Mathematics and Statistics 11(5): 856-867, 2023 http://www.hrpub.org
DOI: 10.13189/ms.2023.110513

Properties and Applications of Klongdee Distribution in Actuarial Science

Adisak Moumeesri[1], **Weenakorn Ieosanurak**[2,*]

[1]Department of Statistics, Faculty of Science, Silpakorn University, Thailand

[2]Department of Mathematics, Faculty of Science, Khon Kaen University, Thailand

Received July 3, 2023; Revised September 12, 2023; Accepted September 23, 2023

Cite This Paper in the following Citation Styles

(a)*: [1] Adisak Moumeesri, Weenakorn Ieosanurak, "Properties and Applications of Klongdee Distribution in Actuarial Science," Mathematics and Statistics, Vol.11, No.5, pp. 856-867, 2023. DOI: 10.13189/ms.2023.110513*

(b)*: Adisak Moumeesri, Weenakorn Ieosanurak (2023). Properties and Applications of Klongdee Distribution in Actuarial Science. Mathematics and Statistics, 11(5), 856-867. DOI: 10.13189/ms.2023.110513*

Abstract We have introduced a novel continuous distribution known as the Klongdee distribution, which is a combination of the exponential distribution with parameter $(\frac{\theta}{\alpha})$ and the gamma distribution with parameters $(2, \frac{\theta}{\alpha})$. We thoroughly examined various statistical properties that provide insights into probability distributions. These properties encompass measures such as the cumulative distribution function, moments about the origin, and the moment-generating function. Additionally, we explored other important measures including skewness, kurtosis, C.V., and reliability measures. Furthermore, we explore parameter estimation using nonlinear least squares methods. The numerical results presented compare the unweighted and weighted least squares (UWLS and WLS) methods, maximum likelihood estimation (MLE), and method of moments (MOM). Based on our findings, the MLE demonstrates superior performance compared to other parameter estimation methods. Moreover, we demonstrate the application of this distribution within an actuarial context, specifically in the analysis of collective risk models using a mixed Poisson framework. By incorporating the proposed distribution into the mixed Poisson model and analyzing a real-life dataset, it has been determined that the Poisson-Klongdee model outperforms alternative models in terms of performance. Highlighting its capability to mitigate the problem of overcharges, the Poisson-Klongdee model has been proven to be a valuable tool.

Keywords Exponential Distribution, Gamma Distribution, Parameter Estimations, Bonus-malus System, Actuarial Science

1 Introduction

A mixing distribution in probability theory and statistics refers to a probability distribution that results from the combination of two or more component distributions. The key concept behind a mixing distribution is that the observed random variable is generated by mixing these component distributions, where each component is assigned a specific weight or mixing proportion.

Mixture distributions find utility in various domains like finance, economics, biology, and signal processing. They provide a versatile approach to modeling intricate data that cannot be suitably characterized by a single distribution. By blending various distributions together, mixture distributions can effectively capture a broad spectrum of data patterns, including multiple modes, heavy tails, and asymmetry. This adaptability renders them a valuable instrument for faithfully depicting and studying intricate real-world data.

Estimating and analyzing mixing distributions present intriguing challenges as it involves estimating both the mixing proportions and the parameters of the component distributions. To tackle these challenges, various statistical methods and techniques have been developed. These include maximum likelihood estimation, Bayesian inference, and expectation-maximization algorithms. These methods provide valuable tools for accurately estimating the parameters and mixing proportions of the component distributions in a mixing distribution. They enable researchers and analysts to perform robust analyses and make reliable inferences based on the observed data.

Let X be a continuous random variable that follows an exponential distribution with parameter $\lambda > 0$, denoted as $X \sim \text{Exp}(\lambda)$. Its probability density function (PDF) is given by:

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0. \tag{1}$$

The cumulative distribution function (CDF) of the distribution has been derived as follows:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0. \tag{2}$$

Let X be a continuous random variable that follows the gamma distribution with two positive parameters, α and λ, denoted as $X \sim \text{Gamma}(\alpha, \lambda)$. The probability density function (PDF) for this distribution can be written as:

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x \geq 0. \tag{3}$$

Where, $\Gamma(\alpha)$ denotes the gamma function. Notably, when $\alpha = 1$, the gamma distribution simplifies to the exponential distribution with parameter λ, represented as $\text{Exp}(\lambda)$.

The CDF of a continuous random variable X following a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ is given by:

$$F(x) = \frac{\gamma(\alpha, \lambda x)}{\Gamma(\alpha)}, \quad x > 0. \tag{4}$$

Here, $\gamma(\alpha, \lambda x)$ denotes the lower incomplete gamma function, and $\Gamma(\alpha)$ represents the gamma function.

The two-parameter Lindley distribution combines characteristics from both the exponential and gamma distributions [1] - [8].

Ekhosuehi, Nzei, and Opone [3] proposed a Lindley distribution with two parameters by modifying the blending ratio between the exponential and gamma distributions, expressed as $f(x) = wf_1(x) + (1 - w)f_2(x)$, where $w = \frac{1}{1+\beta}$, and f_1 and f_2 represent the probability density functions of the exponential and gamma distributions, respectively. In other words,

$$f(x, \beta, \theta) = \frac{1}{1+\beta}\theta e^{-\theta x} + \frac{\beta}{1+\beta}\frac{\theta \cdot (\theta x)^{\beta-1}}{\Gamma(\beta)}e^{-\theta x}.$$

If we have a continuous random variable X that obeys the Janardan distribution (JD) [4] with two positive parameters, θ and α, we write it as $X \sim \text{JD}(\alpha, \theta)$. The probability density function (PDF) for this distribution is as follows:

$$f(x; \alpha, \theta) = \frac{\theta^2}{\alpha(\theta + \alpha^2)}(1 + \alpha x)^{-\frac{\theta}{\alpha}x}, \quad x > 0. \tag{5}$$

Its CDF is expressed as

$$F(x) = 1 - \frac{\alpha(\theta + \alpha^2) + \theta\alpha^2 x}{\alpha(\theta + \alpha^2)} exp(-\frac{\theta}{\alpha}x), \quad x > 0. \tag{6}$$

The Janardan distribution's probability density function (PDF) can be expressed as a mixture of two familiar distributions: $\text{Exp}(\frac{\theta}{\alpha})$ and $\text{Gamma}(2, \frac{\theta}{\alpha})$. The following demonstrates this,

$$f(x, \alpha, \theta) = pf_1(x) + (1 - p)f_2(x), \tag{7}$$

where $p = \dfrac{\theta^2}{\alpha(\theta + \alpha^2)}$, $f_1(x) = \dfrac{\theta}{\alpha}e^{-\frac{\theta}{\alpha}x}$, and $f_2(x) = \dfrac{\theta^2}{\alpha^2}xe^{-\frac{\theta}{\alpha}x}$.

Gaining a comprehensive understanding of mixing distributions and effectively utilizing them can yield valuable insights into the underlying processes that generate the data. This understanding leads to more accurate modeling and inference. The objective of this paper is to delve into the properties, estimation methods, and applications of mixing distributions. Our goal in doing this is to explore a continuous distribution with two parameters, namely $\theta > 0$ and $\alpha > 0$, which is introduced as follows:

$$f(x, \alpha, \theta) = (1 - p)f_1(x) + pf_2(x). \tag{8}$$

Its CDF, the first four moments, and certain associated measures have been suggested.

The structure of the paper is as follows: In Section 2, we provide a comprehensive definition of the Klongdee distribution and explore its properties. The derivation of moments, generating function, and other related measures are presented in Section 2.1. Section 2.2 focuses on the examination of reliability measures, including the survival function, failure rate function, and mean residual life function specifically tailored to the Klongdee distribution. Parameter estimation methods are discussed in Section 3, where four different approaches are introduced. In Section 4, we present numerical simulations that serve to validate the proposed methods, and we also delve into an actuarial application within this section. Finally, a conclusive summary is provided in Section 5.

2 Materials and Methods

In this section, we aim to introduce the Klongdee distribution, which is a continuous distribution characterized by two parameters, namely α and θ. The Klongdee distribution is named as a tribute to our advisor, whose surname is Klongdee, in recognition of their remarkable contributions and invaluable guidance throughout our research endeavors.

The PDF of the Klongdee distribution can be expressed as a mixture of the exponential distribution $(\frac{\theta}{\alpha})$ and the gamma distribution $(2, \frac{\theta}{\alpha})$ as follows:

$$f(x; \alpha, \theta) = (1 - p)f_1(x) + pf_2(x), \tag{9}$$

where $f_1(x) = \dfrac{\theta}{\alpha}e^{-\frac{\theta}{\alpha}x}$, $f_2(x) = \left(\dfrac{\theta}{\alpha}\right)^2 xe^{-\frac{\theta}{\alpha}x}$, and $p = \dfrac{\theta}{\theta + \alpha^2}$.

Definition 1 *Define a continuous random variable X to have a Klongdee distribution with two parameters α and θ, denoted as $X \sim KD(\alpha, \theta)$, if its probability density function (PDF) is expressed as follows:*

$$f(x; \alpha, \theta) = \frac{\theta}{\theta + \alpha^2}\left(\alpha + (\frac{\theta}{\alpha})^2 x\right)e^{-\frac{\theta}{\alpha}x}, \tag{10}$$

for all $x > 0, \alpha > 0, \theta > 0$.

Figure 1 presents density plots depicting the Klongdee distribution for specific values of α and θ. The observations from Figure 1 clearly indicate that the Klongdee distribution exhibits a notable characteristic of having a light tail.

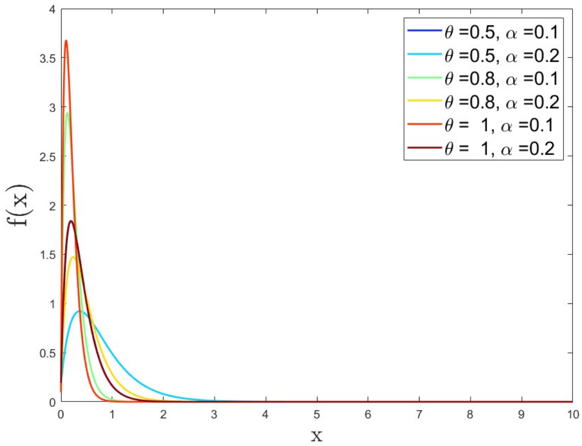

Figure 1. Graphs showing the PDF of the Klongdee distribution with different parameter values.

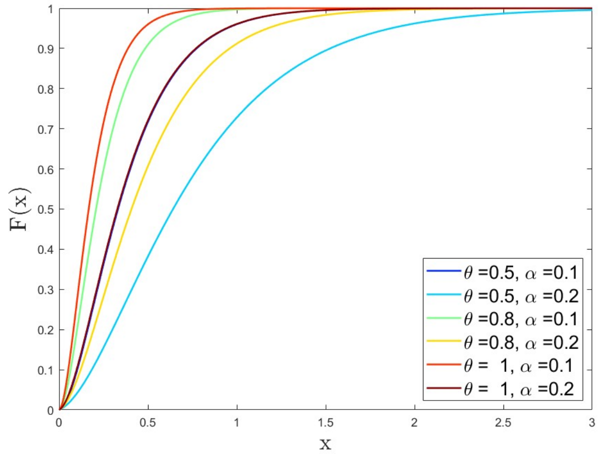

Figure 2. Graph showing the CDF of the Klongdee distribution with different parameter values.

To obtain the first derivative of Equation (10) with respect to x, we differentiate the equation accordingly:

$$\frac{d}{dx}f(x;\alpha,\theta) = \frac{\theta^2}{\alpha^3(\theta+\alpha^2)}(\alpha\theta - \alpha^3 - \theta^2 x)e^{-\frac{\theta}{\alpha}x}. \quad (11)$$

Now, based on Equation (11), we obtain
1. By setting $f'(x) = 0$, we can find the critical point of the function. Solving for x, we have: $x = \dfrac{\alpha\theta - \alpha^3}{\theta^2}$. For the case where $\alpha < \theta$, the value $x_0 = \frac{\alpha\theta-\alpha^3}{\theta^2}$ represents the unique critical point where $f(x)$ attains its maximum.
2. For the case when $\alpha > \theta$, we observe that $f'(x) \leq 0$, indicating that $f(x)$ is a decreasing function with respect to x. Consequently, the mode of the distribution described by Equation (10) is given by the value of x that maximizes the PDF, which occurs at the lower bound of the support.

$$\text{Mode} = \begin{cases} \dfrac{\alpha\theta - \alpha^3}{\theta^2} & \text{if } \alpha < \theta, \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

We proceed to derive the CDF of the Klongdee distribution as presented in Theorem 2.1.

Theorem 2.1 *The CDF of the Klongdee distribution is given by:*

$$F(x) = 1 - \frac{(\alpha^3 + \alpha\theta + \theta^2 x)}{\alpha(\alpha^2+\theta)}e^{\frac{-\theta}{\alpha}x}, \quad (13)$$

$x > 0, \theta > 0, \alpha > 0.$

Figure 2 displays the CDF plots of the Klongdee distribution for specific values of α and θ.

2.1 Statistical properties and tools

In this section, we explore and derive several properties of the Klongdee distribution, including the r^{th} moment about the origin (statistical quantities describing distribution shape and characteristics), the moment generating function (efficient calculation of moments for random variables), the coefficient of variation (a measure of relative variability, C.V.), the coefficient of skewness (measure of distribution asymmetry), and the coefficient of kurtosis (measure of tail shape relative to normal distribution), and Reliability Measures (assessment of system/process failure probability over time).

Theorem 2.2 *The r^{th} moment about the origin of the Klongdee distribution is defined as follows:*

$$\mu_r' = \frac{r!\alpha^r(\alpha^2 + (r+1)\theta)}{\theta^r(\alpha^2+\theta)}, r = 1, 2, 3, \ldots.$$

We can compute the first four moments around the origin for the Klongdee distribution in the following manner:

1. The first moment about the origin (mean):
   $\mu_1' = E[X] = \frac{\alpha(\alpha^2+2\theta)}{\theta(\alpha^2+\theta)}$.
2. The second moment about the origin :
   $\mu_2' = \frac{2\alpha^2(\alpha^2+3\theta)}{\theta^2(\alpha^2+\theta)}$.
3. The third moment about the origin:
   $\mu_3' = \frac{6\alpha^3(\alpha^2+3\theta)}{\theta^3(\alpha^2+\theta)}$.
4. The fourth moment about the origin:
   $\mu_4' = \frac{24\alpha^4(\alpha^2+3\theta)}{\theta^4(\alpha^2+\theta)}$.

By verifying that $\theta = \alpha^2$, we establish a direct relationship between the Klongdee distribution and the Janardan distribution. As a result, the moments about the origin of the Klongdee distribution simplify to the corresponding moments of the Janardan distribution.

To find moments around the mean, we can use the relationship between moments around the mean and moments around the origin. This relationship enables us to calculate the moment around the mean in the following way.

$$\mu_2 = \frac{\alpha^2(\alpha^4 + 2\theta^2 + 4\alpha^2\theta)}{\theta^2(\theta+\alpha^2)^2},$$
$$\mu_3 = \frac{2\alpha^3(2\theta^3 + \alpha^6 + 6\alpha^4\theta + 6\alpha^2\theta^2)}{\theta^3(\theta+\alpha^2)^3},$$
$$\mu_4 = \frac{3\alpha^4(8\theta^4 + 3\alpha^8 + 44\alpha^4\theta^2 + 24\alpha^6\theta + 32\alpha^2\theta^3)}{\theta^4(\theta+\alpha^2)^4}.$$

Specifically, the 2nd moment about the mean corresponds to the variance, which is denoted by

$$\sigma^2 = \mu_2 = \frac{\alpha^2(\alpha^4 + 2\theta^2 + 4\alpha^2\theta)}{\theta^2(\theta+\alpha^2)^2}. \tag{14}$$

Other properties found for the Klongdee distribution are as follows:

1. The coefficient of variation (C.V): C.V. $= \frac{\sigma}{\mu_1'} = \frac{\sqrt{\alpha^4+2\theta^2+4\alpha^2\theta}}{\alpha^2+2\theta}$.
2. The coefficient of skewness ($\sqrt{\beta_1}$): $\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2(2\theta^3+\alpha^6+6\alpha^4\theta+6\alpha^2\theta^2)}{(\alpha^4+2\theta^2+4\alpha^2\theta))^{3/2}}$.
3. The coefficient of kurtosis (β_2): $\beta_2 = \frac{3(8\theta^4+3\alpha^8+44\alpha^4\theta^2+24\alpha^6\theta+32\alpha^2\theta^3)}{(\alpha^4+2\theta^2+4\alpha^2\theta)^2}$.

The moment generating function (MGF) of the Klongdee distribution can be derived as follows, as follows:

$$M_X(t) = E[e^{tX}] = \frac{\theta(\alpha^2\theta + \theta^2 - \alpha^3 t)}{(\theta+\alpha^2)(\theta-\alpha t)^2}, \quad \frac{\theta}{\alpha} > t.$$

2.2 Reliability Measures

In this section, our goal is to obtain formulas for the reliability measures of the Klongdee distribution. These measures encompass the survival function, failure rate function, and mean residual life function. They are essential for gaining insights into the behavior and characteristics of the Klongdee distribution, shedding light on its reliability and lifespan properties.

The survival function, denoted as $S(x)$, represents the probability that a Klongdee random variable exceeds a specified time t. It can be obtained by subtracting the CDF from 1:

$$S(x) = 1 - F(x).$$

The failure rate function, denoted as $h(t)$ or $h(x)$, provides insights into the instantaneous failure rate at time t. It is defined as the ratio of the PDF to the survival function: $h(x) = \dfrac{f(x)}{S(x)}$.

Lastly, the mean residual life function, denoted as $m(x)$, represents the expected remaining lifetime given that a Klongdee random variable has survived up to time x. It is defined as the ratio of the expected remaining lifetime to the survival function: $E[X - x|X > x]$.

Theorem 2.3 *For $x > 0$, with parameters $\theta, \alpha > 0$, reliability measures of the Klongdee distribution are defined as follows:*

1. *Survival function:* $S(x) = \frac{(\alpha^3+\alpha\theta+\theta^2 x)}{\alpha(\alpha^2+\theta)} e^{\frac{-\theta}{\alpha}x}$.
2. *Failure rate function:* $h(x) = \frac{\theta\alpha(\alpha+(\frac{\theta}{\alpha})^2 x)}{\alpha^3+\alpha\theta+\theta^2 x}$.
3. *Mean residual life function:* $m(x) = \frac{\alpha(\alpha^3+2\alpha\theta+\theta^2 x)}{\theta(\alpha^3+3\alpha\theta+\theta^2 x)}$.

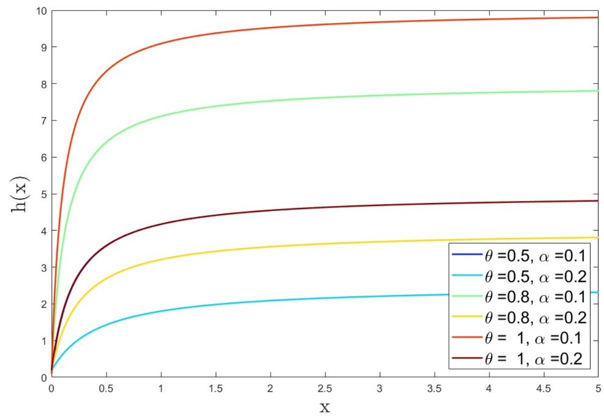

Figure 3. The failure rate function of the Klongdee distribution for various parameter values.

Figure 3 displays the failure rate function of the Klongdee distribution. It is worth noting some important properties of the failure rate function in relation to the Klongdee distribution:

At $x = 0$, the failure rate function takes the value $h(0) = \frac{\theta\alpha}{\theta+\alpha^2}$. Interestingly, this value is equal to the PDF evaluated at $x = 0$, denoted as $f(0)$. This indicates that the failure rate at the origin is equal to the density of the distribution at the origin.

The mean residual life function at $x = 0$, denoted as $m(0)$, corresponds to the derivative of the moment generating function evaluated at $x = 0$. In other words, $m(0) = \mu_1'$, where μ_1' represents the derivative of the first moment about the origin.

The derivative of the failure rate function, denoted as $h'(x)$, is greater than zero. This indicates that the failure rate function $h(x)$ is an increasing function of x, α, and θ. In other words, as x, α, or θ increases, the failure rate also increases. This implies that the failure rate tends to increase with time and with higher parameter values.

On the other hand, the mean residual life function, denoted as $m(x)$, is a decreasing function of x, α, and θ. This means that as x, α, or θ increases, the mean residual life decreases. This implies that the expected remaining lifetime tends to decrease with time and with higher parameter values.

These properties offer valuable insights into how the failure rate function and mean residual life function behave within the Klongdee distribution.

3 Estimation of Parameters

Parameter estimation is a critical task in statistics and data analysis, with the primary objective of deducing the unknown parameters of a statistical model based on observed data. Its significance spans across diverse disciplines, including economics, engineering, biology, and social sciences, where comprehending the underlying parameters of a system holds vital importance for informed decision-making and drawing meaningful conclusions. The aim of parameter estimation is to determine the most optimal estimate or approximation for the unknown parameters, ensuring the best possible fit between the model and the observed data. This involves selecting an appropriate estimation method. In this section, we employ four techniques to estimate the parameters of the Klongdee distribution. These techniques include the UWLS method using the CDF, the WLS method using the CDF, the method of moments (MOM), and maximum likelihood estimation (MLE). By utilizing these techniques, we aim to obtain estimates that accurately capture the true values of the parameters. We then apply these estimation methods to the available data, ensuring that the derived estimates faithfully represent the underlying parameter values of the Klongdee distribution.

3.1 Unweighted least squares method via the CDF

This method is a popular technique for parameter estimation in statistical analysis. It aims to minimize the sum of squared differences between the observed CDF values and the corresponding CDF values predicted by the distribution. By focusing on the overall fit of the CDF, this method offers a straightforward and intuitive approach to estimating distribution parameters.

The UWLS method via the CDF assumes that the observed data points, denoted as $X_1, X_2, \ldots, X_n$, are generated from random variables that follow the Klongdee distribution, represented as $X_i \sim \mathrm{KD}(\alpha, \theta)$. This assumption ensures that the data conforms to the specific Klongdee distribution with parameters α and θ. Additionally, it is assumed that the observed data points are independent of each other, meaning that the values of X_i do not depend on or influence each other within the dataset.

$$\log(F(x)) = \log\left(1 - \frac{(\alpha^3 + \alpha\theta + \theta^2 x)}{\alpha(\alpha^2+\theta)} e^{\frac{-\theta}{\alpha}x}\right), \tag{15}$$

$x > 0, \alpha > 0, \theta > 0$.

Consider n ordered observations, where $0 < x_1 < x_2 < \ldots < x_n$. In the context of the Klongdee distribution with parameters α and θ, these observations can be treated as independent and identically distributed random variables, denoted as $X_1, X_2, \ldots, X_n$.

$$\begin{aligned}\log(F(x_i)) &= \log(\alpha(\alpha^2+\theta) \\ &\quad - \log((\alpha^3+\alpha\theta+\theta^2 x)e^{\frac{-\theta x}{\alpha}})) \\ &\quad - \log(\alpha(\alpha^2+\theta)),\end{aligned} \tag{16}$$

where $x, \alpha, \theta > 0$.

$$F_n(x_i) = \frac{i-d}{n-2d+1}, \quad i = 1, 2, \ldots, n. \tag{17}$$

For a real number d such that $0 \leq d \leq 1$, we select four commonly used expressions as estimators of $F(x_i)$, where $F(x_i)$ represents the CDF at the i-th ordered observation x_i.

$$u_{ik} = \begin{cases} \frac{i}{n+1}, & k=1 \\ \frac{i-0.3}{n+0.4}, & k=2 \\ \frac{i-0.375}{n+0.25}, & k=3 \\ \frac{i-0.5}{n}, & k=4 \\ \frac{i-r}{n+r}, & k=5. \end{cases} \tag{18}$$

For $i = 1, 2, 3, \ldots, n$ and a real number r in the range of (0, 1), our objective is to estimate the parameters α and θ using the UWLS method. This entails minimizing the following function:

$$E_k(\alpha,\theta) = \sum_{i=1}^{n}\left(\log(u_{ik}) - \log\left(1 - \frac{(\alpha^3+\alpha\theta+\theta^2 x_i)}{\alpha(\alpha^2+\theta)} e^{\frac{-\theta}{\alpha}x_i}\right)\right)^2. \tag{19}$$

By solving the given equations for $k = 1, 2, 3, 4, 5$, we can determine the values of the unknown variables.

$$\frac{\partial}{\partial\alpha}E_k(\alpha,\theta) = 0,$$
$$\frac{\partial}{\partial\theta}E_k(\alpha,\theta) = 0.$$

Then, for $k = 1, 2, 3, 4, 5$, we can proceed with the following calculations.

$$A_k(x_i,\alpha,\theta) = \frac{(\alpha^5 - \alpha^3\theta + \alpha^2\theta^2 x_i + \theta^3 x_i)x_i e^{\frac{-\theta}{\alpha}x_i}}{(\alpha(\alpha^2+\theta) - (\alpha^3+\alpha\theta+\theta^2 x_i)e^{\frac{-\theta}{\alpha}x})}, \tag{20}$$

$$B_k(x_i,\alpha,\theta) = \frac{(\alpha^5 + \alpha^2\theta^2 x_i\theta^3 x_i))x_i e^{\frac{-\theta}{\alpha}x_i}}{(\alpha(\alpha^2+\theta) - (\alpha^3+\alpha\theta+\theta^2 x_i)e^{\frac{-\theta}{\alpha}x})}. \tag{21}$$

We obtain

$$\log(\hat{\alpha}) = \frac{\sum_{i=1}^{n}\left[\log\left(\hat{\alpha}(\hat{\alpha}^2+\theta) - (\hat{\alpha}^3+\hat{\alpha}\hat{\theta}+\hat{\theta}^2 x_i)e^{\frac{-\hat{\theta}}{\hat{\alpha}}x_i} - \log(u_{ik})\right)\right] A_k(x_i,\hat{\alpha},\hat{\theta})}{\sum_{i=1}^{n} A_k(x_i,\hat{\alpha},\hat{\theta})} - \log(\hat{\alpha}^2+\theta), \tag{22}$$

and

$$\log(\hat{\theta}) = -\frac{\sum_{i=1}^{n}[\log(\hat{\alpha}(\hat{\alpha}^2+\hat{\theta}) - (\hat{\alpha}^3+\hat{\alpha}\hat{\theta}+\hat{\theta}2x_i)e^{\frac{-\hat{\theta}}{\hat{\alpha}}x_i}) - \log(u_{ik})]B_k(x_i,\hat{\alpha},\hat{\theta})}{\sum_{i=1}^{n} B_k(x_i,\hat{\alpha},\hat{\theta})} + \log(\hat{\alpha}) + \log(\hat{\alpha}^2\theta + \hat{\theta}^2). \tag{23}$$

We utilize an iterative method to estimate α and θ from Equation (22) and (23), yielding the respective estimators $\hat{\alpha}$ and $\hat{\theta}$. This iterative procedure involves repeatedly updating the estimates until convergence is achieved.

3.2 Weighted least squares method via the CDF

The weighted least squares method via the CDF is a statistical technique used for parameter estimation. In this method, each observation in the dataset is assigned a weight, which reflects its relative importance in the estimation process. The weights are typically determined based on the characteristics of the data or the specific needs of the analysis. By incorporating these weights, the method emphasizes the observations with higher weights, giving them more influence in the estimation of the parameters. The weighted least squares method via the CDF aims to minimize the weighted sum of squared differences between the observed CDF values and the corresponding CDF values predicted by the distribution, providing a more tailored and flexible approach for parameter estimation compared to the UWLS method.

In order to address the potential issues associated with using the same weight for all data points, it is important to introduce a weighting factor to the values. This is done in Equation (19). To determine the weighting factors, a variance stabilizing transformation can be utilized. Bickel and Doksum [9] proposed an approach that relates the variance of $\log(u_{ik})$, denoted as $Var(\log(u_{ik}))$, to the uncertainty of u_{ik}, denoted as $Var(u_{ik})$. By considering this relationship, we obtain the following expression:

$$Var(\log(u_{ik})) = \left(\frac{\partial \log(u_{ik})}{\partial u_{ik}}\right)^2 Var(u_{ik}),$$

which gives

$$Var(\log(u_{ik})) = \left(\frac{1}{\partial u_{ik}}\right)^2 Var(u_{ik}).$$

This relationship allows us to estimate the appropriate weighting factor for each observation based on the corresponding variance. By incorporating these weighting factors into the estimation procedure, we can obtain more accurate and reliable parameter estimates.

Hence, in order to minimize the given function, we employ w_{ik} as the weighting factor, defined as $w_{ik} = u_{ik}^2$.

$$E_k(\alpha,\theta) = \sum_{i=1}^{n} w_{ik}\left(\log(u_{ik}) - \log\left(1 - \frac{(\alpha^3+\alpha\theta+\theta^2 x_i)}{\alpha(\alpha^2+\theta)} e^{\frac{-\theta}{\alpha}x_i}\right)\right)^2, \tag{24}$$

$k = 1, 2, 3, 4, 5$ and we solve Equation (24) in the same manner as before.
Then,

$$\log(\hat{\alpha}) = \frac{\sum_{i=1}^{n} w_{ik}\left[\log\left(\hat{\alpha}(\hat{\alpha}^2+\theta) - (\hat{\alpha}^3+\hat{\alpha}\hat{\theta}+\hat{\theta}^2 x_i)e^{\frac{-\hat{\theta}}{\hat{\alpha}}x_i} - \log(u_{ik})\right)\right] A_k(x_i,\hat{\alpha},\hat{\theta})}{\sum_{i=1}^{n} w_{ik} A_k(x_i,\hat{\alpha},\hat{\theta})} - \log(\hat{\alpha}^2+\theta), \tag{25}$$

and

$$\log(\hat{\theta}) = -\frac{\sum_{i=1}^{n} w_{ik}[\log(\hat{\alpha}(\hat{\alpha}^2+\hat{\theta}) - (\hat{\alpha}^3+\hat{\alpha}\hat{\theta}+\hat{\theta}2x_i)e^{\frac{-\hat{\theta}}{\hat{\alpha}}x_i}) - \log(u_{ik})]B_k(x_i,\hat{\alpha},\hat{\theta})}{\sum_{i=1}^{n} w_{ik} B_k(x_i,\hat{\alpha},\hat{\theta})} + \log(\hat{\alpha}) + \log(\hat{\alpha}^2\theta+\hat{\theta}^2). \tag{26}$$

We employ an iterative method to estimate α and θ from Equations (25) and (26), resulting in the estimators $\hat{\alpha}$ and $\hat{\theta}$, respectively.

3.3 Method of moments

The method of moments (MOM) is a widely used method for estimating the parameters of a statistical distribution based on sample moments. The main principle behind MOM is to equate the theoretical moments of the distribution with the corresponding sample moments and solve for the unknown parameters.

In the case of the Klongdee distribution, MOM involves equating the theoretical moments of the distribution (such as the mean, variance, skewness, etc.) with the sample moments calculated from the available data. By equating these moments, we can derive equations that allow us to estimate the parameters α and θ.

In order to estimate the two parameters of the Klongdee distribution, we can utilize the first moment about the origin (mean). By equating the theoretical first moment of the Klongdee distribution, denoted as $E[X]$, with the sample mean, denoted as $\bar{x}$, we can proceed with the estimation process.

$$E[X] = \frac{\alpha(\alpha^2+2\theta)}{\theta(\theta+\alpha^2)},$$
$$E[X] = \bar{X}.$$

Let us assume that $\theta = b\alpha^2$. By making this assumption, we obtain the following expression:

$$\bar{X} = \frac{\alpha(\alpha^2+2\theta)}{\theta(\theta+\alpha^2)}.$$

Therefore,

$$\hat{\alpha} = \frac{1+2b}{b(b+1)\bar{X}} \quad \text{and} \quad \hat{\theta} = \frac{1}{b}\left(\frac{1+2b}{b(b+1)\bar{X}}\right)^2,$$

Here, $\hat{\alpha}$ and $\hat{\theta}$ represent the estimators of the parameters α and θ, respectively.

3.4 Maximum Likelihood Estimates

The Maximum Likelihood Estimates (MLE) method assumes that the observed data points, denoted as $x_1, x_2, \ldots, x_n$, are generated from random variables that follow the Klongdee distribution. Additionally, it is assumed that these data points are independent and identically distributed (i.i.d). The observed frequency in the sample corresponding to $X = x$ is denoted as f_x, where $x = 1, 2, \ldots, k$. Here, k represents the largest observed value that has a non-zero frequency. It is important to note that the sum of all frequencies, $\sum_{x=1}^{k} f_x$, equals the total sample size n. These assumptions are crucial for applying the MLE method to estimate the parameters of the Klongdee distribution based on the observed data.

The likelihood function, denoted as L, of the Klongdee distribution is expressed as follows:

$$L = \left(\frac{\theta}{\theta + \alpha^2}\right)^n \prod_{x=1}^{k} \left(\alpha + (\frac{\theta}{\alpha})^2 x\right)^{f_x} e^{-\frac{\theta}{\alpha}(n\bar{x})}. \tag{27}$$

Therefore, the log-likelihood function is obtained by taking the natural logarithm of the likelihood function, resulting in:

$$\ln L = n \ln \theta - n \ln(\theta + \alpha^2) + \sum_{x=1}^{k} f_x \ln(\alpha + (\frac{\theta}{\alpha})^2 x) - \frac{\theta}{\alpha}(n\bar{x}). \tag{28}$$

By differentiating Equation (28) with respect to θ and α, we obtain the following partial derivatives:

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \frac{n}{\theta + \alpha^2} + \sum_{x=1}^{k} f_x \frac{2\theta x}{\alpha^3 + \theta^2 x} - \frac{n\bar{x}}{\alpha}, \tag{29}$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{-2n\alpha}{\theta + \alpha^2} + \sum_{x=1}^{k} f_x \frac{\alpha^3 - 2\theta^2 x}{\alpha^4 + \alpha\theta^2 x} + \frac{\theta n\bar{x}}{\alpha^2}. \tag{30}$$

It appears that the two equations (29) and (30) cannot be directly solved. Nevertheless, the Fisher's scoring method can be employed to solve these equations, considering that we have

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{-n}{\theta^2} + \frac{n}{(\theta + \alpha^2)^2} + \sum_{x=1}^{k} f_x \left(\frac{2\alpha^3 x - 2\theta^2 x^2}{(\alpha^3 + \theta^2 x)^2}\right), \tag{31}$$

$$\frac{\partial \ln L}{\partial \theta \partial \alpha} = \frac{2n\alpha}{(\theta + \alpha^2)^2} + \sum_{x=1}^{k} f_x \frac{-6\alpha^2 \theta x}{(\alpha^3 + \theta^2 x)^2} + \frac{n\bar{x}}{\alpha^2}, \tag{32}$$

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = \frac{2n\alpha^2 - 2n\theta}{(\theta + \alpha^2)^2} - \sum_{x=1}^{k} f_x \left(\frac{\alpha^6 - 10\alpha^3\theta^2 x - 2\theta^4 x^2}{(\alpha^4 + \alpha\theta^2 x)^2}\right) - \frac{2\theta n\bar{x}}{\alpha^3}. \tag{33}$$

The equation for estimating $\hat{\theta}$ and $\hat{\alpha}$ can be solved using the following expressions:

$$\begin{bmatrix} \frac{\partial^2 lnL}{\partial \theta^2} & \frac{\partial^2 lnL}{\partial \theta \partial \alpha} \\ \frac{\partial^2 lnL}{\partial \alpha \partial \theta} & \frac{\partial^2 lnL}{\partial \alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\alpha} - \alpha_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial lnL}{\partial \theta} \\ \frac{\partial lnL}{\partial \alpha} \end{bmatrix}, \tag{34}$$

where θ_0 and α_0 are the initial values of θ and α respectively. This equation is solved iteratively until sufficiently accurate estimates of $\hat{\theta}$ and $\hat{\alpha}$ are obtained.

4 Numerical results

In this section, we present the numerical results in three distinct categories. The first part highlights the application of the method to real data, demonstrating its practical utility in real-world scenarios. The second part showcases a simulated study, providing insights into the method's performance under controlled conditions. Lastly, we delve into claim modeling and insurance premium pricing, specifically examining its applicability within a bonus-malus system. By organizing the results in this manner, we provide a comprehensive overview of the method's effectiveness and its potential applications across various domains.

4.1 Application to real data

In this section, we proposed two real datasets: one about the waiting times of 100 bank customers [2], and the other about the survival times of 121 patients with breast cancer [10].

Table 1 presents the fittings of the Klongdee distribution, which pertain to the waiting times (in minutes) of 100 bank customers. The parameters have been estimated using the method of moments. For the purpose of comparison, the expected frequencies based on the Lindley distribution (LD) and Janardan distribution (JD) are also provided alongside those obtained from the Klongdee distribution (KD). The results highlight that the Klongdee distribution exhibits a superior fit to the data when compared to the Lindley and Janardan distributions. Moreover, Table 1 presents the expected frequencies for further analysis and comparison.

By looking at Table 1, we can see that the Kolmogorov-Smirnov (KS) statistics are very similar for all three distributions. Also, when the chi-square (χ^2) value decreases, it means that the observed and expected frequencies match better. The KD distribution has the lowest χ^2 value, which shows it agrees better with the expected frequencies compared to other methods.

The survival times of 121 patients with breast cancer obtained from a large hospital in a period from 1929 to 1938 as shown in Table 2. Table 3 presents the fittings of the Klongdee distribution, which pertain to the survival times of 121 patients with breast cancer obtained from a large hospital in a period from 1929 to 1938. It's evident from the experiment that when using the WLS method with $k = 2$, the chi-square value is minimized. The chi-square values for the WLS and UWLS methods are quite similar. Conversely, when employing the WLS method with $k = 4$, the KS test yields the lowest value, while the KS test values remain similar for the WLS and UWLS methods. However, it's worth noting that the MOM method produces a notably higher KS test result.

4.1.1 A simulated study

Based on the waiting time data presented in Table 1, we estimate the parameter values $(\theta, \alpha) = (0.0077, 0.0422)$ using the method of moments. To generate samples from the specified distribution with the parameter values $(\theta, \alpha) = (0.0077, 0.0422)$, we utilize the acceptance-rejection method [11]. In this process, we set the sample size to $1,000$ and perform $1,000,000$ iterations for each method to ensure sufficient sample generation and accurate representation of the distribution.

Table 1. Waiting times (in minutes) of 100 bank customers for observed and expected frequencies.

Waiting time (minutes)	Observed frequency	Expected frequency		
		LD	JD	KD (MOM)
0-5	30	30.39	30.16	29.92
5.01-10	32	30.69	30.92	29.79
10.01-15	19	19.21	19.32	19.18
15.01-20	10	10.28	10.28	10.63
20.01-25	5	5.08	5.05	5.46
25.01-30	1	2.39	2.37	2.67
30.01-35	2	1.09	1.07	1.27
35.01-40	1	0.49	0.47	0.59
total	100	99.62	99.64	99.51
Estimated parameters		$\hat{\theta} = 0.1897$	$\hat{\theta} = 0.2139$ $\hat{\alpha} = 1.1189$	$\hat{\theta} = 0.0077$ $\hat{\alpha} = 0.0422$
χ^2		2.1711	2.2499	1.9956
d.f.		6	5	5
KS test		0.90038	0.90043	0.90049

Table 2. Survival times of 121 patients with breast cancer.

Survival times	Observed frequency
0.3 -19.5125	32
19.5025 - 38.7250	26
38.7150 - 57.9375	28
57.9275 - 77.1500	13
77.1400 - 96.3625	8
96.3525 - 115.5750	6
115.5650 - 134.7875	6
134.7775 - 154.000	2
total	121

Suppose we have a requirement to sample a random value x_i from the Klongdee distribution, denoted as $f(x)$, in order to calculate the function value for any given x. To accomplish this, we define an auxiliary distribution function as a uniform distribution. We select a value for the "envelope constant" ($m > 0$), which is used to scale the auxiliary distribution and create the "blanket function" denoted as $m \cdot g(x)$. The choices of g and m must satisfy the condition $m \cdot g(x) \geq f(x)$ for all x. For the waiting time data, it is possible to choose $m \geq 2.8$. In this simulation, we specifically choose the value of $m = 2.8$ to ensure the blanket function adequately covers the Klongdee distribution.

In this section, we perform a comparative analysis of four distinct parameter estimation methods: the UWLS method via the CDF, the WLS method via the CDF, the MLE and the MOM. The results of this simulation study are presented in Table 4, providing a comprehensive overview of the performance of each method. Additionally, Table 5 displays the corresponding chi-square values associated with each estimation method, further aiding in the evaluation of their effectiveness.

In our analysis, we employ the Chi-squared test as a metric to assess the performance of the estimation methods [12]. The Chi-squared test statistic (χ^2) is utilized, which is defined as follows:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

where, O_i represents the observed value and E_i represents the expected value for each category. By calculating the Chi-squared test statistic, we can evaluate the discrepancy between the observed and expected values and determine the goodness of fit of the estimation methods. Furthermore, we employ the Kolmogorov-Smirnov test, which can be described as follows:

$$\max |F_O(x) - F_T(x)|,$$

where $F_O(x)$ denotes the CDF observed in a random sample of n observations, while $F_T(x)$ pertains to the theoretical frequency distribution.

The analysis of Table 4 reveals that among the UWLS methods, the one with $k = 4$ stands out as the most favorable. Similarly, within the WLS methods, the one with $k = 1$ emerges as the top performer. These findings suggest that these specific parameter configurations yield the most accurate and reliable estimations within their respective methods.

We proceed to compare the optimal UWLS, the optimal WLS, MLE, and MOM methods. Using a significance level of 0.05, we obtain $\chi^2_{0.05,5} = 11.07$ from Table 5. Upon examination of the table, it becomes apparent that the estimates obtained through UWLS ($k = 4$), WLS ($k = 1$), and MLE are all below 11.07 and display a high degree of similarity. However, the Chi-squared value associated with the MLE method surpasses the others, indicating its superiority in terms of goodness of fit.

4.1.2 Bonus-Malus System

In this section, we mix the Poisson distribution with the proposed distribution and apply the mixed distribution to an actual dataset. Our goal is to demonstrate that our mixed model is superior to other competing models in terms of how well it fits the data. Additionally, we present a model for calculating automobile insurance premiums under the bonus-malus system.

Table 3. Survival times of 121 patients with breast cancer for observed and expected frequencies.

Observed frequency	Expected frequency											
	UWLS					WLS					MOM	MLE
	k=1	k=2	k=3	k=4	k=5	k=1	k=2	k=3	k=4	k=5		
32	31.47	31.47	30.35	30.35	30.35	32.02	30.19	30.62	31.38	33.08	28.72	28.09
26	28.88	28.88	27.57	27.57	27.57	27.99	28.28	25.46	30.60	28.55	25.76	27.77
28	21.61	21.61	21.09	21.09	21.09	20.93	21.58	19.51	22.53	21.00	20.27	21.80
13	14.70	14.70	14.84	14.84	14.84	14.46	14.97	14.23	14.78	14.26	14.87	15.46
8	9.47	9.47	9.93	9.93	9.93	9.53	9.84	10.05	9.09	9.22	10.45	10.34
6	5.89	5.89	6.42	6.42	6.42	6.08	6.23	6.93	5.37	5.78	7.13	6.66
6	3.57	3.57	4.06	4.06	4.06	3.79	3.85	4.70	3.08	3.53	4.76	4.18
2	2.12	2.12	2.52	2.52	2.52	2.32	2.33	3.14	1.74	2.12	3.12	2.57
$\sum = 121$												
$\hat{\theta}$	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0006	0.0008	0.0005	0.0008	0.0006	0.0005
$\hat{\alpha}$	0.0189	0.0189	0.0199	0.0199	0.0199	0.0228	0.0177	0.0288	0.0147	0.0237	0.0210	0.0146
χ^2	4.2852	4.2852	4.1062	4.1062	4.1062	4.2573	**4.0626**	5.1946	5.2456	4.6063	5.0436	4.3233
df	5	5	5	5	5	5	5	5	5	5	5	5
KS test	0.0334	0.0334	0.0578	0.0578	0.0578	0.0418	0.0492	0.0860	0.0329	0.0300	0.0930	0.0689

Table 4. Estimation Results for $\theta = 0.0077$ and $\alpha = 0.0422$ with a Sample Size of 1,000 and Waiting Time Data using UWLS, WLS, MLE, and MOM Methods.

Methods								
MLE		MOM		UWLS			WLS	
θ	α	θ	α	k	θ	α	θ	α
				1	0.008620	0.042426	0.007211	0.041406
				2	0.008110	0.042066	0.006687	0.042586
0.005480	0.029499	0.008684	0.041493	3	0.007787	0.042681	0.007731	0.041440
				4	0.007353	0.042069	0.006233	0.042617
				5	0.008153	0.043147	0.008287	0.042139

To enhance the modeling of the claim frequency distribution in automobile insurance, each policyholder is assigned a risk parameter that signifies their risk of experiencing an accident. This risk parameter is considered a random variable that varies among policyholders and follows a prior distribution. One proposed approach for modeling the frequency distribution involves mixing the Poisson distribution with the Klongdee distribution.

Mixing distribution : Assuming that the probability mass function (PMF) for the count of claims, denoted by y, is represented by the Poisson distribution with a parameter value of λ, the PMF can be formulated as follows:

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots, \lambda > 0. \tag{35}$$

The expected value or mean of the Poisson random variable is $E[Y|\lambda] = \lambda$.

The average number of claims made by a policyholder reflects their underlying risk, represented as a constant value denoted by λ. In the proposed model, λ is assumed to adhere to the Klongdee distribution with parameters α and θ. This distribution can be characterized by the probability density function (PDF) of λ, as shown below:

$$\pi(\lambda) = \frac{\theta}{\theta + \alpha^2}\left(\alpha + (\frac{\theta}{\alpha})^2\lambda\right)e^{-\frac{\theta}{\alpha}\lambda}, \tag{36}$$

for all $\lambda > 0, \alpha > 0, \theta > 0$. Outlined below is the procedure to derive the mixed Poisson distribution in conjunction with the Klongdee distribution:

$$\begin{aligned} f(y) &= \int_0^\infty f(y|\lambda)\,\pi(\lambda)\,d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \cdot \frac{\theta}{\theta + \alpha^2}\left(\alpha + (\frac{\theta}{\alpha})^2\lambda\right)e^{-\frac{\theta}{\alpha}\lambda}\,d\lambda \\ &= \frac{\theta\alpha^y[\alpha^2(\alpha+\theta) + \theta^2(y+1)]}{(\theta+\alpha^2)(\alpha+\theta)^{y+2}}, \end{aligned} \tag{37}$$

where $\alpha > 0, \theta > 0$ and $y = 0, 1, 2, \ldots$.

Model fitting : [13] presented a dataset on the distribution of automobile insurance policyholders according to the number of claims. This dataset is observed to be highly skewed towards the right and over-dispersed, where the variance is greater than the mean. The Poisson distribution (PD) is widely recognized as an unsuitable option for automobile insurance claims due to its mean and variance restriction. As a result, a mixed Poisson distribution with a prior distribution is preferred in such cases. Table 6 presents a comparison between the mixed Poisson model and the proposed model, including the Poisson-Lindley distribution (PLD), using maximum likelihood parameter estimation. The Chi-square statistic values indicate that the Poisson-Klondgee distribution (PKD) offers a superior fit to the dataset compared to other distributions.

Bonus − Malus premium : Many countries have implemented a bonus-malus system (BMS) that rewards policyholders with no claims and punishes those with claims. Thus, the following year's premium is determined by the policyholder's history up to the current year, regardless of the claim size. [14] calculated the BMS premium using the following formula:

$$\text{Premium}_{t+1} = \frac{E_{\pi^*(\lambda|n)}[l(\lambda)]}{E_{\pi(\lambda)}[l(\lambda)]} \times 100, \tag{38}$$

where $l(\lambda) = \sum_{n=0}^{\infty} nP(N = n|\lambda)$, $\pi(\lambda)$ represents the prior distribution, and $\pi^*(\lambda|n)$ signifies the posterior distribution. Consequently, when working with a sequence of independent and identically distributed claims denoted as $n = (n_1, n_2, \ldots, n_t)$, it becomes clear that deriving the posterior distribution is a simple process achieved by dividing the mixing distribution by the marginal distribution, as depicted below.

$$\pi^*(\lambda|n) = \frac{P(n|\lambda)\pi(\lambda)}{\int_0^\infty P(n|\lambda)\pi(\lambda)d\lambda}. \tag{39}$$

Table 5. Chi-squared Test Statistic (χ^2) Results for UWLS ($k = 4$), WLS ($k = 1$), MLE, and MOM Methods with Waiting Time Data.

Waiting time (minutes)	f	Methods			
		UWLS ($k = 4$)	WLS ($k = 1$)	MLE	MOM
0 - 5	275	288.7408	286.8614	288.3657	342.1238
5 - 10	303	291.7746	291.3072	305.4621	319.1345
10 - 15	198	193.2380	193.6271	196.3265	181.9561
15 - 20	124	110.4687	111.0644	107.4397	88.4414
20 - 25	60	58.5476	59.0564	54.2461	39.6778
25 - 30	22	29.6272	29.9814	26.0914	16.9610
30 - 35	11	14.5315	14.7525	12.1485	7.0194
35 - 40	7	6.9688	7.0974	5.5265	2.8384
Estimated parameters	θ	0.0074	0.0072	0.0055	0.0087
	α	0.04207	0.0414	0.0295	0.0415
	χ^2	5.7185	5.6608	4.9595	49.9615
	KS test	0.0172	0.0181	0.0158	0.0832

Table 6. Number of claims in automobile insurance.

Number of claim	**Observed frequency**	**Expected frequency**		
		PD	**PLD**	**PKD**
0	63,232	63,094.32	63,252.68	63,234.36
1	4333	4590.55	4292.03	4326.26
2	271	167.00	290.30	277.25
3	18	4.05	19.58	17.05
4	2	0.07	1.32	1.02
5	0	0.00	0.09	0.06
total	67,856	67,856	67,856	67,856
Estimated parameters		$\hat{\theta} = 0.0728$	$\hat{\theta} = 14.6238$	$\hat{\theta} = 643.2161$ $\hat{\alpha} = 34.7107$
χ^2		177.9421	2.2507	1.2083

$$\pi^*(\lambda|n) = \frac{\alpha^2(t+\frac{\theta}{\alpha})^{n+2}}{\Gamma(n+1)\left[\alpha^3(t+\frac{\theta}{\alpha})+\theta^2(n+1)\right]}\, e^{-(t+\frac{\theta}{\alpha})\lambda}\left(\alpha+\frac{\theta^2}{\alpha^2}\lambda\right)\lambda^n. \tag{40}$$

$$\text{Premium}_{t+1} = \frac{\theta(n+1)(\alpha^2+\theta)}{(\alpha t+\theta)(\alpha^2+2\theta)} \cdot \frac{\alpha^2(\alpha t+\theta)+\theta^2(n+2)}{\alpha^2(\alpha t+\theta)+\theta^2(n+1)} \cdot 100. \tag{41}$$

The posterior distribution of the Poisson-Klongdee distribution can be expressed in the following form:

By utilizing Equation (41), we compute the bonus-malus premiums only considering the frequency component. The outcomes we obtain are showcased in Table 7.

According to Table 7, policyholders who do not submit any claims in the first year receive a bonus equal to 5.95% of the base premium. On the other hand, policyholders who make a single claim during the first year face a penalty of 76.19% of the base premium. Claim-free policyholders enjoy lower premiums, whereas premiums increase for policyholders who file claims.

In order to facilitate comparisons, we have computed the bonus-malus premiums using the traditional Poisson-Lindley model (see [15], for details). The results are displayed in Table 8.

Based on the results showcased in Table 8, policyholders who refrain from submitting any claims in the first year receive a bonus equivalent to 6.74% of the base premium. In contrast, policyholders who file a single claim in the first year incur a malus amounting to 85.92% of the base premium.

The Poisson-Klongdee model demonstrates a lower level of penalization in comparison to traditional Poisson-Lindley models, highlighting its ability to alleviate the issue of overcharges.

5 Conclusions

A two-parameter continuous distribution called the Klongdee distribution has been introduced. This distribution's properties, including its CDF, expected value, r^{th} moment, and parameter estimation using nonlinear least squares methods, MLE, and MOM, have been proposed.

Table 7. Bonus-malus premium using Poisson-Klongdee model.

t	Number of claims					
	0	1	2	3	4	5
0	100.00					
1	94.05	176.19	253.60	328.66	402.37	475.26
2	88.73	166.77	240.44	311.90	382.09	451.48
3	83.94	158.26	228.54	296.73	363.71	429.92
4	79.63	150.55	217.72	282.92	346.98	410.29
5	75.71	143.52	207.84	270.31	331.68	392.35
6	72.14	137.09	198.79	258.74	317.65	375.87
7	68.88	131.19	190.47	248.10	304.72	360.70

Table 8. Bonus-malus premium using Poisson-Lindley model.

t	Number of claims					
	0	1	2	3	4	5
0	100.00					
1	93.26	185.92	278.08	369.81	461.17	552.22
2	87.37	174.23	260.67	346.74	432.50	517.99
3	82.17	163.92	245.30	326.37	407.17	487.72
4	77.56	154.75	231.63	308.24	384.61	460.77
5	73.43	146.55	219.40	292.01	364.41	436.63
6	69.72	139.18	208.39	277.39	346.21	414.87
7	66.37	132.50	198.42	264.16	329.74	395.17

The simulation results are divided into two parts. In the first part, the Klongdee distribution is applied to a data set representing waiting times in order to test its goodness of fit. The results show that the Klongdee distribution provides better fits compared to the earlier fits of the Lindley distribution and Janardan distribution. In the second part, we obtain the parameters for generating data based on the results from the first part. We then compare the performance of four methods using the Chi-squared test. Our analysis concludes that the MLE estimators outperform the UWLS, WLS, and MOM estimators in terms of performance.

In the context of actuarial science, we propose the mixed Poisson with Klongdee distribution as a model for claim modeling. We utilize this mixed distribution to develop a pricing model for insurance premiums based on the BMS. The findings indicate that the Poisson-Klongdee distribution has the ability to address the problem of overcharging.

The Klongdee distribution is specifically designed to suit right-skewed data. However, its application to datasets exhibiting different skewness characteristics can result in inadequate fit and inaccurate outcomes. It's imperative to thoroughly evaluate the skewness of the data before selecting the distribution. When dealing with data that lacks right-skewness, considering alternative distributions like the normal or gamma distribution is advisable, as it has the potential to yield more accurate results. In forthcoming research, we aim to investigate the feasibility of relaxing specific assumptions in our applications, notably the assumption of independence. This exploration will involve assessing the potential implications of such relaxations on our conclusions. Furthermore, our future endeavors will encompass expanding the model to encompass a wider array of assumptions and complexities. This extension aims to elevate the applicability and robustness of our findings to a broader range of scenarios. In conclusion, comprehending the intricacies of mixture distributions is instrumental in refining decision-making processes. By adeptly capturing complex patterns and revealing latent structures within datasets, these distributions empower us to formulate strategies that are not only well-informed but also precise in real-world applications.

Acknowledgements

We would like to extend our sincere appreciation to our advisor, Associate Professor Watcharin Klongdee, for the invaluable mentorship and support generously shared with us.

REFERENCES

[1] D. V. Lindley, Fiducial distributions and Bayes' theorem, Journal of the Royal Statistical Society. Series B (Methodological), Vol.20, 102–107, 1985.

[2] M. E. Ghitany, B. Atieh, S. Nadarajah, Lindley distribution and its application, Mathematics and Computers in Simulation, Vol.78, 493–506, 2008.

[3] N. Ekhosuehi, L. C. Nzei, F. Opone. A New Mixture of Exponential-Gamma Distribution, Journal of Science, Vol.33, No.2, 548-564, 2020.

[4] R. Shanker, S. Sharma, U. Shanker, R. Shanker, Janardan distribution and its application to waiting times data, Indian Journal of Applied Research, Vol.3, 500–502, 2013.

[5] R. Shanker, Akash Distribution and Its Application, International Journal of Probability and Statistics, Vol.4, No.3, 65-75, 2015.

[6] R. Shanker, H. Fesshaye, S. Sharmbhu, On Two - Parameter Lindley distribution and its applications to model lifetime data, Biometrics & Biostatistics International Journal, Vol.3, 9-15, 2016.

[7] S. Sarma, I. Ahmed, A. Begum. A New Two Parameter Gamma-Exponential Mixture, Journal of Mathematical and Computational Science, Vol.11, No.1, 414-426, 2021.

[8] S. Boonthiem, A. Moumeesri, W. Klongdee, W. Ieosanurak. A new Sushila distribution: properties and applications, European Journal of Pure and Applied Mathematics, Vol.15, No.3, 1280-1300, 2022.

[9] P.J. Bickel, K.A. Doksum, Mathematical statistics: basic ideas and selected topics, volumes I-II package, CRC Press, 2015.

[10] E. T. Lee, Statistical Methods for Survival Data Analysis (2nd Edition), John Wiley and Sons Inc., 1992.

[11] D. Gamerman, H. F. Lopes, Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference, CRC press, 2006.

[12] R.B. D'Agostino, M.A. Stephens, Goodness-of-Fit Techniques, Marcel Dekker, New York, 1986.

[13] P. De Jong, G. Heller, Generalized linear models for insurance data, Cambridge University Press, 2008.

[14] E. Gomez, F. Vazquez, Robustness in Bayesian Models for Bonus–Malus Systems, Intelligent And Other Computational Techniques In Insurance: Theory and Applications, 435-463, 2003.

[15] A. Moumeesri, W. Klongdee, T. Pongsart, Bayesian Bonus–Malus Premium with Poisson–Lindley Distributed claim frequency and Lognormal–Gamma distributed claim Severity in automobile Insurance, WSEAS Transactions on Mathematics, Vol.9, 443–451. 2020.