Available online at http://www.iorajournal.org/index.php/ijgor/index

International Journal of Global Operations Research

e-ISSN: 2722-1016
p-ISSN: 2723-1739

Vol. 6, No. 1, pp. 18-25, 2025

Collective Risk Model Analysis with Negative Binomial Distribution for Claim Amount and Discrete Uniform Distribution for Claim Amount

Faqih Sholahuddin Al Ayubi[1*]

[1]*University of Padjajaran, Sumedang, Indonesia*

** Corresponding author email: faqih21001@mail.unpad.ac.id*

Abstract

Insurance claims are requests submitted by policyholders to obtain protection against financial losses due to a risk. Individual claims arise whenever there is a risk, while aggregate claims are the number of individual claims in a certain period. These claims are important in managing insurance company expenses, especially for calculating aggregate losses, which are the total losses borne by the company. This study aims to analyze the average and variance of the number of claims distributed Negative Binomial and the amount of claims distributed Discrete Uniform using claim payment data from PT. Jasa Raharja (Persero) Representative of Purwakarta during 2018-2020. The collective risk model is used with the help of Easyfit software to determine the best distribution. The results of the analysis show the number of claims distributed Negative Binomial with an average of 13.2 claims and a variance of 7.6 claims, while the amount of claims distributed Discrete Uniform with an average of IDR2,108,950,000 and a variance of IDR567,850,000. The average aggregate claim is IDR 27,800,000,000 with a variance of IDR 6,270,000,000 during the period. The conclusion of this study confirms the effectiveness of the collective risk model in modeling aggregate claims for insurance data.

Keywords: Negative binomial distribution, discrete uniform distribution, aggregate loss claim model, collective risk model

1. Introduction

Humans in their daily lives are always faced with risks, which arise due to uncertainty in various aspects of life. Risks can cause financial losses, so they require structured handling. Insurance is an effective risk management solution, where individual risks are transferred to insurance companies by paying premiums. Through this mechanism, insurance companies are required to have the ability to predict possible risks and losses that may occur in the future.

Insurance claims are the rights of policyholders to obtain protection against losses that occur, in accordance with the agreed contract. These claims can be modeled using an individual risk or collective risk approach. In the collective risk model, individual claims are summed up to produce an aggregate claim. This aggregate claim is very important for insurance companies because it provides an overview of total losses over a certain period.

Previous research by Yohandoko et al. (2023) showed that the Negative Binomial distribution is suitable for modeling the number of claims, while the Discrete Uniform distribution is often used for the size of claims. This study extends this research by analyzing claim payment data from PT. Jasa Raharja (Persero) Purwakarta Representative during 2018-2020. This study aims to determine the best distribution for the number and size of claims and to calculate the average and variance of aggregate claims.

Through a statistical approach and the use of Easyfit software, this study provides important guidance for insurance companies in understanding claim patterns to help better financial planning. Analysis of claim frequency and severity data helps predict potential future losses and provides relevant recommendations for corporate risk management.

2. Literature Review

Research on the analysis of collective risk models by forming an aggregate claim model has been studied by Yohandoko *et al*. (2023) who studied the analysis of risks borne by insurance companies in an insurance period. The

study used total claim data from the company to form a large distribution model and the number of claims. The results of the study obtained the average and variance of claims during the insurance period.

Based on several studies that have been conducted previously, this study will apply a collective risk model and claim payment data according to all types of insurance and nature of injury in 2018-2020 of PT. Jasa Raharja (Persero) Purwakarta Representative to form a large distribution model and the number of claims which are then used to determine the average and variance of claims in insurance companies during the period 2018-2020.

3. Methodology

3.1 Materials

The data used in this study comes from an article entitled 'Analysis of the Number of Aggregate Claims with Negative Binomial Distribution and the Largest Claims with Discrete Uniform Distribution Using the Convolution Method' published in the Journal of Mathematics: Theory and Applications, Volume 1, No. 2, pages 50-56 in 2019. The data used are recapitulation data on claim payments according to all types of guarantees and the nature of injuries of PT. Jasa Raharja (Persero) Representative of Purwakarta during the period 2018-2020. This data contains the number and amount of claims according to all types of guarantees and the nature of injuries of PT. Jasa Raharja (Persero) Representative of Purwakarta from January 2018 to December 2020.

3.2 Methods

This research is a quantitative research based on calculations and statistical data to determine the relationship with the phenomenon being studied. The methods used in data collection are literature and documentation methods. The documentation method used, by taking data from claim payments based on all types of guarantees and the nature of the injury from 2018 to 2020. The data is secondary and official data that researchers obtained from an article entitled 'Analysis of the Number of Aggregate Claims Distributed Negative Binomial and Large Claims Distributed Discrete Uniform Using the Convolution Method' published in the Journal of Mathematics: Theory and Applications, Volume 1, No. 2, pages 50-56 in 2019. The literature method used in this study is data and information both through written and electronic documents that can support the writing process.

In conducting data analysis, the first step taken is to conduct statistical tests and parameter estimates on the data on the number of claims and the amount of insurance claims of PT. Jasa Raharja (Persero) Representative of Purwakarta based on all types of guarantees and the nature of injuries in 2018-2020 to determine the candidate distribution and its parameters for the number and amount of claims. The second step is to conduct the Kolmogorov-Smirnov test to determine the distribution that fits the data. After obtaining the best distribution for the number and amount of claims, the next step is to calculate the average number and amount of claims, as well as the variance of the number and amount of claims. The last step is to carry out calculations to obtain the average and variance of the occurrence of aggregate claims.

3.2.1 Formula / Equation

The negative binomial distribution is a repeated experiment that runs continuously until a certain number of successes occur (Sudaryono, 2012). The negative binomial distribution is denoted by $\mathrm{b}*(\mathrm{x};\mathrm{n},\mathrm{p})$. According to Maharani *et al*. (2025), If X represents the number of repetitions of the experiment that result in x successes, then the probability of success in the nth repetition preceded by $\mathrm{n}-1$ successes and $\mathrm{n}-\mathrm{x}$ failures, the distribution of the random variable X is the number of repetitions until x successes occur. However, because each repeater is independent of each other, it needs to be multiplied by all the probabilities p and failures $\mathrm{q}=1-\mathrm{p}$.

According to Maharani et al. (2019), if a negative binomial experiment has a probability of success p and a probability of failure q, the probability distribution of the random variable X is the number of repetitions until x successes occur, then mathematically the probability for the negative binomial distribution is as follows:

$$p(x) = P(X = x) = \binom{x-1}{n-1} p^n q^{x-n} \quad (1)$$

with:

$p(x)$: negative binomial distribution probability with random variables$X$and many repetitions of the experiment until the time$n$,

x : many attempts to get it$p$success to- n,

n : number of successful events that occur,

p : the chance of a successful event occurring,

q : the chance of a failed event occurring.

The formula for the mean and variance of the negative binomial distribution is as follows:

$$E(x) = \mu = \frac{n}{p} \tag{2}$$

$$Var(x) = \frac{n(1-p)}{p^2} \tag{3}$$

Discrete Uniform Distribution

A random variable X that has a discrete uniform distribution (X~Discrete Uniform(a, b)), indicates that the variable X has a discrete uniform distribution with integer parameters a and b, where a < b. The probability density function (fkp) of the discrete uniform distribution is as follows:

$$f(x) = \frac{1}{b-a+1} \qquad x = a, a+1, \dots, b. \tag{4}$$

with:

$f(x)$: the chance if the random variable X distributed discrete uniform,
a : first order number,
b : last sequence number.

Mean and variance formulas of a distribution discrete uniform are as follows:

$$E(x) = \mu = \frac{a+b}{2} \tag{5}$$

$$Var(x) = \frac{(b-a+1)^2 - 1}{12} \tag{6}$$

Model *Aggregate Loss*

According to Melantika (2023), there are two approaches to modeling aggregate loss, namely the individual risk model and the collective risk model. The individual risk model emphasizes the losses from each individual contract and states Saggregate loss as follows:

$$S_n = X_1 + X_2 + \cdots + X_n \tag{7}$$

with:

X_i $(i = 1,2, \dots, n)$: the amount of loss from n contracts, at individual risks assumed to be independent,
n : number of contracts.

According to Maharani (2025), to obtain aggregate loss, it is done by recording each claim amount and adding up all the claims. Aggregate loss is expressed by a random variable S and the number of claims in one period in a portfolio is expressed by N. The amount of each claim can be expressed in random variables $X_1, X_2, \dots$ so that the collective risk model is obtained as follows:

$$S = X_1 + X_2 + \cdots + X_N \qquad N = 0,1,2, \dots \tag{8}$$

According to Klugman *et al.* (2004), assumptions that must be taken into account in aggregate loss for the collective risk model are as follows:

a. Given $N = n$ random variables $X_1, X_2, \dots, X_N$ are random variables that are distributed identically and independently,
b. Given $N = n$ joint distribution of random variables $X_1, X_2, \dots, X_N$ does not depend on the value of n,
c. The distribution of N random variables does not depend on the values of random variables $X_1, X_2, \dots, X_N$.

The average occurrence of aggregate claims for the collective risk model is given by:

$$E[S] = E[N]E[X] \tag{9}$$

The variance in the occurrence of aggregate claims in the collective risk model is a conditional variance, namely:

$$Var[S] = E[Var(X|I)] + Var[E(X|I)] \tag{10}$$

If frequency and severity are independent, then the compound variance is:

$$Var[S] = E[N]Var[X] + Var[N]E[X]^2 \tag{11}$$

4. Results and Discussion

This study uses total insurance claim data based on all types of guarantees and the nature of injuries of PT. Jasa Raharja (Persero) Representative of Purwakarta in 2018-2020. The number of claims or frequency data shows the number of claim incidents and the severity or amount of claims data shows the amount of claim payments made by

PT. Jasa Rahaja (Persero) Representative of Purwakarta as an insurance company. Visualization of frequency and severity data can be seen in Figures 1 and 2.

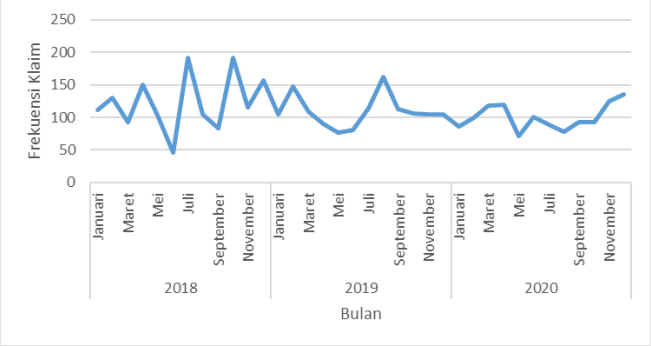

Figure 1: Number of Insurance Claims of PT. Jasa Raharja (2018-2020)

Figure 1 shows a graph of the number of claims or insurance frequency data of PT. Jasa Raharja Representative Office of Purwakarta which experienced an increase and decrease in the number of claims each month during 2018-2020.

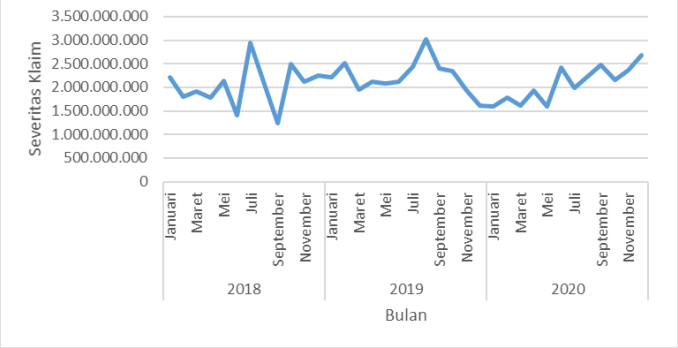

Figure 2: Amount of Insurance Claims of PT. Jasa Raharja (2018-2020)

Figure 2 shows a graph of the large insurance claims or severity data of PT. Jasa Raharja (Persero) in 2018-2020 which experienced an increase and decrease in the large claims each month. Then, parameter estimation and statistical tests were carried out with the help of Easyfit software to determine the distribution of the number and amount of claims that correspond to the insurance data of PT. Jasa Raharja (Persero) Purwakarta Representative which can be seen in tables 1 and 2.

Table 1: Estimation of Candidate Parameters for Claims Amount Distribution

No	Distribution	Parameter
1	D. Uniform	$a = 57, \quad b = 164$
2	Geometric	$p = 0.00894$
3	Logarithmic	$\theta = 0.99863$
4	Negative Binomial	$n = 14, \quad p = 0.11452$
5	Poisson	$\lambda = 110.81$
6	Bernoulli	No fit (max data > 1)
7	Binomial	No fit
8	Hypergeometric	No fit

Based on table 1 which shows the parameter estimates of candidate distributions from the data on the number of insurance claims of PT. Jasa Raharja (Persero) Representative of Purwakarta, five candidate distributions were obtained that had parameter values from all eight distributions obtained. The five candidate distributions are D. Uniform, Geometric, Logarithmic, Negative Binomial, and Poisson. Meanwhile, the other three distributions that showed No Fit, namely the Bernoulli, Binomial, and Hypergeometric distributions showed that there was no match with the data on the number of claims. Then Kolmogorov-Smirnov was carried out to find out whether the data came from a certain population distribution or not and Table 2 was obtained below.

Table 2: Statistical Value of Prospective Distribution of Number of Claims

No	Distribution	Kolmogorov-Smirnov	
		Statistics	Ranking
1	D. Uniform	0.15741	2
2	Geometric	0.44854	4
3	Logarithmic	0.69274	5
4	Negative Binomial	0.09945	1
5	Poisson	0.28622	3
6	Bernoulli	No fit (max data>1)	
7	Binomial	No fit	
8	Hypergeometric	No fit	

Table 2 shows the ranking based on the Kolmogorov-Smirnov test, namely Negative Binomial, D. Uniform, Poisson, Geometric, and Logarithmic, while the other three distributions show a mismatch. Then the p-value calculation is carried out to determine the best distribution for the number of claims data and the results are obtained as in Table 3.

Table 3: P-Value of Candidate Distribution of Number of Claims

No	Distribution	P-Value	Ranking
1	Negative Binomial	0.83423	1
2	D. Uniform	0.30177	2
3	Poisson	0.00419	3
4	Geometric	5.1659E-7	4
5	Logarithmic	3.8668E-16	5

Then the Kolmogorov-Smirnov hypothesis test was carried out with,

Hypothesis: H_0: Data comes from a certain population distribution
H_1: Data does not come from a certain population distribution.
Real level: $\alpha = 5\%$
Test statistic: Kolmogorov-Smirnov test
Test criteria : $p - value > \alpha = 0.05$ then H_0 is accepted (failed to reject H_0)
Based on value $p - value$ in table 3, the results of the Kolmogorov-Smirnov test are as in Table 4 below:

Table 4: Kolmogorov-Smirnov Test Results for Prospective Distribution of Claim Amounts

No	Distribution	P-Value	Test Results
1	Negative Binomial	0.83423	H_0accepted
2	D. Uniform	0.30177	H_0accepted
3	Poisson	0.00419	H_0rejected
4	Geometric	5.1659E-7	H_0rejected
5	Logarithmic	3.8668E-16	H_0rejected

The results of the Kolmogorov-Smirnov test above show that the hypothesis decision for Negative Binomial and D.Uniform is H_0 accepted or failed to reject H_0 because the $p - value$ obtained has a $p - value > \alpha$, which is 0.05 which means the data comes from a certain population. Meanwhile, the candidate distributions Poisson, Geometric, and Logarithmic have so they reject H_0 which means the data does not come from a certain distributed population. Based on the largest $p - value$ and the ranking order of the two candidate distributions that have been obtained previously, the Negative Binomial distribution is the best distribution in modeling the number of insurance claims of PT. Jasa Raharja (Persero) Representative of Purwakarta.

Next, parameter estimation and statistical tests will be carried out for severity data or large claim data as done on frequency data or claim amount data. The results of parameter estimation for severity data can be seen in Table 5.

Table 5: Estimation of Candidate Parameters for Large Claim Distributions

No	Distribution	Parameter
1	D. Uniform	a =1.4334E+9 2.7845E+9b =
2	Geometric	p =4.7418E-10
3	Logarithmic	θ = 1,0
4	Negative Binomial	n = 29, p =1.3863E-8
5	Poisson	λ =2.1089E+9
6	Bernoulli	No fit (max data > 1)
7	Binomial	No fit
8	Hypergeometric	No fit

Based on table 5 which shows the parameter estimates of candidate distributions from the large insurance claim data of PT. Jasa Raharja (Persero) Representative of Purwakarta, five candidate distributions were obtained that had parameter values from all eight distributions obtained. The five candidate distributions are D. Uniform, Geometric, Logarithmic, Negative Binomial, and Poisson, while the other three distributions that showed No Fit, namely Bernoulli, Binomial, and Hypergeometric distributions showed that they did not match the large claim data. Then Kolmogorov-Smirnov was carried out to find out whether the data came from a certain distributed population or not and Table 6 was obtained below.

Table 6: Statistical Value of Candidate Large Claim Distribution

No	Distribution	Kolmogorov-Smirnov	
		Statistics	Ranking
1	D. Uniform	0.11204	1
2	Geometric	0.47433	2
3	Logarithmic	N/A	N/A
4	Negative Binomial	N/A	N/A
5	Poisson	0.55556	3
6	Bernoulli	No fit (max data>1)	
7	Binomial	No fit	
8	Hypergeometric	No fit	

Table 6 shows that there are three candidate distributions that have statistical values, while the other five candidate distributions do not have statistical values. Based on the Kolmogorov-Smirnov test ranking order, the candidate distribution D. Uniform is ranked first, Geometric is ranked second, and Poisson is ranked third, while Logarithmic and Negative Binomial have N/A results which mean they have no value. Meanwhile, the Bernoulli, Binomial, and Hypergeometric distributions do not fit the climate big data or No Fit. Then the $p - value$ calculation is carried out to determine the best distribution for the big claim data and the results are obtained as in Table 7.

Table 7: P-Value of Candidate Large Claim Distribution

No	Distribution	P-Value	Ranking
1	D. Uniform	0.71502	1
2	Geometric	8.5852E-8	2
3	Poisson	1.5654E-10	3

Then the Kolmogorov-Smirnov hypothesis test was carried out with,

Hypothesis :H_0: Data comes from a population with a certain distribution
H_1:The data does not come from a specific distributed population.
Real level :$\alpha = 5\%$
Test statistics : Kolmogorov-Smirnov test
Test criteria : $p - value > \alpha = 0{,}05$then accepted (failed to reject $H_0 H_0$)

Based on the $p - value$ in table 3, the results of the Kolmogorov-Smirnov test are as in Table 8 below:

Table 8: Kolmogorov-Smirnov Test Results for Large Claim Distribution Candidates

No	Distribution	P-Value	Test Results
1	D. Uniform	0.71502	H_0 accepted
2	Geometric	8.5852E-8	H_0 rejected
3	Poisson	1.5654E-10	H_0 rejected

The results of the Kolmogorov-Smirnov test above show that the hypothesis decision for D.Uniform is H_0 accepted or fails to reject H_0 because the p-value obtained has a $p - value > \alpha$, which is 0.05, which means that the data comes from a certain population. Meanwhile, the Geometric and Poisson distribution candidates have $\mathrm{p-values} < \alpha$ so that they reject H_0, which means that the data does not come from a certain distributed population. Based on the largest $p - value$ and the ranking order that has been obtained previously, the D. Uniform distribution is the best distribution in modeling the amount of insurance claims of PT. Jasa Raharja (Persero) Representative of Purwakarta.

After obtaining the best distribution for frequency data or number of claims, namely the Negative Binomial distribution, and the best distribution for severity data or large claim data, namely the D. Uniform distribution, the average number of claims, average claim amount, claim amount variance, and claim amount variance can be calculated.

The number of claims is distributed Negative Binomial, so the average (mean) is given by equation (2) and the variance is given by equation (3). With the parameters obtained from the parameter estimates above, namely $n = 14$ and , $p = 0.11452$, so the average and variance of the number of claims can be written as follows:

$$E[N] = \frac{14}{0{,}11452} = 122{,}2493888$$

$$Var[N] = \frac{14(1 - 0{,}11452)}{0{,}11452^2} = 945{,}2444006$$

The claim size is distributed D. Uniformly, then the average (mean) is given by equation (5) and the variance is given by equation (6). With the parameters obtained from the parameter estimates above, namely $\mathrm{a} = 1.4334\mathrm{E} + 9 = 1.4334 \times 10^9$ and $\mathrm{b} = 2.7845\mathrm{E} + 9 = 2.7845 \times 10^9$, so the average and variance of the claim size can be written as follows:

$$E[X] = \frac{(1{,}4334 \times 10^9) + (2{,}7845 \times 10^9)}{2} = 2108950000$$

$$Var[X] = \frac{(2{,}7845 \times 10^9 - 1{,}4334 \times 10^9 + 1)^2 - 1}{12} = 1{,}52123 \times 10^{17}$$

After obtaining the average and variance of the number and amount of claims, the average and variance of the occurrence of aggregate claims during the period can be calculated. In the collective risk model, the average occurrence of aggregate claims is given by equation (9) and the variance of the occurrence of aggregate claims is given by equation (11). Thus, the results of the average value and variance of the occurrence of claims in claim payments according to all types of guarantees and the nature of injuries of PT. Jasa Raharja (Persero) Representative of Purwakarta during the period 2018-2020 are as follows:

$$E[S] = 122{,}2493888 \times 2108950000 = 2{,}57818 \times 10^{11}$$

$$Var[S] = 122{,}2493888 \times 1{,}52123 \times 10^{17} + 945{,}2444006 \times (2108950000)^2$$

$$Var[S] = 4{,}22273 \times 10^{21}$$

Based on this value, the average occurrence of aggregate claims is IDR with a variance of IDR $2{,}57818 \times 10^{11} 4{,}22273 \times 10^{21}$

5. Conclusion

Based on the above processing, it is obtained that the frequency data or number of claims is distributed Negative Binomial and the severity data or amount of claims is distributed D. Uniform. Based on the data calculation, the

average occurrence of aggregate claims is IDR 2.57818×10^{11} with a variance of IDR 4.22273×10^{21} during the 2018-2020 insurance period.

References

Kartikasari, M., D. (2017). Premium Pricing of Liability Insurance Using Random Sum Model, *17*(1), 46-54.Kalfin, Sukono, Sudradjat Supian, and Mustafa Mamat. (2023). Model for Determining Insurance Premiums Taking into Account the Rate of Economic Growth and Cross-Subsidies in Providing Natural Disaster Management Funds in Indonesia. *Sustainability*, 15(24): 16655. https://doi.org/10.3390/su152416655

Maharani, Asthie. (2025). Analysis of Aggregated Claim Numbers with Geometric Distribution and Claim Sizes with Weibull Distribution Using Convolution Method. International Journal of Mathematics, Statistics, and Computing 3 (February):12–20. https://doi.org/10.46336/ijmsc.v3i1.178.

Meyers, Glenn, and John A. Beekman. (1987). An Improvement to the Convolution Method of Calculating ψ(u). *Insurance: Mathematics and Economics 6* (4): 267–74. https://doi.org/10.1016/0167-6687(87)90031-X.

Oktavia, Rini, Rahma Zuhra, Hafnani Hafnani, Nurmaulidar Nurmaulidar, and Intan Syahrini. (2023). Application of Poisson and Negative Binomials Models to Estimate the Frequency of Insurance Claims. *Jurnal Natural, 23*(1): 21–27. https://doi.org/10.24815/jn.v23i1.26623.

Yohandoko, Setyo, Agung Prabowo, Usman Yakubu, and Chun Wang. (2023). Life Insurance Aggregate Claims Distribution Model Estimation. Operations Research: International Conference Series 4 (December):117–25. https://doi.org/10.47194/orics.v4i4.271.