Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/121129

Bivariate Discrete Burr Lifetime Distribution: A Mathematical and Statistical Framework for Modeling Medical and Engineering Data

Mohamed El-Dawoody[1,*] *and Mohamed S. Eliwa*[2,3]

[1]Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[2]Department of Statistics and Operation Research, College of Science,Qassim University, Buraydah 51482, Saudi Arabia
[3]Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

Received: 3 Oct. 2023, Revised: 26 Oct. 2023, Accepted: 30 Oct. 2023
Published online: 1 Nov. 2023

Abstract: The central role of modeling discrete bivariate data in enhancing understanding, facilitating informed decision-making, and advancing knowledge spans various fields. This modeling allows the depiction of the intricate relationship between two variables and finds applications in diverse domains. The focus of this study is on introducing a novel statistical model, specifically the bivariate discrete Burr distribution, an unexplored entity in existing statistical literature. This model is presented as the discrete counterpart of the Burr distribution, and we explore its essential statistical characteristics. This exploration includes the derivation of the joint probability mass function, joint survival function, joint hazard rate function along with its reversed counterpart, conditional expectations, and positive quadrant dependence. For parameter estimation of the model, maximum likelihood estimation is employed. Additionally, an extensive simulation study is conducted to evaluate the bias and mean square error of the maximum likelihood estimators. Finally, two real-world datasets are examined to demonstrate the practical applicability of the model.

Keywords: Statistical model; Discrete bivariate distributions; Joint hazard rate function; Positively quadrant dependent; Simulation; Statistics and numerical data.
2000 AMS Classification: 60E05; 62E10.

1 Introduction

Given the crucial role of probability distributions in characterizing natural accidents and phenomena, as well as in modeling diverse real-world data across various fields like engineering, medicine, biological and industrial studies, insurance, and economics, researchers have increasingly directed their efforts towards unveiling numerous discrete and continuous statistical distributions that precisely capture these phenomena. In recent decades, there has been a notable shift in focus towards introducing distributions that can effectively represent the intricacies of real-life data, including truncated, asymmetric, skewed, and upper-recorded data. This research specifically aims to enhance flexibility and adaptability in lifetime distribution modeling. The emphasis is on proposing an extension to the Burr distribution to address the diverse forms and types of real-world data. This extension is designed to provide a more suitable way of describing such data. The Burr distribution is particularly significant due to its capacity to encapsulate the characteristics found in various types of continuous distributions. Several authors have proposed it as a suitable model for lifetime data, with Gupta et al. [1] analyzing failure time data using the Burr distribution. Various studies, such as those by Wingo [2], have delved into maximum likelihood estimation (MLE) of Burr distribution parameters and techniques for fitting this distribution to life test data. Ghitany and Awadhi [3] provided examples of survival studies involving different leukemia treatments, incorporating censored data from the Burr distribution. The Pareto distribution can be regarded as a specific case within the Burr distribution framework and was initially introduced by Pareto to describe income distribution across

* Corresponding author e-mail: m.eldawoody@psau.edu.sa

a population. This distribution finds wide applications in economic studies, particularly when modeling phenomena with long-tailed random variable distributions. Notably, Arnold [4], and Kotz et al. [5] have applied the Pareto distribution to various naturally occurring phenomena. The Burr XII distribution, first introduced by Burr [6], encompasses distributions such as exponential, Weibull, and log-logistic for specific parameter values and provides insights into the curve shape characteristics of distributions like normal, lognormal, gamma, logistic, and several Pearson-type distributions. Rodriguez [7] extensively explored the relationship between the Burr XII distribution and other continuous distributions. Krishna and Pundir [8] introduced the discrete Burr (DB) and discrete Pareto (DP) distributions. The survival function (SF), say $S(x;.)$, and probability mass function (PMF), say $f(x;.)$, of the DB distribution are presented as follows

$$S_{DB}(x;\beta,\theta)=\theta^{\log(1+x^{\beta})};\ x=0,1,2,..., \tag{1}$$

and

$$f_{DB}(x;\beta,\theta)=\theta^{\log(1+x^{\beta})}-\theta^{\log(1+(1+x)^{\beta})};\ x=0,1,2,..., \tag{2}$$

where $0<\theta<1$ and $\beta>0$. The DP distribution is a special case of the DB when putting $\beta=1$. Over the past two decades, several papers have emerged in the literature focusing on discrete distributions derived from the discretization of continuous distributions. Many researchers have shown interest in identifying bivariate distributions that possess specific characteristics, allowing them to apply these novel lifetime distributions to predict and describe the lifespans of various devices. Bivariate discrete and continuous distributions find applications across diverse fields, including engineering, reliability analysis, sports, meteorology, drought studies, and more. Notable references include Lee and Cha [9], Nekoukhou and Kundu[10], Kundu and Nekoukhou [11-12], Ali et al. [13], Eliwa and El-Morshedy [14-15], Tahir et al. [16], Eliwa et al. [17], Lee and Cha [18], El-Morshedy et al. [19], Nekoukhou et al. [20], Alotaibi et al. [21], El-Morshedy et al. [22], Barbiero [23], Mohammed et al. [24], El-Sherpieny et al. [25-27], Al-Essa et al. [28], among others. In recent years, numerous methods have been proposed for the development of new discrete distributions. One of the most significant approaches involves discretizing well-established continuous distributions. Consider $f(x_1;x_2)$ and $S(x_1;x_2)$ as the joint probability density function (PDF) and joint survival function of a given bivariate continuous distribution over the domain $(0;\infty)^2$. In this context, a new discrete bivariate distribution can be created using the following method

$$\Pr(X_1=x_1,X_2=x_2)=S(x_1,x_2)-S(x_1+1,x_2)-S(x_1,x_2+1)+S(x_1+1,x_2+1). \tag{3}$$

The primary objective of this paper is to introduce a novel bivariate distribution derived from the discrete Burr distribution, which we refer to as the "bivariate discrete Burr distribution". Importantly, this new distribution exhibits marginal distributions consistent with the discrete Burr. The bivariate discrete Burr distribution is a versatile tool for examining and interpreting various data shapes across multiple domains, particularly data that exhibits extreme observations, also known as a "heavy-tailed model". Moreover, its bivariate hazard rate function can be employed to simulate diverse failure patterns.

The paper is organized as follows: Section 2 covers the derivation of the joint SF, joint PMF, and various associated functions. Section 3 introduces several statistical and reliability properties. In Section 4, the focus shifts to the estimation of model parameters. Section 5 involves a simulation to evaluate the effectiveness of the estimation method. The practical applicability of the proposed model is demonstrated through the analysis of two real datasets in Section 6. Finally, in Section 7, the paper concludes with some closing remarks.

2 The BDB Distribution: Description and Clarifications

Let's consider independent random variables $M_1\sim DB(\beta,\theta_1)$, $M_2\sim DB(\beta,\theta_2)$, and $M_3\sim DB(\beta,\theta_3)$. Now, if we define X_1 as the minimum of M_1 and M_3, and X_2 as the minimum of M_2 and M_3, the bivariate vector $\mathbf{X}=(X_1,X_2)$ follows a BDB distribution with the parameter vector $\tau=(\beta,\theta_1,\theta_2,\theta_3)^T$. This bivariate discrete distribution is referred to as

$BDB(\beta,\theta_1,\theta_2,\theta_3)$. If $\mathbf{X} \sim BDB(\beta,\theta_1,\theta_2,\theta_3)$, then the joint SF is given by

$$\begin{aligned}
S_{X_1,X_2}(x_1,x_2) &= \Pr(X_1 \geq x_1, X_2 \geq x_2) \\
&= \Pr(\min\{M_1,M_3\} \geq x_1, \min\{M_2,M_3\} \geq x_2) \\
&= \Pr(M_1 \geq x_1, M_2 \geq x_2, M_3 \geq \max\{x_1,x_2\}) \\
&= S_{DB}(x_1;\beta,\theta_1) S_{DB}(x_2;\beta,\theta_2) S_{DB}(z;\beta,\theta_3) \\
&= \theta_1^{\log(1+x_1^{\beta})} \theta_2^{\log(1+x_2^{\beta})} \theta_3^{\log(1+z^{\beta})} \\
&= \begin{cases} S_1(x_1,x_2) & \text{if} \quad x_1 < x_2 \\ S_2(x_1,x_2) & \text{if} \quad x_2 < x_1 \\ S_3(x) & \text{if} \quad x_1 = x_2 = x, \end{cases}
\end{aligned} \tag{4}$$

where $x_1, x_2 \in \mathbb{N}$, $z = \max\{x_1,x_2\}$ and $S_1(x_1,x_2)$, $S_2(x_1,x_2)$, $S_3(x)$ are given by

$$S_1(x_1,x_2) = S_{DB}(x_1;\beta,\theta_1) S_{DB}(x_2;\beta,\theta_2\theta_3),$$

$$S_2(x_1,x_2) = S_{DB}(x_1;\beta,\theta_1\theta_3) S_{DB}(x_2;\beta,\theta_2)$$

and

$$S_3(x) = S_{DB}(x;\beta,\theta_1\theta_2\theta_3).$$

The discrete joint SF is a critical tool for understanding and quantifying the joint behavior of multiple random variables, making it valuable in fields ranging from risk assessment and reliability to decision-making and data analysis. The marginal SF of X_i ; $(i = 1,2)$ is given by

$$S_{X_i}(x_i) = S_{DB}(x_i;\beta,\theta_i\theta_3);\ x_i \in \mathbb{N}_\circ, \tag{5}$$

where

$$S_{X_i}(x_i) = \Pr(\min\{M_i,M_3\} \geq x_i) = (\theta_i\theta_3)^{\log(1+x_i^{\beta})}.$$

The joint cumulative distribution function (CDF) can be derived using the following relationship

$$F_{X_1,X_2}(x_1,x_2) = F_{X_1}(x_1) + F_{X_2}(x_2) + S_{X_1,X_2}(x_1,x_2) - 1, \tag{6}$$

Here, $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ correspond to the marginal distributions of the random vector $\mathbf{X}$. The discrete joint CDF is a fundamental tool for understanding and quantifying the combined behavior of multiple random variables. Its significance extends to various fields, making it an essential component of statistical and probabilistic analysis in both practical applications and theoretical research.

Another crucial statistical concept is known as the joint PMF. The importance of the joint PMF lies in its ability to provide crucial information about the simultaneous occurrence of multiple events in a discrete random variable system. Here are some key points highlighting its significance:

–Multivariate Probability Analysis: Joint PMF is used to analyze the joint probability of multiple random variables. It provides a complete description of the probability distribution for all possible combinations of outcomes in a multivariate system.
–Event Dependency: It helps in understanding the dependency between different random variables. By examining how their joint PMF behaves, we can assess whether events are independent, positively correlated, or negatively correlated.
–Statistical Inference: Joint PMF plays a fundamental role in statistical inference and hypothesis testing involving multiple variables. For instance, it is crucial in chi-squared tests for independence and analysis of contingency tables.
–Risk Assessment: In various fields like finance and insurance, the joint PMF is used to model and analyze the risk associated with multiple events occurring simultaneously. This is vital for portfolio risk management and actuarial calculations.
–Engineering and Reliability: Engineers use joint PMF to assess the reliability of complex systems where the failure of multiple components can lead to system failure. It helps in modeling and analyzing the reliability of such systems.
–Decision Making: In decision theory and optimization problems, understanding the joint PMF can aid in making informed decisions when multiple random factors are involved.
–Machine Learning and Data Analysis: Joint PMF is essential in machine learning and data analysis, especially when dealing with multiple correlated variables. It helps in modeling, feature selection, and dimensionality reduction.
–Quality Control: In quality control and manufacturing, joint PMF is used to analyze the occurrence of defects or faults at multiple points in a production process, allowing for process improvement.

–Social Sciences: In the social sciences, joint PMF is used in survey analysis, especially when studying relationships between multiple variables, like income, education, and employment.
–Diverse Applications: Joint PMF finds applications in various fields such as epidemiology, environmental science, genetics, and more, where understanding the joint occurrence of events or variables is essential for research and decision-making.

In summary, the joint PMF is a fundamental concept in probability theory and statistics that enables a comprehensive understanding of the relationships and interactions between multiple random variables, leading to valuable insights and informed decision-making across a wide range of disciplines. If the random vector $\mathbf{X} \sim BDB(\beta, \theta_1, \theta_2, \theta_3)$, then the joint PMF can be formulated as

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2) & \text{if } x_1 < x_2 \\ f_2(x_1,x_2) & \text{if } x_2 < x_1 \\ f_3(x) & \text{if } x_1 = x_2 = x, \end{cases} \tag{7}$$

where

$$\begin{aligned} f_1(x_1,x_2) &= \left(\theta_1^{\log(1+x_1^\beta)} - \theta_1^{\log(1+(1+x_1)^\beta)}\right)\left((\theta_2\theta_3)^{\log(1+x_2^\beta)} - (\theta_2\theta_3)^{\log(1+(1+x_2)^\beta)}\right) \\ &= f_{DB}(x_1;\beta,\theta_1)f_{DB}(x_2;\beta,\theta_2\theta_3), \end{aligned}$$

$$\begin{aligned} f_2(x_1,x_2) &= \left((\theta_1\theta_3)^{\log(1+x_1^\beta)} - (\theta_1\theta_3)^{\log(1+(1+x_1)^\beta)}\right)\left(\theta_2^{\log(1+x_2^\beta)} - \theta_2^{\log(1+(1+x_2)^\beta)}\right) \\ &= f_{DB}(x_1;\beta,\theta_1\theta_3)f_{DB}(x_2;\beta,\theta_2) \end{aligned}$$

and

$$\begin{aligned} f_3(x) &= p_1\left((\theta_1\theta_3)^{\log(1+x^\beta)} - (\theta_1\theta_3)^{\log(1+(1+x)^\beta)}\right) - p_2\left(\theta_1^{\log(1+x^\beta)} - \theta_1^{\log(1+(1+x)^\beta)}\right) \\ &= p_1 f_{DB}(x;\beta,\theta_1\theta_3) - p_2 f_{DB}(x;\beta,\theta_1). \end{aligned}$$

where $p_1 = \theta_2^{\log(1+x^\beta)}$, $p_2 = (\theta_2\theta_3)^{\log(1+(1+x)^\beta)}$. We can derive the expressions $f_1(x_1,x_2)$, $f_2(x_1,x_2)$ and $f_3(x)$ by using the following relation

$$f_{X_1,X_2}(x_1,x_2) = S_{X_1,X_2}(x_1,x_2) - S_{X_1,X_2}(x_1+1,x_2) - S_{X_1,X_2}(x_1,x_2+1) + S_{X_1,X_2}(x_1+1,x_2+1).$$

Figure 1 displays graphical representations of the joint PMF for varying parameter values in the context of the BDB distribution.

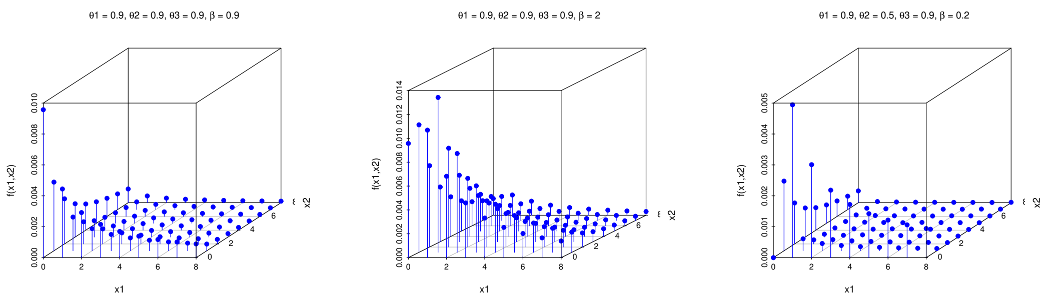

Figure 1. The PMFs of the BDB distribution.

Remarkably, the PMFs can be employed to examine and assess non-symmetric data exhibiting various forms of kurtosis. Moreover, it can serve as a heavy-tailed distribution to address outlier observations. The marginal PMF of $X_i, (i = 1,2)$, corresponding to (5) is

$$\begin{aligned} f_{X_i}(x_i) &= (\theta_i\theta_3)^{\log(1+i^\beta)} - (\theta_i\theta_3)^{\log(1+(1+i)^\beta)}, x_i \in \mathbb{N}_\circ \\ &= f_{DB}(x_i;\beta,\theta_i\theta_3). \end{aligned} \tag{8}$$

The following interpretations can be proved for BDB distribution. It's important to observe that when β equals 1, the random variables (X_1,X_2) exhibit discrete Pareto marginal distributions. This scenario gives rise to a bivariate discrete Pareto distribution characterized by parameters θ_1, θ_2 and θ_3, with its joint SF as follows

$$S_{X_1,X_2}(x_1,x_2) = \theta_1^{\log(1+x_1)}\theta_2^{\log(1+x_2)}\theta_3^{\log(1+z)}, \tag{9}$$

where $x_1,x_2 \in \mathbb{N}$ and $z = \max\{x_1,x_2\}$. Suppose $(X_{m1},X_{m2}) \sim BDB(\beta,\theta_{m1},\theta_{m2},\theta_{m3})$ which $m = 1,2,...,n$ and they are independently distributed. If $C_1 = \min(X_{11},X_{21},...,X_{n1})$ and $C_2 = \min(X_{12},X_{22},...,X_{n2})$, then

$$(C_1,C_2) \sim BDB\left(\beta,\prod_{m=1}^{n}\theta_{m1},\prod_{m=1}^{n}\theta_{m2},\prod_{m=1}^{n}\theta_{m3}\right). \tag{10}$$

3 Distributional Characteristics and its Attributes

3.1 Positive quadrant dependent

Positive quadrant dependent (PQD) is a term used in statistics to describe a specific type of dependence between random variables. When random variables are said to be PQD, it means that they tend to increase or decrease together in the positive direction. In other words, when one variable takes on a higher value than its expected or average value, the other variable is more likely to also take on a higher value than its expected or average value. Similarly, when one variable takes on a lower value than its expected or average value, the other variable is more likely to take on a lower value as well. This positive dependence implies that there is a positive correlation between the variables in the first quadrant of a Cartesian coordinate system (where both variables are positive). It suggests that when one variable experiences an extreme positive deviation from its mean, the other variable is more likely to also exhibit a positive deviation from its mean. PQD is commonly encountered in various fields, including finance, economics, and environmental science. For example, in finance, PQD may be observed when the stock prices of two companies tend to rise or fall together due to similar market conditions or industry trends. Understanding PQD is important in statistical modeling, risk assessment, and decision-making, as it can impact the analysis of joint probabilities, correlations, and dependencies between variables. Researchers and analysts often use different statistical tools and models to account for PQD when working with data that exhibits this type of relationship. If $X_1 = \min\{M_1,M_3\}$ and $X_2 = \min\{M_2,M_3\}$, then X_1 and X_2 are PQD where

$$\Pr(X_1 > x_1, X_2 > x_2) \geq \Pr(X_1 > x_1)\Pr(X_2 > x_2), \text{ for all } x_1 \text{ and } x_2,$$

where $M_1 \sim DB(\beta,\theta_1)$, $M_2 \sim DB(\beta,\theta_2)$, and $M_3 \sim DB(\beta,\theta_3)$. Or, equivalently, if and only if

$$\Pr(X_1 \leq x_1, X_2 \leq x_2) \geq \Pr(X_1 \leq x_1)\Pr(X_2 \leq x_2), \text{ for all } x_1 \text{ and } x_2.$$

3.2 Joint probability generating function

The joint probability generating function (PGF) is a mathematical concept used in probability theory and statistics to describe and analyze the joint probability distribution of multiple random variables. It is a powerful tool for understanding the behavior and relationships between these variables. Its versatile applications span across various fields. In actuarial science, the joint PGF can be employed to model the joint distribution of claims and assess risk in insurance portfolios (insurance and risk assessment). In queueing systems, the joint PGF helps analyze the number of customers in a queue at different time points (queueing theory). In reliability analysis, it is used in reliability theory to study the number of failures in a system over time. In statistics and data analysis, it is used to compute moments and correlations between multiple random variables, which is valuable in statistical analysis. In economics, it can be used to model and analyze the joint distribution of economic variables. If the random vector $\mathbf{X} \sim BDB(\beta,\theta_1,\theta_2,\theta_3)$, then the joint PGF can be expressed as

$$\begin{aligned}
G(u,v) &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}P(X_1=k,X_2=l)u^kv^l \\
&= \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty} f_{DB}(k;\beta,\theta_1)f_{DB}(l;\beta,\theta_2\theta_3)u^kv^l + \sum_{l=0}^{\infty}\sum_{k=l+1}^{\infty} f_{DB}(k;\beta,\theta_1\theta_3)f_{DB}(l;\beta,\theta_2)u^kv^l \\
&+ \sum_{k=0}^{\infty}\theta_2^{\log(1+k^{\beta})}f_{DB}(k;\beta,\theta_1\theta_3)(uv)^k - \sum_{k=0}^{\infty}(\theta_2\theta_3)^{\log(1+(1+k)^{\beta})}f_{DB}(k;\beta,\theta_1)(uv)^k,
\end{aligned}$$

where $|u| < 1$ and $|v| < 1$. The joint PGF allows us to extract the marginal PGFs for individual random variables.

3.3 Conditional distribution and expectation

These concepts are widely used in fields such as probability theory, statistics, finance, economics, machine learning, and engineering, where understanding the impact of one variable on another in the presence of specific conditions is essential for modeling, analysis, and decision support. The conditional PMF of $(X_1 \mid X_2 = x_2)$, say $f_{X_1|X_2=x_2}(x_1 \mid x_2)$, is given by

$$f_{X_1|X_2=x_2}(x_1 \mid x_2) = \begin{cases} f_1(x_1 \mid x_2) & \text{if} \quad 0 \leq x_1 < x_2 \\ f_2(x_1 \mid x_2) & \text{if} \quad 0 \leq x_2 < x_1 \\ f_3(x_1 \mid x_2) & \text{if} \quad 0 \leq x_1 = x_2 = x, \end{cases} \tag{11}$$

where

$$f_1(x_1 \mid x_2) = f_{DB}(x_1; \beta, \theta_1),$$
$$f_2(x_1 \mid x_2) = \frac{f_{DB}(x_1; \beta, \theta_1\theta_3) f_{DB}(x_2; \beta, \theta_2)}{f_{DB}(x_2; \beta, \theta_2\theta_3)}$$

and

$$f_3(x_1 \mid x_2) = \frac{\theta_2^{\log(1+x^\beta)} f_{DB}(x; \beta, \theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x)^\beta)} f_{DB}(x; \beta, \theta_1)}{f_{DB}(x; \beta, \theta_2\theta_3)}.$$

The conditional SF of $(X_1 \mid X_2 \geq x_2)$, say $S_{X_1|X_2 \geq x_2}(x_1)$, can be reported as

$$S_{X_1|X_2 \leq x_2}(x_1) = \begin{cases} \theta_1^{\log(1+x_1^\beta)} & \text{if} \quad 0 \leq x_1 < x_2 \\ (\theta_1\theta_3)^{\log(1+x_1^\beta)} \theta_3^{-\log(1+x_2^\beta)} & \text{if} \quad 0 \leq x_2 < x_1 \\ \theta_1^{\log(1+x^\beta)} & \text{if} \quad 0 \leq x_1 = x_2 = x. \end{cases} \tag{12}$$

Thus, the conditional expectation of $(X_1 \mid X_2 = x_2)$, say $E(X_1 \mid X_2 = x_2)$, can be listed as

$$\begin{aligned} E(X_1 \mid X_2 = x_2) &= \sum_{x_1=0}^{\infty} x_1 f_{X_1|X_2=x_2}(x_1 \mid x_2) \\ &= \sum_{x_1=0}^{x_2-1} x_1 f_1(x_1 \mid x_2) + \sum_{x_1=x_2+1}^{\infty} x_1 f_2(x_1 \mid x_2) + \sum_{x_1=0}^{\infty} x_2 f_3(x_1 \mid x_2). \\ &= \sum_{x_1=0}^{x_2-1} x_1 \left(\theta_1^{\log(1+x_1^\beta)} - \theta_1^{\log(1+(1+x_1)^\beta)} \right) + \frac{\left(\theta_2^{\log(1+x_2^\beta)} - \theta_2^{\log(1+(1+x_2)^\beta)} \right)}{\left((\theta_2\theta_3)^{\log(1+x_2^\beta)} - (\theta_2\theta_3)^{\log(1+(1+x_2)^\beta)} \right)} \\ &\times \sum_{x_1=x_2+1}^{\infty} x_1 \left((\theta_1\theta_3)^{\log(1+x_1^\beta)} - (\theta_1\theta_3)^{\log(1+(1+x_1)^\beta)} \right) \\ &+ \frac{x_2 \theta_2^{\log(1+x^\beta)} \left((\theta_1\theta_3)^{\log(1+x_2^\beta)} - (\theta_1\theta_3)^{\log(1+(1+x_2)^\beta)} \right)}{\left((\theta_2\theta_3)^{\log(1+x_2^\beta)} - (\theta_2\theta_3)^{\log(1+(1+x_2)^\beta)} \right)} \\ &- \frac{x_2 (\theta_2\theta_3)^{\log(1+(1+x)^\beta)} \left(\theta_1^{\log(1+x_2^\beta)} - \theta_1^{\log(1+(1+x_2)^\beta)} \right)}{\left((\theta_2\theta_3)^{\log(1+x_2^\beta)} - (\theta_2\theta_3)^{\log(1+(1+x_2)^\beta)} \right)}. \end{aligned} \tag{13}$$

3.4 Joint hazard rate function

The joint hazard rate function (HRF) provides insights into how the risk of a concurrent event involving multiple variables changes with the values of those variables. It is often used in reliability analysis to understand the likelihood of system failures, where X_1 and X_2 might represent the lifetimes of different components or subsystems. Understanding the joint hazard rate function is valuable for assessing and mitigating risks in situations where multiple factors or components interact to influence the occurrence of events or failures. If the random vector $\mathbf{X} \sim BDB(\beta, \theta_1, \theta_2, \theta_3)$, then the joint HRF can be expressed as

$$h_{X_1,X_2}(x_1,x_2) = \begin{cases} h_1(x_1,x_2) & \text{if } \; x_1 < x_2 \\ h_2(x_1,x_2) & \text{if } \; x_2 < x_1 \\ h_3(x) & \text{if } \; x_1 = x_2 = x, \end{cases} \tag{14}$$

where

$$h_1(x_1,x_2) = \frac{\left(\theta_1^{\log(1+x_1^{\beta})} - \theta_1^{\log(1+(1+x_1)^{\beta})}\right)\left((\theta_2\theta_3)^{\log(1+x_2^{\beta})} - (\theta_2\theta_3)^{\log(1+(1+x_2)^{\beta})}\right)}{\theta_1^{\log(1+x_1^{\beta})}(\theta_2\theta_3)^{\log(1+x_2^{\beta})}},$$

$$h_2(x_1,x_2) = \frac{\left((\theta_1\theta_3)^{\log(1+x_1^{\beta})} - (\theta_1\theta_3)^{\log(1+(1+x_1)^{\beta})}\right)\left(\theta_2^{\log(1+x_2^{\beta})} - \theta_2^{\log(1+(1+x_2)^{\beta})}\right)}{(\theta_1\theta_3)^{\log(1+x_1^{\beta})}\theta_2^{\log(1+x_2^{\beta})}},$$

and

$$h_3(x) = \frac{\theta_2^{\log(1+x^{\beta})}\left((\theta_1\theta_3)^{\log(1+x^{\beta})} - (\theta_1\theta_3)^{\log(1+(1+x)^{\beta})}\right)}{(\theta_1\theta_2\theta_3)^{\log(1+x_1^{\beta})}} - \frac{(\theta_2\theta_3)^{\log(1+(1+x)^{\beta})}\left(\theta_1^{\log(1+x^{\beta})} - \theta_1^{\log(1+(1+x)^{\beta})}\right)}{(\theta_1\theta_2\theta_3)^{\log(1+x_1^{\beta})}}.$$

Figure 2 exhibits visual depictions of the joint HRF with varying parameter values within the framework of the BDB distribution.

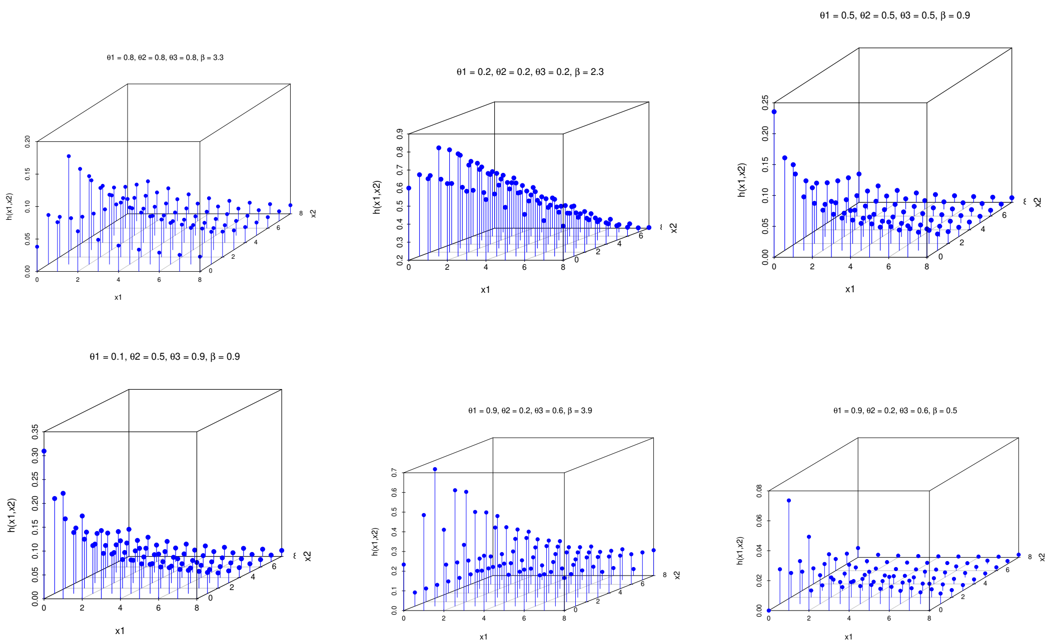

Figure 2. The joint HRFs of the BDB distribution.

Notably, the utilization of joint HRFs allows for the exploration and evaluation of diverse failure patterns across multiple domains.

3.5 Joint reversed hazard rate function

The ”reversed hazard rate function” or ”inverse hazard rate function” is a concept used in survival analysis and reliability theory. It represents the expected time or survival time until an event occurs, given that the event has not occurred until time t. The reversed hazard rate function is used to describe the remaining life expectancy, conditional on the event not occurring up to time t (see Bismi, [29]). If the random vector $\mathbf{X} \sim BDB(\beta, \theta_1, \theta_2, \theta_3)$, then the joint HRF can be listed as

$$R_{X_1,X_2}(x_1,x_2) = \begin{cases} r_1(x_1,x_2) & \text{if } \ x_1 < x_2 \\ r_2(x_1,x_2) & \text{if } \ x_2 < x_1 \\ r_3(x) & \text{if } \ x_1 = x_2 = x, \end{cases} \tag{15}$$

where

$$r_1(x_1,x_2) = \frac{\left(\theta_1^{\log(1+x_1^\beta)} - \theta_1^{\log(1+(1+x_1)^\beta)}\right)\left((\theta_2\theta_3)^{\log(1+x_2^\beta)} - (\theta_2\theta_3)^{\log(1+(1+x_2)^\beta)}\right)}{1-(\theta_1\theta_3)^{\log(1+x_1^\beta)} + (\theta_2\theta_3)^{\log(1+x_2^\beta)}\left(\theta_1^{\log(1+x_1^\beta)} - 1\right)},$$

$$r_2(x_1,x_2) = \frac{\left((\theta_1\theta_3)^{\log(1+x_1^\beta)} - (\theta_1\theta_3)^{\log(1+(1+x_1)^\beta)}\right)\left(\theta_2^{\log(1+x_2^\beta)} - \theta_2^{\log(1+(1+x_2)^\beta)}\right)}{1-(\theta_2\theta_3)^{\log(1+x_2^\beta)} + (\theta_1\theta_3)^{\log(1+x_1^\beta)}\left(\theta_2^{\log(1+x_2^\beta)} - 1\right)},$$

and

$$r_3(x) = \frac{\theta_2^{\log(1+x^\beta)}\left((\theta_1\theta_3)^{\log(1+x^\beta)} - (\theta_1\theta_3)^{\log(1+(1+x)^\beta)}\right)}{1+\theta_3^{\log(1+x^\beta)}\left((\theta_1\theta_2)^{\log(1+x^\beta)} - \theta_1^{\log(1+x^\beta)} - \theta_2^{\log(1+x^\beta)}\right)} - \frac{(\theta_2\theta_3)^{\log(1+(1+x)^\beta)}\left(\theta_1^{\log(1+x^\beta)} - \theta_1^{\log(1+(1+x)^\beta)}\right)}{1+\theta_3^{\log(1+x^\beta)}\left((\theta_1\theta_2)^{\log(1+x^\beta)} - \theta_1^{\log(1+x^\beta)} - \theta_2^{\log(1+x^\beta)}\right)}.$$

The reversed hazard rate at any time t is the reciprocal of the hazard rate at the same time. This relationship allows you to derive one function from the other. If you know one function, you can easily calculate the other. Both the hazard rate and reversed hazard rate functions are valuable tools in analyzing survival data, reliability, and risk assessment in various fields, such as medicine, engineering, finance, and more. They provide a comprehensive view of how events or failures occur over time and how survival probabilities change as time progresses.

4 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a method used in statistics and machine learning to estimate the parameters of a statistical model. The main idea behind MLE is to find the parameter values that maximize the likelihood of the observed data, given a specific probabilistic model. The likelihood function is a function of the model parameters and is used to describe how well the model explains the observed data. It is denoted as $l(\delta|x)$, where δ represents the parameter(s) of the model, and x represents the observed data. The likelihood function is often transformed into its natural logarithm, creating the log-likelihood (L) function . This transformation simplifies the optimization process and does not change the parameter values that maximize the likelihood. The goal is to find the parameter values δ that maximize the L function. This is typically done using optimization techniques, such as gradient descent, the Newton-Raphson method, or other numerical optimization algorithms where

$$\widehat{\delta} = \arg\max L(\delta|x). \tag{16}$$

One of the attractive features of MLE is its efficiency and asymptotic properties, which means that as the sample size increases, the MLE estimates become increasingly accurate and tend to follow a normal distribution. However, MLE can also be sensitive to the choice of the initial parameter values, and in some cases, it may not have unique solutions or closed-form solutions, requiring the use of numerical optimization techniques. In this section, the unknown parameters β, θ_1, θ_2 and θ_3 of the BDB distribution are estimated by using the method of maximum likelihood. Suppose that, we have a sample of size t, of the form $\{(x_{11},x_{21}),(x_{12},x_{22}),...,(x_{1t},x_{2t})\}$ from the BDB distribution. We use the following notations: $I_1 = \{x_{1j} < x_{2j}\}$, $I_2 = \{x_{2j} < x_{1j}\}$, $I_3 = \{x_{1j} = x_{2j} = x_j\}$, $I = I_1 \cup I_2 \cup I_3$, $|I_1| = t_1$, $|I_2| = t_2$, $|I_3| = t_3$ and $t = t_1 + t_2 + t_3$.Derived from the collected observations, the likelihood function takes the form of

$$l(\beta,\theta_1,\theta_2,\theta_3|x) = \prod_{j=1}^{t_1} f_1(x_{1j},x_{2j}) \prod_{j=1}^{t_2} f_2(x_{1j},x_{2j}) \prod_{j=1}^{t_3} f_3(x_j). \tag{17}$$

The log-likelihood function transforms into

$$\begin{aligned} L(\beta,\theta_1,\theta_2,\theta_3|x) = & \sum_{j=1}^{t_1} \ln\left(m_1(x_{1j};\theta_1)\right) + \sum_{j=1}^{t_1} \ln\left(m_1(x_{2j};\theta_2\theta_3)\right) \\ & + \sum_{j=1}^{t_2} \ln\left(m_1(x_{1j};\theta_1\theta_3)\right) + \sum_{j=1}^{t_2} \ln\left(m_1(x_{2j};\theta_2)\right) \\ & + \sum_{j=1}^{t_3} \ln\left(\theta_2^{\log(1+x_j^\beta)} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} m_1(x_j;\theta_1)\right), \end{aligned} \tag{18}$$

where

$$m_1(x;\gamma) = \gamma^{\log(1+x^\beta)} - \gamma^{\log(1+(1+x)^\beta)}.$$

To find the maximum likelihood estimates (MLEs) for the parameters β, θ_1, θ_2 and θ_3, you can calculate the first partial derivatives of (18) with respect to each of these parameters and then set these derivatives equal to zero. This process yields the likelihood equations in the following format

$$\begin{aligned} \frac{\partial L}{\partial \beta} = & \sum_{j=1}^{t_1} \frac{m_2(x_{1j};\theta_1) - m_2(1+x_{1j};\theta_1)}{m_1(x_{1j};\theta_1)} + \sum_{j=1}^{t_1} \frac{m_2(x_{2j};\theta_2\theta_3) - m_2(1+x_{2j};\theta_2\theta_3)}{m_1(x_{2j};\theta_2\theta_3)} \\ & + \sum_{j=1}^{t_2} \frac{m_2(x_{1j};\theta_1\theta_3) - m_2(1+x_{1j};\theta_1\theta_3)}{m_1(x_{1j};\theta_1\theta_3)} + \sum_{j=1}^{t_2} \frac{m_2(x_{2j};\theta_2) - m_2(1+x_{2j};\theta_2)}{m_1(x_{2j};\theta_2)} \\ & + \sum_{j=1}^{t_3} \frac{\theta_2^{\log(1+x_j^\beta)} \left(m_2(x_j;\theta_1\theta_3) - m_2(1+x_j;\theta_1\theta_3)\right) + m_2(x_j;\theta_2) m_1(x_j;\theta_1\theta_3)}{\theta_2^{\log(1+x_j^\beta)} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} m_1(x_j;\theta_1)} \\ & - \sum_{j=1}^{t_3} \frac{(\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} \left(m_2(x_j;\theta_1) - m_2(1+x_j;\theta_1)\right) + m_2(1+x_j;\theta_2\theta_3) m_1(x_j;\theta_1)}{\theta_2^{\log(1+x_j^\beta)} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} m_1(x_j;\theta_1)}, \end{aligned} \tag{19}$$

$$\begin{aligned} \frac{\partial L}{\partial \theta_1} = & \sum_{j=1}^{t_1} \frac{m_3(x_{1j};\theta_1) - m_3(1+x_{1j};\theta_1)}{m_1(x_{1j};\theta_1)} + \\ & \sum_{j=1}^{t_2} \frac{\theta_3^{\log(1+x_{1j}^\beta)} m_3(x_{1j};\theta_1) - \theta_3^{\log(1+(1+x_{1j})^\beta)} m_3(1+x_{1j};\theta_1)}{m_1(x_{1j};\theta_1\theta_3)} + \\ & \sum_{j=1}^{t_3} \frac{\theta_2^{\log(1+x_j^\beta)} \left(\theta_3^{\log(1+x_{1j}^\beta)} m_3(x_j;\theta_1) - \theta_3^{\log(1+(1+x_j)^\beta)} m_3(1+x_{1j};\theta_1)\right)}{\theta_2^{\log(1+x_j^\beta)} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} m_1(x_j;\theta_1)} \\ & - \sum_{j=1}^{t_3} \frac{(\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} \left(m_3(x_{1j};\theta_1) - m_3(1+x_{1j};\theta_1)\right)}{\theta_2^{\log(1+x_j^\beta)} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^\beta)} m_1(x_j;\theta_1)}, \end{aligned} \tag{20}$$

$$\frac{\partial L}{\partial \theta_2} = \sum_{j=1}^{t_1} \frac{\theta_3^{\log(1+x_{2j}^{\beta})} m_3(x_{2j};\theta_2) - \theta_3^{\log(1+(1+x_{2j})^{\beta})} m_3(1+x_{2j};\theta_2)}{m_1(x_{2j};\theta_2\theta_3)}$$
$$+ \sum_{j=1}^{t_2} \frac{m_3(x_{2j};\theta_2) - m_3(1+x_{2j};\theta_2)}{m_1(x_{2j};\theta_2)}$$
$$+ \sum_{j=1}^{t_3} \frac{m_3(x_j;\theta_2) m_1(x_j;\theta_1\theta_3) - \theta_3^{\log(1+(1+x_j)^{\beta})} m_3(1+x_j;\theta_2) m_1(x_j;\theta_1)}{\theta_2^{\log(1+x_j^{\beta})} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^{\beta})} m_1(x_j;\theta_1)}, \tag{21}$$

and

$$\frac{\partial L}{\partial \theta_3} = \sum_{j=1}^{t_1} \frac{\theta_2^{\log(1+x_{2j}^{\beta})} m_3(x_{2j};\theta_3) - \theta_2^{\log(1+(1+x_{2j})^{\beta})} m_3(1+x_{2j};\theta_3)}{m_1(x_{2j};\theta_2\theta_3)} +$$
$$\sum_{j=1}^{t_2} \frac{\theta_1^{\log(1+x_{1j}^{\beta})} m_3(x_{1j};\theta_3) - \theta_1^{\log(1+(1+x_{1j})^{\beta})} m_3(1+x_{1j};\theta_3)}{m_1(x_{1j};\theta_1\theta_3)} +$$
$$\sum_{j=1}^{t_3} \frac{\theta_2^{\log(1+x_j^{\beta})} \left(\theta_1^{\log(1+x_j^{\beta})} m_3(x_j;\theta_3) - \theta_1^{\log(1+(1+x_j)^{\beta})} m_3(1+x_{1j};\theta_3)\right)}{\theta_2^{\log(1+x_j^{\beta})} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^{\beta})} m_1(x_j;\theta_1)} -$$
$$\sum_{j=1}^{t_3} \frac{\theta_2^{\log(1+(1+x_j)^{\beta})} m_3(1+x_j;\theta_3) m_1(x_j;\theta_1)}{\theta_2^{\log(1+x_j^{\beta})} m_1(x_j;\theta_1\theta_3) - (\theta_2\theta_3)^{\log(1+(1+x_j)^{\beta})} m_1(x_j;\theta_1)}, \tag{22}$$

where

$$m_2(x;\gamma) = \frac{\ln x x^{\beta} \ln \gamma \gamma^{\log(1+x^{\beta})}}{1+x^{\beta}},$$
$$m_3(x;\gamma) = \ln(1+x^{\beta}) \gamma^{\log(1+x^{\beta})-1}.$$

To determine the MLEs for the parameters β, θ_1, θ_2 and θ_3, you can derive these estimates by solving the system of four nonlinear equations presented in (19) through (22). Solving these equations is challenging, necessitating the use of a numerical technique to obtain the MLEs.

5 Evaluating the Characteristics of the MLE Method Using Simulation Technique

Applying the Monte Carlo Markov Chain (MCMC) technique for the simulation of a bivariate discrete probability distribution is a valuable approach widely employed across various domains, including statistics, machine learning, and probabilistic modeling. MCMC stands out as a potent statistical method designed to generate random samples from intricate probability distributions. Its particular strength comes to the forefront when dealing with probability distributions that lack a closed-form expression but can be assessed pointwise. MCMC encompasses a suite of algorithms, such as the Metropolis-Hastings algorithm and Gibbs sampling, which systematically sample from the target distribution to approximate its key properties. This approach allows the simulation of complex and high-dimensional probability distributions, rendering it relevant for tackling real-world challenges. It facilitates the estimation of distribution characteristics, summary statistics, and other properties of interest.

Nevertheless, the task of selecting appropriate transition probabilities and proposal mechanisms can be intricate, especially in scenarios featuring high-dimensional distributions. Addressing convergence and mixing concerns is paramount in MCMC simulations, and diagnostic tools like the Gelman-Rubin statistic are often enlisted to scrutinize the convergence of the chains. Shifting the focus to the evaluation of MLE within an MCMC framework, this endeavor plays a pivotal role in the realm of statistical modeling. Ensuring the effective integration of MLE, a frequently used parameter estimation method, within the context of MCMC, known for its capacity to sample from intricate probability distributions, is of paramount importance. The procedure involves selecting an appropriate MCMC algorithm, with common choices encompassing the Metropolis-Hastings algorithm, Gibbs sampling, and Hamiltonian Monte Carlo, tailored to the specific problem at hand. Generating synthetic data closely resembling real-world data is integral, allowing control over the true parameter values. MLE is seamlessly incorporated within the MCMC framework, utilizing MLE estimates as the initial values for the MCMC simulation. Running the MCMC algorithm with MLE-initialized values yields posterior samples for the parameters of interest, along with the opportunity to obtain MLE estimates from the MCMC procedure. Evaluation of MLE performance under MCMC encompasses various aspects, including assessments of bias, variance, convergence, efficiency, and more. This rigorous evaluation procedure not only serves to gauge the accuracy and reliability of parameter estimates but also sheds light on the synergy between MLE and MCMC within the context of specific modeling tasks. In this section, we employ MLE techniques to estimate the parameters $\theta_1, \theta_2, \theta_3$ and β of the BDB distribution. The population parameters are generated using the **R**-software package. We obtain sampling distributions for various sample sizes, specifically, $n = 40$, 90, 120, 140, 200, and 300, based on 1000 replications. This section conducts an evaluation of the MLE technique by examining properties such as bias and mean square error (MSE). A general approach for generating a bivariate vector X from the BDB distribution begins with the generation of a value Y from the continuous DB distribution. Subsequently, this value is discretized to derive the variable X. The estimated MLEs are presented in Tables 1 and 2 for two instances of the BDB distribution: BDB(0.7, 0.3, 0.3, 0.6) and BDB(0.8, 0.4, 0.5, 0.7).

Table 1. The bias and MSE values for the BDB$(0.7, 0.3, 0.3, 0.6)$.

	Size	θ_1		θ_2		θ_3		β	
Method	n	**bias**	**MSE**	**bias**	**MSE**	**bias**	**MSE**	**bias**	**MSE**
	25	0.098	0.075	0.097	0.056	0.081	0.029	0.068	0.074
MLE	**40**	0.055	0.033	0.072	0.037	0.044	0.017	0.047	0.046
	90	0.038	0.029	0.069	0.029	0.037	0.014	0.034	0.030
	120	0.025	0.021	0.064	0.022	0.024	0.012	0.025	0.026
	140	0.022	0.019	0.052	0.014	0.021	0.011	0.022	0.017
	200	0.015	0.015	0.034	0.010	0.017	0.009	0.016	0.015
	300	0.006	0.011	0.013	0.007	0.013	0.004	0.014	0.010

Table 2. The bias and MSE values for the BDB$(0.8, 0.4, 0.5, 0.7)$.

	Size	θ_1		θ_2		θ_3		ζ	
Method	n	**bias**	**MSE**	**bias**	**MSE**	**bias**	**MSE**	**bias**	**MSE**
MLE	**25**	0.076	0.058	0.083	0.074	0.039	0.061	0.053	0.066
	40	0.049	0.037	0.066	0.039	0.028	0.036	0.034	0.040
	90	0.042	0.031	0.053	0.024	0.022	0.027	0.031	0.036
	120	0.039	0.022	0.042	0.019	0.016	0.014	0.026	0.029
	140	0.030	0.014	0.030	0.013	0.009	0.012	0.021	0.016
	200	0.022	0.012	0.019	0.008	0.002	0.007	0.018	0.012
	300	0.011	0.007	0.010	0.002	0.001	0.006	0.009	0.005

Tables 1 and 2 yield the following insights: As the sample size, n, approaches infinity, the bias consistently diminishes to zero. Similarly, the mean square errors (MSEs) exhibit a consistent trend of decreasing to zero as $n \to \infty$. These findings underscore the estimators' consistency, implying that MLE is a robust and effective method for data analysis.

6 Examination of Data Fit

In this section, we analyze two real datasets to assess the efficacy of the BDB distribution. For each dataset, we perform a comparative analysis by pitting the BDB distribution against several alternative distributions. The evaluation is based on multiple criteria, including the Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), and the maximized L.

6.1 Dataset I: Medical data

These data, as documented in Davis [30], pertain to the effectiveness of steam inhalation in alleviating common cold symptoms. Figures 3-6 display nonparametric plots of data set I. Notably, the plots reveal the presence of some joint extreme values and outliers.

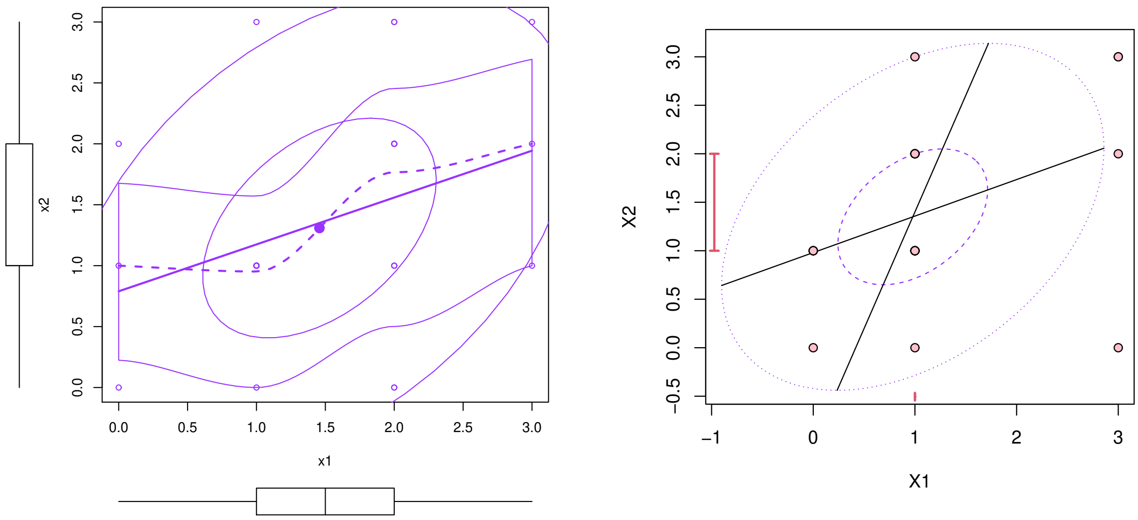

Figure 3. Scatter (left panel) and box (right panel) sketches of data set I.

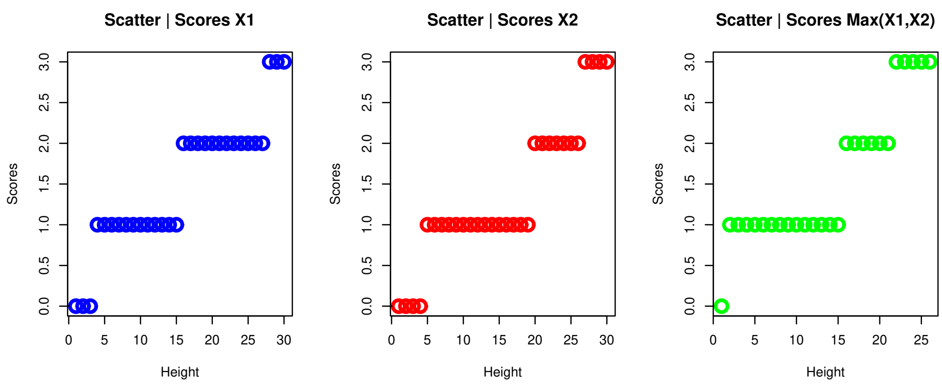

Figure 4. Visualizing marginal distributions with scatter sketches: Data set I.

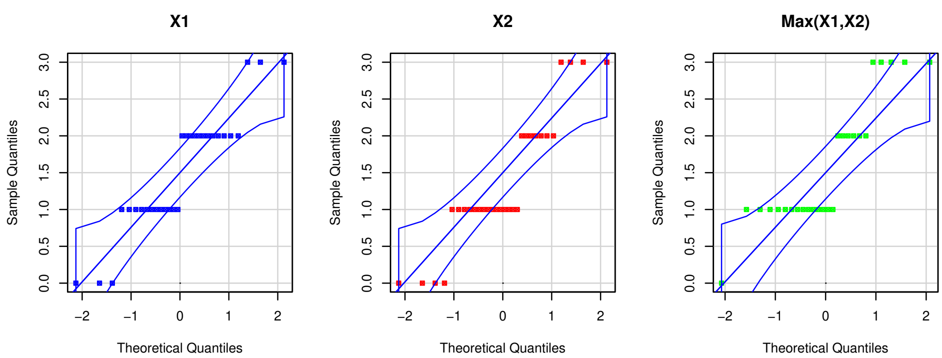

Figure 5. Exploring data set I through QQ sketches for marginal distributions.

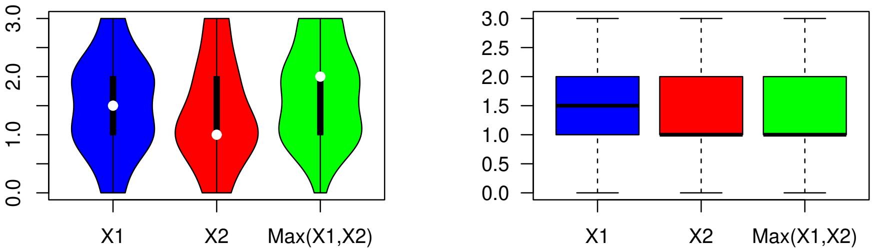

Figure 6. Investigating marginal distributions in data set I using box and violin sketches.

Before embarking on the analysis of the bivariate data, the initial step involves the examination of the marginal distributions. This assessment revealed that all p-values associated with the marginals exceeded 0.05. Subsequently, an evaluation was conducted to gauge the suitability of the BDB distribution for modeling dataset I. The empirical results, as shown in Table 3, include the proposed model and several robust competing models, such as the bivariate discrete exponentiated (BDE), bivariate discrete inverse exponentiated (BDIE), bivariate discrete inverse Rayleigh (BDIR), and bivariate discrete inverse Weibull (BDIW) distributions. It becomes readily apparent that the proposed model outperforms all other competing distributions.

Table 3. The MLEs and goodness-of-fit for dataset I.

	Distribution				
Statistic	**BDE**	**BDIE**	**BDIR**	**BDIW**	**BDB**
$\widehat{\beta}$	$--$	$--$	$--$	2.453	7.612
$\widehat{\theta}_1$	0.846	0.501	0.262	0.192	0.886
$\widehat{\theta}_2$	0.792	0.622	0.405	0.337	0.924
$\widehat{\theta}_3$	0.693	0.383	0.363	0.360	0.828
$-L$	88.0	92.48	78.66	76.51	73.816
AIC	182.0	190.96	163.32	161.02	155.632
CAIC	182.92	191.88	164.24	162.62	157.230
BIC	186.20	195.16	167.52	166.62	161.240
HQIC	183.34	192.30	164.66	162.81	157.430

6.2 Dataset II: Engineering data

These data, as documented in Casiena [31]. Figures 7-10 showcase nonparametric data plots for data set II. It is evident that there are some noteworthy joint extreme values and outliers.

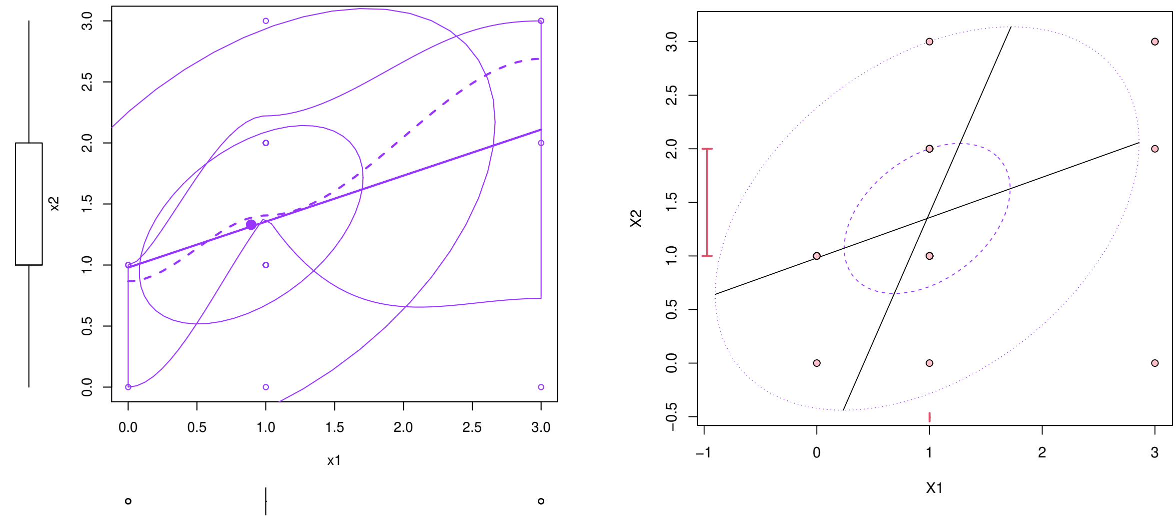

Figure 7. Scatter (left panel) and box (right panel) sketches of data set II.

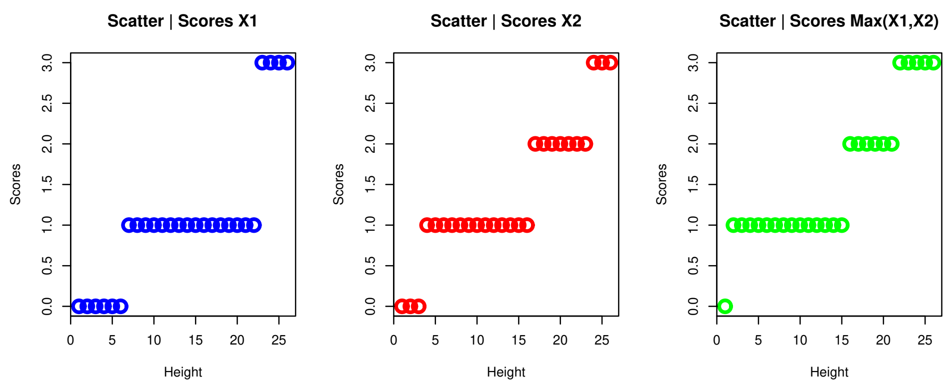

Figure 8. Representing marginal distributions using scatter plots: Data set II.

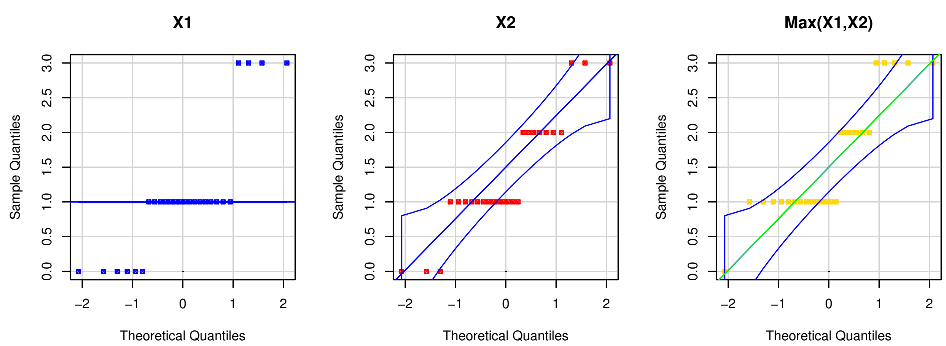

Figure 9. Examining marginal distributions in data set II with QQ plots.

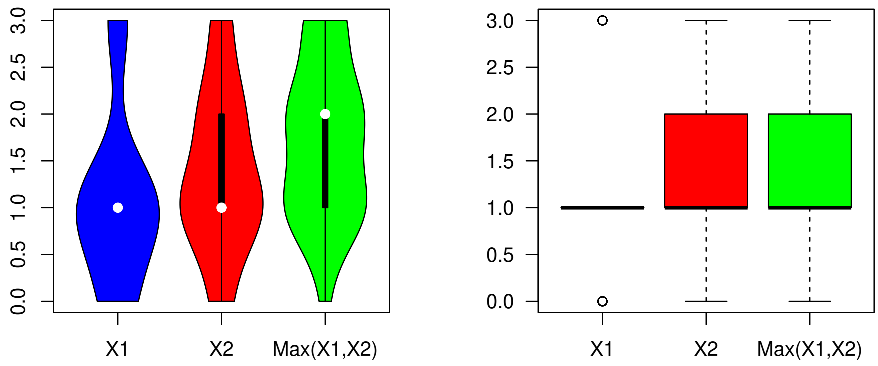

Figure 10. Analyzing data set II's marginal distributions with box and violin plots.

Before commencing the analysis of the bivariate data, the initial step involves scrutinizing the marginal distributions. This assessment revealed that all p-values associated with the marginals exceeded 0.05. Subsequently, an evaluation was carried out to assess the appropriateness of employing the BDB distribution for modeling dataset II. The empirical findings, presented in Table 4, encompass the proposed model and several robust competing models, namely, the bivariate discrete Rayleigh (BDR), bivariate discrete Weibull (BDW), BDIE, BDIR, BDIW and BDB distributions. It becomes evidently clear that the proposed model surpasses all other competing distributions.

Table 4. The MLEs and goodness-of-fit for dataset II.

	Distribution						
Statistic	**BDE**	**BDR**	**BDW**	**BDIE**	**BDIR**	**BDIW**	**BDB**
$\widehat{\beta}$	−−	−−	2.125	−−	−−	2.738	7.698
$\widehat{\theta}_1$	0.652	0.790	0.807	0.669	0.493	0.420	0.736
$\widehat{\theta}_2$	0.812	0.872	0.882	0.388	0.212	0.141	0.922
$\widehat{\theta}_3$	0.713	0.905	0.917	0.514	0.561	0.587	0.855
$-L$	75.35	63.99	63.89	78.54	64.10	61.96	60.62
AIC	156.70	133.98	133.78	163.07	134.21	131.82	129.24
CAIC	157.79	135.07	134.87	163.99	135.29	133.82	131.14
BIC	160.47	137.75	137.55	167.28	137.98	136.95	134.27
HQIC	157.79	135.07	134.87	164.42	135.29	133.37	130.69

7 Summary and Findings

This investigation centered on the introduction of an innovative statistical model, namely the bivariate discrete Burr distribution. Following the proposal of the mathematical structure for this novel model, a comprehensive exploration of its statistical properties ensued. The results revealed the model's potential applicability to assess a wide range of data types, particularly those featuring outlier observations. Moreover, its adaptability to address various forms of failures was highlighted, attributable to the flexibility inherent in its hazard rate function. Furthermore, the introduced bivariate distribution showcased a noteworthy property known as positive quadrant dependence, thereby augmenting its relevance and applicability across diverse fields. In the realm of parameter estimation, the study employed the maximum likelihood estimation method. To gauge the effectiveness of this approach under different circumstances, an extensive simulation study was conducted, evaluating both bias and mean square error of the maximum likelihood estimators. Finally, the practical utility of the model was underscored through the examination of two real-world datasets. This served to illuminate the model's versatility and relevance in real-world scenarios, further establishing its potential as a valuable tool in statistical analysis and decision-making contexts.

Acknowledgement: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/01/25737).

References

[1] P. L. Gupta, R. C. Gupta and S. J. Lvin, Analysis op failure time data by burr distribution, Communications in Statistics-Theory and Methods, 25(9), 2013-2024, (1996).

[2] D. R. Wingo, Maximum likelihood methods for fitting the Burr type XII distribution to life test data, Biometrical journal, 25(1), 77-84, (1983).

[3] M. E. Ghitany and S. Al-Awadhi, Maximum likelihood estimation of Burr XII distribution parameters under random censoring, Journal of Applied Statistics, 29(7), 955-965, (2002).

[4] B. C. Arnold, Pareto Distributions, International Co-operative, Publishing House, Baltimore, MD, (1983).

[5] S. Kotz, N. Balakrishnan and N. L. Johnson, Continuous multivariate distributions, Volume 1: Models and applications, John Wiley & Sons, (1983).

[6] I. W. Burr, Cumulative frequency functions, The Annals of mathematical statistics, 13(2), 215-232, (1942).

[7] R. N. Rodriguez, A guide to the Burr type XII distributions, Biometrika, 64(1), 129-134, (1977).

[8] H. Krishna and P. S. Pundir, Discrete Burr and discrete Pareto distributions, Statistical methodology, 6(2), 177-188, (2009).

[9] H. Lee and J. H. Cha, On two general classes of discrete bivariate distributions, The American Statistician, 69(3), 221-230, (2015).

[10] V. Nekoukhou and D. Kundu, Bivariate discrete generalized exponential distribution, Statistics, 51(5), 1143-1158, (2017).

[11] D. Kundu and V. Nekoukhou, On bivariate discrete Weibull distribution, Communications in statistics-Theory and methods, 48(14), 3464-3481, (2019).

[12] D. Kundu and V. Nekoukhou, Univariate and bivariate geometric discrete generalized exponential distributions, Journal of Statistical Theory and Practice, 12, 595-614, (2018).

[13] S. Ali, M. Shafqat, I. Shah and S. Dey, Bivariate discrete Nadarajah and Haghighi distribution: properties and different methods of estimation, Filomat, 33(17), 5589-5610, (2019).

[14] M. S. Eliwa and M. El-Morshedy, Bivariate Gumbel-G family of distributions: statistical properties, Bayesian and non-Bayesian estimation with application, Annals of data science, 6(1), 39-60, (2019).

[15] M. S. Eliwa and M. El-Morshedy, Bayesian and non-Bayesian estimation of four-parameter of bivariate discrete inverse Weibull distribution with applications to model failure times, football and biological data, Filomat, 34(8), 2511-2531, (2020).

[16] M. H. Tahir, M. A. Hussain, G. M. Cordeiro, M. El-Morshedy and M. S, Eliwa, A new Kumaraswamy generalized family of distributions with properties, applications, and bivariate extension, Mathematics, 8(11), 1989, (2020).

[17] M. S. Eliwa, Z. A. Alhussain, E. A. Ahmed, M. M. Salah, H. H. Ahmed and M. El-Morshedy, Bivariate Gompertz generator of distributions: statistical properties and estimation with application to model football data, Journal of the National Science Foundation of Sri Lanka, 48(2), (2020).

[18] H. Lee and J. H. Cha, A new general class of discrete bivariate distributions constructed by using the likelihood ratio, Statistical Papers, 61, 923-944, (2020).

[19] M. El-Morshedy, M. S. Eliwa, A. El-Gohary and A. A. Khalil, Bivariate exponentiated discrete Weibull distribution: statistical properties, estimation, simulation and applications, Mathematical sciences, 14, 29-42, (2020).

[20] V. Nekoukhou, A. Khalifeh and H. Bidram, A bivariate discrete inverse resilience family of distributions with resilience marginals, Journal of Applied Statistics, 48(6), 1071-1090, (2021).

[21] R. Alotaibi, M. Khalifa, E. M. Almetwally and I. Ghosh, Classical and bayesian inference of a mixture of bivariate exponentiated exponential model, Journal of Mathematics, 2021, 1-20, (2021).

[22] M. El-Morshedy, M. H. Tahir, M. A. Hussain, A. Al-Bossly and M. S, Eliwa, A new flexible univariate and bivariate family of distributions for unit interval (0, 1), Symmetry, 14(5), 1040, (2022).

[23] A. Barbiero, Discrete analogues of continuous bivariate probability distributions, Annals of Operations Research, 312(1), 23-43, (2022).

[24] B. I. Mohammed, N. Makumi, R. Aldallal, T. E. Dyhoum and H. M. Aljohani, A New Model of Discrete-Continuous Bivariate Distribution with Applications to Medical Data, Computational and Mathematical Methods in Medicine, (2022).

[25] E. S. A. El-Sherpieny, H. Z. Muhammed and E. M. Almetwally, Data Analysis by Adaptive Progressive Hybrid Censored Under Bivariate Model, Annals of Data Science, 1-42, (2022a).

[26] E. S. A. El-Sherpieny, H. Z. Muhammed and E. M. Almetwally, Bivariate Chen distribution based on copula function: Properties and application of diabetic nephropathy, Journal of Statistical Theory and Practice, 16(3), 54, (2022b).

[27] E. S. A. El-Sherpieny, H. Z. Muhammed and E. M. Almetwally, Accelerated life testing for bivariate distributions based on progressive censored samples with random removal, J. Stat. Appl. Probab, 11(2), 203-223, (2022c).

[28] L. A. Al-Essa, M. S. Eliwa, H. S. Shahen, A. A. Khalil, H. N. Alqifari and M. El-Morshedy, Bivariate Discrete Odd Generalized Exponential Generator of Distributions for Count Data: Copula Technique, Mathematical Theory, and Applications, Axioms, 12(6), 534, (2023).

[29] N. G. Bismi and V. K. Ramachandran, Bivariate burr distributions (Doctoral dissertation, Cochin University of Science and Technology), (2005).

[30] C. S. Davis, Statistical methods for the analysis of repeated measures data, Springer-Verlag, New York, (2002).

[31] M. B. Casiena, Analysis of data collectively within the realm of physical conditions, Journal of Optics. In press, (2022).