RT&A, No 2 (84)
Volume 20, June 2025

AN APPLICATIONS-BASED STOCHASTIC MODELING OF MIXTURE DISTRIBUTION FOR CANCER DATA

Poovitha R[1] and Pandiyan P[2]

•

[1]Research Scholar, Department of Statistics,
Annamalai University, Tamilnadu, India
[2]Professor, Department of Statistics,
Annamalai University, Tamilnadu, India
Corresponding Authour: [1]pranaveesai29@gamil.com, [2]pandiyanau@gmail.com

Abstract

This study proposed a new stochastic model by the mixture of Gamma and Zeghdoudi distribution (MGZ Distribution). The Reliability Analysis and statistical features such as moments, Moment generating function, order statistics, stochastic ordering, entropy, and the Maximum Likelihood Estimation of the model parameters were derived. Two real-time datasets of cancer patients' survival times were used to demonstrate the model's goodness-of-fit using the AIC, BIC, and AICC model selection methods. The new MGZ model has been compared with the other classical models like Exponential, Lindley and Shanker. The result shows that the proposed model is more flexible than the other models with the lowest values of AIC, BIC and AICC.

Keywords: Mixture distribution, Moments, Order Statistics, Maximum likelihood Estimation, Goodness of fit Criteria.

1. Introduction

The statistical distributions have been extensively utilized for analyzing time-to-event data, also referred to as survival or reliability data, in different areas of applicability, including medical science. In recent years, an impressive set of new statistical distributions has been explored by statisticians. The development of an expanded class of classical distribution is required in analysis, biomedicine, reliability, insurance, and finance. Recently, many researchers have been working in this area and have proposed new methods to develop improved probability distributions with utility. A statistical analysis is often employed, which heavily relies on the assumed probability model or distributions.

A Zeghdoudi distribution [8] random variable X with a parameter $\theta > 0$ is described by its probability density function (pdf)

$$f_1(x;\theta) = \frac{\theta^3}{2+\theta} x(1+x)e^{-\theta x}, \quad x > 0, \quad \theta > 0 \tag{1}$$

Considering the gamma distribution [18] with parameters $\beta = 3, \theta$, the pdf can be defined as:

$$f_2(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x}, \quad x > 0, \quad \theta > 0 \tag{2}$$

Lindley [5] introduced the fiducial distribution and Bayes theorem. Rama Shanker [17] proposed a mixture of exponential (θ) and gamma $(2, \theta)$ distributions, leading to the Shanker distribution. Shanker [14] also combined exponential (θ), gamma $(2,\theta)$, and gamma $(4,\theta)$ distributions with respective mixing proportions $\frac{\theta^3}{\theta^3+\theta^2+6}$, $\frac{\theta^2}{\theta^3+\theta^2+6}$, and $\frac{6}{\theta^3+\theta^2+6}$ This distribution is known as the Uma distribution. The Akash distribution is a two-component mixture of an exponential and a gamma distribution with mixing proportions $\frac{\theta^2}{\theta^2+2}$, and $\frac{2}{\theta^2+2}$.

Shanker [16] later proposed the Komal distribution with applications in survival analysis. Rama Shanker [13] also worked on combining exponential (θ) and gamma $(2,\theta)$ distributions with mixing proportions $\frac{\theta(\theta+1)}{\theta^2+\theta+1}$, and $\frac{1}{\theta^2+\theta+1}$ This article introduces the (MGZD), which was developed using a blend of Gamma and Zeghdoudi distributions. The probability density function (pdf) and cumulative distribution function (cdf) of the proposed distribution, along with several statistical characteristics and a reliability analysis, are discussed in the following sections. The maximum likelihood estimation (MLE) method is used to estimate model parameters. Cancer remains one of the leading causes of morbidity and mortality worldwide, impacting millions of lives annually. Despite advancements in cancer detection, treatment, and prevention, the global cancer burden continues to rise due to population aging, expansion, and shifts in risk factors. According to Sung et al. [19], the most recent global cancer data in 2020 showed approximately 10 million cancer-related deaths and 19.3 million new cases. The most common diagnoses were breast, lung, colon, prostate, and stomach cancers. The American Cancer Society [3] reported that advancements in early detection and treatment have improved survival rates, with breast cancer 5-year survival rates exceeding 85% in high-income countries. The World Health Organization [20] estimated the global economic impact of cancer at over $1 trillion annually, covering direct treatment costs and indirect losses such as productivity declines. Additionally, regional disparities in cancer incidence and mortality exist due to differences in healthcare access, risk factors, and early detection services. Approximately 70% of cancer deaths occur in low- and middle-income countries.

Statistical models are essential for studying cancer due to its complexity and heterogeneity. These models estimate survival rates, assess treatment effectiveness, and predict cancer risk. By incorporating new data and trends, these models improve predictions, enhance patient outcomes, and guide public health policies. Lastly, the results of fitting various well-known distributions to cancer survival data using (MGZD) are presented. All computations in this research were performed using the statistical programming language R.

2. New Mixture Distribution

This section introduces the MGZ distribution, a new distribution created by combining two existing distributions. Let X be a random variable with a mixed distribution. Its probability density function (p.d.f), $f(x)$, is expressed as follows

$$f(x) = \sum_{i=1}^{k} \omega_i f_i(x)$$

where $f_i(x)$ is the probability density function for all i, and ω_i, $i = 1, \ldots, k$, denote mixing proportions that are non-negative and satisfy

$$\sum_{i=1}^{k} \omega_i = 1$$

The components of the mixture are - $f_1(x) \sim \text{Gamma}(3,\theta)$ with parameters $\beta = 3, \theta$ - $f_2(x) \sim \text{Zeghdoudi}(\theta)$ with parameter θ. These two independent random variables have mixing proportions $\frac{\beta}{\beta+1}$ and $\frac{1}{\beta+1}$, respectively. Thus, the density function of X is given by

$$f(x;\theta,\beta) = \frac{\theta^3}{\beta+1}\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x} \quad (3)$$

The function defined in Equation (3) represents a valid probability density function $f(x;\theta,\beta)$ for all $x > 0$

$$\int_0^\infty f(x;\theta,\beta)\,dx = \int_0^\infty \frac{\theta^3}{\beta+1}\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x}\,dx$$

after simplification, we get

$$\int_0^\infty f(x;\theta,\beta)\,dx = 1$$

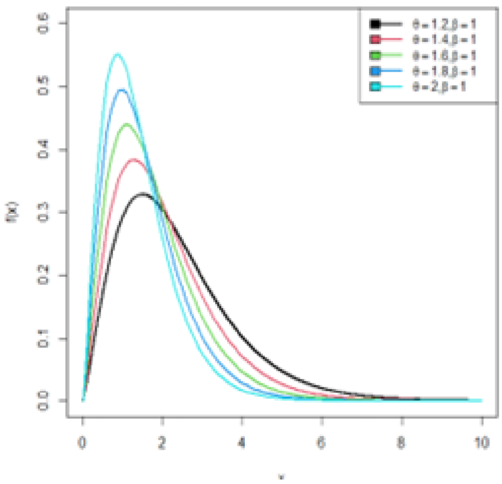

Figure 1: *PDF plot of MGZD different θ and β.*

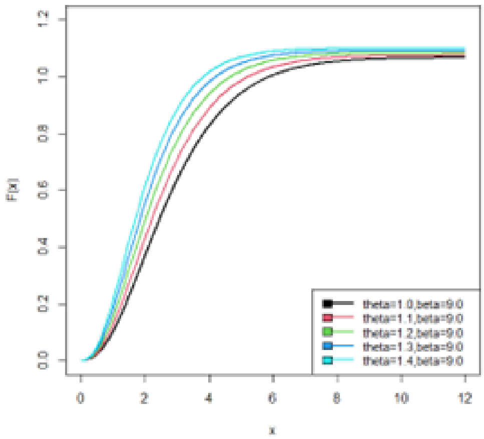

Figure 2: *CDF plot of MGZD different θ and β.*

The cumulative distribution function (c.d.f) of MGZD is defined as

$$F(x;\theta,\beta) = \int_0^x \frac{\theta^3}{\beta+1}\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x}\,dx$$

after simplification, we obtain

$$F(x;\theta,\beta) = \frac{1}{\beta+1}\left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta}\right], \quad x > 0, \theta, \beta > 0 \tag{4}$$

3. Reliability Analysis

This section provides the reliability function, hazard function, reverse hazard function, cumulative hazard function, odds rate, Mills ratio, and mean residual function for the MGZ distribution.

3.1. Survival Function

The survival function of the MGZ distribution is defined as

$$S(x) = 1 - F(x;\theta,\beta)$$

$$S(x) = 1 - \frac{1}{\beta+1}\left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta}\right]$$

3.2. Hazard Rate Function

An important metric for describing life phenomena is the hazard rate function of the MGZ distribution, which is defined as

$$h(x) = \frac{f(x;\theta,\beta)}{1-F(x;\theta,\beta)}$$

$$h(x) = \left(\frac{\theta^3 \left(\frac{\beta}{2}x^2 + \frac{x}{(2+\theta)}(1+x) \right) e^{-\theta x}}{(\beta+1) - \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]} \right)$$

3.3. Reverse Hazard Function

The reverse hazard rate function of the MGZ distribution is defined as

$$h_r(x) = \frac{f(x;\theta,\beta)}{F(x;\theta,\beta)}$$

$$h_r(x) = \left(\frac{\theta^3 \left(\frac{\beta}{2}x^2 + \frac{x}{(2+\theta)}(1+x) \right) e^{-\theta x}}{\left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]} \right)$$

3.4. Cumulative Hazard Function

The cumulative hazard function of the MGZ distribution is defined as

$$H(x) = -\ln(1 - F(x;\theta,\beta))$$

$$H(x) = \ln \left(\frac{1}{\beta+1} \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right] - 1 \right)$$

3.5. Odds Rate Function

The odds rate function of the MGZ distribution is defined as

$$O(x) = \frac{F(x;\theta,\beta)}{1 - F(x;\theta,\beta)}$$

$$O(x) = \left(\frac{\left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]}{(\beta+1) - \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]} \right)$$

3.6. Mean Residual Function

The mean residual function of the MGZ distribution is defined as

$$M(x) = \frac{1}{S(x;\theta,\beta)} \int_x^\infty t f(t;\theta,\beta)dt - x$$

$$M(x) = \frac{1}{1 - \frac{1}{\beta+1} \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]} \int_x^\infty t \frac{\theta^3}{\beta+1} \left(\frac{\beta}{2}x^2 + \frac{x}{(2+\theta)}(1+x) \right) e^{-\theta x}\, dt - x$$

using the substitution method and

$$\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt,$$

we obtain

$$M(x) = \left[\frac{\frac{\beta\Gamma(4,\theta x)}{2} + \frac{\theta\Gamma(3,\theta x)+\Gamma(4,\theta x)}{2+\theta}}{\theta(\beta+1) - \left(\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right)} \right] - x$$

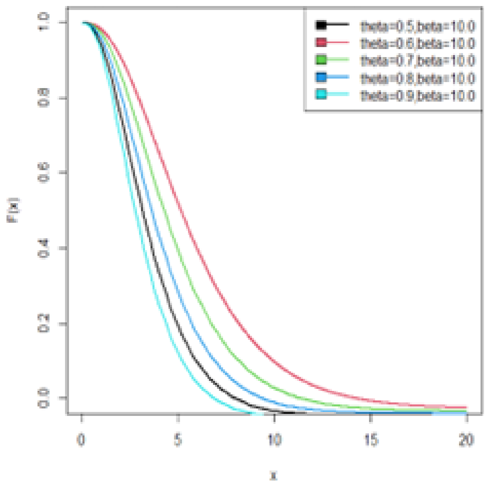

Figure 3: *Survival function of MGZD.*

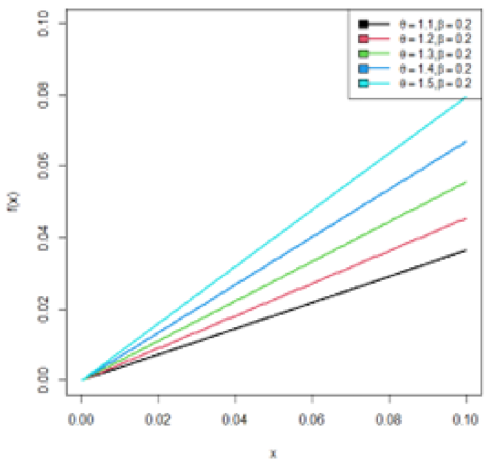

Figure 4: *Hazard function of the MGZD.*

4. Statistical Properties

In this section, we derive the structural properties, moments, moment-generating function, characteristic function, and the r^{th} moment for the MGZ distribution. We also investigate the mean, variance, coefficient of variation, skewness, kurtosis, and dispersion.

4.1. Moments

The r^{th} moment of a random variable X is defined as

$$E(X^r) = \mu_r' = \int_0^\infty x^r f(x;\theta,\beta)\,dx$$

$$\mu_r' = \int_0^\infty x^r \frac{\theta^3}{\beta+1}\left(\frac{\beta}{2}x^2 + \frac{x}{(2+\theta)}(1+x)\right)e^{-\theta x}\,dx$$

on simplifying the integral and using, $\frac{\Gamma(t)}{a^t} = \int_0^\infty x^{t-1}e^{-ax}dx$, we obtain

$$\mu_r' = \frac{1}{\theta^r(\beta+1)}\left[\frac{\beta\Gamma(r+3)}{2} + \frac{\theta\Gamma(r+2)+\Gamma(r+3)}{2+\theta}\right] \tag{5}$$

where $\Gamma(.)$ is the gamma function. By substituting $r = 1, 2, 3$, and 4 in Equation (5), we obtain

$$E(X) = \text{Mean} = \left(\frac{6(\beta+1)+\theta(3\beta+2)}{\theta(\beta+1)(2+\theta)}\right)$$

$$E(X^2) = \left(\frac{24(\beta+1)+6\theta(2\beta+1)}{\theta^2(\beta+1)(2+\theta)}\right)$$

$$E(X^3) = \left(\frac{120(\beta+1)+6\theta(10\beta+4)}{\theta^3(\beta+1)(2+\theta)}\right)$$

$$E(X^4) = \left(\frac{720(\beta+1)+10\theta(36\beta+12)}{\theta^4(\beta+1)(2+\theta)}\right)$$

4.2. Variance

The variance is given by $\sigma^2 = E(X^2) - [E(X)]^2$. On substituting the values, we get

$$\sigma^2 = \left[\frac{24(\beta+1)+6\theta(2\beta+1)}{\theta^2(\beta+1)(2+\theta)}\right] - \left[\frac{6(\beta+1)+\theta(3\beta+2)}{\theta(\beta+1)(2+\theta)}\right]^2$$

after simplification

$$\sigma^2 = \left(\frac{3\beta^2(\theta^2+4\theta+4)+6\beta(\theta^2+4\theta+4)+2(\theta^2+6\theta+6)}{\theta^2(\beta+1)^2(2+\theta)^2} \right)$$

4.3. Coefficient of Variation

The coefficient of variation (C.V.) is defined as

$$C.V. = \frac{\sigma}{\mu}$$

$$C.V. = \frac{\sqrt{3\beta^2(\theta^2+4\theta+4)+6\beta(\theta^2+4\theta+4)+2(\theta^2+6\theta+6)}}{6(\beta+1)+\theta(3\beta+2)}$$

4.4. Skewness

The skewness is given by

$$\text{Skewness} = \sqrt{\beta_1} = \frac{E(X^3)}{\sigma^3}$$

after simplification

$$\text{Skewness} = \left(\frac{(120(\beta+1)+6\theta(10\beta+4))}{\left[(3\beta^2(\theta^2+4\theta+4)+6\beta(\theta^2+4\theta+4)+2(\theta^2+6\theta+6)\right]^{3/2}} \right)$$

4.5. Kurtosis

The kurtosis is defined as

$$\text{Kurtosis} = \beta_2 = \frac{E(X^4)}{\sigma^4}$$

$$\text{Kurtosis} = \left(\frac{[720(\beta+1)+10\theta(36\beta+12)]}{\left[(3\beta^2(\theta^2+4\theta+4)+6\beta(\theta^2+4\theta+4)+2(\theta^2+6\theta+6)\right]^{2}} \right)$$

4.6. Moment Generating Function

The moment-generating function (MGF) of a random variable X is denoted by $M_X(t)$ and is defined as

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x;\theta,\beta)\,dx, \quad t \in \mathbb{R}$$

$$M_X(t) = \frac{\theta^3}{\beta+1}\int_0^\infty e^{tx}\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x}\,dx$$

simplifying and by applying the Taylor series $e^x = \sum_{j=0}^{\infty} \frac{t^j}{j!}\ \forall\ \infty < x < \infty$, we obtain

$$M_X(t) = \int_0^\infty \left(1 + tx + \frac{(tx)^2}{2!} + \dots\right)\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x}\,dx$$

$$M_X(t) = \sum_{j=0}^{\infty}\frac{t^j}{j!}\int_0^\infty x^j f(x;\theta,\beta)\,dx$$

$$M_X(t) = \sum_{j=0}^{\infty}\frac{t^j}{j!}E(x^j)$$

$$M_X(t) = \sum_{j=0}^{\infty}\frac{t^j}{j!\theta^j(\beta+1)}\left[\frac{\beta\Gamma(j+3)}{2} + \frac{\theta\Gamma(j+2)+\Gamma(j+3)}{2+\theta}\right]$$

4.7. Characteristic Function

The characteristic function of a random variable X is denoted by $\phi_X(t)$ and is defined as

$$\phi_X(t) = E(e^{itX}) = \int_0^\infty e^{itx} f(x;\theta,\beta)\, dx$$

using the definition of the moment-generating function

$$\phi_X(t) = M_X(it)$$

$$\phi_X(t) = \sum_{j=0}^{\infty} \frac{(it)^j}{j!\theta^j(\beta+1)} \left[\frac{\beta\Gamma(j+3)}{2} + \frac{\theta\Gamma(j+2)+\Gamma(j+3)}{2+\theta} \right]$$

5. Order Statistics

Let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ be the order statistics of a random sample $X_1, X_2, ..., X_n$ from a continuous population with probability density function $f_X(x)$ and cumulative density function $F_X(x)$. Then, the pdf of the r^{th} order statistic $X_{(r)}$ can be written as

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} [F(x;\theta)]^{r-1} [1-F(x;\theta)]^{n-r} f(x;\theta), \quad x > 0. \tag{6}$$

simplifying Equation (6) and substituting Equations (3) and (4), we obtain

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} \sum_{j=0}^{\infty} (-1)^j \binom{n-r}{j} \left(\frac{1}{\beta+1} \right)^{r+j-1}$$

$$\times \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]^{r+j-1}$$

$$\times \frac{\theta^3}{\beta+1} \left(\frac{\beta}{2} x^2 + \frac{x}{2+\theta}(1+x) \right) e^{-\theta x}$$

First order statistic ($X_{(1)}$)

$$f_{1:n}(x) = \frac{n!}{(n-1)!} \sum_{j=0}^{\infty} (-1)^j \binom{n-1}{j} \left(\frac{1}{\beta+1} \right)^{j}$$

$$\times \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]^{j}$$

$$\times \frac{\theta^3}{\beta+1} \left(\frac{\beta}{2} x^2 + \frac{x}{2+\theta}(1+x) \right) e^{-\theta x}$$

n^{th} Order statistic ($X_{(n)}$)

$$f_{n:n}(x) = \frac{n!}{(n-1)!} \sum_{j=0}^{\infty} (-1)^j \binom{n-1}{j} \left(\frac{1}{\beta+1} \right)^{n+j-1}$$

$$\times \left[\frac{\beta\gamma(3,\theta x)}{2} + \frac{\theta\gamma(2,\theta x)+\gamma(3,\theta x)}{2+\theta} \right]^{n+j-1}$$

$$\times \frac{\theta^3}{\beta+1} \left(\frac{\beta}{2} x^2 + \frac{x}{2+\theta}(1+x) \right) e^{-\theta x}$$

6. Stochastic Ordering

A crucial technique in reliability and finance for evaluating the relative performance of models is stochastic ordering. Let X and Y be two random variables with probability density functions $f(x)$ and $f(y)$, cumulative distribution functions $F(x)$ and $F(y)$, and survival functions $S(x) = 1 - F(x)$ and $S(y) = 1 - F(y)$. Then,

- Mean residual life order (MRLO) is denoted as

$$X \leq_{MRLO} Y, \quad \text{if} \quad m_X(x) \leq m_Y(y), \quad \forall x$$

- Hazard rate order (HRO) is denoted as

$$X \leq_{HRO} Y, \quad \text{if} \quad \frac{S_X(x)}{S_Y(y)} \quad \text{is decreasing for } x \geq 0$$

- Stochastic order (SO) is denoted as

$$X \leq_{SO} Y, \quad \text{if} \quad S_X(x) \leq S_Y(y), \quad \forall x$$

Likelihood ratio ordering proof, assume that X and Y are two independent random variables with probability density functions $f_X(x;\theta,\beta)$ and $f_Y(x;\lambda,\alpha)$. If $\theta < \lambda$ and $\beta < \alpha$, then the likelihood ratio is

$$\Lambda = \frac{f_X(x;\theta,\beta)}{f_Y(x;\lambda,\alpha)}$$

substituting the density functions

$$\Lambda = \frac{\left[\frac{\theta^3}{\beta+1}\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x}\right]}{\left[\frac{\lambda^3}{\alpha+1}\left(\frac{\alpha}{2}x^2 + \frac{x}{2+\lambda}(1+x)\right)e^{-\lambda x}\right]}$$

therefore,

$$\log(\Lambda) = \log\left[\frac{\theta^3(\alpha+1)}{\lambda^3(\beta+1)}\right] + \log\left[\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right] - \log\left[\frac{\alpha}{2}x^2 + \frac{x}{2+\lambda}(1+x)\right] - (\theta - \lambda)x$$

differentiating with respect to x

$$\frac{\partial \log(\Lambda)}{\partial x} = \left(\frac{\left[\beta x + \frac{2x+1}{2+\theta}\right]}{\left[\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right]} - \frac{\left[\alpha x + \frac{2x+1}{2+\lambda}\right]}{\left[\frac{\alpha}{2}x^2 + \frac{x}{2+\lambda}(1+x)\right]} + (\lambda - \theta)\right)$$

hence,

$$\frac{\partial \log(\Lambda)}{\partial x} < 0$$

if $\theta < \lambda$ and $\beta < \alpha$,

Thus, the likelihood ratio is decreasing, proving that the MGZ distribution follows likelihood ratio ordering.

7. Bonferroni and Lorenz Curves

In this section, the Bonferroni and Lorenz curves are derived using the MGZ distribution. These curves are powerful tools for analysing distributions and have diverse applications in fields such as economics, insurance, income inequality, reliability, and medicine. For a random variable X and its probability density function $f(x)$, the Bonferroni and Lorenz curves are constructed. Here, $f(x)dx$ represents the probability that a randomly selected unit falls within a given interval.

$$B(p) = \frac{1}{p\mu_1'} \int_0^q x f(x;\theta,\beta)\, dx \quad \text{and}$$

$$L(p) = \frac{1}{\mu_1'} \int_0^q x f(x;\theta,\beta)\, dx,$$

where $q = F^{-1}(p)$, with $q \in [0,1]$ and

$$\mu_1' = \left(\frac{6(\beta+1) + \theta(3\beta+2)}{\theta(\beta+1)(2+\theta)} \right)$$

Thus, the Bonferroni and Lorenz curves for our distribution are given by

$$B(p) = \frac{1}{p\mu_1'} \int_0^q x \frac{\theta^3}{\beta+1} \left(\frac{\beta}{2}x^2 + \frac{x}{(2+\theta)}(1+x) \right) e^{-\theta x}\, dx.$$

integrating by the substitution method, we obtain

$$B(p) = \left(\frac{\left[\frac{\beta\gamma(3,\theta q)}{2} + \frac{\theta\gamma(2,\theta q)+\gamma(3,\theta q)}{2+\theta} \right]}{p\left[\frac{6\beta(\theta+1)+3\theta\beta+2\theta}{2+\theta} \right]} \right) \tag{7}$$

The Lorenz curve is given by

$$L(p) = pB(p)$$

$$L(p) = \left(\frac{\left[\frac{\beta\gamma(3,\theta q)}{2} + \frac{\theta\gamma(2,\theta q)+\gamma(3,\theta q)}{2+\theta} \right]}{\left[\frac{6\beta(\theta+1)+3\theta\beta+2\theta}{2+\theta} \right]} \right)$$

8. Entropies

In this section, we derived the Renyi entropy, and Tsallis entropy from the MGZ distribution. Entropy and information are widely recognized as measures of uncertainty or randomness within a probability distribution. The entropy functionals of probability distributions are rooted in a variational definition of uncertainty measures, providing a robust framework for analyzing probabilistic systems.

8.1. Renyi Entropy

Entropy is defined as a random variable X is a measure of the variation of the uncertainty. It is used in many fields, such as engineering, statistical mechanics, finance, information theory, biomedicine, and economics. The Renyi entropy of order λ is defined as

$$R_\lambda = \frac{1}{1-\lambda} \log \int_0^\infty [f(x;\theta,\beta)]^\lambda\, dx, \quad \lambda > 0, \lambda \neq 1.$$

$$R_\lambda = \frac{1}{1-\lambda} \log \int_0^\infty \left[\frac{\theta^3}{\beta+1} \left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x) \right) e^{-\theta x} \right]^\lambda dx.$$

using the gamma function $\frac{\Gamma(t)}{a^t} = \int_0^\infty x^{t-1} e^{-ax} dx$, and the binomial theroem $(a+b)^z = \sum_{j=0}^\infty \binom{z}{j} a^j b^{z-j}$, and simplifying the integral, we obtain

$$R_\lambda = \frac{1}{1-\lambda} \log \left(\frac{\theta^3}{\beta+1} \right)^\lambda \sum_{j=0}^\infty \sum_{k=0}^\infty \binom{\lambda}{j} \binom{\lambda-j}{k} \left(\frac{\beta}{2} \right)^j \left(\frac{1}{2+\theta} \right)^{\lambda-j} \frac{\Gamma(2\lambda-k+1)}{(\lambda\theta)^{2\lambda-k+1}}. \tag{8}$$

8.2. Tsallis Entropy

The Boltzmann-Gibbs (B-G) statistical properties introduced by Tsallis have received significant attention. This generalization of B-G statistics was first proposed by Tsallis (1988) through the introduction of Tsallis entropy. For continuous random variables, it is defined as

$$T_\lambda = \frac{1}{\lambda - 1}\left[1 - \int_0^\infty [f(x;\theta,\beta)]^\lambda \, dx\right], \quad \lambda > 0, \lambda \neq 1.$$

$$T_\lambda = \frac{1}{\lambda - 1}\left[1 - \int_0^\infty \left(\frac{\theta^3}{\beta+1}\right)^\lambda \left[\left(\frac{\beta}{2}x^2 + \frac{x}{2+\theta}(1+x)\right)e^{-\theta x}\right]^\lambda \, dx\right].$$

simplifying the integral using the gamma function $\frac{\Gamma(t)}{a^t} = \int_0^\infty x^{t-1}e^{-ax}dx$, and the Taylor series $e^x = \sum_{j=0}^\infty \frac{t^j}{j!} \ \forall \ \infty < x < \infty$, we obtain

$$T_\lambda = \frac{1}{\lambda - 1}\left[1 - \left(\frac{\theta^3}{\beta+1}\right)^\lambda \sum_{j=0}^\infty \sum_{k=0}^\infty \binom{\lambda}{j}\binom{\lambda-j}{k}\left(\frac{\beta}{2}\right)^j \left(\frac{1}{2+\theta}\right)^{\lambda-j} \frac{\Gamma(2\lambda-k+1)}{(\lambda\theta)^{2\lambda-k+1}}\right] \tag{9}$$

9. Estimation of Parameter

The MGZ distribution parameter's maximum likelihood estimates and Fisher's information matrix are provided in this section.

9.1. Maximum likelihood Estimation (MLE) and Fisher's Information Matrix

Consider $x_1, x_2, x_3, , x_n$ be a random sample of size n from the MGZ distribution with parameter θ, β the likelihood function, which is defined as

$$L(x;\theta,\beta) = \prod_{i=1}^n f(x_i;\theta,\beta)$$

$$L = \prod_{i=1}^n \frac{\theta^3}{\beta+1}\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right)e^{-\theta x_i}$$

then, the log-likelihood function is

$$\log L = 3n\log\theta - n\log(\beta+1) + \sum_{i=1}^n \log\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right) - \theta\sum_{i=1}^n x_i$$

differentiating with respect to θ and β, we obtain

$$\frac{\partial \log L}{\partial\theta} = \left(\frac{3n}{\theta}\right) + \sum_{i=1}^n \left(\frac{\left[\frac{x_i + x_i^2}{(2+\theta)^2}\right]}{\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right)}\right) - \sum_{i=1}^n x_i = 0 \tag{10}$$

$$\frac{\partial \log L}{\partial\beta} = -\left(\frac{n}{\beta+1}\right) + \sum_{i=1}^n \left(\frac{\left(\frac{x_i^2}{2}\right)}{\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right)}\right) = 0 \tag{11}$$

The maximum likelihood estimate of the parameters for the MGZ distribution is provided by equations (10) and (11). The equation, however, cannot be solved analytically, so we used R programming and a data set to solve it numerically. The asymptotic normality results are used to derive the confidence interval. Given that if $\lambda = (\theta, \beta)$ represents the MLE of $\lambda = (\theta, \beta)$, the results can be expressed as follows

$$\sqrt{n}(\hat{\lambda} - \lambda) \to N_2(0, I^{-1}(\lambda))$$

in this case, $I(\lambda)$ represents Fisher's Information Matrix,

$$I(\lambda) = -\frac{1}{n}\begin{bmatrix} E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right] & E\left[\frac{\partial^2 \log L}{\partial \theta \partial \beta}\right] \\ E\left[\frac{\partial^2 \log L}{\partial \beta \partial \theta}\right] & E\left[\frac{\partial^2 \log L}{\partial \beta^2}\right] \end{bmatrix}$$

The second derivatives of the log-likelihood function are

$$\frac{\partial^2 \log L}{\partial \theta^2} = -\left(\frac{3n}{\theta^2}\right) + \sum_{i=1}^{n}\left(\frac{\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right)\left(\frac{2\theta(x_i+x_i^2)+4(x_i+x_i^2)}{(2+\theta)^4}\right) - \left(\frac{x_i+x_i^2}{(2+\theta)^2}\right)^2}{\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right)^2}\right)$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = \left(\frac{n}{(\beta+1)^2}\right) - \sum_{i=1}^{n}\left(\frac{\left(\frac{x_i^2}{2}\right)^2}{2\left(\frac{\beta}{2}x_i^2 + \frac{x_i}{2+\theta}(1+x_i)\right)^2}\right)$$

$$\frac{\partial^2 \log L}{\partial \theta \partial \beta} = \sum_{i=1}^{n}\left(\frac{x_i^3(1+x_i)}{2\left(\theta^2 + 4(\theta+1)\right)}\right)$$

10. Applications

In this section, the proposed MGZ distribution is applied to the real data sets.

Dataset 1: Survival time of breast cancer patients (in months) of patients. The secondary data has been taken from a private tertiary care hospital in Puducherry. 63, 61, 58, 57, 55, 26, 5, 20, 54, 58, 30, 24, 35, 27, 47, 67, 70, 78, 39, 26, 65, 77, 49, 50, 27, 41, 65, 52, 48, 62, 38, 26.

Dataset 2: We have used the data representing the size of the Tumour (in mm) of 210 patients with node-positive breast cancer, as reported by Royston and Altman (2013). The entire data set contains patient records from a 1984-1989 trial conducted by the German Breast Cancer Study Group (GBSG) of 720 patients with node-positive breast cancer; it retains the 686 patients with complete data for the predictive variables. We used the initial 210 sample values for the analysis. 18, 20, 40, 25, 30, 52, 21, 20, 20, 30, 12, 30, 21, 25, 57, 25, 21, 27, 12, 45, 18, 18, 30, 30, 25, 14, 21, 30, 23, 30, 55, 11, 40, 25, 15, 29, 15, 45, 40, 30, 60, 21, 20, 58, 30, 28, 25, 18, 40, 35, 30, 21, 45, 40, 20, 19, 52, 80, 40, 20, 80, 30, 33,16, 25, 39, 52, 50, 39, 42, 65, 28, 25, 23, 32, 15, 40, 35, 25, 30, 22, 45, 20, 27, 29, 22, 24, 25, 23, 30, 50, 45, 7, 8, 18, 16, 28, 21, 15, 20, 12, 20, 30, 50, 25, 30, 25, 32, 35, 20, 28, 26, 35, 23, 80, 20, 20, 34, 30, 23, 35,70, 20, 45, 24, 28, 35, 25, 60, 35, 40, 23, 40, 21, 35, 55, 120, 23, 50, 55, 16, 22, 21, 12, 14, 60, 55, 120, 23, 50, 55, 16, 22, 21, 12, 14, 60, 30, 25, 40, 48, 35, 25, 28, 35, 30, 35, 22, 12, 21, 35, 52, 36, 25, 31, 20, 30, 25, 44, 21, 30, 32, 50, 21, 20, 60, 35, 30, 20, 23, 49, 55, 20, 25, 22, 18, 30, 20, 40, 35, 13, 70, 27, 21, 25, 80, 36, 32, 25, 19.

Dataset 3: Age of the patients with advanced lung cancer in years (J. F. Lawless (2003)). 69, 64, 38, 63, 65, 49, 69, 68, 43, 70, 81, 63, 63, 52, 48, 61, 42, 35, 63, 56, 55, 67, 63, 65, 46, 53, 69, 68, 43, 55, 42, 64, 65, 65, 55, 66, 60, 67, 53, 62, 67, 72, 48, 68, 67, 61, 60, 62, 38, 50, 63, 64, 43, 34, 66, 62, 52, 47, 63, 68, 45, 41, 66, 62, 60, 66, 38, 53, 37, 54, 60, 48, 52, 70, 50, 62, 65, 58, 62, 64, 63, 58, 64, 52, 35, 63, 70, 51, 40, 69, 36, 71, 62, 60, 44, 54, 66, 49, 72, 68, 62, 71, 70, 61, 71, 59, 67, 60, 69, 57, 39, 62, 50, 43, 70, 66, 61, 81, 58, 63, 60, 62, 42, 69, 63, 45, 68, 39, 66, 63, 49, 64, 65, 64, 67, 65, 37.

We observe its flexibility over some well-known existing distributions. The results for the analysis in this present study are obtained using R software. We have also calculated the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Akaike Information Criteria Corrected (AICC), and -2logL for the considered distributions to observe their fit. The distribution with the value of lowest AIC and BIC and AICC is considered as the best. The Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Akaike Information Criteria Corrected (AICC), and -2logL are used to compare the goodness of fit of the fitted distribution. The following formula can be used to determine AIC, BIC, AICC, and -2logL.

$$\text{AIC} = 2k - 2\log L, \quad \text{BIC} = k\log n - 2\log L, \quad \text{and AICC} = \text{AIC} + \frac{2k(k+1)}{(n-k-1)}$$

Where, k= number of parameters, n sample size, and -2logL is the maximized value of the log-likelihood function.

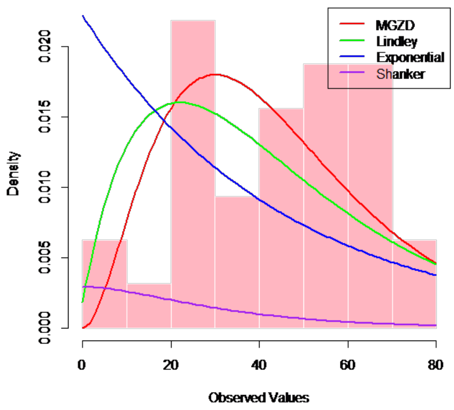

Figure 5: *Fitting density curve for dataset 1*

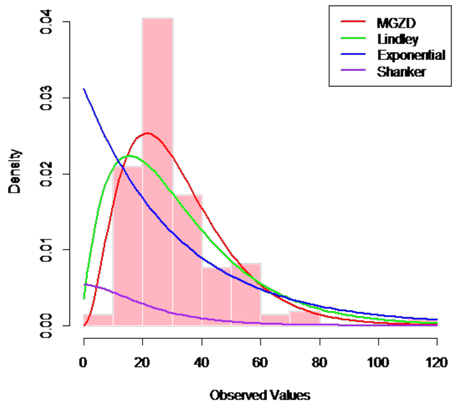

Figure 6: *Fitting density curve for dataset 2*

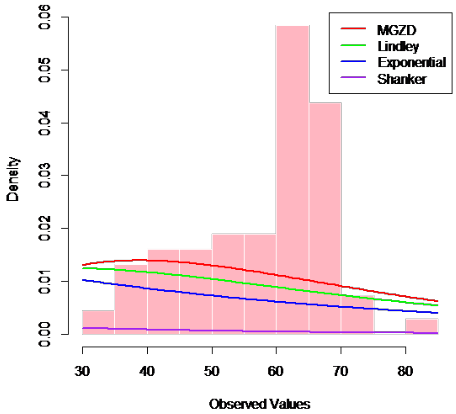

Figure 7: *Fitting density curve for dataset 3*

Table 1: *The MLEs of MGZD parameters and AIC, BIC, AICC, and -2logL for Data Set 1.*

Distributions	ML Estimates	-2logL	AIC	BIC	AICC
Mixture of Gamma and Zeghdoudi	$\hat{\theta} = 0.05542\,(0.0067)$ $\hat{\beta} = 0.00403\,(0.00017)$	203.2229	207.2229	209.4939	207.8229
Exponential	$\hat{\theta} = 0.0185\,(0.0038)$	229.6043	231.6043	232.3970	231.8043
Lindley	$\hat{\theta} = 0.0363\,(0.00535)$	213.2929	215.2929	216.4284	215.4929
Shanker	$\hat{\theta} = 0.0369\,(0.00543)$	212.536	214.536	215.6727	214.736

Table 2: *The MLEs of MGZD parameters and AIC, BIC, AICC, and -2logL for Data Set 2.*

Distributions	ML Estimates	-2logL	AIC	BIC	AICC
Mixture of Gamma and Zeghdoudi	$\hat{\theta} = 0.09355\,(0.00372)$ $\hat{\beta} = 0.00382\,(0.00593)$	1717.616	1721.616	1728.310	1721.674
Exponential	$\hat{\theta} = 0.03119\,(0.00215)$	1876.483	1878.483	1881.830	1878.502
Lindley	$\hat{\theta} = 0.06059\,(0.00295)$	1770.519	1772.519	1775.866	1772.539
Shanker	$\hat{\theta} = 0.07918\,(0.00348)$	2753.765	2755.765	2759.112	2755.784

Table 3: *The MLEs of MGZD parameters and AIC, BIC, AICC, and -2logL for Data Set 3.*

Distributions	ML Estimates	-2logL	AIC	BIC	AICC
Mixture of Gamma and Zeghdoudi	$\hat{\theta} = 0.05143\,(0.00253)$ $\hat{\beta} = 14.3464\,(41.4323)$	1233.039	1237.039	1242.878	1237.128
Exponential	$\hat{\theta} = 0.01716\,(0.00146)$	1388.006	1390.006	1392.926	1390.036
Lindley	$\hat{\theta} = 0.03374\,(0.00203)$	1291.502	1293.502	1296.422	1293.532
Shanker	$\hat{\theta} = 0.04516\,(0.00243)$	2092.173	2094.173	2097.093	2094.203

From table 1, 2 and table 3, the (MGZD) mixture gamma and zeghdoudi distributions has the lowest AIC, BIC and AICC values, thus making it to be fitted better than the (Exponential, Lindley and Shanker) distributions. Figure 5, 6 and 7 also explicates the better model.

11. Conclusion

This study introduces a novel two-parameter distribution known as the Modified Gamma-Zeghdoudi Distribution (MGZD). This distribution is a compound mixture of the Zeghdoudi and Gamma distributions, two well-known statistical models. Essential features of the MGZD, such as survival and hazard functions, have been analyzed. We have explored some of the important statistical properties, such as entropies, Bonferroni and Lorenz curves, moments (mean, variance, skewness, and kurtosis), order statistics, and stochastic ordering. We have estimated the parameters of the MGZD are estimated using the maximum likelihood estimation method.

To assess the effectiveness, the MGZD is applied to three real-world data sets and compared with well-known distributions. The results illustrate that the MGZD gives a better fit to the data, proving its superiority over other known classical models in enclosing the above data structure.

References

[1] Alam, F., Alrajhi, S., Nassar, M., and Afify, A. Z. (2021). Modeling liver cancer and leukemia data using arcsine-Gaussian distribution. *Computers, Materials and Continua (CMC)*, 67(2), 2185–2202.

[2] Al-Kadim, K. A., and Mahdi, A. A. (2018). Exponentiated transmuted exponential distribution. *Journal of University of Babylon for Pure and Applied Sciences*, 26(2), 78–90.

[3] American Cancer Society. (2021). Breast Cancer Survival Rates. Available at: https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html

[4] David, H. A. (1970). *Order Statistics*. Wiley and Sons, New York.

[5] Ghitany, M. E., Atieh, B., and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, 78(4), 493–506.

[6] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

[7] Messaadia, H., and Zeghdoudi, H. (2018). Zeghdoudi distribution and its applications. *International Journal of Computing Science and Mathematics*, 9(1), 58–65.

[8] Macdonald, P. (1971). An estimation procedure for mixtures of distributions. *Journal of the Royal Statistical Society*, 20, 102–107.

[9] R©nyi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Vol. 4, 547–562. University of California Press.

[10] R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[11] Royston, P., and Altman, D. G. (2013). External validation of a Cox prognostic model: Principles and methods. *BMC Medical Research Methodology*, 13(1), 33. https://doi.org/10.1186/1471-2288-13-33

[12] Sakthivel, M.,and Pandiyan, P. (2024). Stochastic Modelling on New Mixture Distribution with Properties and their Application. *International Journal of Mathematics, Statistics, and Computer Science*, 2, 259–275. https://doi.org/10.59543/ijmscs.v2i.9015

[13] Shanker, R. (2016). Aradhana distribution and its applications. *International Journal of Statistics and Applications*, 6(1), 23–34.

[14] Shanker, R. (2022). Uma distribution properties and application. *Biometrics Biostatistics International Journal*, 11(5), 165–169.

[15] Shanker R. (2015). Shanker distribution and its applications. *International Journal of Statistics and Applications*, a; 5(6):338–348, doi: org/10.5923/j.statistics.20150506.08.

[16] Shanker R, Hagos F, Sujatha S. (2015).On modeling of lifetimes data using exponential and Lindley distributions. *Biom Biostat Int J*; 2(5):140–147, doi.org/10.15406/bbij.2015.02.00042.

[17] Shanker R. (2015). Akash distribution and its applications. *International Journal of Probability and Statistics*. b;4(3):65–75, doi: 10.5923/j.statistics.20150506.08.

[18] Shanker R. (2016) Garima distribution and its application to model behavioral science data. *Biometrics and Biostatistics International Journal*, 4(7):1–9, doi.org/10.15406/bbij.2016.04.00116

[19] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.

[20] World Health Organization. (2011). *Global Status Report on Noncommunicable Diseases 2010*. Available at: https://www.who.int/nmh/publications/ncd_report_full_en.pdf