International Journal of Mathematical Analysis and Modelling

(Formerly Journal of the Nigerian Society for Mathematical Biology)

Volume 8 Issue 1, 2025

ISSN (Print): 2682 - 5694
ISSN (Online): 2682 – 5708
www.tnsmb.org

Applied estimate for conditional bivariate Pareto distribution

S.S. Umar*, B. Onoghojobi†, A.L. Ozioko‡, and H.M. Rabiu§

Abstract

A bivariate conditional Pareto distribution is derived from the marginal and conditional Pareto distributions. Its joint properties are derived including its posterior and the prior distribution, methods for estimating its parameters are discussed. It is applied to a bivariate dataset of Fertilizer A and Fertilizer B, where the conditional dependence of the two variables is suspected. The result shows that the model is compatible with the data. It is proposed for application when there is a need to account for such a conditional structure in a bivariate model.

Keywords and phrases: Conditional Pareto distribution; Marginal distribution; Joint moment; Pareto bivariate; prior distribution.

1 Introduction

The Pareto distribution, introduced by Italian economist Vilfredo Pareto in the late 19th century, is a key model in statistics and economics, characterized by its power-law behavior and ability to describe phenomena where a small proportion of observations accounts for a significant portion of the total [3]. Originally applied to wealth distribution, the Pareto distribution has become a cornerstone in various fields, including finance, telecommunications, and environmental science, due to its heavy-tailed nature, which captures the occurrence of rare but impactful events [3]. The univariate Pareto model has been extensively utilized for modeling risk and insurance claims, providing insights into the distribution of extreme losses and helping in the design of robust financial products [9, 8]. Recent advancements in extreme value theory have further cemented its relevance, allowing researchers to understand better the dynamics of high-impact events and their implications in diverse sectors [4, 9].

As the need for more sophisticated statistical models increases, the extension of the Pareto distribution into a bivariate framework has gained traction, addressing scenarios where two related random variables exhibit heavy-tailed characteristics and interdependencies. The bivariate Pareto distribution enables joint modeling of such variables, offering valuable insights into complex systems where extreme values in one variable might influence or correlate with extreme values in another [1, 5]. This modeling approach has significant applications in risk management, particularly in finance, where it is crucial to analyze correlated losses and their potential impacts on portfolios [2, 7]. Furthermore, the application of copula models and multivariate extreme value theory enhances the bivariate Pareto distribution's capacity to capture dependencies between variables with heavy tails, thereby improving risk assessment and management strategies [10, 9]. Recent empirical studies have demonstrated the effectiveness of the bivariate Pareto model in various domains, including telecommunications and environmental data, where understanding the joint behavior of extreme events is critical [6].

*Department of Statistics, Federal Polytechnic Auchi, Edo State, Nigeria

†Department of Statistics, Federal University Lokoja, Nigeria; E-mail:benson.onoghojobi@fulokoja.edu.ng

‡Corresponding Author; Department of Mathematics, Federal University Lokoja, Nigeria; E-mail: arinze.luke@fulokoja.edu.ng

§Department of Statistics, Federal University Lokoja, Nigeria; E-mail: muhammedhas1995@gmail.com

2 Bivariate Pareto Distribution Model

The probability density function of the Pareto distribution is defined as:

$$f(x; x_m, \alpha) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \tag{1}$$

where:

- x: The variable of interest(e.g., income, wealth, or any other measure).
- x_m: The minimum value x can take.
- α: The shape parameter.

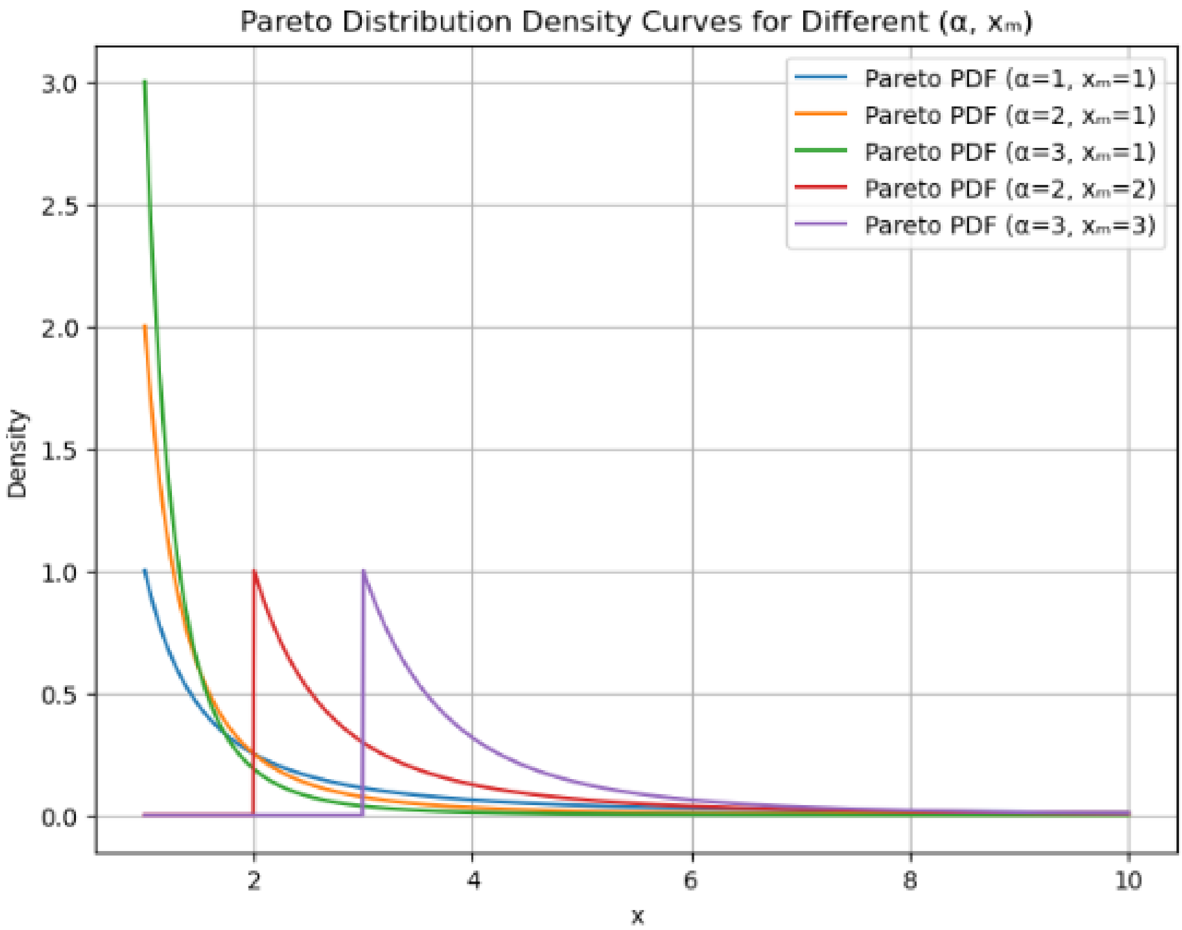

Figure 1: Density curve

Consider another random variable Y, assumed to depend on X, such that the distribution of Y given a realization of X at x is also a Pareto distribution. Then the probability density function of Y given $X = x$ is given by:

$$f_{Y|X}(y|x) = \frac{\beta y_m^\beta}{y^{\beta+1}} \tag{2}$$

where:

- y: The variable of interest (e.g., income, wealth, or any other measure).

- y_m: The minimum value Y can take given $X = x$.
- β: The shape parameter for the conditional distribution of Y.

If X and Y are independent and both have probability distributions, then the distribution of Y given a particular value x for X is represented by this conditional density function. while conducting various statistical analyses and wishing to comprehend how Y behaves while X is fixed at a specific value, this distribution can be helpful. We must first define the joint distribution of X and Y and then condition it on a particular value of X in order to produce a specific bivariate conditional pareto distribution. Let us take an example where X and Y have a combined distribution that is a bivariate pareto distribution. For two independent random variables X and Y that both follow normal distributions, the joint probability density function (PDF) can be written as the product of their individual pdfs.

$$f(x, y) = f_{X,Y}(x, y) = f(x) \cdot f_{Y|X}(y|x) \tag{3}$$

The BCPD is, therefore, given by:

$$f_{X,Y}(x, y) = \frac{\alpha x_m^\alpha \beta y_m^\beta}{x^{\alpha+1} y^{\beta+1}} \tag{4}$$

2.1 Theorem

Theorem 1: The relation

$$f_{X,Y}(x, y) = \frac{\alpha x_m^\alpha \beta y_m^\beta}{x^{\alpha+1} y^{\beta+1}} \tag{5}$$

This function is a true bivariate probability density function for $x, y \in ((x_m, y_m), \infty)$.
Proof: Let

$$M = \int_0^\infty \int_0^\infty f(x, y)\, dx\, dy. \tag{6}$$

then,

$$M = \int_{x_m}^\infty \int_{y_m}^\infty \frac{\alpha x_m^\alpha \beta y_m^\beta}{x^{\alpha+1} y^{\beta+1}}\, dx\, dy \tag{7}$$

We can rewrite M as follows:

$$M = \alpha x_m^\alpha \beta y_m^\beta \int_{x_m}^\infty \int_{y_m}^\infty \frac{1}{x^{\alpha+1} y^{\beta+1}}\, dx\, dy.$$

Since the integrand is a product of functions of x and y, we can separate the double integral into the product of two single integrals:

$$M = \alpha x_m^\alpha \beta y_m^\beta \left(\int_{x_m}^\infty \frac{1}{x^{\alpha+1}}\, dx \right) \left(\int_{y_m}^\infty \frac{1}{y^{\beta+1}}\, dy \right).$$

First, we evaluate the integral:

$$\int_{x_m}^\infty \frac{1}{x^{\alpha+1}}\, dx.$$

This is a standard integral. The antiderivative of $x^{-\alpha-1}$ is:

$$-\frac{1}{\alpha} x^{-\alpha}.$$

Thus, we have:

$$\int_{x_m}^\infty \frac{1}{x^{\alpha+1}}\, dx = \left[-\frac{1}{\alpha} x^{-\alpha} \right]_{x_m}^\infty = 0 + \frac{1}{\alpha x_m^\alpha} = \frac{1}{\alpha x_m^\alpha}.$$

Now, we evaluate the integral:

$$\int_{y_m}^{\infty} \frac{1}{y^{\beta+1}}\, dy.$$

Using a similar approach, the antiderivative is:

$$-\frac{1}{\beta} y^{-\beta}.$$

Thus, we have:

$$\int_{y_m}^{\infty} \frac{1}{y^{\beta+1}}\, dy = \left[-\frac{1}{\beta} y^{-\beta}\right]_{y_m}^{\infty} = 0 + \frac{1}{\beta y_m^{\beta}} = \frac{1}{\beta y_m^{\beta}}.$$

Substituting these results back into the expression for M:

$$M = \alpha x_m^{\alpha} \beta y_m^{\beta} \left(\frac{1}{\alpha x_m^{\alpha}}\right)\left(\frac{1}{\beta y_m^{\beta}}\right).$$

Simplifying the expression:

$$M = \alpha x_m^{\alpha} \beta y_m^{\beta} \cdot \frac{1}{\alpha x_m^{\alpha}} \cdot \frac{1}{\beta y_m^{\beta}} = 1.$$

Thus, we conclude that:

$$M = 1.$$

Since $M = 1$, the relation (5) is a true bivariate probability density function.

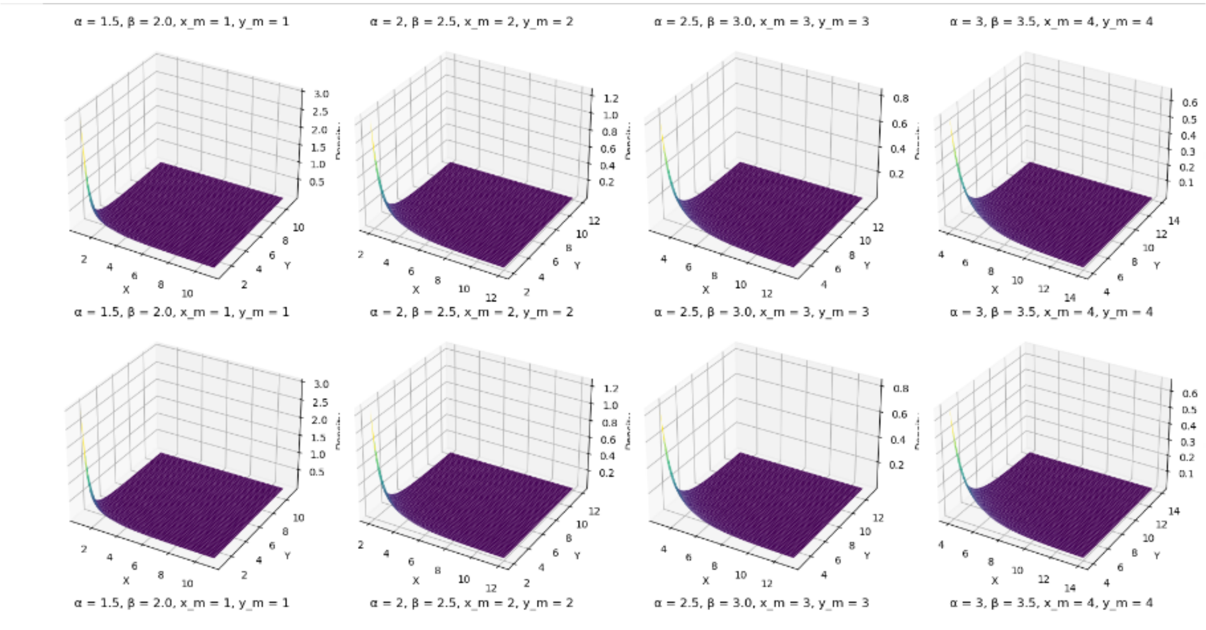

Figure 2: Bivariate Pareto distribution plot

2.2 Generalization

In a more generalized setting, we can introduce:

- **Additional Variables:** Extend the model to include n variables, where each follows a Pareto distribution, leading to a multivariate Pareto distribution.
- **Dependency Structure:** Consider copulas to represent the dependency structure between X and Y more flexibly. This allows for various forms of correlation and dependence.

The generalized joint PDF can be expressed as:

$$f_{X,Y}(x, y; \theta) = c(\theta) \cdot \frac{\alpha x_m^{\alpha} \beta y_m^{\beta}}{x^{\alpha+1} y^{\beta+1}} \cdot \mathbb{I}(x \geq x_m, y \geq y_m),$$

where $c(\theta)$ is a normalization constant that depends on the parameter θ representing the correlation structure between X and Y.

3 Properties of Pareto Distribution

3.1 Marginal Distribution of Y

Theorem 2
The marginal probability density function (PDF) of Y from the joint PDF $f_{X,Y}(x, y)$, we integrate out X:

$$f_Y(y) = \int_{x_m}^{\infty} f_{X,Y}(x, y)\, dx.$$

Proof:
The joint PDF is given by:

$$f_{X,Y}(x, y) = \frac{\alpha x_m^{\alpha} \beta y_m^{\beta}}{x^{\alpha+1} y^{\beta+1}} \quad \text{for } x \geq x_m \text{ and } y \geq y_m.$$

Setting up the integral for the marginal of Y:

$$f_Y(y) = \int_{x_m}^{\infty} \frac{\alpha x_m^{\alpha} \beta y_m^{\beta}}{x^{\alpha+1} y^{\beta+1}}\, dx.$$

Factoring out constants:

$$f_Y(y) = \frac{\alpha x_m^{\alpha} \beta y_m^{\beta}}{y^{\beta+1}} \int_{x_m}^{\infty} \frac{1}{x^{\alpha+1}}\, dx.$$

We compute the integral:

$$\int_{x_m}^{\infty} \frac{1}{x^{\alpha+1}}\, dx = \frac{1}{\alpha x_m^{\alpha}}.$$

Substituting this result back, we have:

$$f_Y(y) = \frac{\alpha x_m^{\alpha} \beta y_m^{\beta}}{y^{\beta+1}} \cdot \frac{1}{\alpha x_m^{\alpha}}.$$

This simplifies to:

$$f_Y(y) = \frac{\beta y_m^{\beta}}{y^{\beta+1}} \quad \text{for } y \geq y_m.$$

Thus, the marginal PDF of Y is:

$$f_Y(y) = \begin{cases} \frac{\beta y_m^{\beta}}{y^{\beta+1}} & \text{for } y \geq y_m, \\ 0 & \text{otherwise.} \end{cases}$$

3.2 Posterior Distribution

To derive the posterior distribution of the bivariate Pareto distribution, we need to combine the likelihood of the observed data with a prior distribution on the parameters.
The joint probability density function (PDF) of the bivariate Pareto distribution is given by:

$$f_{X,Y}(x,y) = \begin{cases} \frac{\alpha x_m^\alpha \beta y_m^\beta}{x^{\alpha+1} y^{\beta+1}}, & x \geq x_m,\, y \geq y_m \\ 0, & \text{otherwise} \end{cases}$$

where:

- x and y are the random variables,
- x_m and y_m are the minimum values,
- α and β are the shape parameters.

Likelihood Function

Given a sample of size n from the bivariate Pareto distribution, the likelihood function based on the observed data $D = \{(x_i, y_i)\}_{i=1}^n$ is:

$$L(\alpha, \beta | D) = \prod_{i=1}^{n} \frac{\alpha x_m^\alpha \beta y_m^\beta}{x_i^{\alpha+1} y_i^{\beta+1}}.$$

Taking the logarithm of the likelihood function:

$$\log L(\alpha, \beta | D) = n \log(\alpha) + n \log(\beta) + n \log(x_m^\alpha y_m^\beta) - (\alpha + 1) \sum_{i=1}^{n} \log(x_i) - (\beta + 1) \sum_{i=1}^{n} \log(y_i).$$

Prior Distribution

Assume conjugate priors for α and β, where:

$$\alpha \sim \text{Gamma}(\alpha_0, \beta_0), \quad \beta \sim \text{Gamma}(\alpha_1, \beta_1).$$

Posterior Distribution

Using Bayes' theorem, the posterior distribution is given by:

$$f(\alpha, \beta | D) \propto L(\alpha, \beta | D) \cdot f(\alpha) \cdot f(\beta).$$

Substituting the likelihood and prior:

$$f(\alpha, \beta | D) \propto \left(\prod_{i=1}^{n} \frac{\alpha x_m^\alpha \beta y_m^\beta}{x_i^{\alpha+1} y_i^{\beta+1}} \right) \cdot \left(\alpha^{\alpha_0 - 1} e^{-\beta_0 \alpha} \right) \cdot \left(\beta^{\alpha_1 - 1} e^{-\beta_1 \beta} \right).$$

Simplifying the posterior distribution:

$$f(\alpha, \beta | D) \propto \alpha^{\alpha_0 + n - 1} \cdot \beta^{\alpha_1 + n - 1} \cdot e^{-\beta_0 \alpha - \beta_1 \beta} \cdot \left(\prod_{i=1}^{n} x_m^\alpha y_m^\beta \right) \cdot \left(\prod_{i=1}^{n} \frac{1}{x_i^{\alpha+1} y_i^{\beta+1}} \right).$$

Thus, the posterior distributions for α and β are:

$$\alpha | D \sim \text{Gamma}(\alpha_0 + n, \beta_0 + \sum_{i=1}^{n} \log(x_i)),$$

$$\beta | D \sim \text{Gamma}(\alpha_1 + n, \beta_1 + \sum_{i=1}^{n} \log(y_i)).$$

4 Method of Parameter Estimation

The parameters of the BCBD can be estimated using the maximum likelihood method of estimation. From the density in (5), For a sample of size n, the likelihood function $L(\alpha, \beta|D)$ based on observed data $D = \{(x_i, y_i)\}_{i=1}^n$ is:

$$L(\alpha, \beta|D) = \prod_{i=1}^n f_{X,Y}(x_i, y_i) = \prod_{i=1}^n \frac{\alpha x_m^\alpha \beta y_m^\beta}{x_i^{\alpha+1} y_i^{\beta+1}}.$$

This can be rewritten as:

$$L(\alpha, \beta|D) = \left(\alpha x_m^\alpha \beta y_m^\beta\right)^n \prod_{i=1}^n \frac{1}{x_i^{\alpha+1} y_i^{\beta+1}}.$$

Taking the logarithm of the likelihood function:

$$\log L(\alpha, \beta|D) = n\log(\alpha) + n\log(\beta) + n\log(x_m^\alpha y_m^\beta) -$$

$$(\alpha+1)\sum_{i=1}^n \log(x_i) - (\beta+1)\sum_{i=1}^n \log(y_i).$$

Simplifying gives:

$$\log L(\alpha, \beta|D) = n\log(\alpha) + n\log(\beta) + n\alpha\log(x_m) + n\beta\log(y_m) -$$

$$(\alpha+1)\sum_{i=1}^n \log(x_i) - (\beta+1)\sum_{i=1}^n \log(y_i).$$

Partial Derivatives

To find the MLE, take the partial derivatives of the log-likelihood function with respect to α and β, and set them to zero.

Partial Derivative with Respect to α

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + n\log(x_m) - \sum_{i=1}^n \log(x_i) = 0.$$

Rearranging gives:

$$\frac{n}{\alpha} + n\log(x_m) = \sum_{i=1}^n \log(x_i).$$

Solving for α:

$$\alpha = \frac{n}{\sum_{i=1}^n \log(x_i) - n\log(x_m)}.$$

Partial Derivative with Respect to β

$$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} + n\log(y_m) - \sum_{i=1}^n \log(y_i) = 0.$$

Rearranging gives:

$$\frac{n}{\beta} + n\log(y_m) = \sum_{i=1}^n \log(y_i).$$

Solving for β:

$$\beta = \frac{n}{\sum_{i=1}^n \log(y_i) - n\log(y_m)}.$$

5 Real data application

In this study, we analyzed crop yield data from two different fertilizer regimens using the Bivariate Pareto Distribution to model the relationship between the yields. The dataset comprised 30 observations each for Fertilizer A and Fertilizer B, with yields recorded in kilograms per hectare. We then estimated the distribution parameters using Maximum Likelihood Estimation (MLE). The resulting parameters provided insights into the yield patterns of the two fertilizers. Fertilizer A showed variability with a tendency towards higher yields, but also a significant probability of lower yields, indicating a broad distribution. Fertilizer B, on the other hand, displayed a sharp concentration around higher yields, suggesting it is more consistent in producing high yields. The Kolmogorov-Smirnov (K-S) test is used to compare a sample distribution with a reference probability distribution. the K-S test values for Fertilizer A (0.5974) and Fertilizer B (0.4464) indicate how well the data fit the Bivariate Conditional Pareto distribution.both fertilizers show a moderate fit with the Bivariate Pareto distribution, with Fertilizer B having a good fit. A strong positive correlation of 0.995 between fertilizer A and fertilizer B, with an extremely small p-value of $3.81x10^{-29}$ indicating that the correlation is statistically significant, providing strong evidence that the relationship between the two variables is not due to random chance.

Dataset:

'Fertilizer A Yield (kg/ha)': 3600, 4050, 3450, 2750, 4400, 3650, 4550, 3100, 3950, 3350, 3700, 4200, 3300, 3150, 4500, 2850, 4050, 3800, 2950, 3600, 4300, 3100, 3400, 3650, 3850, 2900, 4000, 3450, 3700, 4100 ,

'Fertilizer B Yield (kg/ha)': 3250, 3900, 3000, 2250, 4100, 3400, 4250, 2700, 3600, 2900, 3450, 3950, 2900, 2750, 4150, 2400, 3700, 3450, 2600, 3200, 3950, 2650, 3000, 3300, 3500, 2500, 3600, 3100, 3350, 3800

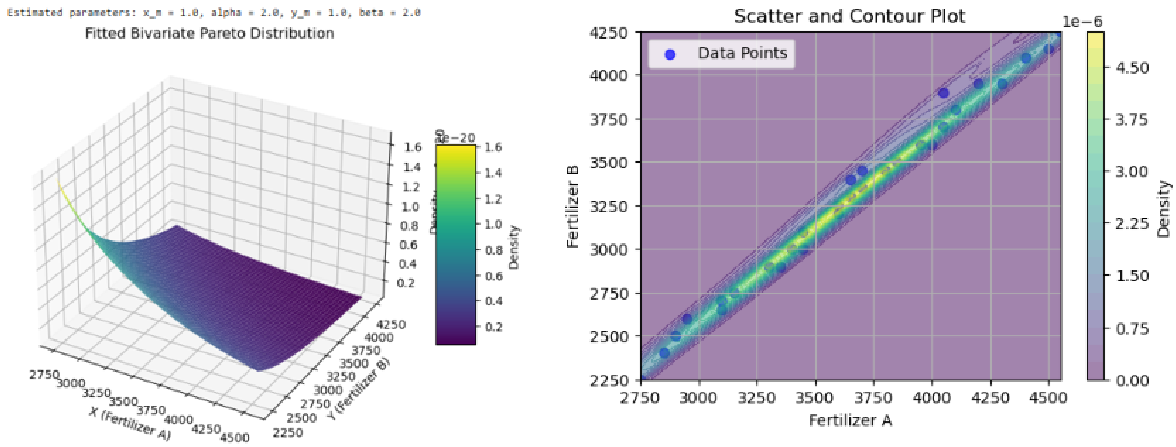

Figure 3: the bivariate plot and the contour plot of fertilizer A and B

The bivariate plot of the Pareto distribution applied to the dataset of Fertilizer A and Fertilizer B usage provides a clear visualization of how the quantities of both fertilizers are distributed jointly. The 3D surface plot shows the density of observations across different combinations of Fertilizer A (X-axis) and Fertilizer B (Y-axis). The steep slope in the surface indicates a higher density of observations at the lower levels of both fertilizers, consistent with the heavy-tailed nature of the Pareto distribution, where smaller quantities of both fertilizers are more common. As the values of Fertilizer A and Fertilizer B increase, the density decreases significantly, suggesting that larger quantities of fertilizers are less frequently observed. The contour plot further reinforces this by displaying contours of constant density, which clearly show how the probability density diminishes as the usage levels of both fertilizers rise. The contour lines reflect the sparsity of extreme values, indicating that the majority of observations are concentrated in the lower ranges of Fertilizer A and B usage, while larger quantities are rare. Together, these plots illustrate how the bivariate Pareto distribution effectively models the joint distribution of Fertilizer A and Fertilizer B usage, capturing the tendency for small amounts of both fertilizers to be more prevalent and the rarity of higher usage quantities.

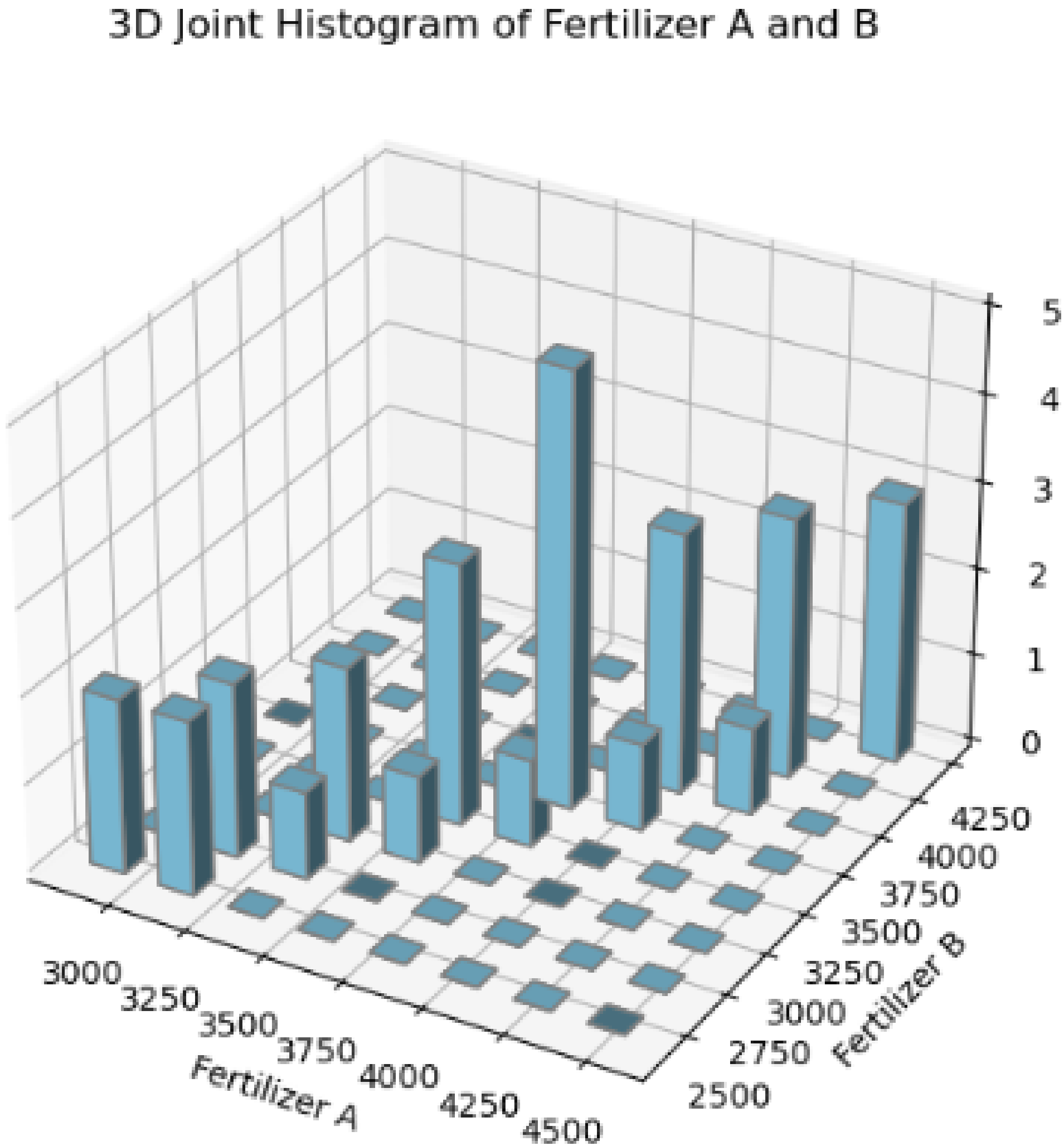

Figure 4: The Data histogram for fertilizer A and B

6 Conclusion

In this work, we have thoroughly explored the application of the bivariate Pareto distribution, focusing on its estimation and visualization through maximum likelihood estimation (MLE) for fitting parameters to a given dataset. We began by defining univariate and bivariate Pareto probability density functions (PDFs) for two variables representing Fertilizer A and B. Using the dataset, we employed the MLE technique to estimate key parameters such as the shape parameters (α, β) and scale parameters (x_m, y_m). The estimated parameters were then used to visualize the bivariate Pareto distribution through a 3D surface plot, effectively illustrating the joint behavior of the two variables. The results provided valuable insights into the dependence structure of the data, assuming independence between the two variables. Based on these findings, we recommend conducting further validation of the model by comparing it to alternative distributions and using larger, more diverse datasets to assess the robustness of the bivariate Pareto model.

References

[1] Asimit, A. V., & Vernic, R. (2020). Bivariate Pareto distributions: a review and applications. *Statistical Modelling*, 20(5), 330-352.

[2] Beirlant, J., Goegebeur, Y., Kauffman, R. J., & Teugels, J. L. (2004). *Statistics of Extremes: Theory and Applications*. Wiley 3(12) 11-19.

[3] Embrechts, P., Klappelberg, C., & Mikosch, T. (2012). *Modelling Extremal Events for Insurance and Finance*. Springer.4(22), 34-37

[4] Falk, M., Klose, M., & Reiss, R. D. (2021). *Statistical Extreme Value Theory: An Introduction*. Springer.

[5] Ghosh, D., & Sinha, A. (2021). Bivariate Pareto distribution: Statistical properties and applications. *Communications in Statistics: Theory and Methods*, 50(4), 871-885.

[6] Ghosh, D., Ranjan, A., & Sinha, A. (2023). Applications of Bivariate Pareto Distribution in Telecommunications Data. *Journal of Applied Statistics*, 50(2), 254-271.

[7] Kauffman, R. J., & Wang, Z. (2019). Joint modeling of extreme risks: An application of bivariate Pareto distribution. *Journal of Risk and Financial Management*, 12(2), 78.

[8] Klappelberg, C., Lindner, A., & Maller, R. (2019). *Stochastic Modelling and Statistical Inference of Extreme Events*. Springer.

[9] Moehrle, M. G., Schmid, A., & Ziegler, A. (2021). Modeling Extreme Losses with a Pareto Distribution: Applications in Risk Management. *Insurance: Mathematics and Economics*, 100, 210-221.

[10] Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Economic Surveys*, 26(4), 636-659.