AUT Journal of Mathematics and Computing

AUT J. Math. Com., 1(1) (2020) 69-87
DOI: 10.22060/ajmc.2018.14869.1012

A generalization of Marshall-Olkin bivariate Pareto model and its applications in shock and competing risk models

Shirin Shoaee*[a], Esmaeil Khorram[b]

[a] *Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran*
[b] *Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran*

ABSTRACT: Statistical inference for extremes has been a subject of intensive research during the last years. In this paper, we generalize the Marshall-Olkin bivariate Pareto distribution. In this case, a new bivariate distribution is introduced by compounding the Pareto Type II and geometric distributions. This new bivariate distribution has natural interpretations and can be applied in fatal shock models or in competing risks models. We call the new proposed model Marshall-Olkin bivariate Pareto-geometric (MOBPG) distribution, and then investigate various properties of the new distribution. This model has five unknown parameters and the maximum likelihood estimators cannot be afforded in explicit structure. We suggest to use the EM algorithm to calculate the maximum likelihood estimators of the unknown parameters, and this structure is quite flexible. Also, Monte Carlo simulations are performed to investigate the effectiveness of the proposed algorithm. Finally, we analyze a real data set to investigate our purposes.

Review History:

Received:02 August 2018
Revised:20 October 2018
Accepted:05 November 2018
Available Online:1 February 2020

Keywords:

Bivariate model
Competing risk model
Expectation-Maximization algorithm
Pareto Type II distribution
Shock model

1. Introduction

Statistical modeling of extreme values has been extensively developed during the last decades. Also, Many studies apply this model to a wide range of important problems such as extreme wind speeds, wave heights, floods, insurance claims and price fluctuations.

The Pareto distribution is a skewed, heavy-tailed distribution that is sometimes used in business, economics, actuarial science, queueing theory and Internet traffic modeling. The two-parameter Pareto Type II distribution has the following cumulative distribution function (CDF) and probability density function (PDF), respectively:

$$f(x,\alpha,\lambda) = \alpha\lambda(1+\alpha x)^{-\lambda-1}, \qquad x>0, \alpha>0, \lambda>0, \quad F(x,\alpha,\lambda) = 1-(1+\alpha x)^{-\lambda}, \qquad x>0, \alpha>0, \lambda>0.$$

Also, $Pa(\alpha,\lambda)$ explains the Pareto distribution with the shape parameter $\lambda>0$ and scale parameter $\alpha>0$. The modeling of a lifetime is an important problem in a variety of scientific and technological fields. Several methods have been proposed for multivariate survival data. For an extensive discussion of multivariate models and their properties and applications, one may refer to [15].

In this paper, we introduce Marshal-Olkin formulation ([22]) of Pareto distribution. In the classical Marshall-Olkin model, a system consists of two components which are exposed to shocks arriving from three sources. These shocks are built randomly. The shock of the first source modifies the first component, the shock of the second source modifies the second component and the shock of the third source modifies both components. The independent exponential shock times are discussed by [22]. They conclude the joint survival function of the components

*Corresponding author,
E-mail addresses: Sh_Shoaee@sbu.ac.ir, eskhor@aut.ac.ir

of the system. In recent years, many articles have been devoted to the generalization of Marshall and Olkin's model. [22] proposed a bivariate Weibull distribution. [2], suggested a bivariate distribution using the Gompertz and an exponential distribution. [27] introduced a bivariate distribution by generalized exponential an exponential distribution. [19] improved the proposed model by Sarhan and Balakrishnan assuming different generalized exponential distributions for the components. In this paper, we discuss a bivariate Pareto distribution whose marginals have Pareto distributions using the proposed idea of Marshall and Olkin. This new bivariate distribution is called the Marshall-Olkin bivariate Pareto (MOBP) distribution and several other properties of the MOBP model are investigated.

Also, for modeling survival data, [24] introduced a procedure to add an extra parameter to a family of univariate distributions, and considered the generalization of exponential and Weibull models in details. In this method, they presented a class of univariate distributions. So that this class is acquired by a minimum and maximum of independent and identically distributed continuous random variables, and the sample size has the geometric distribution. Therefore, because of an additional parameter, the new class of univariate distributions has more flexible than the exponential or Weibull class, respectively. In fact, continuous compounding with discrete distributions have been presented and discussed in recent years. This method introduces distributions with great flexibility and is useful in expanding statistical models which use in a variety of applications. Then, the same method for several distributions is studied by many researchers, see for example [11], [12], [26], [10], [8], [5] and their references.

In fact, it can be seen that due to the lack of analyzes for bivariate distributions, there has not been much work on these distributions and their expansion. Therefore, assuming that the failure times follow the Pareto Type II distribution, the main aim of this paper is to generalize the Marshall-Olkin bivariate Pareto (MOBP) distribution by the structure of [24]. In fact, this procedure obtains distributions with more flexibility. This new bivariate distribution is called the Marshall-Olkin bivariate Pareto-geometric (MOBPG) distribution. Also, the marginals and conditionals are univariate Pareto-geometric distributions (UPG), and they are also very flexible.

We discuss different properties of this new distribution. Because of the presence of five parameters, this model is very flexible and the joint probability density function (PDF) has different shapes. Thus, it can be applied exactly to analyze bivariate data. Also, the generation of random samples from the proposed bivariate model is very simple, so, simulation studies can be done very easily.

The estimation of maximum likelihood of the unknown parameters in the MOBPG distribution cannot be computed in closed formations. We should solve five non-linear equations simultaneously. So, there are serious problems to solve them. One can mention the detection of the initial starting value for the algorithm and the convergence of the algorithm. To solve this subject, we investigate it as a missing value problem and suggest using the expectation maximization (EM) algorithm to calculate the MLEs. In this algorithm, at each E-step, we solve one-dimensional non-linear optimization problems. Therefore, the execution of the proposed EM algorithm is very simple. This model is very flexible and the performance is very simple, so, it provides us a choice of another bivariate model, which may provide a better fit than the available models.

For study purposes, this article is organized as follows: a bivariate Pareto distribution of the Marshall-Olkin type is obtained, in Section 2. Also, various properties of the new bivariate distribution are discussed in this section. We generalize the Marshall-Olkin bivariate Pareto distribution, and a new bivariate model of the proposed method of Marshall and Olkin (1997) is introduced. In addition, various properties of this distribution are investigated in Section 3. The MLE of the unknown parameters are computed in Section 4. The results of the simulation experiments and a real data set have been presented in Section 5 and finally, we conclude the paper in Section 6.

2. Model Formulation

The important purpose of this section is to create a bivariate Pareto distribution so that the marginals have Pareto distributions using the same structure of Theorem 3.2 proposed by [22]. This distribution has wide applications in modeling, data related to finance, insurance, environmental sciences and the internet network.

2.1. Marshall-Olkin Bivariate Pareto Distribution

In this subsection, we will introduce the bivariate Pareto distribution. Suppose $U_0 \sim Pa(\alpha, \lambda_0)$, $U_1 \sim Pa(\alpha, \lambda_1)$, $U_2 \sim Pa(\alpha, \lambda_2)$ and they are independent. Then, we consider $X_1 = \min\{U_0, U_1\}$, and $X_2 = \min\{U_0, U_2\}$. Therefore, the bivariate vector (X_1, X_2) is a bivariate Pareto distribution with the parameters $\alpha, \lambda_0, \lambda_1, \lambda_2$ and is displayed as $MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$.

Theorem 1. *If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, so, their joint survival function has the following structure for $z = \max\{x_1, x_2\}$,*

$$\bar{F}_{X_1,X_2}(x_1, x_2) = P(X_1 \geq x_1, X_2 \geq x_2) = P(U_1 \geq x_1, U_2 \geq x_2, U_0 \geq z) \quad = (1 + \alpha x_1)^{-\lambda_1}(1 + \alpha x_2)^{-\lambda_2}(1 + \alpha z)^{-\lambda_0}.$$

Corollary 2.1. *The joint survival function of the $MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$ can be obtained as:*

$$\begin{aligned}\bar{F}_{X_1,X_2}(x_1, x_2) &= \bar{F}_{CH}(x_1, \alpha, \lambda_1)\bar{F}_{CH}(x_2, \alpha, \lambda_2)\bar{F}_{CH}(z, \alpha, \lambda_0) \\ &= \begin{cases} \bar{F}_{Pa}(x_1, \alpha, \lambda_1 + \lambda_0)\bar{F}_{Pa}(x_2, \alpha, \lambda_2) & \text{if} \quad x_2 < x_1 \\ \bar{F}_{Pa}(x_1, \alpha, \lambda_1)\bar{F}_{Pa}(x_2, \alpha, \lambda_2 + \lambda_0) & \text{if} \quad x_1 < x_2 \\ \bar{F}_{Pa}(x, \alpha, \lambda_0 + \lambda_1 + \lambda_2, \lambda) & \text{if} \quad x_1 = x_2 = x. \end{cases}\end{aligned} \tag{1}$$

Theorem 2. *If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then the joint PDF of (X_1, X_2) is:*

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} f_1(x_1, x_2) & \text{if} \quad x_2 < x_1 \\ f_2(x_1, x_2) & \text{if} \quad x_1 < x_2 \\ f_0(x) & \text{if} \quad 0 < x_1 = x_2 = x < \infty, \end{cases} \tag{2}$$

where

$$\begin{aligned} &f_1(x_1, x_2) = \alpha^2\lambda_2(\lambda_0 + \lambda_1)(1 + \alpha x_1)^{-(\lambda_0+\lambda_1)-1}(1 + \alpha x_2)^{-\lambda_2-1}. \\ &f_2(x_1, x_2) = \alpha^2\lambda_1(\lambda_0 + \lambda_2)(1 + \alpha x_1)^{-\lambda_1-1}(1 + \alpha x_2)^{-(\lambda_0+\lambda_2)-1}. \\ &f_0(x) = \alpha\lambda_0(1 + \alpha x)^{-(\lambda_0+\lambda_1+\lambda_2)-1}. \end{aligned}$$

Proof: The phrases of $f_1(.,.)$ and $f_2(.,.)$ can be received by $-\frac{\partial^2 \bar{F}_{X_1,X_2}(x_1,x_2)}{\partial x_1 \partial x_2}$ for $x_1 < x_2$ and $x_2 < x_1$ respectively. But $f_0(.)$ can not be established in a similar method. Now considering that

$$\int_0^\infty \int_{x_2}^\infty f_1(x_1, x_2)dx_1dx_2 + \int_0^\infty \int_{x_1}^\infty f_2(x_1, x_2)dx_2dx_1 + \int_0^\infty f_0(x)dx = 1.$$

So,

$$\int_0^\infty \int_{x_2}^\infty f_1(x_1, x_2)dx_1dx_2 = \lambda_2 \int_0^\infty \alpha(1 + \alpha x_2)^{-(\lambda_0+\lambda_1+\lambda_2)-1}dx_2 = \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}.$$

and similarly,

$$\int_0^\infty \int_{x_1}^\infty f_2(x_1, x_2)dx_2dx_1 = \lambda_1 \int_0^\infty \alpha(1 + \alpha x_1)^{-(\lambda_0+\lambda_1+\lambda_2)-1}dx_1 = \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_2}.$$

Note that

$$\int_0^\infty f_0(x)dx = \lambda_0 \int_0^\infty \alpha(1 + \alpha x)^{-(\lambda_0+\lambda_1+\lambda_2)-1}dx = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}.$$

Therefore, the result was obtained..

It is necessary to mention that the MOBP distribution has two parts, similar to the Marshall-Olkin bivariate exponential or bivariate Weibull model. It has an absolute continuous part and a singular part. We know that, the PDF of the MOBP distribution is $f_{X_1,X_2}(.,.)$ in Theorem 2. It is quite clear that the first two phrases are densities with respect to the two dimensional Lebesgue measure and the third expression is a density function with respect to the one dimensional Lebesgue measure. In the MOBP distribution, we have a singular part. It means that if X_1 and X_2 are MOBP distribution, then $X_1 = X_2$ has a positive probability. In many applied states, it occurred that X_1 and X_2 are both continuous random variables, but $X_1 = X_2$ has a positive probability, see [22]. In the following, we will obtain some structures to investigate the absolute continuous part and the singular part of the MOBP distribution.

Theorem 3. *If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$. Then,*

$$\bar{F}_{X_1,X_2}(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}\bar{F}_a(x_1, x_2) + \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}\bar{F}_s(x_1, x_2), \tag{3}$$

where for $z = \max\{x_1, x_2\}$,

$$\bar{F}_s(x_1, x_2) = (1 + \alpha z)^{-(\lambda_0+\lambda_1+\lambda_2)},$$

and

$$\bar{F}_a(x_1, x_2) = \frac{\lambda_0 + \lambda_1 + \lambda_2}{\lambda_1 + \lambda_2}(1 + \alpha x_1)^{-\lambda_1}(1 + \alpha x_2)^{-\lambda_2}(1 + \alpha z)^{-\lambda_0} - \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}(1 + \alpha z)^{-(\lambda_0+\lambda_1+\lambda_2)}.$$

proof 1. *Suppose A is the following event:*

$$A = \{U_0 < U_1\} \cap \{U_0 < U_2\}.$$

Then $P(A) = \lambda_0/(\lambda_0 + \lambda_1 + \lambda_2)$ and $P(A') = (\lambda_1 + \lambda_2)/(\lambda_0 + \lambda_1 + \lambda_2)$. Therefore,

$$\bar{F}_{X_1,X_2}(x_1, x_2) = P(X_1 \geq x_1, X_2 \geq x_2|A)P(A) + P(X_1 \geq x_1, X_2 \geq x_2|A')P(A').$$

Moreover, for z as defined before,

$$\begin{aligned} P(X_1 \geq x_1, X_2 \geq x_2|A) &= [P(A)]^{-1}P(U_1 \geq U_0, U_2 \geq U_0, U_0 \geq z) \\ &= (1 + \alpha z)^{-(\lambda_0+\lambda_1+\lambda_2)}. \end{aligned}$$

Then, we calculate $P(X_1 \geq x_1, X_2 \geq x_2|A')$ using subtraction. It is immediate that $(1 + \alpha z)^{-(\lambda_0+\lambda_1+\lambda_2)}$ is the singular part as its mixed second partial derivatives is zero when $x_1 \neq x_2$, and $P(X_1 \geq x_1, X_2 \geq x_2|A')$ is the absolute continuous part as its mixed second partial derivatives is a bivariate density function.

Corollary 2.2. *The joint PDF of (X_1, X_2) can be obtained as follows for $z = \max\{x_1, x_2\}$;*

$$f_{X_1,X_2}(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_0 + \lambda_1 + \lambda_2} f_a(x_1, x_2) + \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_s(z), \tag{4}$$

where

$$f_a(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_0 + \lambda_1 + \lambda_2} \begin{cases} f_{Pa}(x_1, \alpha, \lambda_0 + \lambda_1) f_{Pa}(x_2, \alpha, \lambda_2) & \text{if} \quad x_2 < x_1 \\ f_{Pa}(x_2, \alpha, \lambda_1) f_{Pa}(x_2, \alpha, \lambda_0 + \lambda_2) & \text{if} \quad x_1 < x_2, \end{cases}$$

and

$$f_s(z) = f_{Pa}(z, \alpha, \lambda_0 + \lambda_1 + \lambda_2).$$

Thus, here $f_a(x_1, x_2)$ and $f_s(z)$ are the absolutely continuous and singular part, respectively. Also, the surface plot of the joint probability density function of MOBP is drawn in Figure 1.

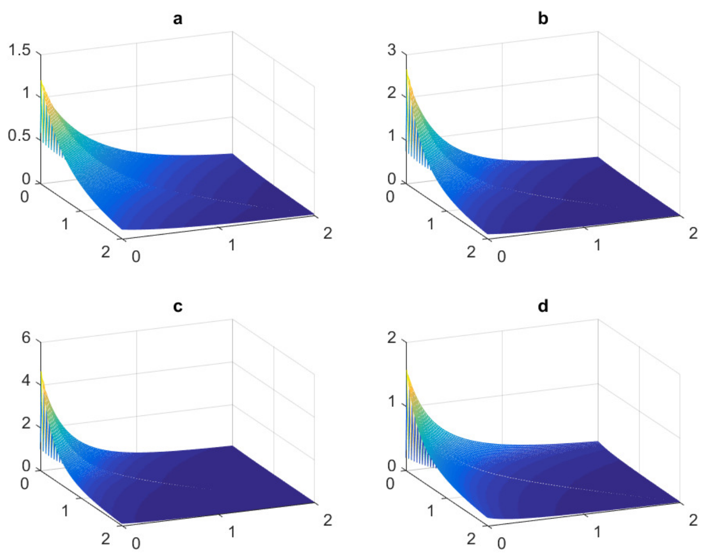

Figure 1: The shape of the joint probability density function of MOBP for various values of parameters $\Theta = (\alpha, \lambda_0, \lambda_1, \lambda_2)$: (a) $\Theta = (0.5, 1, 1.5, 2)$ (b) $\Theta = (0.75, 1, 1.5, 2)$ (c) $\Theta = (1, 1, 1.5, 2)$ (d) $\Theta = (2, 0.1, 0.5, 0.7)$.

It is easily seen from Theorem 3 that for fixed λ_1 and λ_2, as $\lambda_0 \to 0$,

$$\bar{F}_{X_1,X_2}(x_1, x_2) \to (1 + \alpha x_1)^{-\lambda_1}(1 + \alpha x_2)^{-\lambda_2},$$

therefore, X_1 and X_2 become independent. Moreover, since

$$A = (U_1 < U_0) \cap (U_2 < U_0) = \{\max\{U_1, U_2\} > U_0\} = \{X_1 = X_2\},$$

and $P(A) = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}$, therefore, as $\lambda_0 \to \infty$, $P(A) = P(X_1 = X_2) \to 1$, i.e. X_1 and X_2 are asymptotically almost surely equal. This implies that for fixed λ_1 and λ_2, as λ_0 varies from 0 to ∞, the correlation between X_1 and X_2 varies from 0 to 1.

2.2. Different Properties

In this subsection, we discuss various basic properties of the MOBP model. First, we discuss the marginal and conditional distributions of the MOBP model. An algorithm is presented to generate a random sample of MOBP distribution. Also, the ageing properties and the bivariate hazard gradient are discussed.

Proposition 2.1. *If* $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$. *Then,*

1) $X_1 \sim Pa(\alpha, \lambda_0 + \lambda_1)$ *and* $X_2 \sim Pa(\alpha, \lambda_0 + \lambda_2)$.
2) $P(X_1 < X_2) = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}$.
3) $\min\{X_1, X_2\} \sim MOBP(\alpha, \lambda_0 + \lambda_1 + \lambda_2)$.

Proof: They can easily be obtained.

Algorithm to generate from MOBP:

Step 1 Generate v_0, v_1 and v_2 from $\mathrm{U}(0, 1)$,

Step 2 Compute

$$U_0 = \frac{(1-v_0)^{-\frac{1}{\lambda_0}} - 1}{\alpha}, \qquad U_1 = \frac{(1-v_1)^{-\frac{1}{\lambda_1}} - 1}{\alpha}, \qquad U_2 = \frac{(1-v_2)^{-\frac{1}{\lambda_2}} - 1}{\alpha},$$

Step 3 Obtain

$$X_1 = \min\{U_0, U_1\}, \qquad \text{and} \qquad X_2 = \min\{U_0, U_2\}.$$

Hence, the marginal distributions of the bivariate vector (X_1, X_2) are the Pareto distributions, therefore, they can have a decreasing failure rate function.

Proposition 2.2. *If* $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$. *Then,*

1) *The conditional survival function of* X_1 *given* $X_2 \geq x_2$, *say* $\bar{F}_{X_1|X_2\geq x_2}(x_1)$ *is an absolutely continuous survival function as follows:*

$$P(X_1 \geq x_1 | X_2 \geq x_2) = \bar{F}_{X_1|X_2\geq x_2}(x_1) = \begin{cases} (1+\alpha x_1)^{-(\lambda_0+\lambda_1)}(1+\alpha x_2)^{\lambda_0} & if \quad x_2 < x_1 \\ (1+\alpha x_1)^{-\lambda_1} & if \quad x_1 < x_2. \end{cases} \tag{5}$$

2) *The conditional survival function in (5) has a presentation*

$$\bar{F}_{X_1|X_2\geq x_2}(x_1) = pG(x_1) + (1-p)H(x_1),$$

where,

$$G(x_1) = \frac{1}{p}\begin{cases} (1+\alpha x_1)^{-(\lambda_0+\lambda_1)}(1+\alpha x_2)^{\lambda_0} & if \quad x_2 < x_1 \\ (1+\alpha x_1)^{-\lambda_1} - \frac{\lambda_0}{\lambda_0+\lambda_2}(1+\alpha x_2)^{-\lambda_1} & if \quad x_1 < x_2, \end{cases} \qquad H(x) = \begin{cases} 1 & if \quad x < x_2 \\ 0 & if \quad x > x_2, \end{cases}$$

and

$$p = 1 - \frac{\lambda_0}{\lambda_0+\lambda_2}(1+\alpha x_2)^{-\lambda_1}.$$

Proof: The proofs can be obtained in a routine manner.

2.3. Ageing Properties and Bivariate Hazard Gradient

(a) If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then for $\alpha > 1$, it can be shown that

$$\frac{P(X_1 > x_1 + t, X_2 > x_2 + t)}{P(X_1 > x_1, X_2 > x_2)},$$

increases in x_1 and x_2 for $t > 0$. Thus, in this state (X_1, X_2) has the multivariate decreasing failure rate (MDFR) property.

(b) [13] defined the bivariate hazard gradient as follows:

$$h_{X_1,X_2}(x_1, x_2) = (-\frac{\partial}{\partial x_1}, -\frac{\partial}{\partial x_2}) \ln P(X_1 > x_1, X_2 > x_2).$$

If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then for $\alpha \geq 1$ and all values of $x_1 > 0, x_2 > 0$, both the components of $h_{X_1,X_2}(x_1, x_2)$ are decreasing functions of x_1 and x_2.

2.4. Dependence

We can describe various concepts of positive and negative dependence for multivariate distributions which are presented in the literature, see for example, [6].

(a) A random vector (X_1, X_2) has a positive upper orthant dependent if for all $x_1 > 0$ and $x_2 > 0$,

$$P(X_1 \geq x_1, X_2 \geq x_2) \geq P(X_1 > x_1)P(X_2 > x_2). \tag{6}$$

(see [20]). If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then (X_1, X_2) satisfies (6). So, (X_1, X_2) is positive upper orthant dependent.

(b) As we know, the connection between joint survival function and the joint CDF is as follows:

$$\bar{F}_{X_1,X_2}(x_1, x_2) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1,X_2}(x_1, x_2).$$

If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then $\bar{F}_{X_1,X_2}(x_1, x_2) \geq \bar{F}_{X_1}(x_1)\bar{F}_{X_2}(x_2)$ for all x_1, x_2. So, $F_{X_1,X_2}(x_1, x_2) \geq F_{X_1}(x_1)F_{X_2}(x_2)$ for all x_1, x_2, therefore, they will be positive quadrant dependent, i.e., for every pair of increasing functions $h_1(.)$ and $h_2(.)$ ([7]) the following relation is satisfied:

$$Cov(h_1(X_1), h_2(X_2)) > 0.$$

(c) A random vector (X_1, X_2) has the right tail increasing (RTI) property if for $i \neq j$,

$$P(X_i > x_i | X_j > x_j). \tag{7}$$

be non-decreasing in x_j for all $x_i > 0$. If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then (X_1, X_2) satisfies 7. So, (X_1, X_2) has the right tail increasing property.
In the following, the bivariate right corner set increasing (RCSI) is explained, which can be an alternative bivariate dependence concept.

(d) A random vector (X_1, X_2) has the right corner set increasing (RCSI) property if

$$P(X_1 > x_1, X_2 > x_2 | X_1 \geq \tilde{x}_1, X_2 \geq \tilde{x}_2). \tag{8}$$

increases in $\tilde{x}_1, \tilde{x}_2$ for every choice of (x_1, x_2). If $(X_1, X_2) \sim MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$, then, (X_1, X_2) satisfies (8). So, (X_1, X_2) has the RCSI property.

3. Generalization: Marshall-Olkin Bivariate Pareto-Geometric Distribution

In this section, the Marshall-Olkin bivariate Pareto-geometric distribution is introduced. For this purpose, suppose $\{(X_{1n}, X_{2n}); n = 1, 2, \ldots\}$ is a sequence of i.i.d. non-negative bivariate random variables with common joint distribution function $F_X(.,.)$ where $X = (X_1, X_2)$ and N is a geometric random variable independent of $\{(X_{1n}, X_{2n}), n = 1, 2 \ldots\}$. Also, a random variable N has a geometric distribution function which it is denoted by $GM(\theta)$ and is defined as follow:

$$P(N = n) = \theta(1-\theta)^{n-1}, \qquad n = 1, 2, 3, \ldots, \quad 0 < \theta < 1. \tag{9}$$

In the following, we will attend the bivariate random variable $Y = (Y_1, Y_2)$:

$$Y_1 = \min\{X_{11}, \ldots, X_{1N}\}, \qquad and \qquad Y_2 = \min\{X_{21}, \ldots, X_{2N}\}.$$

The joint survival function of $Y = (Y_1, Y_2)$ becomes:

$$\begin{aligned} P(Y_1 > y_1, Y_2 > y_2) &= \sum_{n=1}^{\infty} P(Y_1 > y_1, Y_2 > y_2 | N = n) P(N = n) \\ &= \sum_{n=1}^{\infty} [\bar{F}_X(y_1, y_2)]^n \theta (1-\theta)^{n-1} = \frac{\theta \bar{F}_X(y_1, y_2)}{1 - (1-\theta)\bar{F}_X(y_1, y_2)}. \end{aligned} \tag{10}$$

In this case, we call Y a bivariate F-geometric (BFG) distribution. Then the survival function of Y_i is:

$$\bar{F}_{Y_i}(y_i) = \frac{\theta \bar{F}_{X_i}(y_i)}{1 - (1-\theta)\bar{F}_{X_i}(y_i)}, \qquad i = 1, 2. \tag{11}$$

where, $\bar{F}_{X_i}, i = 1, 2$ are the marginal survival functions of $\bar{F}_X$, i.e. $\bar{F}_{X_1}(x) = \bar{F}_X(x, 0)$, $\bar{F}_{X_2}(x) = \bar{F}_X(0, x)$. Therefore, the random variable $Y = (Y_1, Y_2)$ has the bivariate Pareto-geometric distribution with parameters $\theta, \alpha, \lambda_0, \lambda_1, \lambda_2$, if the distribution F_X in (10) is $MOBP(\alpha, \lambda_0, \lambda_1, \lambda_2)$. Therefore, the joint survival function of (Y_1, Y_2) becomes;

$$\begin{aligned} \bar{G}_Y(y_1, y_2) &= P(Y_1 > y_1, Y_2 > y_2) \\ &= \begin{cases} \frac{\theta(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}}{1-(1-\theta)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}} & \text{if} \quad y_2 \le y_1 \\ \frac{\theta(1+\alpha y_1)^{-\lambda_1}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}}{1-(1-\theta)(1+\alpha y_1)^{-\lambda_1)}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}} & \text{if} \quad y_1 < y_2. \end{cases} \end{aligned} \tag{12}$$

(13)

It will be denoted by $(Y_1, Y_2) \sim MOBPG(\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$.

Proposition 3.1. *If $(Y_1, Y_2) \sim MOBPG(\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$, then,*

$$\bar{G}_Y(y_1, y_2) = \sum_{n=1}^{\infty} p_n \bar{F}_{MOBP}(y_1, y_2; \alpha, n\lambda_0, n\lambda_1, n\lambda_2), \tag{14}$$

where $p_n = P(N = n) = \theta(1-\theta)^{n-1}$.

Theorem 4. *Let $(Y_1, Y_2) \sim MOBPG(\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$, then the joint PDF of (Y_1, Y_1) is*

$$g_Y(y_1, y_2) = \begin{cases} g_1(y_1, y_2) & if \quad y_2 < y_1 \\ g_2(y_1, y_2) & if \quad y_1 < y_2 \\ g_0(y_1, y_2) & if \quad y_1 = y_2 = y. \end{cases}$$

where

$$g_1(y_1,y_2) = \frac{\theta\alpha^2\lambda_2(\lambda_0+\lambda_1)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)-1}(1+\alpha y_2)^{-\lambda_2-1}}{[1-(1-\theta)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}]^3} \times [1+(1-\theta)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}].$$

$$g_2(y_1,y_2) = \frac{\theta\alpha^2\lambda_1(\lambda_0+\lambda_2)(1+\alpha y_1)^{-\lambda_1-1}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)-1}}{[1-(1-\theta)(1+\alpha y_1)^{-\lambda_1)}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}]^3} \times [1+(1-\theta)(1+\alpha y_1)^{-\lambda_1}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}].$$

$$g_0(y) = \frac{\theta\alpha\lambda_0(1+\alpha y)^{-(\lambda_0+\lambda_1+\lambda_2)-1}}{[1-(1-\theta)(1+\alpha y)^{-(\lambda_0+\lambda_1+\lambda_2)}]^2}.$$

Proof: The phrases for $g_1(.,.)$ and $g_2(.,.)$ can be calculated using $-\frac{\partial^2 \bar{G}(y_1,y_2)}{\partial y_1 \partial y_2}$ for $y_2 < y_1$ and $y_1 < y_2$ respectively. But $g_0(.)$ can not be calculated in the similar method. Using the facts that

$$\int_0^\infty \int_{y_2}^\infty g_1(y_1,y_2)dy_1dy_2 + \int_0^\infty \int_{y_1}^\infty g_2(y_1,y_2)dy_2dy_1 + \int_0^\infty g_0(y)dy = 1.$$

Therefore, the result follows.

Proposition 3.2. *If $(Y_1,Y_2) \sim MOBPG(\theta,\alpha,\lambda_0,\lambda_1,\lambda_2)$, then,*

$$g_1(y_1,y_2) = \sum_{n=1}^\infty p_n f_{Pa}(y_1,\alpha,n(\lambda_0+\lambda_1)) f_{Pa}(y_2,\alpha,n\lambda_2).$$

$$g_2(y_1,y_2) = \sum_{n=1}^\infty p_n f_{Pa}(y_1,\alpha,n\lambda_1) f_{Pa}(y_2,\alpha,n(\lambda_0+\lambda_2)).$$

$$g_0(y) = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}\sum_{n=1}^\infty p_n f_{Pa}(y,\alpha,n(\lambda_0+\lambda_1+\lambda_2)).$$

where $p_n = P(N=n) = \theta(1-\theta)^{n-1}$ and $f_{Pa}(.,n\lambda)$ is the PDF of Pareto distribution with parameters α and $n\lambda$. Note that $f_{Pa}(.,n\lambda)$ is the PDF of the random variable $\min(U_1,\cdots,U_n)$ *where U_i's are independent random variables from a Pareto distribution with parameters α and λ.*

The MOBPG model has various forms. We have presented the surface plots of the PDF of MOBPG distribution in Figure 2 for various values of parameters.

Also, the joint PDF of (Y_1,Y_2) can be written as follows:

$$g(y_1,y_2) = \frac{\lambda_1+\lambda_2}{\lambda_0+\lambda_1+\lambda_2} g_a(y_1,y_2) + \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} g_s(y).$$

here

$$g_a(y_1,y_2) = \frac{\lambda_0+\lambda_1+\lambda_2}{\lambda_1+\lambda_2} \times \begin{cases} g_1(y_1,y_2) & \text{if} \quad y_2<y_1 \\ g_2(y_1,y_2) & \text{if} \quad y_1<y_2. \end{cases}$$

and

$$g_s(y) = \frac{\theta\alpha\lambda_0(1+\alpha y)^{-(\lambda_0+\lambda_1+\lambda_2)-1}}{[1-(1-\theta)(1+\alpha y)^{-(\lambda_0+\lambda_1+\lambda_2)}]^2}, \qquad \text{if} \quad y_1=y_2=y.$$

where, $g_a(.,.)$ and $g_s(.)$ are the absolute continuous part and the singular part, respectively. If $\lambda_0 = 0$, it does not have any singular part, and it becomes an absolute continuous distribution. For $\theta = 1$, it is immediate that MOBP can be obtained as a special case of MOBPG.
In the following, the joint PDF of Y_1, Y_2 and N is obtained, where (Y_1,Y_2) has the MOBPG distributions and N is the geometric distribution.

$$\begin{aligned} P(Y_1>y_1,Y_2>y_2,N=n) &= P(Y_1>y_1,Y_2>y_2|N=n)P(N=n) \\ &= \begin{cases} \theta(1-\theta)^{n-1}(1+\alpha y_1)^{-n(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-n\lambda_2} & \text{if} \quad y_2 \le y_1 \\ \theta(1-\theta)^{n-1}(1+\alpha y_1)^{-n\lambda_1}(1+\alpha y_2)^{-n(\lambda_0+\lambda_2)} & \text{if} \quad y_1<y_2. \end{cases} \end{aligned}$$

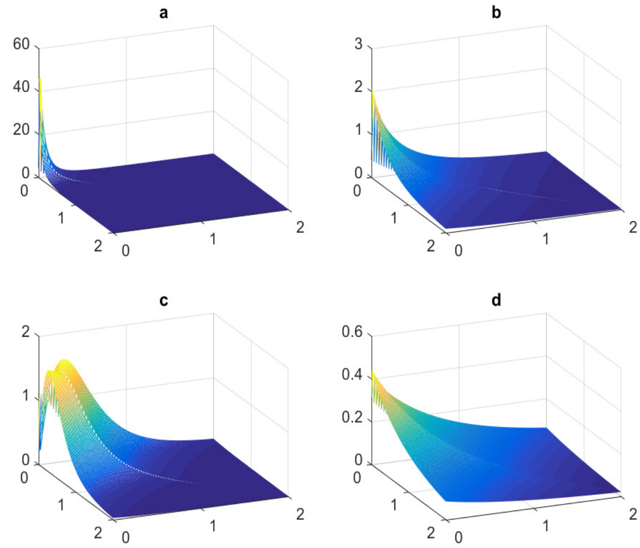

Figure 2: The shape of the joint probability density function of MOBPG for various values of parameters $\Theta = (\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$: (a) $\Theta = (0.2, 0.3, 3, 3, 3)$ (b) $\Theta = (0.5, 0.2, 1, 2, 3)$ (c) $\Theta = (2, 0.75, 3, 3, 3)$ (d) $\Theta = (1, 0.3, 1, 1.5, 2)$.

Therefore, the joint PDF of Y_1, Y_2 and N is:

$$f_{Y_1,Y_2,N}(y_1, y_2, n) = \begin{cases} \theta(1-\theta)^{n-1} f_{1n}(y_1, y_2) & \text{if} \quad y_2 < y_1 \\ \theta(1-\theta)^{n-1} f_{2n}(y_1, y_2) & \text{if} \quad y_1 < y_2 \\ \theta(1-\theta)^{n-1} f_{0n}(y) & \text{if} \quad y_1 = y_2 = y. \end{cases} \tag{15}$$

where,

$$\begin{aligned} f_{1n}(y_1, y_2) &= n^2\alpha^2\lambda_2(\lambda_0+\lambda_1)(1+\alpha y_1)^{-n(\lambda_0+\lambda_1)-1}(1+\alpha y_2)^{-n\lambda_2-1}. \\ f_{2n}(y_1, y_2) &= n^2\alpha^2\lambda_1(\lambda_0+\lambda_2)(1+\alpha y_1)^{-n\lambda_1-1}(1+\alpha y_2)^{-n(\lambda_0+\lambda_2)-1}. \\ f_{0n}(y) &= n\alpha\lambda_0(1+\alpha y)^{-n(\lambda_0+\lambda_1+\lambda_2)-1}. \end{aligned} \tag{16}$$

Also, the conditional probability mass function of N given $Y_1 = y_1$ and $Y_2 = y_2$ can be calculated as follows:

$$f_N(n|y_1, y_2) = \begin{cases} c_1(y_1, y_2) n^2 (1-\theta)^{n-1}(1+\alpha y_1)^{-(n-1)(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-(n-1)\lambda_2} & \text{if} \quad y_2 < y_1 \\ c_2(y_1, y_2) n^2 (1-\theta)^{n-1}(1+\alpha y_1)^{-(n-1)\lambda_1}(1+\alpha y_2)^{-(n-1)(\lambda_0+\lambda_2)} & \text{if} \quad y_1 < y_2 \\ c_0(y) n (1-\theta)^{n-1}(1+\alpha y)^{-(n-1)(\lambda_0+\lambda_1+\lambda_2)} & \text{if} \quad y_1 = y_2 = y. \end{cases} \tag{17}$$

where,

$$\begin{aligned} c_1(y_1, y_2) &= \frac{[1-(1-\theta)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}]^3}{1+(1-\theta)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}}. \\ c_2(y_1, y_2) &= \frac{[1-(1-\theta)(1+\alpha y_1)^{-\lambda_1}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}]^3}{1+(1-\theta)(1+\alpha y_1)^{-\lambda_1}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}}. \\ c_0(y) &= [1-(1-\theta)(1+\alpha y)^{-(\lambda_0+\lambda_1+\lambda_2)}]^2. \end{aligned}$$

We will be using the following equations for (17)

$$f_N(n|y_1, y_2) = \begin{cases} \frac{[1-\xi_1(y_1,y_2,\theta,\gamma)]^3}{[1+\xi_1(y_1,y_2,\theta,\gamma)]} n^2 \xi_1^{n-1}(y_1, y_2, \theta, \gamma) & \text{if} \quad y_2 < y_1 \\ \frac{[1-\xi_1(y_1,y_2,\theta,\gamma)]^3}{[1+\xi_2(y_1,y_2,\theta,\gamma)]} n^2 \xi_2^{n-1}(y_1, y_2, \theta, \gamma) & \text{if} \quad y_1 < y_2 \\ [1-\xi_0(y_1, y_2, \theta, \gamma)]^2 n \xi_0^{n-1}(y_1, y_2, \theta, \gamma) & \text{if} \quad y_1 = y_2 = y. \end{cases}$$

where $\gamma = (\alpha, \lambda_0, \lambda_1, \lambda_2)$ and

$$\xi_1(y_1, y_2, \theta, \gamma) = (1-\theta)(1+\alpha y_1)^{-(\lambda_0+\lambda_1)}(1+\alpha y_2)^{-\lambda_2}.$$
$$\xi_2(y_1, y_2, \theta, \gamma) = (1-\theta)(1+\alpha y_1)^{-\lambda_1}(1+\alpha y_2)^{-(\lambda_0+\lambda_2)}.$$
$$\xi_0(y_1, y_2, \theta, \gamma) = (1-\theta)(1+\alpha y)^{-(\lambda_0+\lambda_1+\lambda_2)}.$$

Also, we can compute

$$E(N|y_1, y_2) = \begin{cases} \frac{(1-\xi_1(y_1,y_2,\theta,\gamma))^2-6(1-\xi_1(y_1,y_2,\theta,\gamma))+6}{(1-\xi_1(y_1,y_2,\theta,\gamma))^2} & \text{if} \quad y_2 < y_1 \\ \frac{(1-\xi_2(y_1,y_2,\theta,\gamma))^2-6(1-\xi_2(y_1,y_2,\theta,\gamma))+6}{(1-\xi_2(y_1,y_2,\theta,\gamma))^2} & \text{if} \quad y_1 < y_2 \\ \frac{1+\xi_0(y_1,y_2,\theta,\gamma)}{1-\xi_0(y_1,y_2,\theta,\gamma)} & \text{if} \quad y_1 = y_2 = y. \end{cases}$$

Theorem 5. *Let* $(Y_1, Y_2) \sim MOBPG(\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$, *then*

(I) $Y_1 \sim UPG(\theta, \alpha, \lambda_0 + \lambda_1)$.

(II) $Y_2 \sim UPG(\theta, \alpha, \lambda_0 + \lambda_2)$.

(III) $Y = \min\{Y_1, Y_2\} \sim UPG(\theta, \alpha, \lambda_0 + \lambda_1 + \lambda_2)$.

(IV) $P(Y_1 < Y_2) = \frac{\lambda_1}{\lambda_0+\lambda_1+\lambda_2}$.

proof: The proofs of parts of 1, 2 and 3 are similar. The proof of 1 and 2 are immediate from (11) with $\bar{F}_{X_i}$ having the Pareto distribution. Also, we have

$$\begin{aligned}\bar{F}_{X_1}(y_1) &= P(\min\{U_0, U_1\} > y_1) = P(U_0 > y_1, U_1 > y_1)\\ &= \bar{F}_{Pa}(y_1, \alpha, \lambda_0)\bar{F}_{Pa}(y_1, \alpha, \lambda_1) = \bar{F}_{Pa}(y_1, \alpha, \lambda_0 + \lambda_1).\end{aligned}$$

The proof of 3 is obtained by (11). The part IV can be proven as follows:

$$\begin{aligned}P(Y_1 < Y_2) &= \sum_{n=1}^{\infty} P(Y_1 < Y_2, N = n)\\ &= \sum_{n=1}^{\infty}\int_0^{\infty}\int_{y_1}^{\infty}\theta(1-\theta)^{n-1}f_{2n}(y_1, y_2)dy_2dy_1\\ &= \sum_{n=1}^{\infty}\theta(1-\theta)^{n-1}\int_0^{\infty}\int_{y_1}^{\infty}f_{2n}(y_1, y_2)dy_2dy_1\\ &= \theta\sum_{n=1}^{\infty}(1-\theta)^{n-1}\frac{\lambda_1}{\lambda_0+\lambda_1+\lambda_2} = \frac{\lambda_1}{\lambda_0+\lambda_1+\lambda_2}.\blacksquare\end{aligned}$$

We apply the following algorithm to check the purpose.
Algorithm to generate from MOBPG:

- Generate n from $GM(\theta)$.
- Generate $\{u_{01}, \dots, u_{0n}\}$ from $Pa(\alpha, \lambda_0)$, similarly, $\{u_{11}, \dots, u_{1n}\}$ from $Pa(\alpha, \lambda_1)$ and $\{u_{21}, \dots, u_{2n}\}$ from $Pa(\alpha, \lambda_2)$.
- Obtain $x_{1k} = \min\{u_{0k}, u_{1k}\}$ and $x_{2k} = \min\{u_{0k}, u_{2k}\}$, for $k = 1, \dots, n$.
- Compute the desired (y_1, y_2) as $y_1 = \min\{x_{11}, \dots, x_{1n}\}$ and $y_2 = \min\{x_{21}, \dots, x_{2n}\}$.

4. Maximum Likelihood Estimators

In this section, the maximum likelihood estimators of the unknown parameters of the MOBPG distribution using the EM algorithm are considered. Also, the confidance intervals of the parameters are calculated.

4.1. EM Algorithm

Suppose $\{(y_{11}, y_{21}), \ldots, (y_{1m}, y_{2m})\}$ is a random sample from MOBPG with parameters $\Theta = (\theta, \alpha, \lambda_0, \lambda_1, \lambda_2)$, and we will obtain the MLE's of the unknown parameters. We introduce the following useful notation:
$I_0 = \{i : y_{1i} = y_{2i} = y_i\}$, $I_1 = \{i : y_{1i} > y_{2i}\}$ and $I_2 = \{i : y_{1i} < y_{2i}\}$. Also, $|I_0| = m_0$, $|I_1| = m_1$, $|I_2| = m_2$ and $m = m_0 + m_1 + m_2$. According to the above notation, the log-likelihood function is computed as follows:

$$\ell(\Theta) = \sum_{i\in I_0} \ln g_0(y_i) + \sum_{i\in I_1} \ln g_1(y_{1i}, y_{2i}) + \sum_{i\in I_2} \ln g_2(y_{1i}, y_{2i}). \tag{18}$$

where g_0, g_1 and g_2 are defined in Theorem 4. We will compute the MLE's of the parameters by maximizing $\ell(\Theta)$ in equation (18) with respect to the unknown parameters. Clearly, the MLEs of the unknown parameters cannot be calculated directly. They can be obtained only by solving five equations with five unknowns. We suggest to use the EM algorithm, which this algorithm is more flexible and computationally done easily. So, they are considered as a missing value problem. For this purpose, suppose that, the joint PDF of (Y_1, Y_2, N) is provided in (15). Note that,

$$(Y_1, Y_2|N) \sim MOBP(\alpha, n\lambda_0, n\lambda_1, n\lambda_2).$$

We can show the complete data as follows: $\{(y_{1i}, y_{2i}, n_i); i = 1, ..., m\}$. We maximize the conditional log-likelihood function to calculate the MLE's of unknown parameters, and it is as follows:

$$\ell_1(\alpha, \lambda_0, \lambda_1, \lambda_2) = \sum_{i\in I_0} \ln f_{0n_i}(y_i) + \sum_{i\in I_1} \ln f_{1n_i}(y_{1i}, y_{2i}) + \sum_{i\in I_2} \ln f_{2n_i}(y_{1i}, y_{2i}).$$

The functions, f_{0n_i}, f_{1n_i} and f_{2n_i} are determined in (16). We will describe this subject as a missing value problem using the EM algorithm. For this purpose, we consider the new set of random variables as follows:

$$\{U_i|N = n\} \sim Pa(\alpha, n\lambda_i), \qquad i = 0, 1, 2. \tag{19}$$

Also, we assume that they are conditionally independent. We easily know that:

$$\{Y_1|N = n\} = \min\{U_0, U_1\}|N = n, \qquad \text{and} \qquad \{Y_2|N = n\} = \min\{U_0, U_2\}|N = n.$$

Assume that for the bivariate random vector (Y_1, Y_2), there are associated random vectors

$$(\Delta_1, \Delta_2) = \begin{cases} (0,0) & \text{if} \quad Y_1 = U_0,\ Y_2 = U_0 \\ (0,1) & \text{if} \quad Y_1 = U_0,\ Y_2 = U_2 \\ (1,0) & \text{if} \quad Y_1 = U_1,\ Y_2 = U_0 \\ (1,1) & \text{if} \quad Y_1 = U_1,\ Y_2 = U_2\ . \end{cases} \tag{20}$$

Here Y_i's are the same as defined above. Therefore, a sample is obtained from $(Y_1, Y_2, \Delta_1, \Delta_2, N)$ which is the complete observation. Therefore, suppose that $(Y_1, Y_2, \Delta_1, \Delta_2, N)$ be a random sample of size m. So, we can solve a one dimensional optimization problem to calculate the estimation of the unknown parameters. It is clear that if we know (Y_1, Y_2), the associated (Δ_1, Δ_2) may not always be known. Thus, if $Y_1 = Y_2$, then, $\Delta_1 = \Delta_2 = 0$ is known. Hence, if $Y_1 \neq Y_2$, therefore, (Δ_1, Δ_2) is not known. If $(Y_1, Y_2) \in I_1$, the feasible values of (Δ_1, Δ_2) are (0,1) or (1,1), and if $(Y_1, Y_2) \in I_2$, the feasible values of (Δ_1, Δ_2) are (1,0) or (1,1), with positive probabilities, see for example [17]. We discuss the 'pseudo' log-likelihood function for performance of the EM algorithm. For this purpose, we use the proposed structure of [9] or [16]. In this method, the conditional 'pseudo' log-likelihood function is formed by conditioning on N, and N is replaced by $E(N|Y_1, Y_2)$.

In the 'E' step, if the observations belong to I_0, then, we kept the log-likelihood contribution intact and their related (Δ_1, Δ_2) are completely known. Also, if the observations depend to I_1 or I_2, then they are treated as missing observations.
If $(y_1, y_2) \in I_1$, the 'pseudo observation' is formed by fractioning (y_1, y_2) to two partially complete 'pseudo observations' of the form $(y_1, y_2, u_1(\Theta))$ and $(y_1, y_2, u_2(\Theta))$. The fractional mass $u_1(\Theta)$ and $u_2(\Theta)$ assigned to the 'pseudo observation' are the conditional probabilities that (Δ_1, Δ_2) takes values (0,1) or (1,1), respectively, given that $(Y_1, Y_2) \in I_1$.
Similarly, if $(Y_1, Y_2) \in I_2$, 'pseudo observations' are formed as $(y_1, y_2, v_1(\Theta))$ and $(y_1, y_2, v_2(\Theta))$, where $v_1(\Theta)$ and $v_2(\Theta)$ are the conditional probabilities that (Δ_1, Δ_2) takes values (1,0) and (1,1), respectively. The information in Table 1 will be applied to build 'E' step of the EM algorithm.

Table 1: All possible cases of U_0, U_1, U_2, corresponding probabilities and (Δ_1, Δ_2).

Different Case	probability	(Δ_1, Δ_2)	$Y_1 \& Y_2$	Set
$U_0 < U_1 < U_2$	$\frac{\lambda_1\lambda_0}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	(0,0)	$Y_1 = Y_2$	I_0
$U_0 < U_2 < U_1$	$\frac{\lambda_2\lambda_0}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	(0,0)	$Y_1 = Y_2$	I_0
$U_1 < U_0 < U_2$	$\frac{\lambda_1\lambda_0}{(\lambda_0+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	(1,0)	$Y_1 < Y_2$	I_2
$U_1 < U_2 < U_0$	$\frac{\lambda_1\lambda_2}{(\lambda_0+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	(1,1)	$Y_1 < Y_2$	I_2
$U_2 < U_0 < U_1$	$\frac{\lambda_0\lambda_2}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+\lambda_2)}$	(0,1)	$Y_2 < Y_1$	I_1
$U_2 < U_1 < U_0$	$\frac{\lambda_1\lambda_2}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+\lambda_2)}$	(1,1)	$Y_2 < Y_1$	I_1

Therefore,

$$v_1(\Theta) = \frac{\lambda_0}{\lambda_0 + \lambda_2}, \qquad v_2(\Theta) = \frac{\lambda_2}{\lambda_0 + \lambda_2}, \qquad u_1(\Theta) = \frac{\lambda_0}{\lambda_0 + \lambda_1}, \qquad u_2(\Theta) = \frac{\lambda_1}{\lambda_0 + \lambda_1}.$$

In the following, we consider the 'pseudo' log-likelihood contribution of an observation $y \in I_0$, with conditioning on $N = n$. Then, the log-likelihood contribution is:

$$\ln n + \ln \lambda_0 - n(\lambda_0 + \lambda_1 + \lambda_2) \ln(1 + \alpha y) + \ln \alpha - \ln(1 + \alpha y).$$

The above equation follows from (19), and in this case, $U_0 = y$, $U_1 > y$ and $U_2 > y$.
Also, when $(y_1, y_2) \in I_1$ and $(y_1, y_2) \in I_2$, the 'pseudo' log-likelihood contribution, conditioning on $N = n$, becomes:

$$2 \ln n + 2 \ln \alpha + \ln \lambda_2 + u_1 \ln \lambda_0 + u_2 \ln \lambda_1 - (n\lambda_2 + 1) \ln(1 + \alpha y_2) - (n(\lambda_0 + \lambda_1) + 1) \ln(1 + \alpha y_1).$$

and

$$2 \ln n + 2 \ln \alpha + \ln \lambda_1 + v_1 \ln \lambda_0 + v_2 \ln \lambda_2 - (n\lambda_1 + 1) \ln(1 + \alpha y_1) - (n(\lambda_0 + \lambda_2) + 1) \ln(1 + \alpha y_2).$$

respectively. We know that, $\ln n$ and n are missing. So, we compute the 'pseudo' log-likelihood function at the E-step of the EM algorithm by replacing them with $E(\ln N|y_1, y_2)$ and $E(N|y_1, y_2)$, respectively. Also, we will use the following notations in the k-th step of the EM algorithm.

- $\Theta^{(k)} = (\theta^{(k)}, \alpha^{(k)}, \lambda_0^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)})$ is defined for the estimates of the parameters in the k-th step.
- $E(N|y_{1i}, y_{2i}, \Theta) = a_i$.
- $E(N|y_{1i}, y_{2i}, \Theta^{(k)}) = a_i^{(k)}$.
- $u_1(\Theta^{(k)}) = u_1^{(k)}$, $u_2(\Theta^{(k)}) = u_2^{(k)}$, $v_1(\Theta^{(k)}) = u_1^{(k)}$ and $v_2(\Theta^{(k)}) = u_2^{(k)}$.

Now we can discuss the EM algorithm.
E-Step: At the k-step of the EM algorithm, we can write the 'pseudo' log-likelihood function without the additive constant as follows:

$$\begin{aligned}
\ell_{pseudo}(\Theta) &= (m_0 + 2m_1 + 2m_2) \ln \lambda_0 + (m_2 + m_1 u_2^{(k)}) \ln \lambda_1 + (m_2 v_2^{(k)} + m_1) \ln \lambda_2 \\
&- \lambda_0 \{\sum_{i \in I_0} a_i^{(k)} \ln(1 + \alpha y_i) + \sum_{i \in I_2} a_i^{(k)} \ln(1 + \alpha y_{2i}) + \sum_{i \in I_1} a_i^{(k)} \ln(1 + \alpha y_{1i})\} \\
&- \lambda_1 \{\sum_{i \in I_0} a_i^{(k)} \ln(1 + \alpha y_i) + \sum_{i \in I_2} a_i^{(k)} \ln(1 + \alpha y_{1i}) + \sum_{i \in I_1} a_i^{(k)} \ln(1 + \alpha y_{1i})\} \\
&- \lambda_2 \{\sum_{i \in I_0} a_i^{(k)} \ln(1 + \alpha y_i) + \sum_{i \in I_2} a_i^{(k)} \ln(1 + \alpha y_{2i}) + \sum_{i \in I_1} a_i^{(k)} \ln(1 + \alpha y_{2i})\} \\
&- \{\sum_{i \in I_0} \ln(1 + \alpha y_i) + \sum_{i \in I_2} \ln(1 + \alpha y_{1i}) + \sum_{i \in I_2} \ln(1 + \alpha y_{2i}) + \sum_{i \in I_1} \ln(1 + \alpha y_{2i}) + \sum_{i \in I_1} \ln(1 + \alpha y_{1i})\} \\
&+ (m_0 + 2m_1 + 2m_2) \ln \alpha + m \ln \frac{\theta}{1 - \theta} + \ln(1 - \theta) \sum_{i=1}^{m} a_i^{(k)}. \qquad (21)
\end{aligned}$$

M-Step: In the 'M'-step, we maximize $\ell_{pseudo}(\Theta)$ with respect to the unknown parameters. Also for fixed α, This maximization is determined according to the unknown parameters as follows:

$$\widehat{\lambda}_0(\alpha) = \frac{m_0 + m_2 v_1^{(k)} + m_1 u_1^{(k)}}{\sum_{i\in I_0} a_i^{(k)} \ln(1+\alpha y_i) + \sum_{i\in I_2} a_i^{(k)} \ln(1+\alpha y_{2i}) + \sum_{i\in I_1} a_i^{(k)} \ln(1+\alpha y_{1i})}. \tag{22}$$

$$\widehat{\lambda}_1(\alpha) = \frac{m_2 + m_1 u_2^{(k)}}{\sum_{i\in I_0} a_i^{(k)} \ln(1+\alpha y_i) + \sum_{i\in I_2} a_i^{(k)} \ln(1+\alpha y_{1i}) + \sum_{i\in I_1} a_i^{(k)} \ln(1+\alpha y_{1i})}. \tag{23}$$

$$\widehat{\lambda}_2(\alpha) = \frac{m_1 + m_2 v_2^{(k)}}{\sum_{i\in I_0} a_i^{(k)} \ln(1+\alpha y_i) + \sum_{i\in I_2} a_i^{(k)} \ln(1+\alpha y_{2i}) + \sum_{i\in I_1} a_i^{(k)} \ln(1+\alpha y_{2i})}. \tag{24}$$

and $\hat{\theta}$ is:

$$\hat{\theta} = \frac{m}{\sum_{i=1}^{m} a_i^{(k)}}. \tag{25}$$

We can compute the pseudo-maximum likelihood estimate of α through maximization the pseudo-profile log-likelihood function $\ell_{pseudo}(\widehat{\theta}, \alpha, \widehat{\lambda}_0(\alpha), \widehat{\lambda}_1(\alpha), \widehat{\lambda}_2(\alpha))$. Another method is to solve the following non-linear equation,

$$g(\alpha) = \alpha. \tag{26}$$

where $g(\alpha) = \frac{m_0+2m_1+2m_2}{h(\alpha)}$, and

$$\begin{aligned}
h(\alpha) = {} & \hat{\lambda}_0(\alpha)\{\sum_{i\in I_0} a_i^{(k)} \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i^{(k)} \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} a_i^{(k)} \frac{y_{1i}}{1+\alpha y_{1i}}\} \\
& + \hat{\lambda}_1(\alpha)\{\sum_{i\in I_0} a_i^{(k)} \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i^{(k)} \frac{y_{1i}}{1+\alpha y_{1i}} + \sum_{i\in I_1} a_i^{(k)} \frac{y_{1i}}{1+\alpha y_{1i}}\} \\
& + \hat{\lambda}_2(\alpha)\{\sum_{i\in I_0} a_i^{(k)} \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i^{(k)} \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} a_i^{(k)} \frac{y_{2i}}{1+\alpha y_{2i}}\} \\
& - \{\sum_{i\in I_0} \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} \frac{y_{1i}}{1+\alpha y_{1i}} + \sum_{i\in I_2} \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} \frac{y_{1i}}{1+\alpha y_{1i}}\}.
\end{aligned}$$

Thus, we can apply an easy, iterative process to solve the equation (26). For example, the Newton - Raphson method or the proposed method by [17] and [18] can be used. Now, we use the following steps to obtain the estimation of parameters by the EM algorithm: **ALGORITHM**

- Step 1: Take some initial value of Θ, say $\Theta^{(0)} = (\theta^{(0)}, \alpha^{(0)}, \lambda_0^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)})$.
- Step 2: Compute $a_i^{(0)} = E(N|y_{1i}, y_{2i}; \Theta^{(0)})$.
- Step 3: Compute u_1, u_2, v_1, and v_2.
- Step 4: Find $\hat{\alpha}$ by solving equation (26), say $\hat{\alpha}^{(1)}$.
- Step 5: Compute $\hat{\lambda}_i^{(1)} = \hat{\lambda}_i(\hat{\alpha}^{(1)})$, $i = 0, 1, 2$ from (22)-(24).
- Step 6: Find $\hat{\theta}$ from (25).
- Step 7: Replace $\Theta^{(0)}$ by $\Theta^{(1)} = (\theta^{(1)}, \alpha^{(1)}, \lambda_0^{(1)}, \lambda_1^{(1)}, \lambda_2^{(1)})$, go back to step 1 and continue the process until convergence takes place.

4.2. Confidence Interval

For constructing confidence intervals, the observed Fisher Information Matrix obtained from the EM algorithm using the method of [21] is applied. Using a similar equation in [21], the Fisher Information Matrix can be computed as follows:

$$I_{obs} = B - SS^T.$$

where matrix B defines the Hessian matrix and the vector S obtains the gradient vector of the pseudo-log-likelihood function. We provide the elements of matrix B and S as follows:

$$\begin{aligned}
B_{11} &= \frac{m_0 + 2m_1 + 2m_2}{\hat{\alpha}^2} + \hat{\lambda}_0\{\sum_{i\in I_0} a_i \frac{y_i^2}{(1+\alpha y_i)^2} + \sum_{i\in I_2} a_i \frac{y_{2i}^2}{(1+\alpha y_{2i})^2} + \sum_{i\in I_1} a_i \frac{y_{1i}^2}{(1+\alpha y_{1i})^2}\} \\
&+ \hat{\lambda}_1\{\sum_{i\in I_0} a_i \frac{y_i^2}{(1+\alpha y_i)^2} + \sum_{i\in I_2} a_i \frac{y_{1i}^2}{(1+\alpha y_{1i})^2} + \sum_{i\in I_1} a_i \frac{y_{1i}^2}{(1+\alpha y_{1i})^2}\} \\
&+ \hat{\lambda}_2\{\sum_{i\in I_0} a_i \frac{y_i^2}{(1+\alpha y_i)^2} + \sum_{i\in I_2} a_i \frac{y_{2i}^2}{(1+\alpha y_{2i})^2} + \sum_{i\in I_1} a_i \frac{y_{2i}^2}{(1+\alpha y_{2i})^2}\} \\
&- \{\sum_{i\in I_0} \frac{y_i^2}{(1+\alpha y_i)^2} + \sum_{i\in I_2} \frac{y_{1i}^2}{(1+\alpha y_{1i})^2} + \sum_{i\in I_2} \frac{y_{2i}^2}{(1+\alpha y_{2i})^2} + \sum_{i\in I_1} \frac{y_{2i}^2}{(1+\alpha y_{2i})^2} + \sum_{i\in I_1} \frac{y_{1i}^2}{(1+\alpha y_{1i})^2}. \\
B_{12} &= B_{21} = \sum_{i\in I_0} a_i \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} a_i \frac{y_{1i}}{1+\alpha y_{1i}}. \\
B_{13} &= B_{31} = \sum_{i\in I_0} a_i \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i \frac{y_{1i}}{1+\alpha y_{1i}} + \sum_{i\in I_1} a_i \frac{y_{1i}}{1+\alpha y_{1i}}. \\
B_{14} &= B_{41} = \sum_{i\in I_0} a_i \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} a_i \frac{y_{2i}}{1+\alpha y_{2i}}. \\
B_{15} &= B_{51} = 0. \\
B_{22} &= \frac{m_0 + m_2 v_1 + m_1 u_1}{\hat{\lambda}_0^2}, \qquad B_{23} = B_{32} = 0, \qquad B_{24} = B_{42} = 0, \qquad B_{25} = B_{52} = 0. \\
B_{33} &= \frac{m_2 + m_1 u_2}{\hat{\lambda}_1^2}, \qquad B_{34} = B_{43} = 0, \qquad B_{35} = B_{53} = 0. \\
B_{44} &= \frac{m_2 v_2 + m1}{\hat{\lambda}_2^2}, \qquad B_{45} = B_{54} = 0. \\
S_1 &= \frac{m_0 + 2m_1 + 2m_2}{\hat{\alpha}} - \hat{\lambda}_0\{\sum_{i\in I_0} a_i \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} a_i \frac{y_{1i}}{1+\alpha y_{1i}}\} \\
&+ \hat{\lambda}_1\{\sum_{i\in I_0} a_i \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i \frac{y_{1i}}{1+\alpha y_{1i}} + \sum_{i\in I_1} a_i \frac{y_{1i}}{1+\alpha y_{1i}}\} \\
&+ \hat{\lambda}_2\{\sum_{i\in I_0} a_i \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} a_i \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} a_i \frac{y_{2i}}{1+\alpha y_{2i}}\} \\
&- \{\sum_{i\in I_0} \frac{y_i}{1+\alpha y_i} + \sum_{i\in I_2} \frac{y_{1i}}{1+\alpha y_{1i}} + \sum_{i\in I_2} \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} \frac{y_{2i}}{1+\alpha y_{2i}} + \sum_{i\in I_1} \frac{y_{1i}}{1+\alpha y_{1i}}\}.
\end{aligned}$$

$$\begin{aligned}
S_2 &= \frac{m_0 + m_2 v_1 + m_1 u_1}{\hat{\lambda}_0} - \{\sum_{i\in I_0} a_i \ln(1+\alpha y_i) + \sum_{i\in I_2} a_i \ln(1+\alpha y_{2i}) + \sum_{i\in I_1} a_i \ln(1+\alpha y_{1i})\}. \\
S_3 &= \frac{m_2 + m_1 u_2}{\hat{\lambda}_1} - \{\sum_{i\in I_0} a_i \ln(1+\alpha y_i) + \sum_{i\in I_2} a_i \ln(1+\alpha y_{1i}) + \sum_{i\in I_1} a_i \ln(1+\alpha y_{1i})\}. \\
S_4 &= \frac{m_1 + m_2 v_2}{\hat{\lambda}_2} - \{\sum_{i\in I_0} a_i \ln(1+\alpha y_i) + \sum_{i\in I_2} a_i \ln(1+\alpha y_{2i}) + \sum_{i\in I_1} a_i \ln(1+\alpha y_{2i})\}. \\
S_5 &= \frac{m}{\hat{\theta}(1-\hat{\theta})} - \frac{\sum_{i=1}^m a_i}{(1-\hat{\theta})}.
\end{aligned}$$

Now, the asymptotic normality results are explained to create the asymptotic confidence intervals of $\alpha, \lambda_0, \lambda_1, \lambda_2$ and θ. It can be presented as follows:

$$\sqrt{n}[(\hat{\theta}-\theta), (\hat{\alpha}-\alpha), (\hat{\lambda}_0-\lambda_0), (\hat{\lambda}_1-\lambda_1), (\hat{\lambda}_2-\lambda_2)] \to N_5(0, I^{-1}(\Theta)), \qquad as \qquad n \to \infty. \tag{27}$$

Where $I^{-1}(\Theta)$ is the variance-covariance matrix, $\Theta = (\hat{\theta}, \hat{\alpha}, \hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_1, \hat{\lambda}_2)$ and $\Theta = (\theta, \alpha, \lambda_0, \lambda_1, \lambda_1, \lambda_2)$. Since Θ is unknown in Equation (27), $I^{-1}(\Theta)$ is estimated by $I^{-1}(\hat{\Theta})$; the asymptotic variance-covariance matrix that is defined above and this can be used to obtain the asymptotic confidence intervals of $\alpha, \lambda_0, \lambda_1, \lambda_2$ and θ.

5. Data analysis and comparison study

In this section, some results based on Monte Carlo simulations and also one real data set in order to evaluate the new suggested model and the performance of EM algorithm are obtained. We use Matlab software for data analysis.

5.1. Numerical experiments

In this subsection, several simulation experiments are performed to test the efficiency of the EM algorithm for various sample sizes and parameter values.

To calculate the estimation of the unknown parameters, we apply the EM algorithm as proposed in Section 3 and thus the BPG model is fitted to the simulated data set. We used three sets of parameter values $\Theta_i = (\theta_i, \alpha, \lambda_0, \lambda_1, \lambda_2) = (\theta_i, 3, 1, 1, 1)$, where $\theta_i = 0.3, 0.5, 0.7$ for $i = 1, 2, 3$, respectively. For each situation, the initial value for the EM algorithm is $\theta_i = 0.2, 0.4, 0.6$ for $i = 1, 2, 3$, $\alpha = \lambda_0 = \lambda_1 = \lambda_2 = 1$. We stop iteration when the absolute value of the difference of the two consecutive iterates for all the five parameters are less than 10^{-5}. We repeat the procedure 1000 times, and compute the Bias estimations, the related mean squared errors (MSEs), the average confidence lengths and the coverage percentages.

Some of the points are really obvious from the simulation results. In all situations, we observe that the biases and the mean square errors decrease with increasing the size of the sample, which verifies the consistency properties of the MLEs. The results are presented in Table 2. In Table 3, we presented the average confidence lengths and the corresponding coverage percentages. The nominal level for the confidence intervals is 0.95 in each case. From Table 3, it is evident that as the sample size increases, the average confidence lengths decrease and the corresponding coverage percentage increase.

Table 2: The Bias estimates and the associated mean squared errors (MSEs).

Θ	n	α		λ_0		λ_1		λ_2		θ	
		$\|Bias\|$	MSE	$\|Bias\|$	MSE	$\|Bias\|$	MSE	$\|Bias\|$	MSE	$\|Bias\|$	MSE
	40	0.3320	0.2084	0.1009	0.4375	0.1083	0.4347	0.1095	0.4054	0.1204	0.0150
Θ_1	80	0.1920	0.1448	0.0296	0.2366	0.0216	0.2780	0.0301	0.2706	0.1173	0.0135
	120	0.1559	0.1074	0.0151	0.1186	0.0190	0.1311	0.0269	0.1314	0.1072	0.0115
	40	0.3319	0.2396	0.0661	0.3335	0.0823	0.4199	0.0946	0.4201	0.1252	0.0163
Θ_2	80	0.1084	0.1834	0.0279	0.1196	0.0345	0.1451	0.0293	0.1406	0.1239	0.0155
	120	0.0816	0.0968	0.0199	0.0741	0.0179	0.0832	0.0197	0.0735	0.1154	0.0105
	40	0.3961	0.1578	0.0989	0.2307	0.1076	0.3217	0.1019	0.2383	0.1171	0.0162
Θ_3	80	0.2352	0.1022	0.0229	0.0690	0.0286	0.0853	0.0231	0.1024	0.1149	0.0159
	120	0.1441	0.0851	0.0168	0.0342	0.0172	0.0417	0.0162	0.0425	0.1057	0.0150

5.2. Real data set

We discuss the MOBPG distribution for fitting the one real data set. We apply the proposed EM algorithm and then the estimation of parameters and their corresponding log-likelihood values are computed. Also, we obtain the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). We also compute the Kolmogorov-Smirnov(K-S) distances between the fitted distribution, the empirical distribution function and the corresponding p-values for Y_1, Y_2 and $\min\{Y_1, Y_2\}$. Finally, we make use the likelihood ratio test (LRT) and the corresponding p-values to choose a better model.

Data set: This data set contains the indemnity payments (Loss) and allocated loss adjustment expense (ALAE) relating to 50 general liability claims from an insurance company in Iran. This data set is reported in Table 4. Before analyzing the data we divide all data by 10^8.

Table 3: The average confidence lengths and the corresponding coverage percentages.

Θ	n	α		λ_0		λ_1		λ_2		θ	
		Length	CP	Length	CP	Length	CP	Length	CP	Length	CP
	40	0.9130	0.8890	1.0228	0.8920	1.1186	0.9250	1.0205	0.9000	0.4898	0.9110
Θ_1	80	0.7610	0.9210	0.9728	0.9200	1.0545	0.9400	1.0138	0.9380	0.3727	0.9420
	120	0.6554	0.9510	0.6887	0.9490	0.7148	0.9560	0.8983	0.9590	0.2061	0.9600
	40	0.9724	0.8950	1.0547	0.9290	1.0938	0.9150	1.0179	0.9020	0.5068	0.9000
Θ_2	80	0.8045	0.9230	0.8016	0.9400	0.8944	0.9290	0.8437	0.9200	0.4139	0.9200
	120	0.6944	0.9500	0.6305	0.9580	0.6857	0.9300	0.6276	0.9450	0.3587	0.9420
	40	0.7937	0.9120	0.9918	0.9090	1.0398	0.9250	0.9771	0.9280	0.5014	0.9300
Θ_3	80	0.7069	0.9260	0.5864	0.9420	0.6043	0.9430	0.6450	0.9340	0.4906	0.9510
	120	0.5832	0.9580	0.3913	0.9600	0.4184	0.9590	0.4166	0.9520	0.4089	0.9590

Table 4: The real data set.

S.N.	Y_1	Y_2	S.N.	Y_1	Y_2	S.N.	Y_1	Y_2
1	3406788	386367	18	2559763	1129319	35	49732966	5084502
2	366797	16546499	19	48271944	4513586	36	12272011	3210349
3	9113373	9046402	20	7162599	7416202	37	2475933	182092
4	5816565	2469313	21	7402695	9107311	38	6132238	16562877
5	511398	767917	22	13450355	3479296	39	49487877	35107066
6	1175114	33849400	23	94415944	271460	40	17903766	219030
7	719650	3854188	24	4662978	8338641	41	8740014	18542911
8	7594948	10395944	25	1664301	1664301	42	24558000	1987411
9	10121699	7405536	26	2564265	1266798	43	10324500	4367226
10	3000947	3000947	27	13324122	22958511	44	2797877	1975131
11	11806400	2110433	28	728146	1794489	45	63904333	2915554
12	164582	365291	29	8366643	9129838	46	2257089	4657025
13	4513141	5872939	30	654520	654520	47	13181611	4384871
14	3412786	10322111	31	9825560	1102267	48	20017400	7944262
15	1958098	1719298	32	2869349	1098717	49	27989044	156936
16	761409	1684610	33	1658672	396442	50	4418335	2440770
17	1833065	4397685	34	1244447	2928196			

Before going to analyze the data by MOBPG distribution, the Pareto distribution to Y_1, Y_2 and $\min\{Y_1, Y_2\}$ are fitted, separately. In addition to the Pareto distribution, we fit exponential and Weibull distributions. It will help us to investigate the recent various models and compute the initial values also. The estimation of parameters of the ML method, the corresponding Kolmogorov-Smirnov distances (K-S) and the associated p-values are computed. The results are presented in the Table 5. Based on the p-values the exponential, Weibull and Pareto distribution cannot be rejected for the marginals and for the minimum also. We see that the Pareto distribution has a better fit than two other distributions.

Table 5: The MLE's of parameters, the Kolmogorov-Smirnov (K-S) and the associated p-values for real data set.

Models		$\hat{\alpha}$	$\hat{\lambda}$	K-S	P-value
	Y_1	-	9.2878	0.1793	0.0761
Exponential	Y_2	-	16.6012	0.1282	0.3418
	$\min\{Y_1, Y_2\}$	-	25.9384	0.1121	0.5076
	Y_1	0.7565	5.7325	0.1081	0.5532
Weibull	Y_2	0.8450	11.6456	0.0846	0.8284
	$\min\{Y_1, Y_2\}$	0.8796	18.6737	0.0655	0.9506
	Y_1	8.1545	1.9218	0.0748	0.9172
Pareto	Y_2	7.0243	3.3170	0.0698	0.9503
	$\min\{Y_1, Y_2\}$	7.7711	4.3571	0.0591	0.9896

Now we will fit the MOBPG model. Also, for a better comparison of models, we fit the Marshall-Olkin bivariate Weibull-geometric (MOBWG) distribution and the Marshall-Olkin bivariate exponential-geometric (MOBEG) distribution. We fit these models by the suggested EM algorithm to the bivariate data set and obtain the estimation of parameters and their log-likelihood values. For each fitted model, the Akaike Information Criterion (AIC) and the Bayesian information criterion (BIC) are computed. The results are given in Table 6.

For further study of these models, the Kolmogorov-Smirnov (K-S) distances and the corresponding p-values for three random variables Y_1, Y_2 and $\min\{Y_1, Y_2\}$ are computed. We present these results in Table 7. From these

Table 6: The MLE's of parameters, the corresponding log-likelihood, AIC and BIC for real data set.

Model	$\hat{\alpha}$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\theta}$	$\log(l)$	AIC	BIC
MOBEG	-	1.0918	6.8670	14.3191	0.6355	133.6083	-259.2166	-251.5685
MOBWG	0.8160	0.8931	5.1817	9.2854	0.4045	135.4399	-260.8798	-251.3197
MOBPG	1.5893	0.5906	3.5326	6.1979	0.4238	144.4461	-278.8922	-269.3321

results, it is seen that using the Marshal-Olkin bivariate Pareto-geometric model is more suitable for this real data set.

Table 7: The Kolmogorov-Smirnov (K-S) distances and the corresponding p-values for three random variables Y_1, Y_2 and $\min\{Y_1, Y_2\}$for real data set.

Model	Y_1		Y_2		$\min\{Y_1, Y_2\}$	
	K-S	P-value	K-S	P-value	K-S	P-value
MOBEG	0.1216	0.4051	0.0667	0.9359	0.0569	0.9632
MOBWG	0.1752	0.0770	0.2426	0.0040	0.2643	0.0012
MOBPG	0.0863	0.8110	0.0713	0.9315	0.0846	0.8292

Finally, the likelihood ratio test (LRT) and the corresponding p-values are obtained for testing the Marshall-Olkin bivariate (MOB) models by the exponential, Weibull and Pareto distributions against the Marshall-Olkin bivariate models compounding with the geometric distribution. On the other hand, for example, our purpose is to test the null hypothesis H_0 : MOBP against the alternative hypothesis H_1 : MOBPG. Also, in the same way for Weibull and exponential distributions. We report the statistics and the corresponding p-values in Table 8. Hence, for any usual significance level, we reject proposed models in H_0 in favor of the alternative models.

Table 8: The log-likelihood, AIC, BIC, LRT and the corresponding p-values for different models.

Test	Models					
	MOBEG	MOBE	MOBWG	MOBW	MOBPG	MOBP
AIC	-259.2166	-220.6672	-260.8798	-225.8360	-278.8922	-229.5176
BIC	-251.5685	-214.9311	-251.3197	-218.1879	-269.3321	-221.8695
$\log(l)$	133.6083	113.3336	135.4399	116.9180	144.4461	118.7588
LRT	40.5494		37.0438		51.3746	
P-value	1.9171×10^{-10}		1.1551×10^{-9}		7.6317×10^{-12}	

6. Conclusions

In this paper, we discussed two bivariate distributions based on the Pareto distribution and used the proposed methods of [22] in the bivariate case and [24] in the univariate cases. In the second case, their method was generalized to bivariate case and a new bivariate distribution was introduced. In fact, this bivariate distribution was obtained by compounding geometric distribution and bivariate Pareto distribution. We called these new distributions the Marshall-Olkin bivariate Pareto (MOBP) distribution and bivariate Pareto-geometric (MOBPG) distribution, respectively. It was also stated that the MOBP distribution can be computed as a particular expression of the MOBPG distribution. Then, several properties of these distributions were established. The estimations of unknown parameters were calculated using the maximum likelihood method. However, We observed that we cannot directly solve the related log likelihood equations. Thus, the maximum likelihood estimation was numerically computed through the associated nonlinear equation using the EM algorithm. Therefore, we proposed that the EM algorithm was applied to calculate the estimation of the unknown parameters by the ML method. It was also shown that the proposed EM algorithm has a desirable performance. Also, this new model was exactly suitable for data analysis.

Acknowledgements The authors are highly grateful to the referees and editors for their suggestions which led to the improvement in the quality of this article.

References

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

[2] A. Al-Khedhairi, A. El-Gohary, A new class of bivariate Gompertz distributions and its mixture, International Journal of Mathematical Analysis, 2 (2008) 235-253.

[3] A.V. Asimit, E. Furman, R. Vernic, On a multivariate Pareto distribution, Insurance: Mathematics and Economics, 2 (2010) 308–316.

[4] A.V. Asimit, E. Furman, R. Vernic, Statistical inference for a new class of multivariate Pareto distributions, Communications in Statistics-Simulation and Computation, 2, (2016) 456-471.

[5] S.F. Bagheri, E. Bahrami Samani, M. Ganjali, The generalized modified Weibull power series distribution: Theory and applications, Computational Statistics and Data Analysis, 94 (2016) 136-160.

[6] N. Balakrishnan, C.D. Lai, Continuous bivariate distributions, New York, Springer, 2 2009.

[7] R.E. Barlow, F. Proschan, Statistical Theory of Reliability and Life Testing, Probability Models, Maryland, Silver Spring, 1981.

[8] W. Barreto-Souza, Bivariate gamma-geometric law and its induce Levy process, Journal of Multivariate Analysis, 109 (2012) 130-145.

[9] G. Dinse, Non-parametric estimation of partially incomplete time and types of failure data, Biometrics, 38 (1982) 417-431.

[10] M. Chahkandi, M. Ganjali, On some lifetime distributions with decreasing failure rate, Computational Statistics and Data Analysis, 53 (2009) 4433-4440.

[11] M. Ghitany, E. Al-Hussaini, R. Al-Jarallah, Marshall-Olkin extended Weibull distribution and its application to censored data, Journal of Applied Statistics, 32 (2005) 1025-1034.

[12] M. Ghitany, F. Al-Awadhi, L. Alkhalfan, Marshall-Olkin extended lomax distribution and its application to censored data, Communications in Statistics - Theory and Methods, 36 (2007) 1855-1866.

[13] N.L. Johnson, S. Kotz, A vector of multivariate hazard rate, Journal of Multivariate Analysis, 5 (1975) 53-66.

[14] R.A. Johnson, D.W. Wiechern, Applied Multivariate Statistical Analysis, New Jersey, Prentice Hall, 1992.

[15] S. Kotz, N. Balakrishnan, N.L. Johnson, Continuous multivariate distributions, New York, John Wiley and Sons, 2000.

[16] D. Kundu, Parameter estimation for partially complete time and type of failure data, Biometrical Journal, 46 (2004) 165–179.

[17] D. Kundu, A. Dey, Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm, Computational Statistics and Data Analysis, 35 (2009) 956–965.

[18] D. Kundu, R.D. Gupta, Estimation of $R = P(Y < X)$ for Weibull distribution, IEEE Transactions on Reliability, 55 (2006) 270–280.

[19] D. Kundu, R.D. Gupta, Modified Sarhan-Balakrishnan singular bivariate distribution, Journal of Statistical Planning and Inference, 140 (2010) 526–538.

[20] E.L. Lehmann, Some concepts of dependence, Annals of Mathematical Statistics, 37 (1966) 1137–1153.

[21] T.A. Louis, Finding the observed information matrix when using the EM algorithm, Journal of the Royal Statistical Society, 44 (1982) 226–233.

[22] A.W. Marshall, I. Olkin, A multivariate exponential distribution, Journal of the American Statistical Association, 62 (1967) 30–44.

[23] A.W. Marshall, I. Olkin, Families of multivariate distributions, Journal of the American Statistical Association, 83 (1988) 834–841.

[24] A.W. Marshall, I. Olkin, A new method of adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika, 84 (1997) 641–652.

[25] S.G. Meintanis, Test of fit for Marshall-Olkin distributions with applications, Journal of Statistical Planning and inference, 137 (2007) 3954–3963.

[26] H. Pham, C.D. Lai, On recent generalizations of the Weibull distribution, IEEE Transactions on Reliability, 56 (2007) 454–458.

[27] A.M. Sarhan, N. Balakrishnan, A new class of bivariate distributions and its mixture, Journal of Multivariate Analysis, 98 (2007) 1508–1527.

[28] R.B. Silva, M. Bourguignon, C.R.B. Dias, G.M. Cordeiro, The compound class of extended Weibull power series distributions, Computational Statistics and Data Analysis, 58 (2013) 352–367.

[29] P. Veenus, KRM. Nair, Characterization of a bivariate Pareto distribution, Journal of Indian Statistical Association, 32 (1994) 15–20.

Please cite this article using:

Shirin Shoaee, Esmaeil Khorram, A generalization of Marshall-Olkin bivariate Pareto model and its applications in shock and competing risk models, AUT J. Math. Com., 1(1) (2020) 69-87
DOI: 10.22060/ajmc.2018.14869.1012