Available online at http://iorajournal.org/indx.php/orics /index

Operations Research: International Conference Series

e-ISSN: 2722-0974
p-ISSN: 2723-1739

Vol. 4, No. 4, pp. 117-125, 2023

Life Insurance Aggregate Claims Distribution Model Estimation

Setyo Yohandoko [1*], Agung Prabowo[2], Usman Abbas Yakubu[3], Chun Wang[4]

[1,2] *The Faculty of Mathematics and Natural Sciences Jenderal Soedirman University, Purwokerto, Indonesia*

[3]*Department of Mathematics, Yusuf Maitama Sule University, Kano, Nigeria*

[4]*JCU Singapore Business School, James Cook University Singapore, 149 Sims Drive, 387380, Singapore*

**Corresponding author email*: *setyo.yohandoko@mhs.unsoed.ac.id*

Abstract

Risk is a hazard or consequence that can occur in an ongoing process or future events. As the party responsible for assuming and managing risks, the insurance company must be prepared to provide compensation in the event of claims; otherwise, they may face bankruptcy. Hence, it is important to understand the characteristics of risks handled by the insurance company. The risk's characteristics can be analyzed through the distribution model of previous-period claims. The sum of aggregate claims over several periods forms the aggregate claims distribution. The aggregate claims distribution used to determine the amount of pure premium and gross premium that must be obtained by the insurance company. In this research, the determination of distribution model estimation was examined for data cases on aggregate claims of life insurance in Indonesia 2016-2020. The result of this research conduct that the appropriate distribution model is the inverse Gaussian 3P distribution (three parameters).

Keywords: Aggregate claims, Inverse Gaussian 3P dirstribution, Life Insurance, Maximum Likelihood Estimation

1. Introduction

Risk is a hazard or consequence that can occur in an ongoing process or future events (Hanafi, 2016). Life insurance is among the services provided by insurance companies. Life insurance is insurance that aims to cover a person from unexpected financial losses caused by the death of a person too soon or his life is too long (Hasan,2014). The insured party will pay a premium as a payment for the coverage, and the amount of the premium has been predetermined and agreed upon by both parties. In the event that the insured party faces a disaster resulting in loss of life (death), the life insurance company, as the risk insurer, will pay a set of claims based on the agreed-upon amount. A claim is compensation for a risk of loss (Bowers et al., 1997). Insurance companies, as institutions that provide assurance and manage risks, must be prepared to provide compensation in the event of a claim by insurance participants; otherwise, the insurance company could face bankruptcy (Riaman et al., 2013).

Several previous studies have examined insurance risk. Arrow (1971) obtained the optimal solution for the undesirable consequences of no risk transfer at all, or total risk transfer, his analysis proposed partial risk transfer, or, coinsurance, where the insurer pays a certain proportion of the loss. Dassios & Embrechts (1989) suggested the Davis method, to ensure that Davis theory can be used for more realistic insurance model analysis, it is important to understand in detail its scope and flexibility. Feng (2023) discussed four risk management techniques commonly used in practice, namely risk avoidance, risk retention, risk transfer, and risk mitigation. Utility theory is introduced and used to explain why certain insurance product designs have become popular to meet risk management needs.

Therefore, it is very important to understand the character of the risks handled by the insurance company. The character of risk can be studied based on the distribution model of claims that occurred in previous periods. The accumulation of aggregate claims over several insurance periods can form a probability distribution known as the aggregate claim distribution. The distribution of aggregate claims can be used to predict the average value of aggregate claims that occur over a period of time.

2. Literature Review

Riaman (2013) in his article, analyzed collective risk in credit life insurance using an aggregated claim model. In the paper, he explained the utilization of the Collective Risk Model in credit life insurance to measure the risks faced by insurance companies. In credit life insurance, if a debtor passes away, the insurance company is responsible for refunding the remaining credit issued by the bank. In this context, the insurance company must pay close attention to the risks arising from the bank during the insurance period, as neglecting these risks could lead to losses. Therefore, the Collective Risk Model is employed to measure the risk by forming an aggregated claim model based on extensive individual claim data and the number of claim occurrences.

The results of applying the Collective Risk Model to life insurance companies indicated that the aggregate claims model for credit life insurance follows a compound negative binomial distribution. The extent of risk borne by the insurance company depends on the magnitude of individual claims and the number of claims that occur during the insurance period. Overall, the analysis in the paper demonstrates the use of the Collective Risk Model as a tool for measuring risk in credit life insurance. By understanding risks characteristics, insurance companies can take appropriate steps in risk management and selecting suitable premium rates.

3. Materials and Methods

3.1 Materials

The main problem in this research is the estimation of the aggregate claims distribution model of life insurance based on aggregate claims data of life insurance in Indonesia from 2016 until 2020. The type of data is secondary data in which obtained from the official website of the Financial Services Authority (OJK), Indonesian Insurance Statistics of 2020. The aggregate claims data is analyzed, so that can be known the probability distribution which is appropriate with aggregate claims data. The determination of aggregate claims distribution is assisted with Easyfit version 5 through the conformity distribution test, Kolmogorov-Smirnov dan Anderson-Darling. Then the appropriate distribution determined for its mean by using the maximum likelihood estimation method and its variance by using the approach of exponential family distribution. Both mean and variance of a distribution are utilized in further calculations to determine the gross insurance premium.

3.1.1 Insurrance Risk Model

In risk underwriting, insurance companies must understand the characteristics of the risks they managed, so that they can avoid losses. The characteristics of risk can be studied through insurance risk models. An insurance risk model is a mathematical equation used to calculate the potential losses that may occur for an insurance company. In statistical-mathematical studies of insurance risk, the risk is modeled through an equation that follows a certain probability distribution, from which the parameter values can be estimated. These parameter values serve as references for predicting the characteristics of the risk. There are two approaches to insurance risk modeling, individual risk models and collective risk models. The collective risk model assumes that it is a random process that generates insurance contract claims. In essence, the collective risk model is quite similar to the individual risk model, especially in terms of assumptions made, such as assuming independent random variables for the claim sizes, constant currency values, and closed risk models. In the collective risk model, the fundamental concept is the random process that generates claims from the portfolio of policies. This process has the characteristic that the portfolio is viewed as a whole and not on an individual basis. In the field of insurance, the collective risk model can be viewed as the total claims, which are defined as:

$$S = X_1 + X_2 + \cdots + X_N. \tag{1}$$

In the equation X_i represents the size of the i-th claim, and N represents the number of claims (non-negative integers). If there are no claims, then $N = 0$. N is a discrete random variable that influences the aggregate claim amount (Bowers et al., 1997). In the collective risk model, risk characteristics can also be studied through the aggregate claims model. The aggregate claims model is defined as:

$$\sum_{i=1}^{n} S_i = S_1 + S_2 + \cdots + S_n. \tag{2}$$

The aggregate claims model is used to determine mean or expected value of total claims $E(S)$ and its variance $Var(S)$, through the test of distribution suitability and parameter estimation.

3.1.2 Inverse Gaussian 3P Distribution

The inverse Gaussian distribution is a distribution that describes the first passage time of a Brownian motion. The inverse Gaussian distribution shares some properties analogous to the Gaussian (Normal) distribution. The probability density function of the inverse Gaussian distribution with parameters μ and λ, denoted as $X \sim IG(\mu, \lambda)$ is as follows:

$$f(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right) \tag{3}$$

(Folks dan Chhikara, 1978).

The probability density function for the three-parameter inverse Gaussian distribution, $X \sim IG(\mu, \lambda, \gamma)$, is as follows:

$$f(x;\mu,\lambda,\gamma) = \sqrt{\frac{\lambda}{2\pi(x-\gamma)^3}}\exp -\frac{\lambda(x-\gamma-\mu)^2}{2\mu^2(x-\gamma)}. \quad (4)$$

With $x > 0$, $\mu > 0$ representing the mean or expectation, $\lambda > 0$ as the scale parameter, and $\gamma > 0$ as the location parameter with $\gamma < x < \infty$. The scale parameter λ describes the density or spread of the inverse Gaussian distribution. A larger value of the scale parameter results in a wider distribution. On the other hand, the location parameter γ represents the shifting of the distribution curve to the left or right.

3.2 Methods

The distribution of aggregate claims is determined with Kolmogorov-Smirnov and Anderson-Darling's goodness of fit test. The aggregate claims distribution conformity test is a test of the hypothesis that the aggregate claims collection data has a certain distribution. The selection of the distribution which is the most appropriate with the aggregate claims data is determined by selecting the distribution that has the smallest statistical value of the Kolmogorov-Smirnov test and the smallest statistical value of the Anderson-Darling test or the distribution that has the highest rank in Easyfit table of distributions. Furthermore, the mean and variance of the selected distribution model are estimated. The Maximum Likelihood Estimation (MLE) method is used to determine the mean, while the approach of the exponential family distribution is used to determine its variance.

3.2.1 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is a statistical test used to determine whether a sample distribution comes from a population with a specific data distribution or follows a particular statistical distribution. It assumes that the sample consists of random variable values from a population with a continuous distribution. The Kolmogorov-Smirnov test compares the cumulative distribution frequency of the observed data (sample) with the theoretical cumulative frequency of the assumed distribution. One of the advantages of the Kolmogorov-Smirnov test is that it does not require categorical data like the Chi-Square test. This test can be applied to single (continuous) data and can be used with small sample sizes. The statistics of the Kolmogorov-Smirnov test are defined as follows (Mood, Graybill and Boes 1974):

$$D = \max|F_S(x_i) - F_t(x_i)| \quad (5)$$

with $i = 1,2, \dots, n$,
$F_S(x_i)$: cumulative sample frequency, and
$F_t(x_i)$: cumulative theoretical distribution frequency.

3.2.2 Anderson-Darling test

The Anderson-Darling test is a common goodness-of-fit test used to determine whether a sample comes from a specified distribution. This method tests the hypothesis that the sample has been drawn from a population with a specified continuous probability distribution function (Press et al., 1992). The Anderson-Darling test exhibits better sensitivity compared to the Kolmogorov-Smirnov test. The Anderson-Darling test generally compares the theoretical distribution function with the observed data distribution function (sample). This test places more emphasis on the tails of the distribution. Anderson-Darling Test. The statistics of the Anderson-Darling test are defined as follows:

$${W_n}^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\,[\ln u_j + \ln(1-u_{n-j+1})] \quad (6)$$

3.2.3 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a method used to estimate one or more population parameters from observed data (sample) in a way that maximizes the likelihood of obtaining that observed data. According to Casella and Berger (2002), if $X_1, \dots, X_n$ a sample that is identically and independently distributed (IID) from a population with a probability density function $f(x|\theta_1, \dots, \theta_k)$, then its likelihood function is defined as follows:

$$\begin{aligned} L(\theta) &= L(\theta_1, \dots, \theta_k|\, x_1, \dots, x_n) \\ &= f(x_1|\theta_1, \dots, \theta_k) f(x_2|\theta_1, \dots, \theta_k) \dots f(x_n|\theta_1, \dots, \theta_k) \\ &= \textstyle\prod_{i=1}^{n} f(x_i|\theta_1, \dots, \theta_k), \end{aligned} \quad (7)$$

where θ_i are unknown parameters and $\theta\_i \in S$, $i = 1,2,\dots,k$. When a set of observations $x_1, x_2, \dots, x_n$ is given, a value $\hat{\theta} \in S$ that maximizes the likelihood function $L(\theta)$ is called the parameter estimate θ. The value $\hat{\theta}$ is the one that satisfies the following equation:

$$f\left(x_1, x_2, \dots, x_n \middle| \hat{\theta}\right) = {}^{max}_{\theta \in S} f(x_1, x_2, \dots, x_n | \theta). \tag{8}$$

3.2.4 The Approach of Exponential Family Distribution

The method of moments for the exponential family of distributions is an estimation method used to estimate the population mean and variance of a distribution. A distribution is considered a member of the exponential family or Exponential Family (EF) if its probability density function can be written as:

$$f(x;\theta) = \exp\left(\frac{\theta x - b(\theta)}{a\phi} + c(x,\phi)\right) \tag{9}$$

with $a\phi$, $b(\theta)$ and $c(x,\phi)$ are real-valued functions, and $c(x,\phi)$ is a function that contains a logarithm. If ϕ is known, the exponential family distribution function $f(x;\theta)$ is a distribution function with the canonical parameter θ (McCullagh and Nelder, 1989). The parameter estimation to determine the mean and variance of a distribution that belongs to the exponential family of distributions, as shown in equation (9), is as follows:

$$L(\theta) = f(x;\theta)$$

$$L(\theta) = \exp\left(\frac{\theta x - b(\theta)}{a\phi} + c(x,\phi)\right)$$

$$\ln L(\theta) = \ln \exp\left(\frac{\theta x - b(\theta)}{a\phi} + c(x,\phi)\right)$$

$$\ln L(\theta) = \frac{\theta x - b(\theta)}{a\phi} + c(x,\phi).$$

$$s(\theta) = \frac{\partial}{\partial\theta} \ln L(\theta)$$

$$s(\theta) = \frac{\partial}{\partial\theta}\left[\frac{\theta x - b(\theta)}{a\phi} + c(x,\phi)\right]$$

$$s(\theta) = \frac{x - \frac{\partial}{\partial\theta}[b(\theta)]}{a\phi}$$

$$E_X[s(\theta)] = E_X\left[\frac{x - \frac{\partial}{\partial\theta}[b(\theta)]}{a\phi}\right] = 0$$

$$\frac{E_X[X] - \frac{\partial}{\partial\theta}[b(\theta)]}{a\phi} = 0$$

$$E[X] = \frac{\partial}{\partial\theta}[b(\theta)]. \tag{10}$$

$$Var_X\big(s(\theta)\big) = Var_X\left(\frac{x - \frac{\partial}{\partial\theta}[b(\theta)]}{a\phi}\right)$$

$$Var_X\big(s(\theta)\big) = \frac{1}{(a\phi)^2} Var_X\left(x - \frac{\partial}{\partial\theta}[b(\theta)]\right) = \frac{Var(X)}{(a\phi)^2}$$

$$-\frac{\partial}{\partial\theta} s(\theta) = -\frac{\partial}{\partial\theta}\left[\frac{x - \frac{\partial}{\partial\theta}[b(\theta)]}{a\phi}\right] = \frac{\frac{\partial^2}{\partial\theta^2}[b(\theta)]}{a\phi}$$

$$-E_X\left[\frac{\partial}{\partial\theta}s(\theta)\right] = E_X\left[\frac{\frac{\partial^2}{\partial\theta^2}[b(\theta)]}{a\phi}\right]$$

$$Var_X\big(s(\theta)\big) = -E_X\left[\frac{\partial}{\partial\theta}s(\theta)\right]$$

$$\frac{Var(X)}{(a\phi)^2} = \frac{\frac{\partial^2}{\partial\theta^2}[cb(\theta)]}{a\phi}$$

$$Var(X) = a\phi\frac{\partial^2}{\partial\theta^2}[b(\theta)]. \tag{11}$$

(Rothman, 2021).

4. Results And Discussion

4.1 Distribution Conformity Test

Based on the results of the Kolmogorov-Smirnov and Anderson-Darling goodness of fit tests, the results were obtained in Table 1 and Table 2.

Table 1 Kolmogorov-Smirnov test results for five distribution candidates

#	Distributions	Kolmogorov-Smirnov			
		p-value	Statistics Value	Critical Value ($\alpha = 0.05$)	Decision
56	Wakeby	0.96393	0.19876	0.56328	Accepted
28	Invers Gaussian 3P	0.94805	0.20812	0.56328	Accepted
15	Frechet	0.90333	0.22899	0.56328	Accepted
35	Log-Logistic	0.87158	0.24092	0.56328	Accepted
19	Gen. Extrem Value	0.86836	0.24204	0.56328	Accepted

From Table 1, it can be observed that the aggregate claims data meet the criteria for all five candidate distributions.

Table 2 Anderson-Darling test results for five distribution candidates

#	Distributions	Anderson-Darling		
		Statistics Value	Critical Value ($\alpha = 0.05$)	Decision
56	Wakeby	3.6029	2.5018	Rejected
28	Inverse Gaussian 3P	0.3409	2.5018	Accepted
15	Frechet	2.8929	2.5018	Rejected
35	Log-Logistic	3.9795	2.5018	Rejected
19	Gen. Extrem Value	0.8080	2.5018	Accepted

From Table 2, it can be observed that the aggregate claims data only meet the criteria for the Inverse Gaussian 3P distribution and the General Extreme Value distribution.

Based on the results of the goodness of fit test in Table 1 and Table 2, it can be concluded for the probability distribution which is appropriate with aggregate claims data ordered by its rank shown in Tabel 3 as follows.

Table 3 Goodness of fit test outputs for aggregate claims distribution

#	Distributions	Kolmogorov-Smirnov		Anderson-Darling	
		Statistics	Ranking	Statistics	Ranking
28	Inverse Gaussian 3P	0.2081	2	0.3409	1
19	Gen. Extrem Value	0.2420	5	0.8080	4

Based on the results of the Kolmogorov-Smirnov test and the Anderson-Darling test, it is known that the probability distribution that appropriate with the aggregate claims data is the Inverse Gaussian 3P distribution. The Inverse Gaussian 3P distribution give the second lowest statistical value of the Kolmogorov-Smirnov test which is 0.20812 and give the lowest statistical value of the Anderson-Darling test which is 0.34099. The Inverse Gaussian 3P distribution ranks second highest for the statistical value of the Kolmogorov-Smirnov test and ranks highest for the statistical value of the Anderson-Darling test.

4.2 Claims Aggregate Distribution Model

Based on the estimation results of the claim aggregate distribution using Easyfit version 5, the probability distribution that appropriate with a set of life insurance aggregate claims data is the Inverse Gaussian 3P distribution with the probability density function denoted as equation (3) as:

$$f(x;\mu,\lambda,\gamma) = \sqrt{\frac{\lambda}{2\pi(x-\gamma)^3}}\exp - \frac{\lambda(x-\gamma-\mu)^2}{2\mu^2(x-\gamma)}.$$

The parameter values for the Inverse Gaussian 3P distribution obtained from the output of the goodness-of-fit test using EasyFit version 5 are $\lambda = 2{,}4492 \cdot 10^{11}$, $\mu = 2{,}2287 \cdot 10^{13}$, $\gamma = 2{,}3782 \cdot 10^{12}$. The loss model distribution based on the probability density function of the Inverse Gaussian 3P distribution that has been parameterized is:

$$f(x) = 1{,}9873 \cdot 10^{6} \exp - \frac{2{,}4492 \cdot 10^{16}(x - 2{,}466 \cdot 10^{17})^2}{9{,}9342 \cdot 10^{26}x - 2{,}3625 \cdot 10^{39}} \qquad (12)$$

4.3 The Estimation of Mean and Variance Model Distribution

4.3.1 Mean Estimation for Inverse Gaussian 3P Distribution

Mean estimation for the Inverse Gaussian 3P distribution is performed using the maximum likelihood estimation method as follows:

$$f(x_i;\mu,\lambda,\gamma) = \lambda^{\frac{1}{2}}\,(2\pi)^{-\frac{1}{2}}\,(x_i-\gamma)^{-\frac{3}{2}}\exp - \frac{\lambda(x_i-\gamma-\mu)^2}{2\mu^2(x_i-\gamma)}$$

$$L(\mu,\lambda,\gamma) = \prod_{i=1}^{n}\left[\lambda^{\frac{1}{2}}\,(2\pi)^{-\frac{1}{2}}\,(x_i-\gamma)^{-\frac{3}{2}}\exp - \frac{\lambda(x_i-\gamma-\mu)^2}{2\mu^2(x_i-\gamma)}\right]$$

$$= \lambda^{\frac{n}{2}}(2\pi)^{-\frac{n}{2}}\prod_{i=1}^{n}(x_i-\gamma)^{-\frac{3}{2}}\exp - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{[(x_i-\gamma)-\mu]^2}{(x_i-\gamma)}.$$

$$\ln L\,(\mu,\lambda,\gamma) = \ln\lambda^{\frac{n}{2}} + \ln(2\pi)^{-\frac{n}{2}} + \ln\left(\prod_{i=1}^{n}(x_i-\gamma)^{-\frac{3}{2}}\right) + \ln\left(\exp - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{[(x_i-\gamma)-\mu]^2}{(x_i-\gamma)}\right)$$

$$\ln L\,(\mu,\lambda,\gamma) = \frac{n}{2}\ln\lambda - \frac{n}{2}\ln 2\pi - \frac{3n}{2}\sum_{i=1}^{n}\ln(x_i-\gamma) - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{[(x_i-\gamma)-\mu]^2}{(x_i-\gamma)}$$

$$\frac{[(x_i-\gamma)-\mu]^2}{(x_i-\gamma)} = \frac{(x_i-\gamma)^2 - 2\mu(x_i-\gamma)+\mu^2}{(x_i-\gamma)}$$

$$= \frac{(x_i-\gamma)^2}{(x_i-\gamma)} - \frac{2\mu(x_i-\gamma)}{(x_i-\gamma)} + \frac{\mu^2}{(x_i-\gamma)}$$

$$= (x_i-\gamma) - 2\mu + \frac{\mu^2}{(x_i-\gamma)}$$

$$\ln L\,(\mu,\lambda,\gamma) = \frac{n}{2}\ln\lambda - \frac{n}{2}\ln(2\pi) - \frac{3n}{2}\sum_{i=1}^{n}\ln(x_i-\gamma) - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}(x_i-\gamma) + \frac{n\lambda}{\mu} - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{(x_i-\gamma)}$$

$$\frac{\partial \ln L[f(x_i;\mu,\lambda,\gamma)]}{\partial\mu} = \frac{\partial\left[-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}(x_i-\gamma) + \frac{n\lambda}{\mu} - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{(x_i-\gamma)}\right]}{\partial\mu}$$

$$\frac{\partial \ln L[f(x_i;\mu,\lambda,\gamma)]}{\partial\mu} = \frac{\lambda\sum_{i=1}^{n}(x_i-\gamma)}{\mu^3} - \frac{n\lambda}{\mu^2}$$

$$\frac{\partial \ln L[f(x_i;\mu,\lambda,\gamma)]}{\partial\mu} = 0$$

$$\frac{\lambda\sum_{i=1}^{n}(x_i-\gamma)}{\mu^3} - \frac{n\lambda}{\mu^2} = 0$$

$$\frac{\lambda\sum_{i=1}^{n}(x_i-\gamma)}{\mu^3} = \frac{n\lambda}{\mu^2}$$

$$\frac{\sum_{i=1}^{n}(x_i-\gamma)}{\mu} = n$$

$$\mu = \frac{\sum_{i=1}^{n}(x_i-\gamma)}{n}$$

$$\mu = \frac{\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\gamma}{n}$$

$$\mu = \frac{\sum_{i=1}^{n}x_i}{n} - \frac{n\gamma}{n}$$

$$\mu = \bar{x} - \gamma \qquad (13)$$

4.3.2 Variance Estimation for Inverse Gaussian 3P Distribution

Mean estimation for the Inverse Gaussian 3P distribution is performed using the maximum likelihood estimation method as follows:

$$\ln f(x) = \ln\lambda^{\frac{1}{2}} + \ln 2\pi^{-\frac{1}{2}} + \ln\,(x-\gamma)^{-\frac{3}{2}} - \frac{\lambda(x-\gamma-\mu)^2}{2\mu^2(x-\gamma)}$$

$$\ln f(x) = -\frac{\lambda(x-\gamma-\mu)^2}{2\mu^2(x-\gamma)} + \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{3}{2}\ln(x-\gamma).$$

$$f(x) = \exp\left(-\frac{\lambda(x-\gamma)}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2(x-\gamma)} + \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{3}{2}\ln(x-\gamma)\right)$$

$$\frac{\theta(x-\gamma)}{a\phi} = -\frac{\lambda(x-\gamma)}{2\mu^2}.$$

$$f(x) = \exp\left(-\frac{\lambda(x-\gamma)}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2(x-\gamma)} + \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{3}{2}\ln(x-\gamma)\right)$$

$$f(x) = \exp\left(\frac{\theta(x-\gamma) - 2\theta^{\frac{1}{2}}}{-2\phi} + \frac{1}{2\phi(x-\gamma)} + \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{3}{2}\ln(x-\gamma)\right)$$

$$f(x) = \exp\left(\frac{\theta(x-\gamma) - 2\theta^{\frac{1}{2}}}{-2\phi} + c\big((x-\gamma),\phi\big)\right)$$

$$E[S] = \frac{\partial b(\theta)}{\partial\theta} = 2\frac{\partial\theta^{\frac{1}{2}}}{\partial\theta} = \theta^{-\frac{1}{2}} = \frac{1}{\theta^{\frac{1}{2}}} = \mu$$

$$Var(S) = a\phi\frac{\partial^2 b(\theta)}{\partial\theta^2} = -2\phi\frac{\partial\theta^{-\frac{1}{2}}}{\partial\theta} = \phi\theta^{-\frac{3}{2}}. \quad (14)$$

5. Conclusion

Based on this research, the estimation of life insurance aggregate claims distribution model, the distribution which is appropriate with a set of data aggregate claims of life insurance in Indonesia is Inverse Gaussian 3P distribution. The estimated mean μ using the maximum likelihood method is denoted as $\bar{x} - \gamma$, and its variance $Var(S)$ estimated using the exponential family distributions approach is denoted as $\phi\theta^{-\frac{3}{2}}$. The estimation of the mean and variance is used for determining the amount of pure premium and its variance that life insurance companies in Indonesia need to provide.

Aknowledgements

The completion of this research would not have been possible without the valuable contributions and support from various individuals and institutions. We would like to express our heartfelt gratitude to all those who have played a significant role in this study. First and foremost, we extend our deepest appreciation to our supervisor, Prof. Dr. Sukono, M.M., M.Si. for his continuous guidance, valuable insights, and unwavering support throughout the research process. Their expertise and encouragement have been instrumental in shaping the direction of this study. Our sincere thanks go to our colleagues and friends who provided valuable feedback and constructive criticism during the course of this study. Their insights have helped improve the quality of this paper. Last but not least, we express our heartfelt appreciation to our families for their understanding, encouragement, and patience throughout this research journey.

Refrences

Andriyani, A. (2017). Penentuan Premi Asuransi Kendaraan Menggunakan Pendekatan Distribusi Peluang. *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya II (KNPMP II)*. Surakarta 18 Maret 2017. Universitas Muhammadiyah Surakarta. 37-49.

Arrow, K. J. (1971). Insurance, risk and resource allocation. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

Bain, L. J., dan Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics Second Edition.* California: Duxbury.

Bowers Jr., N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., dan Nesbitt, C. J. (1997). *Actuarial Mathematics (2nd ed.).* Illinois: The Society of Actuaries.

Casella, G., dan Berger, R. L. (2002). *Statistical Inference (Second Edition)*. California: Duxbury.

Chrisan, K. W., Altien, J. R., dan Manurung, T. (2019). Model Distribusi Data Klaim Asuransi Mobil untuk Menentukan Premi Murni. *Jurnal d'CartesiaN*, 8(2), 108-113. https://doi.org/10.35799/dc.8.2.2019.24378

Cleménte, G. P., dan Savelli, N. (2010). A Collective Risk Model for Claims Reserve Distribution. *Atti del XVI Convegno di Teoria del Rischio*, (pp. 59-87). Retrieved Mei 12, 2022, from https://www.researchgate.net/publication/261397269_A_Collective_Risk_Model_for_Claims_Reserve_Distribution

Dassios, A., & Embrechts, P. (1989). Martingales and insurance risk. *Communications in Statistics. Stochastic Models*, *5*(2), 181-217.

Dickson, D. C. M. (2005). *Insurance Risk and Ruin*. Cambridge: Cambridge University Press.

Feng, R. (2023). Economics of Risk and Insurance. In *Decentralized Insurance: Technical Foundation of Business Models* (pp. 55-84). Cham: Springer International Publishing.

Folks, J. L. dan Chhikara, R. S. (1978). The Inverse Gaussian Distribution and its Statistical Application-A Review. *J. R. Statist. Soc. B*, 40(3), 263-289.

Forbes, C., Evans, M., Hastings, N., dan Peacock, B. (2011). *Statistical Distributions.* New Jersey: John Wiley dan Sons.

Hanafi, M. M. (2016). *Manajemen Resiko (edisi ke-3)*. Yogyakarta: UPP STIM YKPN.

Hasan, N. I. (2014). *Pengantar Asuransi Syariah*. Jakarta: Gaung Persada Press Grup.

Hogg, R. V. dan Klugman, S. A. (1984). *Loss Distributions*. New Jersey: John Wiley dan Sons.

Kellison, S. G., dan London, R. L. (2011). *Risk Models and Their Estimation.* Winsted: ACTEX Publications, Inc.

Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Kamus Besar Bahasa Indonesia. [Online]. Retrieved Mei 25, 2022, from KBBI Daring: https://kbbi.kemendikbud.go.id/entri/risiko.

McCullagh, P., dan Nelder, J. A. (1989). *Generalized Linear Models. 2nd Edition*. London: Chapman and Hall.

Mood, A. M., Graybill, F. A., dan Boes, D. C. (1974). *Introduction to the Theory of Statistics, 3rd*. New York: McGraw-Hill.

Otoritas Jasa Keuangan. (2021). *Statistik Perasuransian Indonesia tahun 2020*. Jakarta: OJK.

Prabowo, A., Mamat, M., Sukono, dan Taufiq, A. A. (2019). Pricing of Premium for Automobile Insurance using Bayesian Method. *International Jurnal of Recent Technology and Engineering (IJRTE),*8(3), 6226-6229. https://doi.org/10.35940/ijrte.C5740.098319

Press, W. H., Flannery, B. P., Teukolsky, S. A., dan Vetterling, W. T. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge: Cambridge University Press.

Riaman., Supena, Y., Lesmana, E., Sukono, F., dan Firdaus, R. (2013). Analisis Model Risiko kolektif pada Asuransi Jiwa Kredit Menggunakan Model Klaim Agregasi. *Prosiding Seminar Nasional Sains dan Teknologi Nuklir PTNBR – BATAN Bandung 04 Juli 2013*, pp 634-643. ISSN 1858-3601. Retrieved 25 Mei, 2022, from https://adoc.pub/analisis-model-risiko-kolektif-pada-asuransi-jiwa-kredit-men.html

Rothman, A. (2021). GLM Part I: Rigorous Mathematical Formulation. Retrieved June 21, 2022, from Toward Data Science: https://towardsdatascience.com/generalized-linear-models-a-rigorous-mathematical-formulation-58ac2ec7d9ea

Sumiyati. (2019). *Layanan Lembaga Perbankan Dan Keuangan Mikro*. Jakarta: Kompas Gramedia.

Vaughan, E. J. dan Vaughan, T. (1987). *Fundamental of Risk and Insurance (10th ed.*). New Jersey: John Wiley dan Sons.

Walpole, R. E., Myers, R. H., Myers, S. L. dan Ye, K. (2011). *Probability dan Statistics for Engineers dan Scientists Ninth Edition*. New Jersey: Prentice Hall.