International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Pricing of Premium for Automobile Insurance using Bayesian Method

Agung Prabowo, Mustafa Mamat, Sukono, Afif Amrullah Taufiq

Abstract*: *The aggregate claim model can be used to determine the amount of premium charged to the insured by the insurance company. This model consists of two mutually independent random variables, namely the number of claims that occur per period and the amount of claim for each event. In this study, the number of claims is Poisson distributed, and the amount of claim is distributed by generalized extreme value (GEV). The Bayes method is used to estimate the parameters of each distribution. Parameter estimation results are used to calculate the expectations and variances of the aggregate claim model which are then used to calculate insurance premiums. Based on the estimation results, the amount of premium charged to the insured ranges from IDR 3,831,480 to IDR 6,443,860.

Index Terms*: *Bayesian method, mobile insurance, premium, pricing.

I. INTRODUCTION

The growth of motorcycles number in Indonesia in the period of 2015-2017 was quite rapid [1], accompanied by an increase in the risk of damage to motorcycles. Therefore, an action is needed to reduce the risk by providing insurance [2],[3].
As takeover institution and risk recipient, insurance companies must be able to anticipate whether claims that occur cause losses and make the company go bankrupt [4]. Losses of insurance companies can be predicted if the company knows the characteristics of the loss distribution [5]. In loss insurance, the loss distribution is modeled by the distribution of aggregate claims consisting of two mutually independent variables, namely the number of claims that occur per period and the amount of claim for each incident [6]. The distribution of claim number is usually modeled with a discrete distribution, while the amount of claims is modeled with a continuous non-negative distribution [7][8].
In loss insurance, the aggregate claim model is used to determine the premium paid by the insured to the insurance company [9]. This study aims to estimate the model of aggregate claim on automobil insurance using the Bayes method. According to [2],[10], Bayes method can be used to estimate parameters of the distribution of the number and amount of claims. Then, based on the results of the estimation obtained, the amount of premium to be paid by the insured to the insurance company is determined.

II. RESEARCH METHODOLOGY

This study combines the literature study and the use of secondary data. The premium determination is done based on four principles, namely: pure premium, expectation value, variance value and standard deviation value. Easyfit 5.6 software is used for data processing to determine the distribution of the number and amount of claims. Based on the data, it is known that the number of claim distribution is Poisson with λ parameter, while the amount of claim distribution is generalized extreme value (GEV) with μ, σ, and ξ parameters [11].
The Poisson and GEV likelihood distribution functions are made to estimate the distribution parameters. Parameter estimation is done using the Bayes method. The use of the Bayes method requires the need to specify prior and posterior distributions for each parameter in both distributions [12],[13].
OpenBUGS software that works with the principle of the Monte Carlo Markov Chain (MCMC) and Gibbs sampler algorithm is used to determine the estimated parameter values of both distributions. The estimated parameter values are obtained from the sample data. Markov chain convergence is achieved when the mc-error value is smaller than 5% posterior standard deviation [14],[15].
The estimated value of the four parameters that have been obtained is used to determine the expected value and the variance of the number and amount of claims distribution. The expectation and variance values of aggregate claim distribution are then determined [16]. Premium loading factors are obtained using the standard normal distribution approach. Finally, car insurance premium is determined by using the four premium calculation principles [17].

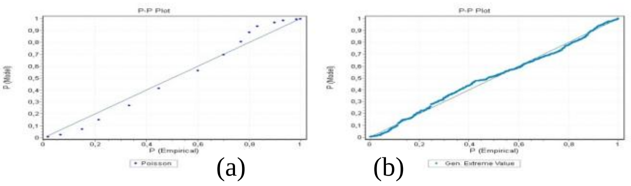
III. LITERATURE STUDY AND DATA ANALYSIS

A. Claim distribution model (The number and amount of claims)

The determination of the N claim number and the $X_1, X_2, \ldots, X_N$ severity of claim distribution forms based on the data are conducted by using the Easy fit 5.6 software. The form of the distribution that represents the number of claims is the Poisson distribution.

Manuscript published on 30 September 2019
* Correspondence Author
Agung Prabowo*, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Indonesia.
Mustafa Mamat, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia.
Sukono, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia.
Afif Amrullah Taufiq, Engineering Division, General Insurance of Bumiputera Muda 1967, Indonesia.

The plot between the values of the empirical cumulative distribution function of the data on the number of claims with the theoretical cumulative distribution function value of the Poisson distribution is shown in Figure 1 (a) in the form of a probability plot [7].



(a) (b)

Figure 1. Probability plot for the number (a) and amount of claims (b)

Based on the probability plot curve in Figure 1 (a), it can be seen that the Poisson distribution model is proper to represent the number of claims. This can be seen from the spread of points that follow the diagonal line. The mass function probability of a Poisson distribution with parameters $\lambda > 0$ is:

$$p(n)=\begin{cases}\dfrac{e^{-\lambda}\cdot\lambda^{n}}{n!} & ;n=0,1,2,....\\ 0 & ;x \text{ lainnya}\end{cases} \tag{1}$$

Based on the probability plot curve in Figure 1 (b), it can be seen that the distribution model of generalized extreme value (GEV) is proper to represent the amount of claim distribution. This can be seen from the points that spread around the diagonal line and the spread follows the diagonal line Figure 1 (b) [4]. The probability density function for the X_i claim distributed with GEV with μ, σ, and ξ parameters given in equation (2) with $-\infty < \mu < \infty$ and $-\infty < \xi < \infty$.

$$f(x)=\begin{cases}\dfrac{1}{\sigma}\left[1+\xi\dfrac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}-1} e^{-\left[1+\xi\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}} & ;\xi\neq 0 \text{ dan } \left\{x:1+\xi\left(\dfrac{x-\mu}{\sigma}\right)>0\right\}\\ \dfrac{1}{\sigma}e^{-\left[\frac{(x-\mu)}{\sigma}+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]} & ;\ \xi=0 \text{ dan } x\in\Re\end{cases} \tag{2}$$

B. Aggregate claim model

A collective risk model can be formed in equation (3)

$$S=\sum_{i=1}^{N}X_i \tag{3}$$

with random variable S stating the random variable of aggregate claims resulted from the portfolio in a certain period. The assumptions used in this model are (1) the X_i amount of claim is a non-negative random variable that is identically distributed and mutually independent, and (2) the random variable N number of claims is independent against X_i amount of claim.

The expectation value and variance of the aggregate claim model (3) can be calculated using equations (4) and (5) as follows [6; 7]:

$$E[S]=E[X]E[N] \tag{4}$$

$$Var(S)=E[N]Var(X)+(E[X])^2Var(N) \tag{5}$$

C. Bayesian method

In the Bayes method, a distribution parameter is considered random variable that has a distribution and is called prior distribution. The main problem in the Bayes method is to choose the prior distribution for an unknown parameter but suitable for the problem. Based on the data, the distribution of N number of claims is Poisson with λ parameter and the distribution of X_i amount of claim is GEV with μ, σ, and ξ parameters [18].

According to [18], the $\lambda > 0$ parameters in the Poisson distribution has prior conjugate distribution in the form of gamma distribution with α and β hyper parameters. The probability density function of the gamma distribution is given in equation (6):

$$\widehat{p}(\lambda)=\frac{\lambda^{\alpha-1}e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\ ;\quad \lambda\geq 0, \alpha>0, \beta>0 \tag{6}$$

According to [11], one of the prior distributions for GEV distribution parameters is the maximal data information (MDI) prior. MDI prior is defined as $\widehat{p}(\theta)\propto e^{E[\ln f(x|\theta)]}$, so the combined prior distribution for μ, σ, and ξ is given in equation (7) with $\sigma>0$, $\mu\in\Re$, $\xi>$ -1 and $\gamma=0{,}57722$ is Euler constant.

$$\widehat{p}(\theta)=\frac{1}{\sigma}e^{-\gamma\left(\xi+1+\frac{1}{\gamma}\right)}\propto\frac{1}{\sigma}e^{-\gamma(1+\xi)} \tag{7}$$

D. Likelihood function

If θ is specified, the likelihood function is a function of the combined density of n $X_1, X_2, ..., X_n$ random variables for θ parameter and denoted by $L(\theta)$. If $x_1, x_2, ..., x_n$ states $X_1, X_2, ..., X_n$ random samples observations with function density opportunity $f(x;\theta)$, according to [12] the likelihood function given in Equation

$$L(\theta)=f(x_1;\theta)f(x_2;\theta).....f(x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta) \tag{8}$$

Based on equation (8), the likelihood function of the Poisson distribution in Equation (1) is

$$L(\lambda)=\frac{e^{-n\lambda}\cdot\lambda^{\sum_{i=1}^{n}y_i}}{\prod_{i=1}^{n}y_i!} \tag{9}$$

Next, based on equation (8), the likelihood function of the GEV distribution in equation (2) is

$$L(\mu,\sigma,\xi)=\frac{1}{\sigma^n}\left(\prod_{i=1}^{n}\left[1+\xi\frac{(x_i-\mu)}{\sigma}\right]^{-\frac{1}{\xi}-1}\right)e^{-\sum_{i=1}^{n}\left[1+\xi\frac{(x_i-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}} \tag{10}$$

E. Posterior distribution for Poisson and GEV parameter

Posterior distribution can be used to determine estimators and interval estimates of unknown parameters. The Markov Chain Monte Carlo (MCMC) method is used to determine the value of parameter estimates from the posterior distribution,.

According to [13] the posterior distribution is a θ conditional probability density function if the x observed value is identified. Posterior distribution can be written as:

$$f(\theta|x)=\frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{\infty}f(x|\theta)f(\theta)d\theta} \tag{11}$$

with $f(x;\theta)$ is a likelihood function and $f(\theta)$ is the prior distribution.

The posterior distribution for λ parameter in the Poisson distribution is obtained by substituting equations (6) and (9) in equation (11) as:

$$f(\lambda|y_1,y_2,\ldots,y_3)=\frac{L(\lambda)\bar{p}(\lambda)}{\int_0^\infty L(\lambda)\bar{p}(\lambda)d\lambda} \tag{12}$$

$$=\frac{1}{C_1}\cdot e^{-\left(n+\frac{1}{\beta}\right)\lambda}\cdot\lambda^{\sum_{i=1}^{n}y_i+\alpha-1}$$

with $C_1=\Gamma\left(\sum_{i=1}^{n}y_i+\alpha\right)\cdot\left(n+\frac{1}{\beta}\right)^{-\left(\sum_{i=1}^{n}y_i+\alpha\right)}$ is the normalization constant.

The combined posterior distribution for μ, σ, and ξ parameters of the GEV distribution is obtain by substituting equations (7) and (10) in equation (11).

$$f(\mu,\sigma,\xi|x_1,x_2,\ldots,x_3)$$

$$=\frac{L(\mu,\sigma,\xi)\bar{p}(\mu,\sigma,\xi)}{\int_{-1}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty}L(\mu,\sigma,\xi)\bar{p}(\mu,\sigma,\xi)d\mu\, d\sigma\, d\xi}$$

$$=\frac{1}{C_2}e^{-\lambda\xi}\frac{1}{\sigma^{n+1}}\left(\prod_{i=1}^{n}\left[1+\xi\frac{(x_i-\mu)}{\sigma}\right]^{-\frac{1}{\xi}-1}\right)e^{-\sum_{i=1}^{n}\left[1+\xi\frac{(x_i-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}} \tag{13}$$

with

$$C_2=\int_{-1}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty}e^{-\lambda\xi}\frac{1}{\sigma^{n+1}}\left(\prod_{i=1}^{n}\left[1+\xi\frac{(x_i-\mu)}{\sigma}\right]^{-\frac{1}{\xi}-1}\right)e^{-\sum_{i=1}^{n}\left[1+\xi\frac{(x_i-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}}d\mu\, d\sigma\, d\xi$$

is the normalization constant.

IV. RESULTS AND DISCUSSION.

A. Parameter estimation

The data used in this study are data on motor vehicle insurance such as sedans, minibuses, jeeps, station wagons, and the like, with insurance coverage amount of IDR 125,000,001 – IDR 200,000,000 of PT. Asuransi Jasa Indonesia Purwokerto branch office from 2013 to 2017. The insurance claim data is further divided into two groups, namely data on the number of claims and data on the amount of claim (Table 1). In data simulation, the specified posterior distribution is used to estimate the value of its distribution parameters. The MCMC method is effective enough to determine the estimated value of the parameters of a random variable obtained from a posterior distribution that is very complex and quite complicated if solved using analytic means. One technique used in the MCMC method is Gibbs sampler [7].

Table 1. Number of claim and claim amount for mobile insurance

Year	Number of Claim	Claim Amount (IDR)
2013	80	509,845,468,5
2014	98	400,612,304,5
2015	115	637,940,404,8
2016	80	403,043,129,0
2017	62	280,397,453,0

Iteration is done 20,000 times and resulted in a sample of 39,000 for each parameter. The parameter estimation results are shown in Table 2.

Table 2. Results of data simulation with *OpenBUGS*

Parameter	Mean $(\times10^3)$	Standard Deviation $(\times10^3)$	5% Standard Deviation $(\times10^3)$	mc-error $(\times10^3)$
λ	7.250	0.2459	0.0123	0.0013
μ	1.910	0.0904	0.0045	0.0018
σ	1.696	0.0888	0.0044	0.0015
ξ	0.480	0.0173	0.0009	0.0001

Based on Table 2, the Markov chain has been converged. This is stated by the mc_error value which is less than 5% standard deviation for all four parameters. The parameter estimation value can be seen in the mean column of each parameter.

B. Aggregate claim model estimates

Based on the results of the simulation data that has been done, it can be concluded that the Poisson distributed number of claims with λ parameter is 7.250, while the GEV distributed amount of claim with μ parameter is 1.910, σ parameter is 1.696, and ξ parameter is 0.480. Furthermore, the expectation value and variance of the Poisson distribution [12] are $E[N]=\lambda=7.250\ (\times10^3)$ and $Var(N)=\lambda=7.250\ (\times10^3)$

The expectation value and variance of GEV distribution [16] are

$$E[X]=\mu+\frac{\sigma(\Gamma(1-\xi)-1)}{\xi}=1.91+\frac{1.696(\Gamma(1-0.4801)-1)}{0.4801}=4.404\ \ (\times10^3)$$

$$Var(X)=\frac{\sigma^2\left(\Gamma(1-2\xi)-(\Gamma(1-\xi))^2\right)}{\xi^2}=\frac{1.696^2\cdot21.675^2}{0.4801^2}=270.488\ \ (\times10^6)$$

Based on Equations (4) and (5), the expectation and variance of the aggregate claim model are

$$E[S]=4.404\cdot7.25=31.929\ \ (\times10^3)$$

$$Var(S)=(7.25\cdot270.488)+(4.404^2\cdot7.25)=2{,}101.653\ \ (\times10^6)$$

with $E[S]$ is the average amount of loss received by insurance company every month, and $Var(S)$ is a measure of the loss distribution received by insurance company every month. The expectation and the loss variance received by insurance company every year are

$$E[S^*]=31.929\times12=383.1480\ \ (\times10^3)$$

$$Var(S^*)=2{,}101.653\times12=25{,}219.836\ \ (\times10^6)$$

C. Premium calculation

The premium is calculated based on four principles, namely the principle of pure premium, the principle of expectation value, the principle of variance, and the principle of standard deviation [6]. Premium calculation based on the pure premium principle can be solved by equation (14):

$$\Pi_{S^*}=E[S^*] \tag{14}$$

The (η) premium loading factor is calculated based on the approximation of aggregate claim using the Central Limit Theorem [5] at the α level of significance using equation (15):

$$P(S^*<\Pi_{S^*})=P\left(\frac{S^*-E(S^*)}{\sqrt{Var(S^*)}}<\frac{\Pi_{S^*}-E(S^*)}{\sqrt{Var(S^*)}}\right)=1-\alpha \tag{15}$$

Dickson [6] provided three premium calculation principles in Equations (16) for principle the expectation value, equations (17) for principle of variance and equations (18) for principle of standard deviation.

$$\Pi_{S^*} = E[S^*] + \eta \cdot E[S^*] \quad (16)$$

$$\Pi_{S^*} = E[S^*] + \eta \cdot Var(S^*) \quad (17)$$

$$\Pi_{S^*} = E[S^*] + \eta \cdot \sqrt{Var(S^*)} \quad (18)$$

The results of calculations using Equations (14) through (18) are given in Table 3. The (η) premium value of loading factor and the amount of premium per person billed based on each principle if there are 100 portfolios with α =0.05 are given in Table 3. The figures in Table 3 are obtained from equations (14) and (16) through (18). The premium for the principle of pure premium is calculated without involving premium loading factor so, in Table 3 there is no any premium loading factor value.

Table 3. Premium loading factor and premium per person

Premium Calculation Principle	Premium loading factor (η)	Premium per person (IDR)
Pure Premium Principle	-	3,831,480
Principle the Expectation Value	0.68182	6,443,860
Principle of Variance	0.01000	6,353,460
Principle of Standard Deviation	1.64500	6,443,860

V. CONCLUSION

Based on the results and discussion, the distribution of claims amount is Poisson with the λ parameter estimation value is 7.250 while the GEV distributed amount of claim with μ parameter is 1.910, σ parameter is 1.696, and ξ parameter is 0.480.

Furthermore, the expected value of aggregate claim for a year is IDR 383,148,000 with a variance of IDR 25,219,836,000. The premium per person (insured) paid to the insurance company based on the principle of pure premium is IDR 3,831,480, the principle of expectation value is IDR 6,443,860, the principle of variance is IDR 6,353,460, and the principle of standard deviation is IDR 6,443,860.

REFERENCES

1. Badan Pusat Statistik 2018 Statistik Indonesia. Jakarta: Badan Pusat Statistik, 2018.
2. Sukono, Riaman, E. Lesmana, R. Wulandari, H. Napitupulu, and S. Supian, “Model Estimation of Claim Risk and Premium for Motor Vehicle Insurance by Using Bayesian Method”, IOP Conf. Series: Materials Science and Engineering, 300 (2018) 012027.
3. Riaman, Sukono, D. Susanti, E. Marbun, and A.T.B. Bon, “Net Premium Estimation by Using Forward Selection Linear Model for Motor Vehicle Insurance”, Proceedings of the International Conference on Industrial Engineering and Operations Management (IEOM), Bandung, Indonesia, March 6-8, 2018, 2711-2717.
4. S.A. Klugman, H.H. Panjer, and G.E. Willmot, “Loss Models: From Data to Decisions” New Jersey: John Wiley & Sons, Inc., 1998.
5. D.C.M. Dickson, “Insurance Risk and Ruin”, Cambridge: Cambridge University Press, 2005.
6. N.L. Bowers Jr., H.U. Gerber, J.C. Hickman, D.A. Jones, and C.J. Nesbitt, “Actuarial Mathematics”, Illinois: The Society of Actuaries, 1997.
7. A. Saputra, Sukono, and E. Rusyaman, “Risk Adjustment Model of Credit Life Insurance Using a Genetic Algorithm”, IOP Conf. Series: Materials Science and Engineering, 332 (2018) 012007.
8. P. Sidi, M. Mamat, Sukono, and S. Supian, “Evaluation Model for Risk Insurance Premiums of Building Damage Caused by Flood: Case Study in Citarum Watershed, Southern Bandung, Indonesia”, Journal of Engineering and Applied Sciences, 12(17), January 2017: 4420-4425.
9. Sukono, I. Aisah, Y.R.H. Tampubolon, H. Napitupulu, S. Supian, and P. Sidi, “Estimated Value of Insurance Premium Due to Citarum River Flood by Using Bayesian Method”, IOP Conf. Series: Materials Science and Engineering, 332 (2018) 012047.
10. W.M. Bolstad, “Introduction to Bayesian Statistics”, New Jersey: A John Wiley & Sons. Inc., 2007.
11. P.J. Northrop and N. Attalides. Statistica Sinica 26(2), 2016, pp. 721-743.
12. L.J. Bain and M. Engelhardt, “Introduction to Probability and Mathematical Statistics”, California: Duxbury Press, 1992.
13. J.O. Berger, “Statistical Decision Theory and Bayesian Analysis”, New York: Springer, 1985.
14. Sukono, M. Suyudi, F. Islamiyati, S. Supian, “Estimation Model of Life Insurance Claims Risk for Cancer Patients by Using Bayesian Method” IOP Conf. Series: Materials Science and Engineering, 166 (2017) 012022.
15. D. Gamerman and H.F. Lopes, “Markov Chain Monte Carlo”, London: Chapman & Hall, 2006.
16. M.A. Mohamed, A.M. Razali and N. Ismail, “Approximation of Aggregate Losses Using Simulation”, Journal of Mathematics and Statistics, 6 (3), 2010, pp. 233-239.
17. Sukono, J. Nahar, M. Mamat, F.T. Putri, and S. Supian, “Estimation of Outstanding Claims Reserving Based on Inflation Risk on Car Insurance Companies by Using the Bootstrap Method”, Far East Journal of Mathematical Sciences (FJMS), Volume 102, Number 4, 2017, pp. 687-706.
18. I. Ntzoufras, “Bayesian Modeling Using WinBUGS”, New Jersey: John Wiley & Sons, Inc., 2009.

AUTHORS PROFILE

Agung Prabowo is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman (UNSOED), Purwokerto with expertise in Number Theory, Statistics, Actuarial Sciences.

Mustafa Mamat is Professor at Universiti Sultan Zainal Abidin (UniSZA), Malaysia since 2013. His research interests include conjugate gradient methods, steepest descent methods, Broyden’s family and quasi-Newton methods..

Sukono is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Currently serves as Head of Master's Program in Mathematics, the field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Afif Amrullah Taufiq is a staff of Engineering Division, General Insurance of Bumiputera Muda 1967, Jl. Wolter Monginsidi No. 63, Kebayoran Baru, South Jakarta, Indonesia.