Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN: 3025-0803

Vol. 2, No. 4, pp. 161-168, 2024

Modeling of Car Insurance Premiums Using the Bayes Method with Poisson and Exponential Distributions

Nur Rizky Farida[1*], Siti Nur Hikmah[2]

[1,2]*Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, West Java, Indonesia*

**Corresponding author email: nur21003@mail.unpad.ac.id*

Abstract

Car insurance is included in the general insurance category. Determination of general insurance premiums can be done through various approaches, one of which is the Bayes method used in this study. The Poisson distribution is chosen to model the frequency of claims, while the Exponential distribution is used to model the amount of claims. Insurance premiums are calculated by multiplying the expected frequency of claims by the expected amount of claims. Based on the results of the analysis of car insurance data using the Bayes method, it was found that the highest premium rates were for Mercedes brand vehicles, while the lowest rates were for Saab brand vehicles. The results of this calculation can be used by insurance companies as a reference in managing car insurance reserve funds.

Keywords: Vehicle insurance, claim frequency, claim size, Bayes method, Poisson distribution, exponential distribution.

1. Introduction

Insurance is an agreement in which the insurer commits to the insured party, by receiving premium payments, to provide compensation for losses, damages, or loss of profits that may occur due to unforeseen events. This definition is contained in the Commercial Law Code (KUHD) of the Republic of Indonesia, Chapter 9 Article 246. Insurance is divided into two main categories, namely life insurance and general insurance. Life insurance aims to protect against financial losses arising from the death or life of an insured person. Meanwhile, general insurance provides protection against financial losses that occur due to damage or loss of assets. One form of general insurance is car insurance.

Car insurance is an increasingly important financial protection instrument, especially in developed countries like America, where vehicle ownership rates continue to increase every year. In 2023, car insurance in the United States will experience significant changes in terms of market dynamics and insurance premium prices. The car insurance industry in America continues to experience increasing pressure due to various factors, including inflation, technological developments, changes in consumer behavior, and increasing risks associated with traffic conditions and vehicle use. One of the main factors affecting the car insurance industry is the increasing number of claims, both in terms of frequency and amount of claims, which directly contributes to increasing premium costs.

In 2023, car insurance premiums in the United States will increase by an average of 8-10%, higher than in previous years. This increase is largely driven by the surge in vehicle repair costs, which have increased by more than 15% since 2022. Higher repair costs are driven by rising prices for parts and labor, as well as the increasing complexity of technology used in modern vehicles, such as sensors and automated driving systems that require specialized equipment and higher skills to repair. In addition, the number of insurance claims has also increased, causing insurance companies to raise premiums.

According to data from the Insurance Information Institute (III) and the National Association of Insurance Commissioners (NAIC), nearly 87% of vehicle owners in the United States have auto insurance policies. This reflects the high level of consumer awareness of the importance of insurance protection. Other factors that contribute to the increase in premiums include the increasing number of claims, damage due to natural disasters, and risk-based rate adjustments that are increasingly tailored to the individual profile of the policyholder, such as driving history, age, and geographic location.

To calculate the insurance premium price, there are several methods that can be used. In this study, the method used is the Bayes method, because this method allows the use of existing information to reduce estimation errors. The Poisson distribution is used to model how often insurance claims occur, while the Exponential distribution is used to model the size of the claim. The insurance premium is calculated by multiplying the estimated frequency of claims by the estimated size of the claim.

The compilation of the state of the art of this research topic was carried out by searching for articles related to vehicle insurance premiums for risks based on frequency and claim size. The articles were sourced from the Scopus and Science Direct databases. The article search was carried out using the keywords "Insurance", "Risk", "Premium", "frequency" and "claim" in 2021-2024. The data obtained from the database is in the form of meta data stored in RIS format.

Table 1: State Of The Art

Writer	Frequency and size of claims	Using the Bayes method	Using the Poisson distribution	Using Gamma distribution	Using exponential distribution	Using claim data
Boylan et al.,	YES	NO	YES	YES	NO	NO
Zhang et al.,	YES	YES	YES	NO	NO	YES
McDonnell et al.,	YES	YES	YES	YES	YES	YES
Shahroodi et al.,	YES	NO	YES	YES	NO	YES
Kester,	YES	NO	NO	NO	NO	NO
Purwandari et al.,	YES	NO	YES	YES	YES	NO
Boucher & Turcotte,	YES	NO	YES	NO	NO	YES
Ieosanurak et al.,	YES	NO	YES	YES	NO	YES
Xie & Shi,	YES	NO	NO	NO	NO	YES
Nobanee et al.,	YES	NO	YES	NO	NO	NO
Wilson et al.,	YES	YES	YES	NO	YES	YES
Jacob & Wu,	YES	NO	NO	YES	NO	YES
Safari-Katesari & Zaroudi,	YES	NO	NO	NO	NO	YES
Reiff et al.,	YES	NO	YES	NO	NO	YES

This study focuses on modeling car insurance premiums using a dataset obtained from Mendeley Data, which includes insurance claim records from various providers for several types of products, such as vehicle, property, and personal injury insurance. Based on the results of the state of the art study, various methods have been applied in insurance premium modeling, such as the Bayes method, Poisson distribution, and Gamma distribution, each of which has its own advantages in modeling the frequency and magnitude of claims.

The state of the art review shows that most studies use the Poisson distribution to model the claim frequency, and the Exponential or Gamma distribution to model the claim magnitude. Some authors such as Zhang et al. and McDonnell et al. also combine the Bayes method to improve the accuracy of premium estimation. Based on these results, this study will implement all the approaches identified in the state of the art table, including the Bayes method, Poisson distribution, and Gamma distribution, by utilizing the available claim data.

By using this approach, it is expected that through this modeling, insurance companies can determine more accurate premiums and in accordance with the risk profile of each customer. In addition, this model can also help reduce the risk of financial losses that companies may face due to unexpected claims. This study aims to contribute to the development of better premium determination methods, not only for vehicle insurance in Indonesia, but also for other insurance products.

2. Literatur Review

2.1. Insurance Claims

An insurance claim is a formal request from a policyholder to an insurance company for payment or benefits based on the terms and conditions of the insurance contract. Claims are usually filed after an incident that causes a loss, such as a car accident, vehicle theft, or vehicle damage.

According to Chen et al. (2020), the frequency of insurance claims is influenced by various external and internal factors. External factors can include road conditions, weather, and accident rates in an area. For example, drivers who frequently drive in areas with high accident risks or on damaged roads tend to file claims more often. On the other hand, internal factors such as driver behavior also play an important role. Drivers who are aggressive, careless, or frequently violate traffic rules have a higher risk of having accidents, which ultimately increases the frequency of claims.

Insurance claims not only help companies in setting premiums, but also in managing their reserve funds. By predicting the size of claims that may occur, insurance companies can prepare funds to face potential losses, so that they can still meet their obligations to policyholders in the future.

2.2. Premium Pricing Determination

Premium is an amount of money that must be paid by the policyholder to the insurance company in return for the protection provided in the insurance policy. Determining the premium price is one of the important components in the insurance industry because premiums are the main source of income for insurance companies. In setting premiums, companies must balance the need to make a profit with the need to provide adequate protection for policyholders (Diba et al., 2017).

Premium pricing depends on two main aspects:

- The claim amount is the amount of compensation or payment that must be issued by the insurance company to cover the losses experienced by the policyholder. The claim amount varies greatly depending on the type of insurance, the type of claim, and the nature of the loss that occurs. For example, in vehicle insurance, the claim amount can vary from minor damage to total loss due to a major accident.

- While claim frequency refers to how often the policyholder files a claim in a certain period. This claim frequency is very important in determining the premium because the more often the claim is filed, the higher the risk for the insurance company, which means the premium must also be adjusted.

To calculate the premium, data on the claim size and claim frequency are modeled in the form of a certain probability distribution. This is important because claims do not occur regularly and are often random, so a statistical approach is needed to predict the possibility of future claims. The premium rate is usually calculated by multiplying the expected claim size (the average of the total predicted loss) by the expected claim frequency (the number of claims expected to be filed). With this approach, insurance companies can anticipate possible losses and set premiums accordingly.

2.3. Premium Price Determination Using the Bayes Method

One approach that is often used to determine premium rates is the Bayes method. This method is very useful in situations where insurance companies have prior information about claims and want to update their knowledge based on new data received from claims that have occurred. The Bayes method allows the incorporation of this information into posterior information, which is then used to estimate more accurate premiums.

The first step in applying the Bayes method is to determine the prior distribution, which is the initial probability distribution of the parameters to be estimated. In the context of insurance claims, these parameters could be the frequency of claims and the size of claims. This prior distribution is then updated using the observed claims data through the likelihood function.

The distribution often used to model claim frequency is the Poisson distribution, because claims usually occur in random time intervals and the number of claims filed in one period is discrete. Meanwhile, to model the size of the claim, the Exponential distribution is used, which is suitable for describing the time or amount of loss that is random and continuous.

Once the prior and likelihood distributions are combined, the result is a posterior distribution, which provides a better estimate of the frequency of claims and the size of claims based on the data at hand. For example, if a company has information that the average number of claims in the past five years has been 10 per year, they can use this data to update their predictions about future claims.

The main advantage of the Bayesian method is its ability to continually update estimates based on new data. This means that as more claims come in, insurance companies can adjust their predictions and set premiums that are more in line with actual conditions.

3. Materials and Methods

3.1. Materials

This study uses a dataset taken from Mendeley Data. This dataset includes a complete collection of insurance claim records from various insurance providers covering several types of insurance products, such as vehicle insurance, property, and personal injury. The dataset was prepared through a data cleaning and processing process using Microsoft Excel software, which involved removing duplicate data, filling in missing data, and normalizing data formats to ensure consistency. Furthermore, the data was classified according to 14 different car brands, namely, Volkswagen, Toyota, Suburu, Nissan, Saab, Mercedes, Jeep, Honda, Ford, Dodge, Chevrolet, BMW, Accura, and Audi.

Table 2: Summary of insurance data according to car brand

Brand	Many police	Number of claims	Average claim frequency ($\bar{x}$)	Large number of claims
Volkswagen	68	19	0.279412	3458130
Toyota	70	13	0.185714	3256660
Subur	80	19	0.2375	4298410
Nissan	78	14	0.179487	4020530
Saab	80	18	0.225	4115630
Mercedes	65	22	0.338462	3404190
Jeep	67	11	0.164179	3457790
Honda	55	14	0.254545	2844320
Ford	72	22	0.30556	4073050
Dodge	80	20	0.25	4475550
Chevrolet	76	21	0.276315	4008740
BMW	72	20	0.277778	4025180
Accuracy	68	13	0.191176	3571280
Audi	69	21	0.304348	3752480

3.2. Methods

The Bayes method is applied in this study to calculate car insurance premiums based on the distribution of claim frequency and claim size. The claim frequency is modeled using the Poisson distribution, with prior parameters that follow the Gamma distribution. This is done to adjust the claim frequency with existing historical claim data. Meanwhile, the claim size is modeled using the Exponential distribution with prior parameters that also follow the Gamma distribution.

The prior and parameters for the Gamma distribution are determined based on historical data on claim frequencies, using the mean and variance of the data. Parameter determination can be done using the following formula,$\alpha\beta\alpha$

$$\alpha = \bar{x} \times \left(\frac{\bar{x} \times (1 - \bar{x})}{s^2} - 1 \right) \tag{1}$$

Where is the average of the claim frequencies from is its variance. The estimator for the parameter is calculated by the maximum likelihood estimation method using the formula,$\bar{x}s^2\beta$

$$\beta = (1 - \bar{x}) \times \left(\frac{\bar{x} \times (1 - \bar{x})}{s^2} - 1 \right) \tag{2}$$

Table 3 shows the best prior values for the claim frequency variable based on each of the studied car brands:

Table 3: Best priors for claim frequency variables

Brand	α	β
Volkswagen	1,067	2,753
Toyota	1,008	4,418
Subur	1,041	3,343
Nissan	1,008	4,608

Saab	1,033	3,559
Mercedes	1,127	2,203
Jeep	0.996	5,075
Honda	1,086	3,183
Ford	1,134	2,578
Dodge	1,051	3,151
Chevrolet	1,069	2,800
BMW	1,068	2,778
Accuracy	1,009	4,269
Audi	1,092	2,496

For the claim amount, the parameter presents the average or mean claim amount for each car brand. This value can be calculated using the average formula of all total claims that occur for the car brand. Parameter determination can be done using the following formula,uu

$$u = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For the variable presenting the large variance of claims, which measures how much the spread or deviation from the average value of the claim for the brand is. Parameter determination can be done with the following formula,vv

$$v = \frac{1}{n}\sum_{i=1}^{n} (x_i - u)^2$$

With this approach, the model can reflect the claim pattern more accurately, which is then used in the premium calculation. Table 4 below shows the best prior values for the claim size variable for each car brand:

Table 4: Best priors for large claim variables

Brand	u	v
Volkswagen	50854.85	708978288
Toyota	46523.71	754362678
Subur	53730.12	580663441
Nissan	51545.26	573381007
Saab	51445.38	648140323
Mercedes	52372.15	796362942
Jeep	51608.80	760375983
Honda	51714.91	720004977
Ford	56570.14	672569269
Dodge	55944.37	683495100
Chevrolet	52746.58	811867572
BMW	55905.28	705647752
Accuracy	52518.82	670132646
Audi	54383.77	760053012

3.2.1Claim Frequency Expectations

If the claim frequency is denoted by X, then

$$X_i|\ \theta \sim Poisson(\theta) iid\ ;\ \ i = 1,2,\dots,n \tag{3}$$

Sample information in the form of a likelihood function is denoted by . The likelihood function of X is,$x = f(x|\theta)$

$$f(x|\theta) = \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \tag{4}$$

While the prior information is in the form of a prior distribution for the parameter □. The prior distribution used is$\theta \sim Gamma(\alpha, \beta)$.

$$f(\theta) = \frac{\beta^{\alpha}\theta^{\alpha-1}e^{-\beta\theta}}{\Gamma(\alpha)} \quad (5)$$

Next, the posterior distribution will be searched using the Box Tiao method.$f(x|\theta)$

$$f(\theta|x) \propto f(x|\theta)f(\theta) \propto \theta^{\sum_{i=1}^{n} x_i}e^{-n\theta}\theta^{\alpha-1}e^{-\beta\theta} = \theta^{\alpha+\sum_{i=1}^{n} x_i-1}e^{-(\beta+n)\theta} \quad (6)$$

The above equation approximates the probability density function form of the Gamma distribution with parameters$(\alpha + \sum_{i=1}^{n} x_i, \beta + n)$.

$$\theta|x \sim Gamma\left(\alpha + \sum_{i=1}^{n} x_i \ , \beta + n\right) \quad (7)$$

3.2.2 Big Claim Expectations

If the claim amount is denoted by □, then

$$Y_i|\lambda \sim Exp(\lambda) iid \ ; \ i = 1,2,..,n \quad (8)$$

Sample information in the form of a likelihood function is denoted by . The likelihood function of □ is$f(y|\theta)$

$$f(y|\lambda) = \lambda^{n}e^{-\lambda\sum_{i=1}^{n} y_i} \quad (9)$$

While the prior information is in the form of a prior distribution for the parameter □. The prior distribution used is$\lambda \sim Gamma(u, v)$.

$$f(\lambda) = \frac{v^{u}\lambda^{u-1}e^{-\lambda v}}{\Gamma(u)} \quad (10)$$

Next, the posterior distribution of □(□|□) will be searched using the Box Tiao method.

$$f(\lambda|y) \propto f(y|\lambda)f(\lambda) \propto \lambda^{n}e^{-\lambda\sum_{i=1}^{n} y_i}\lambda^{u-1}e^{-\lambda v} = \lambda^{u+n-1}e^{-\lambda v} = \lambda^{u+n-1}e^{-(v+\sum_{i=1}^{n} y_i)\lambda} \quad (11)$$

The above equation approximates the probability density function form of the Gamma distribution with parameters$(u + n, v + \sum_{i=1}^{n} y_i)$.

$$\lambda|y \sim Gamma\left(u + n, v + \sum_{i=1}^{n} y_i\right) \quad (12)$$

The following table shows premium rates based on expected claim frequency and size:

Table 5: Premium rates based on car brand

Brand	Claim Frequency	Big Claim	Premium
Volkswagen	0.2836	715,494	202914,098
Toyota	0.1882	615,666	115868,341
Subur	0.2405	921,855	221706,128
Nissan	0.1817	895,809	162768,495
Saab	0.2278	79,148	18029,9144
Mercedes	0.3441	656.108	225766,763
Jeep	0.1665	677,049	112728,659
Honda	0.2593	71,619	18570,8067
Ford	0.3102	83,711	25967,1522
Dodge	0.2532	81,343	20596,0476
Chevrolet	0.2801	647,434	181346,263
BMW	0.2817	788,776	222198,199
Accuracy	0.1938	780,563	151273,109
Audi	0.3089	81,291	25110,789

Based on Table 5, it can be seen that the lowest premium is 18029 USD (Saab) and the highest premium is 18029 USD (Saab).225766 USD (Mercedes). This result is consistent with historical data for both brands. The Saab brand has the smallest claim size compared to other brands, while the Mercedes brand has the highest claim frequency compared to other brands.

4. Result and Discussion

4.1. Result

Based on calculations using the Bayesian method with Poisson and Exponential distributions, it was found that the highest premiums are for Mercedes-brand vehicles, while the lowest premiums are for Saab-brand vehicles. This result is due to the higher claim frequency associated with Mercedes vehicles and relatively lower claim amounts for Saab. Using a dataset that includes 14 different car brands, the calculations show a significant variation in premiums across brands, influenced by each brand's claim frequency and claim size.

The premium calculation results indicate that the Bayesian method, utilizing the Poisson distribution for claim frequency and the Exponential distribution for claim size, can generate more accurate and relevant premium estimates aligned with the risk profile of each car brand.

4.2 Discussion

The results of this study indicate that the Bayesian method is an effective approach for determining motor vehicle insurance premiums, as it allows for the integration of historical information (prior) with current claim data (likelihood). This combination produces a posterior distribution that provides better estimates for claim frequency and size than conventional methods that do not consider data updates.

The choice of Poisson and Exponential distributions to model claims aligns with the discrete nature of claim frequency and the continuous nature of claim amounts, typical of insurance data. Additionally, the findings show that certain car brands with higher claim risks have higher premiums, supporting the insurance principle of allocating premium costs based on risk profile.

This study can serve as a reference for insurance companies in setting fairer and more accurate premium rates and aid in managing insurance reserve funds.

5. Conclusion

The Bayes method can be used to calculate motor vehicle insurance premiums by finding the posterior distribution of claim frequency and claim size. The premium is calculated by multiplying the expected claim frequency by the expected claim size obtained from the posterior distribution. Choosing the right prior is essential in the Bayes method to obtain optimal results.

Based on data analysis, the results showed that the lowest car insurance premium was for the Saab brand and the highest premium was for the Hyundai brand.Mercedes.The results of this premium calculation can be used as a consideration by insurance companies in managing vehicle insurance reserve funds.

References

Boucher, J. P., & Turcotte, R. (2020). A longitudinal analysis of the impact of distance driven on the probability of car accidents. Risks, 8(3), 1–19. https://doi.org/10.3390/risks8030091

Boylan, J., Meyer, D., & Chen, W.S. (2024). A systematic review of the use of in-vehicle telematics in monitoring driving behavior. Accident Analysis and Prevention, 199. https://doi.org/10.1016/j.aap.2024.107519

Diba, F., Saepudin, D., & Rohmawati, AA (2017). Modeling and Simulation of Insurance Company Bankruptcy Probability with Exponentially Distributed Premium Value and Claim Size Analysis.

Ieosanurak, W., Khomkham, B., & Moumeesri, A. (2023). CLAIM MODELING AND INSURANCE PREMIUM PRICING UNDER A BONUS–MALUS SYSTEM IN MOTOR INSURANCE. International Journal of Applied Mathematics and Computer Science, 33(4), 637–650. https://doi.org/10.34768/amcs-2023-0045

Jacob, A., & Wu, Z. (2020). An Alternative Pricing System through Bayesian Estimate and Method of Moments in Bonus-Malus Framework for the Ghanaian Auto Insurance Market. Journal of Risk and Financial Management, 13(7). https://doi.org/10.3390/jrfm13070143

Kania, DS (2022). Calculation of Motor Vehicle Insurance Risk Premiums Based on Claim Frequency and Amount Data. Journal of Statistical Research, 111–118. https://doi.org/10.29313/jrs.v2i2.1295

Kester, J. (2022). Insuring future automobility: A qualitative discussion of British and Dutch car insurers' responses to connected and automated vehicles. Research in Transportation Business and Management, 45. https://doi.org/10.1016/j.rtbm.2022.100903

Kireina Waha, C., Rindengan, A.J., Manurung, T., Kunci, K., Premium, :, Binomial, M., Log-Logistic, N., Claim, B., Claim, B.,

& Car, A. (2019). Distribution Model of Car Insurance Claim Data to Determine Pure Premium. https://ejournal.unsrat.ac.id/index.php/decartesian

McDonnell, K., Murphy, F., Sheehan, B., Masello, L., & Castignani, G. (2023). Deep learning in insurance: Accuracy and model interpretability using TabNet. Expert Systems with Applications, 217. https://doi.org/10.1016/j.eswa.2023.119543

Nobanee, H., Elsaied, FA, Noor, HA, Resiq, TR ed A., Ahmed, FA, & Ismail, DA (2023). A Bibliometric Analysis of Auto Insurance: Current Status, Development, and Future Research Directions. Journal of Systems and Management Sciences, 13(2), 384–408. https://doi.org/10.33168/JSMS.2023.0227

Purwandari, T., Hidayat, Y., Sukono, Kalfin, Ibrahim, RA, & Subiyanto. (2024). Development of the Black–Scholes Model for Determining Insurance Premiums to Mitigate the Risk of Disaster Losses Using the Principles of Mutual Cooperation and Regional Economic Growth. Risks, 12(7). https://doi.org/10.3390/risks12070110

Putri, AA, Suryono, A., & Id, C. (2017). STUDY OF REPORTING PROCEDURES AND HANDLING OF TOTAL LOSS ONLY MOTOR VEHICLE INSURANCE CLAIMS (Study at PT Asuransi Jasa Indonesia {Persero} Surakarta Branch): Vol. V (Issue 2). http://www.jasindo.

Reiff, M., Šoltés, E., Komara, S., Šoltésová, T., & Zelinová, S. (2022). Segmentation and estimation of claim severity in motor third-party liability insurance through contrast analysis. Equilibrium. Quarterly Journal of Economics and Economic Policy, 17(3), 803–842. https://doi.org/10.24136/eq.2022.028

Safari-Katesari, H., & Zaroudi, S. (2020). Count copula regression model using generalized beta distribution of the second kind. Statistics in Transition New Series, 21(2), 1–12. https://doi.org/10.21307/STATTRANS-2020-011

Shahroodi, K., Avakh Darestani, S., Soltani, S., & Eisazadeh Saravani, A. (2024). Developing strategies to retain organizational insurers using a clustering technique: Evidence from the insurance industry. Technological Forecasting and Social Change, 201. https://doi.org/10.1016/j.techfore.2024.123217

Wilson, A. A., Nehme, A., Dhyani, A., & Mahbub, K. (2024). A Comparison of Generalized Linear Modeling with Machine Learning Approaches for Predicting Loss Costs in Motor Insurance. Risks, 12(4). https://doi.org/10.3390/risks12040062

Xie, S., & Shi, K. (2023). Generalized Additive Modeling of Auto Insurance Data with Territory Design: A Rate Regulation Perspective. Mathematics, 11(2). https://doi.org/10.3390/math11020334

Zhang, Y., Ji, L., Aivaliotis, G., & Taylor, C. (2024). Bayesian CART models for insurance claims frequency. Insurance: Mathematics and Economics, 114, 108–131. https://doi.org/10.1016/j.insmatheco.2023.11.005