefore generate synthetic portfolios in two steps :

- Generate the total number of claims for each portfolio
- Generate a sum of claims for each portfolio, with n_i claims.

This gives us a distribution of sum of claims, as such :

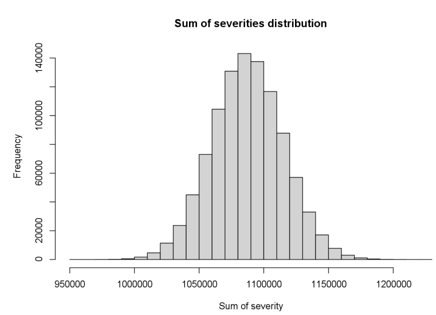

Figure 10: Histogram of Sum of Severities

Now, we can take the empirical quantile

$$\begin{aligned} F_S^{-1}(99.5\%) &\approx S_{\lfloor 99.5\% \cdot n \rfloor : n} \\ &\approx 1'159'189 \end{aligned}$$

such that n is the total sum of severities generated.
Note : we could have implemented the antithetic variance reduction method to generate those sums of severities (and claims). The issue with using a Gamma distribution is in generating a random number from that distribution. For that, we need access to the

cumulant generating function, which includes the lower incomplete gamma function $\gamma(\alpha, \beta x)$. This function is implemented in packages that are not up to date. To not complexify the problem, we decided not to approximate the integral ourselves. However, the standard deviation of the mean of total severities is equal to 0.699. Considering the mean of 1′087′478, implementing variance reduction is negligible.

Model approximation

Modelling our portfolio as a collective risk models gives us the specification :

$$\begin{aligned} N_t &\sim Poi(n\lambda) \\ Y_i &\sim \Gamma(n\alpha, \beta) \\ S &= \sum_{1}^{N_t} Y_i \end{aligned}$$

The model therefore exhibits the following moments :

$$\begin{aligned} \mathbb{E}(S) &= \mathbb{E}(N_t)\mathbb{E}(Y) \\ \mathrm{Var}(S) &= \mathbb{E}(N_t)\mathrm{Var}(Y) + \mathbb{E}(Y)^2\mathrm{Var}(N_t) \end{aligned}$$

Due to the high count of individuals, we can use a normal approximation. The Value at risk is then defined as

$$\begin{aligned} VaR_{0.5\%}(S) &= F_S^{-1}(1 - 0.5\%) \\ &\approx \Phi^{-1}(99.5\%) \\ &\approx 1'158'946 \end{aligned}$$

The two values (simulated and from the approximation) are close. We will be using the one from the model because it is the lowest one.
The total amount of premiums amounts to 1′240′850, which covers the solvency capital. To be precise, we could reduce the average premium by 7.067% and still cover the capital requirements.

5 Model Explanation

We believe the claims to follow a Poisson distribution, meaning that a high number of claims per individual is very unlikely, but expect each individual to have at least one claim. This is how the claims are expected to behave :

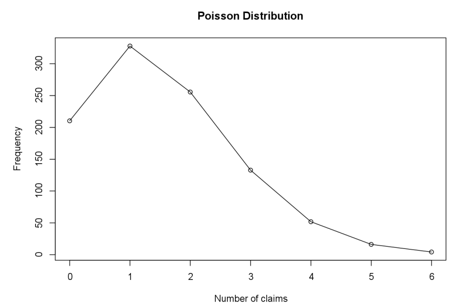

Figure 11: Poisson Distribution

You can thus expect a lot of individuals to have between 0 and 2 claims.

Concerning the severity of the claims, the Gamma distribution behaves very nicely : it is light-tailed, so you can expect claims to not deviate heavily, and it is quite symmetrical, so you can plan around the mean of the claims. In 95% of cases, you can expect claims to be between CHF 614.94 and CHF 784.46, with an average of CHF 697.57 The premiums are priced enough to cover the solvency capital, there is thus

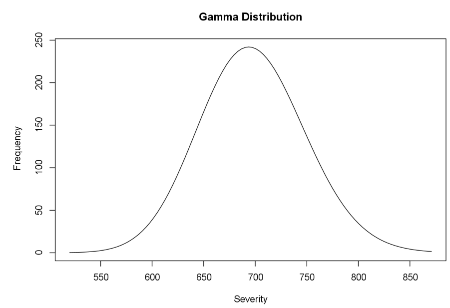

Figure 12: Gamma distribution

nothing to change on that front. The necessary capital is CHF 1′158′946, which is covered by the sum of premiums amounting to CHF 1′240′850.

The surplus buffer of CHF 81′904, representing the excess of premiums over the necessary capital, raises the possibility of reducing premiums for policyholders. However, this decision requires a comprehensive evaluation that considers multiple factors to maintain financial stability and meet profitability goals.

One important consideration is the need to ensure smooth premium adjustments over time. By maintaining a consistent premium structure, policyholders can have confidence in the stability and predictability of their insurance coverage. This promotes trust and long-term relationships with customers.

While actuarial modelling provides valuable insights into the financial standing of the company, it is essential to look beyond these numbers. Administrative costs, operational expenses, and other factors should be taken into account to assess the overall financial health of the organisation. This broader perspective ensures that premium reductions are sustainable and do not compromise the company's ability to meet its financial obligations.

Moreover, profitability goals are crucial for the company's success and the satisfaction of its shareholders. Any decision to lower premiums must align with these objectives and contribute to long-term profitability. It is important to strike a balance between providing affordable premiums to policyholders and generating sufficient revenue to cover costs and deliver a return on investment.