Final Project – Internal model

Actuarial statistics

Asier Gil Sedano

Index

Acronyms

EIOPA – European Insurance and Occupational Pensions Authority

SCR – Solvency capital requirement

i.i.d. – Independent and identically distributed

PDF – Probability density function

CDF – Cumulative distribution function

ECDF – Empiric cumulative distribution function

DGP – Data generation process

MSE – Mean square error

MM – Method of moments

PM – Percentile matching

ML – Maximum likelihood

VaR – Value at risk

TVaR – Tail value at risk

CLT – Central limit theorem

1. Overview

In this Project, an insurance company internal model is constructed considering life and non-life insurances. The objective of the model is to properly evaluate the premium risk (the risk of having more claims than expected) so the economic capital can be estimated.

The insurance company to be modelled is composed by the following number of policies:

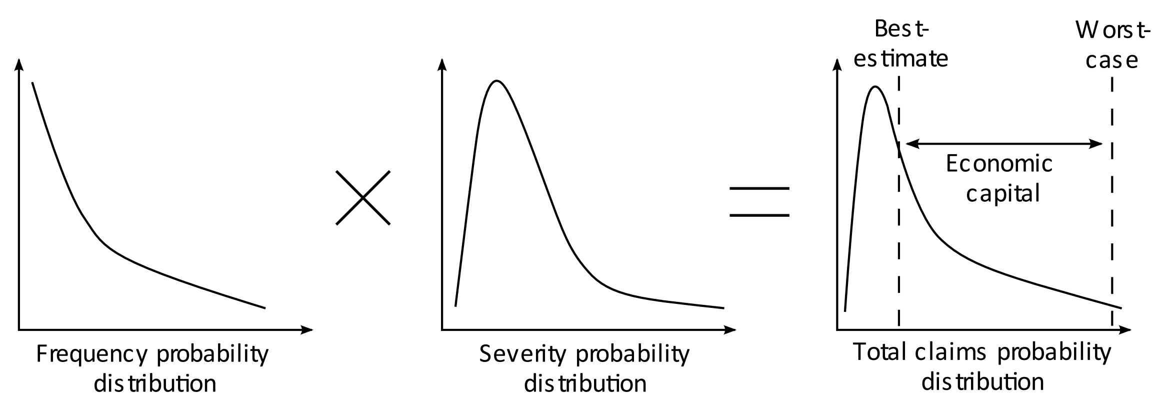
- Car insurances: 25,234
- Life insurances: 20,809

For the car claims frequency and severity, the data of last year 1,116 policyholders with 2,230 claims are available. As for the life insurances, the USA-2015 mortality table and the policyholders ages are provided. It is known that the insurance amount of each life policyholder is 1M USD.

The project has four main parts:

1. Estimation methods validation
2. Probability density function goodness of fit testing
3. Aggregated models construction for life and non-life
4. Montecarlo simulation for cost scenarios

With this model, the expected cost will be evaluated together with the $VaR_{99.5}$ (maximum expected cost in one year with a confidence interval of 99.5%) and $TVaR_{99.5}$ (expected cost if the cost surpasses the $VaR_{99.5}$). This way the worst-case scenarios will be quantified, and the company solvency capital requirement will be determined.



2. Available data

2.1. Non-life

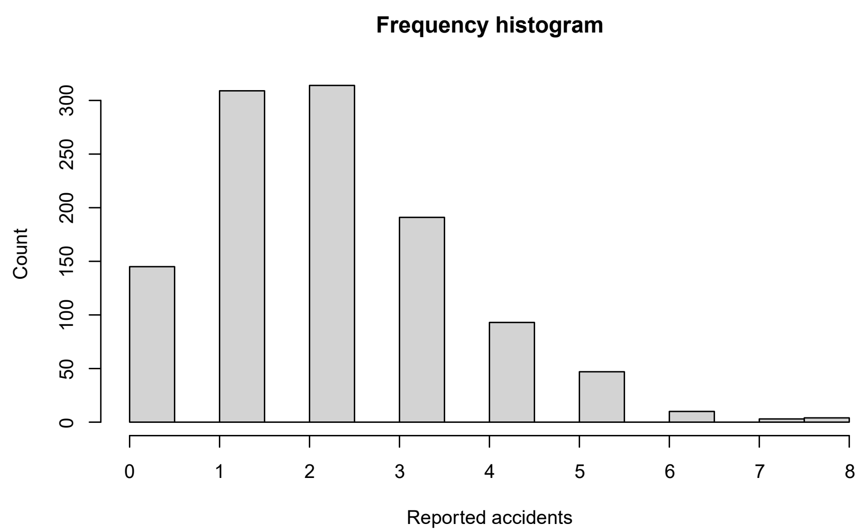
The available data for non-life policies is composed by the frequency (number of claims per policy) and the severity (cost of each claim). For each of them, it is assumed they are i.i.d. so the original distribution can be estimated.

To estimate the maximum cost (distribution tail) and its expected value, Montecarlo simulation will be performed with the estimated distribution. This makes the estimation task critical since the result will completely depend on the accuracy of this estimation.

2.1.1. Frequency

The available data for number of claims per policy has the following histogram.

It can be observed it is a discrete right-skewed distribution with a relatively long tail. The Poisson and negative binomial distributions will be tested.



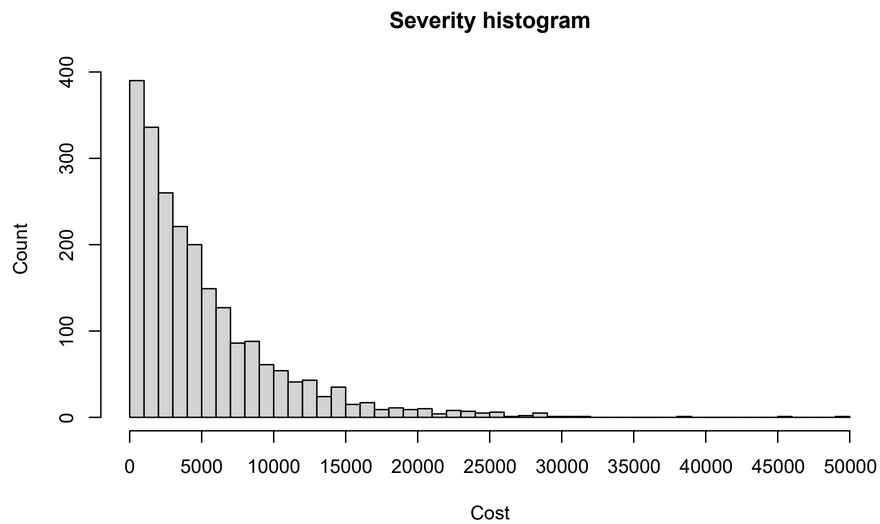
2.1.2. Severity

The available data for the severity (cost) of each policy report has the following histogram.

From the histogram it can only be observed that the distribution shows that as the cost increases, the probability decreases, but it is not clear if it could be an exponential function or say, a gamma function.

Considering the histogram does not provide any clear shape about the type of distribution it could be a smoothing will be applied.

The smoothing and the plots are programmed in the lines 18 to 82.



2.1.2.1 Smoothing

To obtain a more reliable representation of the sample that accounts for the values not shown in the sample but also possible, a smoothing function is applied to the sample of size n. This way, it can be obtained a better overview of the PDF than that provided by the histogram. The smoothed sample function is calculated as follows.

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Being K a kernel function and h the range of application. The kernel applied is set as the uniform probability density function $U(0,2)$. The h is calculated as follows for the Uniform Kernel:

$$h = 0.9 \cdot \min\left(\hat{\sigma}; \frac{IQR}{1.34}\right)\frac{1}{\sqrt[5]{n}}$$

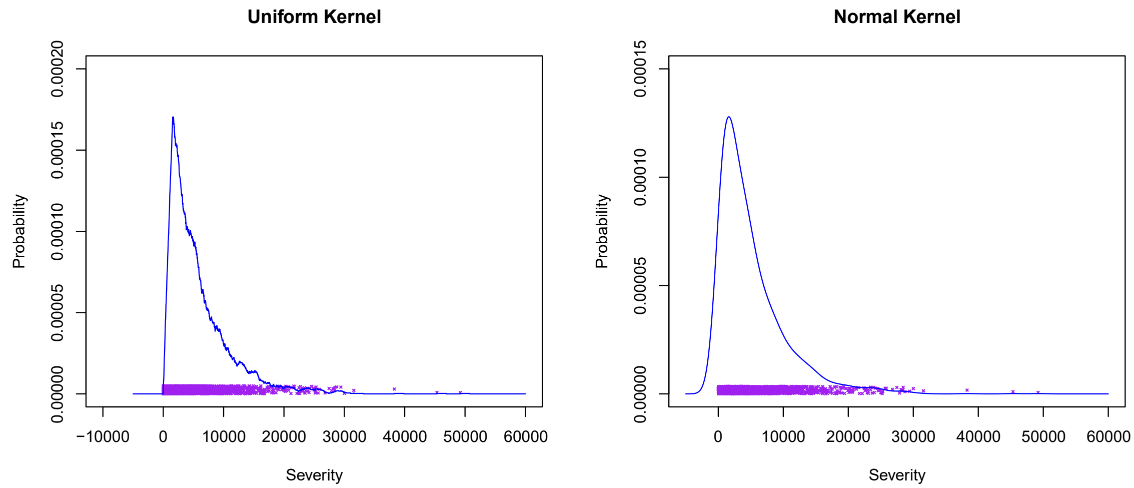
Being IQR the Interquartile Range.

Additionally, the smoothing has been applied by using a normal Kernel $N(0,1)$. The h used in this case that minimizes the MSE is defined as:

$$h = 1.06\frac{\hat{\sigma}}{\sqrt[5]{n}}$$

Some other popular Kernel functions could also be tested such as the Epanechnikov or the tri-cube function Kernels.

With this smoothing applied, the following PDF are obtained where it can be observed that the function is positively skewed, and negative values shall have 0 probability (note the normal Kernel does not account for this issue). Also, it can be observed it has a long tail and there is a significant probability to obtain very large values, so it shall be a long-tailed distribution.



Based in these observations, a positively skewed distribution with a long tail will be tested. The distributions that will be considered are the Lognormal, Exponential, Erlang and Gamma distributions.

2.2. Life

For the life insurances, the information of the [linked dataset] is provided. For simplicity, no differentiation by the gender of the insurance policyholder is considered so the death rate is unique. The available data are the USA-2015 mortality tables with the different death probabilities by age.

Also, the death rates and the number of policyholders by age are provided. Being *qx* the death rate of a person with a specific age, and *ni* the number of policyholders of that age.

For the simulation, the death event will be modelled as a Bernoulli experiment with a death probability of *qx*. Considering the number of policyholders of an age x as the number of times the experiment is performed, the random variable of the number of people dying at that age ($\boldsymbol{X_x}$), will follow a binomial distribution of size $n_x = ni$ and probability $q_x = qx$.

$$\boldsymbol{X_x} \sim B(n_x, q_x)$$

Therefore, to simulate the life insurance cost scenarios, Montecarlo simulation will be performed for each age.

3. Estimation methods validation

According to EIOPA, the following three methods are valid for the PDF parameters estimation:

- Method of moments
- Percentile matching
- Maximum likelihood

To decide which of the methods is best for the parameter's estimation, we must perform a Data Generating Process (DGP). The DGP consist in the generation of samples from a known distribution. This way the bias, the MSE and the consistency can be computed and tested for the different estimation methods.

These three methods are evaluated for all the PDFs that are going to be tested.

- Frequency
  - Poisson
  - Negative Binomial
- Severity
  - Log Normal
  - Exponential
  - Erlang
  - Gamma

The Bias is computed as the expected difference between the estimator and the real parameter.

$$Bias = E[\hat{\phi}] - \phi$$

The Mean Square Error is computed as the second order moment centered in the estimator:

$$MSE(\hat{\phi}) = E\left[(\hat{\phi} - \phi)^2\right] = Var(\hat{\phi}) + Bias(\hat{\phi})^2$$

As for the consistency, it is checked that the estimator converges to the real parameter when the size of the sample increases. For that, the estimator is computed for increasing sample sizes so this convergency can be checked. This is done graphically but it can be summarized as follows.

$$\lim_{n\to\infty} Pr(|\hat{\phi}_n - \phi| > \varepsilon) = 0$$

When checking the estimator, we need to confirm out estimator is consistent, unbiased and minimizes the MSE. The minimization of the estimation MSE is critical since it will minimize the uncertainty of our estimation.

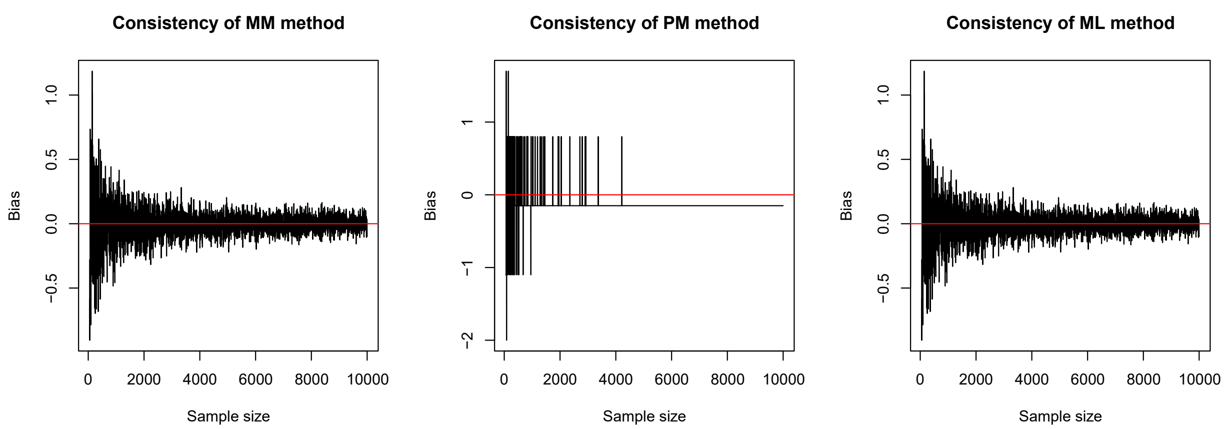
This validation has been computed with a DGP in the code lines 84 to 973. The plots are also programmed in these lines.

3.1. Poisson distribution

The following table summarizes the results obtained for a 1116 sample size DGP of $\lambda = 25$ Poisson distribution. Also, the estimation performance up to a 10.000 sample size is provided (consistency).

method	Bias ($\hat{\lambda}$)	MSE ($\hat{\lambda}$)
MM	-0.00129	0.02320
PM	-0.09110	0.06441
ML	-0.00129	0.02320

Consistency o $\hat{\boldsymbol{\lambda}}$ estimator.



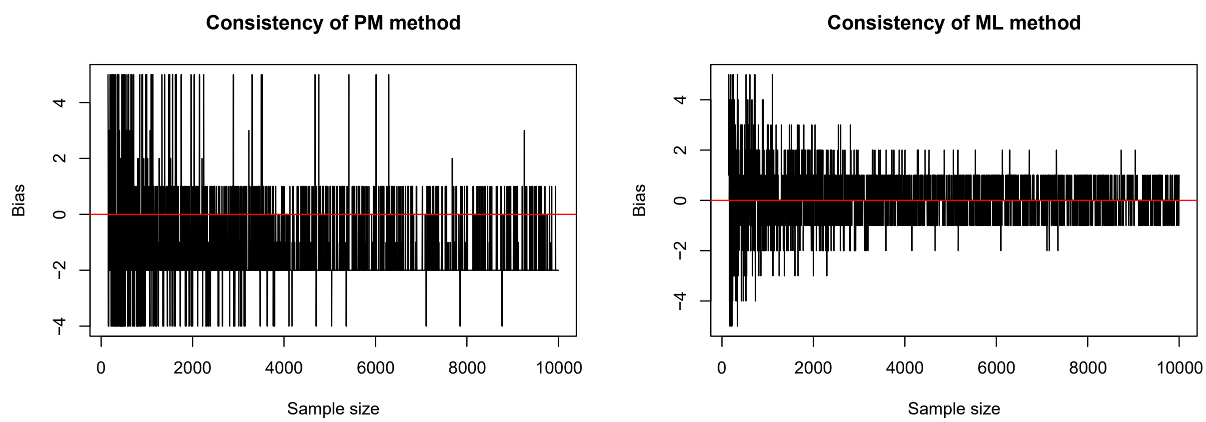
It can be observed that both the MM and the ML methods are consistent while the PM method is not. As for the bias and the MSE, the MM and the ML have the same results, with a better performance than the PM method. The **method of moments** will be used for the estimation of the parameters (indeed the ML and the MM for a Poisson parameter are the same estimation).

3.2. Negative binomial

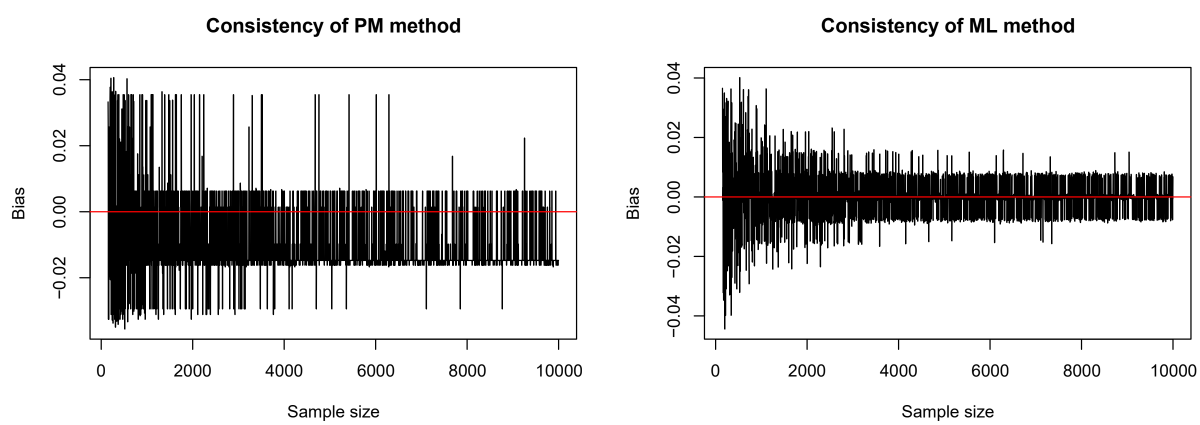
The following table summarizes the results obtained for a 1,116 sample size DGP of $r = 25$ and $p = 0.25$ parameters Negative Binomial distribution. Also, the estimation performance up to a 10.000 sample size is provided (consistency). Considering the negative binomial r parameter shall be a natural number; the method of moments estimation is not considered for this distribution.

method	Bias ($\hat{r}$)	Bias ($\hat{p}$)	MSE ($\hat{r}$)	MSE ($\hat{p}$)
PM	-0.95933	-0.00765	5.22143	0.00029
ML	0.15367	0.00097	2.24107	0.00013

Consistency of $\hat{\boldsymbol{r}}$ estimator.



Consistency of $\hat{\boldsymbol{p}}$ estimator.



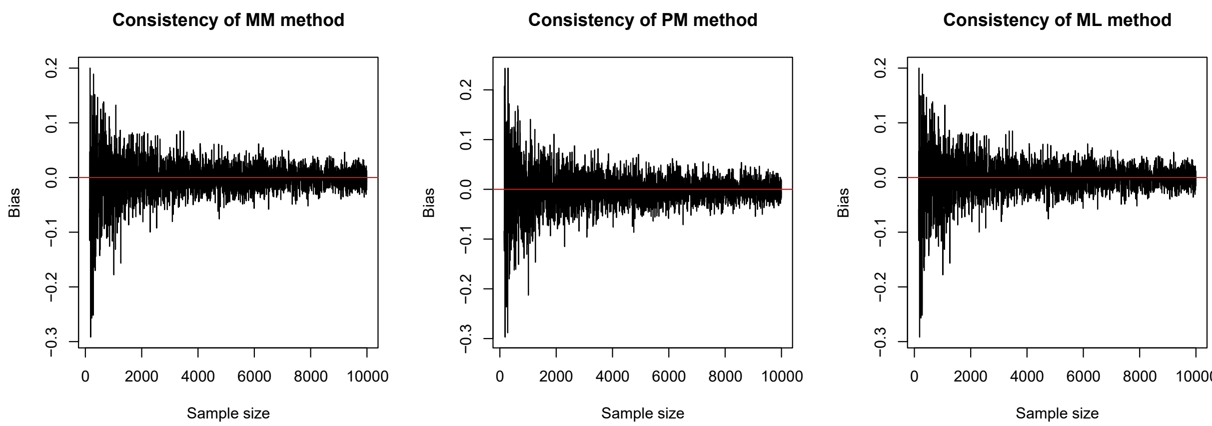
It can be observed that the ML is consistent while the PM is not. As for the bias and the MSE, the ML method has a greater performance than the PM. The **maximum likelihood will be used** for the estimation of the parameters.

3.3. Lognormal distribution

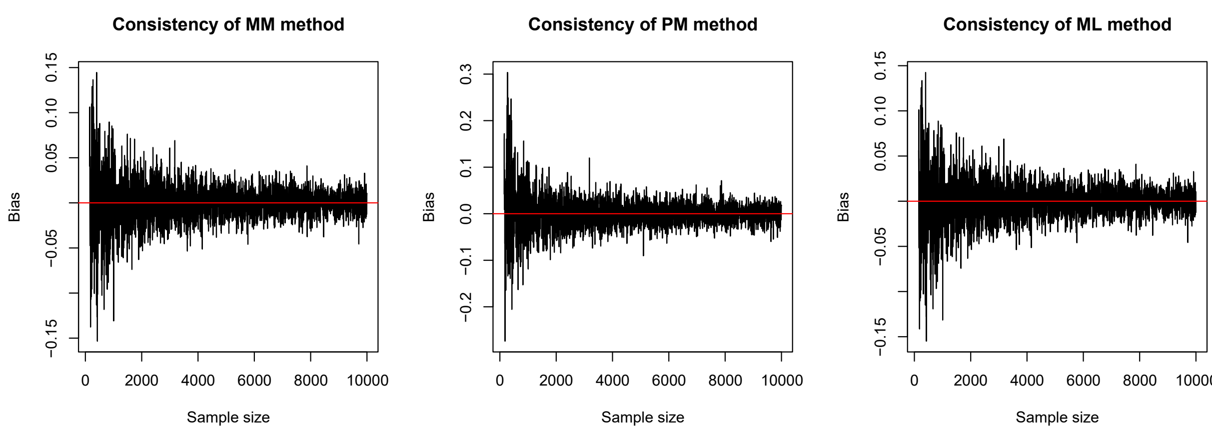
The following table summarizes the results obtained for a 2,230 sample size DGP of $\mu = 6$ and $\sigma = 1.5$ log-normal distribution. Also, the estimation performance up to a 10.000 sample size is provided (consistency).

method	Bias ($\hat{\mu}$)	Bias ($\hat{\sigma}$)	MSE ($\hat{\mu}$)	MSE ($\hat{\sigma}$)
MM	-0.00015	-0.00027	0.001068	0.000494
PM	-0.00043	-0.00114	0.001366	0.001061
ML	-0.00015	-0.00061	0.001068	0.000494

Consistency of $\hat{\boldsymbol{\mu}}$ estimator.



Consistency of $\hat{\boldsymbol{\sigma}}$ estimator.



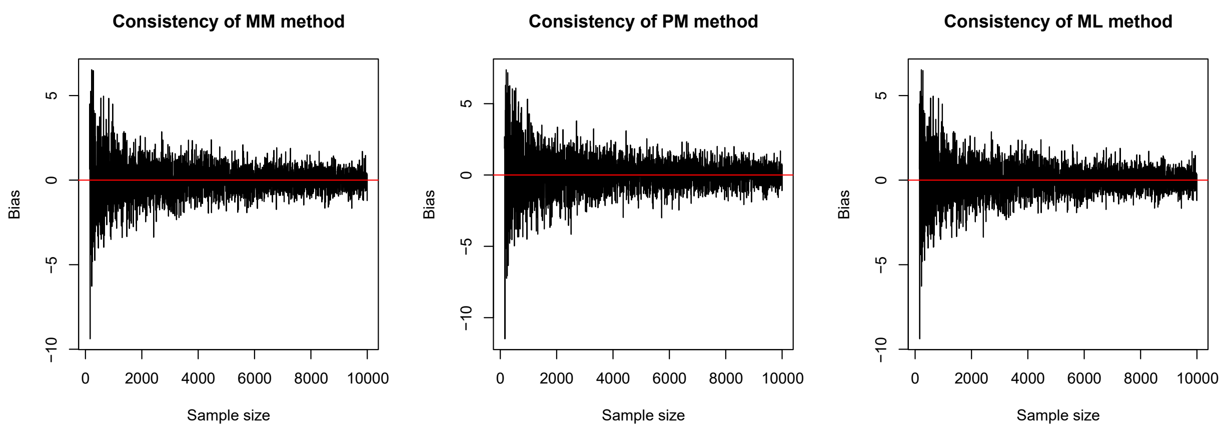
The results the same for the MM and the ML. All the methods are consistent, the bias and MSE are the same for the MM and the ML, while it is worse for the PM. Therefore, for the Lognormal function the **maximum likelihood** will be used for the parameter's estimation, method of moments could be also used.

3.4. Exponential distribution

The following table summarizes the results obtained for a 2,230 sample size DGP of a scale parameter $\beta = 50$ exponential distribution. Also, the estimation performance up to a 10.000 sample size is provided (consistency).

method	Bias ($\widehat{\beta}$)	MSE ($\widehat{\beta}$)
MM	0.002856	1.123751
PM	-0.00133	1.725935
ML	0.002841	1.123746

Consistency of $\widehat{\boldsymbol{\beta}}$ estimator.



The results are the same for the MM and the ML. All the methods are consistent, the bias and MSE are the same for the MM and the ML, while it is worse for the PM. Therefore, **maximum likelihood method will be used for the estimation**.

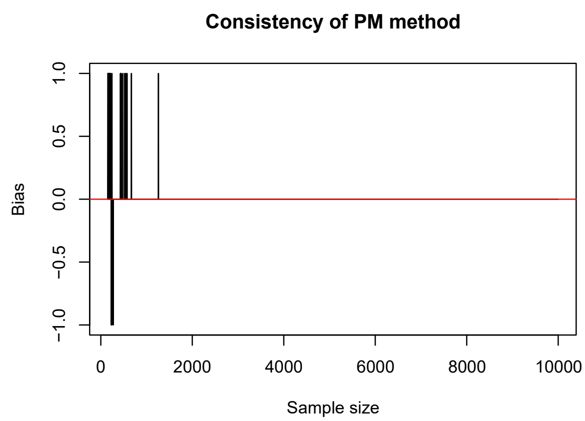
3.5. Erlang Distribution

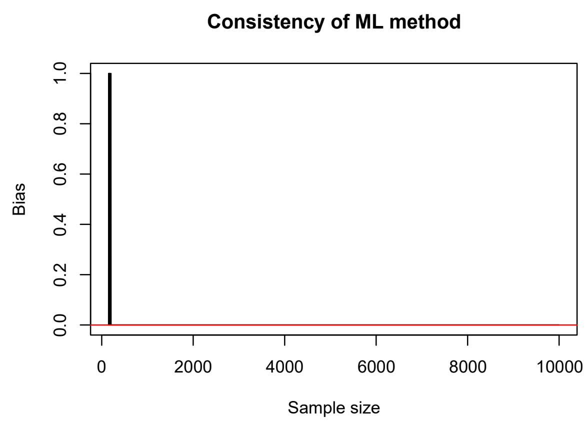
The following table summarizes the results obtained for a 2230 sample size DGP of $k = 3$ and $\beta = 50$ Erlang distribution. Also, the estimation performance up to a 10.000 sample size is provided (consistency).

Considering the Erlang is a special case of the Gamma distribution where k parameter shall be a natural number; the method of moments estimation is not considered for this distribution.

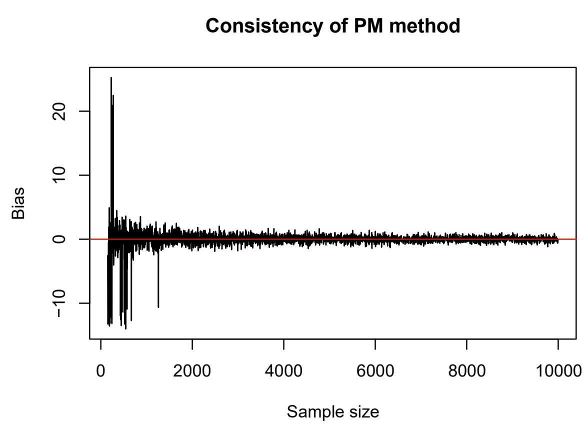
method	Bias ($\hat{k}$)	Bias ($\hat{\beta}$)	MSE ($\hat{k}$)	MSE ($\hat{\beta}$)
PM	0.000367	-0.00967	0.000367	0.593631
ML	0	-0.00025	0	0.374383

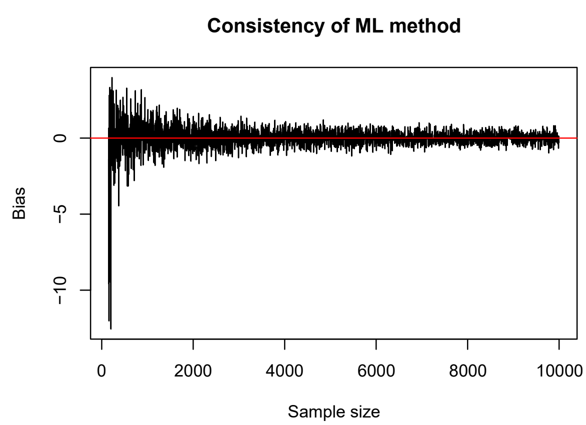
Consistency of $\hat{\boldsymbol{k}}$ estimator.





Consistency of $\hat{\boldsymbol{\beta}}$ estimator.





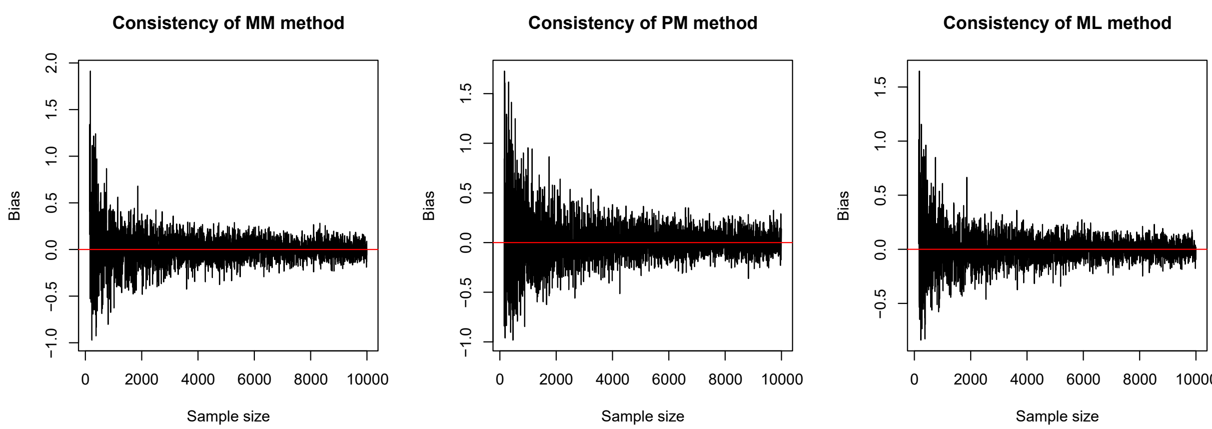
Both the PM and ML methods are consistent. As for the Bias and the MSE, the ML shows better results than the PM method. Therefore, the **maximum likelihood method will be used**.

3.6. Gamma distribution

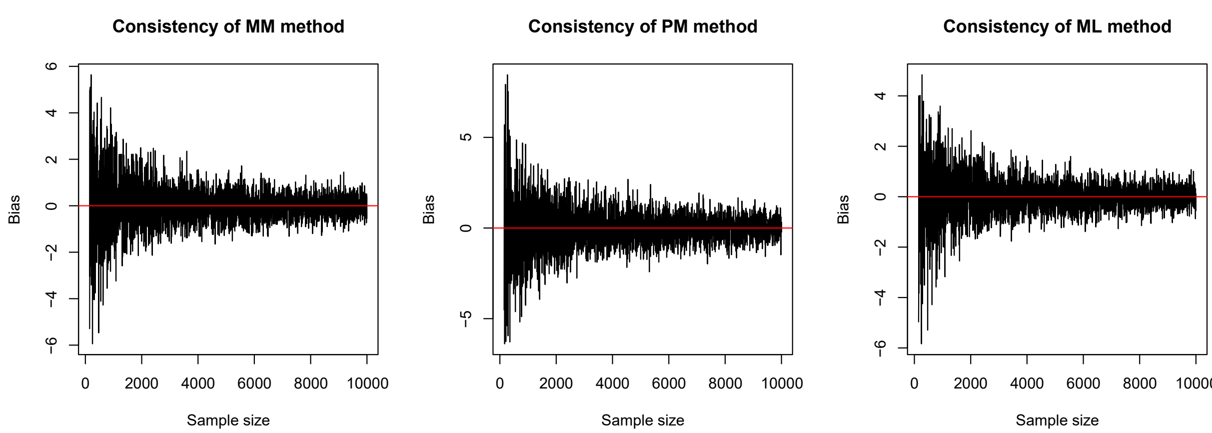
The following table summarizes the results obtained for a 2230 sample size DGP of $\alpha = 5$ and $\theta = 25$ Gamma distribution. Also, the estimation performance up to a 10.000 sample size is provided (consistency).

method	Bias ($\hat{\alpha}$)	Bias ($\hat{\theta}$)	MSE ($\hat{\alpha}$)	MSE ($\hat{\theta}$)
MM	0.00373	0.00336	0.02680	0.72627
PM	0.01152	-0.01912	0.04397	1.22691
ML	0.00307	0.00038	0.02054	0.56685

Consistency of $\hat{\boldsymbol{\alpha}}$ estimator.



Consistency of $\hat{\boldsymbol{\theta}}$ estimator.



All methods are consistent. As for the bias and the MSE, the ML has the best results, followed by the MM and PM. Therefore, **maximum likelihood** will be used for the estimation of the parameters, but the method of moments could be also used since same performance is expected.

4. Distribution estimation

For the distribution estimation, the PDF parameters will be estimated with the methods described in the previous point. Once the parameters are estimated, the distribution and the sample will be graphically compared to observe the possible deviations from one to the other.

After the graphical comparison, different test will be performed by computing different statistics. These tests are the following.

Chi square test (χ^2)

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Being O_i the observed value and E_i the expected value.

Kolmogorov Smirnoff test (D)

$$D_n = max\left|F_n(x) - \hat{F}(x)\right|$$

Being $F_n(x)$ the empirical distribution function and $\hat{F}(x)$ the estimated distribution function.

Cramer Von Mises test (W^2)

$$W_n^2 = n \int \left(F_n(x) - \hat{F}(x)\right)^2 w(x)\, \hat{f}(x) dx$$

Being $w(x)$ a weight function equal to 1 and $\hat{f}(x)$ the estimated PDF.

For the discrete variables, the test is defined as follows:

$$W_n^2 = \sum_{i=1}^{k} \left(F_n(x_i) - \hat{F}(x_i)\right)^2 w(x_i)\, \hat{f}(x_i)$$

Anderson Darling test (A^2)

It is the same test as the Cramer Von Mises with the following weight function.

$$w(x) = \frac{1}{\hat{F}(x)\left(1 - \hat{F}(x)\right)}$$

For the discrete distributions (Frequency) the four tests will be performed. As for the continuous distributions (Severity), all but the chi square test will be used. Once the test is performed and the statistic is computed, the p-value will be calculated being the null hypothesis (H_0) that the data comes from the estimated distribution.

The graphical comparisons and the test are coded in the lines 975 to 1589.

4.1. Frequency

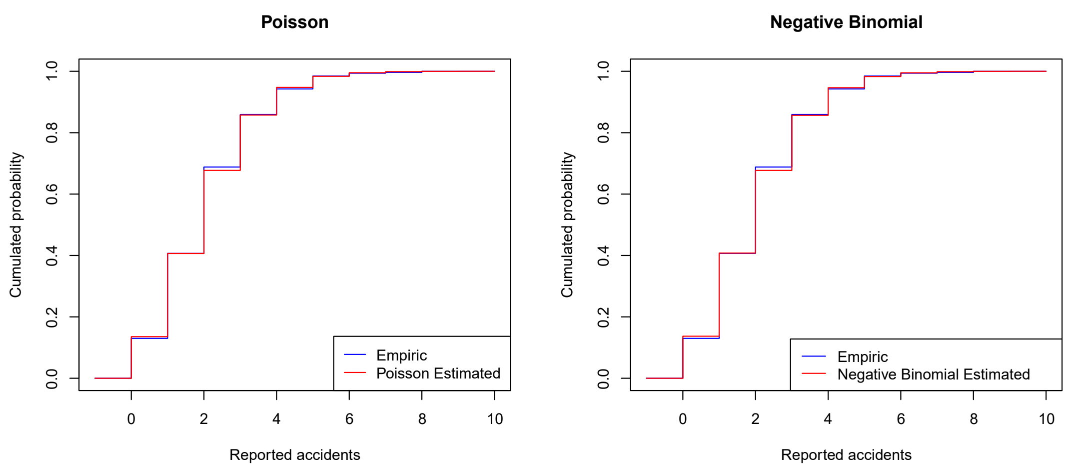
The parameter estimated by the method of moments for the Poisson distribution is $\hat{\lambda} = 1.998208$.

As for the negative binomial distribution, the estimated parameters by maximum likelihood are:

$\hat{r} = 195$ and $\hat{p} = \ 0.9898562$.

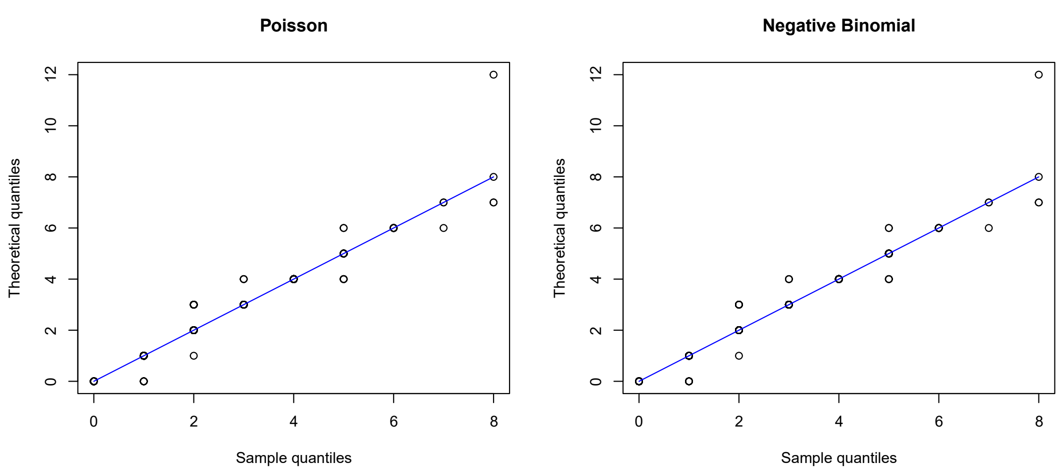
4.1.1. Graphical comparison

First, the CDF of the estimated function and the sample is graphically compared:



Both estimated distributions are similarly close to the sample, so there is no evidence to indicate what estimation could fit it better.

Additionally, the Q-Q plots are also provided with no conclusion for both distributions.



4.1.2. Goodness of fit tests

Four tests have been performed for the discrete function's goodness of fit test being the null hypothesis the data comes from the estimated distribution. The calculation is programmed in the code lines 1158 to 1197.

The following results have been obtained.

	Chi Square		Kolmogorov-Smirnov		Cramer Von Mises		Anderson Darling	
distribution	χ^2	**p-value**	D	**p-value**	W^2	**p-value**	A^2	**p-value**
Poisson	14.089	0.1195	0.01101	0.8036	0.0447	0.8428	0.2975	0.8761
Negative binomial	13.297	0.1474	0.01104	0.7978	0.0478	0.8264	0.3106	0.8660

The goodness of fit tests shows almost the same results for the Poisson and the Negative Binomial.

For the Chi square test, there is a low statistical significance (0.125-0.147) to say it comes from any of the Poisson or Negative binomial. As for the rest of the tests, it is shown a much higher statistical significance. Considering that for three of the tests, there is a high significance level (0.796-0.876), it can be assumed the frequency follows one of these two distributions.

When determining which of the two distributions fits better the sample, the test Chi square shows it is more statistically significant that it follows a Negative binomial while the rest of the tests show it is more likely it follows the Poisson.

Since the goodness of fit tests do not provide any clear conclusion, a cross validation will be performed to the sample so both distributions can be evaluated.

4.1.3. Cross validation

The cross validation consists in randomly dividing the sample dataset in different datasets, named *train dataset* and *test dataset*. This way the model can be calibrated with the *train dataset* and the performance of the model is tested with the *test dataset*. This way, different models can be tested so the accuracy of the model can be tested one against the other, in this case, by computing the MSE. The MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Being y_i the observed number of claims and $\hat{y}_i$ the estimated number of claims.

As a summary, the following steps for the cross validation have been performed. Code lines 1199 to 1250.

1. The sample is randomly divided into two samples, being one sample the *training dataset* (75% of the data) and the other sample the *test dataset* (25% of the data).
2. The Poisson and Negative Binomial distributions parameters are estimated with the train dataset.
3. The MSE of the *test dataset* and the estimated distribution is computed
4. The process is repeated N times.

The distribution with a lower expected MSE will be considered a better estimation.

The results obtained from the cross validation are summarized in the following table.

distribution	$E[MSE]$	$sd(MSE)$
Poisson	20.903	14.527
Negative binomial	22.147	15.377

It is observed that the Poisson distribution performance is better than that of the negative binomial, therefore the frequency will be **modelled as a Poisson distribution**.

Note that the Negative binomial arises a Poisson distribution when

$$\lim_{r\to\infty} NB\left(r, 1 - \frac{\lambda}{r+\lambda}\right) \sim Pois(\lambda)$$

Therefore, considering the estimated parameters of the Negative binomial are $\hat{r} = 195$ and $\hat{p} = 0.9898562$ and the Poisson estimated parameter is $\hat{\lambda} = 1.998208$, it can be confirmed that the estimated Negative binomial is arising as a Poisson distribution. The better results in some statistical tests like the Chi square, are only due to an overfitting that is reflected as a worse performance in the cross validation.

To account for these issues related to the overfitting some other statistical tests like the Akaike information criterion could be considered when testing the distributions.

4.2. Severity

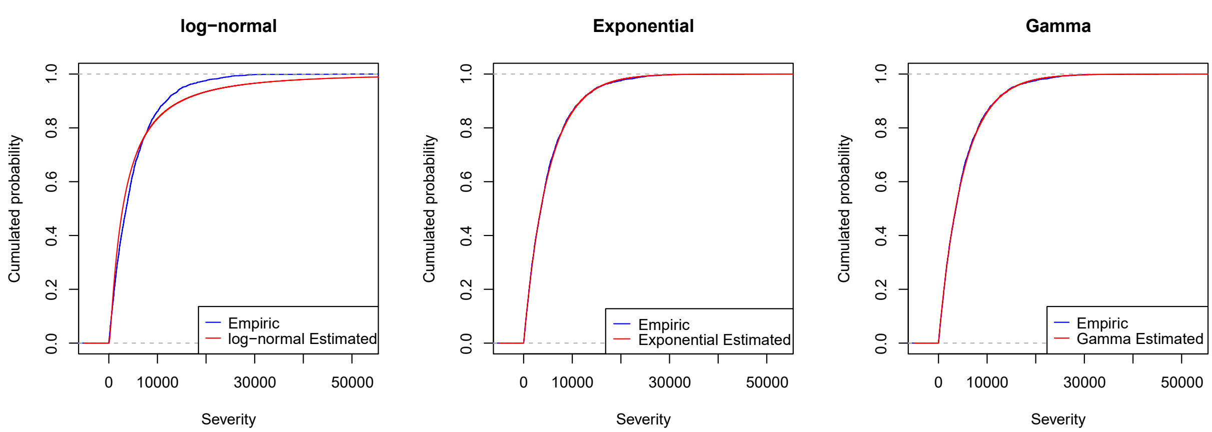
The following table summarizes the estimated parameters for the distribution that will be tested.

distribution	method	parameter	parameter
Lognormal	ML	$\hat{\mu} = 7.953878$	$\hat{\sigma} = 1.291855$
Exponential	MM	$\hat{\lambda} = 5106.187$	
Erlang	ML	$\hat{k} = 1$	$\hat{\beta} = 5106.187$
Gamma	ML	$\hat{\alpha} = 0.9891021$	$\hat{\theta} = 5162.448$

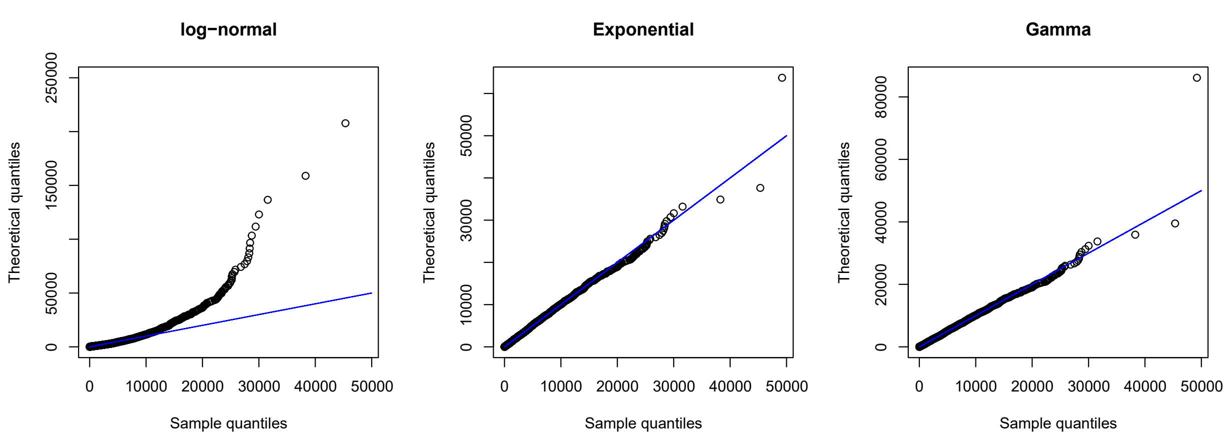
Note the Erlang distribution arises to the exponential distribution when $\hat{k} = 1$. Therefore, from now on, the exponential distribution only will be analyzed since **both estimated distributions are equal**.

4.2.1. Graphical comparison

The CDF of the estimated functions are plotted with the sample so they can be compared.



The Lognormal distribution does not fit well the empiric distribution. Both the Exponential/Erlang and the Gamma distributions graphically fit the sample. Additionally, the Q-Q plots are also provided with same observation.



4.2.2. Goodness of fit tests

Three tests have been performed for the function's goodness of fit test being the null hypothesis the data comes from the estimated distribution.

The results are summarized in the following table. Note that due to the iterative way the p-value is calculated, the minimum p-value considered is 3.33E-5. Therefore, all p-values lower than that value will appear as 0. The calculation is programmed in the code lines 1484 to 1522.

	Kolmogorov-Smirnov		Cramer Von Mises		Anderson Darling	
distribution	D	**p-value**	W^2	**p-value**	A^2	**p-value**
Lognormal	0.07514	0	4.46776	0	26.81728	0
Exponential	0.01303	0.7825	0.04138	0.9279	0.25999	0.9631
Gamma	0.01253	0.8384	0.04019	0.9399	0.25829	0.9658

The goodness of fit tests shows there no statistical significance that the data follows the Lognormal estimated distribution. As for the Exponential and the Gamma distribution, there is sufficient statistical significance to consider it is following one of these distributions.

It is important to note that the Gamma distribution has two parameters while the Exponential distribution has only one parameter. As it was happening with the frequency data, this slightly better results could probably be due to an overfitting issue of the gamma model. Therefore, a cross validation will be performed to ensure this slightly higher results are not due to an overfitting issue.

Also, it is important to remark that the Gamma distribution arises to the Exponential distribution when $\hat{\alpha}$ tends to 1. Considering that $\hat{\alpha} = 0.9891021$, the overfitting issue is likely to be happening.

4.2.3. Cross validation

The cross validation methodology is as described for the frequency estimation. As for the MSE estimation, since the severity is a continuous variable, the data has been divided in ranges, so the estimated value is computed as the number of claims estimated in that cost range. See code lines 1524 to 1562 for more details.

The results obtained from the cross validation are summarized in the following table.

distribution	$E[MSE]$	$sd(MSE)$
Exponential	7.788	3.451
Gamma	8.219	3.810

In this case the exponential distribution provides a better result, meaning the overfitting issue could justify the better results in the tests for the Gamma distribution. Therefore, **the exponential function will be used in the model as the severity model** since a better performance is expected.

5. Aggregated model

5.1. Single car insurance policy cost

Once the distribution functions have been determined for the non-life insurances, the aggregated model is constructed. First, one policy cost is modelled. The cost of one policy will be determined by the number of claims and the cost of all these claims.

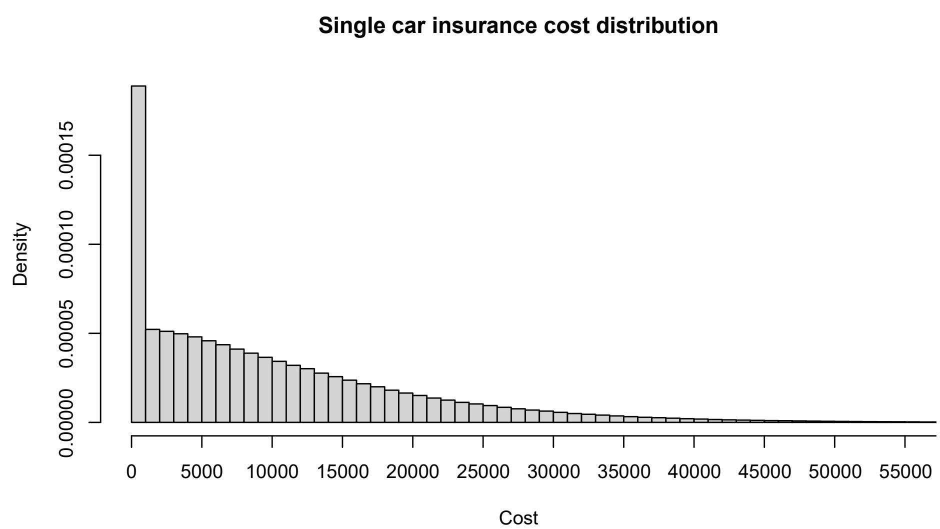
As it has been stated before, the random variable of the frequency of the claims $\boldsymbol{N}$ follows a Poisson distribution with an estimated parameter $\hat{\lambda} = 1.998208$

$$\boldsymbol{N} \sim Pois(\hat{\lambda})$$

As for the severity (cost) random variable of each claim $\boldsymbol{C}$, it has been stated it follows an Exponential distribution with an estimated scale parameter $\hat{\beta} = 5106.187$

$$\boldsymbol{C} \sim Exp(\hat{\beta})$$

Therefore, the cost will be simulated by the sum of $\boldsymbol{N}$ random variables $\boldsymbol{C}$. By the use of Montecarlo simulation to generate frequency and severity scenarios, the following distribution is obtained for the random variable single cost $\boldsymbol{X}$.



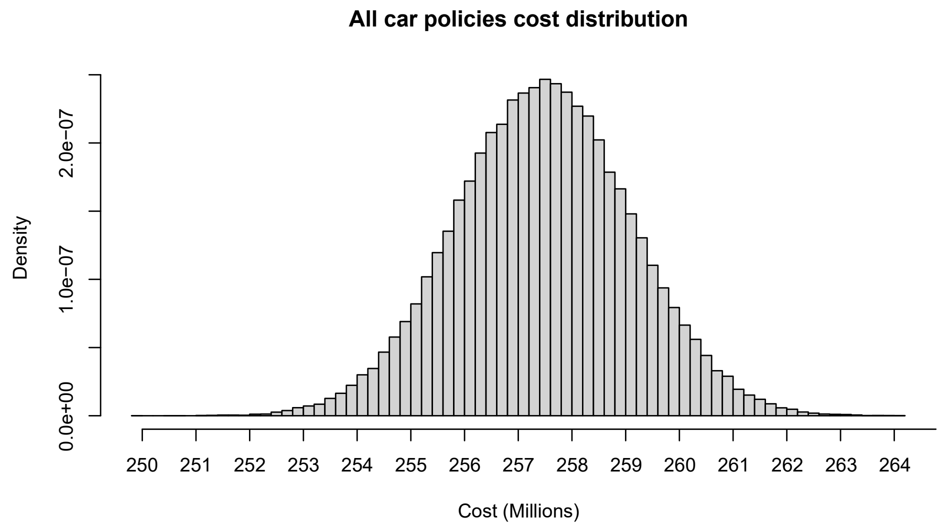
With an expected value of 10,199 USD and a standard deviation of 10,205 USD. The $VaR_{99.5}$ is 49,507 USD and the $TVaR_{99.5}$ is 57,272 USD.

It can be observed that the possibility of having no cost is high and the distribution has a long tail.

The calculation is programmed in the code lines 1591 to 1616.

5.2. All car policies total cost

If this same aggregated model is now simulated to compute the total cost of the 25,234 policies, the following result is obtained (code lines 1618 to 1659).



The expected cost due to all the car claims is 257.47M USD and the standard deviation is 1.62M USD. The $VaR_{99.5}$ is 261.67M USD and the $TVaR_{99.5}$ is 262.18M USD.

It can be observed that the total cost distribution seems to be a normal distribution. This makes sense since the total cost is the sum of the 25,234 random variables that are the single cost of one policy.

The Central limit theorem states that the random variable $\boldsymbol{S_n}$ defined as the sum of n equal and independent random variables $\boldsymbol{X}$ with a mean μ and a variance σ^2, if n is sufficiently large, the total sum will be normal with a mean $n\mu$ and variance $n\sigma^2$ even if $\boldsymbol{X}$ is not normally distributed. This can be summarized as follows:

$$\boldsymbol{S_n} = \boldsymbol{X_1} + \boldsymbol{X_2} + \cdots + \boldsymbol{X_n} = \sum_{i=1}^{n} \boldsymbol{X_n}$$

$$\boldsymbol{S_n} \sim N(n\mu, n\sigma^2)$$

Since the aggregated model is simulating the cost of 25,234 random variables, it can be considered sufficiently large to fit the CLT.

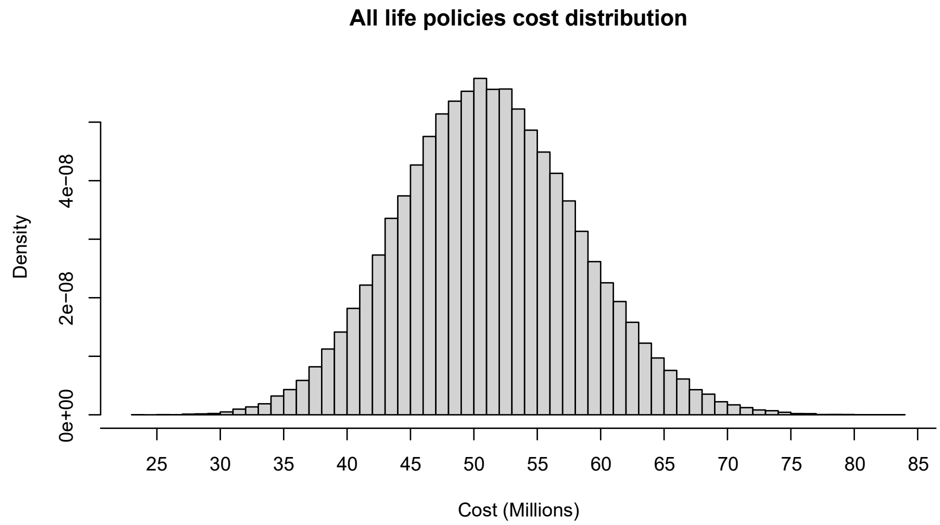
Therefore, the expected cost shall be 25,234 times the expected cost of the policy (10,199 USD), 257,36M USD. The standard deviation $\sqrt{25{,}234}$ times the standard deviation (10,205 USD), 1.62M USD.

5.3. Life policies cost

As previously stated, the age x deaths random variable $\boldsymbol{Z_x}$, follows a binomial distribution of size $n_x = n_i$ (number of policyholders of an age x) and probability $p = q_x$ (probability of dying at an age x).

$$\boldsymbol{Z_x} \sim B(n, q_x)$$

After simulating the number of deaths of all policyholders, the following cost distribution is obtained (code lines 1661 to 1712). Note the insurance amount is 1M USD for all the policyholders.



With an expected cost of 51.64M USD and standard deviation of 7.06M USD. The $VaR_{99.5}$ is 71M USD and the $TVaR_{99.5}$ is 73.98M USD.

The total policies cost random variable $\boldsymbol{Y}$ will be the sum of all the independent random variables. **Considering they are independent**; the total policies cost will be estimated as:

$$\boldsymbol{Y} = \sum_{x=0}^{x_{max}} \boldsymbol{Z_x}$$

With the following moments.

$$E[\boldsymbol{Y}] = E\left[\sum_{x=0}^{x_{max}} \boldsymbol{Z_x}\right] = \sum_{x=0}^{x_{max}} E[\boldsymbol{Z_x}] = \sum_{x=0}^{x_{max}} n_x q_x$$

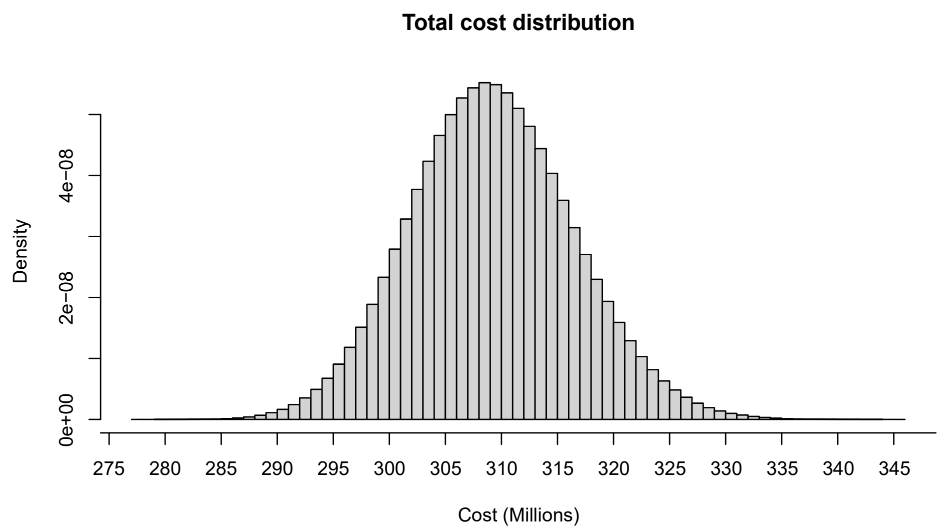
$$Var(\boldsymbol{Y}) = Var\left(\sum_{x=0}^{x_{max}} \boldsymbol{Z_x}\right) = \sum_{x=0}^{x_{max}} Var(\boldsymbol{Z_x}) = \sum_{x=0}^{x_{max}} n_x q_x (1 - q_x)$$

If we compute the moments with the provided data of the policyholders and the mortality rates, the expected cost of all the life policies is 51.60M USD and the standard deviation 7.07M USD.

5.4. Aggregated model of non-life and life cost

For the total cost of all the insurance company insurances, the random variables of the non-life and life cost need to be added. It is considered that the cost variable for each insurance type is independent. Therefore, as previously indicated, under the independency assumption, the expected cost will be the sum of the expected individual costs and the variance will be the sum of the variances.

From the simulation the following distribution is obtained (code lines 1714 to 1744)



The expected cost is 309.11M USD and a standard deviation of 7.24M USD. The $VaR_{99.5}$ is 328.74M USD and the $TVaR_{99.5}$ is 331.27M USD.

Therefore, the economic capital will be the difference between the expected cost and the $VaR_{99.5}$ which is 19.62M USD.

6. Summary

1. For the frequency of the claims modelling the estimated Poisson distribution has been used, with an estimated parameter by the method of moments $\hat{\lambda} = 1.998208$

$$\boldsymbol{N} \sim Pois(\hat{\lambda})$$

2. For the severity (cost) of the claims modelling the estimated Exponential distribution has been used, with an estimated scale parameter by the maximum likelihood method $\hat{\beta} = 5106.187$

$$\boldsymbol{C} \sim Exp(\hat{\beta})$$

3. For the life policies, it has been modelled as a binomial distribution with the parameters given in the mortality table.

$$\boldsymbol{Z_x} \sim B(n, q_x)$$

4. The expected cost of the company by insurance type and total, together with the VaR and TVaR with the 99.5% confidence interval are provided in the following table.

Policies	Expected cost (M USD)	$VaR_{99.5}$ (M USD)	$TVaR_{99.5}$ (M USD)
non-life	257.47	261.67	262.18
life	51.64	71.00	73.98
total	309.11	328.74	331.27

5. The economic capital (expected cost – $VaR_{99.5}$) with a 99.5% confidence interval is 19.63M USD.