REPORT I

Insurance Models for Industry

Authors
Martyna Hirak
Klaudia Jaworek

Introduction

This report solves four problems related to insurance models using computer simulations.

In the first task, we analyze the insurer's surplus model and we compare numerical and analytical expected value of ruin time.

In the second task, we compare different types of insurance (integral and reducing deductibles, proportional insurance) using the Pareto distribution.

The third task investigates a mixed Poisson distribution where the parameter Λ follows a gamma distribution, showing that it is a negative binomial distribution.

The final task examines the distribution of $S = \sum_{k=1}^{N} X_k$, where X follows a uniform or logarithmic distribution and N follows either a Poisson or geometric distribution, comparing the results with Panjer's recursion.

Task 1.

For this task, we have selected the second problem from the property insurance section of the actuarial exams held on June 12th, 2017, with the description outlined below.

Task description

Consider a model of an insurer's surplus with discrete time: $U_n = u + X_1 + X_2 + ... + X_n$, where

- u is the initial surplus (non-negative),
- $X_1, X_2, X_3, ...$ are i.i.d. and represent the differences between premiums received and claims paid in successive years,
- the distribution of the random variable X_1 is: $\mathbb{P}(X_1 = 3) = p_3, \mathbb{P}(X_1 = 2) = p_2, \mathbb{P}(X_1 = 1) = p_1, \mathbb{P}(X_1 = 0) = p_0, \mathbb{P}(X_1 = -1) = 1 - p_0 - p_1 - p_2 - p_3$.

Let $N = \min\{n : U_n < 0\}$ denote the time of ruin. Assume that the parameters of the process are: $p_0 = p_1 = p_2 = p_3 = 1/12$, and $u = 9/2$. In these conditions, ruin is certain, so $\mathbb{P}(N < \infty) = 1$, and therefore the expected time to ruin $\mathbb{E}(N)$ is well-defined and finite.

What is the value of $\mathbb{E}(N)$?

To address this task, 10^6 paths of the process U_n were generated, stopping at the first time the process fell below zero ($U_n < 0$). The expected value $\mathbb{E}(N)$ was then estimated using the sample mean of the obtained ruin times, resulting in $\mathbb{E}(N) \approx 30.21036$.

To ensure that our result obtained through simulation is correct, we can compare it with the analytical result. Notice that the increments of the surplus are integers, and the negative increment can only be -1, so the time of ruin $N = \min\{n : u + X_1 + X_2 + ... + X_n < 0\}$ for $u = 4.5$ is equivalent to the moment $N = \min\{n : X_1 + X_2 + ... + X_n = -5\}$. For convenience, let us denote $S_n = X_1 + X_2 + ... + X_n$ and $N_k = \min\{n : S_n = k\}$. Given that the only negative increment is -1, and considering that the process S_n has stationary and independent increments with $S_0 = 0$ almost surely, we can conclude that $\mathbb{E}(N_{-k}) = k \cdot \mathbb{E}(N_{-1})$ for any $k \in \mathbb{N}$. The expected time to reach -1 is calculated using the law of total expectation, which conditions on the possible outcomes of the first step. Each outcome influences the overall expected time based on the subsequent state of the process, specifically

$$\begin{aligned}\mathbb{E}(N_{-1}) &= 1 + p_0 \cdot \mathbb{E}(N_{-1}) + p_1 \cdot \mathbb{E}(N_{-2}) + p_2 \cdot \mathbb{E}(N_{-3}) + p_3 \cdot \mathbb{E}(N_{-4}) = \qquad (1)\\ &= 1 + (p_0 + 2p_1 + 3p_2 + 4p_3)\mathbb{E}(N_{-1}). \qquad (2)\end{aligned}$$

For the probability values specified in the task we obtain $N_{-1} = 6$, and then $N = 30$.

The close agreement between the value obtained through simulation and the analytical solution confirms the accuracy of the numerical approximation.

Task 2.

Task description

Consider a loss X with a Pareto distribution characterized by parameters λ and α, having the cumulative distribution function

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha}, \quad \alpha > 1, \lambda > 0, x > 0.$$

For this distribution, illustrate the following on graphs:

(a) Comparison of integral and reducing deductibles (depending on the cut-off level).

(b) Proportional insurance (depending on the proportion).

(c) Insurance with a reducing deductible modifying the loss as follows:

$$h_{d_1,d_2}(x) = \begin{cases} 0, & \text{for } x \leq d_1, \\ \frac{d_2(x-d_1)}{d_2-d_1}, & \text{for } d_1 \leq x \leq d_2, \\ 0, & \text{for } x \geq d_2 \end{cases}$$

(depending on d_1 and d_2).

Analyze three different parameter sets (α, λ).

A deductible is a mechanism used in insurance to adjust the amount of loss that is compensated, typically to reduce the insurer's liability or align incentives. In this task, the following types of deductibles are considered

- integral deductible: $h(x, d) = \begin{cases} 0, & \text{for } x < d, \\ x, & \text{for } x \geq d, \end{cases}$

- reducing deductible: $h(x, d) = \begin{cases} 0, & \text{for } x < d, \\ x - d, & \text{for } x \geq d, \end{cases}$

- proportional deductible: $h(x,d) = (1-d)x, \quad d \in [0,1]$,
- reducing deductible modifying the loss: $h(x, d_1, d_2) = \begin{cases} 0, & \text{for } x \leq d_1, \\ \frac{d_2(x-d_1)}{d_2-d_1}, & \text{for } d_1 \leq x \leq d_2, \\ 0, & \text{for } x \geq d_2. \end{cases}$

Each type of deductible adjusts the loss compensation in distinct ways, balancing fairness, risk-sharing, and cost control. The choice of deductible type and parameters (d, d_1, d_2) depends on the specific goals of the insurance policy, such as incentivizing risk management or ensuring equitable compensation. The graphs in Figure 1 visually demonstrate how different deductibles modify the relationship between the original loss x and the compensated amount $h(x)$.

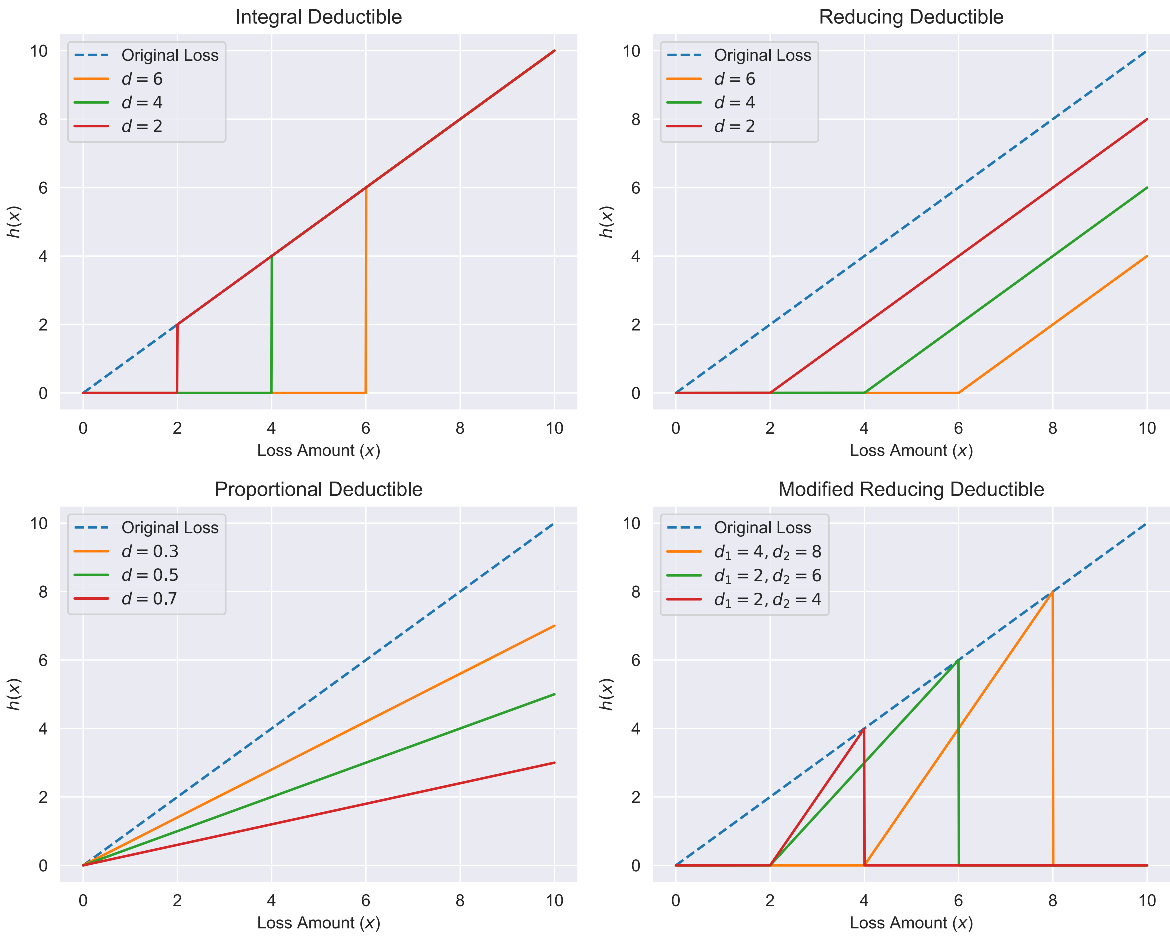

Figure 1: Comparison of different types of deductibles with various parameters.

The expected value of the deductible-modified loss corresponds to the net premium, which represents the insurer's expected cost of covering claims. When the loss variable X follows a Pareto distribution with α and λ parameters, the net premiums are as follows:

- integral deductible: $\mathbb{E}(h(X,d)) = \mathbb{E}(X) - \mathbb{E}(\min\{X,d\}) + d(1-F(d)) = \left(\frac{\lambda}{\lambda+d}\right)^{\alpha}\left(\frac{\lambda+d\alpha}{\alpha-1}\right)$,
- reducing deductible: $\mathbb{E}(h(X,d)) = \mathbb{E}(X) - \mathbb{E}(\min\{X,d\}) = \left(\frac{\lambda}{\lambda+d}\right)^{\alpha}\left(\frac{\lambda+d}{\alpha-1}\right)$,
- proportional deductible: $\mathbb{E}(h(X,d)) = (1-d)\mathbb{E}(X) = (1-d)\frac{\lambda}{\alpha-1}$,

- reducing deductible modifying the loss: $\mathbb{E}(h(X, d_1, d_2)) = \frac{d_2}{d_2 - d_1}(L(d_2) - L(d_1)) - d_2(1 - F(d))$, where $L(d) = \mathbb{E}(\min\{X, d\}) = \lambda^\alpha(\frac{\lambda^{1-\alpha} - (\lambda+d)^{1-\alpha}}{\alpha - 1})$.

Figures 2, 3, 4 and 5 show the comparison between exact values of net premium and numerical approximation based on 10^4 simulated random variables for varouis types of deductibles. The plots provide insights into how the deductible structure and the distribution parameters affect the premiums and the overall behavior of the insurance model.

For the same parameters, the net premium associated with the integral deductible (Figure 2) is always greater than or equal to that of the reducing deductible (Figure 3). This difference is mathematically expressed as $d(1 - F(d))$, where F is the cumulative distribution function of the underlying Pareto distribution. The parameters of the Pareto distribution have a significant impact on the variance of the claims. As these parameters change, they impact the accuracy of the resulting premiums. A higher variance in the claims distribution typically leads to less accurate premium estimates. An important observation is the behavior of the net premium when d is set to zero. In both the integral and reducing deductible cases, that means that the insurer covers the entire loss. As we see on the plots, the net premium behaves as expected: it is a decreasing function of d.

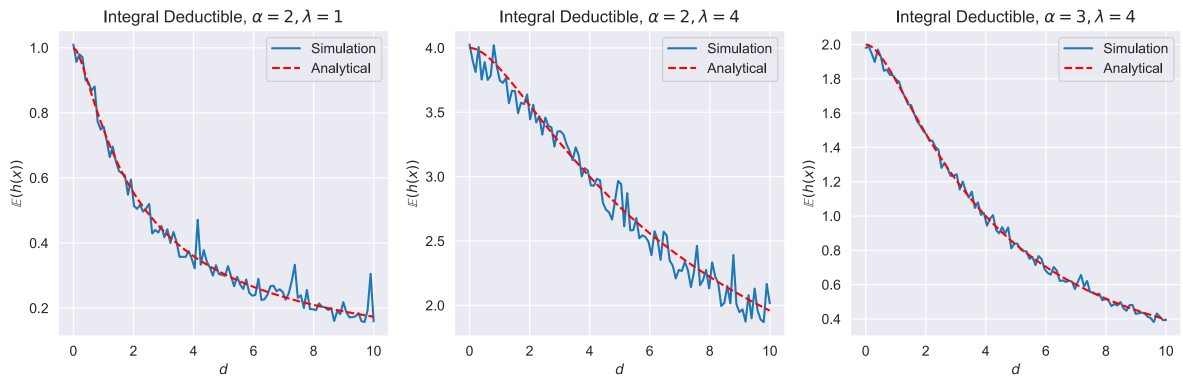

Figure 2: Comparison of net premium dependency on the integral deductible parameter d across various parameters of Pareto distribution of loss. Comparison of estimated results based on 10^4 random variables against exact values.

Similar conclusions can be drawn when analyzing the case of a proportional deductible. When the deductible $d = 0$, the insurer bears the full cost of the loss. The net premium is a decreasing function of d, reflecting the reduced liability for the insurer as the deductible increases. For proportional insurance, the relationship between the net premium and d simplifies to a linear function. Furthermore, we observe a close alignment between the numerical and analytical solutions.

In the case of insurance with a reducing deductible, where the loss is adjusted according to the given formula, the net premium is observed to be a decreasing function of d_1 and an increasing function of d_2. This indicates that the cost to the insurer increases as the difference $d_2 - d_1$ becomes larger. Once again, we observe a close alignment between the numerical approximation and the exact solution, further validating the accuracy and reliability of the numerical approach in this context.

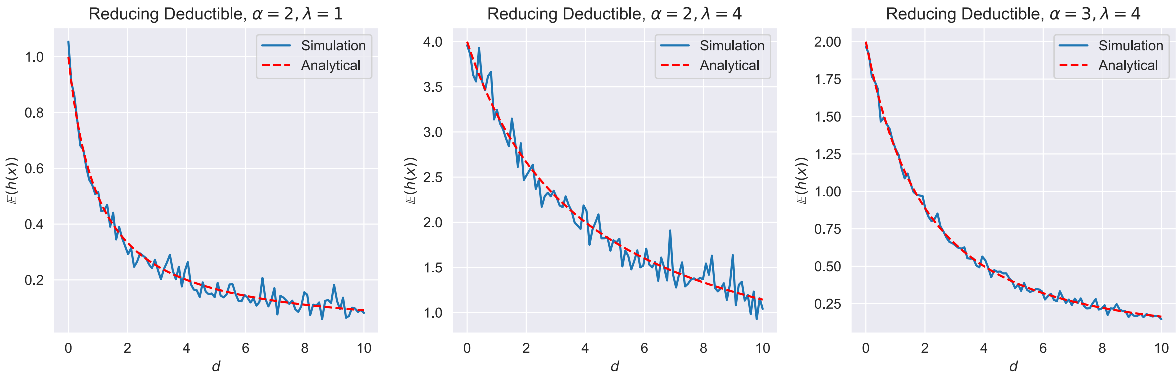

Figure 3: Comparison of net premium dependency on the reducing deductible parameter d across various parameters of Pareto distribution of loss. Comparison of estimated results based on 10^4 random variables against exact values.

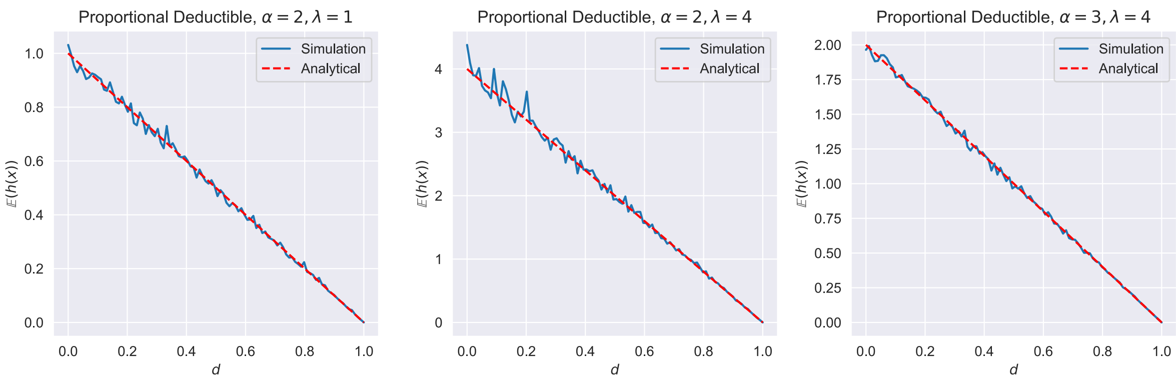

Figure 4: Comparison of net premium dependency on the proportional deductible parameter d across various parameters of Pareto distribution of loss. Comparison of estimated results based on 10^4 random variables against exact values.

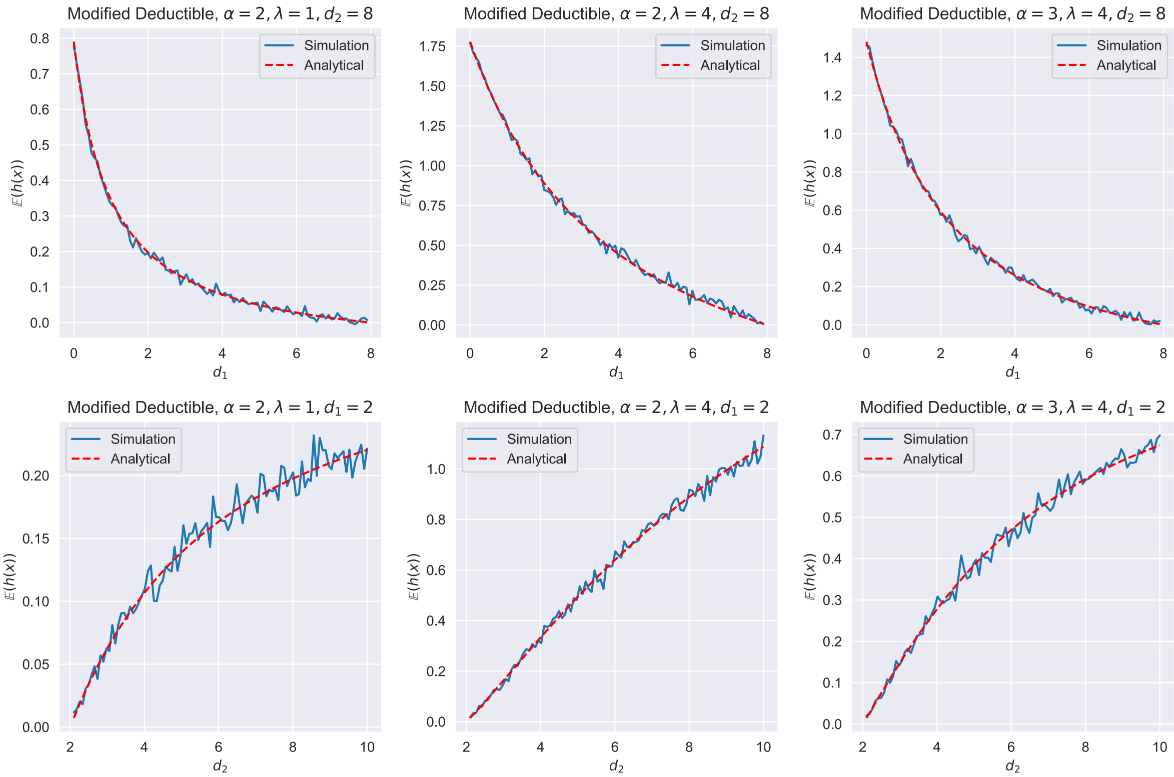

Figure 5: Comparison of net premium dependency on the modified reducing deductible parameters d_1 and d_2 across various parameters of Pareto distribution of loss. Comparison of estimated results based on 10^4 random variables against exact values.

Task 3.

Task description

Using Monte Carlo simulations, demonstrate that the mixed Poisson distribution with Λ following a gamma distribution is a negative binomial distribution.

A Mixed Poisson Distribution extends the standard Poisson distribution by allowing its rate parameter λ to be random instead of fixed. Unlike the standard Poisson distribution, where λ is fixed, here λ changes according to a distribution. This allows the model to handle overdispersion (variance larger than the mean), which the standard Poisson distribution cannot.

For different values of the parameters $\alpha \in \{2, 3, 5\}$ and $\beta \in \{0.5, 1, 2\}$, 10^5 random variables were generated from a mixed Poisson distribution with Λ following a gamma distribution. The probability density function of the gamma distribution is given by

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \quad \alpha, \beta > 0 \tag{3}$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$.

The generated random variables will be compared with the negative binomial distribution with the probability mass function given by the formula

$$P(X = k) = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \ldots \quad r > 0, \quad p \in (0, 1). \tag{4}$$

where $r = \alpha, p = \frac{\beta}{1+\beta}$.

We consider the hypothesis H_0 that the mixed Poisson distribution with Λ following a gamma distribution is a negative binomial distribution.

Initially, the histograms of the random variables were compared with the corresponding probability mass functions of the negative binomial distribution. The results are shown in Figure 6. For each considered value of α and β, no significant differences are observed between the histogram and the probability mass function. The plots suggest the validity of hypothesis H_0.

Next, for each considered value of α and β, 10^5 random variables were generated from the negative binomial distribution with the density following formula 4 and compared with the random variables from the mixed Poisson distribution using the Kolmogorov-Smirnov test at a 5% confidence level. As shown in Table 1, for each parameter, the p-value is higher than 5%, therefore there is no basis to reject hypothesis H_0.

Based on the Kolmogorov-Smirnov test and visual comparison shown on Figure 6, we accept that the mixed Poisson distribution with Λ following a gamma distribution is a negative binomial distribution.

$\alpha\backslash\beta$	**0.5**	**1.0**	**2.0**
2.0	0.66	0.19	0.81
3.0	0.82	1.00	0.94
5.0	0.44	0.40	0.99

Table 1: p-value from the Kolmogorov-Smirnov test for different values of α and β.

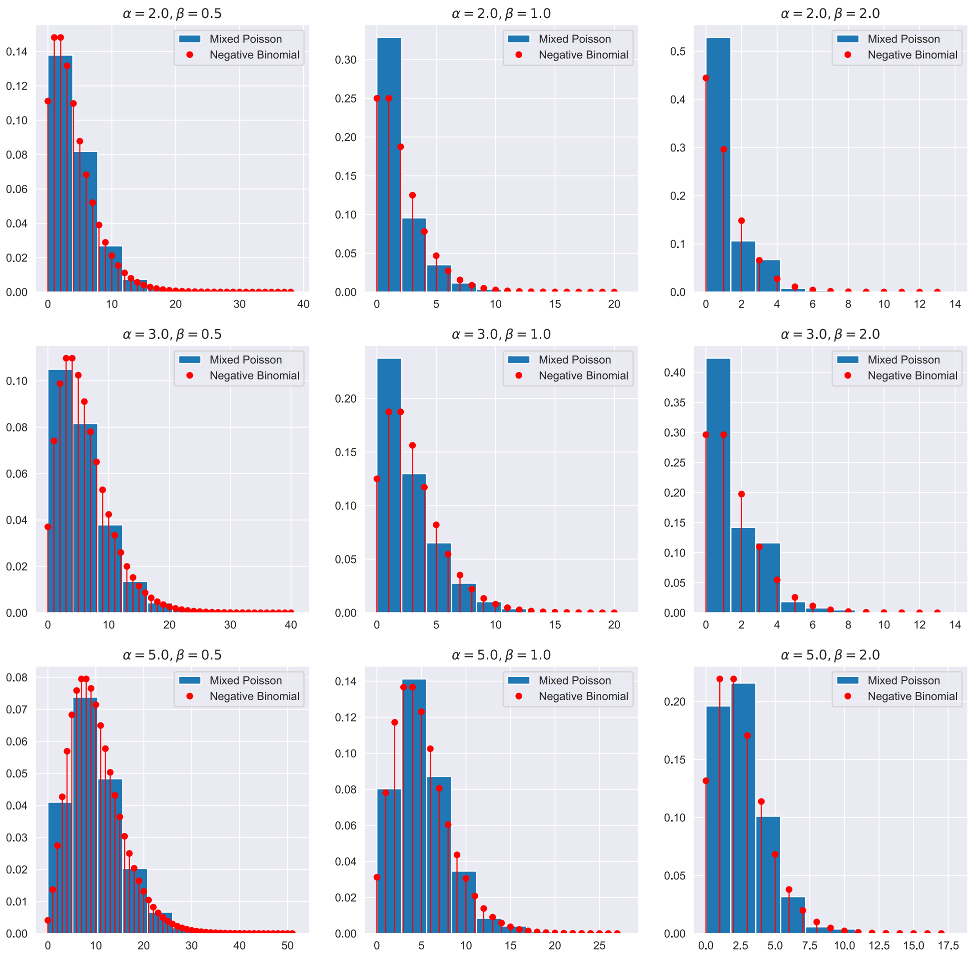

Figure 6: Comparison of the histogram of 10^5 random variables from a mixed Poisson distribution and the probability mass function of the negative binomial distribution for different α and β values.

Task 4.

Task description

Using Monte Carlo simulations, investigate the distribution of $S = \sum_{k=1}^{N} X_k$, where:

1. X follows a uniform distribution over $\{1, 2, \ldots, 10\}$.
2. X follows a logarithmic distribution with parameter p such that $E[X] = 5$ (solve for p numerically).

Consider two cases for N:

1. N follows a Poisson distribution $P(\lambda = 5)$.
2. N follows a geometric distribution with p such that $E[N] = 5$.

Compare these results with those obtained using Panjer's recursion (a total of 4 cases).

To begin the simulations, the parameters p for the considered logarithmic and geometric distributions must be determined. For the logarithmic distribution, p will be calculated numerically, while for the geometric distribution, it can be derived directly using the fact that

$$\mathbb{E}[X] = \frac{p}{1-p} = 5. \tag{5}$$

From this, we obtain that p for the considered geometric distribution is $\frac{1}{6}$.

Solving numerically for p in the considered logarithmic distribution

The objective was to determine the parameter p for X from a logarithmic distribution such that $E[X] = 5$.

For each $p \in \{0.01, 0.02, \ldots, 0.99\}$, 10^6 random variables were generated from the logarithmic distribution. The expected value of X was then estimated using the sample mean. The results are shown in Figure 7. Most importantly, the plot indicates that within the considered range of p, there exists a value of p for which $E[X] = 5$. While this value may not belong to $\{0.01, 0.02, \ldots, 0.99\}$, it is certainly within the interval $[0.01, 0.99]$. Moreover, since the function is strictly increasing, there exists a unique value of p such that its value equals 5.

The value closest to 5 corresponds to $p = 0.93$. To determine a more precise value of p, the method was repeated for p in the vicinity of $p = 0.93$, specifically for p within the $\{0.92, 0.921, \ldots, 0.939, 0.94\}$. Once again, for each p, 10^6 random variables were generated from the logarithmic distribution, and the expected value of X was estimated using their sample mean. $p = 0.93$ was again identified as the closest value for achieving $E[X] = 5$ (with $\hat{E}[X] \approx 4.999$). Therefore, for the simulations, we set $p = 0.93$, rounded to three decimal places.

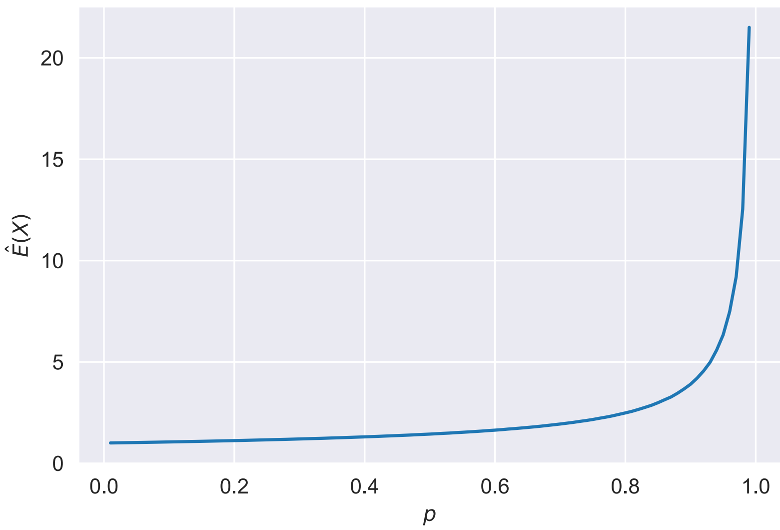

Figure 7: Estimated expected value of logarithmic distribution for different p values.

Investigation of distributions

The goal is to examine four distributions of the random variable $S = \sum_{k=1}^{N} X_k$:

- $S_{PU} : N \sim \mathcal{P}oiss(\lambda = 5)$ and $X \sim \mathcal{U}(\{1, 2, \ldots, 10\})$,
- $S_{GU} : N \sim \mathcal{G}eom(p = \frac{1}{6})$ and $X \sim \mathcal{U}(\{1, 2, \ldots, 10\})$,
- $S_{PL} : N \sim \mathcal{P}oiss(\lambda = 5)$ and $X \sim \mathcal{L}og(p = 0.93)$,
- $S_{GL} : N \sim \mathcal{G}eom(p = \frac{1}{6})$ and $X \sim \mathcal{L}og(p = 0.93)$.

To generate each S, 10^5 values of N were sampled, followed by generating N random variables X for each sampled N. The values of S were then computed as the sum of these N variables X. This process produced 10^5 samples of S for each distribution, and their distributions are shown in Figure 8.

From Figure 8, it can be observed that the distributions of S where N follows a Poisson distribution are characterized by less dispersion compared to those where N follows a geometric distribution. When N is geometrically distributed, higher losses occur more frequently, but there is also a higher likelihood of losses close to zero. Moreover, the dispersion is further increased when X follows a logarithmic distribution. From the insurance company's perspective, it is safer to deal with a distribution where the probability of incurring a high loss is very low.

In Table 2, the summary statistics of the studied distributions are presented. The mean values are similar: for S_{GL} and S_{PL} they are close to 25, while for S_{GU} and S_{PU} they are approximately 27.5. These values can be easily verified analytically using the formula $\mathbb{E}[S] = \mathbb{E}[N] \cdot \mathbb{E}[X]$.

The standard deviation values confirm the observations from the histograms: for N following a Poisson distribution, they are lower than for N following a geometric distribution. Additionally, the standard deviation is lower when X comes from a uniform distribution. Furthermore, the maximum values for N from a geometric distribution are 2–3 times higher than those for N from a Poisson distribution.

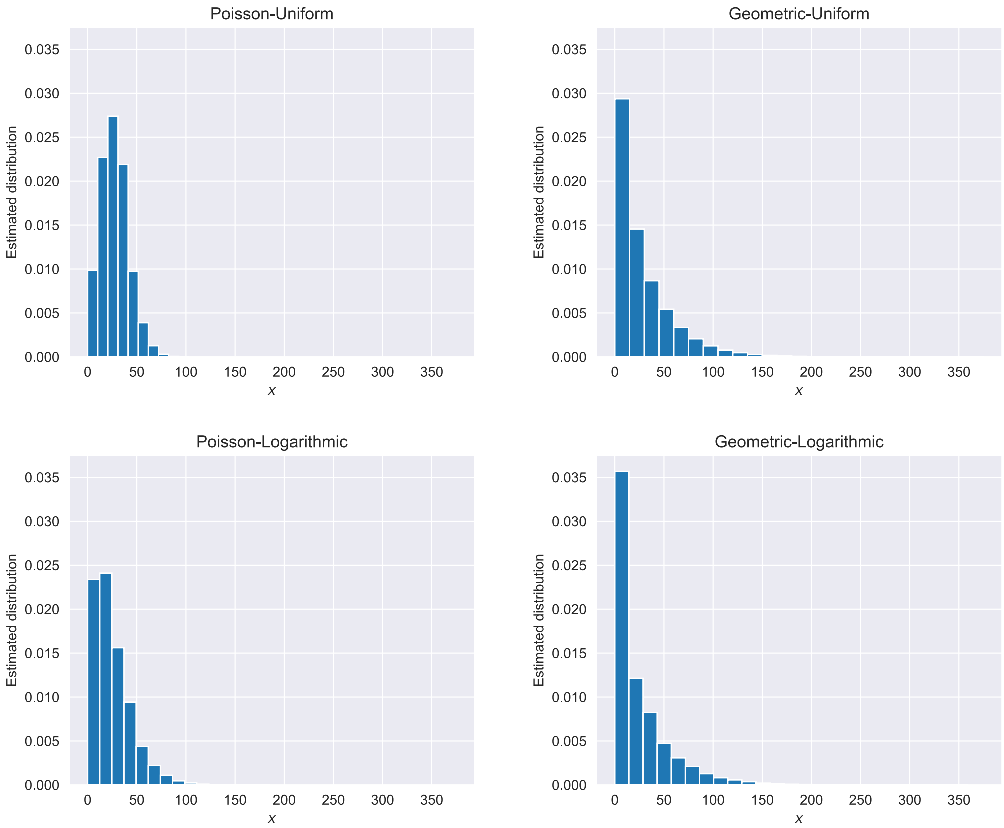

Figure 8: Histograms of 10^5 random variables from different S distributions.

Comparison with the results of Panjer's formula

In the next step, the results were compared with the Panjer formula. The comparison is presented in Figure 9. The results from the Panjer formula are presented as a continuous line to improve the clarity of the plots. However, these results should not be interpreted as continuous, as the formula returns values only for integer values of x.

For N from the Poisson distribution, the values returned by the Panjer formula match the histogram exactly. The histogram is not as well approximated for N from the geometric distribution. The convergence of the Panjer formula and the Monte Carlo method was further investigated in the plot shown in Figure 10.

On Figure 10, a comparison between the Panjer formula and the histogram generated with the number of bins corresponding to the number of distinct generated values of S is presented. This ensures the histogram is as accurate as possible. Upon visualizing the results, one can observe that the Panjer formula values align with those obtained through the Monte Carlo method, even for N following a geometric distribution.

Distribution	Mean	Std	Min	Max
Geometric–Logarithmic	24.88	31.24	0	359
Geometric–Uniform	27.44	30.85	0	375
Poisson–Logarithmic	25.09	19.00	0	185
Poisson–Uniform	27.55	13.97	0	103

Table 2: Summary statistics for generated random variables by distribution.

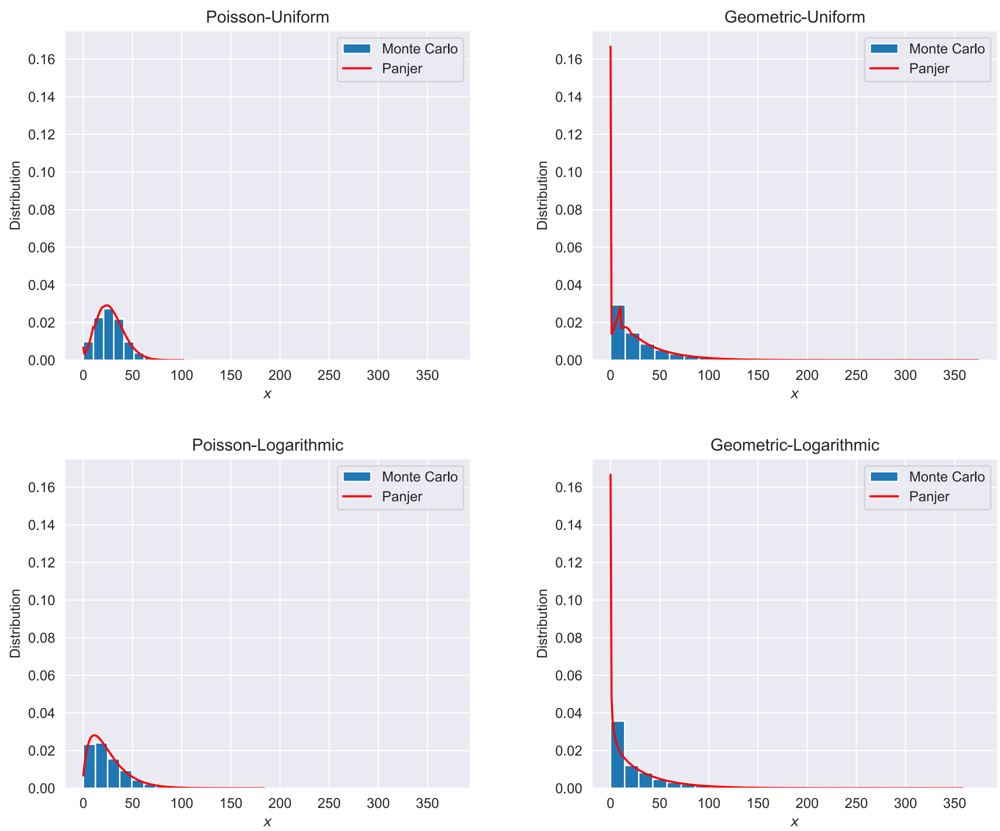

Figure 9: Comparison of the histogram of 10^5 random variables from different S distributions with the results from the Panjer formula.

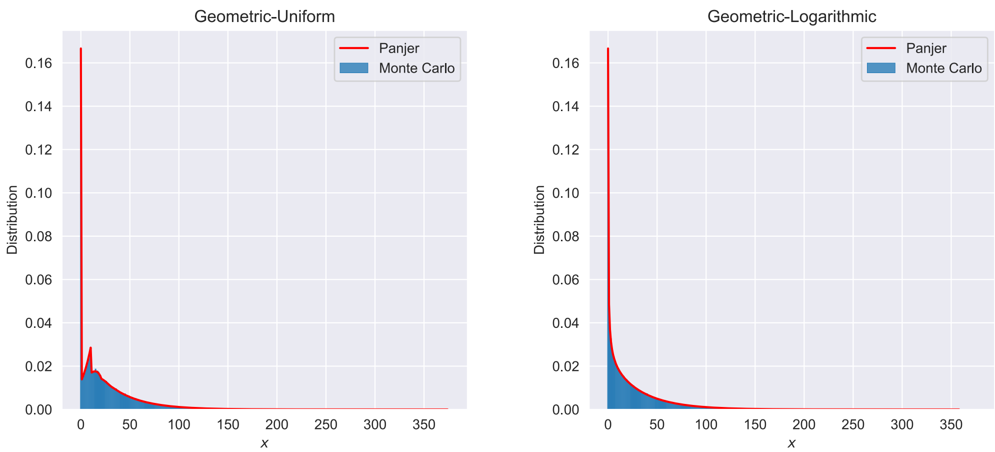

Figure 10: Comparison of the histogram of 10^5 random variables from different S distributions with the results from the Panjer formula for $N \sim \mathcal{G}eom$.

Summary

Task 1 evaluates the expected time to ruin for an insurer's surplus process under specified conditions. Using simulation of 10^5 sample paths, the expected time to ruin was estimated as $\mathbb{E}(N) \approx 30.21036$. An analytical approach confirmed this result, deriving $\mathbb{E}(N) = 30$ based on properties of the process and the law of total expectation. The close alignment between the simulated and analytical results demonstrates the accuracy of the numerical approximation and confirms the robustness of the calculated expected time to ruin.

In **Task 2**, we analyzed the effects of different deductible structures - integral, reducing, proportional, and modified reducing - on net premiums for losses following a Pareto distribution. Analytical results were compared to numerical simulations, showing close alignment. The choice of deductible type and parameters significantly influences the insurer's cost and risk-sharing dynamics.

Task 3 investigates the relationship between a mixed Poisson distribution, where the rate parameter Λ follows a gamma distribution, and the negative binomial distribution.Monte Carlo simulations for various gamma parameters α and β confirm that the mixed Poisson distribution aligns closely with the negative binomial distribution. This was verified through histograms and the Kolmogorov-Smirnov test, supporting the hypothesis of equivalence between the two distributions.

In **task 4** the objective is to investigate the distribution of $S = \sum_{k=1}^{N} X_k$, where X follows either a uniform or logarithmic distribution, and N follows either a Poisson or geometric distribution. The logarithmic parameter p was determined numerically, while the geometric parameter p was derived analytically. Simulations showed that S has higher variance when N follows a geometric distribution, with further increases when X is logarithmic.

The distributions generated through Monte Carlo simulations were compared to those obtained using Panjer's formula. For N from the Poisson distribution, Panjer's formula closely matched the histogram. For N from the geometric distribution, slight deviations were observed, but the overall alignment remained strong. The Panjer formula's accuracy was confirmed through detailed visualization, showing convergence with the Monte Carlo method even for geometric N.