International Journal of Innovative Mathematics, Statistics & Energy Policies
12(2):60-71, April-June, 2024

 www.seahipaj.org **ISSN: 2467-852X**

A Bayesian Approach to Weibull Distribution with Application to Insurance Claims Data

Nuraddeen Yusuf Adamu

Department of Mathematics,
Isa Kaita College of Education, Dutsin-Ma, Katsina State, Nigeria
Email: ***anisultan11@gmail.com***

ABSTRACT
This study investigates the application of statistical distributions in actuarial science and financial risk management, emphasizing their role in modeling and fitting various data sets to assess risk exposure on investments. It highlights the importance of key risk indicators derived from these distributions in evaluating companies' susceptibility to risks stemming from fluctuations in underlying variables such as equity prices, interest rates, and exchange rates. The Weibull distribution emerges as a prevalent model, particularly in forecasting stock price movements and predicting uncertainties. The study proposes a Bayesian approach to modeling Weibull distribution parameters, employing gamma priors, and validates the proposed distribution of claim amounts through computational analysis grounded in actuarial measures. Comparisons with traditional methods like maximum likelihood estimation (MLE) and simulated annealing algorithm (SA) reveal the superior performance of Bayesian estimators in capturing underlying distributions. Real-world actuarial data sets further demonstrate the practical applicability of the proposed model in effectively modeling insurance claim data. Overall, this study underscores the crucial role of statistical distributions in mitigating financial risks and enhancing decision-making processes in actuarial science and financial risk management.
Keywords: Bayesian approach; Weibull distribution; Insurance claims; Data analysis; Actuarial science; Financial risk management

1. INTRODUCTION
The modeling of uncertain events, such as fluctuations in stock prices, predicting market demand for commodities, and estimating insurance claims, is a widely studied area in business, finance, and economics. Selecting the appropriate statistical distribution is crucial for accurately representing data in these fields. In actuarial science, choosing the right distribution for modeling uncertainty and forecasting insurance claims is essential for assessing market risk exposure in investment portfolios. Over the years, extensive research has been conducted to identify suitable statistical distributions for precise forecasting. Early contributions, such as Ter Berg's work on Gamma and Poisson distribution models (1980), laid the groundwork for subsequent studies. Researchers like [1] further expanded this area with investigations into distributions such as the Poisson inverse Gaussian, [2], and generalized Poisson distributions [3]. Additionally, efforts have been made to develop approximations for aggregate claim distributions [4], [5].

1.1. Problem Statement
In accordance with industrial trends, the expansion of insurance firms has been substantial, with a corresponding surge in demand for insurance coverage among both businesses and households [6]–[8]. Consequently, insurance companies are adapting to meet evolving needs such as policy requirements, risk management, investments, and claims processing. Ensuring timely reimbursement is paramount, as policyholders rely on insurance contracts to mitigate financial losses. Accurately estimating expected claims facilitates informed decision-making regarding investment strategies and claims settlements, thereby addressing these challenges. This study aims to explore probability models based on the Weibull distribution for various insurance claims.

1.2. Objectives of the Study

The primary objective is to investigate a Bayesian approach to modeling the Weibull distribution, assuming a gamma prior for the shape parameter. This involves adequately modeling the occurrence of claims under insurance policies to estimate their expected number. Notably, previous studies have not explored modeling insurance claim data using the Weibull distribution with shape parameters following a gamma prior. Specific objectives include:

i) Developing an insurance claims model assuming a Weibull distribution with a shape parameter governed by a gamma distribution.
ii) Applying Bayesian modeling to estimate model parameters.
iii) Comparing the performance of Bayesian estimation with maximum likelihood and simulated annealing algorithms using insurance claims data.

2. LITERATURE REVIEW

Statistical distributions play a critical role in modeling various datasets, particularly in finance and actuarial science, where they help assess companies' exposure to risk stemming from market fluctuations. The selection of appropriate statistical distributions for modeling uncertainty and forecasting insurance claims is crucial in evaluating market risk exposure. Extensive research has been conducted over the years, including early contributions by [9] on Gamma and Poisson distribution models. Subsequent studies by [1] explored distributions such as the Poisson inverse Gaussian and the Delaporte distribution, while others focused on generalized Poisson distributions [3], the Poisson-Goncharov distribution [10], and approximations for aggregate claim distributions [4],[5]. Recent research has investigated various mathematical and statistical distribution models for insurance claims, such as Aggregate Claims Amount Probability Distributions [11], modeling Severity and Frequency of Auto Insurance Claims [12], and Bayesian approaches in estimating insurance claim uncertainties [13]–[17]. However, many studies have primarily focused on posterior parameter estimation rather than distribution hypotheses testing or estimating insurance claim uncertainties. Additionally, there have been studies on modified Weibull distributions for internal rate of return modeling [18].

3. METHODOLOGY

3.1. Estimating Expected Insurance Claims Return

The study employed a Bayesian approach to estimate the expected return of insurance claims using a posterior distribution. This involved integrating the likelihood function of the best-fit model for claims return amounts with a Gamma prior distribution. The expected insurance claims amount was then determined from the expectation of the posterior distribution.

3.2. Weibull Distribution Model

Let X represent the insurance claims amounts. Assuming X follows a Weibull distribution denoted as X~Weibull, the probability density function (PDF) is defined as:

$$f_{Weibull}(x_t|,\alpha,\beta) = \begin{cases} \beta\alpha {x_t}^{(\beta-1)} e^{-(\alpha x_t)^{\beta}}, & x_t > 0 \\ 0 & , x_t \leq 0 \end{cases} \quad (1)$$

The cumulative distribution function (CDF) of Equation (1) is:

$$F_{Weibull}(x_t,|\alpha,\beta) = P(X_t \leq x_t) \begin{cases} 1 - e^{-(\alpha x_t)^{\beta}}, & x_t > 0 \\ 0 & , x_t \leq 0 \end{cases} \quad (2)$$

Here, β represents the shape parameter determining the curve's slope, and α is the scale parameter representing the characteristic life of the Weibull distribution.

The expected value and variance of insurance claims data following the Weibull distribution are given by:

$$E(X) = \alpha\Gamma\left(1 + \frac{1}{\beta}\right) \qquad (3)$$

$$Var(X) = \alpha^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{2}{\beta}\right)\right] \qquad (4)$$

where β is the shape, α is the scale parameter and Γ is the gamma function.

3.3. Determining Likelihood Function

The maximum likelihood function is commonly used for parameter estimation. Suppose. Suppose that Let $X_t, t = 1,2, \dots, n$ are independent and identically distributed random variables following a Weibull distribution $W(\beta, \alpha)$, where the distribution parameters β and α are unknown. The likelihood function $L(\alpha, \beta | x_1, x_2, \dots, x_n)$ is built from Equation (1) as

$$L(\alpha, \beta | x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i |, \alpha, \beta) = \prod_{i=1}^{n} \beta\alpha x_i^{(\beta-1)} e^{-(\alpha x_t)^\beta} \qquad (5)$$

Simplifying the likelihood function in (5), gives

$$L(\alpha, \beta | x_1, x_2, \dots, x_n) = \alpha^n \beta^n \prod_{t=1}^{n} x_t^{(\beta-1)} e^{\left(-\alpha \sum_{i=1}^{n} x_t^\beta\right)} \qquad (6)$$

The log-likelihood function is obtained by taking the natural logarithm of Equation (6):

$$\log(L) = n\log(\beta) + n\log(\alpha) + (\beta - 1)\sum_{t=1}^{n} \log(x_t) - \alpha \sum_{t=1}^{n} \log(x_t)^\beta \qquad (7)$$

By differentiating Equation (7) with respect to scale and shape parameters and equating to zero, gives:

$$\frac{\partial \log(L)}{\partial \beta} = \frac{n}{\alpha} + \sum_{t=1}^{n} x_t^\beta = 0 \qquad (8)$$

$$\frac{\partial \log(L)}{\partial \beta} = \left(\frac{n}{\alpha}\right) + \sum_{t=1}^{n} (\log x_t) - \alpha \sum_{t=1}^{n} \left(\log x_t^\beta\right) = 0 \qquad (9)$$

The maximum likelihood estimators (MLE) of parameters are obtained by solving these equations. To optimize the log-likelihood function, methods like nonlinear optimization or simulated annealing algorithm can be used. The estimator for the scale parameter $(\hat{\alpha})$

$$\hat{\alpha} = \frac{n}{\sum_{t=1}^{n} x_t^{\beta}} \tag{10}$$

The estimator of shape parameter can be obtained by substituting Equation (9) in Equation (10) as follows:

$$n + \hat{\beta}\sum_{t=1}^{n} \log x_t = \frac{n\hat{\beta}\sum_{t=1}^{n} x_t^{\beta} \log x_t}{\sum_{t=1}^{n} x_t^{\beta}} \tag{11}$$

Equation (11) can be solved either through numerical methods or by utilizing an Excel spreadsheet package to acquire the estimated value of $\hat{\beta}$. Once $\hat{\beta}$ is obtained, $\hat{\alpha}$can be easily derived. In this research, we have utilized a simulated annealing approach to determine the estimated value of $\hat{\beta}$.

3.4. Bayesian Estimation Method

The Bayesian approach has gained considerable attention across disciplines for analyzing lifetime data, serving as an alternative to traditional maximum likelihood methods. Originating from Thomas Bayes, this method relies on Bayes' theorem, offering a straightforward mechanism for updating probabilities with new information. In Bayesian modeling, observed data acts as new information, allowing for the revision of prior assumptions about parameters of interest, which are treated as random variables. When estimating parameters using data $x_1, x_2, \dots, x_n$ for a statistical model defined by the probability (density) function $P(x_t|\xi)$ parameters are viewed as random variables in Bayesian analysis and are assigned distributions. If prior knowledge about parameters is unavailable, non-informative prior distributions can be employed. In this study, we adopt a Gamma prior for shape parameters, with no specific prior assumed for scale parameters.

3.5. Prior Assumptions on Insurance Claim Amount Data

It's worth noting that if the distribution's shape parameter is known, the scale parameter will possess a conjugate prior distribution, which in this case is a gamma prior. In situations where both distribution parameters are unknown, they lack conjugate priors. Initially, we consider the known shape parameter of the Weibull distribution. Therefore, we set priors on λ and ψ. The prior for λ is a gamma distribution with parameters ψ and λ, denoted as Gamma(ψ, λ), with the probability density function (PDF) presented in Equation (12). The posterior distribution is determined by multiplying the likelihood function of insurance claims amounts $x_1, x_2, \dots, x_n$by the prior distribution under the Bayesian approach. In this scenario, the prior distribution of the scale parameter is assumed to follow a Gamma distribution with the PDF as described in Equation (12).

3.6. Estimating Expected Future Insurance Claims Amounts from the Posterior Distribution

When the likelihood is based on insurance claims amounts, obtaining the Bayes estimator involves multiplying the prior distributions with the likelihood function. This results in a posterior distribution from which expected future insurance claims amounts can be estimated. The marginal distribution of parameters given insurance claim data is found by integrating over both parameters:

$$P(x_t) = \int_0^{\infty} P(x_t, \beta, \alpha) \tag{12}$$

This integration simplifies to a form involving a new variable ρ, defined as the product of α and the sum of β and λ:

$$\rho = \alpha\left(\sum_{t=1}^{n} x_t^{\beta} + \lambda\right) \quad (13)$$

The likelihood function is then proportional to the marginal function, leading to the derivation of the joint posterior distribution of the two parameters:

$$P(\alpha|x_t) = \frac{P(x_t,\beta,\alpha)}{P(x_t)} \quad (14)$$

This distribution resembles a Gamma distribution:

$$\frac{\alpha^{n+\psi-1}}{\Gamma(n+\psi)} e^{-\alpha\left(\sum_{t=1}^{n} x_t^{\beta}\right)}\left(\sum_{t=1}^{n} x_t^{\beta} + \lambda\right)^{n+\psi} \sim Gamma\left(n+\psi, \sum_{t=1}^{n} x_t^{\beta} + \lambda\right) \quad (15)$$

From this, the Bayes estimate under the squared error loss function is calculated:

$$\hat{\alpha} = E(\alpha) = \frac{n+\psi}{\sum_{t=1}^{n} x_t^{\beta} + \lambda} \quad (16)$$

It's worth noting that the distributions obtained for both parameters do not match known distributions, and closed-form solutions are not attainable. Therefore, parameter estimations under the quadratic loss function rely on powerful optimization techniques. Various methods can be employed to approximate these estimations. In this study, the comparison focuses on maximum likelihood and simulated annealing algorithms with Bayes estimators, using RMSE and MAE as measuring criteria.

In this section, the behavior of Bayesian estimators within a finite sample size n is explored. The effectiveness of these estimators is compared with the Simulated Annealing algorithm and Maximum Likelihood estimators for determining the shape and scale parameters of the Weibull distribution, focusing on error accumulation criteria. Assumptions are made regarding the distribution of shape parameters, which are assumed to follow Gamma priors. The simulations, executed in MAT LAB programming by the author, involve generating random data from the Weibull distribution to simulate insurance claim amounts. The simulation proceeds through these steps:

i. Samples of varying sizes, ranging from $n = 10$ to $n = 1000$, are generated from the insurance claim distribution.
ii. Maximum Likelihood Estimates (MLE) are computed for the proposed model parameters.
iii. We estimate Weibull parameters for each sample size, considering different values of $\beta = 1$ and α, such as 1, 2, and 3.
iv. Evaluation metrics such as Root Mean Squared Errors (RMSEs) and Mean Absolute Errors (MAEs) are then computed for assessment.

3.7. Evaluation Metrics

To assess the performance of the Bayesian approach, a comparison will be made with the Simulated Annealing algorithm and Maximum Likelihood methods using error accumulation metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These metrics are defined as follows:

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{\left(\beta_t - \widehat{\beta_t}\right)^2 + \left(\alpha_t - \widehat{\alpha_t}\right)^2}{n}} \quad (17)$$

$$MAE = \frac{\sum_{t=1}^{n} \left|\beta_t - \widehat{\beta_t}\right| + \left|\alpha_t - \widehat{\alpha_t}\right|}{n} \quad (18)$$

Here, β_t and $\alpha_t$represent the exact values of the shape and scale parameters, respectively, while $\widehat{\beta_t}$ and $\widehat{\alpha_t}$ denote the estimated values of these parameters. The method with lower error accumulation, as indicated by RMSE and MAE, is considered the best fit model for the insurance claims dataset.

4. Results and Discussion

Tables 1 and 2 present the simulation results. Figures 1 through 4 illustrate the RMSE for different estimation methods of the Weibull distribution. Additionally, Figures 6 through 8 depict the MAE for various methods of the Weibull distribution across different values of shape and scale parameters, considering various sample sizes $n = 20, 60, 120, 180, 240, 300$

Table 1: Estimators for Parameters of the Weibull Distribution ($\beta = 1$ and α, such as 1, 2, and 3.)

β_t	α_t	n	$\widehat{\beta_t}$ (BE)	$\widehat{\alpha_t}$ (BE)	$\widehat{\beta_t}$ (MLE)	$\widehat{\alpha_t}$ (MLE)	$\widehat{\beta_t}$ (SA)	$\widehat{\alpha_t}$ (SA)
1	1	20	1.0101	1.0473	1.1001	1.0570	1.1001	1.0570
1	1	60	0.007	0.001	0.006	0.001	0.005	0.0008
1	1	120	0.006	0.001	0.005	0.0008	0.004	0.0007
1	1	180	0.006	0.0008	0.005	0.0007	0.004	0.0006
1	1	240	0.005	0.0007	0.004	0.0006	0.004	0.0005
1	1	300	0.005	0.0006	0.004	0.0005	0.003	0.0004
1	2	20	0.008	0.001	0.007	0.001	0.007	0.001
1	2	60	0.006	0.0015	0.006	0.001	0.005	0.0008
1	2	120	0.005	0.0008	0.004	0.0007	0.004	0.0007
1	2	180	0.005	0.0008	0.004	0.0007	0.004	0.0006
1	2	240	0.005	0.0007	0.004	0.0006	0.003	0.0005
1	2	300	0.004	0.0006	0.004	0.0005	0.003	0.0004
1	3	20	0.012	0.002	0.009	0.0015	0.008	0.001
1	3	60	0.011	0.0025	0.009	0.002	0.007	0.001
1	3	120	0.009	0.0025	0.007	0.001	0.006	0.0008
1	3	180	0.008	0.002	0.007	0.0015	0.005	0.0008
1	3	240	0.007	0.0018	0.007	0.001	0.006	0.001
1	3	300	0.006	0.0017	0.006	0.001	0.004	0.0008

Table 2: RMSE and MAE for Weibull Distribution Estimation Methods

β_t	α_t	n	RMSE (BE)	MAE (BE)	RMSE (MLE)	MAE (MLE)	RMSE (SA)	MAE (SA)
1	1	20	0.013	0.002	0.010	0.002	0.008	0.001
1	1	60	0.009	0.0015	0.007	0.001	0.006	0.001
1	1	120	0.007	0.001	0.006	0.0008	0.005	0.0008
1	1	180	0.006	0.0008	0.005	0.0007	0.004	0.0006
1	1	240	0.005	0.0007	0.004	0.0006	0.004	0.0005
1	1	300	0.005	0.0006	0.004	0.0005	0.003	0.0004
1	2	20	0.009	0.002	0.008	0.001	0.007	0.001
1	2	60	0.007	0.0015	0.006	0.001	0.005	0.0008
1	2	120	0.006	0.001	0.005	0.0008	0.004	0.0007
1	2	180	0.005	0.0008	0.004	0.0007	0.004	0.0006
1	2	240	0.005	0.0007	0.004	0.0006	0.003	0.0005
1	2	300	0.004	0.0006	0.004	0.0005	0.003	0.0004
1	3	20	0.018	0.004	0.015	0.003	0.012	0.002
1	3	60	0.015	0.0035	0.013	0.0025	0.011	0.002
1	3	120	0.013	0.0025	0.011	0.002	0.009	0.0015
1	3	180	0.011	0.002	0.009	0.0015	0.008	0.001
1	3	240	0.010	0.0018	0.008	0.0013	0.007	0.001
1	3	300	0.009	0.0017	0.007	0.001	0.006	0.0008

The simulation study results obtained in Tables 1 and 2 show that the Bayesian methods yield similar results to the Maximum Likelihood method, particularly when the sample size is small. In such cases, Bayesian estimates tend to have a better fit, followed by Maximum Likelihood, while the Simulated Annealing algorithm exhibits poorer performance. However, as the sample size increases, the Simulated Annealing algorithm performs better, surpassing the Maximum Likelihood method. Notably, as the sample size increases, the accumulation of errors (RMSE and MAE) decreases across all methods.

For small sample sizes, both Bayesian methods and Maximum Likelihood tend to produce better fit estimates compared to Simulated Annealing. However, with larger sample sizes, the Simulated Annealing algorithm outperforms traditional Bayesian and Maximum Likelihood methods in estimating the Weibull distribution parameters based on the Insurance claims dataset. Despite this, the Bayesian approach is typically preferred due to its ability to consider more levels of variability in the model and provide entire posterior distributions for parameters, enabling additional analyses such as predictive distributions of claim amounts.

Figures 1 to 8 depict the results of RMSE and MAE reported in Table 2 for various sample sizes. It is evident that RMSE and MAE accumulation in Bayesian methods closely align with Maximum Likelihood when the sample size is small. However, the error accumulation in the Simulated Annealing algorithm decreases significantly with increasing sample size. Overall, as sample size increases, all methods show a reduction in RMSE and MAE values.

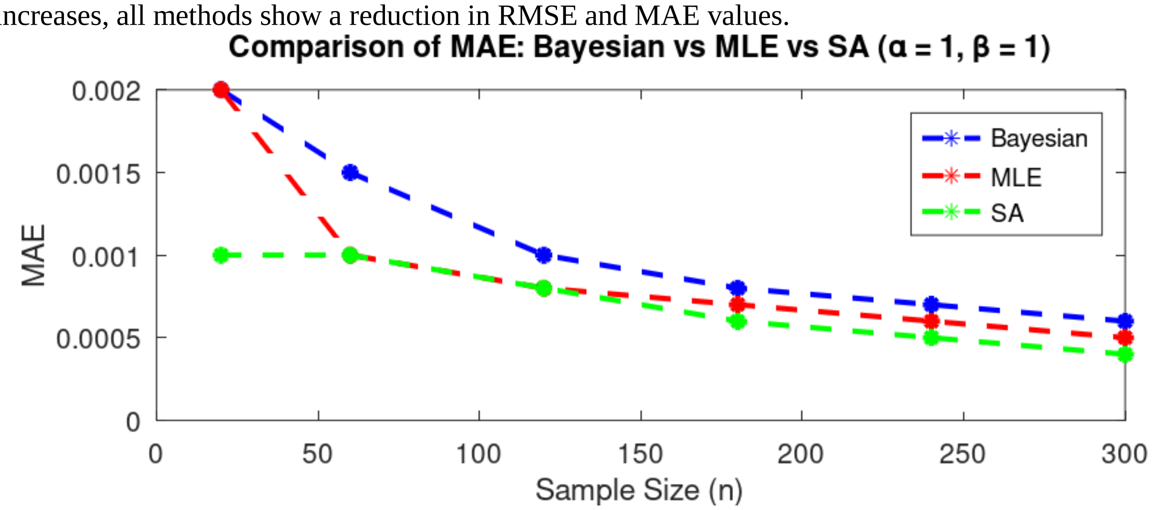

Fig. 1.: Comparison of MAE for $\beta = 1$ and $\alpha = 1$

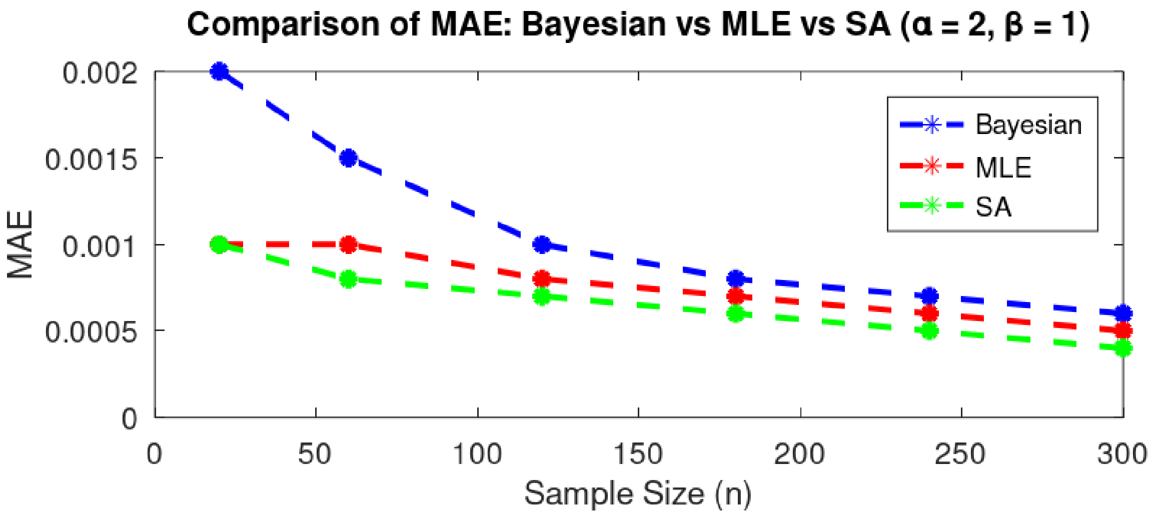

Fig. 2.: Comparison of MAE for $\beta = 1$ and $\alpha = 2$

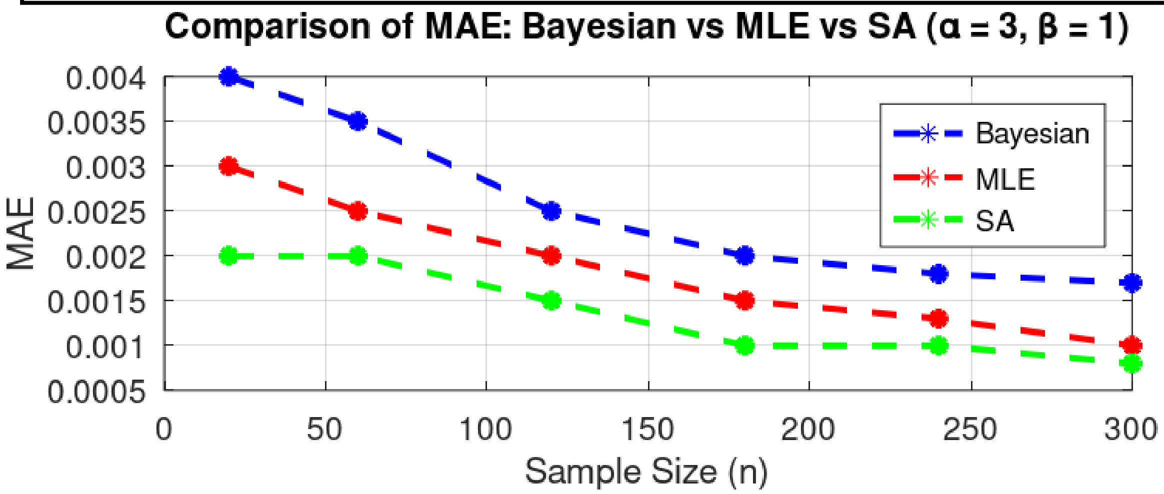

Fig. 3.: Comparison of MAE for $\beta = 1$ and $\alpha = 3$

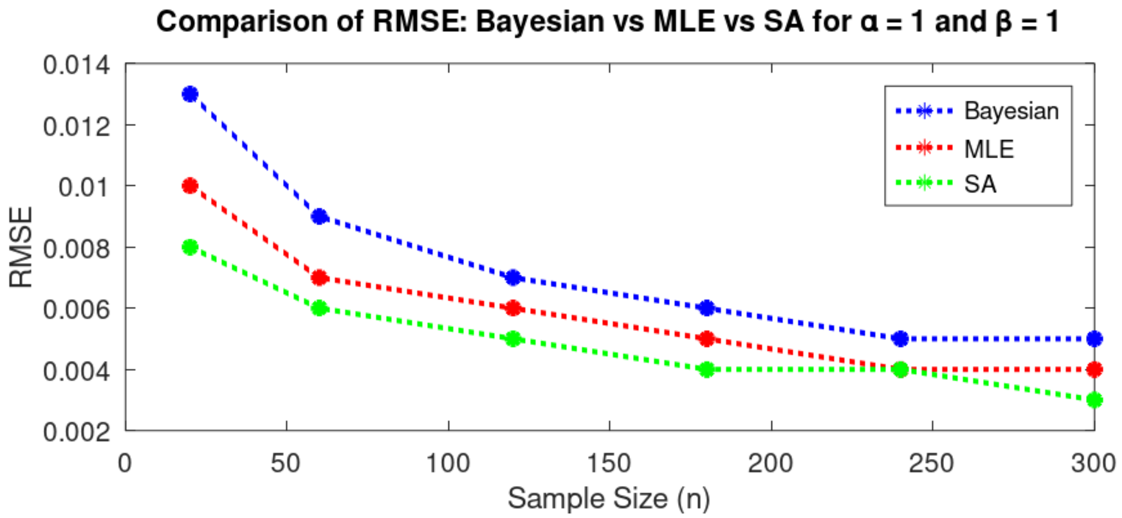

Fig. 4.: Comparison of RMEE for $\beta = 1$ and $\alpha = 1$

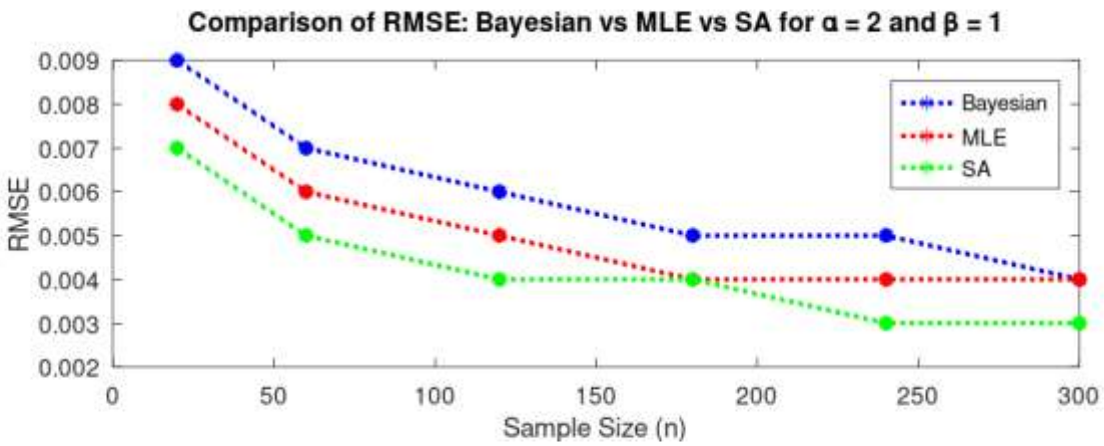

Fig. 5.: Comparison of RMEE for $\beta = 1$ and $\alpha = 2$

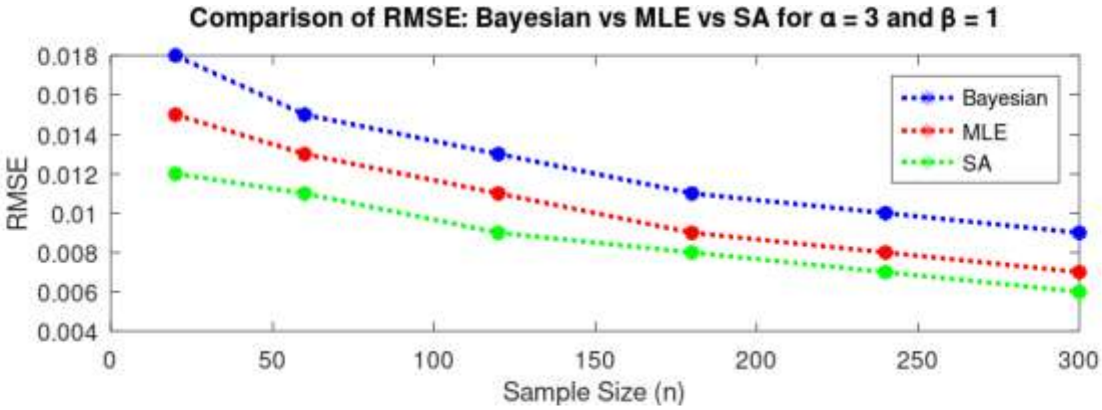

Fig. 6.: Comparison of RMEE for $\beta = 1$ and $\alpha = 3$

In conclusion, the study suggests that Bayesian methods, Maximum Likelihood estimation, and Simulated Annealing algorithm perform comparably in estimating the parameters of the Weibull distribution. As the sample size increases, all methods exhibit a decrease in RMSE and MAE values, with the Bayesian estimation approach offering a viable alternative, particularly for small sample sizes.

4.1. Actuarial Risk Measures

Value-at-Risk (VaR) serves as a statistical measure indicating the potential loss of market value for a financial asset over a specified time horizon, with a predetermined tolerance level. Alternatively, it can represent the maximum loss an individual might face within a certain timeframe under normal market conditions and within a predefined tolerance level (Bello et al., 2020; De Luca et al., 2020; Filippi et al., 2020; Molino and Sala, 2021). Mathematically, VaR is expressed as:

$$VaR_\alpha(X) = inf\{x|F_X(x) \geq \alpha\} = F_X{}^{-1}(\alpha) \qquad (18)$$

where $F_X{}^{-1}(\alpha)$ is the quantile function of random variable X, and α is the level of the quantile of the probability distribution of this random variable. Here, X represents the return of a financial asset at time t. Risk exposure, described using probability distributions, is a crucial aspect of actuarial and financial theory. Actuaries and financial risk managers frequently use key risk indicators to assess the degree of exposure to specific risks posed by changes in underlying factors such as stock prices, equity prices, interest rates, and currency exchange rates (Afify et al., 2020)

4.1.1. Value at Risk Measure of Claims Weibull Distribution

VaR, applied to claims modeled with the Weibull distribution, serves as a statistical tool to estimate the expected loss over a specific period for a particular stock or portfolio, given a confidence level. It's defined as:

$$P(X > \pi_\alpha) = 1 - \alpha \qquad (19)$$

where $\pi_\alpha = F^{-1}(\alpha), \alpha \in (0,1)$ and F is the cumulative distribution function (CDF) of the Weibull distribution.

4.1.2. Tail Value at Risk Measure

The Tail Value at Risk (TVaR) or Expected Shortfall quantifies the expected loss beyond a certain probability threshold. It represents the expected value of losses exceeding the VaR threshold and is defined as:

$$TVaR_\alpha(X) = E[X|X > \pi_\alpha] = \frac{\int_{x=\pi_\alpha}^{\infty} xf(x)dx}{1-F(\pi_\alpha)} \qquad (20)$$

4.1.3. An Application to Insurance Data

This study utilizes car accident data from 500 incidents recorded by one of Turkey's largest insurance firms from January 2009 to December 2009 for real-life exemplification purposes. The dataset includes the total number of claims in car accidents. Estimations of the scale and form parameters for each estimation method are presented in Table 1.

4.1.4. Numerical Study of the Risk Measures

The V and TV for a random variable X with the two-parameter Weibull distribution are defined in Equations (18) and (20), respectively. In this subsection, we provide a numerical study of the risk measures such as V and TV for the Weibull distribution model for different sets of parameters. The process involves generating a random sample of claim amounts according to the Weibull distribution, estimating parameters via Bayesian Estimation (BE), Simulated Annealing (SA), and Maximum Likelihood Estimation (MLE), and computing V and TV, along with standard errors, for each estimator. The numerical results of the risk measures are provided in Table 2. Additionally, the results are displayed graphically corresponding to each table in Figures 9 and 10.

4.1.5. Comparison of Estimation Methods

The comparison of estimation methods aims to identify the most suitable approach for estimating V and TV, considering factors such as accuracy, computational efficiency, and robustness to sample size variations. The study evaluates the performance of BE, MLE, and SA in estimating V and TV for Weibull distribution models. The simulation results demonstrate that SA or BE are chosen as the best estimators, with SA being particularly recommended for estimating parameters and quantiles thereof for large samples. With a high sample size, the SA approach appears to be the most recommendable estimator in general.

Table 3: VaR and TVaR for Various Estimation Methods

Estimator	αb	β	LL	V b baRa(X)	TVaR	SE
BE	817.2431	6.358192	-70111.1	7943.848	23831.544	0.00745097
MLE	825.7635	8.1895894	-71092.7	7740.74	23222.22	0.00774074
SA	887/511	5.967336	-69378.4	7132.48	21397.44	0.00694187

This table presents VaR and TVaR values along with their standard errors for different estimation methods based on insurance claims data (Abubakar and Sabri, 2020). Estimators include BE, MLE, and SA, each providing estimates of VaR and TVaR for the Weibull distribution model.

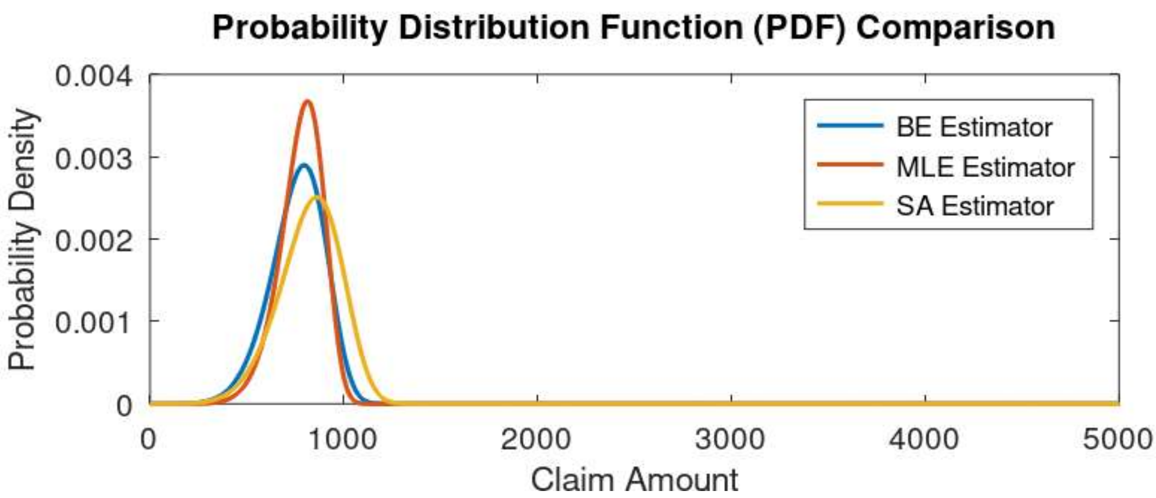

Fig. 7: PDF for Comparing Estimation Methods

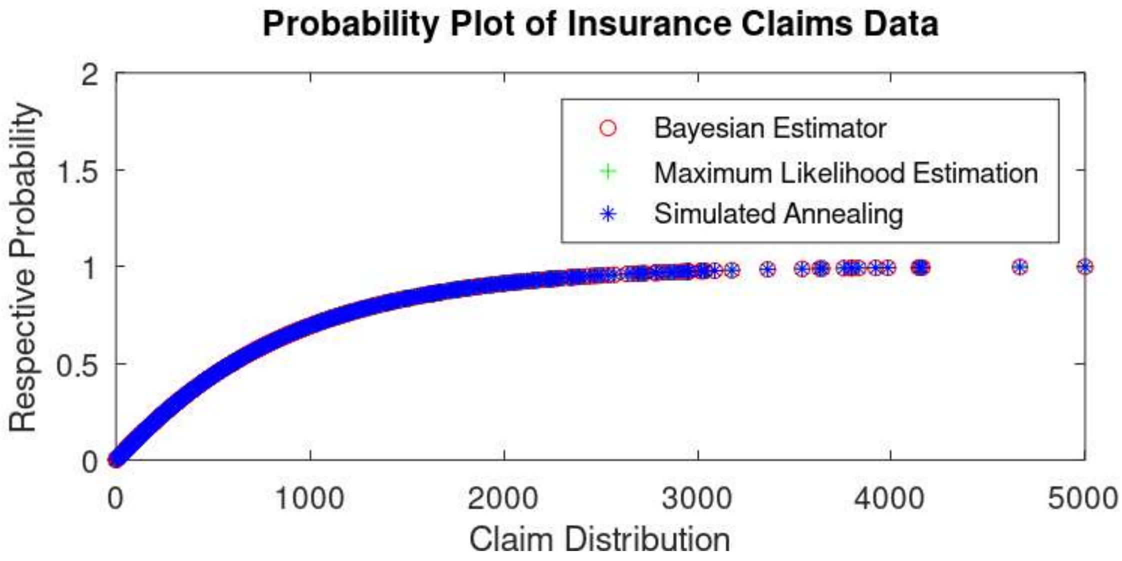

Fig. 8: Probability plot comparing estimation methods for insurance claims data

CONCLUSION

Statistical distributions are essential in financial sciences for data modeling and analysis (Henclova, 2006; Hersch, 2019). This article aimed to model claim sizes using the Weibull statistical distribution, with the shape parameter assuming the Gamma distribution. The parameters of the Weibull distributions were estimated using three methods: the Bayesian approach, the Maximum Likelihood approach, and the Simulated Annealing algorithm. A simulation study was conducted to examine the performance of the Bayesian approach compared to maximum likelihood and simulated annealing in estimating the parameters of the Weibull distribution for different sample sizes and parameter values.

Given the size of insurance claim amounts and the general insurance industries, Weibull distributions can be used to model insurance claim distributions These are useful when analyzing claims rather than using a lengthy schedule of raw claims data. Analysis can take the form of estimating the likelihood of claims falling into a specific range, as well as reinsurance agreements in place and other mathematical analyses. It also demonstrates that the Bayesian approach and the Maximum Likelihood method are satisfactory and agreed at estimating the probabilities of lower claims, whereas the Simulated Annealing algorithm is a better fit at estimating the probabilities of larger claims. As a result, it is advisable to use Simulated Annealing because it does not undervalue probabilities for large claims. This is especially useful when setting up reserves. Interestingly, all estimation methods can be used concurrently; for example, when the organization is interested in the probabilities of low claims, it employs the Bayesian approach or the Maximum Likelihood distribution, whereas when it is interested in large claims, it employs a Simulated Annealing algorithm. Further research on a Weibull distribution can be conducted when both parameters follow specific distributions. According to the findings, the Weibull distribution with the shape following the Gamma distribution fit the insurance claims amount data better. This shows that no method is superior to another; it all depends on the distributions used. The simulated annealing algorithm necessitates more variables in parameter estimation than the Bayesian and Maximum Likelihood methods. Priority setting is primarily subjective and relies on a guide provided by the classical method. In conclusion, the estimation methods work in conjunction to ensure that good conclusions are reached in this type of insurance claims data analysis.

ACKNOWLEDGEMENT

This research was sponsored by Tertiary Education Trust Fund through the Institution Based Research (IBR) grant intervention to Isa Kaita College of Education, Dutsin-Ma, Katsina State

REFERENCES

[1] G. E. Willmot, "The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial," *Scand. Actuar. J.*, vol. 1987, no. 3–4, pp. 113–127, 1987.

[2] G. E. Willmot and B. Sundt, "On evaluation of the delaporte distribution and related distributions," *Scand. Actuar. J.*, vol. 1989, no. 2, pp. 101–113, 1989.

[3] C. S. Kumar and B. U. Nair, "On generalized hyper-Poisson distribution of order k and its application," *Commun. Stat. - Theory Methods*, vol. 45, no. 22, pp. 6720–6731, 2016.

[4] J. Dhaene and B. Sundt, "On Error Bounds for Approximations to Aggregate Claims Distributions," *ASTIN Bull. J. IAA*, vol. 27, no. 2, pp. 243–262, 1997.

[5] C. Hipp, "Improved approximations for the aggregate claims distribution in the individual model," *ASTIN Bull. J. IAA*, vol. 16, no. 2, pp. 89–100, 2017.

[6] R. Kazemi, A. Jalilian, and K. Akram, "Fitting Skew Distributions to Iranian Auto Insurance Claim Data," *Appl. Appl. Math. An Int. J.*, vol. 12, no. 2, pp. 790–802, 2017.

[7] M. V Wuthrich, "Non-Life Insurance : Mathematics & Statistics," *Available SSRN 2319328.*, 2023.

[8] J. Zhao, H. Faqiri, Z. Ahmad, W. Emam, M. Yusuf, and A. M. Sharawy, "The Lomax-Claim Model: Bivariate Extension and Applications to Financial Data," *Complexity*, vol. 2021, pp. 1–17, 2021.

[9] P. Ter Berg, "On the Loglinear Poisson and Gamma Model," *ASTIN Bull. J. IAA*, vol. 11, no. 1, pp. 35–40, 1980.

[10] M. Denuit, "A New Distribution of Poisson-Type for the Number of Claims," *ASTIN Bull. J. IAA*, vol. 27, no. 2, pp. 229–242, 1997.

[11] P. O. Goffard, S. Loisel, and D. Pommeret, "Polynomial Approximations for Bivariate Aggregate Claims Amount Probability Distributions," *Methodol. Comput. Appl. Probab.*, vol. 19, no. 1, pp. 151–174, 2017.

[12] C. O. Omari, S. G. Nyambura, and J. M. W. Mwangi, "Modeling the Frequency and Severity of Auto Insurance Claims Using Statistical Distributions," *J. Math. Financ.*, vol. 08, no. 01, pp. 137–160, 2018.

[13] L. Hong and R. Martin, "A Flexible Bayesian Nonparametric Model for Predicting Future Insurance Claims," *North Am. Actuar. J.*, vol. 21, no. 2, pp. 228–241, 2017.

[14] A. Katsis and I. Ntzoufras, "Bayesian hypothesis testing for the distribution of insurance claim counts using the Gibbs sampler," *J. Comput. Methods Sci. Eng.*, vol. 5, no. 3, pp. 201–214, 2005.

[15] A. Moumeesri, W. Klongdee, and T. Pongsart, "Bayesian bonus-malus premium with Poisson-Lindley distributed claim frequency and Lognormal-Gamma distributed claim severity in automobile insurance," *WSEAS Trans. Math.*, vol. 19, pp. 443–451, 2020.

[16] J. Koneˇ and P. Richt, "Semi-Stochastic Gradient Descent Methods," vol. 1, 2015.

[17] Y. Zhang, V. Dukic, and J. Guszcza, "A Bayesian non-linear model for forecasting insurance loss payments," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 175, no. 2, pp. 637–656, 2012.

[18] H. Abubakar and S. R. M. Sabri, "A Bayesian Approach to Weibull Distribution with Application to Insurance Claims Data," *J. Reliab. Stat. Stud.*, vol. 16, no. 1, pp. 1–24, 2023.