One generalized mixture Pareto distribution and estimation of the parameters by the EM algorithm for complete and right-censored data

Mohamed Kayid
Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia.
e-mail: drkayid@ksu.edu.sa

ABSTRACT A new mixture generalized Pareto distribution is introduced. Then, some of its attributes are explored. The maximum likelihood method and expectation maximization (EM) algorithm have been applied to estimate the parameters for complete and right-censored data. In a simulation study, the bias, absolute bias and mean squared error of the maximum likelihood estimator are compared with those related to the EM estimator. The results show that the absolute bias and mean squared error of the EM estimator are smaller than the related values for the maximum likelihood estimator. Finally, to illustrate its usefulness, the model has been applied to describe real data sets.

INDEX TERMS Generalized Pareto distribution; Mixture model; Maximum likelihood estimator; EM algorithm.

I. INTRODUCTION

THE Pareto distribution is a power-law distribution that was originally used to describe wealth distribution. Additionally, it may be useful in describing observations from scientific, geophysical, actuarial, social, and quality control events and many other fields. The Pareto model may be applied to situations, where an equilibrium is found in the distribution of "small" to "large" values. In many cases, we may point to: the sizes of files transferred on the internet network by TCP/IP, which consists of many smaller files and few larger ones; the hard disk drive error rates, which consist of many small error rates and few large ones; the sizes of human settlements, which consist of many small values related to hamlets/villages and few large values related to cities; the oil reserves volumes in oil fields, which consist of many small fields and few large fields and many other examples. Among many studies in this field, Burroughs and Tebbens [1] analyzed observations from earthquakes and forest fire areas, and Schroeder et al. [2] fitted the data from disk-driven sector errors to this distribution.

To provide more flexible distributions, various generalizations of the Pareto distribution have been proposed. Among them, Akinsete et al. [3] studied the beta Pareto distribution, Nassar and Nada [4] and Mahmoudi [5] considered the beta generalized Pareto, Zea et al. [6] presented the beta exponentiated Pareto distribution, Alzaatreh et al. [7] introduced the gamma Pareto distribution, Elbatal [8] investigated the Kumaraswamy exponentiated Pareto distribution, Bourguignon et al. [9] introduced the Kumaraswamy Pareto distribution, Mead [10] defined one generalized beta exponentiated Pareto distribution, Clifton et al. [11] applied one extended version of the Pareto distribution for novelty detection, and Papastathopoulos and Tawn [12] applied the extended generalized Pareto models to the tail estimation problem. Recently, Jayakumar et al. [13], Tahir et al. [14], Korkmaz et al. [15], Elbatal and Aryal [16], and Chananet and Phaphan [17] proposed new distributions based on the Pareto model.

One generalized Pareto distribution, i.e., the \( GP(\alpha, \beta, \theta) \) distribution, can be defined by the probability density function (pdf)

\[
f(x) = \frac{\Gamma(\alpha + \theta)}{\Gamma(\alpha) \Gamma(\theta)} \frac{1}{\beta} \left(1 + \frac{x}{\beta}\right)^{\theta-1} \left(1 + \frac{x}{\beta}\right)^{\alpha - \theta}, \quad x > 0,
\]

where \( \alpha > 0, \beta > 0, \theta > 0 \), see Beirlant et al. [18] and Wiborg [19]. When \( \alpha \to \infty \) and \( \lambda = \frac{\theta \beta}{\alpha} \) are fixed constants,
this model tends to the gamma distribution with pdf
\[
f(x) = \frac{1}{\Gamma(\theta)} \frac{\theta^\alpha}{\lambda^\alpha} x^{\alpha-1} e^{-\frac{x}{\lambda}}, \quad x > 0.
\] (2)

In this paper, we propose a new model based on a mixture of the GP distribution (1). The proposed model extends mixtures of the exponential, gamma, or Pareto distributions and has sufficient flexibility to describe many real situations. The paper is organized as follows. In Section 2, the new model is defined, and some of its statistical and reliability attributes are studied. In Section 3, the parameters of the model are estimated for the complete and right-censored data by the maximum likelihood estimator (MLE) and EM algorithm. The behavior of the estimators of the parameters has been investigated by a simulation study in Section 4. In Section 5, the proposed model is fitted to real data sets to show its applicability.

II. THE NEW MODEL

We propose a generalized Pareto mixture \(GPM(\alpha, \beta, \gamma)\) by the pdf
\[
f(x) = \frac{\alpha}{\beta (\gamma + 1)} (1 + \frac{x}{\beta})^{-\alpha-1} \gamma + (\alpha + 1) \frac{x}{\beta + x},
\] (3)

which is a mixture of \(GPM(\alpha, \beta)\) and \(GPM(\alpha, \beta, 2)\) with weights \(\gamma/(\gamma + 1)\) and \(1/(\gamma + 1)\), respectively.

Since the GP model tends to the gamma distribution, a special limiting case of the GPM is the quasi Lindley distribution studied by Shanker [20].

For a random lifetime \(X\), the most important function in reliability engineering and survival analysis is the reliability function \(R(x) = P(X \geq x)\) and gives the probability that an object works (survives) at least until a specified time \(x\). The reliability function of GPM is
\[
R(x) = \frac{1}{\gamma + 1} (1 + \frac{x}{\beta})^{-\alpha} \gamma + (\alpha + 1) \frac{1}{\beta + x},
\] (4)

Proposition 1. When \(\alpha > k\), the \(k^{th}\) moments of \(GPM(\alpha, \beta, \gamma)\) are finite and equal to
\[
E(X^k) = \frac{1}{\gamma + 1} \frac{\alpha + 1}{\alpha - k} \sum_{j=0}^{k} (-1)^{k-j} \left( \begin{array}{c} k \\ j \end{array} \right) \gamma - (\gamma + 1) + \frac{1}{\alpha - j} \left( \begin{array}{c} k + 1 \\ j \end{array} \right) \alpha + 1 - \frac{\alpha \beta}{x + \beta}. \] (5)

For \(\alpha \leq k\), it is infinite.

Proof. Because \(GPM(\alpha, \beta, \gamma)\) is a mixture of \(X_1 \sim GPM(\alpha, \beta, 1)\) and \(X_2 \sim GPM(\alpha, \beta, 2)\), we have
\[
E(X^k) = \frac{\gamma}{\gamma + 1} E(X_1^k) + \frac{1}{\gamma + 1} E(X_2^k).
\] (6)

But
\[
E(X_1^k) = \int_0^{\infty} \frac{\alpha}{\beta (1 + \frac{x}{\beta})^{\alpha+1}} dx = \alpha \beta \int_1^{\infty} \frac{(y - 1)^k}{y^{\alpha+1}} dy
\]
\[
= \alpha \beta \sum_{j=0}^{k} \left( \begin{array}{c} k \\ j \end{array} \right) (-1)^{k-j} \frac{1}{\alpha - j} \frac{1}{y^{\alpha+1}} dy.
\] (7)

It is easy to check that for \(\alpha \leq k\), (7) is infinite, but for \(\alpha > k\), it can be simplified as follows:
\[
E(X_1^k) = \alpha \beta \sum_{j=0}^{k} \left( \begin{array}{c} k \\ j \end{array} \right) (-1)^{k-j} \frac{1}{\alpha - j}.
\] (8)

Using a similar approach, we can check that for \(\alpha \leq k\), \(E(X_2^k)\) is infinite, and for \(\alpha > k\), we have
\[
E(X_2^k) = \alpha (\alpha + 1) \beta \sum_{j=0}^{k} \left( \begin{array}{c} k + 1 \\ j \end{array} \right) (-1)^{k+1-j} \frac{1}{\alpha - j + 1}.
\] (9)

Then, the result follows by (7), (8) and (9). \(\Box\)

The quantile function \(q(p) = F^{-1}(p)\) for \(GPM(\alpha, \beta, \gamma)\) has no closed form and can be numerically computed by solving the following equation:
\[
\frac{(1 - p)(\gamma + 1)}{\beta^\alpha} = \frac{1}{(\gamma + 1) + \frac{\alpha \beta}{x + \beta} + q(p)}.
\] (10)

When the corresponding moments exit, the Pearson’s moment coefficient of skewness of a random variable \(X\) is defined to be
\[
B = E\left(\frac{X - \mu}{\sigma}\right)^3 = E(X^3) - 3\mu \sigma^2 - \mu^3\sigma^3
\] (11)

and the kurtosis of \(X\) is
\[
K = E\left(\frac{X - \mu}{\sigma}\right)^4 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}.
\] (12)

where \(\mu = E(X)\), and \(\sigma^2\) is the variance of \(X\).

Moreover, the skewness and kurtosis of a distribution can be described in terms of the quantile function. MacGillivray [21] suggested the following relation for skewness:
\[
B = \frac{q(1 - u) + q(u) - 2q(0.5)}{q(1 - u) - q(u)},
\]
where \(u \in (0, 0.5)\). When \(u = 0.25\), it reduces to Bowley’s measure of skewness, Bowley [22]. Moreover, Moor [23] introduced the kurtosis in terms of the quantile function by
\[
K = \frac{q(\delta) - q(\delta/2) + q(\delta/2) - q(\delta)}{q(\delta) - q(\delta/2)}.
\]

In the economics literature, for a cumulative distribution function (cdf) \(F\), the Lorenz curve is:
\[
L(p) = \frac{p y_p}{\mu} - \frac{q(p)}{0} \int_0^{q(p)} x dF(x) - \int_0^{q(0.5)} x dF(x).
\]

VOLUME 4, 2016
and provides a graphical representation of wealth inequality. In fact, $L(p)$ shows the proportion of overall income or wealth of the $100 \times p$ percent of people with lower income or wealth. This plot will be a convex plot joining two points $(0, 0)$ and $(1, 1)$. For an ideal society where every person has identical income, the Lorenz curve is a straight line that joins these points. For $GPM(\alpha, \beta, \gamma)$, $\mu$ is not finite for $\alpha \leq 1$, but $\mu_p$ is finite, so $L(p)$ is zero and does not describe the wealth distribution. Fortunately, Prendergast and Staudte [24] have proposed alternatives in terms of the quantile function, which replace $\mu$ by the median of the distribution and $\mu_p$ by $q(\frac{p}{2})$.

Precisely

$$L_1(p) = p \frac{q(\frac{p}{2})}{q(0, \beta)},$$

$$L_2(p) = p \frac{q(\frac{p}{2})}{q(1 - \frac{p}{2})},$$

and

$$L_3(p) = 2p \frac{q(\frac{p}{2})}{q(\frac{p}{2}) + q(1 - \frac{p}{2})}.$$

A. RELIABILITY MEASURES

The failure rate (FR), mean residual life (MRL) and $p$-quantile residual life ($p$-QRL) concepts play important roles in describing the dynamic attributes of lifetime variables and have been studied from different perspectives in the reliability and survival analysis literature. The FR function of the $GPM(\alpha, \beta, \gamma)$ is

$$h(x) = \frac{f(x)}{R(x)} = \frac{\alpha x + \alpha + 1}{(\gamma + 1)(\beta + x)^2 - \alpha\beta(\beta + x)}.$$

For a distribution with reliability function $R$, the MRL function is:

$$m(x) = \frac{1}{R(x)} \int_x^\infty R(t)dt.$$

The following proposition gives the form of the MRL function for the proposed model.

Proposition 2. When $\alpha > 1$, the MRL function of $GPM(\alpha, \beta, \gamma)$ is finite and of the form

$$m(x) = \frac{(\beta + x)(\gamma + 1 - \alpha\beta + \beta + 1 - \alpha\beta(\alpha - 1))}{\alpha(\beta + x)(\gamma + 1 - \alpha\beta + 1 - \alpha\beta(\alpha - 1))}.$$

When $\alpha \leq 1$, the MRL is infinite.

Proof: It is straightforward to check that

$$\int_x^\infty R(t)dt = \left(\frac{\gamma}{\gamma + 1} + \frac{\alpha + 1}{\gamma + 1}\right) \int_x^\infty (1 + \frac{t}{\beta})^{-\alpha} dt - \frac{\alpha}{\gamma + 1} \int_x^\infty (1 + \frac{t}{\beta})^{-\alpha - 1} dt. \quad (13)$$

The first integral of the right side of (13) is equal to $\frac{\beta}{(\alpha - 1)(1 + \frac{2}{\beta})}^{-\alpha + 1}$ when $\alpha > 1$ and infinite when $\alpha \leq 1$. The second integral is equal to $\frac{\beta}{\alpha}(1 + \frac{2}{\beta})^{-\alpha}$. Then, the result follows by some algebra. □

The $p$-QRL function of a distribution with reliability function $R$ is defined to be

$$q_p(x) = R^{-1}(\hat{p}R(x)) - x = q(1 - \hat{p}R(x)) - x, \quad x > 0,$$

where $\hat{p} = 1 - p$, and $q(.)$ is the quantile function and defined by (10). Similar to the quantile function, the $p$-QRL function of GPM has no closed form and should be numerically computed.

III. ESTIMATION OF THE PARAMETERS

A. THE MLE FOR COMPLETE DATA

Suppose that $x_1, x_2, ..., x_n$ represent realizations from $GPM(\alpha, \beta, \gamma)$. The log-likelihood function of the parameters is

$$l(\alpha, \beta, \gamma; x) = n \ln \alpha - n \ln \beta - n \ln(\gamma + 1)$$

$$- (\alpha + 1) \sum_{i=1}^n \ln(1 + \frac{x_i}{\beta}) + \sum_{i=1}^n \ln[\gamma + (\alpha + 1)\frac{x_i}{x_i + \beta}],$$

and the log-likelihood equations are

$$\frac{\partial}{\partial \alpha} l(\alpha, \beta, \gamma; x) = -n + \sum_{i=1}^n \frac{x_i}{\beta} = 0,$$

$$\frac{\partial}{\partial \beta} l(\alpha, \beta, \gamma; x) = \frac{n}{\beta} + (\alpha + 1) \sum_{i=1}^n \frac{x_i}{\beta^2 + \beta x_i} - (\alpha + 1) \sum_{i=1}^n \frac{\gamma(x_i + \beta)^2 + (\alpha + 1)x_i(x_i + \beta)}{\gamma(x_i + \beta) + (\alpha + 1)x_i(x_i + \beta)} = 0,$$

and

$$\frac{\partial}{\partial \gamma} l(\alpha, \beta, \gamma; x) = -n \frac{1}{\gamma + 1} + \sum_{i=1}^n \frac{x_i + \beta}{(\alpha + 1)x_i + \gamma \beta} = 0.$$

We calculate the MLE by directly maximizing the log-likelihood function (15) or solving the likelihood equations.

Let $l = \ln f(X)$. Then, the Fisher information matrix can be obtained to apply the variance of the MLE. Here, it is defined as follows:

$$K = \begin{bmatrix}
E(-\frac{\partial^2 l}{\partial \alpha^2}) & E(-\frac{\partial^2 l}{\partial \alpha \partial \beta}) & E(-\frac{\partial^2 l}{\partial \alpha \partial \gamma}) \\
E(-\frac{\partial^2 l}{\partial \beta \partial \alpha}) & E(-\frac{\partial^2 l}{\partial \beta^2}) & E(-\frac{\partial^2 l}{\partial \beta \partial \gamma}) \\
E(-\frac{\partial^2 l}{\partial \gamma \partial \alpha}) & E(-\frac{\partial^2 l}{\partial \gamma \partial \beta}) & E(-\frac{\partial^2 l}{\partial \gamma^2})
\end{bmatrix}. \quad (16)$$

The elements of this matrix are complicated expressions and should be numerically computed. For the iid random sample $X_i$, $i = 1, 2, ..., n$ from $GPM(\alpha_0, \beta_0, \gamma_0)$, the MLE $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ weakly converges to the multivariate normal $N((\alpha_0, \beta_0, \gamma_0), n^{-1}K^{-1})$, where $K^{-1}$ is the inverse of the information matrix.
FIGURE 1. The pdf (left) and FR function (right) of the GPM for some values of the parameters.

FIGURE 2. The MRL (left) and median residual life (right) of the GPM for some values of the parameters.

B. EM ALGORITHM FOR COMPLETE DATA

Assume that \( X_i, i = 1, 2, ..., n \) shows an iid random sample of \( GPM(\alpha, \beta, \gamma) \). Every \( X_i \) arises from the GPM distribution. In the EM algorithm, one indicator latent random variable \( Z_i \) is considered, which determines that \( X_i \) has been taken from \( GP(\alpha, \beta, 1) \) or \( GP(\alpha, \beta, 2) \). Precisely, \( X_i | Z_i = 1 \sim GP(\alpha, \beta, 1) \), \( X_i | Z_i = 2 \sim GP(\alpha, \beta, 2) \) and

\[
P(Z_i = j) = \frac{1}{\gamma+1} \gamma^{j=1}, \quad j = 1, 2. \tag{17}
\]

Let \( \Theta = (\alpha, \beta, \gamma) \) for a brief representation. The likelihood function can be written in the following form.

\[
L(\Theta; x, z) = \prod_{i=1}^{n} \prod_{j=1}^{2} f_j(x_i; \alpha, \beta) P(Z_i = j) I(z_i=j), \tag{18}
\]

where indicator \( I(z_i = j) \) is equal to 1 when \( z_i = j \) and otherwise equal to 0, and \( f_j \) shows the pdf of \( GP(\alpha, \beta, j) \).
In this step, the parameters at iteration \( t \) are estimated in the current iteration of the current estimate of the conditional latent variable. Then in step \( M \), the parameters are estimated in the current iteration by maximizing the expectation value prepared in step \( E \).

The iterative process continues until the iteration does not noticeably improve the expectation.

The \( E \) step

Suppose that the estimate of the parameters \( \Theta_t \) is known at iteration \( t \). Then, the conditional distribution of \( Z_i \) can be computed by the well-known Bayes theorem as follows. It is simplified in Appendix A, (33). After the simplification, we have

\[
p_{1,t} = \frac{\gamma_t(x_t + x_i)}{\gamma_t(x_t + x_i) + (\alpha_t + 1)x_i}, \quad i = 1,2, \ldots, n,\]

and \( p_{2,t} = 1 - p_{1,t} \). These probabilities are known as membership probabilities at iteration \( t \) and are used to construct the expectation function

\[
Q(\Theta; \Theta_t) = Q_1(\alpha, \beta; \Theta_t) + Q_2(\gamma; \Theta_t),
\]

where

\[
Q_1(\alpha, \beta; \Theta_t) = n \ln \alpha - \sum_{i=1}^{n} (\alpha + 1) \ln(1 + x_i) \beta \]

\[
+ \sum_{i=1}^{n} p_{2,t} \ln(\alpha + 1) + \ln \left( \frac{x_i}{\beta + x_i} \right),
\]

and

\[
Q_2(\gamma; \Theta_t) = \sum_{i=1}^{n} p_{1,t} \ln \gamma - n \ln(\gamma + 1).
\]

The \( M \) step

In this step, the parameters at iteration \( t + 1 \) are estimated by maximizing \( Q(\Theta; \Theta_t) \) in terms of \( \Theta \). Specifically,

\[
\Theta_{t+1} = \arg \max_\Theta Q(\Theta; \Theta_t),
\]

By (20), this problem can be reduced to two separate maximization problems as follows:

\[
(\alpha_{t+1}, \beta_{t+1}) = \arg \max_{\alpha,\beta} Q_1(\alpha, \beta; \Theta_t),
\]

and

\[
\gamma_{t+1} = \arg \max_{\gamma} Q_2(\gamma; \Theta_t),
\]

The maximization problem (23) has no analytical solution and should be numerically solved. However, by solving the equation \( \frac{\partial}{\partial \gamma} Q_2(\gamma; \Theta_t) = 0 \), we obtain

\[
\gamma_{t+1} = \frac{\sum_{i=1}^{n} p_{1,i,t}}{\sum_{i=1}^{n} p_{2,i,t}}.
\]

The iterative process can be terminated if for a predefined small \( \epsilon > 0 \), \( Q(\Theta_{t+1}; \Theta_t) < Q(\Theta_t; \Theta_t) + \epsilon \).

C. THE MLE FOR RIGHT-CENSORED DATA

Now, suppose that we have a right-censored iid random sample from \( GPM(\alpha, \beta, \gamma) \). The random variable \( X_i \) is said to be censored from right by a censorship random variable \( C_i \), when \( X_i > C_i \), so we only know that the event time is greater than \( C_i \). Thus, the observations are \( T_i = \min(X_i, C_i) \) and \( d_i \), where \( d_i = 1 \); if the event is not censored, \( X_i \leq C_i \), and \( d_i = 0 \); if it is censored, \( X_i > C_i \). Let \( (t_1, d_1), \ldots, (t_i, d_i) \) show a right-censored sample; then, the log-likelihood function is:

\[
l(\alpha, \beta, \gamma; \mathbf{t}, \mathbf{d}) = \sum_{i=1}^{n} d_i \ln f(t_i) + \sum_{i=1}^{n} (1 - d_i) \ln R(t_i),
\]

where \( f \) and \( R \) are the density and reliability functions of the GPM distribution, respectively. The log-likelihood function reduces to

\[
l(\alpha, \beta, \gamma; \mathbf{t}, \mathbf{d}) = \sum_{i=1}^{n} d_i \left[ \ln \alpha - \ln \beta - \ln(\gamma + 1) \right] \]

\[
- \sum_{i=1}^{n} d_i (\alpha + 1) \ln(1 + \frac{t_i}{\beta}) + \ln[\gamma + (\alpha + 1) \frac{t_i}{\beta + t_i}] + \sum_{i=1}^{n} (1 - d_i) \left[ -\ln(1 + \gamma) - \alpha \ln(1 + \frac{t_i}{\beta}) + \ln(\gamma + 1 + \frac{\alpha \beta}{t_i + \beta}) \right].
\]

D. EM ALGORITHM FOR RIGHT-CENSORED DATA

In the presence of latent variable \( Z_i \), which is defined in the previous sections, the likelihood function for the right-censored data is:

\[
L(\Theta; \mathbf{t}, \mathbf{d}, \mathbf{z}) = \prod_{i=1}^{n} \prod_{j=1}^{2} \left[ f_j(t_i; \alpha, \beta) P(Z_i = j) \right]^{I(z_i = j)d_i} \times \prod_{i=1}^{n} \prod_{j=1}^{2} \left[ R_j(t_i; \alpha, \beta) P(Z_i = j) \right]^{I(z_i = j)(1-d_i)},
\]

where \( P(Z_i = j) \) and \( f_j \) are defined by (17) and (19), respectively, and \( R_j \) shows the reliability function of \( f_j \). The log-likelihood function \( \ln L(\Theta; \mathbf{t}, \mathbf{d}, \mathbf{z}) \) has been simplified in Appendix A, (35).
The E Step

Suppose that the estimate of the parameters at iteration \( t \), \( \Theta_t \), is known; then, the conditional distribution of \( Z_i \) is equal to \( p_{ij,t} = d_i P(Z_i = j|X_i = t, \Theta_t) + (1 - d_i) P(Z_i = j|X_i > t, \Theta_t) \) and can be computed by the well-known Bayes theorem (see Appendix A, (36)).

Specifically, taking \( j = 1 \),

\[
p_{11,t} = d_i \frac{\gamma_t (\beta_t + t_i)}{\gamma_t (\beta_t + t_i) + (\alpha_t + 1)t_i} + (1 - d_i) \frac{\gamma_t}{\gamma_t + \alpha_t + 1 - \frac{\alpha \beta}{\beta_t + t_i}},
\]

\[
p_{12,t} = 1 - p_{11,t},
\]

and \( p_{22,t} = 1 - p_{11,t} \). Then, using (35), the expectation function at iteration \( t \) is \( Q(\Theta|\Theta_t) = E_{Z_i,t,d,Z_i,t}(\Theta; t, d, z) \).

In Appendix A, (37) and (38) show that \( Q(\Theta|\Theta_t) \) is a sum of two expressions, one of which solely depends on \( \alpha \) and \( \beta \), and the other depends on \( \gamma \). In other words,

\[
Q(\Theta|\Theta_t) = Q_1(\alpha, \beta|\Theta_t) + Q_2(\gamma|\Theta_t),
\]

where

\[
Q_1(\alpha, \beta|\Theta_t) = \sum_{i=1}^{n} d_i \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{t_i}{\beta}) \right] + p_{22,t} \ln(\alpha + 1) + p_{21,t} \ln(1 + \frac{t_i}{\beta})
\]

\[
+ \sum_{i=1}^{n} (1 - d_i) \left[ -\alpha \ln(1 + \frac{t_i}{\beta}) \right] + p_{22,t} \ln(\alpha + 1 - \frac{\alpha \beta}{t_i + \beta})
\]

\[
Q_2(\gamma|\Theta_t) = -n \ln(\gamma + 1) + \ln(\sum_{i=1}^{n} p_{i1,t}),
\]

The M step

Similar to the uncensored case, the estimation of the parameters at iteration \( t + 1 \) will be computed by maximizing \( Q(\Theta|\Theta_t) \) in terms of \( \Theta \). By (27), \( Q(\Theta|\Theta_t) \) is a sum of two separate statements, so it follows that

\[
(\alpha_{t+1}, \beta_{t+1}) = \arg \max_{\alpha, \beta} Q_1(\alpha, \beta|\Theta_t),
\]

and

\[
\gamma_{t+1} = \arg \max_{\gamma} Q_2(\gamma|\Theta_t).
\]

The maximization problem (30) cannot be analytically solved and should be numerically solved. However, it is easy to check that the solution of (31) is

\[
\gamma_{t+1} = \frac{\sum_{i=1}^{n} p_{i1,t}}{\sum_{i=1}^{n} p_{i2,t}}.
\]

IV. SIMULATION STUDY

Because the GPM is a mixture model, we can apply the following steps to generate a random sample with size \( n \) from it:

1. Generate one random instance of binomial distribution with parameters \( n \) and \( \frac{\gamma_{t+1}}{\gamma_{t+1}} \). Let the generated instance be \( n_1 \).

2. Generate one random sample with size \( n_1 \) from \( GP(\alpha, \beta, 1) \) and one random sample with size \( n_2 = n - n_1 \) from \( GP(\alpha, \beta, 2) \). Then, we combine these two samples to provide one random sample of GPM.

3. To generate a right-censored sample with censorship \( p \), we take the random censorship variable \( C_i \) to follow the degenerate distribution with mean \( t_\ast \). Then, \( t_\ast \) can be computed by solving equation \( t_\ast = q(p) \), where \( q() \) is the quantile function and defined in (10).

The results of a simulation study to estimate the parameters of GPM have been abstracted in Tables 1 and 2 for uncensored data and right-censored data with censorship rate \( p = 0.2 \), respectively. In every run, \( r = 500 \) replicates of samples with sizes \( n = 80 \) and 100 were generated. Then, the parameters were estimated by the maximum likelihood method or EM algorithm. The EM algorithm is an iterative method to find maximum likelihood and can be applied when the model depends on some latent variables, e.g., the mixture or competing risk models. For MLE, we use the built-in “optim” function of R with the default optimization method “Nelder-Mead”. In both MLE and EM approaches, for real parameters \( \alpha, \beta \) and \( \gamma \), the initial points have been generated from the uniform distribution on the intervals \((0, \alpha), (0, \beta)\) and \((0, \gamma)\) respectively. For real data, we should guess some different initial values for parameters and then compare the results by AIC criterion and Kolmogorov–Smirnov (K-S) statistic. Unfortunately, checking the conditions for terminating the EM process in every EM iteration, makes the simulation runs very slow and time consuming. So, we tested the EM algorithm for many times to find a big constant that is sufficiently large for EM iterations to converge. In this way, we found that 20 iterations is sufficient.

For each parameter, the bias (B), absolute bias (AB) and mean squared error (MSE) were computed. Every cell of these tables shows the A, AB and MSE for \( \alpha, \beta \) and mean squared error (MSE) were computed. Every cell of these tables shows the A, AB and MSE for each parameter.

V. APPLICATIONS

Table 3 shows the number of successive failures for the air conditioning system of 13 Boeing 720 jet airplanes that were analyzed by Tahir et al. [25] and many others. We fit the GPM to this data set by the maximum likelihood method and EM algorithm. Additionally, the Gumbel-Lomax (GuLx) and Weibull-Pareto (WP) distributions, which exhibit
TABLE 1. Simulation results for the MLE and EM estimators of the parameters of the GPM distribution for uncensored data. In every cell, the first, second and third lines are related to $\alpha$, $\beta$ and $\gamma$, respectively.

| n   | $\alpha, \beta, \gamma$ | MLE      | EM        |
|-----|-------------------------|----------|-----------|
|     |                         | B AB MSE | B AB MSE  |
| 80  | 0.7, 0.5, 0.5           | -0.03492 | 0.11637   |
|     |                         | 0.02046  | 0.02741   |
|     |                         | 0.15203  | 0.25741   |
|     | 0.19271 0.38371 0.94833| 0.08766  | 0.24854   |
|     |                         | 0.00494  | 0.01600   |
|     |                         | 0.00452  |
| 1,2,0,2,0,3 | 0.08056 0.22707 0.10473 | 0.03739  | 0.18307   |
|     |                         | 0.00297  | 0.01600   |
|     |                         | 0.00452  |
| 1,05,0,7 | 0.05179 0.15407 0.05193 | 0.03230  | 0.12621   |
|     |                         | 0.01840  | 0.24584   |
|     |                         | 0.00452  |
| 0.8, 2, 0.3 | 0.082140 3.138307 5.71420 | 0.08495  | 0.63135   |
|     |                         | 0.05270  | 0.38074   |
|     |                         | 0.00452  |
| 100 | 0.7, 0.5, 0.5           | 0.02231  | 0.10068   |
|     |                         | 0.01744  | 0.09279   |
|     |                         | 0.00452  |
| 1,2,0,2,0,3 | 0.11896 0.26948 0.16718 | 0.03174  | 0.19863   |
|     |                         | 0.04334  | 0.05532   |
|     |                         | 0.00505  |
| 1,05,0,7 | 0.07237 0.12039 0.04334 | -0.00330 | 0.05532   |
|     |                         | 0.00505  |
|     |                         | 0.00505  |
| 0.8, 2, 0.3 | 0.05185 0.03139 0.00209 | -0.0048  | 0.01408   |
|     |                         | 0.00035  |
|     |                         | 0.00035  |

decreasing and upside-down bathtub-shaped FR functions, were fit to this data set. Table 5 shows the estimates of the parameters, K-S statistic, $D_n = \sup_{x} \{F_n(x) - F(x)\}$ where $F_n(x)$ is the empirical distribution function, and its corresponding p-value. Akaike information criterion (AIC) is also included in this table. Large values of the K-S statistic and/or the AIC clearly indicate poor fit, so we can compare the fitted models using these measurements. The K-S statistic calculated for GPM (in particular by the EM algorithm) is smaller than others. Based on this criterion, GPM performs better than the others. However, based on the AIC, GuLx shows a better fit. The empirical and fitted CDFs are shown in Figure 3, left panel, visually confirming that the estimated GPM distribution is very close to the empirical distribution function.

Table 4 shows a dataset corresponding to the remission times (in months) of a random sample of 128 bladder cancer patients, referring to Tahir et al [26]. The results of fitting GPM, GuLx and WP to this dataset are also shown in Table 5. The K-S statistics and the AIC for GPM show smaller values than GuLx and WP, indicating that GPM is a good candidate to describe this dataset. Figure 3, right panel, which draws the empirical and fitted CDFs, confirms a very good fit.

VI. CONCLUSION

A new generalized Pareto model was introduced and some of the statistical and reliability properties were investigated. The proposed model can be applied in a variety of real-world situations, such as reliability and survival data where equilibrium is found in the distribution from "small" to "large" values, etc. The parameters of the proposed model were estimated using the MLE and EM algorithms. Simulation results confirm that EM performs better than MLE.

ACKNOWLEDGMENT

The author thanks four anonymous reviewers for their useful comments that lead to this improved version of the paper. This work is supported by Researchers Supporting Project number (RSP-2021/392), King Saud University, Riyadh.
TABLE 2. Simulation results for the MLE and EM estimators of the parameters of the GPM distribution when the censorship rate is 20 percent. In every cell, the first, second and third lines are related to $\alpha$, $\beta$ and $\gamma$, respectively.

| $n$ | $\alpha, \beta, \gamma$ | MLE | EM |
|-----|--------------------------|-----|----|
|     |                          | B   | AB | MSE | B   | AB | MSE |
| 80  | 0.7, 0.5, 0.5            | 0.05682 | 0.17180 | 0.06061 | -0.00933 | 0.13277 | 0.02971 |
|     |                          | 0.19560 | 0.37706 | 0.34730 | -0.01936 | 0.16777 | 0.04473 |
|     |                          | 5.95804 | 6.38670 | 171.9882 | -0.15251 | 0.23555 | 0.07630 |
|     | 1.2, 0.2, 0.3            | 0.78733 | 1.03488 | 13.47012 | 0.10901 | 0.35045 | 0.28269 |
|     |                          | 0.37793 | 0.44331 | 2.67611 | 0.01741 | 0.08802 | 0.01881 |
|     |                          | 3.61086 | 3.85548 | 99.04898 | -0.08889 | 0.16944 | 0.03970 |
|     | 1.0, 0.5, 0.7            | 0.25037 | 0.48243 | 3.84762 | -0.05258 | 0.27952 | 0.17328 |
|     |                          | 0.04462 | 0.06573 | 0.08483 | -0.00038 | 0.02206 | 0.00108 |
|     |                          | 6.87167 | 7.47129 | 201.6821 | -0.23615 | 0.32885 | 0.14982 |
|     | 0.8, 2, 0.3              | 0.11562 | 0.24517 | 0.18100 | 0.04134 | 0.18238 | 0.07003 |
|     |                          | 1.26735 | 1.88498 | 14.37903 | -0.12257 | 0.78199 | 1.62294 |
|     |                          | 4.62239 | 4.88856 | 131.7597 | -0.07956 | 0.17278 | 0.04225 |
| 100 | 0.7, 0.5, 0.5            | 0.04973 | 0.16121 | 0.05001 | 0.01461 | 0.12647 | 0.02888 |
|     |                          | 0.20270 | 0.36651 | 0.35795 | -0.01850 | 0.16854 | 0.04617 |
|     |                          | 5.25818 | 6.16985 | 165.6812 | -0.17395 | 0.24806 | 0.08129 |
|     | 1.2, 0.2, 0.3            | 0.41433 | 0.64580 | 6.16477 | 0.05642 | 0.29300 | 0.17374 |
|     |                          | 0.21964 | 0.28243 | 1.61510 | -0.00942 | 0.07685 | 0.01292 |
|     |                          | 3.00607 | 3.25485 | 84.9875 | -0.08119 | 0.15752 | 0.03446 |
|     | 1.0, 0.5, 0.7            | 0.11682 | 0.32925 | 0.59276 | -0.00913 | 0.22540 | 0.09748 |
|     |                          | 0.02325 | 0.04175 | 0.02537 | -0.00327 | 0.01835 | 0.00059 |
|     |                          | 0.25681 | 0.80138 | 182.5726 | -0.25698 | 0.32232 | 0.14114 |
|     | 0.8, 2, 0.3              | 0.08435 | 0.18718 | 0.07648 | -0.01617 | 0.15508 | 0.04225 |
|     |                          | 0.89584 | 1.40920 | 6.67489 | -0.00274 | 0.65027 | 0.80846 |
|     |                          | 3.78973 | 4.01250 | 103.7488 | -0.08853 | 0.15401 | 0.03262 |

TABLE 3. Number of successive failure for the air conditioning system.

| $n$ | 50 | 130 | 487 | 57 | 102 | 15 | 14 | 10 | 37 | 320 | 261 | 51 | 44 |
|-----|----|-----|-----|----|-----|----|----|----|----|-----|-----|----|----|
|     | 9  | 254 | 493 | 33 | 18  | 209 | 41 | 58 | 60 | 48  | 56  | 87 | 11 |
|     | 102| 12  | 5   | 14  | 14  | 29  | 37 | 186| 29 | 104 | 7   | 4   | 72 |
|     | 270| 283 | 7   | 61  | 100 | 61  | 502| 220| 120| 141 | 22  | 603| 35 |
|     | 98 | 54  | 100 | 11  | 181 | 65  | 49 | 12 | 239| 14  | 18  | 39  | 3  |
|     | 12 | 5   | 32  | 9   | 438 | 43  | 134| 184| 20 | 386 | 182 | 71  | 80 |
|     | 188| 230 | 152 | 5   | 36  | 79  | 59 | 33 | 246| 1   | 79  | 3   | 27 |
|     | 201| 84  | 27  | 156 | 21  | 16  | 88 | 130| 14 | 118 | 44  | 15  | 42 |
|     | 106| 46  | 230 | 26  | 59  | 153 | 104| 20 | 206| 5   | 66  | 34  | 29 |
|     | 26 | 35  | 5   | 82  | 31  | 118 | 326| 12 | 54 | 36  | 34  | 18  | 25 |
|     | 120| 31  | 22  | 18  | 216 | 139 | 67 | 310| 3  | 46  | 210 | 57  | 76 |
|     | 14 | 111 | 97  | 62  | 39  | 30  | 7  | 44 | 11 | 63  | 23  | 22  | 23 |
|     | 14 | 18  | 13  | 34  | 16  | 18  | 130| 90 | 163| 208 | 1   | 24  | 70 |
|     | 16 | 101 | 52  | 208 | 95  | 62  | 11 | 191| 14 | 71  |     |     |     |

Saudi Arabia.
### TABLE 4. Remission times of the bladder cancer.

| Time (years) | 0.08 | 0.20 | 0.40 | 0.50 | 0.80 | 1.05 | 1.19 | 1.26 | 1.35 | 1.40 | 1.46 |
|-------------|------|------|------|------|------|------|------|------|------|------|------|
| Data set    |      |      |      |      |      |      |      |      |      |      |      |
| Table 3     |      |      |      |      |      |      |      |      |      |      |      |
| GPM MLE     | 4.8842 | 341.9168 | 47.0243 | 0.0456 | 0.8505 | 1965.446 |
| GPM EM      | 4.2001 | 233.7010 | 3.2691 | 0.0446 | 0.8677 |        |
| GuLx MLE    | 21.9694 | 995.0593 | 1.9508 | 0.0481 | 0.8030 | 1961.591 |
| WP MLE      | 3.7422 | 0.2323 | 1     | 0.0433 | 0.8943 | 1965.856 |
| Table 4     |      |      |      |      |      |      |      |      |      |      |      |
| GPM MLE     | 4.0242 | 15.0751 | 0.13901 | 0.03605 | 0.9963 | 825.2818 |
| GPM EM      | 3.9345 | 14.50030 | 0.11655 | 0.03409 | 0.9984 |        |
| GuLx MLE    | 44.5793 | 236.7911 | 1.7973 | 0.1050 | 0.1185 | 841.8991 |
| WP MLE      | 4.9035 | 0.21297 | 0.08 | 0.0450 | 0.9591 | 840.0858 |

### TABLE 5. Results of fitting the data sets to some models.

| Data set | Model | Method | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\gamma}$ | $D_n$ | p-value | AIC  |
|----------|-------|--------|----------------|----------------|----------------|-------|---------|------|
| Table 3  | GPM   | MLE    | 4.8842         | 341.9168       | 47.0243        | 0.0456 | 0.8505  | 1965.446 |
|          | GPM   | EM     | 4.2001         | 233.7010       | 3.2691         | 0.0446 | 0.8677  |        |
|          | GuLx  | MLE    | 21.9694        | 995.0593       | 1.9508         | 0.0481 | 0.8030  | 1961.591 |
|          | WP    | MLE    | 3.7422         | 0.2323         | 1              | 0.0433 | 0.8943  | 1965.856 |
| Table 4  | GPM   | MLE    | 4.0242         | 15.0751        | 0.13901        | 0.03605 | 0.9963 | 825.2818 |
|          | GPM   | EM     | 3.9345         | 14.50030       | 0.11655        | 0.03409 | 0.9984 |        |
|          | GuLx  | MLE    | 44.5793        | 236.7911       | 1.7973         | 0.1050 | 0.1185 | 841.8991 |
|          | WP    | MLE    | 4.9035         | 0.21297        | 0.08           | 0.0450 | 0.9591 | 840.0858 |

### FIGURE 3. Empirical distribution and fitted GPM distribution for data sets of Table 3 (left) and Table 4 (right).

### REFERENCES

[1] S. M. Burroughs and S. F. Tebbens, “Upper-truncated power law distributions”, Fractals, vol. 9, pp. 209-222, 2001.
[2] B. Schroeder, S. Damouras, P. Gill, “Understanding latent sector error and how to protect against them,” ACM Transactions on Storage (TOS), vol. 6, no. 3, 2010.
[3] A. Akinsete, F. Famoye and C. Lee, “The beta-Pareto distribution,” Statistics, vol. 42, pp. 547-563, 2008.
[4] M.M. Nassar, N.K. Nada, “The beta generalized Pareto distribution,” Journal of Statistics: Advances in Theory and Applications, vol. 6, pp. 1-
Appendix A: Some Detailed Information

The log-likelihood function related to the EM algorithm for complete data can be simplified as follows.

\[
l(\Theta; x, z) = \ln L(\Theta; x, z) = \sum_{i=1}^{n} \sum_{j=1}^{2} I(z_i = j) \times \left\{ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{x_i}{\beta}) + I(j = 2) \left[ \ln(\alpha + 1) + \ln\left(\frac{x_i}{\beta} + x_i\right) \right] \right\} + I(j = 1) \ln \gamma - \ln(\gamma + 1) \right\} = \sum_{i=1}^{n} I(Z_i = 1) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{x_i}{\beta}) + \ln \gamma - \ln(1 + \gamma) \right] + I(Z_i = 2) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{x_i}{\beta}) + \ln(\alpha + 1) = \ln \frac{x_i}{x_i + \beta} - \ln(1 + \gamma) \right].
\]

Membership probabilities of the EM algorithm for complete data are simplified as follows.

\[
p_{ij,t} = P(Z_i = j | X_i = x_i, \Theta_t) = \frac{f(X_i = x_i | Z_i = j, \Theta_t) P(Z_i = j | \Theta_t)}{f(X_i = x_i | \Theta_t)} = \frac{\frac{\alpha_t}{\beta_t} \left(1 + \frac{x_i}{\beta_t}\right)^{-\alpha_t-1} \left(\alpha_t + 1\right) \left(\frac{x_i}{\beta_t} + x_i\right)^{I(j=2)} \gamma_t^{I(j=1)} \frac{1}{\gamma_t + 1}}{\sum_{j=1}^{2} \frac{\alpha_t}{\beta_t} \left(1 + \frac{x_i}{\beta_t}\right)^{-\alpha_t-1} \left(\alpha_t + 1\right) \left(\frac{x_i}{\beta_t} + x_i\right)^{I(j=2)} \gamma_t^{I(j=1)} \frac{1}{\gamma_t + 1}} = \frac{\gamma_t (\beta_t + x_i) I(j = 1) + (\alpha_t + 1) x_i I(j = 2)}{\gamma_t (\beta_t + x_i) (\beta_t + x_i I(j = 1) + (\alpha_t + 1) x_i I(j = 2)} \quad i = 1, 2, ..., n, \quad j = 1, 2.
\]

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
The log-likelihood function related to the EM algorithm for complete data can be written as follows.

\[
Q(\Theta | \Theta_t) = E_Z|X_1, \Theta_t[l(\Theta; x, Z)] \\
= \sum_{i=1}^{n} E_{Z_i|X_i, \Theta_t} \left[ \sum_{j=1}^{2} I(z_i = j) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{x_i}{\beta}) + I(j = 2)(\ln(\alpha + 1) + \ln \frac{x_i}{\beta} + \frac{x_i}{\beta} + \frac{x_i}{\beta^2}) + I(j = 1) \times \ln \beta + \ln(\gamma + 1) \right] \right] \\
+ \sum_{i=1}^{n} p_{2i,t} \left( \ln(\alpha + 1) + \ln \frac{x_i}{\beta} + \frac{x_i}{\beta} + \frac{x_i}{\beta^2} \right) \\
+ \sum_{i=1}^{n} p_{1i,t} \ln \gamma - n \ln(\gamma + 1). 
\]

(34)

The log-likelihood function related to the EM algorithm for right-censored data has been simplified as follows.

\[
l(\Theta; t, d, z) = \ln L(\Theta; t, d, z) \\
= \sum_{i=1}^{n} d_i \sum_{j=1}^{2} I(z_i = j) \ln f_j(t_i; \alpha, \beta) P(Z_i = j) \\
+ \sum_{i=1}^{n} (1 - d_i) I(Z_i = j) \ln R_j(t_i; \alpha, \beta) P(Z_i = j) \\
= \sum_{i=1}^{n} d_i I(Z_i = 1) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{t_i}{\beta}) + \ln \gamma - \ln(\gamma + 1) \right] + \sum_{i=1}^{n} d_i I(Z_i = 2) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \times \ln(1 + \frac{t_i}{\beta}) + \ln(\gamma + 1) \right] \\
+ \sum_{i=1}^{n} (1 - d_i) I(Z_i = 1) \ln \left[ (1 + \frac{t_i}{\beta})^{-\alpha} \frac{\gamma}{\gamma + 1} \right] \\
+ \sum_{i=1}^{n} (1 - d_i) I(Z_i = 2) \ln \left[ (1 + \frac{t_i}{\beta})^{-\alpha} \times (\alpha + 1 - \frac{\alpha \beta}{t_i + \beta}) \frac{1}{\gamma + 1} \right]. 
\]

(35)

Expectation function of the EM algorithm for right-censored data is equal to

\[
Q(\Theta | \Theta_t) = E_Z|t, d, \Theta_t[l(\Theta; t, d, z)] \\
= \sum_{i=1}^{n} d_i P(Z_i = 1|t, d, \Theta_t) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{t_i}{\beta}) + \ln \gamma - \ln(\gamma + 1) \right] \\
+ \sum_{i=1}^{n} d_i P(Z_i = 2|t, d, \Theta_t) \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{t_i}{\beta}) + \ln(\gamma + 1) \right] \\
+ \sum_{i=1}^{n} (1 - d_i) P(Z_i = 1|t, d, \Theta_t) \left[ - \ln(1 + \frac{t_i}{\beta}) + \ln \gamma - \ln(\gamma + 1) \right] \\
+ \sum_{i=1}^{n} (1 - d_i) P(Z_i = 2|t, d, \Theta_t) \left[ - \ln(1 + \frac{t_i}{\beta}) + \ln(\gamma + 1) - \ln(\gamma + 1) \right]. 
\]

(36)

After simplification, it reduces to

\[
Q(\Theta | \Theta_t) = \sum_{i=1}^{n} d_i \left[ \ln \frac{\alpha}{\beta} - (\alpha + 1) \ln(1 + \frac{t_i}{\beta}) + \ln \gamma - \ln(\gamma + 1) \right] \\
+ \left[ - \alpha \ln \frac{t_i}{t_i + \beta} + \ln \gamma - \ln(\gamma + 1) \right] \\
+ \left[ - \alpha \ln \frac{t_i}{t_i + \beta} + \ln(\gamma + 1) - \ln(\gamma + 1) \right], 
\]

(37)

Membership probabilities of the EM algorithm for right-
Mohamed Kayid is a Professor of Statistics at the Department of Statistics and Operations Research at King Saud University. He obtained the BSc Degree (Mathematics, 1991) from Minia University, the M.Sc. Degree (Statistics, 1999) from Al-Azhar University, and a Ph.D. Degree (Statistics, 2003) from South Valley University. His research interests are in reliability theory, preventive maintenance, stochastic orders, and their applications, and he has published in various international journals on these issues. He is an elected Reviewer for American Mathematical Society. He has varied research interests in applied statistics. Some of the research methodologies proposed by him and his coauthors have found interesting applications in the context of reliability and engineering. His current research interests can be classified into four broad areas: theory of stochastic ageing and its applications, estimation in restricted parametric space, theory of stochastic orders, ranking and selection problems and related estimation problems. His work in this area has received attention among engineers and been referenced in electrical engineering and computer science journals.