Asian Journal of Statistical Sciences
Vol. 4, No-2, (2024), pp. 121-140

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

Gamma Poisson distribution and its engineering applications

C. Satheesh Kumar[a] and Satheenthar A. S.[b]

[a] *Department of Statistics, University of Kerala, Kariavattom, Thiruvananthapuram, India*
[b] *Department of Statistics, Government College Kariavattom, Thiruvananthapuram, India*

ARTICLE HISTORY
Compiled April 3, 2025

Received 23 December 2023; Accepted 31 July 2024

ABSTRACT
In this paper, we consider an extended version of the exponential Poisson distribution and examine its theoretical properties. We derive expressions for the cumulative distribution function, survival function, failure rate function, pdf of the order statistics and raw moments. We also discuss the maximum likelihood estimation procedures and the Expectation-Maximization algorithm for estimating the parameters of this distribution. Additionally, a statistical test is proposed to assess the significance of the additional parameter introduced in the model. To demonstrate its practical utility, we provide certain real-life data applications. Furthermore, with the help of simulated datasets, it is shown that as the sample size increases, the average bias and mean squared errors of the maximum likelihood estimators decrease in a consistent manner.

KEYWORDS
EM algorithm; Gamma distribution; Maximum likelihood estimation; Poisson distribution

1. Introduction

In recent years, the approach of mixing different lifetime models has gained significant interest, and been widely applied to complex phenomena across various fields. Standard distributions, such as the exponential and gamma distributions are very effective for modeling lifetime data and are commonly used in areas such as reliability analysis, finance, insurance, economics, engineering, and beyond. In fact, these distributions are often modified to better fit real-world data and scenarios, making them more applicable in practical situations. The exponential and gamma distributions have become central to reliability studies. They are frequently used to model the life expectancy of systems, as they reflect different aspects of the failure process, such as wear and tear, aging, or fatigue. Mixture models that combine these classical distributions, like the exponential and zero-truncated Poisson (ZTP) distributions, are also commonly used to characterize a decreasing failure rate. For example see [3], [9], [14], [6] and [8]. Now the gamma distribution (GD) has been used quite extensively in reliability and survival analysis, particularly when the data are not censored. Among these, the

CONTACT Author[b]. Email: sathystat18@gmail.com

gamma distribution (GD) has been used quite extensively in reliability and survival analysis, especially in cases where data is not censored. The gamma distribution has two key parameters such as shape parameter and scale parameter. Depending on the values of these parameters, the hazard function of the gamma distribution exhibits various behaviors such as increasing, decreasing, or constant in modeling failure rates. In real practice, many unforeseen situations arise, which reduces the choice of the failure model. In such cases, extensions of these classical models offer valuable tools for capturing the intricacies of failure and lifetime analysis in practical scenarios. In this paper, we introduce a new class of lifetime distributions by compounding the GD and ZTP distributions. This results in a failure rate that can take various flexible shapes, including non monotone patterns, which enhances the model's ability to fit complex datasets, making it more appealing.
This paper is organized as follows. Section 2 introduces the genesis of the GPD and outlines some of its key properties. Section 3 presents various methods for estimating the parameters of the distribution. In Section 4, we illustrate the practical utility of the model through two real-world data sets. Finally, Section 5 includes a concise simulation study to evaluate the performance of the maximum likelihood estimators for the parameters of the distribution.

2. Genesis and motivation of the new family

Suppose that a system has N subsystems functioning independently at a given time where N has ZTP distribution with parameter λ. It is the conditional probability distribution of a Poisson-distributed random variable, given that the value of the random variable is not zero. The probability mass function (p.m.f) of N is given by for

$$P(N=n) = \frac{e^{-\lambda}\ \lambda^{n}}{(1-e^{-\lambda})\Gamma(n+1)}, \tag{1}$$

where $n = 1, 2, \ldots,$ and $\lambda > 0$. Let $W_1, W_2, \ldots$ be a sequence of independent and identically distributed gamma random variables with the following probability density function (pdf).

$$f(w;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\ e^{-\beta w} w^{\alpha-1}, \tag{2}$$

in which $w > 0,\ \alpha > 0$ and $\beta > 0$.
Let W_i denote the failure time of the i^{th} subsystem and Z denote the time to failure of the first out of the N functioning subsystems. We can write $Z = min\{W_1, W_2, \ldots, W_N\}$. Then the conditional density function of Z given $N = n$ is given by

$$f_1(z \mid N=n;\ \alpha,\ \beta) = n\Big(\frac{\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}\Big)^{n-1} \frac{\beta^{\alpha}}{\Gamma(\alpha)}\ e^{-\beta z} z^{\alpha-1},$$

in which $\Gamma(\alpha,\beta z) = \int_{\beta z}^{\infty} t^{\alpha-1}e^{-t}\,dt$ and $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}\,dt$. Now the marginal density function of Z is the following, for $z > 0,\ \alpha > 0,\ \beta > 0$ and $\lambda > 0$.

$$\begin{aligned} g(z) &= \sum_{n=1}^{\infty} n\Big(\frac{\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}\Big)^{n-1} \frac{\beta^{\alpha}}{\Gamma(\alpha)}\ e^{-\beta z} z^{\alpha-1} \frac{e^{-\lambda}\ \lambda^{n}}{(1-e^{-\lambda})\Gamma(n+1)} \\ &= \frac{\lambda\beta^{\alpha}\ z^{\alpha-1}e^{-\lambda-\beta z}}{(1-e^{-\lambda})\Gamma(\alpha)} \sum_{n=0}^{\infty} \frac{(n+1)\ n!\ (\lambda\frac{\Gamma(\alpha,\beta z)}{\Gamma(\alpha)})^{n}}{(n+1)!\ n!}, \end{aligned} \tag{3}$$

which reduces to

$$g(z) = \frac{\lambda\beta^{\alpha}\ z^{\alpha-1}e^{-\lambda-\beta z+\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}}{(1-e^{-\lambda})\Gamma(\alpha)}. \tag{4}$$

A distribution with pdf (4), we call "the gamma Poisson distribution" or in short "the GPD". Here the parameters α and β control the shape of the distribution whereas λ control the scale of the distribution. The special cases of the GPD's are

- When $\alpha = 1$ the GPD approaches to the exponential Poisson distribution (EPD) of [9]
- When $\lambda \to 0$ the GPD tends to the gamma distribution
- If $\alpha = 1$ and $\lambda \to 0$ GPD leads to exponential distribution with parameter β.

With different values of the parameters, different curvature forms of this pdf are obtained as shown in Figure 1.

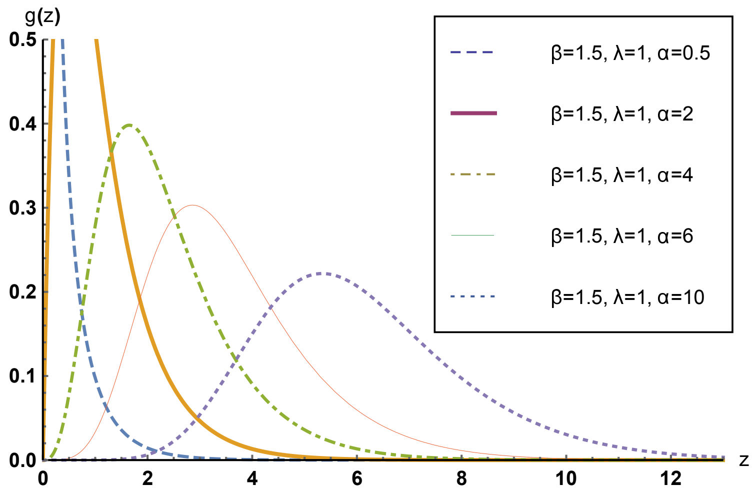


Figure 1. Probability density plots of the GPD for particular values of α= 0.5, 2, 4, 6 and 10.

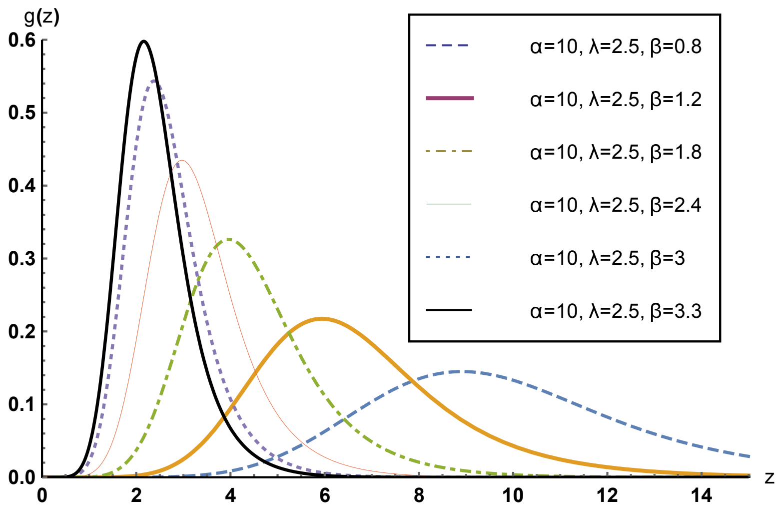


Figure 2. Probability density plots of the GPD for particular values of β= 0.8, 1.2, 1.8, 2.4, 3 and 3.3.

From the Figures 1 and 2, the curves may be decreasing or unimodal, with a broad range of skewness, peakedness, and plateness, but they are often almost symmetrical or positively skewed. Moreover GPD can be used to model a wide range of lifetime phenomena.
Now we have the following results.

Result 2.1. *The cumulative distribution function (c.d.f) of the GPD is given by*

$$G(z) = \frac{1 - e^{-\lambda + \frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}}{1 - e^{-\lambda}}, \tag{5}$$

in which $z > 0$, $\alpha > 0$, $\beta > 0$ *and* $\lambda > 0$.

The proof is included in Appendix A.

Result 2.2. *For* $z > 0$, $\alpha > 0$, $\beta > 0$ *and* $\lambda > 0$, *the survival function of the GPD is given by*

$$S(z) = \frac{e^{-\lambda + \frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}} - e^{-\lambda}}{1 - e^{-\lambda}}. \tag{6}$$

Proof is straightforward; hence omitted.

Result 2.3. *The failure rate function of GPD is given by*

$$h(z) = \frac{\lambda\beta^{\alpha}\ z^{\alpha-1}\ e^{-\beta z}}{\Gamma(\alpha)(1 - e^{-\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}})}. \tag{7}$$

Proof is straightforward from the definition of failure rate function in the light of (4) and (6); hence omitted.
The failure rate plots of the model for particular values of the parameters α and β are included in Figures 3 and 4 respectively.

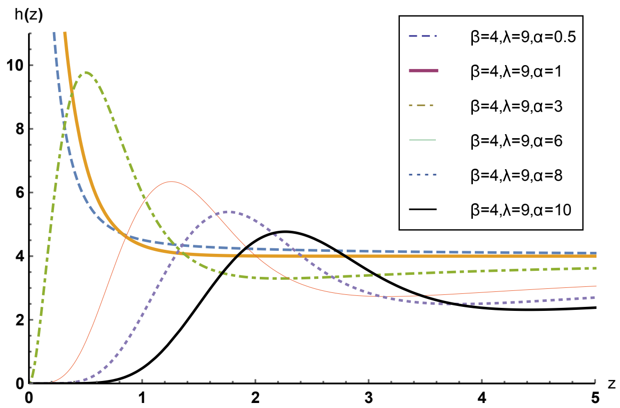


Figure 3. Gamma Poisson failure rates for particular values of α= 0.5, 1, 3, 6, 8 and 10.

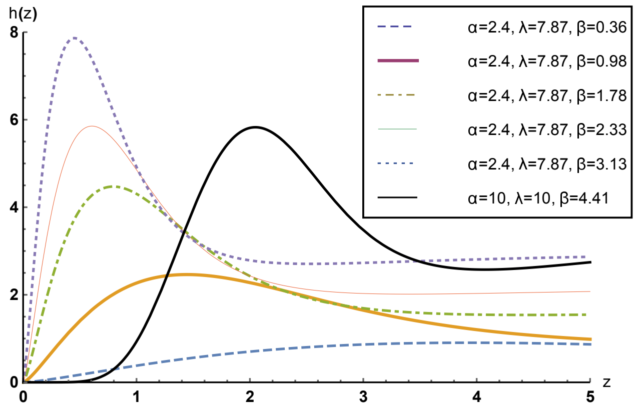


Figure 4. Gamma Poisson failure rates for particular values of β= 0.36, 0.98, 1.78, 2.33, 3.13 and 4.41.

Result 2.4. *The pdf of $g_{i:m}(.)$ of the i^{th} order statistic Z_i of a random sample $Z_1, Z_2, \ldots, Z_m$ taken from GPD is given by*

$$g_{i:m}(z) = \frac{m!\lambda\beta^{\alpha}\ z^{\alpha-1}e^{-\lambda-\beta z+\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}}{\Gamma(\alpha)(i-1)!(m-i)!} \frac{\left(1-e^{-\lambda+\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}\right)^{i-1}}{(1-e^{-\lambda})^{i}} \times \left(\frac{e^{\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}-1}{e^{\lambda}-1}\right)^{m-i}. \tag{8}$$

Proof follows from the following general formula for the pdf of the i^{th} order statistic of a continuous random variable.

$$g_{i:m}(z) = \frac{m!}{(i-1)!(m-i)!} g(z)G(z)^{i-1}(1-G(z))^{m-i},$$

in which $g(.)$ and $G(.)$ are respectively the pdf and cdf of the corresponding distribution.
Next we obtain an expression for the r^{th} raw moment of the GPD through the following proposition.

Result 2.5. *For $r \geq 1$, the r^{th} raw moment $\mu_r^{'}$ of the GPD with pdf (4) is the following.*

$$\mu_r^{'} = \frac{\beta e^{-\lambda}}{1 - e^{-\lambda}} \int_0^{\lambda} [F^{-1}(1 - \frac{u}{\lambda})]^r \; e^u du, \tag{9}$$

where $F^{-1}(.)$ is the inverse cdf of gamma distribution.

Now one can compute the r^{th} raw moment numerically by using some mathematical softwares.

Corollary 2.6. *The first and second raw moments of the GPD are respectively*

$$E(Z) = \frac{\beta e^{-\lambda}}{1 - e^{-\lambda}} \int_0^{\lambda} (F^{-1}(1 - \frac{u}{\lambda})) \; e^u du \tag{10}$$

and

$$E(Z^2) = \frac{\beta e^{-\lambda}}{1 - e^{-\lambda}} \int_0^{\lambda} (F^{-1}(1 - \frac{u}{\lambda}))^2 \; e^u du. \tag{11}$$

Consequently, the variance of the GPD is

$$Var(Z) = \frac{\beta e^{-\lambda}}{1 - e^{-\lambda}} \int_0^{\lambda} (F^{-1}(1 - \frac{u}{\lambda}))^2 \; e^u du - [E(Z)]^2. \tag{12}$$

3. Estimation

In this section we present various estimation procedures for obtaining the estimators of the parameter α, β and λ of the GPD with pdf (4) and discuss asymptotic variances and covariances of the maximum likelihood estimates (MLEs) together with a test procedure.

3.1. Method of maximum likelihood

The log-likelihood function based on a random sample $Z_1, Z_2, \ldots, Z_k$ taken from the GPD is the following, in which $\xi = (\alpha,\ \beta,\ \lambda)$.

$$\begin{aligned}\ell(\xi) = k \ln(\lambda) + k\alpha \ln(\beta) + (\alpha - 1)\sum_{i=1}^{k} \ln z_i - k\ \lambda - \beta \sum_{i=1}^{k} z_i + \frac{\lambda}{\Gamma(\alpha)} \sum_{i=1}^{k} \Gamma(\alpha, \beta z_i) \\ - k \ln \Gamma(\alpha) - k \ln (1 - e^{-\lambda})\end{aligned} \tag{13}$$

The maximum likelihood estimators of $\xi = (\alpha,\ \beta,\ \lambda)$ can be directly obtained by maximizing the log likelihood function (13) or alternatively, by finding solution for the following four likelihood equations, in which

$$\psi^{(k)}(\alpha) = \frac{\partial^k \ \ln\Gamma(\alpha)}{\partial \alpha^k} \tag{14}$$

and

$$T(m,s,x) = G^{m,0}_{m-1,m}\left(\begin{array}{c} 0,0\ldots,0 \\ s-1,-1\ldots,-1 \end{array} \middle| x \right) \tag{15}$$

which is a particular form of the Meijer G-function.

$$\frac{\partial \ell(\xi)}{\partial \alpha} = k\ \ln\ \beta + \sum_{i=1}^{k} \ln\ z_i + \frac{\lambda}{\Gamma(\alpha)} \sum_{i=1}^{k} \big[\ln\ \beta z_i\ \Gamma(\alpha,\beta z_i) + \beta z_i\ T(3,\alpha,\beta z_i) - \psi(\alpha)\Gamma(\alpha,\beta z_i)\big] - k\ \psi^{(0)}(\alpha) = 0, \tag{16}$$

$$\frac{\partial \ell(\xi)}{\partial \beta} = \frac{k\ \alpha}{\beta} - \sum_{i=1}^{k} z_i - \frac{\lambda}{\Gamma(\alpha)} \sum_{i=1}^{k} (\beta z_i)^{\alpha-1} e^{-\beta z_i} = 0 \tag{17}$$

and

$$\frac{\partial \ell(\xi)}{\partial \lambda} = k\left(\frac{1}{\lambda} - 1 - \frac{1}{e^{\lambda}-1}\right) + \sum_{i=1}^{k} \frac{\Gamma(\alpha,\beta z_i)}{\Gamma(\alpha)} = 0. \tag{18}$$

Now for solving likelihood equations one can utilize any of the mathematical softwares like MATHEMATICA, R softwares etc.

3.2. An EM algorithm

The expectation maximization(EM) algorithm introduced by [7] is one of the best methods to obtain MLEs in case of mixture models. This method is a powerful tool for handling the missing data (or incomplete data) situations. It is an iterative method by frequently replacing the missing values in the data with estimated values and updating the parameter estimates. If the amount of information in the missing data is large, it will converge slowly when compared to Newton-Raphson method and it is more reliable than any other method.
To start the algorithm, we have the complete data distribution defined with density function

$$\begin{aligned} g(z,n;\xi) &= g(z \mid N=n;\alpha,\beta)\ P(n;\lambda) \\ &= n\Big(\frac{\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}\Big)^{n-1} \frac{\beta^{\alpha}}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} \frac{e^{-\lambda}\ \lambda^{n}}{\Gamma(n+1)(1-e^{-\lambda})}, \end{aligned} \tag{19}$$

where $n = 1,2,\ldots,\ z>0,\ \alpha>0,\ \beta>0$ and $\lambda>0$.

Now, the E-step for the algorithm can be derived as given below. Let $\xi^{(h)} = (\alpha^{(h)},\ \beta^{(h)},\ \lambda^{(h)})$ be the initial estimate. Then the conditional expectation of $(N \mid Z;\xi^{(h)})$ can be computed as follows.

$$\begin{aligned} P(N \mid Z;\xi^{(h)}) &= \frac{g(z,n;\xi)}{g(z)} \\ &= \frac{\left(\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}\right)^{n-1} e^{-\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}}{(n-1)!} \end{aligned} \tag{20}$$

and

$$E(N \mid Z; \xi^{(h)}) = \sum_{n=1}^{\infty} n \, P(N \mid Z; \xi^{(h)}) = \left(1 + \frac{\lambda\Gamma(\alpha, \beta z)}{\Gamma(\alpha)}\right). \quad (21)$$

The EM cycle is completed with M-step, which is complete data maximum likelihood over ζ, with missing M's replaced by their conditional expectations. Then the log-likelihood function corresponding to the joint density function in (19) is

$$\ell(\xi^{(h)}) = \sum_{i=1}^{k} \ln n_i + \sum_{i=1}^{k} (n_i - 1) \ln (\lambda \frac{\Gamma(\alpha, \beta z_i)}{\Gamma(\alpha)}) + k\alpha \ln(\beta) + (\alpha - 1) \sum_{i=1}^{k} \ln z_i - \beta \sum_{i=1}^{k} z_i - k \lambda + \sum_{i=1}^{k} n_i \ln \lambda - k \ln \Gamma(\alpha) - \sum_{i=1}^{k} \ln \Gamma(n_i + 1) - k \ln (1 - e^{-\lambda}), \quad (22)$$

from which the corresponding estimators are obtained as given below, in which $\psi(.)$ and $T(.)$ are as defined in (14) and (15).

$$\frac{\partial \ell(\xi^{(h)}; z_i)}{\partial \alpha} = 0 \implies \sum_{i=1}^{k} \frac{(n_i - 1)}{\lambda\Gamma(\alpha, \beta z_i)} \big[\ln \beta z_i \, \Gamma(\alpha, \beta z_i) + \beta z_i \, T(3, \alpha, \beta z_i) - \psi(\alpha)\Gamma(\alpha, \beta z_i)\big] - k \ln \beta + \sum_{i=1}^{k} \ln z_i - k\psi(\alpha) = 0 \quad (23)$$

$$\frac{\partial \ell(\xi^{(h)}; z_i)}{\partial \beta} = 0 \implies \sum_{i=1}^{k} \frac{(n_i - 1) e^{-\beta z_i} (\beta z_i)^{\alpha - 1}}{\Gamma(\alpha, \beta z_i)} + \frac{k\alpha}{\beta} - \sum_{i=1}^{k} z_i = 0 \quad (24)$$

and

$$\frac{\partial \ell(\xi^{(h)}; z_i)}{\partial \lambda} = 0 \implies \frac{\sum_{i=1}^{k} n_i - 1}{\lambda} - k + \frac{\sum_{i=1}^{k} n_i}{\lambda} - \frac{k \, e^{-\lambda}}{(1 - e^{-\lambda})} = 0, \quad (25)$$

where $n_i = \left(1 + \frac{\lambda\Gamma(\alpha, \beta z)}{\Gamma(\alpha)}\right)$, obtained from the E-step.

3.3. Asymptotic variances and covariances of the MLEs

The asymptotic variance examines the quality of the MLEs. The large sample approximation of the MLE's of $\xi = (\xi_1, \ \xi_2, \ \xi_3) = (\alpha, \ \beta, \ \lambda)$, is approximately trivariate normal with mean ξ and variance–covariance matrix, which is the inverse of the expected information matrix $H(\xi) = E(I; \xi))$ where $I = I(\xi; z)$ is the observed information matrix with elements $I_{ij} = -\frac{\partial^2 l}{\partial \xi_i \xi_j}$ with $i, j = 1, 2, 3$ and the expectation is to be taken with respect to the distribution of Z. Differentiating (16), (17) and (18) the elements of the symmetric third order observed information matrix are found.
Elements of the observed information matrix are derived in terms of $I_{11}, \ I_{12}, \ I_{13}, \ I_{22}, \ I_{23}$ and I_{33}, which are given in Appendix B. The inverse of expected information matrix evaluated using the parameter estimates gives the

asymptotic variance-covariance matrix of the MLEs.
For interval estimation, let $\widehat{Var}(\hat{\alpha}), \widehat{Var}(\hat{\beta})$ and $\widehat{Var}(\hat{\lambda})$ denote the estimates of the main diagonal elements of the inverse of the observed information matrix, evaluated at the MLE of the parameters. The large-sample 95% confidence intervals (CI) for the parameters α, β and λ are

$$\hat{\alpha} \pm Z_{\frac{\alpha_1}{2}} * \sqrt{\widehat{Var}(\hat{\alpha})}, \qquad \hat{\beta} \pm Z_{\frac{\alpha_1}{2}} * \sqrt{\widehat{Var}(\hat{\beta})}$$

and

$$\hat{\lambda} \pm Z_{\frac{\alpha_1}{2}} * \sqrt{\widehat{Var}(\hat{\lambda})}$$

respectively, where $Z_{\frac{\alpha_1}{2}}$ is the upper $\frac{\alpha_1}{2}$ quantile of the standard normal distribution and α_1 is the 5% level of significance.

3.4. Testing

3.4.1. Generalized Likelihood Ratio Test

Here we consider the generalized likelihood ratio test (GLRT) procedure for testing the significance of the parameters α and λ of the GPD (α, β, λ). Now we test the null hypothesis H_0 : EPD versus the alternate hypothesis H_1 : GPD or equivalently $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$. Similarly, we test the null hypothesis H_0 : GD versus the alternate hypothesis H_1 : GPD or equivalently $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$. The test statistic is

$$LRT = -2 \, ln \left[\frac{\sup_{\xi \in \Xi_0} \mathrm{L}(\hat{\xi^*} \mid y)}{\sup_{\xi \in \Xi_1} \mathrm{L}(\hat{\xi} \mid y)} \right], \tag{26}$$

in which $\hat{\xi}$ is the maximum likelihood estimates of $\xi = (\alpha, \beta, \lambda)$ with no restriction, and $\hat{\xi^*}$ is the maximum likelihood estimates of ξ when $\alpha = 1$ and $\lambda = 0$ respectively. The test statistic given in (26) follows chi-square distribution with one and two degree of freedom respectively.

4. Illustrative examples

In this section, the GPD is applied to model two complete data sets. For illustrating the usefulness of the GPD$(\alpha,\ \beta,\ \lambda)$, here we compare the proposed model with the exponentiated Weibull distribution (EWD) by [12], the generalization of exponential geometric distribution (GEGD) by [10], the generalized exponential Poisson distribution (GEPD) by [4] and Exponentiated Lomax distribution (ELD) by [2].
EWD(α, γ, λ)

$$f_{EWD}(z) = \alpha\gamma\lambda^{\gamma} z^{\gamma-1}(1 - e^{-(\lambda z)^{\gamma}})^{\alpha-1} e^{-(\lambda z)^{\gamma}},$$

in which $z > 0$, $\alpha > 0$, $\gamma > 0$ and $\lambda > 0$.

GEGD(α, β, p)

$$f_{GEGD}(z) = \frac{\alpha\beta(1-p)e^{-\beta z}(1-e^{-\beta z})^{\alpha-1}}{(1-pe^{-\beta z})^{\alpha+1}},$$

in which $z > 0$, $\alpha > 0$, $\beta > 0$ and $p \in (0, 1)$.

GEPD(α, β, λ)

$$f_{GEPD}(z) = \frac{\alpha\ \beta\ \lambda\ e^{-\lambda-\beta z+\lambda e^{-\beta z}}\left(1-e^{-\lambda+\lambda e^{-\beta z}}\right)^{\alpha-1}}{(1-e^{-\lambda})^{\alpha}},$$

in which $z > 0$, $\alpha > 0$, $\beta > 0$ and $\lambda > 0$.

ELD(α, λ, θ)

$$f_{ELD}(z) = \theta\alpha\lambda(1+\alpha z)^{-(\lambda+1)}(1-(1+\alpha z)^{-\lambda})^{\theta-1},$$

in which $z > 0$, $\alpha > 0$, $\lambda > 0$ and $\theta > 0$.

First we consider an uncensored data set from [11]. The data gives 100 observations on breaking stress of carbon fibres (in Gba) is

Data Set 1

3.7, 3.11, 4.42, 3.28, 3.75, 2.96, 3.39, 3.31, 3.15, 2.81, 1.41, 2.76, 3.19, 1.59, 2.17, 3.51, 1.84, 1.61, 1.57, 1.89, 2.74, 3.27, 2.41, 3.09, 2.43, 2.53, 2.81, 3.31, 2.35, 2.77, 3.68, 4.91, 1.57, 2., 1.17, 2.17, 0.39, 2.79, 1.08, 2.88, 2.73, 2.87, 3.19, 1.87, 2.95, 2.67, 4.2, 2.85, 2.55, 2.17, 2.97, 3.68, 0.81, 1.22, 5.08, 1.69, 3.68, 4.7, 2.03, 2.82, 2.5, 1.47, 3.22, 3.15, 2.97, 2.93, 3.33, 2.56, 2.59, 2.83, 1.36, 1.84, 5.56, 1.12, 2.48, 1.25, 2.48, 2.03, 1.61, 2.05, 3.6, 3.11, 1.69, 4.9, 3.39, 3.22, 2.55, 3.56, 2.38, 1.92, 0.98, 1.59, 1.73, 1.71, 1.18, 4.38, 0.85, 1.8, 2.12, 3.65.

The second data set is given by [5] on the fatigue life of 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set consists of 101 observations with maximum stress per cycle 31,000 psi. The data are:

Data Set 2

70, 90, 96, 97, 99, 100, 103, 104, 104, 105, 107, 108, 108, 108, 109, 109, 112, 112, 113, 114, 114, 114, 116, 119, 120, 120, 120, 121, 121, 123, 124, 124, 124, 124, 124, 128, 128, 129, 129, 130, 130, 130, 131, 131, 131, 131, 131, 132, 132, 132, 133, 134, 134, 134, 134, 136, 136, 137, 138, 138, 138, 139, 139, 141, 141, 142, 142, 142, 142, 142, 142, 144, 144, 145, 146, 148, 148, 149, 151, 151, 152, 155, 156, 157, 157, 157, 157, 158, 159, 162, 163, 163, 164, 166, 166, 168, 170, 174, 201, 212.

We have fitted the GPD(α, β, λ) to the data sets with the help of the MATHEMATICA software and the numerical results obtained are given in Tables 1 and 2.

Table 1. Parameter estimates, Kolmogorov-Smirnov (K-S) statistics and P-values obtained from the fit of each of the five distributions for the data sets

Data set	Distribution	Estimates	K-S	P-value
(n= 100)	GPD(α,β,λ)	$(3.3892, 0.2757, 38.8351)$	0.0677	0.7481
	EWD(α,γ,λ)	$(3.9938, 0.5946, 1.4981)$	0.0823	0.5068
	GEGD(α,β,p)	$(8.4709, 0.9742, 0.0761)$	0.1058	0.2124
	GEPD(α,β,λ)	$(9.8798, 0.9917, 0.3415)$	0.1045	0.2247
	ELD(α,λ,θ)	$(0.1360, 9.2265, 9.0242)$	0.1177	0.1251
(n= 101)	GPD(α,β,λ)	$(20.5384, 0.1122, 7.1697)$	0.0594	0.8712
	EWD(α,γ,λ)	$(5.0803, 2.7754, 0.0098)$	0.1004	0.2654
	GEGD(α,β,p)	$(82.9031, 0.0367, 0.0187)$	0.1276	0.0770
	GEPD(α,β,λ)	$(3.2640, 0.0181, 0.0146)$	0.5201	0.0001
	ELD(α,λ,θ)	$(73.2543, 0.4366, 67.5443)$	0.6464	0.0001

Table 2. Information measures such as AIC, BIC, AICc, HQIC and log likelihood values obtained from each of the five distributions for the data sets

Data set	Distribution	AIC	BIC	AICc	HQIC	ℓ
(n= 100)	GPD(α,β,λ)	288.656	296.471	288.906	291.819	-141.328
	EWD(α,γ,λ)	293.956	301.771	294.206	297.119	-143.978
	GEGD(α,β,p)	299.661	307.477	299.911	302.824	-146.830
	GEPD(α,β,λ)	300.846	308.661	301.096	304.009	-147.423
	ELD(α,λ,θ)	305.708	313.524	305.958	308.871	-149.854
(n= 101)	GPD(α,β,λ)	909.440	917.255	909.690	912.603	-451.720
	EWD(α,γ,λ)	913.644	921.459	913.894	916.807	-453.822
	GEGD(α,β,p)	934.489	942.305	934.739	937.652	-464.244
	GEPD(α,β,λ)	1103.399	1111.215	1103.649	1106.562	-548.699
	ELD(α,λ,θ)	1351.982	1359.798	1352.232	1355.145	-672.991

Now, we have computed the variance-covariance matrix of $\hat{\xi} = (\hat{\alpha},\ \hat{\beta},\ \hat{\lambda})$ as $H_1^{-1}(\hat{\xi})$ which is the inverse of expected information matrix based on the Data Set 1.

$$H_1^{-1}(\hat{\xi}) = \begin{pmatrix} 0.1154 & 0.0180 & -0.0534 \\ 0.0180 & 0.0029 & -0.0150 \\ -0.0534 & -0.0150 & 2.5823 \end{pmatrix}.$$

Next, we obtained the Standard error, t-value and P-value for the parameters α, β and λ of the model GPD corresponding to the Data Set 1 are provided in Table 3.

Table 3. Parameter estimates, Standard error, t-value and P-value for the model GPD corresponding to the Data Set 1

Parameters	Estimates	Standard error	t-value	P-value
α	3.3892	0.3397	9.977	0.001
β	0.2757	0.0543	5.071	0.001
λ	38.8351	1.6069	24.167	0.001

Thus 95% confidence interval for the parameters α, β and λ are obtained as

(2.7233, 4.0550), (0.1692, 0.3821) and (35.6856, 41.9846) respectively.
Further, we have calculated the variance-covariance matrix of $\hat{\xi} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ for the Data Set 2 by inverting $H_2(\hat{\xi})$ as given below.

$$H_2^{-1}(\hat{\xi}) = \begin{pmatrix} 0.831244 & 0.005640 & -0.144705 \\ 0.005640 & 0.000045 & -0.002455 \\ -0.144705 & -0.0024552 & 0.656329 \end{pmatrix}.$$

Now, we computed the Standard error, t-value and P-value for the parameters α, β and λ of the model GPD corresponding to the Data Set 2 are presented in Table 4.

Table 4. Parameter estimates, Standard error, t-value and P-value for the model GPD corresponding to the Data Set 2

Parameters	Estimates	Standard error	t-value	P-value
α	20.5384	0.9117	22.53	0.001
β	0.1122	0.0067	16.72	0.001
λ	7.1697	0.8101	8.85	0.001

Similarly, we obtained 95% confidence interval for the parameters α, β and λ are (18.7515, 22.3253), (0.0990, 0.1253) and (5.5819, 8.7575) respectively.
We have ploted both empirical as well as theoritical pdfs and cdfs corresponding to the models GPD, EWD, GEGD, GEPD and ELD to both the data sets considered in the paper and presented in Figures 5 and 6. From Table 1 and 2, it can be observed that the GPD yields better fit to both the data sets. The plots in the Figures 5 and 6 also supports the argument of the suitability of GPD to the data sets.

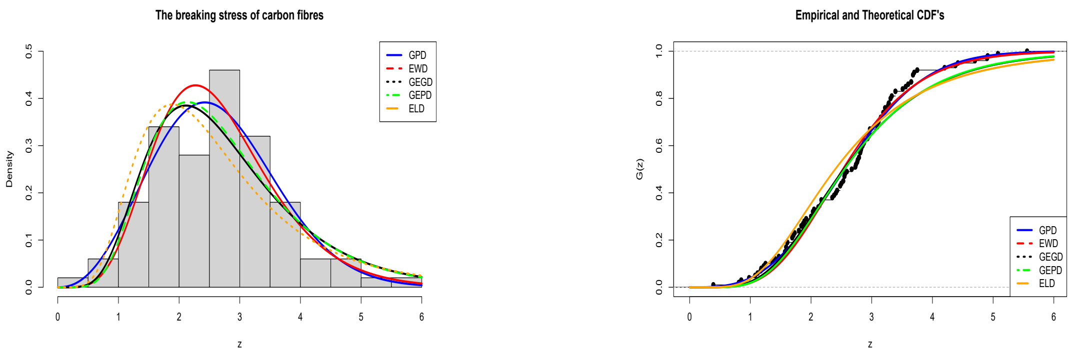


Figure 5. Fitted densities and empirical and theoretical CDFs for the Data Set 1.

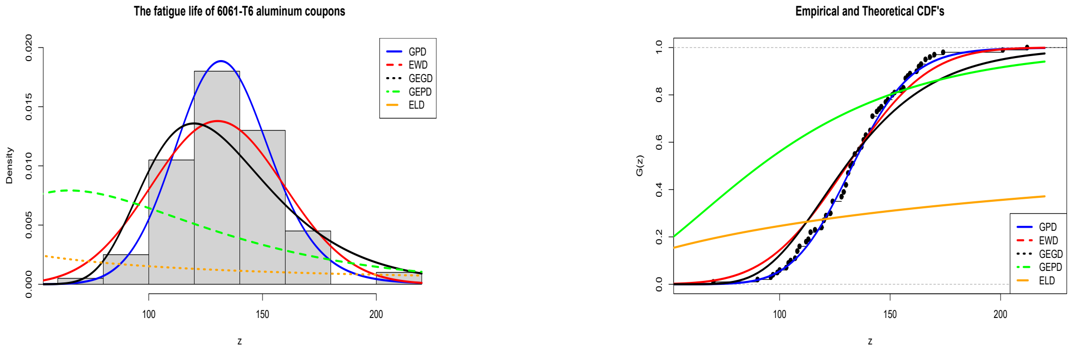


Figure 6. Fitted densities and empirical and theoretical CDFs for the Data Set 2.

Now by adopting the procedure discussed in the Sub-section 3.4, Now we test the null hypothesis H_0 : EPD versus the alternate hypothesis H_1 : GPD or equivalently $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$. Similarly, we test the null hypothesis H_0 : GD versus the alternate hypothesis H_1 : GPD or equivalently $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$. The numerical results obtained are given in Tables 5 and 6 respectively.

Table 5. Testing the significance of the parameter α of the GPD (α, β, λ) using GLRT

Data Set	L($\hat{\xi^*} \mid y$) or ($H_0 : \alpha = 1$)	L($\hat{\xi} \mid y$) or ($H_1 : \alpha \neq 1$)	LRT	P-value
1	5.2146×10^{-86}	3.3119×10^{-62}	109.616	0.0001
2	2.63927×10^{-39}	8.2049×10^{-9}	140.424	0.0001

Table 6. Testing the significance of the parameter λ of the GPD (α, β, λ) using GLRT

Data Set	L($\hat{\xi^*} \mid y$) or ($H_0 : \lambda = 0$)	L($\hat{\xi} \mid y$) or ($H_1 : \lambda \neq 0$)	LRT	P-value
1	6.2291×10^{-64}	3.3119×10^{-62}	7.946	0.0048
2	4.2949×10^{-25}	8.2049×10^{-9}	74.977	0.0001

In order to identify the shape of the hazard rate function of the data sets, we consider a graphical method based on the Total Time on Test (TTT) plot. Thus we included the empirical TTT plot using the following equation.

$$T(r/k) = \frac{\sum_{i=1}^{r} Z_{(i)} + (k-r)Z_{(r)}}{\sum_{i=1}^{k} Z_{(i)}}, \qquad r = 1, 2, \ldots, k,$$

in which $Z_{(i)}$ and $Z_{(r)}$ denote the i^{th} and r^{th} order statistic of the sample. If the empirical TTT transform is convex, concave, convex then concave and concave then convex, the shape of the corresponding hazard rate function is decreasing, increasing, bathtub-shaped and upside-down bathtub shape respectively. For details in this regard see [1]. The TTT plots corresponding to the Data Sets 1 and 2 are presented in Figure 7.

From the Figure 7, it can be seen that Data Sets 1 and 2 possess increasing failure rates.

For assessing how closely a data set fits a particular model, one can plot a probability–probability plot (in short P-P plot). A P-P plot depicts the graph of the expected

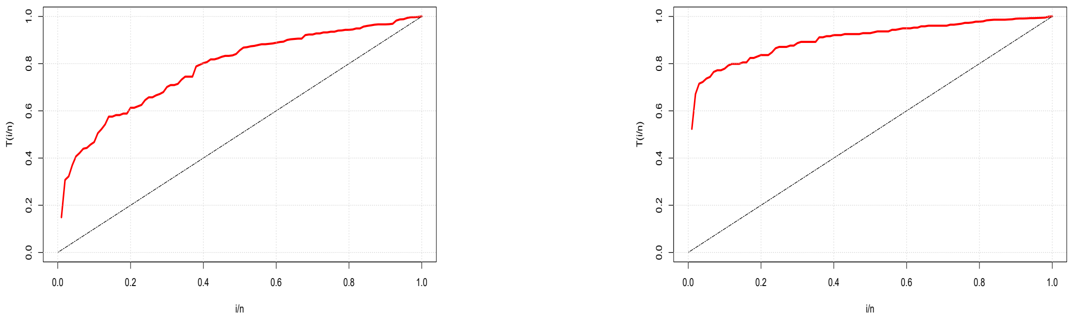
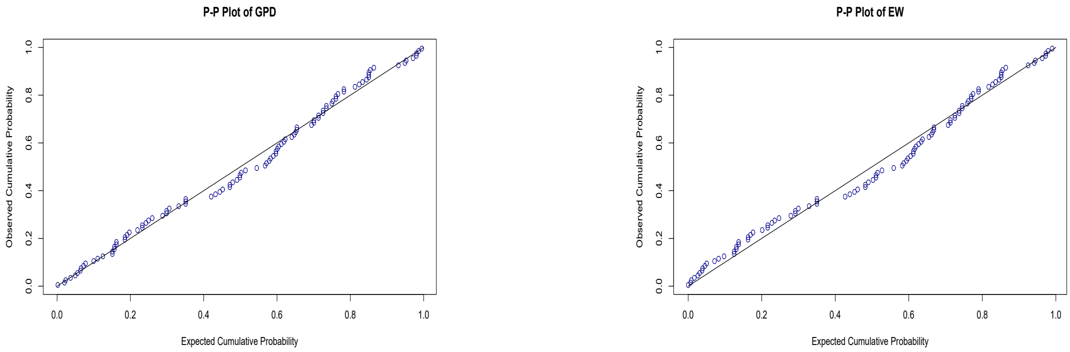


Figure 7. The empirical TTT plot based on the Data Set 1 and 2 respectively.

cumulative probability of a particular model against the observed cumulative probability. Here we obtain P-P plots of the respective models as in Figures 8 and 9. These plots also support the suitability of the proposed model to both the data sets compared to other existing models considered in the paper.



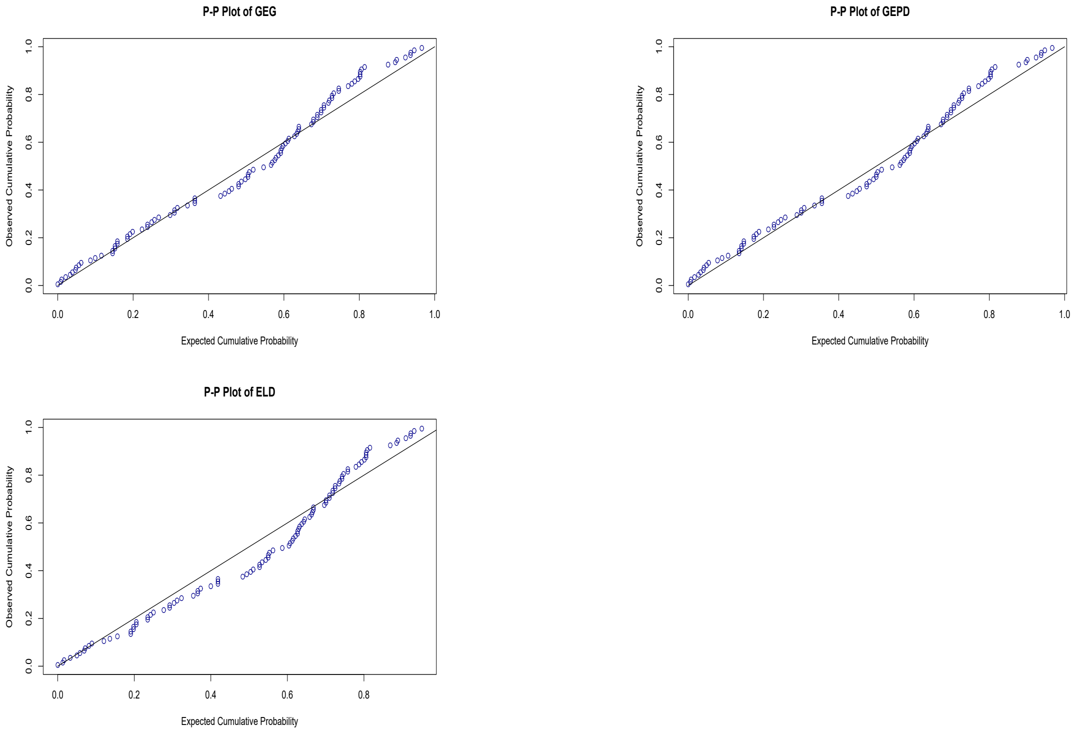


Figure 8. P-P plots of the models for Data Set 1

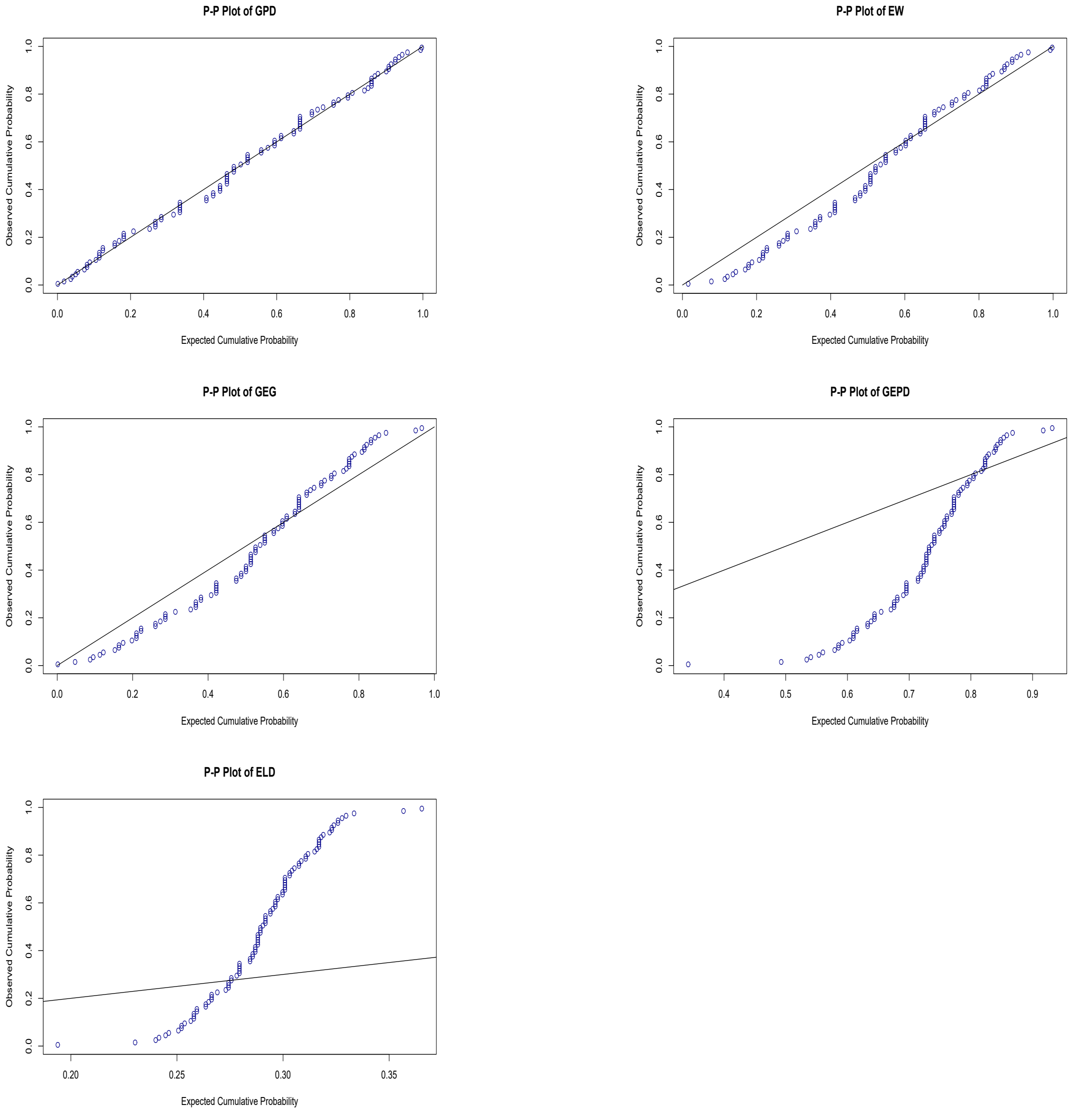


Figure 9. P-P plots of the models for Data Set 2

5. Simulation study

For examining the performance of the maximum likelihood estimators of the parameters α, β and λ, we carry out a brief simulation study. Since we cannot apply inverse transformation method to simulate GPD random samples, we have generated the samples as per the following steps.

(1) Specification of the values of the parameter sets $(\alpha,\ \beta,\ \lambda) = (0.5,\ 4,\ 6)$ in the first case and $(2.3,\ 3,\ 7)$ in the second case.
(2) Specification of the sample size. Here we considered samples of size 50, 100, 500 and 1000.
(3) Generation of pseudo-random sample of GPD utilizing Metropolis–Hastings algorithm for computation of the MLEs of the parameters using EM algorithm.
(4) Generation of 1000 samples.
(5) Computation of Bias and mean squared error (MSE).

The results obtained from the simulation study are presented in Table 7. We have used the R Software ([13]) to find the estimates and sample generation. From Table 7, it is evident that the MSE decreases as sample size increases.

Table 7. The absolute bias and MSEs corresponding to the estimates obtained via EM algorithm for simulated samples.

n	Parameter	$\alpha = 0.5,\ \beta = 4,\ \lambda = 6$		$\alpha = 2.3,\ \beta = 3,\ \lambda = 7$	
		Bias	MSE	Bias	MSE
50	α	0.6253	0.8966	0.2784	0.1132
	β	0.4078	0.1988	0.5399	0.4699
	λ	0.2746	0.1552	0.1143	0.0249
100	α	0.5119	0.4985	0.0607	0.0116
	β	0.3456	0.1228	0.4130	0.2468
	λ	0.1399	0.0678	0.0773	0.0104
500	α	0.1321	0.0199	0.0462	0.0075
	β	0.2246	0.0539	0.2225	0.1053
	λ	0.0260	0.0027	0.0502	0.0042
1000	α	0.0695	0.0055	0.0333	0.0019
	β	0.0786	0.0097	0.0951	0.0180
	λ	0.0064	0.0015	0.0138	0.0005

6. Summary and Discussions

Here we consider a new class of lifetime distribution with decreasing, increasing and upside-down bathtub failure rate as a generalization of the EPD by [9]. We obtained several important statistical properties of the proposed model and discussed various methods of estimation such as method of maximum likelihood and method of EM algorithm for estimating the parameters of the distribution. We have examined the relevance of the proposed model with the help of certain real life data sets and it is shown that the proposed model gives better fit to both the data sets considered in section 4 compared to the existing models such as EWD, GEGD, GEPD and ELD based on the information measures: AIC, BIC, AICc, HQIC as well as K-S statistic, P-values and log likelihood values. The likelihood ratio test procedures are applied for testing the significance of the additional parameters of the model. Simulation studies are also carried out for examining the performance of estimators obtained through the procedure of EM algorithm. Even though several inferential aspects of the proposed model are yet to study, which we hope to present through another publication.

Appendix A. Proof of Result 2

By definition, the c.d.f of the GPD with pdf (4) is

$$\begin{aligned} G(z) &= \int_0^z \mathrm{g(t)}\,\mathrm{d}t \\ &= \int_0^z \frac{\lambda\beta^{\alpha}\ \mathrm{t}^{\alpha-1}\mathrm{e}^{-\lambda-\beta\mathrm{t}+\frac{\lambda\Gamma(\alpha,\beta\mathrm{t})}{\Gamma(\alpha)}}}{\Gamma(\alpha)(1-\mathrm{e}^{-\lambda})}\,\mathrm{d}t. \end{aligned} \tag{A1}$$

On simplification and puting u= $\Gamma(\alpha, \beta t)$ in (A1) to obtain following.

$$G(z) = \frac{e^{-\lambda}}{1-e^{-\lambda}}\Big(e^{\lambda} - e^{\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}\Big), \tag{A2}$$

which gives

$$G(z) = \frac{1-e^{-\lambda+\frac{\lambda\Gamma(\alpha,\beta z)}{\Gamma(\alpha)}}}{1-e^{-\lambda}}.$$

Appendix B. Elements of the Information Matrix

The elements of the information matrix $H(\xi)$ are as given below.

$$\begin{aligned} I_{11} = \frac{\partial^2\ell(\xi)}{\partial\alpha^2} &= \frac{1}{\Gamma(\alpha)}\Big(2\ T(4,\alpha,\beta z_i) + 2\ \log(\beta z_i)\ T(3,\alpha,\beta z_i) \\ &+ \log^2(\beta z_i)\ \Gamma(\alpha,\beta z_i) - 2\ \psi^{(0)}(\alpha)\ \big(\log(\beta z_i)\ \Gamma(\alpha,\beta z_i) + T(3,\alpha,\beta z_i)\big) \\ &+ (\psi^{(0)}(\alpha)^2 - \psi^{(1)}(\alpha))\Gamma(\alpha,\beta z_i)\Big) - k\ \psi^{(1)}(\alpha), \end{aligned} \tag{B1}$$

$$\begin{aligned} I_{12} = \frac{\partial^2\ell(\xi)}{\partial\alpha\partial\beta} &= \frac{k}{\beta} + \frac{\lambda}{\Gamma(\alpha)}\sum_{i=1}^{k}\Bigg\{\frac{\Gamma(\alpha,\beta z_i)}{\beta} - \ln(\beta z_i)\ e^{-\beta z_i}(\beta z_i)^{\alpha-1} \\ &- \pi z_i\Big(\frac{Cosec(\pi\alpha)}{\Gamma(1-\alpha)} - \frac{\alpha Cosec(\pi\alpha)\gamma(\alpha,\beta z_i)}{\Gamma(1-\alpha)\Gamma(\alpha+1)}\Big) + z_i\ T(3,\alpha,\beta z_i)\Bigg\} \\ &+ \psi(\alpha)\ z_i(\beta z_i)^{\alpha-1} \end{aligned} \tag{B2}$$

$$I_{13} = \frac{\partial^2\ell(\xi)}{\partial\alpha\partial\lambda} = \frac{1}{\Gamma(\alpha)}\sum_{i=1}^{k}\big[\ln\ \beta z_i\ \Gamma(\alpha,\beta z_i) + \beta z_i\ T(3,\alpha,\beta z_i) - \psi^{(0)}(\alpha)\Gamma(\alpha,\beta z_i)\big], \tag{B3}$$

$$I_{22} = \frac{\partial^2 \ell(\xi)}{\partial \beta^2} = \frac{-k\ \alpha}{\beta^2} - \frac{\lambda}{\Gamma(\alpha)} \sum_{i=i}^{k} z_i^{\alpha-1}\big((\alpha-1)\beta^{\alpha-2}e^{-\beta z_i} - z_i \beta^{\alpha-1} e^{-\beta z_i}\big), \qquad \text{(B4)}$$

$$I_{23} = \frac{\partial^2 \ell(\xi)}{\partial \beta \partial \lambda} = \frac{1}{\Gamma(\alpha)} \sum_{i=1}^{k} (\beta z_i)^{\alpha-1} e^{-\beta z_i} \qquad \text{(B5)}$$

and

$$I_{33} = \frac{\partial^2 \ell(\xi)}{\partial \lambda^2} = \frac{-k}{\lambda^2} - \frac{e^{k\ \lambda}}{(e^{\lambda}-1)^2} \qquad \text{(B6)}$$

Acknowledgments

Authors are highly thankful to the anonymous Referees for their valuable comments that greatly improved the quality and presentation of the manuscript. The first author is particularly thankful to Department of Science and Technology, New Delhi (Governmet of India) for financial support (MTR/2017/000942). The second author is CSIR-SRF thankful to the Council of Scientific and Industrial Research (CSIR), New Delhi for financial assistance (File No: 09/102(0263)/2019-EMR-I).

References

[1] Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36(1):106–108.

[2] Abdul-Moniem, I. B. and Abdel-Hameed, H. F. (2012). On exponentiated Lomax distribution. *International Journal of Mathematical Archive*, 3(5):2144–2150.

[3] Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters*, 39(1):35–42.

[4] Barreto-Souza, W. and Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. *Statistics & Probability Letters*, 79(24):2493–2500.

[5] Birnbaum, Z. W. and Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6(2):328–347.

[6] Chahkandi, M. and Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics & Data Analysis*, 53(12):4433–4440.

[7] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

[8] Kumar, C. S. and Satheenthar, A. S. (2022). An extended version of exponential Poisson distribution. In *In Proceedings of International Conference on Advances in Mathematicaland Statistical Sciences*, page 10–19. Cape Comorin.

[9] Kuş, C. (2007). A new lifetime distribution. *Computational Statistics & Data Analysis*, 51(9):4497–4509.

[10] Louzada, F., Marchi, V., and Roman, M. (2014). The exponentiated exponential–geometric distribution: a distribution with decreasing, increasing and unimodal failure rate. *Statistics*, 48(1):167–181.

[11] Nichols, M. D. and Padgett, W. (2006). A bootstrap control chart for Weibull percentiles. *Quality and Reliability Engineering International*, 22(2):141–151.

[12] Pal, M., Ali, M. M., and Woo, J. (2006). Exponentiated Weibull distribution. *Statistica*, 66(2):139–147.

[13] R Core Team (2019). R: A language and statistical, statistical computing. *R Foundation forComputing*, Vienna, Austria. URL https://www. R-project. org.

[14] Tahmasbi, R. and Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics & Data Analysis*, 52(8):3889–3901.