IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 5, Ser. 1 (Sept. – October. 2023), PP 21-27
www.iosrjournals.org

Modeling Third Party Liability Insurance Claim Severity Using Generalized Inverse-Lindley Distribution

Syahla Anisah[1], Aceng Komarudin Mutaqin[2]

[1]*(Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Bandung, Indonesia)*

[2]*(Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Bandung, Indonesia)*

Abstract:
In this paper, we will discuss modeling of severity on motor vehicle insurance claims for third party liability guarantees with comprehensive coverage at PT. The 2019 X uses the Generalized Inverse-Lindley distribution. The Generalized Inverse-Lindley distribution is a non-composite distribution with α and λ parameter. The parameters in the Generalized Inverse-Lindley distribution are estimated through the maximum likelihood estimator method using the Newton-Raphson iteration numerical method. Distribution fit testing is carried out using the Kolmogorov-Smirnov test. The data used is secondary data from the recording results of the insurance company PT. X in 2019 is severity on Third Party Liability insurance claims with comprehensive coverage. Based on the results of the application of the Generalized Inverse-Lindley distribution on severity of Third Party Liability motor vehicle insurance claims with comprehensive coverage at PT. X in 2019 in category 3 using the Kolmogorov-Smirnov test can be concluded that the severity of claims in regions 1, 2, and 3 comes from the population distributed Generalized Inverse-Lindley.

Key Word*:Motor Vehicle Insurance, Third Party Liability,* Generalized Inverse-Lindley *Distribution,* Newton-Raphson, Kolmogorov-Smirnov.

Date of Submission: 01-09-2023 Date of Acceptance: 11-09-2023

I. Introduction

Basically, human life is full of uncertainties. Uncertainty arises from the inability of humans to predict the possibilities that will occur in the future. Unvertainty that will come inevitably makes people themselves feel insecure. The insecurity caused by uncertainty is often referred to as risk. Risk is the danger, outcome, or consequence that may occur as a result of ongoing processes or future events [1]. Human activities can pose risks that endanger the safety of body and soul and even property. Vehicles traveling on the highway can be damaged by accidents that can result in losses for drivers and motorized vehicles. The importance of a guarantee through insurance for a driver, in order to minimize losses. This loss can be transferred to the insurer or insurance company. Ismanto [2] said that the types of insurance business fields in Indonesia include liability with one of its insurance products, namely motor vehicle insurance. In general, there are two types of motor vehicle insurance coverage, namely *comprehensive* (*All Risk*) and *total loss only* (TLO). Motor vehicle insurance can also be supplemented with additional guarantees, such as third-party insurance that covers risks arising from compensation claims from third parties. This type of insurance is better known as Legal Liability Against Third Parties or *third party liability* (TPL). So, if the third party is involved in an accident and demands liability, either for themselves or for other, the policyholder can file a claim. In loss modeling , there are two important measures that must be considered, one of which is the severity claims submitted by customers to insurance companies[3]. Severity claim in loss insurance is a positive random variable, which tends to follow a skewed distribution with a thicker right tail. The process of modeling severity insurance claim certainly involves a variety of distributions. Based on various studies, distributions with thicker right tails have been used in loss insurance, including the composite distribution, which is distributions determined by the probability density [4]. As research progresses, non-composite distributions have emerged as a approach to data modeling that may be particularly useful for practitioners dealing with large claims values.

Asgharzadeh et al (2017) discuss modeling severity claims data using a non-composite distribution, namely theGeneralized Inverse-Lindley (GIL) distribution. The results of research conducted by Asgharzadeh et al show that the GIL distribution is better than several composite distributions such as the lognormal distribution, Weibull distribution, Pareto distribution and several other distributions, because it has the smallest Akaike Information Criterion (AIC) value, the smallest Bayesian Information Criterion (BIC), the largest p-value of the Kolmogorov-Smirnov test and the largest p-value of the Cramer-von Mises test[5].

Based on research by Asgharzadeh et al (2017), in this paper will be modeled on severity on motor vehicle insurance claims for Third Party Liability coverage with comprehensive coverage at PT. X in 2019 used the Generalized Inverse-Lindley (GIL) distribution using the Kolmogorov-Smirnov test in evaluating the suitability of its distribution.

II. Material And Methods

The data in this paper is secondary data from the records of the general insurance company PT. X for the 2019 policy. The data includes largeseverity data on motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 for category 3 (sum insured IDR 200,000,000.00 to IDR 400,000,000.00) and for all regions (region 1, region 2, and region 3).

Lindley Distribution

The Lindley distribution was first introduced by Lindley[6]in the context of Bayes' statistics. The Lindley distribution with parameter λ is defined by the following probability density function:

$$f(x;\ \lambda) = \frac{\lambda^2}{\lambda+1}(1+x)e^{-\lambda x}, \quad x > 0, \lambda > 0 \qquad \dots (1)$$

The cumulative distribution function of the Lindley distribution is as follows:

$$F(x;\ \lambda) = 1 - \frac{\lambda+1+\lambda x}{\lambda+1}e^{-\lambda x}, \quad x > 0, \lambda > 0 \qquad \dots (2)$$

Generalized Inverse-Lindley Distribution

Let Y denote a one-parameter Lindley-distributed random variable and $X = Y^{-\frac{1}{\alpha}}$. Then the probability density function of X is as follows:

$$f(x; \alpha, \lambda) = \frac{\alpha\lambda^2}{\lambda+1}(1+x^{-\alpha})x^{-\alpha-1}e^{-\lambda x^{-\alpha}}, \ x > 0, \quad \alpha, \lambda > 0 \qquad \dots (3)$$

where λ is the scale parameter and α is the shape parameter.

The cumulative distribution function of the Generalized Inverse-Lindley distribution is as follows:

$$F(x; \alpha, \lambda) = \left[\frac{1+\lambda+\lambda x^{-\alpha}}{1+\lambda}\right]e^{-\lambda x^{-\alpha}}, \quad x > 0, \quad \alpha, \lambda > 0 \qquad \dots (4)$$

The GIL distribution was proposed by Sharma[7] which offers more flexibility to model extreme values or outside observations in the right tail direction. It also shows that the distribution has a heavy tail, that is, a polynomial tail or wider tail that decays much more slowly and is larger than the tail of the normal distribution. Heavy-tailed distributions are used to measure the probability of disorder or extreme events and are therefore very useful in modeling risk analysis, such as insurance data.

Parameter Estimation of the Generalized Inverse-Lindley Distribution

The Parameters of the generalized inverse-Lindley distribution can be estimated using the maximum likelihood method. The maximum likelihood method is used to obtain parameter estimates by maximizing the log-likelihood function. The likelihood function based on the observed sample $x = \{x_1, x_2, \dots, x_n\}$ is as follows:

$$L(\alpha, \lambda) = \prod_{i=1}^{n} f(x; \alpha, \lambda) = \frac{\alpha^n \lambda^{2n}}{(1+\lambda)^n}\prod_{i=1}^{n}(1+x_i^{-\alpha})\prod_{i=1}^{n}x_i^{-(\alpha+1)}\, e^{-\lambda\sum_{i=1}^{n}x_i^{-\alpha}} \qquad \dots (5)$$

The *log-likelihood* function for the GIL distribution is as follows:

$$l(\alpha, \lambda) = ln\big(L(\alpha, \lambda)\big)$$

$$= n\ln(\alpha) + 2n\ln(\lambda) - n\ln(1+\lambda) + \sum_{i=1}^{n}\ln(1+x_i^{-\alpha}) - (\alpha+1)\sum_{i=1}^{n}\ln(x_i) - \lambda\sum_{i=1}^{n}x_i^{-\alpha} \qquad \dots (6)$$

The maximum likelihood estimator of the GIL distribution parameter is the solution of the first derivative of the *log-likelihood* function to which its parameters are equated to zero, namely:

$$\frac{\partial l(\alpha,\lambda)}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \frac{x_i^{-\alpha} \ln(x_i)}{1 + x_i^{-\alpha}} - \sum_{i=1}^{n} \ln(x_i) + \lambda \sum_{i=1}^{n} x_i^{-\alpha} \ln(x_i) = 0 \quad \dots (7)$$

$$\frac{\partial l(\alpha,\lambda)}{\partial \lambda} = \frac{2n}{\lambda} - \frac{n}{1+\lambda} - \sum_{i=1}^{n} x_i^{-\alpha} = 0 \quad \dots (8)$$

Based on the above equation, it can be seen that the estimated parameter cannot be solved analytically. For cases like this, numerical calculation methods are usually used to obtain the estimated parameters. One of the numerical methods that can be used is the Newton-Raphson iteration method. The Newton-Raphson iteration method requires the initial value of the parameter (α, λ), the first derivative of the log-likelihood function with respect to its parameters and the second derivative of the log-likelihood function with respect to its parameter. The second derivative of the log-likelihood function is given by:

$$\frac{\partial^2 l(\alpha,\lambda)}{\partial \alpha^2} = -\frac{n}{\alpha^2} + \sum_{i=1}^{n} \frac{x_i^{-\alpha} \ln^2(x_i)}{(1 + x_i^{-\alpha})^2} - \lambda \sum_{i=1}^{n} x_i^{-\alpha} \ln^2(x_i) \quad \dots (9)$$

$$\frac{\partial^2 l(\alpha,\lambda)}{\partial \lambda^2} = -\frac{2n}{\lambda^2} + \frac{n}{(1+\lambda)^2} = \frac{-n\lambda^2 - 4n\lambda - 2n}{\lambda^2(\lambda+1)^2} \quad \dots (10)$$

$$\frac{\partial^2 l(\alpha,\lambda)}{\partial \alpha \partial \lambda} = \sum_{i=1}^{n} x_i^{-\alpha} \ln(x_i) \quad \dots (11)$$

Suppose $\boldsymbol{\gamma} = (\alpha, \lambda)^T$ is the parameter vector. At the next iteration $(h+1)$, the updated parameters can be obtained [8]:

$$\boldsymbol{\gamma}^{(h+1)} = \boldsymbol{\gamma}^{(h)} - \left[\left. \frac{\partial^2 l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(h)}} \right]^{-1} \left[\left. \frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(h)}} \right], \qquad h = 0, 1, 2, .. \quad \dots (12)$$

with

$$\Delta = \left\| \boldsymbol{\gamma}^{(h+1)} - \boldsymbol{\gamma}^{(h)} \right\|$$

The iteration process in equation (12) is stopped if the value of $\left\| \boldsymbol{\gamma}^{(h+1)} - \boldsymbol{\gamma}^{(h)} \right\| < \varepsilon$. The value of ε is the error tolerance, e.g. $\varepsilon = 1 \text{ x } 10^{-6}$. The iteration equation in equation (12) requires initial values for α and λ. The way to obtain the initial value is to simplify equation (7) into the following:

$$\lambda = \frac{-\frac{n}{\alpha} + \sum_{i=1}^{n} \frac{\ln(x_i)}{1+x_i^{\alpha}} + \sum_{i=1}^{n} \ln(x_i)}{\sum_{i=1}^{n} x_i^{-\alpha} \ln(x_i)} \quad \dots (13)$$

Next, equation (13) is subsumed into equation (8) to produce the following equation:

$$\frac{2n \sum_{i=1}^{n} x_i^{-\alpha} \ln(x_i)}{-\frac{n}{\alpha} + \sum_{i=1}^{n} \frac{\ln(x_i)}{1+x_i^{\alpha}} + \sum_{i=1}^{n} \ln(x_i)} - \frac{n \sum_{i=1}^{n} x_i^{-\alpha} \ln(x_i)}{-\frac{n}{\alpha} + \sum_{i=1}^{n} \frac{\ln(x_i)}{1+x_i^{\alpha}} + \sum_{i=1}^{n} \ln(x_i) + \sum_{i=1}^{n} x_i^{-\alpha} \ln(x_i)} - \sum_{i=1}^{n} x_i^{-\alpha} = 0 \quad \dots (14)$$

The initial value of the estimated parameter α can be obtained from equation (14) by trial and error. α that satisfies it. While the initial value of the estimated parameter λ can be obtained by subsuming the value of α to equation (13). Note that if a pair of initial values is found α and λ more than one, then select the pair α and λ that maximizes the log-likelihood function the most.

Kolmogorov-Smirnov Test

A distribution fit tests is used to show that a data set can be said to follow a certain distribution. One of the fit tests that can be used to determine a particular distribution is the *Kolmogorov-Smirnov* test. Suppose the realization of a random sample is $x_1, x_2, \ldots, x_n$ which corresponds to an unknown distribution function $F(x)$ and suppose $F^*(x)$ is a complete hypothesized distribution function. The hypothesis formulation for the *Kolmogorov-Smirnov* test is [9]:

H_0: Data comes from a population with a certain distribution
H_1: Data does not come from a population with a certain distribution
The statistics *of the Kolmogorov-Smirnov* test for the above hypothesis is:

$$D = \max_{1 \le i \le n} |F_n(x_i) - F^*(x_i)| \qquad \ldots (15)$$

where:

$$F_n(x_i) = \frac{\text{the number of observations } \le x_i}{n} \qquad \ldots (16)$$

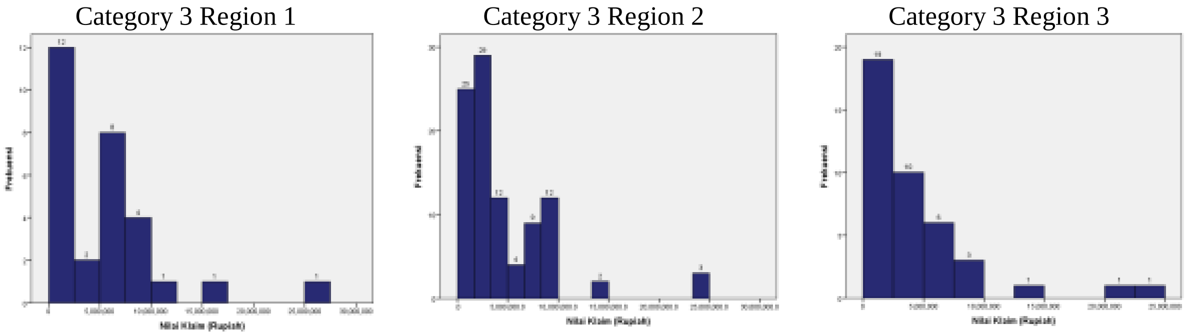
Description:
$F_n(x_i)$: empirical distribution function for the observation data i.
$F^*(x_i)$: cumulative distribution function of the model under test for the the observation data. i

Test criteria ff the test statistics D is smaller than the critical value at the real level α that has been determined, then the null hypothesis is accepted and it is concluded that the data comes from a certain distributed population. For $n > 40$ the critical values is based on asymptotic distribution

III. Result

Data Description

Data descriptions is used to provide an overview of the data distribution. Histograms are used in describing data. Figure 4.1 presents a histogram form of the severity of motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 for category 3 and all regions.



Based on Figure 1, it can be seen that the severity claim data for category 3 has a distribution of the data distribution is skewed to the right, because the tail of the distribution is on the right. Therefore, the GIL distribution is thought to be suitable for modeling the severity of motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 category 3 and all regions.

Parameter Estimation of the Generalized Inverse-Lindley Distribution

The estimation of GIL distribution parameters is done through maximum likelihood estimation using the Newton-Raphson method with the help of RStudio software. The results of pair values α and λ those accompanied by their log-likelihood functions for category 3 and all regions are presented in Table 1.

Table no 1: Parameter Pairs **α** dan **λ** with Log-likelihood Values in Category 3 and All Regions

Category	Region	Initial Value α	Initial Value λ	Log-likelihood Function Value
3	1	0,90	3,9917 x10^5	-480,8492
3	2	0,92	5,7824 x10^5	-1.577,9614
3	3	0,92	5,2857 x10^5	-670.6018

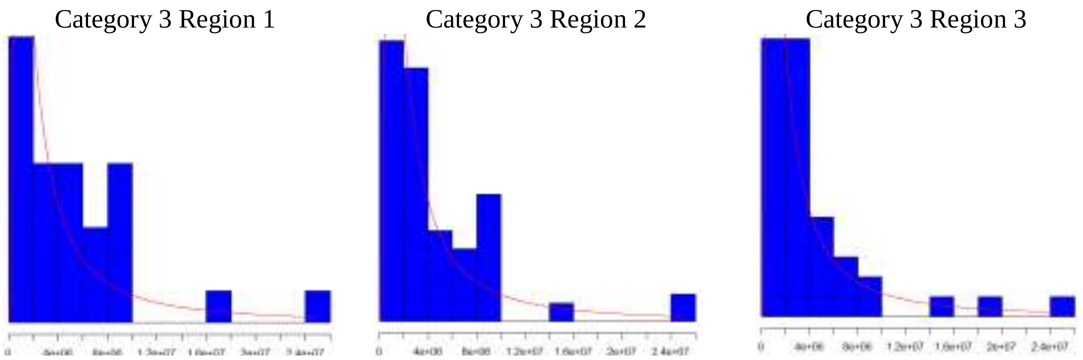
The pair of initial value that have been obtained are entered into the iteration equation in equation (12). The estimated values used $\varepsilon = 1 \text{ x } 10^{-6}$ the results of the Newton-Raphson iteration are presented in Table 2. Table 2 contains pairs of estimated values of the and parameters α and λ, andthe number of iterations.

Table no 2:
Newton-Raphson Iteration Results for Category 3 and All Regions

Category	Region	Estimated Value α	Estimated Value λ	Number of Iterations
3	1	0,8809299	305.923,1000	8
3	2	0,9508589	894.964,5000	2
3	3	0,9373648	673.294,4000	2

Based on Table 2, the estimated parameters α and λ for severity on motor vehicle insurance insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 for category 3 and all regions.

Figure 2 shows the histogram for severity on motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 category 3 with all regions and the GIL distribution density function curve with estimated parameters that have been obtained in Table 4.2.



Based on Figure 2, it can be seen that category 3 data with all regions has a density function curve shape following the histogram pattern.

Generalized Inverse-Lindley Distribution Fit Testing

GIL distribution suitability test on severity of motor vehicle insurance claims forTPL coverage with comprehensive coverage at PT. X in 2019 category 3 with all regions using the Kolmogorov-Smirnov test. The hypotheses for such testing are:

H_0: Severity on motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 category 3 contained in the region t comes from a GIL distributed population, where $t = 1,2,3$.

H_1: Severity on motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 category 3 contained in the region t comes from a GIL distributed population, where $t = 1,2,3$.

The calculation of the Kolmogorov-Smirnov test statistic is obtained using equation (15). The estimated parameter values obtained in Table 2 are used to obtain the estimated cumulative distribution function of the GIL distribution using equation (4) and the empirical distribution function value using equation (16). For example, for the first order data in category 3 and region 1 with the number of claims in category 3 region 1 as many as 29 claims, $x_1 = 300.000$with the help of Microsoft Excel software, the empirical distribution function values is obtained:

$$
\begin{aligned}
F_{29}(x_1) &= F_{29}(300.000) \\
&= \frac{\text{the number of observations } \leq 300.000}{29} \\
&= \frac{1}{29} \\
&= 0{,}0345
\end{aligned}
$$

and the cumulative distribution function value of the GIL distribution is:

$$
\begin{aligned}
F^*(x_1) &= F^*(300.000) \\
&= \left[\frac{1 + 305.923{,}1 + (305.923{,}1)(300.000)^{-0{,}8809299}}{(1 + 305.923{,}1)}\right] e^{-(305.923{,}1)(300.000)^{-0{,}8809299}} = 0{,}0103
\end{aligned}
$$

Estimation of the value of the empirical distribution function and the cumulative distribution function of the GIL distribution for other data is performed using the same method as in the first category 3 of region 1. The results of calculations with the help of Microsoft Excel *software* for each data are presented in Table 3 with column (1) is the severity of motor vehicle insurance claims for TPL coverage with *comprehensive* coverage at PT. X in 2019 category 3 and all regions that have been sorted from smallest to largest and taken unique values only, column (2) is the value of the empirical distribution function of the claim severity, column (3) is the cumulative distribution function of the GIL distribution , and column (4) is the absolute value of the value of the empirical function reduced by the value of the cumulative distribution function of the GIL distribution.

Table no *3:*
Kolmogorov-Smirnov Test Calculation Results on Claim Size Data Category 3 Region 1

No. Tertanggung	x_i	$F_n(x_i)$	$F^*(x_i)$	$\|F_n(x_i) - F^*(x_i)\|$
(1)	(2)	(3)	(4)	(5)
1	300.000	0,0345	0,0103	0,0242
2	385.000	0,0690	0,0254	0,0436
3	685.000	0,1034	0,1095	0,0060
⋮	⋮	⋮	⋮	⋮
27	10.000.000	0,9310	0,8118	0,1192
28	16.800.000	0,9655	0,8763	0,0892
29	25.000.000	1.0000	0,9112	0,0888

Based on the results in Table 3, the Kolmogorov-Smirnov test statistical value for category 3 region 1 data using equation (15) is obtained, namely:
$D = {}_{1\leq i\leq n}^{max}|F_n(x_i) - F^*(x_i)| = 0{,}1639$

The value of the statistics for the fit to the Generalized Inverse-Lindley distribution for the other data was performed using the same method as in category 3 region 1 data. The complete results of the statistical values of the test are presented in Table 4 column (3). Using the real level $\alpha = 0{,}05$, the critical value of the Kolmogorov-Smirnov test of category 3 and all regions were obtained. The complete results are presented in Table 4 column (4). The conclusions of the Generalized Inverse-Lindley distributed distribution fit test are presented in Table 4 column (5).

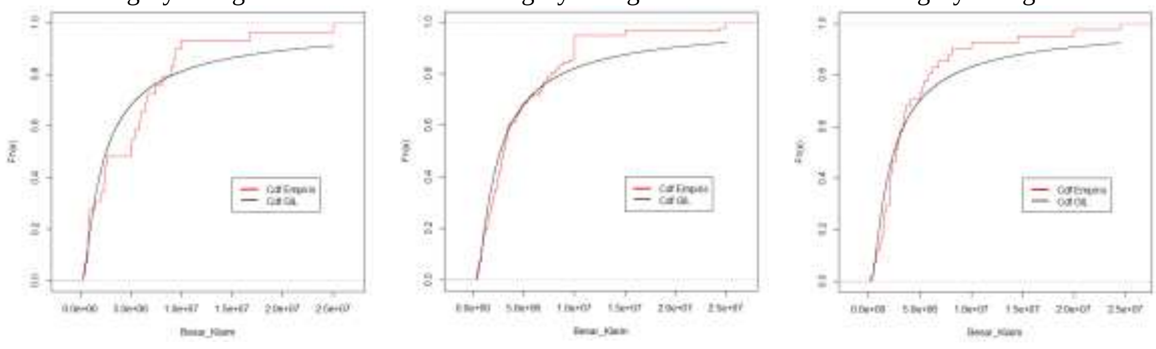
Table no 4: Kolmogorov-Smirnov Fit Test on Category 3 and All Regions

Category	Region	Test Statistics	Critical Value	Conclusion
(1)	(2)	(3)	(4)	(5)
3	1	0,1639	0,2460	H_0 Retrieved
3	2	0,1272	0,1388	H_0 Retrieved
3	3	0,1142	0,2124	H_0 Retrieved

It can be seen that for category 3 and all regions, the null hypothesis is accepted. The conclusion is that the severity on motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 category 3 in all regions comes from a GIL distributed population.

Figure 3 displays the empirical cumulative distribution function graph and the cumulative distribution functions of the GIL distribution for severity on motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 for category 3 and all regions.

Category 3 Region 1 Category 3 Region 2 Category 3 Region 3



Based on Figure 3, it can be seen that the empirical distribution function graph is close to the cumulative distribution function of the GIL distribution for the severity of motor vehicle insurance claims for TPL coverage with *comprehensive* coverage at PT. X in 2019 category 3 and all regions.

IV. Conclusion

Based on the results of the application of the Generalized Inverse-Lindley distribution on the severity of motor vehicle insurance claims for TPL coverage with comprehensive coverage at PT. X in 2019 using the Kolmogorov-Smirnov test it can be concluded that severity on motor vehicle insurance claims for TPL coverage of PT. The X of 2019 comes from the Generalized Inverse-Lindley distributed population for severity claims on category 3 and all regions.

Acknowledgment

This paper was written as part of the research results funded by the List of Contents for Budget Implementation of the Directorate of Research, Technology and Community Service, Directorate General of Higher Education, Research and Technology of the Ministry of Education, Culture, Research, and Technology for Fiscal Year 2023, SP DIPA Number 023.17.1.690523/2023 4th revision dated March 31, 2023.

References

[1] N. Indra, "Pelaksanaan Asuransi Kebakaran Pada Pt. Asuransitakaful Umum Cabang Padang," 2015.
[2] A. N. Hidayah, "Pengaruh Klaim, Hasil Investasi, Risk Based Capital (Rbc) Terhadap Laba Perusahaan Asuransi Syariah Di Indonesia Tahun 2016," 2018.
[3] Y.-K. Tse, "Nonlife Actuarial Models Theory, Methods And Evaluation," 2009.
[4] R. S. Nastiti And A. K. Mutaqin, "Penerapan Model Komposit Weibull-Pareto Pada Data Klaim Asuransi Harta Benda Skripsi Oleh Raisha Shahelia Nastiti Npm: 10060115027 Program Studi Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Bandung 2021," 2021.
[5] A. Asgharzadeh, S. Nadarajah, And F. Sharafi, "Generalized Inverse Lindley Distribution With Application To Danish Fire Insurance Data," Commun. Stat. - Theory Methods, Vol. 46, No. 10, Pp. 5001–5021, May 2017, Doi: 10.1080/03610926.2015.1096394.
[6] M. E. Ghitany, B. Atieh, And S. Nadarajah, "Lindley Distribution And Its Application," Math. Comput. Simul., Vol. 78, No. 4, Pp. 493–506, Aug. 2008, Doi: 10.1016/J.Matcom.2007.06.007.
[7] V. K. Sharma, S. K. Singh, U. Singh, And F. Merovci, "The Generalized Inverse Lindley Distribution: A New Inverse Statistical Model For The Study Of Upside-Down Bathtub Data," Commun. Stat. - Theory Methods, Vol. 45, No. 19, Pp. 5709–5729, Oct. 2016, Doi: 10.1080/03610926.2014.948206.
[8] I. S. Mabrouk, "Statistical Analysis For An Imprecise Flood Dataset Using The Generalized Inverse Lindley Distribution," Int. J. Contemp. Math. Sci., Vol. 14, No. 4, Pp. 163–177, 2019, Doi: 10.12988/Ijcms.2019.9718.
[9] S. Nugroho, Statistika Mutivariat Terapan. 2008.