Risk Theory Final Project

CERVANTES VASCONCELOS MARIA FERNANDA

November 2024

1 Model for the frequency

For the distribution of N, we discussed in class that if $E(N) < V(N)$ (as is the case here), a potential candidate for the distribution is the Negative Binomial. However, we need to verify whether it is an appropriate choice. To assess this, we plotted a bar graph representing the observed data and overlaid it with a scatter plot of a Negative Binomial(r,p) distribution where:

$$p = \frac{E(N)}{V(N)} \qquad r = \frac{pE(N)}{1-p}$$

From the resulting figure, we observe that a Negative Binomial distribution with parameters $r = 5.0773173$ and $p = 0.3028381$ is a good possibility for our distribution of N.

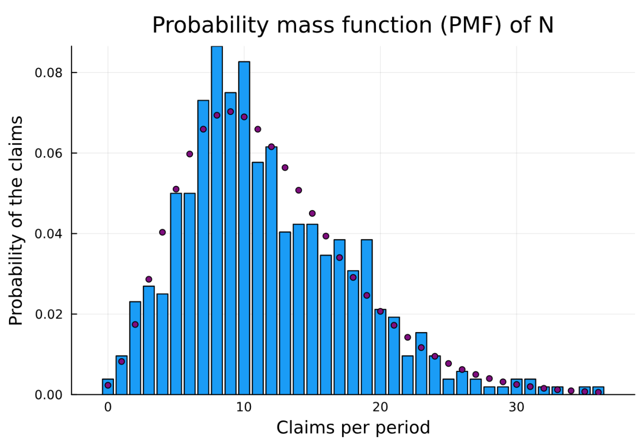

Figure 1: Distribution of N

2 Model for the severity $Y|N = n$

Also, we have that $Y|N = n$ has a distribution function LogNormal$(\mu(n), \sigma(n))$ and that :

$\mu(n) = g_1(n) + \varepsilon_1$ where $\varepsilon_1 \sim$ Normal$(0, \omega_1)$.

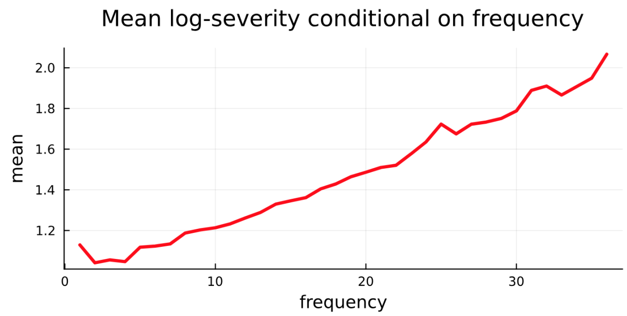

Figure 2: Graph "Mean log-severity conditional on frequency"

In the graph "Mean log-severity conditional on frequency" we can note that $\mu(n)$ can be seen as a linear function, base on this, we propose the function $g_1(n) = \alpha + \beta n$. Under linear regression, $g_1(n)$ is not a random variable, consequently, the mean and the variance of $\mu(n)$ are as follows:

$E(\mu(n)) = E(\alpha + \beta n + \varepsilon_1) = \alpha + \beta n + E(\varepsilon_1) = \alpha + \beta n + 0 = \alpha + \beta n$

$V(\mu(n)) = V(\alpha + \beta n + \varepsilon_1) = V(\varepsilon_1) = \omega_1$

Given that $Z = c + X$, where c is a constant and $X \sim$Normal(μ, σ^2), it follows that $Z \sim$Normal$(c + \mu, \sigma^2)$. So in this case, $\mu(n) \sim$Normal$(\alpha + \beta n, \omega_1)$.

To estimate α and β, we used the method of least squares, we have a function $h(\alpha, \beta) := \sum_{i=1}^{r} (t_i - g_1(n))^2$, where $t_1, ..., t_r$ are the observed mean log-severity, we need to minimize this function to obtain $\hat{\alpha}, \hat{\beta}$.

In this case, we have formulas to obtain $\hat{\alpha}$ and $\hat{\beta}$, but the idea is to differentiate for the values we want to estimate and equal that to zero. With this, we will have a system of equations, which will need to be solved to obtain $\hat{\alpha}$ and $\hat{\beta}$.

$$\hat{\alpha} = \frac{(\sum n_i^2)(\sum \mu(n_i)) - (\sum n_i)(\sum n_i \mu(n_i))}{m(\sum n_i^2) - (\sum n_i)^2}$$

$$\hat{\beta} = \frac{m \sum n_i \mu(n_i) - (\sum n_i)(\sum \mu(n_i))}{m(\sum n_i^2) - (\sum n_i)^2}$$

With the information we have $\hat{\alpha} = 0.9503870245790009$ and $\hat{\beta} = 0.028339161434865128$.

Now that we have the values of α and β, we can see in the next graph that our values and the function we propose are a good choice to estimate $\mu(n)$.

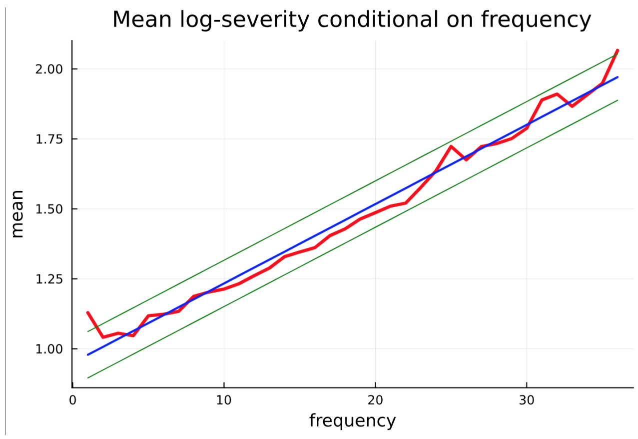

Figure 3: Graph with the linear regression

We need to do the same but for the standard deviation of conditional log-severity, in this case, the standard deviation acts like a constant or a function that becomes constant quickly, we have that

$\sigma(n) = g_2(n) + \varepsilon_2$ where $\varepsilon_2 \sim \text{Normal}(0, \omega_2)$

Also, $\sigma(n)$ will be a Normal. We are going to contrast two possible forms of $g_2(n)$, first we are going to have that $g_2(n) = \frac{1}{a+be^{-cn}}$, to estimate $\hat{a}, \hat{b}$ and $\hat{c}$, we have to minimize $h(a,b,c) := \sum_{i=1}^{r}(t_i - g_2(n))^2$, where $t_1, ..., t_r$ are the observed standard deviation log-severity. We obtain the next values:

$\hat{a}$=3.9216879944095195

$\hat{b}$= 1.6775824244917708

$\hat{c}$=0.17027167227406184.

These values minimize h and it is equal to 0.018471755005940213.

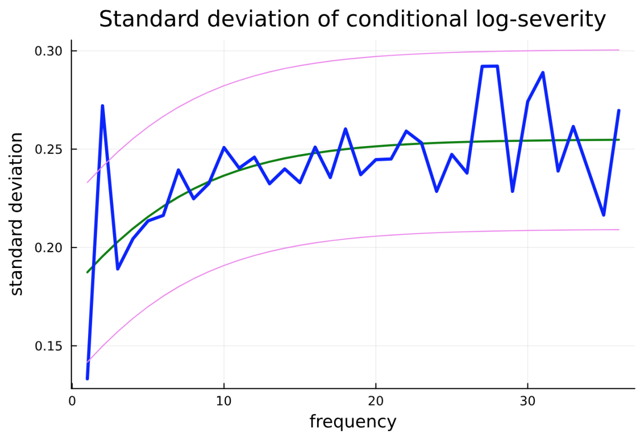

Figure 4: Graph with the linear regression of the standard deviation when $g_2(n) = \frac{1}{a+be^{-cn}}$

Now we are going to try with a constant, we will have that $g_2(n) = d$ and $h(d) := \sum_{i=1}^{r}(t_i - g_2(n))^2$. Minimizing that function, we have that $d = 0.24086151123046873$, and we obtain the next graph.

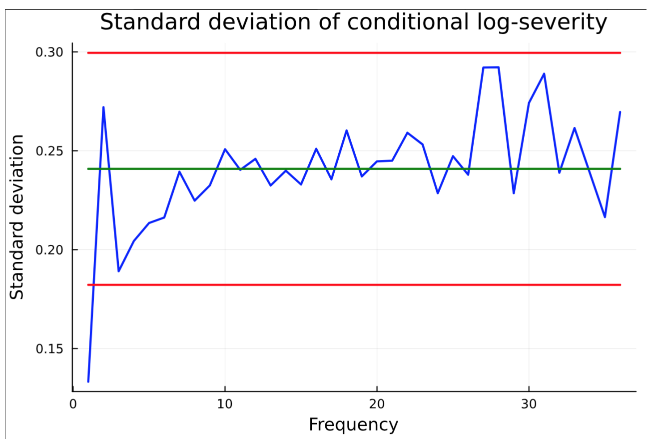

Figure 5: Graph with the linear regression of the standard deviation when $g_2(n) = d$

But when we compare the error, it happens that it is less when we use

$$g_2(n) = \frac{1}{a + be^{-cn}}$$

In this case the error is 0.018471755005940213, and with $g_2(n) = d$ it is 0.030438397296128638. So we are going to work with $g_2(n) = \frac{1}{a+be^{-cn}}$.

So we have that

$\mu(n) \sim$Normal$(\hat{\alpha} + \hat{\beta}n, \omega_1)$

$\sigma(n) \sim$Normal$(\frac{1}{\hat{a}+\hat{b}e^{-\hat{c}n}}, \omega_2)$

$\omega_1 \sim$Normal$(0, 0.0017831372796448587)$

$\omega_2 \sim$Normal$(0, 0.0005432866728878436)$

So we can proceed to the simulations of S_k, which represent the total claims in period $k \in 1, 2, ..., 10$, because we are considering 10 years. Considering an annual low-risk interest rate of 10%, and an ROE of 15% we can calculate our initial capital using the Mexican regulatory solvency capital requirement, where $C_0 = (1 - (ROE - i))SCR(S1)$. For simulating the capital at time t (C_t), we are going to use the next formula $C_t = C_{t-1} + \sum_{j=1}^{n_t} \pi_j(t) - S(t)$.

With this, we can graph our C_t. We simulate each S_k 10,000 times, but we only graph 50 possible scenarios.

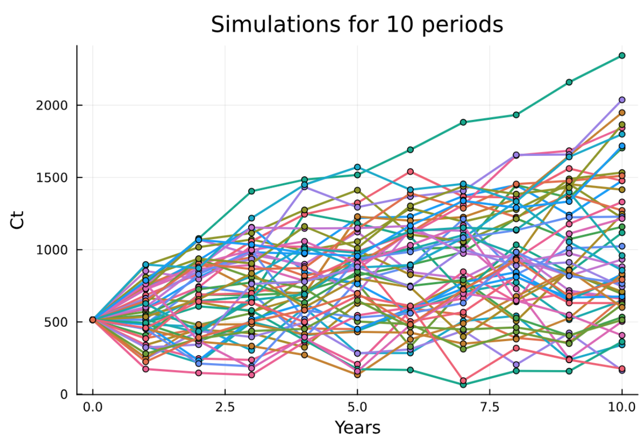

Figure 6: C_t for 10 periods

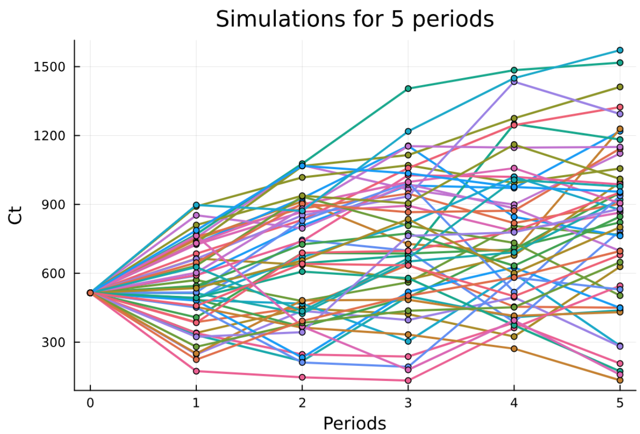

Figure 7: C_t for 5 periods

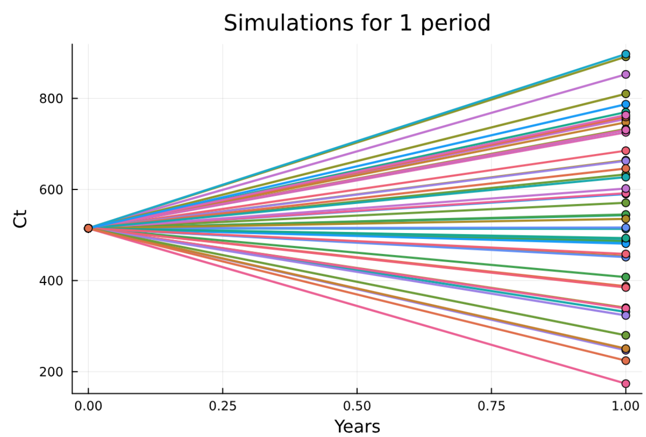

Figure 8: C_t for 1 period

For the probability of ruin in one year, we count the simulations that in the first year have $C_1 < 0$ and divide it by the total of the simulations.
For the other probabilities, we use the next formula

$$P(C_t < 0) = P(C_1 < 0) + P(C_2 < 0 \mid C_1 \geq 0)P(C_1 \geq 0) + P(C_3 < 0 \mid C_2 \geq 0)P(C_2 \geq 0) + \dots$$

The probability of ruin in one year is 0.005, as the Mexican regulations indicate. It will change every time we simulate for five years, but it is approximately 0.09 and 0.18 for ten years.
Additionally, we aim to estimate the severity of ruin. To achieve this, we analyze the quantities where C_t becomes less than 0 for the first time. Specifically, we calculate both the mean and the median of these values. From this analysis, we observe that the median is smaller than the mean, indicating that

the distribution is right-skewed. This is when the right tail is longer than the other. Skewness defines the asymmetry of a distribution. In this case, we have a right-skewed distribution, which means that the mean overestimates the most common values in a positively skewed distribution.