

UNIVERSITY OF CAPE COAST

The background of the title page features a large, faint watermark of the University of Cape Coast crest. The crest is a shield with a red top section containing a yellow eagle with spread wings. Below the eagle is a white section with a yellow sun-like symbol containing a red stylized figure. The bottom section of the shield is blue with white wavy lines. A red ribbon banner wraps around the bottom of the shield with the Latin motto 'VERITAS LUMEN NOBIS' in white capital letters.

AN ALTERNATIVE MULTIVARIATE EXPONENTIAL
POWER DISTRIBUTION

MICHAEL MANFORD

2023



© Michael Manford

University of Cape Coast

UNIVERSITY OF CAPE COAST

AN ALTERNATIVE MULTIVARIATE EXPONENTIAL
POWER DISTRIBUTION

BY

MICHAEL MANFORD

Thesis submitted to the Department of Statistics of the School of Physical
Sciences, College of Agriculture and Natural Sciences, University
of Cape Coast, in partial fulfilment of the requirements for
the award of Doctor of Philosophy degree in Statistics

DECEMBER 2023

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature Date

Name: Michael Manford

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature Date

Name: Prof. Bismark K. Nkansah

Co-Supervisor's Signature Date:

Name: Dr. Arimiyaw Zakaria

ABSTRACT

In the existing multivariate Exponential Power Distribution (MEPD), the shape parameter has the tendency to affect the variability in the data leading to distorted features of the distribution. This study resolves this problem by obtaining a generalized expression in place of the constant coefficient (0.5) of the quadratic form of the Multivariate EPD as a function of the shape parameter, such that the variance-covariance matrix of the data is exactly equal to the scale parameter. The alternative Multivariate EPD (AMEPD) thus obtained is a more generalized multivariate EPD. The maximum likelihood estimation (MLE) of the parameters of the distribution are presented via Newton-Raphson method. The estimates based on the proposed distribution is compared to other methods using information criteria. Results consistently indicate superior performance of the AMEPD to the MEPD for various sample sizes based on both simulated and empirical datasets. The study further explores the application of the new distribution in discriminant analysis under the assumptions of equal and unequal variance-covariance matrices. The application also underscores the importance of accurate specification of the variance-covariance matrix for the desired predictive performance. The proposed approach has the potential to provide a more robust statistical model for complex multivariate distributions that are close to symmetry as it provides greater flexibility in depicting the true underlying characteristics of the distributions.

KEY WORDS

Characteristic Function

Discriminant Analysis

Elliptical Symmetric Distribution

Exponential Power Distribution

Parameter Estimation

Stochastic Representation



ACKNOWLEDGEMENTS

The realization of this study owes much to the indispensable support and contributions of a diverse group of individuals. While it is not feasible to enumerate each person, I wish to convey my heartfelt gratitude to Prof. Bismark K. Nkansah, who served as my Principal Supervisor in the Department of Statistics at the University of Cape Coast (UCC). His pivotal role significantly contributed to the successful completion of this research project. I deeply appreciate his invaluable insights, constructive comments, and unwavering guidance that spanned every phase of this undertaking. Furthermore, I extend my thanks to my co-supervisor, Dr. Arimiyaw Zakaria, for his continual encouragement and enlightening contributions throughout the entire research process.

I extend my acknowledgment and gratitude to all lecturers and staff members from both the Department of Mathematics and the Department of Statistics at UCC for their substantial support and contributions. Special recognition is also extended to my friends and colleagues at Cape Coast Technical University, as well as all those who offered steadfast support throughout my academic journey. I extend my heartfelt gratitude with special recognition to my wife, Roseline, and my children, whose unwavering support, boundless love, and incredible patience have been invaluable throughout this journey. Finally, I wish to express my profound appreciation to my entire family and circle of friends for their unwavering and steadfast support throughout the entire duration of my academic program.

DEDICATION

To my family and friends.



TABLE OF CONTENTS

	Page
DECLARATION	ii
ABSTRACT	iii
KEY WORDS	iv
ACKNOWLEDGEMENTS	v
DEDICATION	vi
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xiv
CHAPTER ONE: INTRODUCTION	
Background to the Study	1
Statement of the Problem	7
Purpose of the Study	10
Objectives of the Study	10
Significance of the Study	11
Delimitation	12
Limitation	13
Organisation of the Study	14
CHAPTER TWO: LITERATURE REVIEW	
Introduction	16
The Concept of Spherically and Elliptically Symmetric Distributions	16

Some Sub-Classes of Elliptical Distributions	20
Review of Some Applications of Multivariate EPD	68
Chapter Summary	75
CHAPTER THREE: RESEARCH METHODS	
Introduction	78
Research Design and Data Characterization	78
Multivariate Dataset	79
Mathematical Preliminaries	83
Definition of Spherical and Elliptical Symmetric Distribution	90
Properties of Spherical and Elliptical Symmetric Distribution	92
Uniform Distribution in the Unit Sphere	99
Distribution of the Quadratic ($\mathbf{R}^2$) Form of Elliptical and Spherical Dis- tribution	104
Some Properties and the Derivative of the Multivariate Normal Distribution	105
Properties of MEPD	111
Method of Deriving Spherical and Elliptical Symmetric Distribution	113
Approaches to Parameter Estimation	118
The Newton-Raphson Optimization Method for Estimating the Shape Pa- rameter	123
Model Selection Techniques for Assessing the Efficiency of the Distribu- tion in Fitting Dataset	129
The Application of Multivariate EPD in Multivariate Data Analysis	133
The Basic Concept of Discriminant and Classification Analysis	133

Chapter Summary	144
CHAPTER FOUR: RESULTS AND DISCUSSION	
Introduction	150
Derivation of AMEPD	150
Properties of the AMEPD	154
Deductions from the AMEPD	161
Bivariate Plots of the AMEPD	161
Simulation Results	164
Real Data Example	168
Maximum Likelihood Estimation of the Parameters of the AMEPD	171
MLE of μ	172
MLE of Σ	173
MLE of q	174
Algorithm (Pseudo-Code) of the MLE	177
Implementation of the MLE	177
Application to Simulated Data	177
MLE based on Datasets	181
Application of the AMEPD in Discriminant Analysis	189
The Discriminant function	190
Demonstration of Discriminant Analysis to Multivariate Dataset	192
Discussion	196
Chapter Summary	199

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATION

Overview	202
Summary	202
Conclusions	207
Recommendations	208
REFERENCES	211
APPENDICES	253
APPENDIX A: R STUDIO CODE FOR AMEPD DENSITY COMPU- TATION	253
APPENDIX B: R STUDIO CODE FOR SIMULATION FROM AMEPD	255
APPENDIX C: R STUDIO CODE FOR AMEPD PARAMETER ESTI- MATION	257
APPENDIX D: R STUDIO CODE FOR PLOTTING 3D DENSITY OF BIVARIATE AMEPD	262
APPENDIX E: R STUDIO CODE FOR DISCRIMINANT ANALYSIS FOR AMEPD POPULATION	265

LIST OF TABLES

	Page
1 Summary of Some Class of Elliptical Distributions	67
2 Confusion Matrix for the Cost of Misclassification	137
3 Confusion Matrix for Observations from Known Populations	144
4 Summary Statistics for Simulated Data for $q = 0.5$	165
5 Summary Statistics for Simulated Data for $q = 1$	166
6 Summary Statistics for Simulated Data for $q = 2$	166
7 Summary Statistics for Simulated Data for $q = 5$	167
8 Summary Statistics for Simulated Data for $q = 10$	167
9 Summary Statistics Data Generated for Price and Specification of Cars Using AMEPD	169
10 Summary Statistics Data Generated for Price and Specification of Cars Using MEPD	170
11 Summary of Parameter Estimation from Simulated Data for $q=0.5$	178
12 Summary of Parameter Estimation from Simulated Data for $q=1$	179
13 Summary of Parameter Estimation from Simulated Data for $q=2$	180
14 Confusion Matrix for Type of Car from AMEPD Populations for Equal Variance-Covariance Matrix	196
15 Confusion Matrix for Type of Car from AMEPD Populations for Unequal Variance-Covariance Matrix	196

LIST OF FIGURES

	Page
1 A Plot of Bivariate Laplace Distribution	23
2 Bivariate Uniform Distribution	27
3 Bivariate Symmetric Kotz Type Distribution.	31
4 Bivariate Plot of Pearson Type II Distribution	39
5 Bivariate Plot of Student's t -Distribution.	42
6 Plot of the Spherical Bessel Distribution of a) the First Kind and b) the Second Kind	44
7 Bivariate Plot of the Spherical-Dirichlet Distribution	47
8 Bivariate Plot of Cauchy Distribution	51
9 Bivariate Logistic Distribution	54
10 Bivariate Normal Distribution Plot	57
11 Gamma Function $[\Gamma(\alpha)]$ across the entire Complex Plane.	85
12 First Illustration of Newton's Method	125
13 Second Illustration of Newton's Method	126
14 Multivariate EPD for $q = 0.5$.	162
15 Multivariate EPD for $q = 1$.	162
16 Multivariate EPD for $q = 2$.	163
17 Multivariate EPD for $q = 5$.	163
18 Multivariate EPD for $q = 10$.	164
19 Distribution of Variables in Carsales Dataset	182

20	Distribution of Variables in River Pra Data	185
21	Distribution of Variables in Concrete Dataset	187
22	Distribution of Six Car Descriptors by Type of Car	193



LIST OF ABBREVIATIONS



AER	Actual Error Rate
AIC	Akaike Information Criterion
AMEPD	Alternative Multivariate Exponential Power Distribution
APER	Average Performance Error Rate
BIC	Bayesian Information Criterion
CDF	Cumulative Density Function
CF	Characteristic Function
ECM	Expected Cost of Misclassification
EPD	Exponential Power Distribution
GED	Generalized Error Distribution
IID	Independent and Identically Distributed
KLD	Kullback-Leibler Divergence
LDA	Linear Discriminant Analysis
LDF	Linear Discriminant Function
MCMC	Markov chain Monte Carlo
MEPD	Multivariate Exponential Power Distribution
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MGF	Moment Generating Function
PDF	Probability Density Function
QDA	Quadratic Discriminant Analysis

CHAPTER ONE

INTRODUCTION

In the last twenty years, researchers have increasingly focused on creating flexible parametric classes of statistical distributions that provide more adaptability compared to the Normal distribution. In various fields such as health, environment, and finance, there has been a push to utilize distributions that offer increased robustness in modeling data displaying deviations from normality. In their comprehensive review, Rigby, Stasinopoulos, Heller, and De Bastiani (2019) have examined a wide range of parametric distribution classes. Exponential Power Distribution (EPD), introduced by Subbotin (1923), is a distribution class thought to have more kurtosis flexibility than the Normal distribution. The multivariate case was the extension of this concept by Gómez, Gómez-Villegas, and Marín (1998). This present study puts forward an alternative Multivariate Exponential Power Distribution (AMEPD) that surpasses the tails' flexibility compared to Gómez et al. (1998).

Background to the Study

Progress in research across diverse fields has generated real-world data that occasionally defies modeling by conventional probability distributions. Consequently, ongoing research remains imperative for the creation of novel distribution classes. These new classes often serve as generalizations or extensions of existing ones, with the primary aim of enhancing the flexibility of established

distribution models. Mostly, models in both univariate and multivariate techniques use the Normal distribution as the usual assumption for hypothesis testing in statistical inference. Unfortunately, the multivariate Normal distribution is often found to provide inadequate fit in many practical situations. Again, in practice, for a continuous response variable, the shape of the distribution can be much more highly platykurtic or leptokurtic (i.e. lighter or heavier tails) than the multivariate Normal distribution. Usually, the appropriate alternative approach is to look for suitable distributional models that are more general rather than the Gaussian one. One of such distributions is the univariate Exponential Power Distribution (EPD). Subbotin (1923) introduced this distribution, constituting a category of distributions perceived to possess greater flexibility than the Normal distribution, particularly in terms of kurtosis. The family of distributions associated with Subbotin's EPD is expressed as

$$f(\varepsilon) = \frac{mh}{2\Gamma(1/m)} \exp[-h^m |\varepsilon|^m], \quad (1)$$

with $-\infty < \varepsilon < \infty$, $h > 0$ and $m \geq 1$. Diananda (1949), Turner (1960), Zeckhauser and Thompson (1970), Azzalini (1985), McDonald and Newey (1988), and Mineo and Ruggieri (2005) are some of the later references.

This class of distributions is also known as the General Error Distribution (GED), noted for its symmetry and adjustable kurtosis. It is alternatively referred to as the Power-Exponential distribution (McDonald & Newey 1988; Johnson, Kotz & Balakrishnan, 1995) and the Box-Tiao distribution (McDonald & Newey 1988). Notably, this distribution shares many structural properties with the Normal distribution, earning it the moniker of General Error Distribu-

tion. Génc (2007) has contributed significantly to this field by developing linear regression models and time series models that incorporate heavy-tailed characteristics, assuming the underlying distribution to be GED. The GED has been a subject of comprehensive exploration in the literature, with Kalke and Richter (2013) and Nadarajah (2005) delving into various distributional properties associated with it. This distribution is particularly versatile, finding application as a general distribution for random errors, as highlighted by Mineo and Ruggieri (2005). Univariate GED has been widely applied across diverse domains, showcasing its adaptability and relevance. Researchers such as Coin (2017), Giller (2005), Nadarajah (2005), Pogány and Nadarajah (2009), and Tahir et al. (2015) have explored different forms of univariate GED in their studies, emphasizing its utility in capturing a range of data patterns. In the field of statistical tools and methodologies, Mineo and Ruggieri (2005) introduced the "normalp package," specifically designed for the statistical environment R. This package provides a set of tools tailored for handling the univariate GED, further facilitating the analysis and interpretation of data exhibiting characteristics aligned with this distribution. The widespread applications and dedicated statistical tools underscore the significance of the GED in modeling random errors and capturing nuanced patterns in data. Its flexibility and resemblance to the Normal distribution make it a valuable tool in various statistical analyses, contributing to advancements in various fields such as economics, finance, and engineering.

The multivariate EPD stands as a significant subset within the broader family of Elliptical Contoured Distributions or Spherical Distributions. Schol-

ars and statisticians have delved into the intricacies of these distributions, marking the multivariate EPD as a burgeoning area of research in recent years. This family of distributions, with roots traced back to foundational works by Kelker (1970), Cambanis, Huang, and Simons (1981), Watson (1984, 1985), Fang, Kotz, and Ng (1991) and Fang and Anderson (1992), provides a rich framework for modeling multivariate data in various fields.

The concept of Elliptical Contoured Distributions revolves around the idea of data distributions that are invariant under linear transformations. Fang et al. (1991) extensively categorized elliptical distributions into various subclasses, each possessing distinct characteristics. These subclasses include multiuniform distributions, symmetric Kotz type distributions, symmetric multivariate Pearson Types VII and II distributions, and the Symmetric Multivariate Laplace Distribution. Additionally, subclasses like the multivariate EPD, symmetrical-Bessel Distribution, multivariate Student's t -Distribution, multivariate Cauchy Distribution, multivariate Logistic Distribution, and multivariate Normal Distribution contribute to this diverse classification. These subclasses cover a range of mathematical formulations designed to capture specific patterns and features present in data distributions. For example, multiuniform distributions are suitable for scenarios involving the coexistence of multiple uniform distributions, while symmetric Kotz type distributions and multivariate Pearson Types VII and II distributions are effective in modeling symmetric data patterns. The Symmetric Multivariate Laplace Distribution is particularly well-suited for situations where data exhibits Laplacian characteristics. Researchers like Rabbani,

Vafadust, Gazor and Selesnick (2006), Eltoft, Kim and Lee (2006b), Yu and Yin (2023) and many others have made significant contributions to comprehending these subclasses. Their work encompasses applications, mathematical properties, and statistical implications associated with elliptical distribution subclasses, thereby enhancing the overall understanding in this field.

In the case of the multivariate EPD, it inherits these properties while also being characterized by an exponential power form. This distribution has gained prominence due to its ability to capture a wide range of shapes and patterns in multivariate data, making it applicable in diverse domains such as finance, biology, and engineering. Researchers and practitioners have been drawn to the multivariate EPD for its flexibility and suitability in modeling complex data structures. The distribution's mathematical elegance, coupled with its capacity to represent heavy-tailed and skewed data, has sparked interest in exploring its applications in various statistical methodologies. Understanding the properties and behaviors of the multivariate EPD is essential for developing robust statistical models, particularly in scenarios where traditional assumptions may not hold.

As an emerging area of research, the exploration of the multivariate EPD involves investigating its statistical properties, estimation methods, and applications across different disciplines. Scholars are actively contributing to the development of methodologies that leverage the multivariate EPD to address challenges in data analysis and interpretation. This growing body of research not only enhances our understanding of the distribution itself but also expands

the toolkit available for statisticians and data scientists working with multivariate data, contributing to advancements in both theoretical and applied statistics.

The multivariate EPD for p -dimensional random variables is first proposed by Gómez et al. (1998), which is given by

$$f(\mathbf{x}|\mu, \Sigma, \beta) = \frac{p\Gamma(\frac{p}{2})|\Sigma|^{-\frac{1}{2}}}{\pi^{\frac{p}{2}}\Gamma(1+\frac{n}{2\beta})2^{1+\frac{p}{2\beta}}} \times \exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\mu)^{\top}\Sigma^{-1}(\mathbf{x}-\mu)\right]^{\beta}\right\}, \quad (2)$$

along with the stochastic representation, the probability characteristics and other statistical properties. The distribution has a variance-covariance matrix defined as

$$D(\mathbf{x}) = \frac{2^{\frac{1}{\beta}}\Gamma\left(\frac{p+2}{2\beta}\right)}{p\Gamma\left(\frac{p}{2\beta}\right)}\Sigma, \quad (3)$$

and the mean vector as $E(\mathbf{x}) = \mu$. The multivariate EPD random variable of the Gómez et al. (1998) may be denoted by $\mathbf{x} \sim \text{MEPD}(\mu, \Sigma, \beta)$. In this study, the distribution in Equation (2) will be referred to as MEPD throughout.

This particular variant of multivariate EPD under consideration has served as a foundational element in numerous studies and diverse applications. Researchers across various domains have explored and applied this form of multivariate EPD to address a wide range of challenges and questions. For instance, Sedighi, Fridrich, and Cогranne (2015) leveraged this distribution in their study, employing its mathematical characteristics to analyze and interpret multivariate data. Sailaja, Rao, and Reddy (2010) utilized the multivariate EPD to uncover patterns and trends in their research, demonstrating its adaptability to different applications. Gómez-Sánchez-Manzano, Gómez-Villegas, and Marín

(2008) explored the practical implications of this distribution, likely uncovering insights into its effectiveness in real-world scenarios.

Pascal, Bombrun, Tournet, and Berthoumieu (2013) delved into parameter estimation of the multivariate EPD. Similarly, Li, Cao, Hu, Xiong, and Wu (2021) incorporated this distribution in their research, contributing to the growing body of literature that recognizes the utility of the multivariate EPD. Najar, Bourouis, Al-Azawi, and Al-Badi (2019) and Korolev (2020) also explored the potential of this distribution, likely shedding light on its applicability in different contexts. The collective body of research represented by these studies attests to the versatility and significance of the multivariate EPD in various scientific and practical domains. Its use as a foundational element in these studies suggests that this particular form of the distribution offers valuable insights and analytical tools for researchers tackling complex multivariate data analyses and modeling challenges. As researchers continue to uncover new applications and properties, the multivariate EPD remains a dynamic and influential tool for statisticians and other scientists.

Statement of the Problem

The introduction of the EPD has achieved to a greater extent a much more robust analysis of data. Thus, many researchers have developed and adopted different forms of the univariate EPD (Vianelli, 1963; Giller, 2005; Rahnamaei, Nematollahi and Farnoosh, 2012; Olosunde, 2013), as little have been done on multivariate case. It has been argued (Olosunde, 2013) that the limitation of the

univariate EPD family is that they exhibit thinner tails. It is therefore inferred that this limitation also transfers in the multivariate EPD in Equation (2), and by extension all works that are based on this form of the multivariate EPD.

It is the position of this study that the limitation in Equation (2) can be regulated by a non-constant coefficient of the quadratic form so that the shape parameter, β does not affect the scale parameter, Σ . An initial attempt at addressing the observed limitation has been made in the univariate case in the work of Nkansah, Manford and Howard (2018), who present the general form of the EPD as

$$P(x|\mu, \sigma, q) = \frac{\sigma^{-1}}{2c^{-1/q}\Gamma\left(1 + \frac{1}{q}\right)} \exp\left\{-c\left|\frac{x-\mu}{\sigma}\right|^q\right\}, \quad (4)$$

where $c = \left[\Gamma\left(\frac{3}{q}\right)\right]^{q/2} \left[\Gamma\left(\frac{1}{q}\right)\right]^{-q/2}$, for $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$, and $q > 0$. This distribution is defined by three parameters: a location parameter denoted as μ , a scale parameter denoted as σ , and a shape parameter denoted as q , with c representing a variable constant that may be dependent on q . When $q = 2$, the distribution reduces to the Normal distribution, while for $q < 2$ or $q > 2$, it produces distributions with heavier or lighter tails, respectively. For instance, the distribution resembles the Double Exponential Distribution when $q = 1$, and approaches a Uniform Distribution as q tends towards infinity. Nkansah et al. (2018) clarified that the formulation for cc is designed to monitor and counteract changes in the distribution's characteristics induced by the value of q . In particular, c is obtained so that it does not affect the variability of the distribution in order to ensure flexibility of the tails. This study focuses on extending the work of Nkansah et al (2018) to the multivariate case as an alternative to that in

Equation (2). In the multivariate case, the proposed alternative multivariate EPD random variable may be denoted as $\mathbf{x} \sim \text{AMEPD}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q)$ or simply AMEPD. In the analysis chapter, it will be realized that by stochastic decomposition, $\mathbf{x}$ may be written as a combination of variables that includes a generating variate (R), with a density function of the form of the Kotz-type distribution. The Kotz distribution has the distribution of the density generator g of the form

$$g(u) = C_n u^{N-1} \exp(-ru^s), \quad r, s > 0, 2N + n > 2, \quad (5)$$

where C_n is a normalizing constant of which $\mathbf{x}$ possesses a symmetric Kotz-type distribution (Kotz, 1975) with density given by

$$\begin{aligned} \varphi(\mathbf{x}) &= C_p |\boldsymbol{\Sigma}|^{-1/2} \left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{N-1} \\ &\quad \times \exp \left\{ -r \left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^s \right\}. \end{aligned} \quad (6)$$

From Equation (4), it is clear that the intended AMEPD will be a mixture of Kotz-type distribution and the distribution presented in Equation (2). It will thus be a family of elliptical contour distribution similar to the one proposed by Olosunde (2013) of the form

$$\begin{aligned} dF(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \omega) &= \frac{d^p \mathbf{x}}{\sqrt{\pi^p |\boldsymbol{\Sigma}|}} \frac{\Gamma\left(1 + \frac{p}{2}\right)}{\Gamma\left(1 + \frac{n}{\omega}\right)} [h(\omega)]^{p/2} \\ &\quad \times \exp \left\{ - \left[h(\omega) (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{\omega/2} \right\} \end{aligned} \quad (7)$$

In Equation (7), the exact expression of $h(w)$ is not specified. Thus, Olosunde's multivariate EPD which is supposed to be a generalized form of Equation (2) is not explicit. Similar concerns can be raised about the distribution of Arellano-Valle and Richter (2012) if the skewed parameter is excluded. It is the

aim of this study to provide an explicit multivariate EPD that can provide an appropriate symmetric fit for a multivariate data. Thus, the study will examine the flexibility and the properties of the multivariate EPD. In the process, we will examine the parameters of the distribution and extend it to hypothesis testing.

Purpose of the Study

The purpose of this study is to embark on the development and extension of an alternative form of multivariate EPD within the field of some multivariate statistical techniques. This endeavor seeks to contribute to the ongoing advancements in statistical modeling and analysis by introducing a novel perspective or variation of the multivariate EPD.

Objectives of the Study

The main objective is to develop an alternative multivariate EPD (AMEPD). Specifically, the aim of the study is:

1. To derive the density of the AMEPD as well as its characteristics so that the effect of shape parameter on the scale parameter is eliminated;
2. To derive the estimators of the parameters of the AMEPD using maximum likelihood estimation method;
3. To compare the efficiency of estimators of the AMEPD to other related distributions using simulated and real data;
4. To demonstrate the application of the AMEPD in a multivariate statistical technique.

Significance of the Study

Dimitrov and Meyers (2010) underscored the pivotal role of statistical distributions in unraveling the complexities inherent in datasets, placing particular emphasis on their significance in elucidating the behavior and nature of data, notably within the healthcare sector. This comprehension, in turn, becomes instrumental in formulating effective control measures against prevailing health issues. Traditionally, the Normal distribution is a commonly employed tool for data analysis; however, its practical unreliability prompted a paradigm shift. In areas characterized by intricate data nuances, such as health, finance, and commerce, the EPD emerged as a crucial instrument, ensuring the dependability of data analysis results. Scholars have extensively explored diverse applications of the EPD in data analysis. Despite the growing interest, challenges persist, particularly in the domain of measurement.

The motivation behind this research emanates from the acknowledgment of the pivotal role of multivariate statistical techniques in unraveling intricate relationships and patterns within datasets containing multiple variables. Through the proposition and extension of an alternative multivariate Exponential Power Distribution, the study aims to introduce a fresh perspective or refinement that can enhance the effectiveness of established statistical methods.

Most studies in this domain have predominantly focused on theoretical aspects, with limited attention directed towards evaluating the adequacy of multivariate EPD using actual datasets. Establishing a meaningful multivariate EPD is imperative for assessing its efficacy in practical data applications. Conse-

quently, this study aims to bridge this gap by utilizing data from the health sector. The research seeks to equip researchers and practitioners with comprehensive insights into the EPD and its real-world applications. Furthermore, it aspires to provide valuable information to statisticians and individuals interested in statistics, enhancing their understanding of the EPD in the context of data analysis. Additionally, this study not only contributes to existing knowledge in EPD studies but also lays the groundwork for further research into other distributions exhibiting skewed and/or symmetrical characteristics.

The development and extension of this alternative multivariate EPD may encompass the exploration of new mathematical formulations, adaptation of existing models to specific contexts, or introduction of modifications to better capture certain features of multivariate data. The study may also venture into diverse applications across fields such as finance, biology, and engineering, thereby showcasing the versatility and efficacy of the proposed alternative. In essence, the significant of the lies in providing researchers, statisticians, and practitioners with a valuable tool that can contribute to a more nuanced and accurate understanding of multivariate data patterns, ultimately fostering advancements in the broader field of statistics.

Delimitation

The distribution that will be developed in this study is primarily designed for continuous dataset. Extending it to a discrete dataset may require additional considerations. It is also suitable for dataset which are symmetric in nature.

Limitation

In the course of this research, certain limitations are anticipated, some of which may be unforeseen. Despite the fact that most practical data often demonstrates skewed and kurtosis distributions, this study specifically focuses on examining the multivariate EPD, which is particularly robust in handling the kurtosis aspect of the distribution. It is important to note that this research does not claim that the multivariate EPD is universally the best distribution for all types of data; rather, it aims to elucidate its characteristics and demonstrate how it can be effectively applied to real-life datasets. It is acknowledged that other distributions exist, and their study and application are also pertinent in various contexts.

Furthermore, the data utilized in this study is secondary, and the researcher is not privy to the specific mode of collection. Consequently, the potential for bias in responses cannot be conclusively determined. Additionally, despite the research relying on a sample dataset from reputable sources, conclusions drawn may be susceptible to sampling errors. Moreover, as is the case with any survey, the research lacked control over extraneous factors that could impact the results. Another challenge lies in the presence of missing observations within the gathered data, but efforts were made to address this by excluding items with missing data.

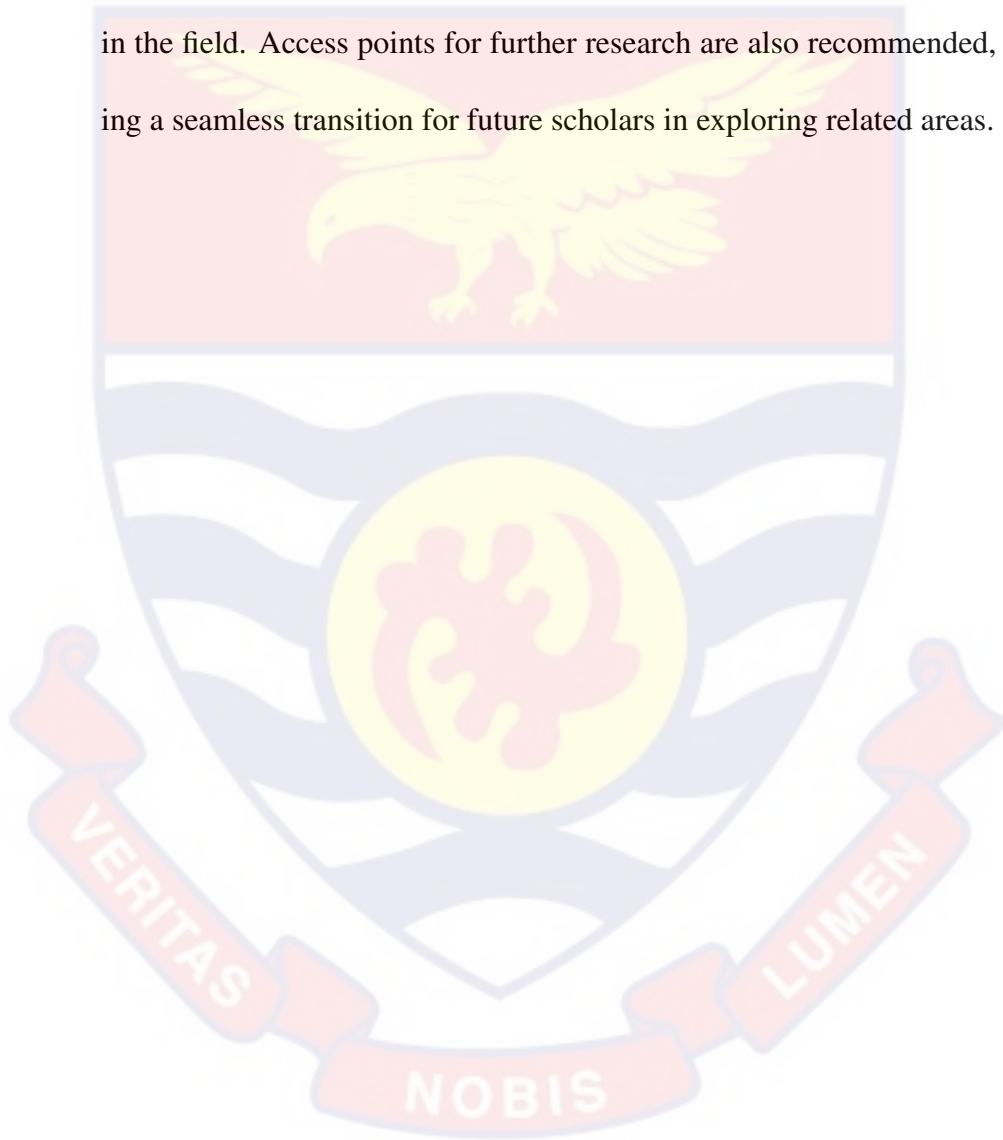
However, in light of these limitations, the findings of the research offer valuable insights into the properties of the multivariate EPD and its practical application. The results serve as a foundation for further exploration in this

emerging and significant area within symmetric elliptical and spherical distributions, particularly in the context of robustness concerning kurtosis. This study opens avenues for additional research endeavors, aiming to refine and expand our understanding of the multivariate EPD and its role in statistical modeling and analysis.

Organisation of the Study

The thesis is structured into five chapters, each serving a distinct purpose. The introductory Chapter One elucidates the study's background, articulates the problem statement, outlines objectives, underscores the significance of the study, and sets delimitations. Chapter Two delves into a comprehensive review of relevant literature, encompassing empirical studies and the perspectives of other authors. Chapter Three systematically discusses the methodological framework adopted for the study. It elaborates on the procedures involved in executing the research, ensuring transparency in detailing the research design, data collection methods, and analytical techniques employed. Chapter Four unveils the results and facilitates in-depth discussions. The study employs various mathematical concepts to derive the Alternative Multivariate Exponential Power Distribution (AMEPD), presenting relevant properties of this new distribution. Bivariate plots illustrating the AMEPD are generated across different values of the shape parameter for a specified mean vector and variance-covariance matrix. Additionally, a meticulous performance assessment is conducted through simulation and analysis of real datasets, with the findings thoroughly analyzed,

discussed, and presented. The final Chapter Five encapsulates the synthesis of the entire research endeavor. It provides a comprehensive summary, drawing informed conclusions from the results and discussions presented in Chapter Four. Furthermore, the chapter offers recommendations for efficient and effective data modeling decisions, contributing valuable insights for future research in the field. Access points for further research are also recommended, facilitating a seamless transition for future scholars in exploring related areas.



CHAPTER TWO

LITERATURE REVIEW

Introduction

In this chapter, relevant literature on multivariate EPD will be reviewed. This will cover the basic properties of elliptical distributions. Some forms of multivariate EPD which have been studied over the years will also be explored. Some characteristics of the multivariate EPD will be presented. Various application of the multivariate EPD will also be examined.

The Concept of Spherically and Elliptically Symmetric Distributions

From Rigby et al., (2017), a statistical distribution represents a simplified model of a population and its behavior, incorporating uncertainties through a random or stochastic component. This randomness is described using the concept of probability, making statistical distributions assume a probabilistic nature for the variables of interest (Rigby et al., 2017). Such distributions arise from the assumption that the observed data is a sample from an unknown larger population, which we aim to study. In the field of statistics, these distributions play a crucial role in accurately representing features of the data. The main goal is to capture the relevant aspects of a phenomenon under investigation and to explain how the data could have been generated as a result of a random experiment (Krishnamoorthy, 2006). Different approaches and theories may lead to various distribution formulations, analytical strategies, and outcomes. Adèr (2008)

added that it is important to note that a statistical distribution characterizes the data-generating process rather than providing a specific description of the data itself. The current study focuses on the Multivariate Exponential Power Distribution, but before delving into that, we will explore spherical and elliptically contoured symmetric distributions. The subsection will presents a comprehensive examination of spherically and elliptically symmetric distributions.

Early research on spherically and elliptically symmetric distributions dates back to Maxwell (1860), Bartlett and White (1934), Hartman and Wintner (1940) and Bennett and Lindley (1961) with review by Chmielewski (1981). However, the modern era of research in this field began with engineering applications discussed by Blake and Thomas (1968) and McGraw and Wagner (1968). Statisticians further contributed to this area with the works of Chu (1973), Dempster (1969), and Kelker (1970), which offered the first organized presentation of this family of distributions. Since then, a considerable and continuously expanding body of research has focused on exploring the properties and applications of these distributions. A paper by Cambanis, Huang, and Simons (1981) provided a systematic treatment of elliptical distributions, while Nolan (2013) offered estimation and theory of elliptical distribution. Additionally, Anderson and Fang (1990) published an influential review paper on this topic, highly regarded among specialists and practitioners in China. Johnson (1987) includes a clear discussion of elliptically contoured distributions. Notably, Diaconis and Freedman (1986) highlighted that elliptically contoured distributions are a minor modification of independently and identically distributed normals, offering

an interesting perspective. There are multiple ways to define distributions that are elliptically and spherically symmetric. An elliptical distribution is an extension of the multivariate normal distribution $N(\mu, \Sigma)$, whereas a spherical distribution is an extension of the standard multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_n)$.

A spherically symmetric distribution, also known as a radially symmetric distribution, is a statistical concept that describes a multivariate probability distribution where the likelihood of an event occurring is the same at a given distance from a central point, regardless of the direction. In other words, the distribution looks the same in all directions when viewed from the central point. Spherically symmetric distributions are commonly used in various fields, including physics and statistics, to model situations where there is no preferred direction. Instances of spherically symmetric distributions comprise the multivariate normal distribution characterized by a diagonal covariance matrix, where each dimension is independent, and the Rayleigh distribution.

In practical terms, spherically symmetric distributions simplify calculations and analysis in certain scenarios where the underlying system exhibits radial symmetry. This symmetry property can be leveraged to simplify integration, likelihood calculations, and other mathematical operations. In simpler terms, a distribution is spherically symmetric if the likelihood of observing a certain value of the random vector $\mathbf{x}$ is the same at the same distance from the origin, regardless of the specific direction in which that value lies. Spherical symmetric distributions are characterized by their radial symmetry, meaning that the distribution's shape remains the same when viewed from any angle around the origin.

This property simplifies statistical analyses and calculations, particularly when dealing with high-dimensional data or multivariate distributions. Examples of spherically symmetric distributions include the multivariate normal distribution with equal variance in all dimensions and the Rayleigh distribution.

A statistical concept known as an elliptically symmetric distribution characterizes a probability distribution that permits varying scales and orientations along distinct dimensions while maintaining the distribution's shape from any angle. Put differently, the distribution can have different degrees of dispersion along different axes, but it remains invariant to rotations. In p -dimensional space $\mathfrak{R}^p$, let $\mathbf{x} = (x, x, \dots, x)$ be a random vector. If the probability density function (PDF) of the distribution of $\mathbf{x}$ is elliptically symmetric under rotations, and the density function f has the form

$$f(\mathbf{x}) = \frac{k}{|\Sigma|^{1/2}} \cdot g\left(\mathbf{x} - \boldsymbol{\mu}\right)^{\top} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right), \quad (8)$$

The normalizing constant is denoted by k in the given expression, a p -dimensional random vector with a median vector $\boldsymbol{\mu}$ (which, if it exists, also represents the mean vector), and a positive definite matrix Σ that is proportionate to the covariance matrix, capturing dispersion along various dimensions when the latter exists (Frahm, Junker, & Szimayer, 2003), $|\Sigma|$ signifies the determinant of the matrix Σ , and $g(\cdot)$ denotes a scalar function encapsulating the distribution's shape, often selected as a well-known function like the multivariate Gaussian density. This means that the distribution is symmetric when looking at it from any direction, but the scale and orientation can differ among different dimensions. Elliptically symmetric distributions are a generalization of spherically

symmetric distributions, allowing for different scales and shapes along different directions while maintaining symmetry to rotations. Multivariate normal distribution with a general covariance matrix is a well-known example of an elliptically symmetric distribution. Elliptically symmetric distributions find applications in various fields, including finance (Kamdem, 2009), signal processing (Zozor Ren & Renaux, 2018), and machine learning (Muzellec & Cuturi, 2018), where data may exhibit similar patterns across different dimensions but with varying levels of spread and correlation. For an extensive knowledge of the characteristics and properties of elliptical distributions, Cambanis et al. (1981), Fang, Kotz, and Ng (1990), and Gupta and Varga (2012) provide extensive insights into the foundational aspects of elliptical distributions, offering a solid grounding in the subject. Furthermore, for more recent perspectives and advancements in the field, Zuo, Yin, and Balakrishnan (2021), Yin (2021), and Yin, Wang, and Sha (2022) contribute to the evolving discourse on elliptical distributions, providing up-to-date insights and exploring novel dimensions of research.

Some Sub-Classes of Elliptical Distributions

In the field of statistical distributions, Fang, Kotz, and Ng (1990) have categorized elliptical distributions into various subclasses, each exhibiting unique characteristics. Among these subclasses are the multiuniform distributions, symmetric Kotz type distributions, symmetric multivariate Pearson Types VII and II distributions, and the Symmetric Multivariate Laplace Distribution. Beyond

these, additional subclasses include the multivariate EPD, symmetrical-Bessel Distribution, multivariate Student's t -Distribution, multivariate Cauchy Distribution, multivariate Logistic Distribution, and multivariate Normal Distribution, among others (Rabbani et al, 2006; Eltoft et al., 2006a; Yu & Yin, 2023). These subclasses represent diverse mathematical formulations that capture specific patterns and characteristics in data distribution. For instance, the multiuniform distributions may describe situations where multiple uniform distributions coexist, while symmetric Kotz type distributions and multivariate Pearson Types VII and II distributions may model symmetric patterns in data. The Symmetric Multivariate Laplace Distribution, on the other hand, may be suitable for scenarios where data exhibits Laplacian characteristics. Rabbani et al. (2006), Eltoft et al. (2006b), and Yu and Yin (2023) have further contributed to the understanding of these subclasses. Their work may delve into specific applications, mathematical properties, or statistical implications associated with these elliptical distribution subclasses. The breadth of these subclasses underscores the versatility of elliptical distributions in modeling a wide range of real-world phenomena, making them valuable tools in statistical analysis and data modeling. Various researchers and practitioners have leverage these subclasses to better understand and describe the underlying patterns in multivariate data, contributing to advancements in various fields such as finance, engineering, and the natural sciences. The following review present some of them.

Symmetric Multivariate Laplace Distribution

The Symmetric Multivariate Laplace Distribution is a specific probability distribution encompassing multiple dimensions or variables, extending from the univariate Laplace distribution initially introduced by Laplace in 1774 (Stigler, 1986). In this multivariate context, the individual variables' marginal distributions follow Laplace distributions, as elucidated by Kotz, Kozubowski, and Podgórski (2012). It is a valuable tool in statistical modeling and multivariate data analysis, particularly when dealing with data that exhibit heavy-tailed and symmetric characteristics. The distribution is often used in robust statistical methods, where it provides an alternative to the Multivariate Normal Distribution in cases where normality assumptions do not hold.

A typical characterization of the symmetric multivariate Laplace distribution has the characteristic function (CF)

$$\varphi(t; \mu, \Sigma) = \frac{\exp(i\mu^\top t)}{1 + \frac{1}{2}t^\top \Sigma t}, \quad (9)$$

It is important to note, as noted by Kotz, Kozubowski, and Podgórski (2012), that the symmetric multivariate Laplace distribution differs from the multivariate normal distribution, where μ represents the vector of means for each variable and Σ signifies the covariance matrix (Fragiadakis & Meintanis, 2011). Even in situations when there is zero correlation and covariance in the covariance matrix, the variables nevertheless follow an elliptical pattern rather than being independent. The PDF for a p -dimensional multivariate Laplace distribution has the

following form in cases when $\mu = \mathbf{0}$:

$$f_{\mathbf{x}}(x_1, \dots, x_p) = \frac{2}{(2\pi)^{p/2} |\Sigma|^{0.5}} \left(\frac{\mathbf{x}^\top \Sigma^{-1} \mathbf{x}}{2} \right)^{v/2} P_v \left(\sqrt{2\mathbf{x}^\top \Sigma^{-1} \mathbf{x}} \right), \quad (10)$$

according to Kotz, Kozubowski and Podgórski (2012), where $v = (2 - p)/2$ and P_v denote the modified Bessel function of the second kind.

When $p = 2$ is a correlated bivariate example and $\mu_1 = \mu_2 = 0$, the PDF reduces to

$$f_{\mathbf{x}}(x_1, x_2) = \frac{1}{\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} P_0 \left(\sqrt{\frac{2 \left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2} \right)}{1 - \rho^2}} \right), \quad (11)$$

where σ_1 and σ_2 are the standard deviations of x_1 and x_2 , respectively, and ρ is the correlation coefficient of x_1 and x_2 . For the uncorrelated bivariate Laplace case, that is $p = 2, \mu_1 = \mu_2 = \rho = 0$ and $\sigma_1 = \sigma_2 = 1$, the PDF becomes

$$f_{\mathbf{x}}(x_1, x_2) = \frac{1}{\pi} P_0 \left(\sqrt{2(x_1^2 + x_2^2)} \right) \quad (12)$$

Figure 1 presents a bivariate plot of the Laplace distribution (Giller, 2024).

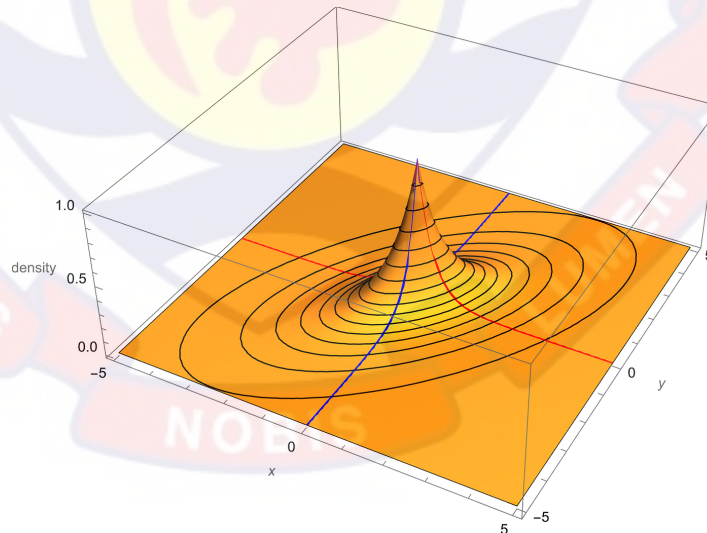


Figure 1: A Plot of Bivariate Laplace Distribution

The Multivariate Laplace distribution has seen extensive use across diverse fields, including applications in image and speech recognition, ocean engi-

neering, and finance. Its initial applications were in the detection of known constant signals amidst random noise, a subject explored by Marks, Wise, Halde-
man and Whited (1978) and Dadi and Marks (1987). Additionally, it has been employed in communication theory, specifically in encoding and decoding analog signals, with a particular emphasis on speech signals. However, it is worth noting that while the Multivariate Laplace distribution accurately models speech signals, true speech signals tend to exhibit significant correlations over time, as observed by Duttweiler and Messerschmitt (1976).

In applications related to image and speech compression, the Laplace distribution has been encountered. Epstein (1948) discussed its potential applications in the fracturing of materials under applied forces. The Multivariate Laplace distribution has found utility in the analysis of financial data, supporting its application in stochastic volatility modeling. Anderson (1992) presented the Multivariate Laplace distribution as a specific instance of the multivariate Linnik distribution, and Klein (1993) scrutinized yield rates, noting deviations from the empirical distribution compared to the Multivariate Laplace distribution. Despite the advantages offered by the Symmetric Multivariate Laplace Distribution, it may not be the most suitable choice for strongly asymmetric data or datasets with intricate dependencies. Furthermore, parameter estimation can be challenging due to the distribution's heavy tails.

As proposed by Kozubowski, Podgórski, and Rychlik (2013), the Symmetric Multivariate Laplace Distribution displays symmetry, exhibiting equal tail behavior on both sides of its center. This symmetry proves beneficial for

modeling non-normally distributed data, particularly in scenarios involving outliers or heavy-tailed data (Kotz, Kozubowski & Podgórski, 2012). Renowned for its heavy-tailed properties, it excels in modeling data with thicker tails than the Multivariate Normal Distribution, especially in the presence of extreme observations (Kollo & Srivastava, 2005). The distribution, characterized by a location vector and scale matrix, enables flexible modeling of central tendency and spread across various data structures (Kotz, Kozubowski & Podgórski, 2001).

The Symmetric Multivariate Laplace Distribution, primarily employed in statistical modeling, proves valuable for handling multivariate data exhibiting heavy-tailed behavior, providing robustness to outliers. Its application in robust statistics underscores its lower sensitivity to outliers compared to the Multivariate Normal Distribution. Notable strengths include its ability to model heavy-tailed data and robustness to outliers, making it a powerful tool in finance, signal processing, and robust statistics (Eltoft, Kim & Lee, 2006b). Its symmetric properties contribute to an accurate representation of real-world phenomena deviating from normality, offering a robust framework for analyzing complex multivariate datasets. Balancing computational efficiency and robustness, it proves practical for applications with potential outliers or heavy-tailed behavior. While suitable for parameter estimation, hypothesis testing, and model fitting, users should be mindful of challenges, particularly in high-dimensional data settings. A comprehensive understanding of its properties and limitations is crucial for accurate modeling and interpretation of results (Lindholm, Wahlström, Lindsten & Schön, 2022).

Multivariate Uniform Distribution

In the field of probability theory and statistics, the multivariate uniform distributions, commonly known as rectangular distributions, form a family of symmetric probability distributions (Ashby, 2014). This particular distribution is applied to articulate experiments in which outcomes fall within defined bounds, representing arbitrary occurrences. Acknowledged as a fundamental probability distribution, it assumes a vital role in modeling data in multidimensional spaces, especially within the domains of statistics, simulation, and optimization. The key characteristic defining this distribution is uniformity, signifying that all points within the specified space possess an equal likelihood of occurrence (Karian & Dudewicz, 2016). Its uniform nature renders it a valuable selection as a reference distribution or a benchmark for comparison across diverse statistical and simulation contexts.

The PDF of a multivariate uniform distribution varies based on the specific dimensionality and the range of values for each variable (Arslan, 2014). Considering the scenario of a multivariate uniform distribution in p -dimensional space, denoted as $U(a, b)$, where a and b represent vectors indicating the lower and upper bounds, respectively. The PDF of a multivariate uniform distribution is expressed as

$$f(\mathbf{x}) = \frac{1}{\text{Vol}(D)} \quad (13)$$

where $\mathbf{x}$ is the vector of random variables, and $\text{Vol}(D)$ is the volume of the region D in the p -dimensional space defined by the bounds a and b .

For a uniform distribution over a hyperrectangle defined by the lower bound vector $\mathbf{a} = (a_1, a_2, \dots, a_p)$ and upper bound vector $\mathbf{b} = (b_1, b_2, \dots, b_p)$, the volume $\text{Vol}(D)$ is given by:

$$\text{Vol}(D) = (b_1 - a_1) \cdot (b_2 - a_2) \cdot \dots \cdot (b_p - a_p) \quad (14)$$

The PDF is non-zero only within the region $\bar{D}$ and is zero outside this region. Figure 2 presents a bivariate plot of the Uniform distribution (Eltoft, Kim & Lee, 2006a).

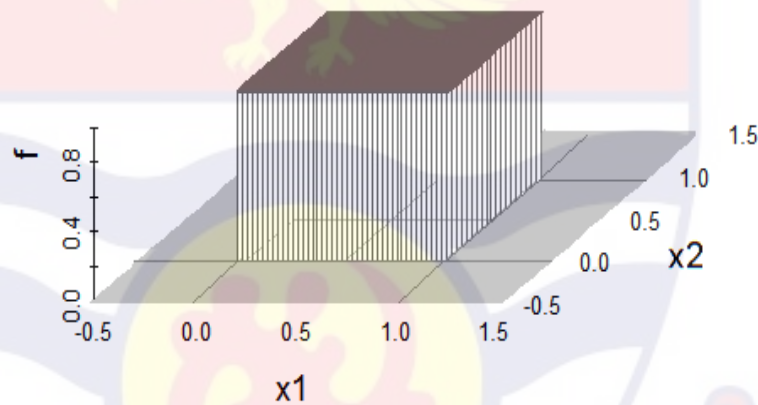


Figure 2: Bivariate Uniform Distribution

The use of the multivariate uniform distribution is widespread in the statistical analysis of directional data and projection pursuit methods (Watson, 1984; Cheng & Spering, 1987). However, it is essential to acknowledge significant limitations associated with this distribution. In many real-world scenarios, data does not adhere to a uniform distribution in multidimensional spaces, potentially restricting its efficacy in accurately representing intricate datasets. It is important to note that the Multivariate Uniform Distribution may encounter challenges in capturing the subtleties and complexities inherent in real-world data. The uniformity observed within the defined hyper-rectangular region may not precisely

mirror the underlying distribution of numerous real-world phenomena, often characterized by more intricate patterns and dependencies. Nevertheless, the Multivariate Uniform Distribution retains its status as a fundamental probability distribution with diverse applications, particularly in multivariate statistical analysis, Monte Carlo simulations, and random number generation. Its defining feature is a uniform PDF within a specific hyper-rectangular region in multidimensional space (Watson, 1984; Cheng & Spering, 1987).

One notable strength of the Multivariate Uniform Distribution is its simplicity and ease of use, making it a popular choice for generating random vectors with equal probabilities over a defined hyper-rectangular region. This simplicity is particularly advantageous in applications where events are equally likely to occur within a specific bounded range. Additionally, its straightforward mathematical properties enable efficient computation and simulation, establishing it as a reliable tool for generating random samples across various contexts (Ashby, 2014).

Moreover, the Multivariate Uniform Distribution serves as a foundational element for various more complex distributions and statistical methods. It acts as a benchmark for comparing and contrasting other multivariate distributions, allowing researchers to assess the performance and characteristics of alternative distributions in specific applications (Li & Peng, 2020). While the Multivariate Uniform Distribution excels at generating random vectors within a defined hyper-rectangular region and provides a simple and efficient means of conducting simulations, it may not capture the intricacies of real-world data. Despite

this, its simplicity and ease of use position it as a valuable distribution for comparison and benchmarking purposes. Recognizing its limitations, particularly in contexts where data exhibit more intricate patterns and dependencies, is essential (Arias-Castro, 2022).

The Multivariate Uniform Distribution proves to be a straightforward yet valuable tool in various statistical and simulation applications, offering a simple method for generating random points with equal probabilities across a defined region. However, it is crucial to acknowledge its limitations and recognize that in many real-world scenarios, data may not conform to such uniformity. Its primary utility lies in cases where no other prior information or distribution assumptions are available, making it a useful reference or starting point for analysis and simulation in multidimensional spaces (Karian & Dudewicz, 2016).

Symmetric Kotz Type Distributions

Symmetric Kotz-type distributions are a class of probability distributions that are used in statistical modeling and data analysis. They are named after the mathematician Y. Kotz, who contributed to the development of this distribution family. These distributions are characterized by their symmetric shape, making them suitable for modeling data with symmetric patterns. Given that $\mathbf{x} \sim EC_p(\mu, \Sigma, g)$, the density generator g is of the form

$$g(u) = C_p u^{N-1} \exp(-ru^s), \quad r, s > 0, 2N + p > 2, \quad (15)$$

where C_p is a normalizing constant (Kotz, 1975). Consequently, $\mathbf{x}$'s density is

provided by

$$C_p |\Sigma|^{-1/2} \left[(\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right]^{N-1} \exp \left\{ -r \left[(\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right]^s \right\} \quad (16)$$

where

$$\begin{aligned} C_p &= \Gamma(p/2) \pi^{-p/2} \left\{ \int_0^\infty u^{p/2-1} u^{N-1} \exp(-ru^s) du \right\}^{-1} \\ &= \frac{s \Gamma(p/2)}{\pi^{p/2} \Gamma((2N+p-2)/2s)} r^{(2N+p-2)/2s} \end{aligned} \quad (17)$$

In the case of $N = 1, s = 1$, and $r = \frac{1}{2}$, the distribution becomes multinormal. This is the basic Kotz distribution described in Kotz (1975) when $s = 1$. It is discovered that this family of distributions is helpful in building models where the standard normality assumption is not relevant (Koutras, 1986). We find a Kotz type distribution for $s = 1$ in Equation (16), the CF of which is given by Iyengar and Tong (1988) as

$$\exp \left(i \mathbf{t}^\top \mu \right) \psi_{n,N} \left(\mathbf{t}^\top \Sigma \mathbf{t}; r \right) \quad (18)$$

where

$$\psi_{n,N}(u; r) = e^{-u/4r} \sum_{m=0}^{N-1} \binom{N-1}{m} \frac{\Gamma(n/2)}{\Gamma(n/2+m)} \left(-\frac{u}{4r} \right)^m. \quad (19)$$

Symmetric Kotz-type distributions, stemming from the Kotz family, exhibit significant flexibility, enabling the modeling of diverse symmetric data patterns. Figure 3 presents a bivariate plot of Symmetric Kotz-type distribution with $m = 2, s = 1, r = 1/2$ (Yu & Yin, 2023).

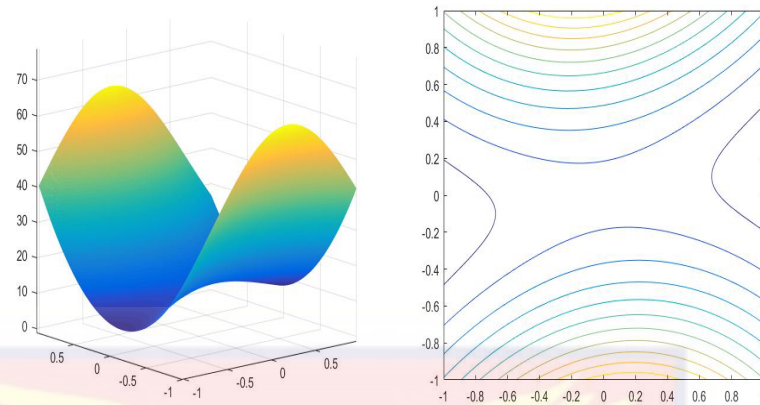


Figure 3: Bivariate Symmetric Kotz Type Distribution.

Symmetric Kotz-type distributions are applicable in finance, engineering, and environmental sciences, these distributions offer customization through adjustable parameters such as shape, scale, and location, tailoring to observed data characteristics. Their inherent symmetry simplifies modeling for symmetric datasets, aiding in central tendency inferences. Particularly adept at capturing heavy-tailed data, Symmetric Kotz-type distributions find utility in fields like risk analysis, reliability engineering, and finance, facilitating modeling of extreme events like market crashes or natural disasters. Accurate parameter estimation poses challenges, especially with limited data, necessitating specialized software and statistical techniques. While less known than some distributions, their application requires careful consideration, and validation through model selection techniques and goodness-of-fit tests is crucial.

A versatile probability distribution applicable in finance, economics, and engineering, the Symmetric Kotz type distribution, derived from the Kotz family, demonstrates flexibility in modeling data with heavy tails and skewness (Nadarajah, 2003). Notably, it excels in capturing diverse shapes and tail be-

haviors, addressing limitations of traditional distributions like the normal distribution. Leveraged for heavy-tailed data, high kurtosis, and skewness, it allows precise fitting through adjustable parameters, contributing to accurate statistical inferences and predictions (Plungpongpun, 2003; Anderson et al. 2013). Applications of the Kotz type distribution are found in a wide range of fields.

In crossover studies when insulin is administered to rabbits, it is used to simulate repeated measurements (Lindsey, 1999; Kotz & Nadarajah, 2002). Furthermore, in an evolutionary process including continuous variables, it functions as a substitute for the normal distribution when modifying individuals (Plungpongpun, 2003). In signal compression methods, the distribution is also utilized to represent the bounds of the minimum conditional entropy context quantizer (MCECQ), which is used for the sequential coding of variables like $X_0, X_1, X_2, \dots$ (Wu, Chou & Xue, 2000). Moreover, Bingham and Kiesel (2002) address its application in the modeling of asset returns inside financial portfolios, and Arellano-Valle, Del Pino, and Martin (2002) highlight its usefulness in Bayesian regression models.

Nadarajah (2003) explained that robust mathematical properties facilitate efficient computation and simulation for various statistical analyses. However, its complexity necessitates a strong understanding for optimal use. The Symmetric Kotz type distribution proves a robust tool for modeling complex datasets with heavy tails and skewness, offering flexibility and accurate representation of real-world data. While valuable, its complexity requires a thorough understanding for optimal utilization. Symmetric Kotz-type distributions, valu-

able for modeling symmetric data patterns, are flexible and adept at capturing heavy-tailed data. Awareness of challenges in parameter estimation and model selection is crucial (Kleiber & Stoyanov, 2013). Overall, they enhance the statistical techniques for data analysis and modeling.

Symmetric Multivariate Pearson Type VII Distributions

A specific type of probability distribution used to model multivariate data is the Symmetric Multivariate Pearson Type VII Distribution. Expanding upon the univariate Pearson Type VII distribution, this distribution is designed for use in situations where data display multivariate features, particularly when handling multivariate heavy-tailed and skewed data. Notably, symmetry is preserved in a particular case of Pearson Type IV.

Let a $p \times 1$ random vector $\mathbf{x} \sim EC_p(\mu, \Sigma, g)$ be given in mathematical terms. Then $\mathbf{x}$ represent a symmetric multivariate Pearson Type VII distribution if the density generator $g(\cdot)$ for $\mathbf{x}$ is given by

$$g(t) = C_p(1 + t/m)^{-N}, \quad N > p/2, \quad m > 0 \quad (20)$$

and

$$C_p = (\pi m)^{-p/2} \Gamma(N) / \Gamma(N - p/2). \quad (21)$$

The Pearson Type IV distribution's skewness is managed by the shape parameter N . A symmetric three-parameter family called the Pearson Type VII distribution is generated by setting its value to zero (Pearson, 1916; Lahcene, 2013). Its

density function may be found using

$$p(x) = \frac{1}{\alpha B(m - \frac{1}{2}, \frac{1}{2})} \left[1 + \left(\frac{x - \lambda}{\alpha} \right)^2 \right]^{-m}, \quad (22)$$

where B represents the Beta function. The density of $u = r^2$, where r is the generating variate of $\mathbf{x}$, can be written as

$$\frac{1}{B(p/2, N - p/2)} m^{-p/2} u^{p/2-1} (1 + u/m)^{-N}, \quad u > 0. \quad (23)$$

This indicates that r^2/m has a beta type II distribution with parameters $p/2$ and $N - p/2$. Essential distributions like the multivariate t distribution (for $N = \frac{1}{2}(p + m)$) and the multivariate Cauchy distribution (for $m = 1$ and $N = \frac{1}{2}(p + 1)$) are examples of symmetric multivariate Pearson Type VII distributions. If the density function of a random variable B is given by

$$\frac{1}{B(\alpha, \beta)} b^{\alpha-1} (1 + b)^{-(\alpha+\beta)}, \quad b > 0, \quad (24)$$

then it is said to have a beta type II distribution with parameters α and β .

This distribution is denoted as $B \sim \text{BeII}(\alpha, \beta)$. Let $\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{y}$ have a symmetric multivariate Pearson Type VII distribution, denoted as $\text{MPVII}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$.

Clearly, $\mathbf{x} = \text{MPVII}(\mathbf{0}, \mathbf{I}, g)$, representing a spherical distribution. If we partition $\mathbf{x}$ into

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix}, \quad \mathbf{x}^{(1)} : r \times 1. \quad (25)$$

then $\mathbf{x}^{(1)} \sim S_r(g_r)$, where the density generator $g_r(\cdot)$ is expressed as

$$g_r(u) = \frac{\Gamma(N - (n - r)/2)}{(\pi m)^{r/2} \Gamma(N - n/2)} \left(1 + \frac{u}{m} \right)^{-N + (n - r)/2}. \quad (26)$$

A multivariate Pearson Type VII distribution, as introduced by Shimizu and Iida (2002), is characterized by a PDF given by

$$f_P(\mathbf{x}) = \frac{\Gamma(v + \frac{p}{2})}{\Gamma(v)} (\pi A^2)^{-\frac{p}{2}} \left(1 + \frac{\mathbf{x}^\top \mathbf{x}}{A^2}\right)^{-\frac{p}{2}-v}, \quad \mathbf{x} \in \mathbb{R}^p, A \neq 0, v > 0, \quad (27)$$

Here, the associated weight is determined by

$$\frac{d}{d\sigma} G(\sigma) = \frac{2^{1-v} A^{2v}}{\Gamma(v)} \sigma^{-1-2v} \exp\left\{-\frac{A^2}{2\sigma^2}\right\}, \quad (28)$$

demonstrating that the distribution of $1/\sigma^2$ is a gamma distribution with scale $2/A^2$ and shape v . A multivariate t -distribution with n degrees of freedom is a notable particular case of the Pearson Type VII distribution with $v = n/2$ and $A^2 = n$.

The Symmetric Multivariate Pearson Type VII Distribution possesses a notable strength in its capacity to model intricate multivariate data characterized by heavy tails and skewness (Shimizu & Iida, 2002). This attribute renders it particularly valuable in domains such as biology (Nayebi & Ganguli, 2017), risk analysis, and reliability engineering (Mott, 2018), where data frequently deviate from normality assumptions. By accommodating both heavy tails and skewness, this distribution offers a more precise representation of real-world phenomena, a critical aspect for accurate risk assessment and decision-making. Moreover, the distribution provides a range of adjustable parameters, allowing for customization to specific multivariate datasets (Yin et al., 2019). This adaptability enables a more accurate fitting to empirical data, enhancing the quality of statistical inferences and predictions.

However, it is important to note that the Symmetric Multivariate Pearson Type VII Distribution is a complex model that may pose challenges, especially

for individuals not well-versed in its mathematical properties. Users must possess a solid understanding of multivariate statistics and the intricacies of this distribution to ensure proper application and interpretation of results. Despite its complexity, the Symmetric Multivariate Pearson Type VII Distribution stands out as a potent tool for modeling multivariate data with heavy tails and skewness, offering an improved representation of real-world phenomena compared to traditional distributions (Garay et al., 2015). Its adaptability and mathematical properties make it invaluable across various fields, particularly when dealing with intricate multivariate datasets (Yin et al., 2019). Nonetheless, its effective utilization demands a thorough comprehension of its mathematical foundations.

Symmetric Multivariate Pearson Type II Distribution

Karl Pearson developed the Pearson Type II Distribution in the late 1800s, and it is a flexible family of continuous probability distributions (Lahcene, 2013; Provosta, Zareamoghaddam, Ahmed & Ha, 2022). The Pearson Type II distribution, often known as the inverse gamma distribution, is a continuous probability distribution that demonstrates symmetry and a link with the gamma distribution (Pearson, 1963). In the Pearson Type I family, it is a special example that is restricted to symmetric distributions (Lahcene, 2013). Widely utilized in statistical modeling, the Pearson Type II distribution holds particular significance in Bayesian statistics due to its diverse range of applications. It is a probability distribution that is part of the Pearson system of distributions, which includes

various types of distributions used in statistical modeling and data analysis (Zografos, 1999). The Pearson Type II Distribution is primarily employed to model data that exhibits a bell-shaped or Gaussian-like distribution. It is characterized by parameters that allow for flexibility in modeling both symmetric and skewed data (Branco & Dey, 2001).

According to Jarder (1997) and Khalafi & Azimmohseni (2014), the univariate Pearson type II distribution can be extended to higher dimensions as the multivariate Pearson type II distribution. The multivariate Pearson type II distribution (MPII) is classified as a multivariate symmetric spherical distribution since it is specified on the closure of a p -sphere. This classification was made by Fang et al. (1991), Liang and Bentler (1998), and Arellano-Valle (2001). A crucial role is taken by the Pearson type II distribution in the class of multivariate symmetric spherical distributions. Notably, the Pearson type II distribution and the related radial distribution can be used to express any marginal distribution of a multivariate symmetric distribution (Arellano-Valle, 2001; Zografos & Nadarajah, 2005; Diaz-Garcia & Jaimez, 2012). The multivariate Pearson type II distribution, defined by two parameters: a shape parameter $a > 0$ and a dimension parameter p , finds use in a variety of domains, including reliability analysis, Bayesian inference, and modelling of random variables with positive support. The characterization of the multivariate Pearson type II distribution extends to its joint density function, marginal distributions, moments and cumulants, characteristic function, and properties like norm-scale invariance, rotation invariance, and complex normality (Joarder, 1997). Furthermore, various simu-

lation methods, including the inverse transform method, the rejection method, or the transformation method, can be employed to simulate the multivariate Pearson type II distribution (Arellano-Valle, 2001).

According to Kotz (1975) and Yin and Dong (2023), the p -dimensional random vector $\mathbf{X}$ is said to follow a symmetric multivariate Pearson type II distribution with parameters $m > -1$, $\mu \in \mathfrak{R}^p$, and $\Sigma : p \times p$ with $\Sigma > 0$ if its PDF is expressed as

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{p}{2} + m + 1\right)}{\pi^{\frac{p}{2}} \Gamma(m + 1) \sqrt{|\Sigma|}} \left(1 - (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right)^m, \quad (29)$$

where $0 \leq (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \leq 1$.

Applications for this distribution include modelling of random variables with positive support (Khalafi & Azimmohseni, 2014), reliability analysis, and Bayesian inference. It is a special case of multivariate symmetric spherical distributions defined on the closure of a p -sphere (Arellano-Valle, 2001). A closed form for the multivariate Pearson type II distribution's CF was developed by Joarder (1997). Applying an integral from Ryzhik and Gradshteyn (2014), presented by

$$\int_0^1 x^{\nu+1} (1-x^2)^{\mu} J_{\nu}(bx) dx = 2^{\mu} \Gamma(\mu+1) b^{-(\mu+1)} J_{\nu+\mu+1}(b), \quad (30)$$

with $\mu = m$ and $\nu = \frac{p-2}{2}$, the CF of the multivariate Pearson type II distribution is expressed as

$$\psi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^T \mu} 2^{\frac{p}{2}+m} \Gamma\left(\frac{p}{2} + m + 1\right) \left(|\Sigma^{\frac{1}{2}} \mathbf{t}|\right)^{-\frac{p}{2}-m} J_{\frac{p}{2}+m}(|\Sigma^{\frac{1}{2}} \mathbf{t}|), \quad \mathbf{t} \in \mathfrak{R}^p. \quad (31)$$

This distribution represents a specific instance within the category of multivariate symmetric spherical distributions, characterized by their definition on

the closure of a p -sphere (Arellano-Valle, 2001). Its practical applications extend to various fields, including Bayesian inference, reliability analysis, and the modeling of random variables that exhibit positive support (Khalafi & Azim-mohseni, 2014). Figure 4 presents a bivariate plot of Pearson Type II distribution with $m = 1$ (Yu & Yin, 2023).

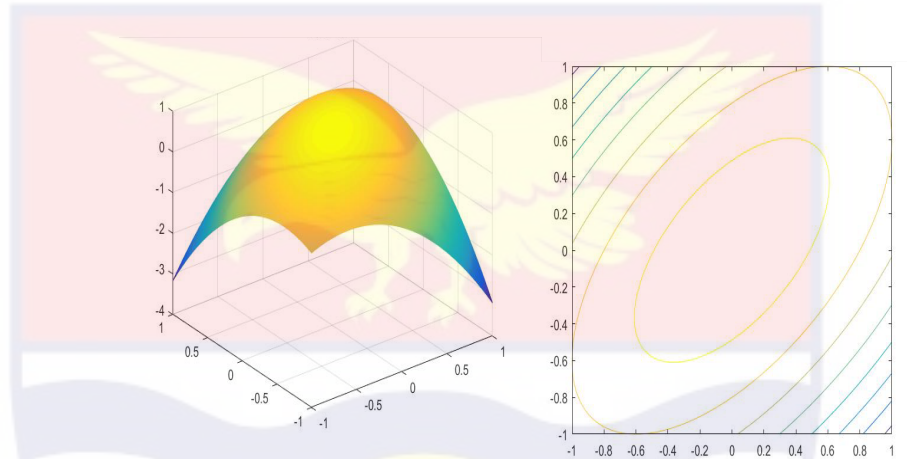


Figure 4: Bivariate Plot of Pearson Type II Distribution

In Bayesian statistics, this distribution is frequently employed as a conjugate prior for the precision (inverse of variance) parameter. Applications for it include modelling random variables with positive support, reliability analysis, and Bayesian inference (Karamikabir & Afshari, 2019; Diaz-Garcia & Jáimez, 2012). The ability of the Pearson Type II Distribution to reflect a variety of data forms, from almost symmetric to somewhat skewed, is one of its advantages (Wang, 2009). This adaptability is especially helpful in situations where a distribution with comparable characteristics is needed but the data's shape does not strictly follow the symmetry of a normal distribution. Researchers and practitioners often use this distribution to model data that exhibits mild to moderate skewness, providing a more accurate representation of real-world data. Addi-

tionally, the Pearson Type II Distribution offers a set of parameters that can be adjusted to fine-tune the distribution to specific data sets (Mecklin & Mundfrom, 2005). This adaptability allows for a more precise fit to empirical data, enhancing the quality of statistical inferences and predictions (Khalafi & Azimmohseni, 2014).

However, it is important to note that the Pearson Type II Distribution may not be the ideal choice for data with extreme skewness, heavy tails, or multimodal behaviour (Bishara & Hittner, 2014). In such cases, other distributional models, such as heavy-tailed distributions or mixtures of distributions, might be more appropriate. The Pearson Type II Distribution is a valuable tool for modeling data with a bell-shaped distribution that ranges from symmetric to moderately skewed (Branco & Dey, 2001). Its flexibility and adaptability make it suitable for a variety of statistical applications, particularly when data exhibit characteristics that deviate from strict normality. Nonetheless, users should be aware of its limitations and consider alternative models when dealing with more complex data distributions (Khalafi & Azimmohseni, 2014).

Multivariate Student's t -Distribution

The multivariate student's t -distribution (Kotz & Nadarajah, 2004), a multivariate version of the univariate student's t -distribution, is a valuable and versatile probability distribution frequently used in statistics, particularly in fields like finance (Nguyen, Ruiz-Gazen, Thomas-Agnan & Laurent, 2019; Kan & Zhou, 2017; Kamdem, 2005), econometrics (Banasaz, 2022), and machine learn-

ing (Chen, Wang & Gorban, 2019). It offers a flexible approach to modeling multivariate data, especially when dealing with heavy-tailed data and uncertainty in parameter estimation. It is often used in statistical inference and multivariate analysis, especially when dealing with small sample sizes. For a p -dimensional random vector $\mathbf{x}$, the multivariate Student's t -distribution PDF with p degrees of freedom is as

$$f(\mathbf{x}; \mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu^{\frac{p}{2}} \pi^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \left(1 + \frac{1}{\nu} (\mathbf{x} - \mu)^{\top} \Sigma^{-1} (\mathbf{x} - \mu)\right)^{-\frac{\nu+p}{2}} \quad (32)$$

where the mean vector is denoted by μ , the variance-covariance matrix is denoted by Σ , the degrees of freedom parameter is denoted by ν , and the gamma function is represented by $\Gamma(\cdot)$. The covariance matrix's determinant is denoted by the notation $|\Sigma|$. It is significant to remember that the multivariate Student's t -distribution converges to the multivariate normal distribution as ν approaches infinity. When working with small sample sizes or in scenarios where the validity of normality assumption may be questioned, this distribution is especially helpful. Compared to the multivariate normal distribution, the Student's t -distribution has larger tails, which makes it more resistant in the face of outliers (Kotz & Nadarajah, 2004). Student's t -distribution with $\nu = 9$ is plotted bivariately in Figure 5 (Yu & Yin, 2023).

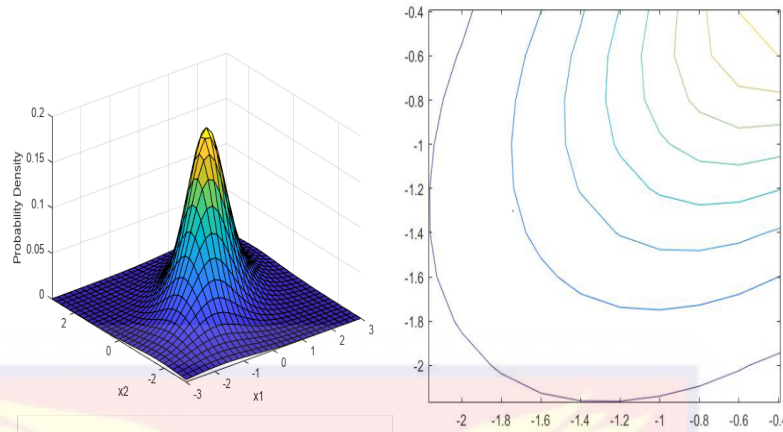


Figure 5: Bivariate Plot of Student's t -Distribution.

One of the most notable strengths of the Multivariate Student's t -Distribution is its ability to model data with heavy tails and outliers (Kotz & Nadarajah, 2004). This distribution provides a more robust alternative to the Multivariate Normal Distribution, which assumes that data are normally distributed and have finite variance. The Student's t -Distribution allows for the accommodation of data with greater variability and is particularly useful when dealing with real-world phenomena that exhibit non-normality and heavier tails. The distribution's flexibility is also a significant advantage. It is characterized by a parameter known as degrees of freedom, which governs the thickness of the tails. As the degrees of freedom increase, the distribution approaches a Multivariate Normal Distribution, making it a versatile tool that can adapt to a wide range of data behaviours (Nguyen et al., 2019).

Additionally, the Multivariate Student's t -Distribution is widely used in stochastic processes (Nkemnole & Abass, 2015), time-series (Jayakumar, Sulthan & Samuel, 2021), Bayesian statistics (de Andrade et al., 2018; Zhang et al., 2019) and hierarchical modeling (Chen et al., 2016), factor analysis (Castro,

Costa, Prates & Lachos, 2014), principal component analysis (Guo & Bondell, 2022), providing a framework for robust inference in situations where the data may not strictly conform to normality assumptions. The conjugate properties of this distribution simplify parameter estimation and posterior analysis in Bayesian frameworks (Zhang et al., 2019). However, it is important to acknowledge that working with the Multivariate Student's t -Distribution can be more computationally intensive compared to the Multivariate Normal Distribution, especially when the degrees of freedom are low (Kan & Zhou, 2017). Interpretation and understanding of the degrees of freedom parameter are essential, as it plays a critical role in shaping the distribution and modeling characteristics (Kotz & Nadarajah, 2004).

The Multivariate Student's t -Distribution is a valuable tool for modeling multivariate data, particularly when dealing with heavy-tailed data and robustness to outliers. Its flexibility and adaptability make it suitable for a wide range of applications, while its Bayesian properties enhance its usefulness in statistical inference. Nonetheless, users should be aware of its computational implications and the importance of understanding the degrees of freedom parameter for accurate modeling and analysis (Kotz & Nadarajah, 2004).

The Bessel Distribution

A probability distribution called the Bessel distribution bears the name of the mathematician Friedrich Bessel. Bessel-K distributions and modified Bessel distributions are the two primary varieties of Bessel distributions (Srivastava &

Nadarajah, 2006). The Bessel function of the second kind, $K_\nu(z)$, is related to the Bessel-K distribution. In this case, z is a non-negative variable and ν is a positive parameter. The distribution of Bessel- K has the PDF

$$f(x; \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} x^{\nu-1} e^{-\frac{x}{2}} I_\nu\left(\frac{x}{2}\right) \quad (33)$$

in where Γ represents the gamma function and I_ν represents the modified Bessel function of the first kind (Nadarajah, 2008). Another significant distribution connected to Bessel functions is the modified Bessel distribution, or $I_\nu(z)$. The modified Bessel distribution's PDF can be obtained by

$$f(x; \nu) = \frac{1}{\Gamma(\nu+1)} \left(\frac{x}{2}\right)^\nu e^{-\frac{x}{2}} I_\nu\left(\frac{x}{2}\right) \quad (34)$$

where Γ is the gamma function, ν is a positive parameter, and I_ν is the modified Bessel function of the first kind. The multivariate Bessel Distribution's PDF can be found using

$$p(\mathbf{x}) = \frac{(-1)^{q+1} (2r)^{-q-n} \pi}{\pi^{\frac{n}{2}} \Gamma\left(q + \frac{n}{2}\right) \sin(q\pi)} \|\mathbf{x} - \boldsymbol{\mu}\|^q I_q\left(\frac{\sqrt{\|\mathbf{x} - \boldsymbol{\mu}\|^2}}{r}\right) \quad (35)$$

with $I_q(z) = \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(k+q+1)} \left(\frac{z}{2}\right)^{q+2k}$

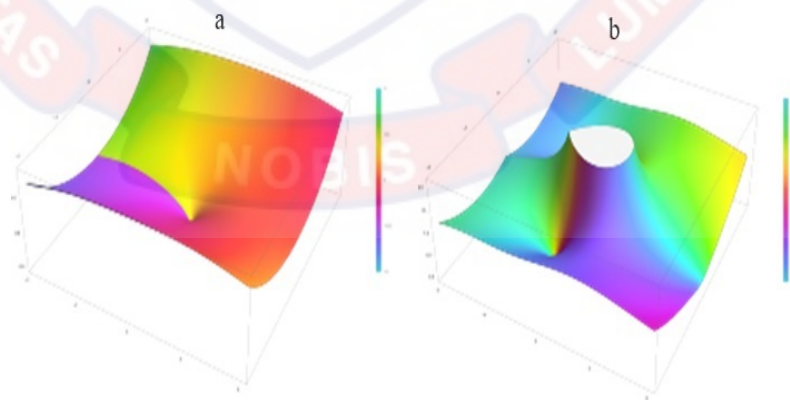


Figure 6: Plot of the Spherical Bessel Distribution of a) the First Kind and b) the Second Kind

Figure 6 presents a bivariate plot of the spherical Bessel Distribution of the first and second kind. Applications of Bessel distributions are found in various fields, including physics, engineering, and statistics (Iliopoulos & Karlis, 2003). Bessel functions arise naturally in problems involving circular symmetry, such as in wave equations, heat conduction, and diffraction of waves. It is important to note that the choice between the Bessel-K and modified Bessel distribution depends on the specific context and the mathematical form that best represents the problem at hand (Srivastava & Nadarajah, 2006; Nadarajah, 2008). These distributions are also commonly used in the context of random matrix theory and circular statistics (Jammalamadaka, Sengupta & Sengupta, 2001; Fyodorov & Keating, 2014; Kent, Ganeiber & Mardia, 2018).

The Spherical-Dirichlet Distribution

The spherical-Dirichlet distribution is a fundamental and versatile probability distribution proposed by Guardiola (2020), used in various fields, including statistics, machine learning and natural language processing. It is particularly valuable for modeling data involving proportions or compositional data, where each data point consists of multiple components that sum to a constant. According to Guardiola (2020), it is generated by converting the Dirichlet distribution on the simplex to the comparable space on the hypersphere. Given the Dirichlet distribution on the simplex given by Ingram (Olkin and Rubin 1964),

let $\mathbf{x}$, then

$$\begin{aligned} f_{\text{Dir}}(\mathbf{y}; \boldsymbol{\alpha}) &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m y_i^{\alpha_i-1} \\ &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^{m-1} y_i^{\alpha_i-1} \left(1 - \sum_{i=1}^{m-1} y_i\right)^{(\alpha_m-1)} \end{aligned} \quad (36)$$

where

$$\alpha_i \in \mathbb{R}^+, \quad \alpha_0 =: \sum_{i=1}^m \alpha_i, \quad 0 \leq y_i \leq 1, \quad \sum_{i=1}^m y_i = 1,$$

Using the square root transformation, convert the Dirichlet distribution from the simplex to the positive orthant of the hypersphere

$$x_i = \sqrt{y_i}, \quad y_i = x_i^2, \quad \frac{\partial y_i}{\partial x_i} = 2x_i, \quad \text{for } i = 1, \dots, (m-1), x_m = \sqrt{y_m}.$$

After calculating the Jacobian for each independent variable, the suggested transformation yields

$$f_{\text{SDir}}(\mathbf{x}; \boldsymbol{\alpha}) = \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^{m-1} x_i^{2\alpha_i-1} \left(1 - \sum_{i=1}^{m-1} x_i^2\right)^{(\alpha_m-1)} \quad (37)$$

where

$$\alpha_0 =: \sum_{i=1}^m \alpha_i, \quad \alpha_i \in \mathbb{R}^+, \quad 0 \leq x_i \leq 1, \quad \sum_{i=1}^m x_i^2 = 1.$$

The spherical-Dirichlet distribution, denoted by $x \sim SDD(\alpha_i)$, is the distribution found in Equation (37). Similar to the analogous parameters of the Dirichlet distribution, the parameters α_i are presented as the concentration parameters. The spherical-Dirichlet distribution is shown as a bivariate in Figure 7 (Guardiola, 2020).

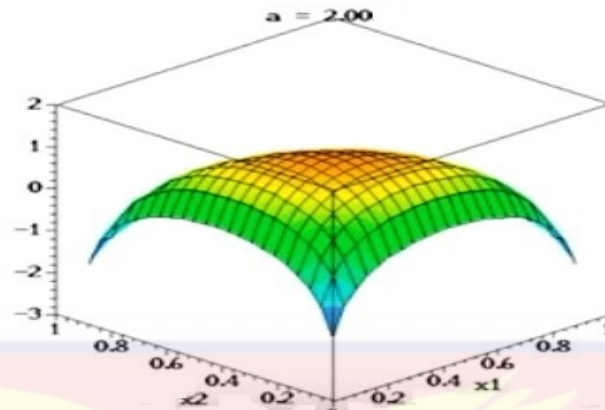


Figure 7: Bivariate Plot of the Spherical-Dirichlet Distribution

The spherical-Dirichlet distribution is a multivariate probability distribution designed to model a vector of random variables, commonly employed to represent probabilities or proportions across various categories or components. This distribution finds applications in diverse fields such as topic modeling, sentiment analysis, and material composition modeling due to its adaptability to different data patterns. Its parameters are adjustable to capture various concentration and dispersion characteristics, making it capable of representing both uniform and highly skewed data distributions. This flexibility is particularly advantageous when dealing with data exhibiting substantial variation in component proportions. In Bayesian statistics (Lan et al., 2020), the spherical-Dirichlet distribution could serve as a conjugate prior for the multinomial distribution. It has been successfully utilized in natural language processing for document-topic modeling, market research for brand choice modeling, and ecology for modeling species composition in ecological communities. The parameters of the Dirichlet distribution are interpretable as concentration parameters, influencing the distribution's concentration and, consequently, the spread of proba-

bilities among different components. Higher concentrations yield more peaked distributions, while lower concentrations result in more dispersed distributions. Despite its versatility, the Dirichlet distribution may not be universally suitable for all types of data. It assumes mutual independence among components, which may not hold true in certain cases. Additionally, it is defined within a simplex, requiring the components' values to sum to a constant, a constraint that may not align with all types of data. The spherical-Dirichlet distribution stands out as a valuable and flexible tool for modeling data characterized by proportions or compositional features. Its usefulness in Bayesian modeling, capacity to represent diverse data patterns, and interpretability contribute to its significance in statistical and machine learning. However, its appropriateness should be carefully evaluated based on specific data characteristics and modeling needs.

Multivariate Cauchy Distribution

The Multivariate Cauchy Distribution is a specialized probability distribution that broadens the univariate Cauchy distribution to encompass multivariate scenarios (Shatskikh, 1996). It is distinguished by its heavy tails and finds application in robust statistics, particularly when dealing with data containing outliers that deviate from the assumptions of conventional distributions like the Multivariate Normal Distribution.

The multivariate Cauchy distribution, much like its counterpart, the multivariate Student's t -distribution, is characterized by heavy-tailed features, signifying a propensity for extreme values. Notably, this distribution lacks well-

defined statistical moments such as mean, variance, and higher moments, setting it apart from more conventional distributions. Furthermore, the absence of a moment generating function (MGF) further distinguishes the Cauchy distribution, making it challenging to describe the entire distribution using conventional statistical measures. In scenarios where random variables are independently and identically drawn from a standard Cauchy distribution, the resulting sample mean follows the same distribution. This interesting property implies that the sample mean itself exhibits heavy-tailed behavior, reinforcing the inherent characteristics of the Cauchy distribution. It's crucial to highlight the significance of acknowledging the assumption of finite variance in the central limit theorem. Lee, Park, and Kim (2014) emphasize this point, underlining the importance of considering variance limitations to ensure the reliability and applicability of statistical inferences.

For a more profound grasp of the multivariate Cauchy distribution, readers can delve into the insightful works of Kotz and Nadarajah (2004) and Adler, Feldman, and Taqqu (1999). Park and Bera (2009) provide an extensive examination of maximum entropy probability distributions, shedding light on the standard Cauchy distribution as the maximum entropy distribution for a random variable X that satisfies the condition $E(\ln(1 + X^2)) = \ln(4)$. This emphasizes the significance of the standard Cauchy distribution in capturing the maximum amount of uncertainty for such random variables. In a novel contribution, Zhang (2010) introduces an unbiased L -estimator rooted in order statistics, expanding the methodological toolkit for statistical estimation. Going back to foundational

work, Ferguson (1962) introduces the multivariate Cauchy distribution by asserting that a random vector $\mathbf{X}$ follows this distribution if, for every real vector $\mathbf{b}$, the inner product $\mathbf{b}^\top \mathbf{X}$ yields a Cauchy-distributed random variable. This formulation implies that the distribution maintains symmetry when its mass is symmetrically distributed around a certain point in p -dimensional space, adding a geometric aspect to its characterization.

Other methods that make use of cumulative distribution functions (CDF) and CF for the Cauchy distribution are still relevant. In modern mathematical statistics, the PDF is widely used, particularly because the multivariate Cauchy density coincides with the multivariate t distribution when the degrees of freedom are set to one. As a result, the following is the formal definition of a multivariate Cauchy distribution, known as $\text{Cauchy}_p(\mu, \Sigma)$ (Kotz and Nadarajah, 2004): If we consider a p -dimensional random vector with this distribution, $\mathbf{X} = (X_1, \dots, X_p)^\top$, then its density function can be written as

$$f_{\mathbf{X}}(\mathbf{x}; \mu, \Sigma) = \frac{\Gamma\left(\frac{1+p}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \pi^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}} \left[1 + (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu)\right]^{\frac{1+p}{2}}}, \quad \mathbf{x} \in \mathbb{R}^p, \quad (38)$$

In this case, the location vector is $\mu \in \mathbb{R}^p$, and the positive-definite scale matrix is Σ . The density function reduces to

$$f_X(x; \mu, \sigma^2) = \frac{1}{\pi} \left[\frac{\sigma}{(x - \mu)^2 + \sigma^2} \right], \quad x \in \mathbb{R}. \quad (39)$$

when $p = 1$. The standard Cauchy distribution is corresponding to the specific case with $\mu = 0$ and $\sigma = 1$. As determined by Kotz and Nadarajah (2004), this distribution's CF is provided by

$$\Psi_X(t) = \exp\left(it^\top \mu - \left\| \Sigma^{\frac{1}{2}} t \right\|\right), \quad \forall t \in \mathbb{R}^p, \quad (40)$$

where $i = \sqrt{-1}$ and $|t| = \sqrt{t^T t}$. Thus, in the univariate situation ($p = 1$), $\Psi_X(t) = \exp(i\mu t - \sigma|t|)$, $\forall t \in \Re$, is the characteristic function.

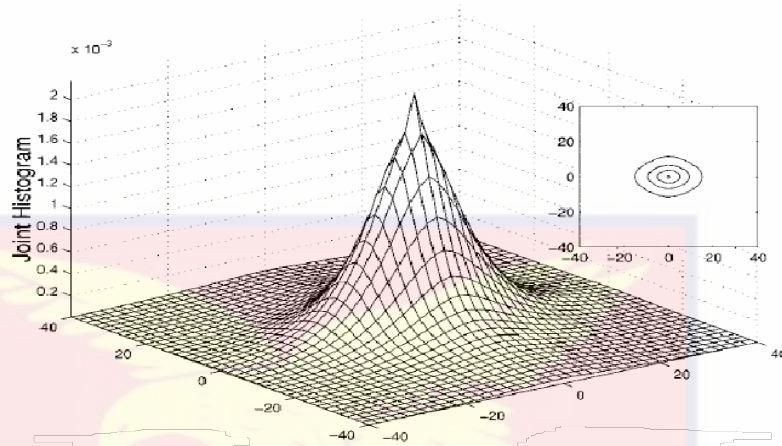


Figure 8: Bivariate Plot of Cauchy Distribution

Figure 8 presents a bivariate plot of the Cauchy distribution (Rabbani, Vafadust, Gazor & Selesnick, 2006). The multivariate Cauchy distribution lacks finite moments beyond the first, rendering the mean, variance, and higher moments nonexistent (Ateya & Madhagi, 2013). This characteristic sets it apart from the well-behaved normal distribution. Similar to its univariate counterpart, the multivariate Cauchy distribution exhibits heavy tails and sensitivity to outliers. Unlike the multivariate Student's t -distribution, it does not converge to a multivariate normal distribution with increasing sample size (Lee, Park & Kim, 2014). These properties contribute to its less frequent use in practical applications compared to the multivariate normal distribution or the multivariate Student's t -distribution. On the positive side, the Multivariate Cauchy Distribution excels in robustness. It effectively models data with extreme variability and outliers, making it valuable in scenarios where data deviates from a typical or symmetric distribution (Kotz & Nadarajah, 2004). The distribution's heavy tails

enable it to handle extreme observations without disproportionately affecting parameter estimation.

In situations where other multivariate distributions, like the Multivariate Normal Distribution, may yield unreliable results due to outliers, the Multivariate Cauchy Distribution provides more robust estimates, helping mitigate the impact of extreme data points (Lee, 2015). However, challenges accompany the Multivariate Cauchy Distribution. Its heavy tails can complicate parameter estimation, and it may not match the predictive accuracy of more typical distributions for data adhering to a standard distribution. Additionally, the heavy-tailed nature can lead to undefined moments, such as the mean and variance, posing challenges in statistical analyses. The Multivariate Cauchy Distribution is a robust tool in statistics, particularly for data with outliers and extreme observations (Kotz & Nadarajah, 2004; Bouhlel & Rousseau, 2022). Its robustness to heavy-tailed data makes it applicable in various scenarios, but caution is necessary, considering challenges in parameter estimation and potential issues with moments like the mean and variance. It should be used judiciously, especially in complex datasets departing from the assumptions of more traditional distributions (Sadreazami, Ahmad & Swamy, 2016).

Multivariate Logistic Distribution

The multivariate logistic distribution serves as an extension of the logistic distribution into multiple dimensions (Ashby, 2014). It finds application in multivariate logistic regression, a statistical approach for examining the re-

relationship between several independent variables and a categorical dependent variable (Glonek & McCullagh, 1995; Ritzema, 1994; Althuwaynee, Pradhan, Park & Lee, 2014). The literature has proposed a number of multivariate logistic distributions, each with unique properties. A prominent illustration is the logit-normal distribution, which applies a logistic transformation on a multivariate normal distribution and expands the idea of the logit-normal distribution to D -dimensional probability vectors (Balakrishnan, 2013). An further variation is the multivariate logistic distribution, which was presented by Fang and Xu. Its marginal distributions take on univariate forms.

A random vector, $\mathbf{x}$, with location parameter μ and scale parameter Σ , has the following PDF for the multivariate logistic distribution

$$f(\mathbf{x}; \mu, \Sigma) = \frac{\exp\left(-(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right)}{\left[1 + \exp\left(-(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right)\right]^{p+1}} \quad (41)$$

where $\mathbf{x}$ is the p -dimensional random vector, μ is the mean vector and Σ is the $p \times p$ scale (covariance) matrix. The CDF of the multivariate logistic distribution lacks a closed form, necessitating the use of numerical methods for probability computations (Kotz, Balakrishnan & Johnson, 2000). Moreover, as is common with many multivariate distributions, it is crucial to specify dependence through the covariance matrix. The Multivariate Logistic Distribution is a probability distribution extensively employed for modeling multivariate data across diverse fields, including statistics, machine learning, and image processing (Ali et al., 2016; Yue, Zhao, Mao & Liu, 2015). This distribution serves as an extension of the univariate logistic distribution to the multivariate field, making it a valuable tool for characterizing data in multiple dimensions. Figure

9 presents a bivariate plot of Logistic Distribution (Gumbel, 1961; Dehghani & Fadaee, 2020).

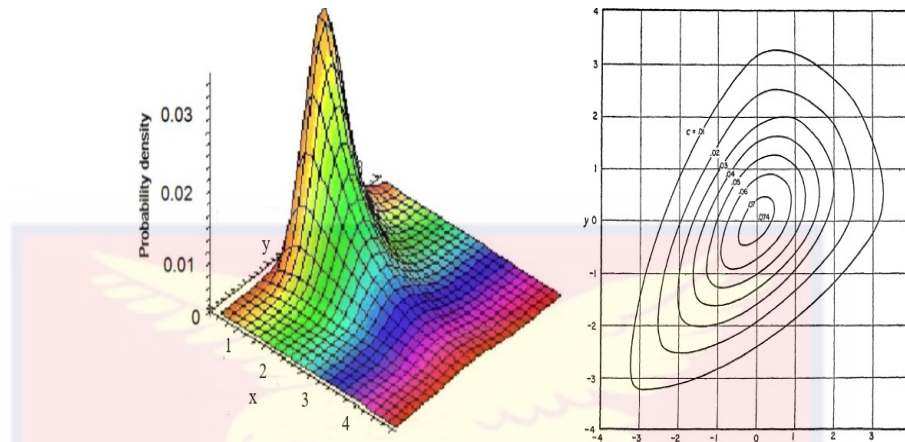


Figure 9: Bivariate Logistic Distribution

A notable strength of the Multivariate Logistic Distribution lies in its capacity to represent data with bounded support, featuring a sigmoidal or “S-shaped” PDF (Di Crescenzo, Paraggio, Román-Román & Torres-Ruiz, 2022). This proves particularly advantageous in scenarios where data exhibit nonlinear relationships or transitions between different categories or states. In machine learning, for example, the Multivariate Logistic Distribution finds utility in multinomial logistic regression, where it models the probabilities of multiple classes in a classification problem (Itano, Ueki, Iizuka & Kuwatani, 2020).

The distribution’s sigmoidal shape is advantageous for modeling probabilities or proportions, ensuring that probabilities sum to 1 in multiclass classification problems (Noorae, Abegaz, Ormel, Wit, & van den Heuvel, 2016). Additionally, the Multivariate Logistic Distribution is computationally manageable, offering a significant advantage in applications requiring parameter estimation and likelihood calculations. It is well-suited for maximum likelihood estimation

(Sur & Candès, 2019). However, it is essential to recognize that the Multivariate Logistic Distribution may not be universally applicable. It is most suitable when data follow a multinomial distribution or when modeling the probabilities of multiple classes or categories (Selig, Shaw & Ankerst, 2020). In situations where data lack a sigmoidal shape or are not naturally bounded between 0 and 1, alternative multivariate distributions may be more appropriate. The Multivariate Logistic Distribution stands as a valuable tool for modeling multivariate data characterized by sigmoidal, bounded, and probabilistic behavior. It finds applications in machine learning (Brungard et al., 2015; Lhéritier, Bocamazo, Delahaye & Acuna-Agost, 2019; Korotcov, Tkachenko, Russo & Ekins, 2017) and statistics (Selig, Shaw & Ankerst, 2020; Pei, Dibeklioglu, Tax & van der Maaten, 2017), particularly in scenarios involving classification problems with multiple classes (Felicísimo, Cuartero, Remondo & Quirós, 2013). Its computational manageability and innate representation of probabilities render it a useful choice when aligned with the characteristics of the analyzed data.

Multivariate Normal Distribution

The Multivariate Normal Distribution holds significant prominence in statistics, probability theory, and various domains of data analysis (Johnson & Wichern, 2013). It is extensively employed for modeling multivariate data and is particularly prevalent in areas like finance (Punzo & Bagnato, 2020; Madan, 2020; DiBmann, Brechmann, Czado, & Kurowicka, 2013), economics (Osnas, Lichstein, Reich & Pacala, 2013; Cleff, 2019) agriculture (Mittal & Mehar, 2015;

Mehar, Mittal & Prasad, 2016), geography (Yan, Zhao & Xu, 2023) and engineering (Yan, 2023; Godwin, 2016; Phoon, 2019). Its primary utility lies in its capacity to describe the collective behavior of multiple random variables, making it indispensable for multivariate statistical analysis and hypothesis testing. This distribution represents a broadened version of the familiar bell-shaped Gaussian density into multiple dimensions, playing a crucial role in applied multivariate analysis (Johnson & Wichern, 2013). While it is acknowledged that real-world data rarely conforms precisely to a multivariate normal distribution, the normal density often proves to be a valuable approximation of the “true” population distribution (Rencher & Christensen, 2012). The appeal of the multivariate normal distribution lies in its mathematical manageability, leading to effective outcomes, an advantage not shared by other data-generating distributions. Despite the fact that real data seldom perfectly follows the multivariate normal distribution, its approximation is frequently practical and advantageous.

The Gaussian distribution’s practicality in real-world applications is rooted in two key factors. Firstly, it can serve as an exact population model under certain circumstances. Secondly, due to a central limit effect, the sampling distributions of numerous multivariate statistics tend to be approximately normal, irrespective of the underlying population’s form (Anderson, 2003). In essence, many real-world challenges naturally align with Gaussian theory. The fundamental significance of the normal distribution lies in its dual functionality: serving both as a population model for specific natural phenomena and as an approximate sampling distribution for a wide range of statistical measures. The

multivariate normal distribution is characterized by its PDF, represented as follows

$$f(x, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{-\frac{1}{2}}} \exp \left[-\frac{1}{2} (x - \mu)^{\top} \Sigma^{-1} (x - \mu) \right] \quad (42)$$

Here, $\mu \in \Re^p$ and Σ is a positive definite symmetric matrix of size $(n \times n)$, representing the location and scale parameters, respectively. The parameters μ and Σ play crucial roles in defining the characteristics of the multivariate normal distribution. Figure 10 illustrates a bivariate plot of the multivariate normal distribution, providing a visual representation of its shape and properties (Ashby & Lee, 1991; Rencher & Christensen, 2012).

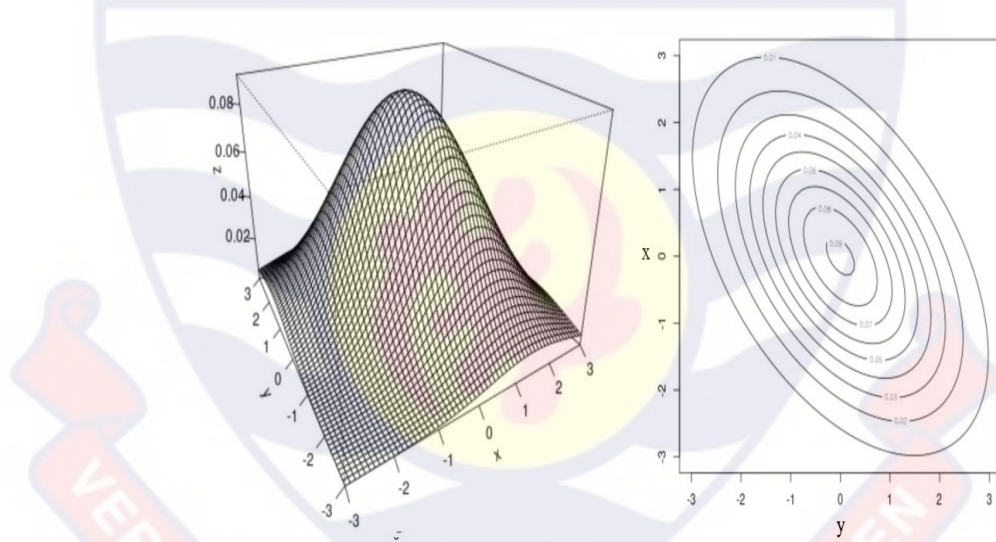


Figure 10: Bivariate Normal Distribution Plot

The Multivariate Normal Distribution, commonly known as the multivariate Gaussian distribution, is a fundamental and extensively utilized probability distribution in statistics and data analysis (Johnson & Wichern, 2013). Its significance extends across diverse fields, encompassing multivariate statistical analysis (Rencher & Christensen, 2012; Tabachnick, Fidell & Ullman, 2018; Everitt & Hothorn, 2011; Chatfield, 2018), machine learning (Maghrebi, Portigliatti &

Prouff, 2016; Sun Yang, Zhang, Ling & Peng, 2020; Murphy, 2023; Zhou, Li & Zare, 2017), time series (Malhotra, Vig, Shroff & Agarwal, 2015; Li, Chen, Goh & Ng, 2018; Pesaran, 2015), and various scientific disciplines. Acknowledged for its flexibility, it is frequently the default choice for modeling multivariate data, accommodating diverse patterns such as spherically symmetric, ellipsoidal, and correlated or dependent data. With its mean vector and covariance matrix as defining features, it aptly represents a wide array of data structures. This distribution is distinguished by well-understood and documented properties, where its mean vector and covariance matrix provide a precise portrayal of central tendency and relationships between variables. The assumption of multivariate normality is pivotal in numerous statistical techniques, including linear regression, factor analysis, and principal component analysis, facilitating the application of potent analytical tools for modeling and analyzing intricate data sets (Rencher & Christensen, 2012; Anderson, 2003; Johnson & Wichern, 2013).

Boasting desirable mathematical properties such as linearity and independence under linear transformations, the Multivariate Normal Distribution proves valuable in statistical modeling, hypothesis testing and Bayesian inference (Alvarez, Niemi & Simpson, 2014; Peterson, Stingo & Vannucci, 2015; Mitropoulos, Talias & Mitropoulos, 2015; Campbell & Broderick, 2019). Its role in the Central Limit Theorem is significant, as it underpins various statistical techniques and justifies the widespread use of the normal distribution (Anderson, 2003). Applications of the Multivariate Normal Distribution span diverse fields, including physics (De Figueiredo et al., 2017; Yan, Ma, Huang & Zhao, 2021),

and image processing (Cannon, 2017; Sisodia, Tiwari & Kumar, 2014), serving as the foundation for statistical methodologies like principal component analysis (PCA) (Bro & Smilde, 2014; Dray & Josse, 2014) and linear discriminant analysis (LDA) (Safo & Ahn, 2016; Boedeker & Kearns, 2019; Zhu, Gao, Yang & Ye, 2022). However, its universal applicability is not guaranteed, particularly when data deviate from normal distribution assumptions or exhibit non-linearity, necessitating consideration of alternative distributions or methods.

Despite its versatility, the assumption of multivariate normality may not always hold in practical applications, especially with real-world data featuring non-linear relationships or non-normally distributed variables (Nkansah et al., 2018). In such instances, alternative distributions or robust statistical techniques become necessary. Nevertheless, the Multivariate Normal Distribution remains a cornerstone in multivariate statistical analysis, providing a robust foundation for comprehending the joint behavior of multiple variables and facilitating the application of diverse analytical tools and methodologies. Its well-documented properties and mathematical tractability make it an indispensable tool for understanding and analyzing complex data sets across a variety of fields. While its assumptions may not universally hold in real-world scenarios, its versatility, mathematical elegance, and widespread applications underscore its significance in multivariate statistics and data analysis. Due to its assumptions and limitations, opting for alternative models such as multivariate EPD is necessary when the data deviate from these assumptions.

The Multivariate Exponential Power Distribution

The multivariate EPD is an intuitive extension of exponential families of distributions which offers the advantage of accommodating thicker tails in comparison to the normal distribution. M. T. Subbotin, in 1923, introduced and established the uni-dimensional version of the EPD, a framework that inherently includes the normal distribution as a special case. In this foundational paper, Subbotin introduced the distribution under the moniker “generalized Laplace distribution.” Subbotin’s pioneering work laid the groundwork for subsequent research and applications of the Exponential Power Distribution. His identification of the distribution as a generalized Laplace distribution reflected its characteristics and established it as a broader distribution encompassing the well-known Laplace distribution. This early recognition set the stage for the further development and analysis of the EPD across various domains and applications. The EPD has garnered attention under various names and terms within the academic and research community. Researchers and scholars have employed diverse terminology to refer to this distribution, showcasing its versatility and applications across different fields. Some of the alternative names used for the EPD include

1. Exponential Power Distribution: Coined by Box and Tiao (1973) and acknowledged by Landsman and Valdez (2003), this term emphasizes the distribution’s characteristic exponential power behavior.
2. Power Exponential Distribution: Gómez et al. (1998), Dang (2014), and

Dang, Browne, and McNicholas (2015) have employed this term to highlight the power-law-like behavior of the distribution.

3. Generalized Error Distribution: Researchers such as Forbes, Evans, Hastings, and Peacock (2011), Giller (2005), and Leemis and McQueston (2008) have referred to the EPD as the generalized error distribution, emphasizing its broader applicability.
4. Generalized Exponential Distribution: Morgan (1996) used this term, underlining the distribution's generalization of the exponential family of distributions.
5. Generalized Normal: Nadaraja (2005) employed this term to describe the EPD, emphasizing its generalization beyond the normal distribution.
6. Generalized Gaussian: The distribution has also been termed as the generalized Gaussian by researchers such as Zhang, Wiesel, and Greco (2013), Boukouvelas, Fu, and Adalı (2015), Varanasi, and Aazhang (1989), Pascal et al. (2013), and Dytso, Bustin, Poor, and Shamai (2018). This reflects the distribution's characteristics resembling the Gaussian distribution in a generalized form.

The adoption of these diverse terms underscores the multifaceted nature of the EPD and its applicability in various statistical and mathematical contexts. Continued exploration of the EPD has led to the proposal of various formulations for its univariate case by several researchers. This diversity in formulations reflects ongoing efforts to understand and capture the nuanced behavior

of the EPD in different contexts. Notable contributions in this regard include Coin (2017)'s work presented a distinct formulation for the univariate case of the EPD, contributing to the ongoing evolution of its parametric representations. Giller (2005)'s research offered another perspective on the univariate form of the EPD, likely introducing unique parameterizations or insights into its statistical properties. Nadarajah (2005)'s contributions added to the spectrum of formulations for the univariate EPD, potentially offering novel perspectives on its representation. The collaborative Pogány and Nadarajah (2009) likely introduced further refinements or alternative expressions for the univariate EPD, enriching the available parametric options. Tahir et al. (2015) contributed to the field by proposing yet another formulation for the univariate EPD, potentially introducing innovative parameter choices or statistical characteristics. Nkansah et al. (2018) is particularly significant as it laid the groundwork for the generalized form of the univariate EPD. Their introduction of a constant, c , that separates the shape parameter, q , from the scale parameter, σ , represents a pivotal step in the evolution of EPD formulations.

These researchers have collectively expanded the landscape of univariate EPD formulations, providing a rich array of options for statisticians, researchers, and practitioners to choose from based on the specific requirements of their applications. The ongoing exploration of different formulations contributes to a more comprehensive understanding of the distribution's behavior and potential applications.

The multivariate EPD has its roots in the cooperative work of Gómez

et al. (1998), and Equation (2) illustrates how it was formulated. Particularly noteworthy is the diversity of behaviours exhibited by this distribution, which depend on the shape parameter β . The distribution adopts a leptokurtic profile, which is distinguished from the Gaussian distribution by a narrow peak and heavy tails, when the value of β falls within the range of $0 < \beta < 1$. However, when $\beta > 1$, the distribution resembles a platykurtic distribution with thinner tails and a wider peak than a Gaussian distribution.

The multivariate EPD is particularly flexible because it takes on well-known shapes for certain values of β : at $\beta = 0.5$, it resembles a Laplace (double-exponential) distribution, and at $\beta = 1$, it resembles a Gaussian distribution. Furthermore, the distribution converges to a multivariate uniform distribution when the shape parameter β gets closer to infinity. Because of its versatility and capacity to switch between various distributional forms, the multivariate EPD is an invaluable instrument with a broad variety of uses in statistical modelling and analysis.

The extension of this distribution to multiple dimensions holds significant importance, especially in the modeling of random phenomena in multidimensional spaces with tail behaviors that deviate from the conventional normal distribution. The multidimensional generalization of the exponential power distribution proves to be a valuable tool in statistical modeling, allowing for a more nuanced representation of the complexities present in diverse datasets.

The application of the multivariate EPD contributes notably to the robustness of various multivariate statistical procedures. By accommodating a broader

range of tail behaviors, this distribution enhances the ability to capture and analyze real-world data that may exhibit deviations from the assumptions of normality. The versatility of the multivariate EPD makes it a valuable asset in fields such as finance, economics, and environmental science, where data often exhibit non-normal characteristics.

It is essential to highlight that the multivariate EPD belongs to the category of elliptically contoured distributions. This subclass of distributions provides a broader framework for understanding and modeling multivariate data with varying degrees of skewness and kurtosis. The incorporation of the multivariate EPD into this family of distributions expands the toolkit available for statisticians and researchers dealing with complex multidimensional datasets.

The multivariate case have been studied by several researchers (Gómez et al., 1998; Gómez-Sánchez-Manzano et al., 2008; Solaro, 2004; Kozubowski, 2002; Korolev, 2020; Ikeshita & Kawaguchi, 2018; Marnissi, Benazza-Benyahia, Chouzenoux & Pesquet, 2013; Olosunde, 2013; Ouzir, Pesquet & Pascal, 2022; Soyinka & Olosunde, 2021; Olosunde & Soyinka, 2020; Barnard, Trindade & Wickamasinghe, 2014; Sedighi, V., Fridrich & Cograanne, 2015; Zhang, Wiesel & Greco, 2013; Pascal et al., 2013; Verdoolaege & Scheunders, 2011, 2012; Bouhlel & Dziri, 2019; Boukouvalas, Fu & Adalı, 2015; Bombrun, Pascal, Tourneret & Berthoumieu, 2012; Verdoolaege, De Backer & Scheunders, 2008; Najar, Bourouis, Bouguila & Belghith, 2019; Hristova, Le Meur, Cozot & Bouatouch, 2017; Omari, Abdelouahad, El Hassouni & Cherifi, 2013; Najar, Bourouis, Al-Azawi & Al-Badi, 2019; Wang, Zhang, Zhao & Sun, 2021; Li,

Cao, Hu, Gan & Wu, 2020; Kwitt, Meerwald & Uhl, 2009).

Olosunde (2013) introduced a comprehensive formulation of the EPD, presented in Equation (7), which unifies various distributions from the existing literature as specific cases. This inclusive EPD incorporates a parameter denoted as ω and a function $h(\omega)$ that governs the tail behavior of the distribution. This feature enhances its flexibility, rendering it a more suitable modeling option compared to both the standard normal distribution and the distribution proposed by Gomez et al. (1998), while still maintaining symmetry. It is noteworthy that Olosunde did not explicitly delineate how the function $h(\omega)$ regulates the tails.

In relationship to the work of Olusunde (2013), the work by Nkansah et al. (2018) introduced a generalized form of the univariate EPD, setting the stage for the extension of the univariate EPD to its multivariate counterpart in the this study. They derived the distribution as presented in Equation (4) with the following properties: Normalising constant k is given as

$$k = \frac{1}{2c^{-1/q} \left[\Gamma\left(\frac{1}{q} + 1\right) \right]}. \quad (43)$$

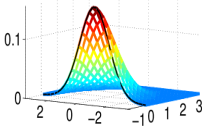
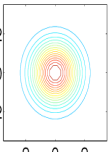
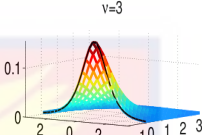
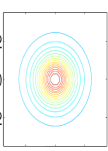
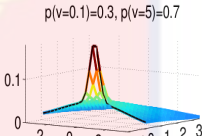
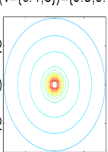
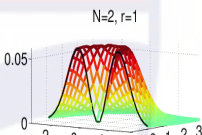
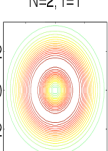
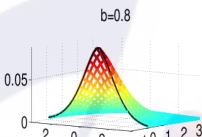
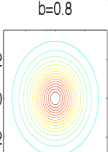
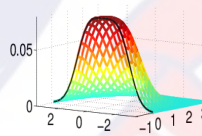
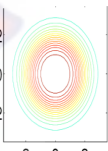
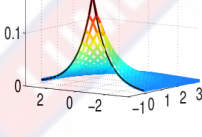
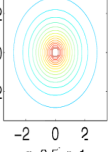
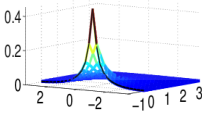
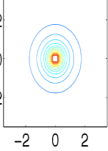
The coefficient of Kurtosis:

$$\alpha_4 = \frac{\left[\Gamma\left(\frac{5}{q}\right) \right] \left[\Gamma\left(\frac{1}{q}\right) \right]}{\left[\Gamma\left(\frac{3}{q}\right) \right]^2} - 3. \quad (44)$$

The key innovation lies in their incorporation of a variable constant, denoted as c , within the formulation. This strategic inclusion ensures that the shape parameter, denoted as q , remains decoupled from the scale parameter, denoted as σ . This fundamental insight proves to be pivotal in guiding the methodology for the extension of the univariate EPD to its multivariate version in the ongoing-

ing investigation. By adopting this constant c , Nkansah et al. have effectively disentangled the influence of the shape parameter on the scale parameter. This separation simplifies the modeling process and enables a more straightforward and interpretable characterization of the distribution. The independence of q and σ facilitates a more flexible and versatile application of the multivariate EPD, allowing for a nuanced representation of complex data structures in multiple dimensions. As a result, this insight not only streamlines the theoretical underpinnings of the multivariate EPD but also enhances its practical utility. The present study leverages this crucial groundwork to extend the univariate EPD to the multivariate context, aiming to provide a comprehensive and robust framework for modeling and analyzing multidimensional data with varying tail behaviors. The independence between shape and scale parameters, guided by the constant c , forms a cornerstone in the development of a versatile and effective multivariate EPD, contributing to the broader landscape of statistical modeling and analysis. Consequently, this study will propose analogous properties to ensure that the shape parameter does not influence the scale parameter in the multivariate exponential power distribution.

Table 1: Summary of Some Class of Elliptical Distributions

Law	Probability density	Param.	2D-visualization	Contours
Gaussian $\mathcal{N}_n(t; \mu, \sigma^2)$	$p(y) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\ y-\mu\ ^2}{2\sigma^2}}$			
Student $\mathcal{T}_n(t; \mu, \sigma^2, \nu)$	$p(y) = \frac{\Gamma(\frac{n+\nu}{2})}{(\pi\sigma^2\nu)^{\frac{n}{2}}\Gamma(\frac{\nu}{2})} \left(1 + \frac{\ y-\mu\ ^2}{\nu\sigma^2}\right)^{-\frac{(\nu+n)}{2}}$	$\nu \geq 1$		
Gaussian mixtures $\mathcal{GM}_n(t; \mu, \sigma^2, G)$	$p(y) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_0^\infty \frac{1}{v^{\frac{n}{2}}} e^{-\frac{\ y-\mu\ ^2}{2v\sigma^2}} G(dv)$	$\int G(dv) = 1$		
Kotz $\mathcal{K}_n(t; \mu, \sigma^2, N, r)$	$p(y) = \frac{\Gamma(\frac{n}{2}) r^{\frac{2N-2+n}{2}}}{\pi^{\frac{n}{2}} \sigma^{n+2N-1} \Gamma(\frac{2N-2+n}{2})} \ y-\mu\ ^{2(N-1)} e^{-r\frac{\ y-\mu\ ^2}{\sigma^2}}$	$N > \frac{2-n}{2}, r > 0$		
Exponential power $\mathcal{EP}_n(t; \mu, \sigma^2, b)$	$p(y) = \frac{n\Gamma(\frac{n}{2})}{(\pi\sigma^2)^{\frac{n}{2}} 2^{1+\frac{1}{2b}} \Gamma(1+\frac{n}{2b})} e^{-\frac{1}{2}\left(\frac{\ y-\mu\ ^2}{\sigma^2}\right)^b}$	$b > 0$		
Logistic $\mathcal{Log}_n(t; \mu, \sigma^2)$	$p(y) = \frac{(\sum_{j=1}^\infty (-1)^{j-1} j^{1-n/2})^{-1}}{(2\pi\sigma^2)^{\frac{n}{2}}} \frac{e^{-\frac{\ y-\mu\ ^2}{2\sigma^2}}}{\left(1 + e^{-\frac{\ y-\mu\ ^2}{2\sigma^2}}\right)^2}$			
Laplace $\mathcal{L}_n(t; \mu, \sigma^2)$	$p(y) = \frac{1}{2^{\frac{n}{2}-1} (\pi\sigma^2)^{\frac{n}{2}} \Gamma(\frac{n}{2})} K_0\left(\sqrt{2}\frac{\ y-\mu\ }{\sigma}\right)$			
Bessel $\mathcal{B}_n(t; \mu, \sigma^2, q, r)$	$p(y) = \frac{(-1)^{q+1} (2r)^{-q-n} \pi}{\pi^{\frac{n}{2}} \Gamma(q+\frac{n}{2}) \sin(q\pi)} \ y-\mu\ ^q I_q\left(\frac{\sqrt{\ y-\mu\ ^2}}{r}\right),$ with $I_q(z) = \sum_{k=0}^\infty \frac{1}{k! \Gamma(k+q+1)} \left(\frac{z}{2}\right)^{q+2k}$	$q > -\frac{n}{2}, r > 0$		

Source: Boisbunon (2012)

Review of Some Applications of Multivariate EPD

Several investigations have focused on enhancing and applying the multivariate EPD. Originating from Subbotin's work in 1923, Gómez et al. (1998) contributed significantly by introducing a multivariate extension of the family of the distribution. This distribution, known for its versatility, has been explored in depth, providing a definition, key properties, and a simulation procedure for generating observations. The investigation expanded with the proposal of a matrix variate generalization by Sánchez-Manzano et al. (2002), enhancing statistical procedures in multivariate analysis.

In the field of elliptical distributions, the multivariate EPD is situated (Fang & Zhang, 1990; Fang, Kotz & Ng, 1991), originally introduced by Kelker in 1970 and studied by Rangaswamy, Weiner, and Ozturk (1993, 1995). Depending on the shape parameter's value, common properties emerge between the multivariate EPD and distributions of spherically invariant random vectors (Gómez-Sánchez-Manzano et al., 2008; Ollila, Tyler, Koivunen & Poor, 2012). Notably, Gómez-Sánchez-Manzano et al. (2008) demonstrated its role as a scale mixture of normal distributions, paving the way for applications in quantitative genetics.

The univariate EPD, widely used in Bayesian analysis, signal processing, astronomy, and insurance. Dytso et al. (2018) provide a comprehensive survey outlining applications of univariate EP distributions. Similarly, the multivariate EPD serves as a valuable tool in applications such as image processing, speech recognition, ocean engineering, and finance. Its significance is particularly pro-

nounced when handling scenarios involving non-Gaussian tails. Despite its historical origins, the multivariate EPD continues to be widely applied across diverse disciplines, emerging as the preferred distribution in situations where non-Gaussian tail behavior is evident in the data. Numerous papers across journals and monographs endorse the multivariate EPD as the preferred distribution, and while it is challenging to cover all, this discussion aims to highlight what is deemed the most interesting and promising applications, taking into account a diverse audience of potential researchers in multivariate EPD.

Gómez-Sánchez-Manzano et al. (2008) and Pascal et al. (2013) enumerate specific fields where multivariate EPD finds application. Kozubowski (2002) delved into the vertical density of multivariate EPD within a family of spherically symmetric distributions on $\mathfrak{R}^p$, revealing that EPDs possess the unique property of having a uniform vertical density. In Solaro's work (2004), two procedures for random variate generation for the family of Multivariate EPDs were proposed and compared. Solaro presented distributions in terms of a mixture of random variables through a transformation approach, offering a simple and efficient procedure. An alternative procedure, based on a general acceptance-rejection (AR) technique for multivariate orthounimodal densities, was also explored. Solaro emphasized that the transformation-based procedure significantly outperformed the alternative in terms of computational efficiency. His study contributed valuable insights into evaluating the computational efficiency of generalizations of AR methods compared to transformation methods for specific multivariate distributions.

The multivariate EPD has been widely applied within the field of image processing. Offering a versatile modeling approach, it stands out as a valuable tool for capturing the statistical properties of diverse images and the features derived from them. Notably, Gaussian and Laplacian distributions are encompassed as special cases within the broader framework of multivariate EPD. Researchers, such as Ollila, Tyler, Koivunen, and Poor (2012), as well as Chang, Yu, and Vetterli (2000), have found that the distribution of coefficients derived from wavelet or curvelet transformations can be effectively and accurately modeled using EPDs. This adaptability highlights the utility of multivariate EPDs in accurately representing the intricate statistical characteristics inherent in image data. This modeling property has been effectively applied in image processing tasks such as image denoising (Boubchir & Fadili, 2006; Moulin & Liu, 1999; Cho & Bui, 2005), context-based image retrieval (Do & Vetterli, 2002), image thresholding (Verdoolaege, De Backer & Scheunders, 2008), and texture classification in industrial problems (Bazi, Bruzzone & Melgani, 2007).

Furthermore, EPDs find applications in radar (Scharcanski, 2007), video coding and denoising (Coban & Mersereau, 1996; Desai & Mangoubi, 2003; Bicego, Gonzalez-Jimenez, Grosso & Castro, 2008; Yang, Wang, Xu & Dai, 2009), and biomedical signal processing (Elguebaly & Bouguila, 2010). Complex EPDs have also been investigated in works by Cam, Belghith, Collet, and Salzenstein (2009), and Novey, Adali and Roy (2009), while Ikeshita and Kawaguchi (2018) focused on independent low-rank matrix analysis based on multivariate complex EPD. Multivariate regression models with Exponential

Power errors were explored by Liu and Bozdogan (2008).

Amidst the growing interest in Exponential Power Distributions (EPDs), the issue of parameter estimation emerges as a pivotal concern. Traditional methodologies employed in the context of univariate EPDs, including maximum likelihood (ML) and the method of moments, have been extensively investigated, as documented by Agro (1995), and Varanasi and Aazhang (1989). In the transition to the multivariate domain, diverse approaches are available for estimating parameters. Least-squares methods, as demonstrated by Cho and Bui (2005), present a viable avenue for parameter estimation in multivariate EPDs. An alternative method involves the minimization of the χ^2 distance between the observed data histogram and the theoretical probabilities associated with the multivariate EPD, as explored by Khelil-Cherif and Benazza-Benyahia (2004). This technique leverages the goodness-of-fit criterion to align observed and expected distributions, contributing to accurate parameter estimation.

Furthermore, the method of moments and ML methods, known for their versatility and robustness, have been proposed as effective estimation strategies in the multivariate EPD context. Verdoolaege and Scheunders (2011, 2012), and Zhang, Wiesel and Greco (2013) have delved into the application of these methods, shedding light on their utility in determining parameters for the multivariate EPD. These estimation techniques play a critical role in enhancing the accuracy and reliability of the multivariate EPD model, allowing it to better capture the underlying statistical characteristics of complex data sets.

In a notable contribution by Bouhlel and Dziri (2019), a closed-form ex-

pression for the Kullback–Leibler divergence (KLD) between zero-mean multivariate EPD was developed. The Lauricella D-hypergeometric function was introduced as a similarity measure, and an efficient method for computing this function was provided. Through Monte-Carlo sampling with a substantial sample size, they demonstrated the convergence of the approximated KLD to its theoretical value. This work is particularly significant in signal and image processing, offering a tailored KLD expression for zero-mean multivariate EPD and confirming the convergence of the Monte-Carlo sampling approximation as the sample size increases.

Olusunde (2013) conducted research investigating the mathematical and statistical properties of the univariate EPD. Parameter estimation was carried out using both moments and maximum likelihood methods. Extending the analysis to the multivariate case, Olusunde applied the univariate EPD and the normal distribution to data related to poultry production. His findings favored a significantly better fit for the generalized exponential power distribution.

In their comprehensive investigation, Pascal, Bombrun, Tourneret, and Berthoumieu (2013) delved into the field of parameter estimation for the multivariate EPD by utilizing the maximum likelihood estimator (MLE). The central challenge of accurately estimating these parameters was effectively addressed through their rigorous methodology. Notably, the authors not only established the existence and uniqueness, up to a scalar factor, of the MLE for the scatter matrix but also introduced a sophisticated estimation algorithm grounded in Newton-Raphson recursion. This algorithm was intricately crafted to cater

specifically to the computation of the MLE for parameters associated with the multivariate EPD. Highlighting the prowess of the maximum likelihood principle, the authors underscored the efficiency of their proposed approach. Their study suggests that the introduced methodology holds significant promise as a foundational and robust method for achieving precise parameter estimation within the context of the multivariate EPD.

In Lindsey's (1999) investigation, the primary focus centered on the multivariate EPD, a pivotal member within the multivariate elliptically contoured family. The study highlighted the considerable importance of the multivariate EPD, positioning it as a valuable extension of the multivariate normal distribution, particularly when addressing the modeling intricacies of repeated measurements. Notably, the distribution was adept at accommodating datasets characterized by both light and heavy tails. Lindsey emphasized the enduring consistency in interpreting the covariance matrix, which streamlined the process of structuring for serial dependence and seamlessly incorporating multiple levels of variance components. The practical application of multivariate EPD analysis was exemplified in a crossover trial involving the administration of insulin to rabbits, featuring repeated measurements within each period. The study deliberately incorporated autocorrelation considerations and systematically addressed two distinct levels of variance components. The outcomes of the analysis revealed a remarkable superiority in the fit of the multivariate EPD when juxtaposed against the skew multivariate Laplace distribution (as documented by Kotz et al., 2001). Intriguingly, the performance of the multivariate Laplace

distribution proved to be less satisfactory than that of the standard multivariate normal distribution. Additionally, the application of a multivariate Student t -distribution, characterized by an estimated 13.8 degrees of freedom, exhibited a fitting comparable to that achieved by the multivariate EPD.

Olosunde and Soyinka (2013) employed discriminant analysis and classification methods, utilizing Fisher's Linear Discriminant Function (LDF), to assess a discriminant function. They aimed to understand differences between two groups of eggs fed distinct combinations of feeds. With a sample of 96 chickens, split into two groups of 48 each (one receiving inorganic copper salt and the other organic copper salt), essential attributes like egg size and cholesterol level were measured. They successfully derived an efficient discriminant function, applicable for classifying eggs with unknown feeds into their respective groups. In their subsequent study (Olosunde & Soyinka, 2020), Olosunde and Soyinka directed attention towards the prevalent assumption of a normal distribution in the categorization of multivariate data based on attributes, a convention deeply ingrained, especially when the distribution of the data deviates from normality in both tails. To tackle this challenge, they introduced the multivariate EPD as an innovative model for discrimination and classification, providing an elliptically contoured alternative to the norm. The distinguishing feature of the multivariate EPD lies in its inclusion of a shape parameter that governs tail behavior, effectively addressing concerns related to both lighter and heavier-tailed data. This expanded conceptualization of the normal distribution resulted in a notable reduction in misclassification errors in discrimination and classification

tasks. In real data applications, Olosunde and Soyinka undertook a comprehensive comparative analysis, pitting their proposed approach against Fisher's LDF. The outcomes of the study revealed the multivariate EPD's superior performance, positioning it as a highly favorable substitute for the normal distribution, particularly when dealing with data displaying tails that are either shorter or longer than the typical normal distribution. The authors emphasized that the discrimination model grounded in the multivariate EPD surpassed those relying on the normal distribution and double EPD, especially when confronted with data exhibiting tails distinct from the customary normal distribution.

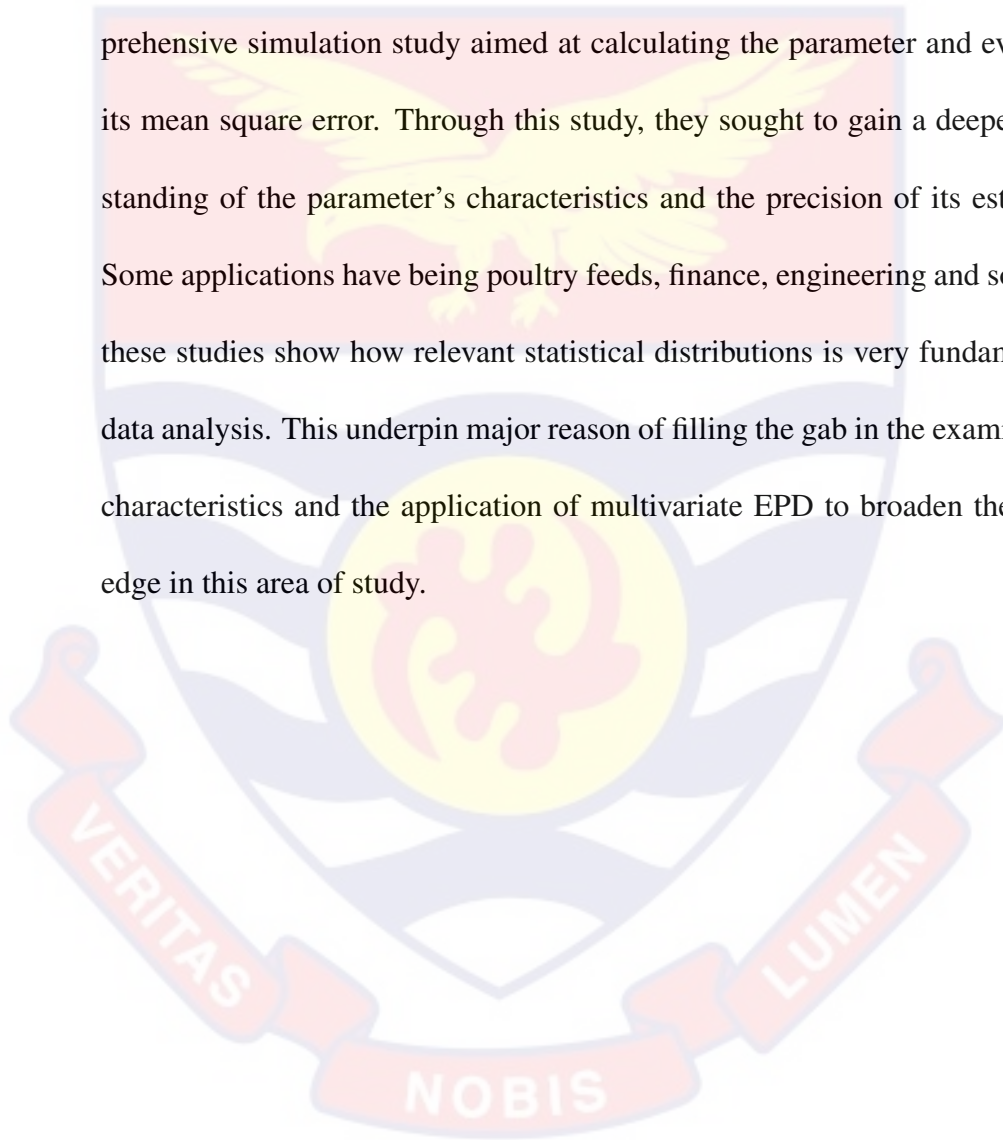
Chapter Summary

Statistical distributions describe the uncertainty nature of variables. It plays a crucial role in representing features of a data as adequate as possible. It was explained that Statistical distributions capture the aspects of a phenomenon that are relevant. In this chapter, relevant literature on multivariate EPD has been reviewed. It covered the basic concept of spherical and elliptical distributions. Some class of elliptical distributions was discussed. The subclasses of elliptical distribution discussed were multiuniform distributions, symmetric Kotz type distributions, symmetric multivariate Pearson Types VII and II distributions and Symmetric Multivariate Laplace Distribution. In addition to exploring various other distribution subclasses, including the Bessel Distribution, multivariate Student's t -Distribution, multivariate Cauchy Distribution, multivariate Logistic Distribution, and multivariate Normal Distribution, considerable atten-

tion has been devoted to the comprehensive study of the multivariate EPD over the years. Researchers have delved into the manifold applications of the multivariate EPD, revealing its versatility in diverse fields. The multivariate Normal distribution, in particular, has emerged as a widely embraced distribution due to its ease of use in mathematical operations and has found practical applications in diverse domains such as IQ scores, velocity of gas molecules, and individual's height. An interesting characteristic of the multivariate EPD is its equivalence to the Normal distribution when the shape parameter β assumes a value of 1. Additionally, the chi-square distribution plays a crucial role in assessing the conformity of a frequency distribution to a specific pattern and demonstrates a noteworthy relationship with the Exponential Power Distribution, particularly when dealing with high degrees of freedom. On another note, the multivariate uniform distribution exhibits a connection with the Exponential Power Distribution as the shape parameter β tends toward infinity. Finally, the multivariate Student's t -distribution serves as a valuable tool for statistically evaluating differences between means in samples.

The multivariate Laplace distribution have similar property as the multivariate Normal but with heavier tail. In this chapter, we reviewed how various researchers have adopted the multivariate EPD in their studies. It has also presented the application of some EPD to data analysis. One researcher used multivariate EPD in discriminant analysis. Another explanation stated that the multivariate EPD is the possible candidate for describing financial market price changes due to its robustness to kurtosis. Several scholars came to the conclusion

in their investigation that the multivariate EPD belongs to the class of alternative distributions and has top rankings among distributions that are frequently utilised in data modelling. Some researchers in their study proposed new ways of estimating the shape parameter. Some used the method of maximum likelihood and approximated fast estimator. Other researchers performed a comprehensive simulation study aimed at calculating the parameter and evaluating its mean square error. Through this study, they sought to gain a deeper understanding of the parameter's characteristics and the precision of its estimation. Some applications have been poultry feeds, finance, engineering and so on. All these studies show how relevant statistical distributions is very fundamental in data analysis. This underpins major reason of filling the gap in the examining the characteristics and the application of multivariate EPD to broaden the knowledge in this area of study.



CHAPTER THREE

RESEARCH METHODS

Introduction

We will delve into several theorems and methodology used to derive the multivariate EPD in this chapter. Our exploration will extend to the examination of various properties associated with this distribution. Furthermore, we will thoroughly discuss the techniques involved in estimating the parameters of the distribution, emphasizing optimization and model selection methods. Additionally, we will explore the practical application of the multivariate EPD within the context of one of the techniques employed in the analysis of multivariate data.

Research Design and Data Characterization

The methodology employed in this research is a quantitative approach, characterized by its reliance on numerical data. Quantitative research methods, specifically experiments, form an integral part of this approach. This methodology involves the identification of specific variables to be studied and encompasses processes such as hypothesis testing and verification. The quantitative method employed in this study provides a structured framework for rigorous analysis and interpretation of numeric data, facilitating a comprehensive exploration of the research objectives.

Multivariate data emerge when a researcher, aiming to comprehend a social or physical phenomenon, chooses more than one variable or attribute ($p > 1$)

to document. The systematically recorded values for each distinct case, item, individual, or experimental unit are represented using the notation x_{ij} , where i denotes the specific item or trial, and j indicates the variable's position. Here, an item could represent a distinct object, subject, item, individual, or experimental unit. It is essential to note that, in the context of utilizing multivariate EPD, all the variables under consideration must exhibit a continuous nature. A multivariate data structure can be expressed as a rectangular array or as a matrix, $\mathbf{X}$ given as

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & x_{2p} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & x_{ip} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & x_{np} \end{pmatrix} \quad (45)$$

Thus, the distributional structure of $\mathbf{X}$ is said to follow the multivariate EPD. However, the multivariate EPD of Gómez et al. (1998) is named MEPD and the one we will derive is named AMEPD in this study.

Multivariate Dataset

A simulated Bivariate Data from a population distributed as AMEPD for various sample sizes (150, 200, 500, 1000, 10000) were used. Again, three real dataset were used for this study. The first dataset is on car sales, the second dataset is on concrete strength and the third one is on data on some heavy metal present in river pra in Ghana.

The first dataset is a Carsales Dataset (A sample data in IBM SPSS 27). It involves 11 variables (Sales, Price, Type of Vehicle, Engine Size (E. Size), Horsepower, Wheelbase, Width, Length, Curb Weight (C. Weight), Fuel Capacity (F. Capacity) and Fuel Efficiency (F. Efficiency)) with 156 observations. We consider this for small sample size analysis. The vehicle variables of sales and price are both expressed in thousands of US dollars. Engine size of vehicle refers to the volume of fuel and air a vehicle's cylinders can hold, measured in cubic centimeters (cc). Typically stated in liters, such as describing a 1390cc engine as 1.4 liters, engine size is crucial for categorizing cars. Engines below 1.0-litre usually have three or four cylinders and often incorporate turbochargers for enhanced performance. Traditionally, larger engines were linked to more power, but modern turbocharged engines can exceed their predecessors. Engine power, measured in horsepower (bhp), signifies the speed at which force is generated, representing the engine's ability to perform work. Vehicle characteristics such as wheelbase, width, length, weight, and fuel capacity are crucial automotive metrics. Wheelbase is the distance between front and rear wheels, width is the distance between vertical planes parallel to the vehicle's longitudinal median, and length is the horizontal length of a vehicle. These dimensions are measured in centimeters. Vehicle weight is the total mass with standard equipment and consumables, excluding passengers, cargo, or optional equipment, measured in thousands of kilograms. Vehicle fuel capacity refers to the tank's usable capacity, excluding unusable volume, vapor space, or the fuel tank filler neck volume, measured in liters. Fuel efficiency is a fundamental measure

of how well a car converts fuel energy into kinetic energy for travel, expressed as distance per unit volume of fuel, measured in miles per gallon (MPG).

The second dataset is a Concrete Dataset (Yeh, 2006). It involves 9 variables Cement, Blast-Furnace Slag (BFS), Fly Ash (F. Ash), Water, Super-Plasticizer (SP), Coarse Aggregate (CA), Fine Aggregate (Fine A.), Age of Testing (AoT) (*days*), Concrete Compressive Strength (CCS) (*MPa*) with 1133 observations. We consider this for Large sample size analysis. The Cement component measures the density of cement in the concrete, defined as the weight of cement per unit volume and expressed in kilograms per cubic meter (kg/m^3) (Schwer & Kailasanath, 2011). Blast furnace slag, a calcium-silicate-based product obtained during iron ore melting, is swiftly cooled to a glassy state and ground for construction use (Provis, Palomo & Shi, 2015). Its weight per unit volume is also measured in kilograms per cubic meter. Fly ash, the residue from coal combustion, is produced in large quantities and is commonly used in various engineering applications (Provis et al., 2015). Similar to cement and blast furnace slag, its weight per unit volume is expressed in kilograms per cubic meter. Water density in concrete, denoted as water absorption, is expressed in kg/m^3 , representing the weight of water per unit volume. Superplasticizer admixtures, used to reduce water content and enhance workability, have their density measured in kg/m^3 as well. Coarse aggregates, particles larger than 0.19 inch in diameter, make up a significant portion of concrete and are measured in kg/m^3 . Fine aggregates, comprising natural sand or crushed stone particles of 1/4 inch or smaller, are also expressed in kg/m^3 . The age of the concrete test, mea-

sured in days, indicates the time taken before evaluating compressive strength. Concrete compressive strength, crucial for assessing its load-bearing capacity, is measured in megapascals (MPa) or newtons per square millimeter (N/mm^2). This strength varies for different construction purposes.

The third dataset is a Dataset obtain from Ghana Minerals Commission. This data file contains a monthly records of data on the amount of chemicals such as mercury, copper, lead and arsenic in river Pra. One of Ghana's basins with a significant concentration of illicit mining activity is the Pra Basin. In regions where illicit mining is conducted, there is frequently heavy metal contamination of water sources. The Pra Basin's heavy metal pollution is covered by the dataset. It involves seven (7) variables (town, season, year, amount of mercury (mg/l), amount of copper (mg/l), amount of lead (mg/l) and amount of arsenic (mg/l)) with 121 observations. We consider this for small sample size analysis. Mercury, a naturally occurring element found in the earth's crust, poses health risks in small-scale gold mining along Ghana's river Pra. The mercury-gold amalgam used in this process contributes to about 37% of global mercury emissions, adversely affecting the environment, miner health, and local communities. The variable for mercury is measured in milligrams per liter (mg/l). Lead, a toxic metal naturally present in the Earth's crust, becomes a widespread environmental contaminant through activities like mining, smelting, and product usage. Exposure to lead, particularly in children, causes severe health issues, impacting brain and nervous system development. The variable for lead is measured in milligrams per liter (mg/l). Copper, a conductive metallic element, can

contaminate river ecosystems through human activities like mining and industrial discharges. While naturally occurring, anthropogenic contributions elevate copper levels, posing risks to human health. The variable for copper is measured in milligrams per liter (mg/l). Arsenic, a naturally occurring element widely distributed in the environment, is highly toxic. Human exposure through drinking water and contaminated food can lead to severe health issues, including skin lesions and cancer. The variable for arsenic is measured in milligrams per liter (mg/l).

Mathematical Preliminaries

This subsection present the mathematical properties of two important distributions that plays key role in developing elliptical and spherical distributions. They are the Gamma and the Dirichlet distributions.

The Gamma Distribution

If a random variable X follows the PDF $f(x)$, which is defined as follows

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, & \text{if } 0 < x < \infty \\ 0, & \text{otherwise} \end{cases} \quad (46)$$

then it is said to have a Gamma distribution with the shape parameter, α , and the scale or spread parameter, β . Here, the shape parameter is a positive real number, and the exponential distribution replaces the gamma distribution when $\alpha = 1$. Again, the scale parameter is also a positive real number and larger values of β result in a narrower and taller distribution, while smaller values

spread it out. $\Gamma(\alpha)$ represents the gamma function, also known as the gamma integral, and is given by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \alpha > 0, \quad (47)$$

or sometimes expressed as

$$\Gamma(\alpha) = 2 \int_0^{\infty} x^{2\alpha-1} e^{-x^2} dx, \quad \alpha > 0. \quad (48)$$

Here, $\Gamma(\alpha)$ is referred to as the “gamma function of α ”. Additionally, the incomplete gamma function, denoted as $\Gamma(\alpha, t)$ and defined as

$$\Gamma(\alpha, t) = \int_t^{\infty} x^{\alpha-1} e^{-x} dx, \quad \alpha > 0, \quad (49)$$

for all $\alpha > 0$ and $x \geq 0$. It is noteworthy that if $\alpha > 1$, then $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$. The logarithmic derivative of $\Gamma(\alpha)$, denoted as the psi or digamma function $\psi(\alpha)$, is expressed as

$$\psi(\alpha) = \frac{d}{d\alpha} (\ln(\Gamma(\alpha))) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \quad (50)$$

and can be written as

$$\psi(\alpha) = \int_0^{\infty} \left(\frac{e^{-x}}{x} - \frac{e^{-\alpha x}}{1 - e^{-x}} \right) dx. \quad (51)$$

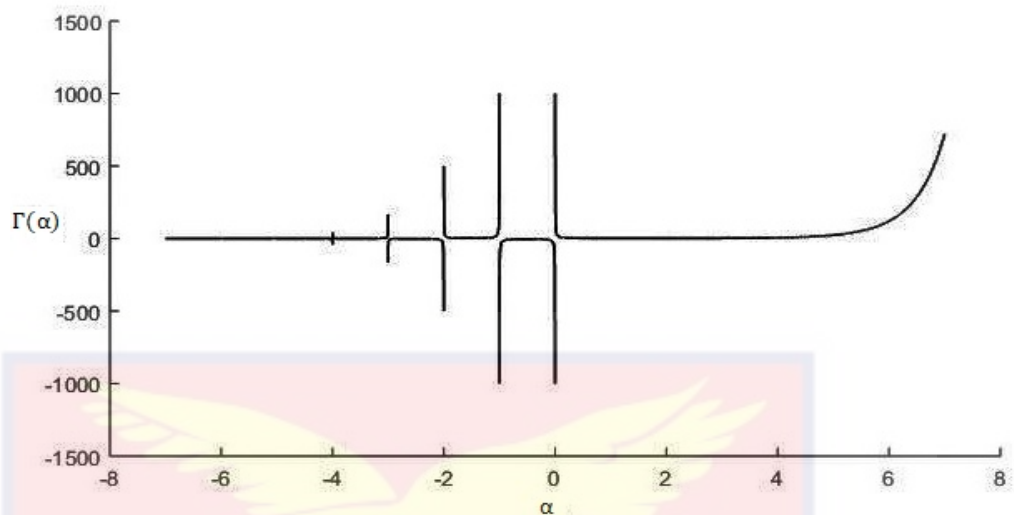


Figure 11: Gamma Function [$\Gamma(\alpha)$] across the entire Complex Plane.

In the construction of the univariate and multivariate instances of the AMEPD, the gamma function is crucial. The gamma function is a special convex extension that can be applied to positive real numbers. It operates as a continuous extension of the factorial function, which is restricted to non-negative integers. Its significance lies in its contribution to comprehending the distinctive characteristics and properties of the AMEPD, as detailed in the subsequent part of this study.

The Dirichlet Distribution

The Dirichlet distribution serves as a fundamental multivariate distribution, representing a generalization of the Beta distribution to multiple dimensions. Applications in Bayesian inference, distribution theory, and statistical modelling are numerous (Dickey & Chen, 1985; Aitchison, 1986). Furthermore, the Dirichlet distribution is especially useful for modelling compositional data and proportions since it can be obtained from the Gamma distribution (Ng,

Tian & Tang, 2011).

To be more specific, the density function of a random variable x that has a gamma distribution and a parameter $\beta > 0$ (written as $x \sim \text{Ga}(\beta)$) is given by

$$(\Gamma(\beta))^{-1} x^{\beta-1} e^{-x}, \quad x > 0. \quad (52)$$

In this case, the variable y assumes a chi-squared distribution with n degrees of freedom when $\beta = n/2$, taking into account $y = 2x$. This is represented as $y \sim \chi_n^2$. To elaborate, suppose that $y_1, \dots, y_m, y_{m+1}$ are independent random variables. They all have the same gamma distribution $\text{Ga}(\alpha_i)$, where $\alpha_i > 0$ for all $i = 1, \dots, m, m+1$, then

$$x_j = \frac{y_j}{\sum_{i=1}^{m+1} y_i}, \quad j = 1, \dots, m. \quad (53)$$

The distribution of $(x_1, \dots, x_m)$ is termed the Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_m, \alpha_{m+1}$. This distribution is denoted as

$$(x_1, \dots, x_m) \sim D_m(\alpha_1, \dots, \alpha_m; \alpha_{m+1})$$

or alternatively as $(x_1, \dots, x_{m+1}) \sim D_{m+1}(\alpha_1, \dots, \alpha_{m+1})$. The density function of $(x_1, \dots, x_m) \sim D_m(\alpha_1, \dots, \alpha_m; \alpha_{m+1})$ is expressed as

$$p_m(x_1, \dots, x_m) = (B_m(\alpha))^{-1} \prod_{i=1}^{m+1} y_i^{\alpha_i-1}, \quad \sum_{i=1}^{m+1} x_i = 1, \quad x_i \geq 0, \quad i = 1, \dots, m+1, \quad (54)$$

where $B_m(\alpha) = (\prod_{i=1}^{m+1} \Gamma(\alpha_i)) / \Gamma(\alpha)$, $\alpha = \sum_{i=1}^{m+1} \alpha_i$ denotes the multivariate beta function associated with parameters the vector, $\alpha = (\alpha_1, \dots, \alpha_{m+1})^\top$. As-

sume $\mathbf{x} = (x_1, \dots, x_m)^\top \sim D_m(\alpha_1, \dots, \alpha_m; \alpha_{m+1})$. Then for every $r_1, \dots, r_m \geq 0$,

we have

$$\mu_{r_1, \dots, r_m} = E \left(\prod_{j=1}^m x_j^{r_j} \right) = \left(\prod_{j=1}^m \alpha_j^{[r_j]} \right) / \alpha^{[r_1 + \dots + r_m]} \quad (55)$$

where $x^{[r]} = x(x+1)\cdots(x+r-1)$ are descending factorials. In particular,

$$E(x_j) = \alpha_j/\alpha \quad \text{and} \quad \text{Var}(x_j) = \frac{\alpha_j(\alpha - \alpha_j)}{\alpha^2(\alpha + 1)} \quad (56)$$

and the covariance between x_i and x_j is

$$-\alpha_i\alpha_j/\alpha^2(\alpha + 1) \quad (57)$$

The fact that the covariance is negative suggests that the components $x_i (i = 1, \dots, m)$ have a negative connection with one another. This distribution is constrained by the condition that the sum of its components equals 1 and each component is non-negative. The Dirichlet distribution holds significance in various applications, particularly in scenarios involving the modeling of proportions or compositional data.

The Dirichlet distribution has some interesting amalgamation properties. The theorems that are pertinent to multivariate EPD are presented here and are expressed in terms of the symmetric form of a Dirichlet distribution. This joint vector $\mathbf{w} = (w_1, \dots, w_k)$ has a Dirichlet distribution $D_k(\alpha_1^*, \dots, \alpha_k^*)$ if the $p = m + 1$ components $x_1, \dots, x_p, \sum_{i=1}^p x_i = 1$, of a $D_p(\alpha_1, \dots, \alpha_p)$ distribution in the symmetric form are lumped into k components $w_i, \sum_{i=1}^k w_i = 1$, where α_i^* is the total of the α_j s for the elements that make up w_i . Suppose we now divide $\mathbf{x} \sim D_p(\alpha)$ into k subvectors, $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$, and α into the corresponding subvectors, $\alpha^{(1)}, \dots, \alpha^{(k)}$. The sums of the components of $\mathbf{x}^{(i)}$ and $\alpha^{(i)}$ should be denoted by w_i and α_i^* , respectively. $\mathbf{y}^{(i)} = \mathbf{x}^{(i)}/w_i$. Then,

- i. The vectors $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}$, and $\mathbf{w}$ are mutually independent.
- ii. $\mathbf{w}$ is distributed as $D_k(\alpha_1^*, \dots, \alpha_k^*)$.

iii. $y^{(i)}$ is distributed as $D_{n_i}(\alpha^{(i)})$, $i = 1, \dots, k$.

If $(y_1, \dots, y_m) \sim D_m(\alpha_1, \dots, \alpha_m; \alpha_{m+1})$ then for any $s < m$, we have

$$(y_1, \dots, y_s) \sim D_s(\alpha_1, \dots, \alpha_s; \alpha_{s+1}^*) \quad (58)$$

where $\alpha_{s+1}^* = \alpha_{s+1} + \dots + \alpha_m + \alpha_{m+1}$. In particular, for any $i (i = 1, \dots, m)$, we have

$$y_i \sim \text{Be}(\alpha_i, \alpha - \alpha_i). \quad (59)$$

The marginal densities of the random variable $\mathbf{x}/|\mathbf{x}|$, where $\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{I}_n)$, can be derived from the density of the Dirichlet distribution. Specifically, for $0 < k < p$, the random variable $(x_1^2/|\mathbf{x}|^2, \dots, x_k^2/|\mathbf{x}|^2)$ follows a Dirichlet distribution with parameters $D_k(1/2, \dots, 1/2; (p-k)/2)$. This implies that the density of $(|x_1|/|\mathbf{x}|, \dots, |x_k|/|\mathbf{x}|)$ is given by

$$\frac{\Gamma(p/2)2^k}{\Gamma((p-k)/2)\pi^{k/2}} \left(1 - \sum_{i=1}^k x_i^2\right)^{(p-k)/2-1}, \quad x_i \geq 0, \quad i = 1, \dots, k, \quad \sum_{i=1}^k x_i^2 < 1 \quad (60)$$

Similarly, the density of $(x_1/|\mathbf{x}|, \dots, x_k/|\mathbf{x}|)$ is expressed as

$$\frac{\Gamma(p/2)}{\Gamma((p-k)/2)\pi^{k/2}} \left(1 - \sum_{i=1}^k x_i^2\right)^{(p-k)/2-1}, \quad \sum_{i=1}^k x_i^2 < 1. \quad (61)$$

These results are particularly valuable for understanding the marginal distributions of components in the given multivariate normal distribution.

According to Kotz et al. (2000), assume that

$$(x_1, \dots, x_m) \sim D_m(\alpha_1, \dots, \alpha_m; \alpha_{m+1})$$

, then

$$x_j^* = x_j / \left(1 - \sum_{i=1}^s y_i\right), \quad j = s+1, \dots, m, \quad (62)$$

$$(x_{s+1}^*, \dots, x_m^* \mid x_1, \dots, x_s) \sim D_{m-s}(\alpha_{s+1}, \dots, \alpha_m; \alpha_{m+1}). \quad (63)$$

Now, we have

$$B_m(\alpha) = \int_{D(x_1, \dots, x_m)} \prod_{i=1}^m x_i^{\alpha_i-1} \left(1 - \sum_{i=1}^m x_i\right)^{\alpha_{m+1}-1} dx_1 \cdots dx_m, \quad (64)$$

where

$$D(x_1, \dots, x_m) = \left\{ (x_1, \dots, x_m) : x_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m x_i < 1 \right\}. \quad (65)$$

The integral Equation (64) is attributed to P.G.L. Dirichlet (1805-59). Joseph Liouville (1809-82). It is pivotal in the developments of Liouville distributions. For any measurable function $f(\cdot)$ that is nonnegative, consider the following

$$I_m(f \mid \alpha) = \int_{R_+^m} f\left(\sum_{i=1}^m x_i\right) \left(\prod_{i=1}^m x_i^{\alpha_i-1} dx_i\right) \quad (66)$$

where $\alpha = (\alpha_1, \dots, \alpha_m)^\top$ and $R_+^m = \{(x_1, \dots, x_m) : x_i \geq 0, i = 1, \dots, m\}$. It is true that

$$I_m(f \mid \alpha) = B_{m-1}(\alpha) \cdot I_1\left(f \mid \sum_{i=1}^m \alpha_i\right) \quad (67)$$

provided that either integral exists. For $\alpha_i > 0, i = 1, \dots, m$, we have

$$\int_{R^m} f\left(\sum_{i=1}^m x_i^2\right) \left(\prod_{i=1}^m |x_i|^{2\alpha_i-1} dx_i\right) = I_m(f \mid \alpha). \quad (68)$$

Setting $\alpha_i = 1/2, i = 1, \dots, m$ in Equation (68) we immediately obtain

$$\begin{aligned} \int_R f\left(\sum_{i=1}^m x_i^2\right) dx_1 \cdots dx_m &= \frac{\pi^{m/2}}{\Gamma(m/2)} \int_0^\infty y^{m/2-1} f(y) dy \\ &= \pi^{m/2} (\Gamma(m/2))^{-1} I_1(f \mid m/2). \end{aligned} \quad (69)$$

where $I_1(f \mid m/2)$ is given by Equation (66). The formula in Equation (69) has numerous applications in the derivation of AMEPD which will be demonstrated in later Chapter.

Definition of Spherical and Elliptical Symmetric Distribution

The collection of spherically symmetric and elliptically contoured distributions can be conceptualized in various ways, as discussed in various sources (Kelker, 1970; Dempster, 1969; Eaton, 1986; Cambanis, Huang & Simons, 1981; Muirhead, 1982; Fang & Zhang, 1990; Provost & Cheong, 2002). A distribution is considered spherically contoured when it is multivariate, and its variables are independent, sharing a common variance. Additionally, the covariance matrix for this distribution takes the form of a diagonal matrix, with identical elements along the diagonal. Mathematically, in p -dimensional space, a distribution is considered spherically symmetric when its PDF is solely dependent on the distance from the central point. If we denote a random vector in this space as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)$ and define $R = \sqrt{\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_p}$ as the Euclidean distance from the origin, the PDF of a spherically symmetric distribution can be expressed mathematically as follows

$$f(\mathbf{x}) = g(R). \quad (70)$$

Here, $g(R)$ is a function that exclusively relies on the scalar distance R . This mathematical expression signifies that the probability of finding the random vector $\mathbf{X}$ at a particular distance R from the origin is determined solely by the function $g(R)$. It implies that, regardless of the specific direction in which $\mathbf{X}$ points, the probability remains the same as long as the distance R is constant.

Definition 1

A $p \times 1$ random vector $\mathbf{X}$ is said to have a spherically symmetric distribution,

$S_p(\phi)$, when it can be represented stochastically as follows:

$$\mathbf{x} \stackrel{d}{=} r\mathbf{U}^{(p)} \quad (71)$$

Here, the notation $\stackrel{d}{=}$ indicates that $\mathbf{X}$ has the same distribution as $r\mathbf{U}^{(p)}$, where r is a one-dimensional random variable representing the radius, and $\mathbf{U}^{(p)}$ is the uniform distribution on the unit sphere. Moreover, r and $\mathbf{U}^{(p)}$ are independent, satisfying the conditions:

$$r = \|\mathbf{x}\| \geq 0, \quad U^{(p)} \stackrel{d}{=} \frac{\mathbf{x}}{\|\mathbf{x}\|}. \quad (72)$$

These conditions in the definition ensure that the radius r corresponds to the Euclidean norm of the vector $\mathbf{x}$, and $U^{(p)}$ is the normalized vector pointing in the same direction as $\mathbf{x}$.

Definition 2

If a $p \times 1$ random vector $\mathbf{X}$ conforms to the stochastic representation, it is said to have an elliptically contoured distribution, which is represented as $EC_p(\mu, \Sigma, \phi)$, with parameters μ of dimension $(p \times 1)$ and covariance matrix Σ of size $(p \times p)$ of the form

$$\mathbf{x} \stackrel{d}{=} \mu + \mathbf{A}^\top \mathbf{y}, \quad \mathbf{y} \sim \mathbf{S}_p(\phi) \quad (73)$$

In this case, $\mathbf{A}$ is a $k \times p$ matrix with the property that $\mathbf{A}^\top \mathbf{A} = \Sigma$, where $\text{rank}(\Sigma) = k$. The distribution has the following features if $\mathbf{X} \sim EC_p(\mu, \Sigma, \phi)$ with $\text{rank}(\Sigma) = p$:

- i. The $\mathbf{X}$'s c.f., represented as $\phi(\mathbf{t}) = E(e^{i\mathbf{t}^\top \mathbf{X}})$, is as

$$\phi(\mathbf{t}) = e^{i\mathbf{t}^\top \mu} \phi(\mathbf{t}^\top \Sigma) \quad (74)$$

for some scalar function ϕ .

- ii. An alternate stochastic representation of the form exists for $\mathbf{X}$ and it is given as

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + r\mathbf{A}^\top \mathbf{U}^{(p)} \quad (75)$$

as $r \perp\!\!\!\perp \mathbf{U}^{(p)}$, for $r \geq 0$.

Properties of Spherical and Elliptical Symmetric Distribution

Renowned researchers like Cambanis et al. (1981) and Kelker (1970) have done a great deal of research on spherical and elliptical distributions. Berkane and Bentler (1986), Dickey and Chen (1985, 1987), and Anderson and Fang (1982) are among the other researchers. These scholars have delved into various aspects, contributing significantly to our understanding of the properties and characteristics associated with these distributions. Their work has laid the foundation for exploring critical theorems and properties relevant to the AMEPD and EPDs in general. This part will discuss these theorems and properties.

Theorem 1. (Fang & Zhang, 1990) If one of the following equivalent requirements is met by the characteristic function $\psi(\mathbf{t})$ of a p -vector $\mathbf{x}$, then the p -vector $\mathbf{x}$ exhibits a spherical distribution:

- i. For any $\Gamma \in \mathcal{O}(p)$, $\psi(\Gamma'\mathbf{t}) = \psi(\mathbf{t})$;
- ii. For a scalar variable, there is a function $\phi(\cdot)$ by which $\psi(\mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t})$

□

In simpler terms, the spherical distribution property of $\mathbf{x}$ is confirmed if its characteristic function remains unchanged under any orthogonal transformation (condition i), or alternatively, if there is a scalar function $\phi(\cdot)$ such that the characteristic function only depends on the squared Euclidean norm of $\mathbf{t}$, (condition ii).

Conversely, condition (i) indicates that $\psi(\mathbf{t})$ remains unchanged with respect to the group $\mathcal{O}(p)$ where the maximal invariant is $\mathbf{t}^\top \mathbf{t}$. Therefore, $\psi(\mathbf{t})$ must solely depend on $\mathbf{t}^\top \mathbf{t}$, leading to condition (ii). Moving forward, we will express $\mathbf{x} \sim S_p(\phi)$ to signify that $\mathbf{x}$ possesses a CF in the form $\phi(\mathbf{t}^\top \mathbf{t})$, where $\phi(\cdot)$ is a function of a scalar variable termed the characteristic generator of the spherical distribution. This concept will be frequently employed in the subsequent discussion. This clarification demonstrates that the CF of a spherical distribution is generated by a scalar function. Consequently, the entire set of characteristic generators for an $p \times 1$ random vector is denoted as

$$\Phi_p = \{ \phi(\cdot) : \phi(\mathbf{t}_1^2 + \cdots + \mathbf{t}_p^2) \text{ is an } p\text{-dimensional CF} \}.$$

It is evident that $\Phi_1 \supset \Phi_2 \supset \Phi_3 \supset \cdots$. Let

$$\Phi_\infty = \bigcap_{n=1}^{\infty} \Phi_p$$

Theorem 2. (Schoenberg, 1938) A function $\phi \in \Phi_p$ if and only if it takes the form

$$\phi(x) = \int_0^\infty \Omega_p(xr^2) dF(r) \quad (76)$$

Here, $F(\cdot)$ represents a CDF defined over the interval $[0, \infty)$, and $\Omega_p(\mathbf{y}^\top \mathbf{y})$ is

defined as

$$\Omega_p(\mathbf{y}^\top \mathbf{y}) = \int_{S: \mathbf{x}^\top \mathbf{x} = 1} e^{i\mathbf{y}^\top \mathbf{x}} dS / S_p \quad (77)$$

Here, S_p denotes the surface area of the unit sphere in $\Re^p$, and $\Omega_p(t^\top t)$ represents the characteristic function of $\mathbf{u}^{(p)}$. If $\mathbf{x} = (x_1, \dots, x_p)^\top$ is a random vector of dimension p , it is identified as a spherical random vector if and only if, for any $\mathbf{a} \in \Re^p$, the inner product of $\mathbf{a}$ and $\mathbf{x}$ satisfies

$$\mathbf{a}^\top \mathbf{x} = |\mathbf{a}|x_1. \quad (78)$$

Furthermore, if $\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} \end{bmatrix} \sim S_p(\phi)$, where $\mathbf{x}^{(1)}$ is an $m \times 1$ random vector, it follows that $\mathbf{x}^{(1)} \sim S_m(\phi)$. $\square$

This implies that if $\mathbf{x} \sim S_p(\phi)$, then each marginal distribution of $\mathbf{x}$ exhibits a spherical nature, and all the marginal CF share a common generator. The following theorem is due to Anderson and Fang (1982).

Theorem 3. (Fang & Zhang, 1990) If $\mathbf{x} \sim S_p(\phi)$, partition $\mathbf{x}$ into m parts $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)})^\top$ each with $p_1, \dots, p_m$ components respectively. Then

$$\begin{bmatrix} \mathbf{x}^{(1)} \\ \vdots \\ \mathbf{x}^{(m)} \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} rd_1 \mathbf{u}_1 \\ \vdots \\ rd_m \mathbf{u}_m \end{bmatrix} \quad (79)$$

where $d_i \geq 0, i = 1, \dots, m; (d_1^2, \dots, d_m^2) \sim D_m(p_1/2, \dots, p_m/2); \mathbf{u}_1, \dots, \mathbf{u}_m$ and $(d_1^2, \dots, d_m^2)$ are independent and $\mathbf{u}_j \stackrel{d}{=} \mathbf{u}^{(p_j)}, j = 1, \dots, m$. $\square$

Theorem 4. (Fang & Zhang, 1990)

The mean vector and covariance matrix of $\mathbf{u}^{(p)}$ are expressed as

$$E(\mathbf{u}^{(p)}) = \mathbf{0}, \quad \text{Cov}(\mathbf{u}^{(p)}) = \frac{1}{p} \mathbf{I}_p, \quad (80)$$

respectively. If $\mathbf{x} = r\mathbf{u}^{(p)} \sim Sp(\phi)$, and $E(r^2) < \infty$, then the expected value and covariance matrix of $\mathbf{x}$ are given by

$$E(\mathbf{x}) = \mathbf{0}, \quad \text{Cov}(\mathbf{x}) = \frac{E(r^2)}{p} \mathbf{I}_p. \quad (81)$$

□

Theorem 5. (Dickey & Chen, 1985)

The moments of the random vector $\mathbf{x} = r\mathbf{u}^{(p)}$, provided they exist, can be expressed through a one-dimensional integral. For even integers $2s_i \geq 0$, $i = 1, \dots, p$, the expression is given by

$$E\left(\prod_{i=1}^p X_i^{2s_i}\right) = E(r^{2s}) \pi^{-p/2} \frac{\Gamma(p/2)}{\Gamma(p/2 + s)} \prod_{i=1}^p \Gamma\left(\frac{1}{2} + s_i\right), \quad (82)$$

where $s = s_1 + s_2 + \dots + s_p$. It is noteworthy that a random vector $\mathbf{x} \sim Sp(\phi)$ may not necessarily possess a density function in general. □

However, based on Definition 1 and Theorem 1, if a density function for $\mathbf{x}$ does exist, it must be of the form $g(\mathbf{x}^\top \mathbf{x})$ for some nonnegative function $g(\cdot)$ of a scalar variable. In such cases, the integral is expressed as

$$\int_{\mathbb{R}^p} g(\mathbf{x}^\top \mathbf{x}) d\mathbf{x} = \frac{\pi^{p/2}}{\Gamma(p/2)} \int_0^\infty y^{p/2-1} g(y) dy = 1. \quad (83)$$

Therefore, a nonnegative function $g(\cdot)$ can be utilized to define a density $cg(\mathbf{x}^\top \mathbf{x})$ for some spherical distribution if and only if the condition

$$\int_0^\infty y^{p/2-1} g(y) dy < \infty \quad (84)$$

holds. In this context, we adopt the notation $\mathbf{x} \sim Sp_p(g)$ as an alternative to $\mathbf{x} \sim Sp(\phi)$, where $g(\cdot)$ is acknowledged as a density generator for the spherical distribution. It is commonly referred to simply as the PDF generator.

Theorem 6. (Kelker, 1970) If the random vector $\mathbf{x}$ follows the distribution $\mathbf{x} \stackrel{d}{=} r\mathbf{u}^{(n)} \sim S_p(\phi)$, the existence of a density generator $g(\cdot)$ for $\mathbf{x}$ is conditioned on r possessing a density $f(\cdot)$. Their relationship is defined by the expression

$$f(r) = \frac{2\pi^{p/2}}{\Gamma(p/2)} r^{p-1} g(r^2). \quad (85)$$

This relationship implies that if $\mathbf{x}$ follows the distribution $\mathbf{x} \sim S_p(g)$, then the density $h(u)$ of $u = r^2$ can be represented as

$$h(u) = \frac{\pi^{p/2}}{\Gamma(p/2)} u^{p/2-1} g(u). \quad (86)$$

□

Theorem 7. (Fang & Zhang, 1990)

Assume that $\mathbf{x} \stackrel{d}{=} r\mathbf{u}^{(p)} \sim S_p^+(\phi)$. Then all the marginal distributions of $\mathbf{x}$ possess densities. In particular, the marginal density of $(x_1, \dots, x_k)$ ($1 \leq k < p$) is

$$\frac{\Gamma(p/2)}{\Gamma((p-k)/2)\pi^{k/2}} \int_{(\sum_{i=1}^k x_i^2)^{1/2}}^{\infty} r^{-(p-2)} \left(r^2 - \sum_{i=1}^k x_i^2 \right)^{(p-k)/2-1} dF(r), \quad (87)$$

where $F(r)$ is the CDF of r and denotes the marginal distribution of a specific spherical component in the elliptically contoured distribution having parameters, μ , Σ and ϕ where $\text{rank}(\Sigma) = k$. Additionally, the quadratic form

$$Q(\mathbf{x}) = (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \stackrel{d}{=} r^2, \quad (88)$$

where Σ^{-1} is a generalized inverse of Σ . □

Theorem 8. (Fang & Zhang, 1990)

Assuming that $\mathbf{x} \sim EC_p(\mu, \Sigma, \phi)$ and $E(r^2) < \infty$, the following expres-

sions hold,

$$E(\mathbf{x}) = \mu, \quad \text{Cov}(\mathbf{x}) = \frac{E(r^2)}{\text{rank}(\Sigma)} \Sigma = -2\phi^\top(0)\Sigma, \quad (89)$$

$$\Gamma_2(\mathbf{x}) = E(\mathbf{x}\mathbf{x}^\top) = \mu\mu^\top - 2\phi^\top(0)\Sigma$$

Here, $\phi^\top(0)$ represents the derivative of ϕ at the origin, and the following definitions apply

$$\Gamma_1(\mathbf{x}) = i^{-1} \frac{\partial \phi(\mathbf{t})}{\partial \mathbf{t}} \Big|_{\mathbf{t}=\mathbf{0}} \quad (90)$$

$$\Gamma_2(\mathbf{x}) = i^{-2} \frac{\partial^2 \phi(\mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^\top} \Big|_{\mathbf{t}=\mathbf{0}}$$

$$\mathbf{x} \stackrel{d}{=} \mu + r\mathbf{A}^\top \mathbf{u}^{(k)}. \quad (91)$$

Additionally, the representation

$$E(\mathbf{x}) = \mu + E(r)\mathbf{A}^\top E(\mathbf{u}^{(k)}) = \mu, \quad (92)$$

leads to the following results

$$\begin{aligned} \text{Cov}(\mathbf{x}) &= \text{Cov}(r\mathbf{A}^\top \mathbf{u}^{(k)}) = E(r^2) \mathbf{A}^\top \text{Cov}(\mathbf{u}^{(k)}) \mathbf{A} \\ &= E(r^2) \frac{1}{k} \mathbf{A}^\top \mathbf{I}_k \mathbf{A} = \frac{1}{k} E(r^2) \cdot \Sigma, \end{aligned} \quad (93)$$

□

For higher-order moments, cumulants, and Edgeworth expansion, Traat (1989) provides detailed expressions. Ct  roux and Ducharme (1989) explore a measure of multivariate correlation and its asymptotic distribution. Muirhead (1982) introduces the kurtosis parameter γ of an elliptical distribution $EC_p(\mu, \Sigma, \phi)$ as

$$\gamma = \frac{\phi''(0)}{(\phi'(0))^2} - 1 \quad (94)$$

where $\phi''(0)$ and $\phi'(0)$ represent the second and first derivatives of ϕ evaluated at zero. Berkane and Bentler (1986) define the general moment parameters as

$$\gamma(m) = \frac{\phi^{(m)}(0)}{(\phi'(0))^m} - 1, \quad (95)$$

and demonstrate that for $\mathbf{X} \sim EC_p(\mathbf{0}, \Sigma, \phi)$, the moment $\mu_{i_1, i_2, \dots, i_{2m}} = E(X_{i_1} X_{i_2} \cdots X_{i_{2m}})$

can be computed by

$$\mu_{i_1, i_2, \dots, i_{2m}} = (\gamma(m) + 1) \Sigma \mu_{i_1, i_2} \mu_{i_3, i_4} \cdots \mu_{i_{2m-1}, i_{2m}}, \quad (96)$$

where the summation is over all the $(2m)! / (2^m m!)$ possible groupings of the subscripts into pairs. For instance,

$$\mu_{ijkl} = (\gamma + 1) (\mu_{ij} \mu_{kl} + \mu_{ik} \mu_{jl} + \mu_{il} \mu_{jk}). \quad (97)$$

Mardia (1970) introduces the multivariate kurtosis coefficient for a p -dimensional random vector $\mathbf{x}$ with mean $E(\mathbf{x}) = \mu$ and covariance matrix $\text{Cov}(\mathbf{x}) = \Sigma$ as

$$\gamma_1 = E \left((\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right)^2 - p(p+2). \quad (98)$$

According to Berkane and Bentler (1987), when the random vector $\mathbf{x}$ follows an elliptically contoured distribution with parameters μ , Σ , and ϕ , and the covariance matrix of $\mathbf{x}$ is Σ , the coefficient of skewness (γ_1) is equal to the coefficient of skewness in the marginal distribution (γ). It is crucial to remember that, generally speaking, a variable $\mathbf{x}$ that comes after $EC_p(\mu, \Sigma, \phi)$ might not have a density function. Let's look at two scenarios: i) $\mathbf{x}$ has a density; ii) $P(\mathbf{x} = \mu) = 0$ and $\Sigma > 0$. It is evident that $\text{rank}(\Sigma) = p$ is a necessary requirement for $\mathbf{x} \sim EC(\mu, \Sigma, \phi)$ to have a density. In this instance, the stochastically

represented random vector $\mathbf{X}$ can be expressed as

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{y} \quad (99)$$

given $\mathbf{A}$ is a nonsingular matrix that satisfies $\mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$ and $\mathbf{y} \sim S_p(\phi)$. With $g(\mathbf{y}^\top \mathbf{y})$ denoting the density generator, $g(\cdot)$, we may characterise the density of $\mathbf{y}$. Using the equation $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{y}$, the density function of $\mathbf{X}$ may be found by applying

$$|\boldsymbol{\Sigma}|^{-\frac{1}{2}} g\left((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right). \quad (100)$$

In the event that $\mathbf{x}$ lacks a density, $P(r = 0) = 0$ and $\boldsymbol{\Sigma} > 0$, applying the same transformation $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{y}$, results in $P(\mathbf{y} = 0) = 0$ and $\mathbf{y}$ retains all marginal densities, just as $\mathbf{x}$ does. Consequently, by Equation (87) the marginal density of $\mathbf{x}_{(k)} = (X_1, \dots, X_k) - \boldsymbol{\mu}_{(k)}$, $1 \leq k < p$, where $\boldsymbol{\mu}_{(k)} = (\mu_1, \dots, \mu_k)^\top$, is given by

$$\begin{aligned} & \frac{\Gamma\left(\frac{p}{2}\right) |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}}{\Gamma\left(\frac{p-k}{2}\right) \pi^{\frac{k}{2}}} \int_0^\infty \left(\mathbf{x}_{(k)}^\top \boldsymbol{\Sigma}_k^{-1} \mathbf{x}_{(k)}\right) r^{-(p-2)} \\ & \times \left(r^2 - \mathbf{x}_{(k)}^\top \boldsymbol{\Sigma}_k^{-1} \mathbf{x}_{(k)}\right)^{\frac{p-k}{2}-1} dF(r) \end{aligned} \quad (101)$$

where $\boldsymbol{\Sigma}_k$ is the first principal minor of $\boldsymbol{\Sigma}$ of dimension k .

Any function $g(\cdot)$ that satisfies Equation (84) serves as a definition for the density in Equation (100) of an elliptically symmetric distribution. This distribution includes a normalizing constant C_n given by

$$C_p = \frac{\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}} \int_0^\infty r^{p-1} g(r^2) dr} \quad (102)$$

Uniform Distribution in the Unit Sphere

If we consider $\mathbf{u}^{(p)}$ to be uniformly distributed on the unit sphere surface

in $\Re^p$, then the density of u_1 is given by

$$\frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)\sqrt{\pi}}(1-u_1^2)^{\frac{p-1}{2}-1}, \quad u_1^2 < 1, \quad (103)$$

Similarly, in $\Re^p$, the uniform distribution inside a unit sphere is likewise considered to be spherical. Let $\mathbf{x}$ be a random vector in $\Re^p$ that follows the uniform distribution inside the unit sphere. This distribution's density function can be written as

$$P_{\mathbf{x}}(\mathbf{x}) = \begin{cases} n\Gamma\left(\frac{1}{2}p\right) / \left(2\pi^{p/2}\right), & \text{if } \sum_{i=1}^p x_i \leq 1 \\ 0 & , \text{otherwise} \end{cases} \quad (104)$$

Certainly, the representation of $\mathbf{x}$ as $\mathbf{x} \stackrel{d}{=} r\mathbf{u}^{(p)}$ implies that the PDF of r can be expressed as

$$f(r) = \begin{cases} pr^{p-1} & \text{if } 0 \leq r \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (105)$$

Consequently, $r \sim Be(p, 1)$. The marginal density of $(x_1, \dots, x_k)$, $1 \leq k < p$ may be computed with ease using $f(r)$ and the density of $(u_1, \dots, u_k)$

$$\frac{\Gamma\left(\frac{p+2}{2}\right)}{\Gamma\left(\frac{p-k+2}{2}\right)\pi^{\frac{k}{2}}} \left(1 - \sum_{i=1}^k x_i^2\right)^{\frac{p-k}{2}}, \quad \left(\sum_{i=1}^k x_i < 1\right). \quad (106)$$

Through direct calculations, the CF of $\mathbf{x}$ is $\phi(\mathbf{t}^\top \mathbf{t})$, where

$$\phi(u) = \frac{2}{B\left(\frac{1}{2}, \frac{p+1}{2}\right)} \int_0^\infty \cos(ux) (1-x^2)^{\frac{p-1}{2}} dx. \quad (107)$$

For any integers $m_1, \dots, m_p$ with $m = \sum_{i=1}^p m_i$, the mixed moments of $\mathbf{x}$ are

$$E\left(\prod_{i=1}^p x_i^{m_i}\right) = \frac{B(p+m, 1)}{B(p, 1)} \prod_{i=1}^p E(u_i^{m_i}). \quad (108)$$

In the theory of spherical distributions and elliptically contoured distributions, the 'uniform basis' $\mathbf{u}^{(p)}$ is crucial, especially for deriving multivariate EPD. Let $\Omega_p(|t|^2), t \in \Re^p$, be the CF of $u^{(p)}$. Then, the CF $\Omega_p(|\mathbf{t}|^2)$ can be expressed in equivalent forms as follows:

$$\begin{aligned}\Omega_p(|\mathbf{t}|^2) &= \frac{1}{B\left(\frac{p-1}{2}, \frac{1}{2}\right)} \int_0^\pi \exp(i|\mathbf{t}| \cos \theta) \sin^{p-2} \theta d\theta \\ &= \frac{\Gamma\left(\frac{p}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{p-1}{2}\right)} \sum_k \frac{(-1)^k |\mathbf{t}|^{2k} \Gamma\left(\frac{p-1}{2}\right) \Gamma((2k+1)/2)}{(2k)! \Gamma\left(\frac{p+2k}{2}\right)} \\ &= {}_0F_1\left(\frac{p}{2}; \frac{1}{4}|\mathbf{t}|^2\right)\end{aligned}$$

where $\Omega_p(|\mathbf{t}|^2)$ is derived from the integral

$$\int_0^\pi \cos^k \theta \cdot \sin^{p-2} \theta d\theta = \begin{cases} B\left(\frac{p-1}{2}, \frac{k+1}{2}\right), & \text{if } k \text{ is even} \\ 0, & \text{if } k \text{ is odd} \end{cases} \quad (109)$$

based on the relations between beta and gamma functions. Therefore,

$$\frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right) \sqrt{\pi} p} \int_0^\pi \exp(z \cos \theta) \sin^{p-2} \theta d\theta = \sum_{k=0}^\infty \frac{\left(\frac{1}{4}z^2\right)^k}{\left(\frac{1}{2}\right)^{[k]} k!} = {}_0F_1\left(\frac{1}{2}p; \frac{1}{4}z^2\right). \quad (110)$$

Here, ${}_0F_1(\cdot)$ is the generalized hypergeometric function defined as

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{k=0}^\infty \frac{a_1^{[k]} \cdots a_p^{[k]} z^k}{b_1^{[k]} \cdots b_q^{[k]} k!} \quad (111)$$

where $a^{[k]}$ represents ascending factorials, i.e., $a^{[k]} = a(a+1) \cdots (a+k-1)$.

The density of u_1 is given by

$$\frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right) \sqrt{\pi}} (1-u_1^2)^{\left(\frac{p-1}{2}\right)-1}, \quad u_1^2 < 1. \quad (112)$$

Moreover, the function $\Omega_p(|\mathbf{t}|^2)$ with $\mathbf{t} \in \Re^{p+1}$ does not represent an $(p+1)$ -dimensional c.f.. Additionally, for any integers $m_1, \dots, m_p$, with $m = \sum_{i=1}^p m_i$,

the mixed moments of $\mathbf{u}^{(p)}$ can be expressed as follows

$$E\left(\prod_{i=1}^p u_i^{m_i}\right) = \begin{cases} \frac{1}{(\frac{p}{2})^{[l]}} \prod_{i=1}^p \frac{(2l_i)!}{4^{l_i}(l_i)!}, & \text{if } m_i = 2l_i \text{ are even, } i = 1, \dots, p, m = 2l; \\ 0, & \text{if at least one of the } m_i \text{ is odd.} \end{cases} \quad (113)$$

Here, as mentioned earlier, $x^{[l]} = x(x+1)\cdots(x+l-1)$.

It is well known that the marginal distributions of $\mathbf{u}^{(p)}$ are not always uniform, even though all the marginal distributions of a multinormal distribution are normal. In particular, $(u_1, \dots, u_k)$, where $1 \leq k \leq n$, has the joint density given by

$$\frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p-k}{2}) \pi^{\frac{k}{2}}} \left(1 - \sum_{i=1}^k u_i^2\right)^{\frac{p-k}{2}-1}, \text{ if } \sum_{i=1}^k u_i^2 < 1, \quad (114)$$

valid when $\sum_{i=1}^k u_i^2 < 1$. This function is a dependence on $\sum_{i=1}^k u_i^2$. Hence, $\mathbf{u}^{(1)} = (u_1, \dots, u_k)^\top$ exhibits a spherical distribution with a stochastic decomposition

$$\mathbf{u}^{(1)} = r_1 \mathbf{u}^{(k)}, \quad (115)$$

where r_1 follows a density of

$$\left[B\left(\frac{k}{2}, \frac{p-k}{2}\right)\right]^{-1} 2r^{k-1} (1-r^2)^{\frac{p-k}{2}-1}, \quad 0 < r < 1. \quad (116)$$

Thus, $r_1^2 \sim B\left(\frac{k}{2}, \frac{p-k}{2}\right)$, a beta distribution with parameters $\frac{k}{2}$ and $\frac{p-k}{2}$. Similarly, $(u_1^2, \dots, u_k^2) \sim D_{k+1}\left(\frac{1}{2}, \dots, \frac{1}{2}, \frac{p-k}{2}\right)$.

For the conditional distribution of the multiuniform distribution, if we partition $\mathbf{u}^{(p)}$ into

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}_{(1)} \\ \mathbf{u}_{(2)} \end{pmatrix}, \quad \text{where } \mathbf{u}_{(1)} : m \times 1, \quad (117)$$

then the conditional distribution $\mathbf{u}(1)$ given $\mathbf{u}(2) = \mathbf{a}$ is

$$(\mathbf{u}_{(1)} \mid \mathbf{u}_{(2)} = \mathbf{a}) \sim \mathcal{S}_m(\Phi_{\|\mathbf{a}\|^2}) \quad (118)$$

with a stochastic representation

$$(\mathbf{u}_{(1)} \mid \mathbf{u}_{(2)} = \mathbf{a}) \stackrel{d}{=} r(\|\mathbf{a}\|^2) \mathbf{u}^{(m)}, \quad (119)$$

where, for each $\|\mathbf{a}\|^2 \geq 0$, $r(\|\mathbf{a}\|^2)$ and $\mathbf{u}^{(m)}$ are independent. Additionally,

$$r(\|\mathbf{a}\|^2) \stackrel{d}{=} \left((1 - \|\mathbf{a}\|^2)^{1/2} \mid \mathbf{u}_{(2)} = \mathbf{a} \right) \stackrel{d}{=} (1 - \|\mathbf{a}\|^2)^{1/2}, \quad (120)$$

which implies that the conditional distribution of $\mathbf{u}(1)$ given $\mathbf{u}(2) = \mathbf{a}$ is a uniform distribution on the sphere with center 0 and radius $(1 - \|\mathbf{a}\|^2)^{1/2}$.

The multiuniform distribution on a sphere is of significant importance in the realm of directional distributions, particularly in multivariate EPD. Watson (1984) dedicates an entire chapter in his book to this distribution, providing a wealth of additional intriguing results. Poincaré (1912) deserves credit for proving that the first k coordinates of $\mathbf{u}^{(p)}$ are approximately independent normals. Diaconis and Freedman (1986) extend this assertion to hold even when $k = O(p)$ in the limit, implying a finite form of de Finetti's theorem which is an orthogonally invariant probability distribution in a p -dimensional space. This elucidates why spherical symmetry remains a crucial property of independently and identically distributed normal distributions. To address the needs of projection pursuit, Cheng (1987) considers the following problem. Let $\mathbf{x}_1, \dots, \mathbf{x}_p$ be a sequence of p -dimensional Independent and Identically Distributed (IID) random vectors, where $\mathbf{x}_1 \sim F$, and $\mathbf{u} \stackrel{d}{=} \mathbf{u}^{(p)}$ is independent of

$(\mathbf{x}_1, \dots, \mathbf{x}_p)$. Cheng investigates the empirical distribution $\hat{F}p^u(x)$ of a sequence $\mathbf{u}^\top \mathbf{x}_1, \dots, \mathbf{u}^\top \mathbf{x}_p$. Under certain regularity conditions, Cheng derives the limit distribution of $\hat{F}p^u(x)$ as both n and p approach infinity. Jiang (1988) delves into the limiting distribution of $\sum_{i=1}^k u_i z_i$, where $z_1, \dots, z_k$ are points on the unit circle, and $\mathbf{u} = (u_1, \dots, u_k)^\top \sim D_k(\mathbf{b})$, a Dirichlet distribution. Under certain conditions, the limiting distribution of $\sum_{i=1}^k u_i z_i$ is a uniform distribution on a unit circle.

Distribution of the Quadratic (R^2) Form of Elliptical and Spherical Distribution

The PDF of $Q(x) = R^2$ (R is the generating variate of $\mathbf{x}$ as per Theorem 5), is given by

$$\frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} u^{\frac{p}{2}-1} g(u) \quad (121)$$

where $g(u)$ is defined by Equation (15), the density of $Q(x) = R^2$ is thus given by

$$\left(\frac{s r^{\frac{2N+p-2}{2s}}}{\Gamma\left(\frac{2N+p-2}{2s}\right)} \right) u^{N+\frac{p}{2}-2} e^{-ru^s}, u > 0. \quad (122)$$

For $s = 1$, the density in Equation (122) becomes

$$\left(\frac{r^{N+\frac{p}{2}-1}}{\Gamma(N+\frac{p}{2}-1)} \right) u^{N+\frac{p}{2}-2} e^{-ru}, u > 0, \quad (123)$$

which is a gamma distribution writing as $\text{Ga}(N + \frac{p}{2} - 1, r)$. As the parameters of $\text{Ga}(a, b) > 0$, it is crucial to note that the conditions, $2N - p > 2$ and $r > 0$ in Equation (15) are necessary. Since the integral of Equation (122) with respect

to u over $[0, \infty)$ equals one, we have for any $t > 0$

$$\begin{aligned} E(R^{2t}) &= C_p \int_0^\infty u^t u^{N+\frac{p}{2}-2} e^{-ru^s} du \\ &= \frac{C_p}{C_{p+2t}} \\ &= r^{-\frac{t}{s}} \frac{\Gamma\left(\frac{2N+p+2t-2}{2s}\right)}{\Gamma\left(\frac{2N+p-2}{2s}\right)}. \end{aligned} \quad (124)$$

Specifically, we have

$$E(R^2) = r^{-\frac{1}{s}} \frac{\Gamma\left(\frac{2N+p}{2s}\right)}{\Gamma\left(\frac{2N+p-2}{2s}\right)} \quad (125)$$

and

$$E(R^2) = \frac{N + \frac{p}{2} - 1}{r} \quad (126)$$

if $s = 1$. Now utilizing formula in Equation (82) it is easy to obtain any mixed moments of $\mathbf{x}$. In particular,

$$E(\mathbf{x}) = \mu \quad \text{and} \quad D(\mathbf{x}) = r^{-\frac{1}{s}} \frac{\Gamma\left(\frac{2N+p}{2s}\right)}{\Gamma\left(\frac{2N+p-2}{2s}\right)} \frac{\Sigma}{\text{rank}\Sigma}. \quad (127)$$

Some Properties and the Derivative of the Multivariate Normal Distribution

The Multivariate Normal Distribution is a flexible set of distributions derived from applying linear transformations to independent normal random variables, described by

$$\mathbf{y} = \mathbf{Ax} + \mathbf{b}, \quad (128)$$

Utilizing the properties of elliptical and spherical distributions, with $N = s = 1$, and $r = \frac{1}{2}$, the function g becomes $g(u) = C_p \exp(-u/2)$, where $C_p = (2\pi)^{-p/2}$ indicates that $\mathbf{x} \sim N_p(\mu, \Sigma)$.

The standardization of independent normal random variables for $\mathbf{x} = (X_1, X_2, \dots, X_p)$ with $E(X_i) = \mu_i$ and $\text{var}(X_i) = \sigma_i^2$ for $i = 1, 2, \dots, p$, will yield $Z_i = \frac{X_i - \mu_i}{\sigma_i}$. Thus, $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)$ is a vector leading to $X_i = \sigma_i Z_i + \mu_i$. Consequently, we can express $\mathbf{x} = \mathbf{C}\mathbf{z} + \boldsymbol{\mu}$, where $\mathbf{C} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)^\top$. Hence, any linear transformation in Equation (128) can be expressed as

$$\mathbf{y} = \mathbf{B}\mathbf{z} + \mathbf{d}. \quad (129)$$

In this case, $\mathbf{D} = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$ and $\mathbf{B} = \mathbf{A}\mathbf{C}$. Standard normal random variables are transformed linearly using Equation (129). It follows that a random vector $\mathbf{X} = (X_1, X_2, \dots, X_p) \sim N_p(\boldsymbol{\mu}, \Sigma)$, given a vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)^\top$ and a linear transformation $\mathbf{x}$. Consequently, $\mathbf{x}$, any multivariate normal vector, can be represented as

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu} \quad (130)$$

where $\mathbf{A}$ is an $n \times p$ matrix of constants and $\mathbf{z}$ is a $p \times 1$ vector of IID standard normal random variables. The mean vector and variance-covariance matrix of $\mathbf{x}$ are $E(\mathbf{X}) = \boldsymbol{\mu}$ and $D(\mathbf{X}) = \mathbf{A}\mathbf{A}^\top$, respectively.

The following are some observations about the Multivariate Normal Distribution

1. The rank of matrix $\mathbf{A}$ in Equation (130) cannot be more than n .
2. The components of $\mathbf{X}$ are assumed to be linearly dependent if the rank of $\mathbf{A}$ is less than n . This results in a singular variance-covariance matrix $\Sigma_{\mathbf{X}}$. $\mathbf{X}$ in these cases is a unique multivariate Normal distribution.

3. In the case where $\mathbf{A}$ has rank equal to n , the matrix $\Sigma_{\mathbf{X}}$ is non-singular and the component variables of $\mathbf{X}$ are linearly independent.
4. In the event that $n > p$ and $\mathbf{A}$'s rank (which is definitely less than p and cannot exceed $\min(n, p)$), then $\mathbf{X}$ has a singular distribution. Consequently, for each $n > p$, the distribution of $\mathbf{X}$ is unique.
5. Assume that $\mathbf{A}$ has a rank of r , where $r < n$, meaning that $\mathbf{X}$ is unique. In this instance, the subvector $\mathbf{X}(r)$ of $\mathbf{X}$ includes r components and a non-singular distribution. According to this, every element of $\mathbf{X}|\mathbf{X}(r)$ that is not in $\mathbf{X}(r)$ is dependent on every element of $\mathbf{X}(r)$ in a linear fashion. This means that $\mathbf{X}(r)$ fully determines the distribution of the whole vector $\mathbf{X}$. Thus, a non-singular distribution associated with any unique multivariate Normal distribution can be used to specify it.
6. The multivariate Normal distribution, represented as $N(\mu, \Sigma)_p$, is defined in a p -dimensional space by a variance-covariance matrix Σ and a mean vector μ . The following lists a few characteristics of the multivariate Normal random variable:
 - (a) If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A}$ is an $n \times p$ matrix of constants, $\mathbf{b}$ is a constant vector, and $\mathbf{X} \sim N(\mu, \Sigma)_p$, then
 - a. $\mathbf{Y} \sim N(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^\top)$;
 - b. If the distribution of $\mathbf{X}$ is non-singular and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, then the distribution of $\mathbf{Y}$ is also non-singular if and only if $\text{rank}(\mathbf{A}) = n$.

(b) If the variance-covariance matrix $\Sigma_{\mathbf{X}}$ is diagonal, then the component variables of $\mathbf{X}$ are mutually independent.

7. Consider the following situation: Σ is a diagonal matrix provided by $\Sigma = \text{diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp})$. Define $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$ and $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)^\top$. The MGF of $\mathbf{Z}$ is expressed as $M_{\mathbf{Z}}(\mathbf{t}) = \exp(\frac{1}{2}\mathbf{t}^\top \mathbf{t})$, where $\mathbf{t} = (t_1, t_2, \dots, t_n)^\top$, and let $\mathbf{s} = (s_1, s_2, \dots, s_p)^\top$. The MGF of $\mathbf{X}$ is then given by

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{s}) &= E \left[\exp(\mathbf{A}\mathbf{Z} + \boldsymbol{\mu})^\top \mathbf{s} \right] \\ &= e^{\boldsymbol{\mu}^\top \mathbf{s}} M_{\mathbf{AZ}}(\mathbf{s}) \\ &= e^{\boldsymbol{\mu}^\top \mathbf{s}} M_{\mathbf{Z}}(\mathbf{A}^\top \mathbf{s}) \end{aligned} \quad (131)$$

Simplifying further with substitutions yields

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{s}) &= e^{\boldsymbol{\mu}^\top \mathbf{s}} \exp \left(\frac{1}{2} \mathbf{s}^\top \mathbf{A} \mathbf{A}^\top \mathbf{s} \right) \\ &= \exp \left(\boldsymbol{\mu}^\top \mathbf{s} + \frac{1}{2} \mathbf{s}^\top \Sigma_{\mathbf{X}} \mathbf{s} \right) \end{aligned} \quad (132)$$

which, when substituting for $\boldsymbol{\mu}$ and $\Sigma_{\mathbf{X}}$ leads to

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{s}) &= \exp \left(\sum_{i=1}^p \mu_i s_i + \frac{1}{2} \sum_{i=1}^p \sigma_{ii} s_i^2 \right) \\ &= \prod_{i=1}^p \exp \left(\mu_i s_i + \frac{1}{2} \sigma_{ii} s_i^2 \right) \end{aligned} \quad (133)$$

Now, considering $M_{X_i}(s) = \exp(\mu_i s_i + \frac{1}{2} \sigma_{ii} s_i^2)$ implying that $X_i \sim N(\mu_i, \sigma_{ii})$

. Therefore, $M_{\mathbf{X}}(\mathbf{s}) = \prod_{i=1}^p M_{X_i}(s_i)$. Consequently, since the MGF of $\mathbf{X}$ is the product of the mgfs of the components X_i , it indicates that X_i (for $(i = 1, 2, \dots, p)$) are independent.

Theorem 9. $\mathbf{X}$ has a non-singular Multivariate Normal distribution with a variance-covariance matrix Σ and a mean vector μ , there is a non-singular matrix $\mathbf{S}$ such that $\mathbf{S}\mathbf{S}^\top = \Sigma$ and

$$\mathbf{Z} = \mathbf{S}^{-1}(\mathbf{X} - \mu) \quad (134)$$

is a vector of independent standard Normal random variables, also of dimension p . □

Now, given $\mathbf{Z} = \mathbf{S}^{-1}(\mathbf{X} - \mu)$, we have $E(\mathbf{Z}) = \mathbf{S}^{-1}(E(\mathbf{X}) - \mu) = \mathbf{0}$. The variance-covariance matrix is given by

$$\begin{aligned} D(\mathbf{Z}) &= \mathbf{S}^{-1}D(\mathbf{X} - \mu)(\mathbf{S}^{-1})^\top \\ &= \mathbf{S}^{-1}\mathbf{S}\mathbf{S}^\top(\mathbf{S}^\top)^{-1} \\ &= \mathbf{I}_p \end{aligned}$$

Thus, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_p)$, representing a vector of p independent standard Normal random variables. Since every multivariate Normal distribution can be determined by some non-singular distribution, it is sufficient to derive the density function of non-singular distributions.

Now, following Theorem 9, let's define a non-singular transformation $\mathbf{Z} \rightarrow \mathbf{X}$ given by $\mathbf{X} = \mathbf{A}\mathbf{Z} + \mu$, with Equation (134) as the inverse. The Jacobian of the inverse transformation is given by

$$J_{\mathbf{Z}}(\mathbf{X}) = \det(\mathbf{S}^{-1}) = \frac{1}{\det(\mathbf{S})} = \frac{1}{|\Sigma^{\frac{1}{2}}|} \quad (135)$$

The density of the vector $\mathbf{Z}$ is expressed as

$$f(\mathbf{z}) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \exp\left(-\frac{1}{2}\mathbf{z}^\top \mathbf{z}\right) \quad (136)$$

Thus, the density of $\mathbf{X}$ is given by

$$\begin{aligned}\phi(\mathbf{x}) &= f(\mathbf{z}) \times |J_{\mathbf{Z}}(\mathbf{x})| \\ &= f(\mathbf{S}^{-1}(\mathbf{X} - \boldsymbol{\mu})) \times |J_{\mathbf{Z}}(\mathbf{x})| \\ &= \frac{1}{|\Sigma|^{\frac{1}{2}}} \left(\frac{1}{\sqrt{2\pi}} \right)^p \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\top} (\mathbf{S}^{\top})^{-1} \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (137)\end{aligned}$$

Further simplification gives

$$\phi(\mathbf{x}) = \frac{1}{|\Sigma|^{\frac{1}{2}}} \left(\frac{1}{\sqrt{2\pi}} \right)^p \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\top} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad \forall \mathbf{x} \in \mathbb{R}^p \quad (138)$$

Thus, the density function of $\mathbf{X}$ depends on the variance-covariance matrix Σ through the quadratic form

$$Q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^{\top} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (139)$$

The quadratic form in Equation (139) is sufficient to identify the distribution of $\mathbf{X}$ and is equivalent to

$$Q(\mathbf{x}) = \sum_{i=1}^p \sum_{j=1}^p a_{ij} (x_i - \mu_i) (x_j - \mu_j) \quad (140)$$

where $\mathbf{A} = (a_{ij}) = \Sigma^{-1}$, a symmetric matrix. Further expansion gives

$$Q(\mathbf{x}) = \sum_{j=1}^p a_{jj} (x_j - \mu_j)^2 + 2 \sum_{i \neq j}^p a_{ij} (x_i - \mu_i) (x_j - \mu_j) \quad (141)$$

The derivation of the Multivariate Normal Distribution provides valuable insights that will serve as a guide in deriving the AMEPD. Understanding the mathematical principles and properties inherent in the Multivariate Normal Distribution lays the foundation for extending these concepts to the development of the AMEPD. Key components of the Multivariate Normal Distribution, such as the transformation of variables, the role of non-singular matrices, and the

quadratic form involving the variance-covariance matrix, will likely find parallels in the derivation of the AMEPD. By leveraging the knowledge gained from the Multivariate Normal Distribution, one can anticipate similar mathematical structures and techniques being applied in the derivation process of the AMEPD.

Properties of MEPD

The univariate EPD, defined by Gomez et al. (1998), is expressed as

$$f(x; \mu, \phi, \beta) = \frac{1}{\phi \Gamma\left(1 + \frac{1}{2\beta}\right) 2^{1 + \frac{1}{2\beta}}} \exp\left[-\frac{1}{2} \left|\frac{x - \mu}{\phi}\right|^{2\beta}\right] \quad (142)$$

Let μ be a $p \times 1$ vector in $\Re^p$, $\mathbf{X}$ a $p \times 1$ random vector, and Σ a $p \times p$ non-negative definite matrix. If the CF $\phi_{x-\mu}$ of $x - \mu$ is a function of the quadratic form $t^\top \Sigma t$ expressed as $\phi_{x-\mu} = \phi(t^\top \Sigma t)$ for some function ψ , then $\mathbf{X}$ is said to have an elliptically contoured distribution, denoted by $EC_p(\mu, \Sigma, g)$. Equation (2) provides the multivariate EPD that was suggested by Gomez et al. (1998). In this case, the scale parameter is $\mu \in \Re^p$, the shape parameter affecting the distribution's tails is $\beta \in (0, \infty)$, and the scale parameter is Σ , a $(p \times p)$ positive definite matrix.

When $p = 1$, the standard one-dimensional distribution is generalised by the multivariate p -dimensional exponential power density function. Nevertheless, this does not necessarily mean that $\mathbf{X}$ follows the distribution $PE_p(0, I_p, \beta)$, with the exception of the case where $\beta = 1$, which corresponds to the normal distribution, even if the components of a random vector $X = (X_1, \dots, X_p)^\top$ satisfy $X_i \sim PE_1(0, 1, \beta)$ for $i = 1, \dots, p$. The distribution changes to the double

exponential distribution at $\beta = 0.5$. The distribution converges to a multivariate uniform distribution as β gets closer to infinity.

Gomez et al. (1998)'s Proposition:

1. If a random vector $\mathbf{X}$ follows a particular distribution $PE_p(\mu, \Sigma, \beta)$, it can be equivalently represented as $\mu + R\mathbf{A}^\top \mathbf{u}^{(p)}$, where $\mathbf{u}^{(p)}$, a random scaled rotation of a uniformly distributed vector in $\mathfrak{R}^p$, R , non-negative random variable independent of $\mathbf{u}^{(p)}$, and $\mathbf{A}$ is a rotation matrix with p order satisfying $\mathbf{A}^\top \mathbf{A} = \Sigma$.
2. Conversely, if $X = (X_1, \dots, X_p)^\top$ has the same distribution as $\mu + R\mathbf{A}^\top \mathbf{u}^{(p)}$ with $\mathbf{u}^{(p)}$ and R defined as

$$h_R(\tau) = \frac{p}{\Gamma\left(1 + \frac{p}{2\beta}\right) 2^{\frac{p}{2\beta}}} r^{p-1} \exp\left\{-\frac{1}{2} r^{2\beta} I_{(0,\infty)}\right\}(\tau), \quad (143)$$

$\beta \in (0, \infty)$, and $\mathbf{A}$ is a $(p \times p)$ non-singular matrix, then $\mathbf{X}$ is distributed as $PE_p(\mu, \Sigma, \beta)$ with $\Sigma = \mathbf{A}^\top \mathbf{A}$.

3. The CF is given by

$$\begin{aligned} \varphi(x) &= \frac{p}{\Gamma\left(1 + \frac{p}{2\beta}\right) 2^{\frac{p}{2\beta}}} \exp^{it^\top \mu} \int_0^\infty \Psi_p\left(\tau \sqrt{t^\top \Sigma} t\right) \tau^{(p-1)} \\ &\quad \times \exp\left\{-\frac{1}{2} \tau^{2\beta}\right\} d\tau \end{aligned} \quad (144)$$

where $\Psi_1(x) = \cos(x)$ and $\Psi_n(x) = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{1}{2}} \Gamma(\frac{p-1}{2})} \int_0^\pi e^{ix \cos \theta} \sin^{p-2} \theta d\theta$ for every $p > 1$.

4. The location parameter is

$$E(X) = \mu \quad (145)$$

5. The variance is

$$\text{Var}(X) = \frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{p+2}{2\beta}\right)}{n \Gamma\left(\frac{p}{2\beta}\right)} \Sigma \quad (146)$$

6. The multidimensional asymmetry is

$$\gamma_1(X) = 0 \quad (147)$$

7. The kurtosis coefficient is

$$\gamma_2(X) = \frac{p^2 \Gamma\left(\frac{p}{2\beta}\right) \Gamma\left(\frac{p+4}{2\beta}\right)}{\Gamma^2\left(\frac{p+2}{2\beta}\right)} - p(p+2) \quad (148)$$

Note that $\gamma_1(X)$ and $\gamma_2(X)$ are defined as expectations involving moments of $\mathbf{X}$ and $\mathbf{Y}$ where $\mathbf{Y}$ is a random vector independent and equally distributed as $\mathbf{X}$. The term $p(p+2)$ corresponds to the value of the second moment when $\mathbf{X}$ is distributed as a multivariate normal.

Analogously to the unidimensional scenario, μ acts as a location parameter, Σ serves as a scale parameter, and when $\beta = 1$, the covariance matrix equals the scale parameter. The parameter β relates to kurtosis, indicating the distribution's departure from normality.

Method of Deriving Spherical and Elliptical Symmetric Distribution

Since the early 20th century, statisticians have been actively involved in extending univariate distributions to their multivariate counterparts. Various approaches have been proposed to tackle this challenge. In his survey paper "Multivariate distributions at a cross-road" (1975) and the joint paper with Goodman

in 1981, Samuel Kotz systematically classified these methods for extending multivariate distributions based on criteria like the type of dependence, mathematical form analogy, model, and characterizations. While these problems are not completely solved, significant progress has been achieved in this field. Over the last few decades, researchers in statistical distributions have successfully introduced numerous univariate distributions, shifting the focus away from the normal distribution. Encouraged by these achievements, investigators have recently directed their attention and methodologies toward analyzing multivariate distributions beyond the normal one. To derive a spherically or elliptically contoured statistical distribution, you can use the concept of multivariate normal distribution as a starting point. For definiteness assume that we wish to extend the normal distribution $N(\mu, \sigma^2)$ to the multivariate case. The following are some possible avenues.

By means of Density

This process of deriving spherical and elliptical symmetric distributions involves establishing the density functions for these distributions. For a spherical symmetric distribution, the PDF is radially symmetric, meaning it depends only on the distance from the origin (radius). The density function $f(r)$ is typically expressed in terms of the radial coordinate r , where r represents the distance from the origin in a multivariate space. If the distribution is defined in Euclidean space, a common approach is to transform the coordinates to spherical coordinates, where the radial coordinate r plays a central role. Express the

probability density function in terms of the radial coordinate r , which represents the distance from the origin. For elliptical symmetric distributions, the density function is invariant under rotations. In other words, the distribution looks the same when viewed from any direction. The density function often involves a radial component, similar to the spherical case, and an orientation matrix that accounts for the shape of the ellipsoid. Transforming to elliptical coordinates involves expressing the density function in terms of coordinates that account for the shape and orientation of the ellipsoid. Ensure that the density function is properly normalized, i.e., the integral over the entire space equals 1.

For normal distribution case, since the density of $x \sim N(\mu, \sigma^2)$ is

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \times \exp\left(-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right) \quad (149)$$

we can guess that the density of multivariate normal distributions should be of the form

$$C|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad (150)$$

where $\mathbf{x}, \mu \in \mathbb{R}^p, \Sigma : p \times p, \Sigma > 0$ and C is a normalizing constant.

By means of the Characteristic Function

Deriving spherical and elliptical symmetric distributions through the characteristic function involves working with the Fourier transform of the PDF. The characteristic function provides a powerful tool for understanding the statistical properties of a distribution. For a spherical symmetric distribution, the characteristic function, denoted by $\phi(t)$, is defined as the expected value of the

complex exponential of the random variable $\phi(t) = E[e^{itX}]$. The characteristic function is expected to exhibit rotational symmetry. This means that the characteristic function should only depend on the radial coordinate. Express the characteristic function in polar coordinates, separating the radial component and angular components. The radial component of the characteristic function, often denoted by $\psi(t)$, represents the contribution from the radial coordinate. Inverse Fourier transform the radial component to obtain the probability density function in terms of the radial distance. For elliptical symmetric distributions, the characteristic function is expected to depend on the Mahalanobis distance $(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$, which incorporates the covariance structure. Express the Mahalanobis distance in terms of the characteristic function, focusing on the dependence on the covariance matrix. Separate the radial component and the part of the characteristic function related to the covariance structure. Inverse Fourier transform the radial component to obtain the probability density function in terms of the Mahalanobis distance.

Using normal distribution case, it is well known that the CF of $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ is

$$\exp\left\{it\mu - \frac{1}{2}\sigma^2 t^2\right\} = \exp\left\{it\mu - \frac{1}{2}t(\sigma^2)t\right\}. \quad (151)$$

Thus an appropriate form of the CF of the multivariate normal distribution can be

$$\exp\left\{it^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}\right\} \quad (152)$$

where $\mathbf{t}, \boldsymbol{\mu} \in \Re^p$ and $\boldsymbol{\Sigma} > 0$.

By means of Linear Combination

Deriving a spherical or elliptical symmetric distribution through a linear combination involves expressing the distribution as a combination of simpler distributions or basis functions. The idea is to find coefficients for these basis functions in a way that the resulting combination satisfies the desired spherical or elliptical symmetry. For instance, With a standard multivariate normal distribution with mean vector μ and covariance matrix Σ . Apply a linear transformation to the variables to introduce correlations and different variances. The linear transformation is typically done using a matrix A , where $A\Sigma A^T = R$, where R is a positive definite matrix. The resulting distribution will have elliptically contoured PDF. The elliptically contoured distribution allows for more flexible dependence structures compared to the spherically contoured distribution. It can capture correlations between variables and varying variances, resulting in ellipsoidal contours in the multivariate space. As an example, consider the $n \times 1$ random vector $\mathbf{X}$. We term $\mathbf{X}$ to have a multivariate normal distribution if, for every $\mathbf{a} \in \Re^p$, $\mathbf{x}^T \mathbf{a}$, is normal. Naturally, only a few specific distributions will be able to benefit from this last strategy.

By means of Stochastic Decomposition

Stochastic decomposition is a method often used in probability theory to break down a complex stochastic process into simpler, more manageable components. When applied to derive spherical or elliptical symmetric distributions, it involves expressing the desired distribution as a sum or combination of sim-

pler stochastic processes or random variables.

If $x \sim N(\mu, \sigma^2)$, it can be expressed as

$$x = \mu + \sigma y, \quad (153)$$

$y \sim N(0, 1)$ in this case. Therefore, we can also obtain the general multivariate normal distribution if we can extend the standard normal distribution to the multivariate case. Assume that $y_1 \sim N(0, 1)$ and that $\mathbf{y} = (y_1, \dots, y_m)^\top$ has IID components. A natural multivariate extension of the normal distribution is the distribution of $\mathbf{y}$. Let

$$\mathbf{x} = \mu + \mathbf{A}\mathbf{y}, \quad (154)$$

where $\mathbf{A} : p \times m$ and $\mathbf{x}, \mu \in \mathbb{R}^p$. We write $\mathbf{x} \sim N_p(\mu, \Sigma)$, where $\Sigma = \mathbf{A}\mathbf{A}^\top$, assuming that $\mathbf{x}$ is distributed according to a multivariate normal distribution.

Numerous signs point to the convenience and potential of the stochastic decomposition approach. In the initial stages of research, investigators endeavored to broaden univariate distributions to encompass a multivariate framework through the method known as the 'one by one' approach. Kotz et al. (2000) provided an exhaustive exploration of multivariate continuous distributions, offering detailed discussions on numerous specific distributions within this context. We can observe that, in the standard form without parameters like μ, Σ , many of these distributions exhibit a certain form of symmetry.

Approaches to Parameter Estimation

In the historical development of parameter estimation theory, two fundamental themes have emerged: (i) the least squares (LS) method, initially

introduced by Gauss and Legendre, grounded in finite sample considerations with a minimum distance interpretation, and (ii) the maximum likelihood (ML) method, introduced by Fisher, primarily justified through asymptotic considerations, including minimum size asymptotic confidence intervals, ideas of which trace back to Laplace. For this study, the maximum likelihood approach was adopted due to its recognized reliability and efficiency. The maximum likelihood estimation (MLE) approach aims to identify the multivariate EPD that renders the observed data most probable.

We refer to $\hat{\theta}$ as the maximum likelihood estimator (MLE) of θ given an estimator for θ denoted as $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$, where $\hat{\theta}$ maximises the function $L(\theta)$. There are theoretical grounds for the MLE. The true value of θ can be represented by θ_0 . Under certain assumptions, the true model at θ_0 can be effectively distinguished from models at $\theta \neq \theta_0$ by the asymptotic maximum of $L(\theta)$. These assumptions are

A0) There are differences in the cumulative distribution functions (cdfs), that

$$\text{is, } \theta \neq \theta' \Rightarrow F(x_i; \theta) \neq F(x_i; \theta').$$

A1) For all θ , there is a common backing among the probability density functions (PDFs).

A2) An interior point in Ω is θ_0 .

While the second assumption suggests that the support of X_i is constant for a range of θ values, the first assumption states that the parameter uniquely identifies the pdf. Presuming that $E_{\theta_0} [f(X_i; \theta) / f(X_i; \theta_0)]$ and that θ_0 is the true

parameter Assuming that (AO) and (A1) converge in distribution, the asymptotic result is as follows:

$$\lim_{n \rightarrow \infty} P_{s_0} [L(\theta_0, \mathbf{x}) > L(\theta, \mathbf{x})] = 1, \text{ for all } \theta \neq \theta_0. \quad (155)$$

The inequality $L(\theta_0, \mathbf{x}) > L(\theta, \mathbf{x})$ can be expressed as follows using logarithms:

$$\frac{1}{n} \sum_{i=1}^n \log \left[\frac{f(X_i; \theta)}{f(X_i; \theta_0)} \right] < 0. \quad (156)$$

Due to the independence and identically distributed nature of the summations, along with the strictly convex nature of the function $\phi(x) = -\log(x)$, Jensen's inequality and the Law of Large Numbers suggest that when θ_0 is the true parameter

$$\frac{1}{n} \sum_{i=1}^n \log \left[\frac{f(X_i; \theta)}{f(X_i; \theta_0)} \right] \xrightarrow{P} E_{\theta_0} \left[\log \frac{f(X_1; \theta)}{f(X_1; \theta_0)} \right] < \log E_{\theta_0} \left[\frac{f(X_1; \theta)}{f(X_1; \theta_0)} \right]. \quad (157)$$

Additionally,

$$E_{\theta_0} \left[\frac{f(X_1; \theta)}{f(X_1; \theta_0)} \right] = \int_{\mathcal{R}^p} \frac{f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta_0)} f(\mathbf{x}; \theta_0) d\mathbf{x} = 1. \quad (158)$$

The likelihood function is therefore maximised at the true value θ_0 asymptotically. Thus, it makes sense to take into account the value of θ that maximises the likelihood when estimating θ_0 . As a maximum likelihood estimator of θ , we define $\hat{\theta} = \hat{\theta}(\mathbf{x})$ if

$$\hat{\theta} = \text{Argmax}_{\theta} L(\theta; \mathbf{x}). \quad (159)$$

In this instance, the notation $L(\theta; \mathbf{x})$ reaches its maximum value at $\hat{\theta}$, according to Argmax. The log-likelihood is frequently taken and its critical

value is ascertained in order to find the MLE. More specifically, the MLE solves the equation letting $l(\theta) = \log L(\theta)$

$$\frac{\partial l(\theta)}{\partial \theta} = 0. \quad (160)$$

These likelihood-based methods, along with confidence intervals and hypothesis testing, are a part of a formal statistical inference theory. These processes are asymptotically optimal under some assumptions.

Let us consider a random variable X that depends on an unknown parameter θ within a set Ω . This is an example of a univariate random variable, with a PDF $f(x; \theta)$. The continuous scenario is the main focus of this investigation. Let us assume that we have a random sample $X_1, \dots, X_n$ from X . The random variables $X_1, \dots, X_n$ are independent and identically distributed, and they share a common probability distribution $f(x; \theta)$, where $\theta \in \text{'Omega'}$. The value of θ is still unknown. Our methods of inference are predicated on the likelihood function, which is described as

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Omega, \quad (161)$$

where $\mathbf{x} = (x_1, \dots, x_n)^\top$. The logarithm of the PDF function is denoted as

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta), \quad \theta \in \Omega. \quad (162)$$

It is worth noting that using $l(\theta)$ incurs no loss of information since the logarithm is a one-to-one function.

Consider a random vector $\mathbf{X}$ in $\Re^p$ with a PDF $f(x; \theta)$ in the context of the multivariate case. For the $p \times p$ information matrix, let $\mathbf{I}(\theta) = [I_{jk}]$ be the

representation. θ_0 represents the true parameter θ . Other presumptions consist of (A3): All third partial derivatives of $f(x; \theta)$ exist for all $\theta \in \Omega_0$, and there exists an open subset $\Omega_0 \subset \Omega$ such that $\theta_0 \in \Omega_0$. (A4) The following formulas are valid (essentially, expectation and differentiation can be interchanged):

$$E_{\theta} \left[\frac{\partial}{\partial \theta_j} \log f(x; \theta) \right] = 0, \text{ for } j = 1, \dots, p \quad (163)$$

$$I_{jk}(\theta) = E_{\theta} \left[-\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(x; \theta) \right], \text{ for } j, k = 1, \dots, p.$$

(A5) For all $\theta \in \Omega_0$, $\mathbf{I}(\theta)$ is positive definite. (R6) There exist functions $M_{jkl}(x)$ such that

$$\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \log f(x; \theta) \right| \leq M_{jkl}(x), \text{ for all } \theta \in \Omega_0, \quad (164)$$

and

$$E_{\theta_0} [M_{jkl}] < \infty, \text{ for all } j, k, l \in 1, \dots, p. \quad (165)$$

In the case where θ is a vector of p parameters, let $X_1, \dots, X_p$ be IID with a common PDF $f(x; \theta)$, where $\theta \in \Omega \subset \mathbb{R}^p$. Similar to the univariate case, the likelihood function and its logarithm are given by

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad (166)$$

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta),$$

for $\theta \in \Omega$.

In the context of MLE, there is no assurance that the maximum likelihood estimator necessarily exists, and even if it does, uniqueness is not guaranteed. When the MLE either does not exist or is not unique, numerical methods become essential for obtaining a solution. Various numerical techniques can be

employed for this purpose, including Newton's Method, Fisher Scoring, Secant Method, and Fixed-Point Iteration. For situations where the maximum likelihood estimator faces challenges in existence or uniqueness, researchers often turn to numerical optimization methods to find a solution. In this particular study, Newton's Method was chosen as the numerical approach to tackle these complexities. Newton's Method is an iterative optimization algorithm that aims to find successively better approximations to the root of a real-valued function. By iteratively refining the estimates, Newton's Method is adept at converging to solutions, making it a valuable tool in scenarios where traditional analytical methods might fall short. The following present details.

The Newton-Raphson Optimization Method for Estimating the Shape Parameter

The favorable properties of maximum likelihood estimators are well recognized, particularly under asymptotic conditions. However, when confronted with the challenge of estimating the shape parameter q within the framework of the multivariate EPD, complexities may arise due to computational intricacies associated with numerical methods. The determination of the estimate for the shape parameter q remains an open issue. Various methodologies have been proposed for the univariate case, with valuable contributions from Nadarajah (2005) as well as Vasudeva and Kumari (2013), Purczyński and Bednarz-Okrzyńska (2014), and Nkansah et al. (2018). In the existing literature, these approaches primarily rely on the Maximum Likelihood Estimation (MLE) method

or involve the computation of kurtosis indices.

In the multivariate case, previous strategies for estimation have embraced diverse techniques. These include fixed-point algorithms and other optimization methods rooted in geodesic convexity, along with extensions based on regularization (Zhang et al., 2013). Furthermore, estimation strategies have utilized Newton-Raphson recursions (Pascal et al., 2013). Certain investigations on this distribution have concentrated their efforts on a specific case where q is between 0 and 1 (Gómez-Sánchez-Manzano et al., 2008; Bombrun et al., 2012; Pascal et al., 2013). In this study, the selected approach for estimation aligns with the Newton-Raphson method, as recommended by Pascal et al. (2013). This method was chosen as it proves to be a robust and effective technique for tackling the challenges associated with estimating the shape parameter q in the specified multivariate context.

According to Givens and Hoeting (2012), Newton's method, also called the Newton-Raphson iteration, is a very quick method of determining roots. This approach is especially popular for applications that are univariate. The method uses the linear Taylor series expansion to approximate $g'(x^*)$ at iteration t , assuming that g' is continuously differentiable and $g''(x^*) \neq 0$:

$$0 = g'(x^*) \approx g'(x^{(t)}) + (x^* - x^{(t)}) g''(x^{(t)}). \quad (167)$$

It is reasonable to estimate the root of g' by the root of the tangent line since g' is approximated by its tangent line at $x^{(t)}$ (Givens & Hoeting, 2012).

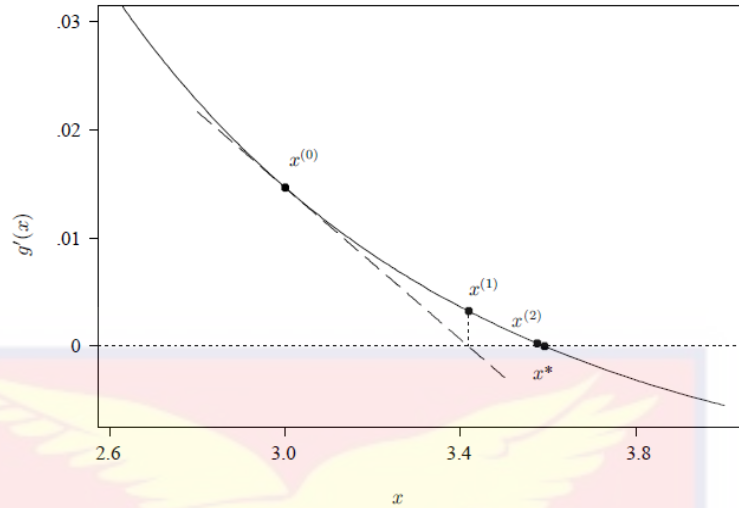


Figure 12: First Illustration of Newton's Method

Consequently, by figuring out x^* in the given expression, we arrive at

$$x^* = x^{(t)} - \frac{g'(x^{(t)})}{g''(x^{(t)})} = x^{(t)} + h^{(t)}. \quad (168)$$

The approximation to x^* in this equation is based on a refinement $h^{(t)}$ and the present estimate $x^{(t)}$. Applying this strategy repeatedly yields Newton's method's iterative equation

$$x^{(t+1)} = x^{(t)} + h^{(t)}. \quad (169)$$

In this case, $h^{(t)} = -\frac{g'(x^{(t)})}{g''(x^{(t)})}$. Finding the maximum of the quadratic Taylor series approximation to $g(x^*)$ analytically justifies the same update, which is $g(x^{(i)}) + (x^4 - x^{(i)})g'(x^{(i)}) + \frac{(x^4 - x^{(i)})^2 g''(x^{(i)})}{2}$ (Givens & Hoeting, 2012).

When $F^r(\theta) = 0$ is solved by $\hat{\theta}$, and the optimisation of g corresponds to an MLE problem, then the updating equation for Newton's method becomes

$$\theta^{(t+1)} = \theta^{(t)} - \frac{P(\theta^{(i)})}{l^v(\theta^{(0)})}. \quad (170)$$

The first iterations of Newton's method applied to the function $g(x) = \frac{\log x}{1+x}$ are shown graphically in Figure 12 (Givens & Hoeting, 2012). The root $x^{(j)}$

becomes the next approximation for the true root x^4 after g' is approximated by its tangent line at $x^{(0)}$ in the first step. The next steps result in $x^{(2)}$, which is already quite near to x^* . For this problem, the Newton increment is defined as

$$h^{(i)} = \frac{(x^{(0)} + 1) \left(1 + 1/x^{(0)} - \log x^{(0)}\right)}{3 + 4/x^{(2)} + 1/(x^{(i)})^2 - 2 \log x^{(2)}}. \quad (171)$$

Starting from $x^{(0)} = 3.0$, Newton's method rapidly converges to $x^{(0)} \approx 3.59112$. For comparison, the bisection method in Figure (12)) requires until iteration 19 to accurately determine the first five decimal places of x^4 . The convergence of Newton's method relies on the function g and the initial value. In Figure (13), there is an example where the method diverges from its starting value (Givens & Hoeting, 2012). To understand the conditions for convergence, a thorough analysis of errors at successive steps is essential. When initiated from $x^{(0)}$, Newton's method diverges by taking steps that progressively deviate from the true root, x^* . Given $g''(x^*) \neq 0$ and the continuity of g'' , there exists a

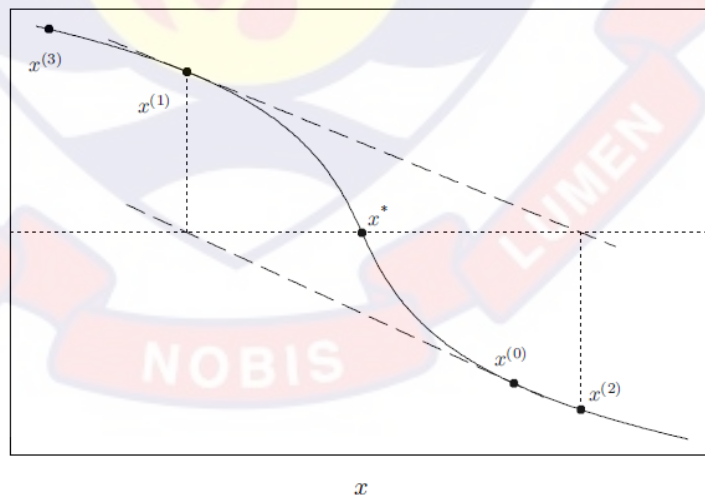


Figure 13: Second Illustration of Newton's Method

neighborhood of x^* where $g''(x^*) \neq 0$ for all x . Within this neighborhood, let's

define $\varepsilon^{(t)} = x^{(t)} - x^*$. By performing a Taylor expansion, we obtain

$$0 = g'(x^*) = g'(x^{(t)}) + (x^* - x^{(t)})g''(x^{(t)}) + \frac{1}{2}(x^* - x^{(t)})^2 g'''(q) \quad (172)$$

where q lies between $x^{(t)}$ and x^* . Upon rearranging terms, we get

$$x^{(t)} + h^{(t)} - x^* = (x^* - x^{(t)})^2 \frac{g'''(q)}{2g''(x^{(t)})}, \quad (173)$$

with $h^{(t)}$ representing the Newton update increment (Givens & Hoeting, 2012).

Since the left-hand side is equivalent to $x^{(t+1)} - x^*$, we can conclude

$$\varepsilon^{(t+1)} = (\varepsilon^{(t)})^2 \frac{g'''(q)}{2g''(x^{(t)})}. \quad (174)$$

In the context of a neighborhood of x^* denoted as $N\delta(x^*) = [x^* - \delta, x^* + \delta]$, for

$\delta > 0$, we define

$$c(\delta) = \max_{x_1, x_2 \in N\delta(x^*)} \left| \frac{g'''(x_1)}{2g''(x_2)} \right|. \quad (175)$$

Since

$$c(\delta) \rightarrow \left| \frac{g'''(x^*)}{2g''(x^*)} \right| \quad (176)$$

as $\delta \rightarrow 0$, it follows that $\delta c(\delta) \rightarrow 0$. Choosing δ such that $\delta c(\delta) < 1$, if $x^{(0)} \in N\delta(x^*)$, Equation (174) implies

$$|c(\delta)\varepsilon^{(i+1)}| \leq (c(\delta)\varepsilon^{(i)})^2. \quad (177)$$

Assuming the initial value is reasonably close, $|\varepsilon^{(0)}| = |x^{(0)} - x^*| \leq \delta$, Equation (177) implies

$$|\varepsilon^{(\theta)}| \leq \frac{(c(\delta)\delta)^z}{c(\delta)}. \quad (178)$$

This expression tends to zero as $t \rightarrow \infty$, which indicates that $x^{(0)}$ approaches x^4 .

Thus, we have established the following theorem: given a continuous g''' and a simple root x^4 of g' , there is a neighbourhood of x^4 such that, starting from any $x^{(0)}$ within that neighbourhood, Newton's method converges to x^4 .

Newton's method undoubtedly shows convergence to the root from any initial point if g' is twice continuously differentiable, convex, and has a root. When the method is called within a range $[a, b]$, it is necessary to confirm the following extra conditions:

- $g''(x^*) \neq 0$ on $[a, b]$.
- $g'''(x^*)$ does not change sign on $[a, b]$.
- $g'(a)g'(b) < 0$, and
- $|g'(a)/g''(a)| < b - a$ and $|g'(b)/g''(b)| < b - a$.

If these conditions are satisfied, Newton's method will converge from any $x^{(0)}$ within the interval.

The efficiency of a root-finding method, such as Newton's method, is commonly quantified by its convergence order. A method possesses a convergence order β if $\lim_{t \rightarrow \infty} \varepsilon^{(t)} = 0$ and

$$\lim_{t \rightarrow \infty} \frac{|e^{(i+1)}|}{|e^{(i)}|^p} = c \quad (179)$$

for certain constants $c \neq 0$ and $\beta > 0$. While higher orders of convergence are desirable, as they imply a quicker attainment of a precise solution, it is noteworthy that achieving high orders can sometimes compromise robustness. Slower

algorithms may be more foolproof compared to their faster counterparts. Equation (174) for Newton's method suggests that, if the method converges, the right-hand side converges to $g'''(x^*) / [2g''(x^*)]$ for $\epsilon^{(t+1)} / (\epsilon^{(t)})^2$. This indicates quadratic convergence ($\beta = 2$), with the constant c defined as

$$c = \left| \frac{g'''(x^*)}{2g''(x^*)} \right|. \quad (180)$$

Quadratic convergence is rapid, doubling the precision of the solution with each iteration.

For bisection, the length reduction of the bracketing interval resembles linear convergence ($\beta = 1$), halving at each iteration. If a root exists in the initial interval, $\lim_{t \rightarrow \infty} |\epsilon^{(t)}| = 0$. However, the distances $x^{(t)} - x^*$ need not consistently decrease, and their ratio can be unbounded. Thus,

$$\lim_{t \rightarrow \infty} \frac{|\epsilon^{(t+1)}|}{|\epsilon^{(t)}|^\beta} \quad (181)$$

may not exist for any $\beta > 0$, and bisection does not formally meet the definition of convergence order. Bisection can be employed as a robust bracketing method to safeguard faster but less reliable methods like Newton's. Rather than generating steps, view bracketing as providing an interval within which a root must lie. If Newton's method exceeds these bounds, the step should be adjusted, truncated, or redirected, potentially reducing the method's convergence order.

Model Selection Techniques for Assessing the Efficiency of the Distribution in Fitting Dataset

In the area of model selection for data analysis, a critical challenge arises

in striking a balance between the competing concerns of variance and bias, commonly referred to as the trade-off between overfitting and underfitting. A model endowed with a greater number of parameters typically exhibits lower bias but higher variance, while a model with fewer parameters tends to be more biased yet with lower variance. Consequently, the overarching objective is to identify a well-fitted model that minimizes information loss.

The evaluation of a model's effectiveness in capturing the underlying patterns in the data can be gauged through the coverage of the parameter's confidence interval. This assessment provides insights into whether the models have been judiciously selected or not. Information criteria offer a systematic approach to comparing non-nested models, establishing an order from the best-performing to the least. These models are then scaled using measures like evidence ratios and Akaike weights. The Akaike Information Criterion (AIC), the Consistent Akaike Information Criterion (CAIC), and the Bayesian Information Criterion (BIC) are a few commonly used information criteria. These criteria are essential to the model selection process because they offer numerical measurements that direct the judgement process.

One popular statistical metric that is essential for choosing and comparing models is the AIC. It was created by Hirotugu Akaike and offers a compromise between a model's complexity and goodness of fit. AIC can be used with likelihood-based models, such as time series analysis, linear regression, and other statistical models. It is described as

$$AIC = -2 \cdot \log(ML) + 2 \cdot (\text{Number of Parameters}) \quad (182)$$

The objective is to reduce AIC's absolute value. AIC takes into account a trade-off between the model's complexity and goodness of fit. To deter overfitting, it penalises models with more parameters. When comparing different models fitted to the same dataset, AIC is very helpful. A better trade-off between model fit and simplicity is indicated by lower AIC values. The model with the lowest AIC is considered the most parsimonious while still providing an adequate fit to the data. AIC is a relative measure; the absolute value itself is not meaningful. It is the differences in AIC between models that are informative. Smaller differences in AIC suggest stronger evidence in favor of the model with the lower AIC. In very large sample sizes, AIC might tend to favor more complex models because the penalty for additional parameters becomes relatively smaller compared to the improvement in fit. In such cases, alternative criteria like the BIC might be considered, as BIC has a stronger penalty for model complexity.

The BIC is a criterion for model selection that penalizes the number of parameters in a statistical model. It is a variant of the AIC and is particularly useful when dealing with relatively small sample sizes. BIC is calculated using the formula

$$\text{BIC} = -2 \cdot \log(\text{ML}) + 2 \cdot (\text{Number of Parameters}) \times \log(\text{Sample Size}) \quad (183)$$

The BIC incorporates a penalty for the number of parameters, which becomes more pronounced as the sample size decreases. This penalty term, Number of Parameters and $\log(\text{Sample Size})$, discourages overly complex models, making BIC especially valuable when choosing between models with different numbers of parameters. The BIC provides a balance between goodness of fit and model

complexity, and its penalty term is particularly useful in situations where the sample size is limited. Lower BIC values indicate preferable models, taking into account both model fit and simplicity.

In essence, the careful consideration of information criteria is instrumental in navigating the intricate landscape of model selection, ensuring that the chosen model strikes an optimal equilibrium between complexity and explanatory power.

When assessing the marginal distributions of the multivariate EPD, a valuable tool for evaluation comes in the form of Q-Q plots, short for Quantile-Quantile plots. These plots are particularly useful for examining the correspondence between the sample observations and the expected quantiles under the assumption of an exponential power distribution. To construct Q-Q plots for the marginal distributions, the sample quantiles are plotted against the expected quantiles that will be observed if the data were truly following an EPD. This graphical representation offers a visual means of assessing how closely the observed data aligns with the theoretical distribution. Deviations from a straight line in a Q-Q plot may indicate departures from the assumed distribution. The tandem use of Q-Q plots serves as a robust methodology for scrutinizing the goodness-of-fit of marginal distributions in the context of the Multivariate Exponential Power Distribution. These visual diagnostics aid in confirming or challenging the assumed distributional form and provide valuable insights into the characteristics of the data.

The Application of Multivariate EPD in Multivariate Data Analysis

The Multivariate EPD finds applicability across various multivariate statistical techniques. It can be effectively utilized in areas such as multivariate multiple regression, multiple regression with unequal variances, hypothesis testing of means, and more. However, the primary emphasis of this particular study lies in its application to discriminant and classification analysis.

The Basic Concept of Discriminant and Classification Analysis

Discriminant analysis is employed to classify observations into predetermined classes by assigning a class to an observation based on a set of predictor variables. The discriminant function model is developed using a training set, where the classes of observations are already known. Descriptive statistics play a crucial role in extracting valuable information from the data during discriminant analysis.

The arithmetic mean of n measurements on the first variable, denoted as $x_{11}, x_{21}, \dots, x_{n1}$, is given by:

$$\bar{x}_1 = \frac{1}{n} \sum_{j=1}^n x_{j1} \quad (184)$$

This mean, termed the sample mean for the first variable, is computed as these measurements represent a subset of the complete set of potential observations. The sample mean can be calculated similarly for each of the p variables, resulting in p sample means:

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk} \quad (185)$$

The sample variance, denoted as S_{11}^2 for n measurements on the first variable, is a measure of spread defined as:

$$S_{11}^2 = \frac{1}{n} \sum_{j=1}^n (x_{j1} - \bar{x}_1)^2 \quad (186)$$

Generalizing for p variables, the sample variance for the k th variable is given by:

$$S_{kk}^2 = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_k)^2 \quad (187)$$

The sample variance squared is $\sqrt{S_{kk}^2}$, which represents the sample standard deviation. Taking into account n pairs of measurements for variables 1 and 2:

$$x_{11}, x_{21}, \dots, x_{n1} \quad x_{12}, x_{22}, \dots, x_{n2}$$

The linear relationship between measurements of variables 1 and 2 is measured by the sample covariance, or S_{12} :

$$S_{12} = \frac{1}{n} \sum_{j=1}^n (x_{j1} - \bar{x}_1) (x_{j2} - \bar{x}_2) \quad (188)$$

Large positive values of S_{12} indicate simultaneous occurrences of large values, while large negative values indicate an inverse relationship. If there is no specific association, S_{12} is approximately zero.

The general sample covariance between the i th and k th variables is given by:

$$S_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ji} - \bar{x}_i) (x_{jk} - \bar{x}_k) \quad i = 1, 2, \dots, p; j = 1, 2, \dots, p \quad (189)$$

This reduces to the sample variance when $i = k$.

The sample correlation coefficient, denoted as τ_{ik} , measures the linear association between variables independent of unit measurements:

$$\tau_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} \quad (190)$$

The product of the square roots of the sample variances provides standardisation for this version of the sample covariance.

Equal And Unequal Covariance Matrices

In the context of discrimination and classification, a crucial assumption involves the equality of covariance matrices across distinct populations. The equality of population covariance matrices must be evaluated prior to combining variations from various samples to create a pooled sample covariance matrix. Box's M-test is the frequently used test for this purpose. For a set of g populations, the null hypothesis is expressed as

$$\Sigma_1 = \Sigma_2 = \cdots \Sigma_g = \Sigma \quad (191)$$

In this case, $l = 1, 2, \cdots g$, and Σ stands for the assumed common variance matrix. It also indicates the covariance matrix for the l th population. According to the alternative hypothesis, the covariance matrices are not equal in at least two cases.

Probability of Misclassification

Let f_1 and f_2 represent the PDFs for the $p \times 1$ vector random variable $\mathbf{x}$ in populations Π_1 and Π_2 , respectively. Objects with associated measurements

$\mathbf{x}$ are classified into sets $\mathbb{R}_1$ and $\mathbb{R}_2$ based on these PDFs. The conditional probabilities $P(2 | 1)$ defined as

$$P(2 | 1) = P(\mathbf{x} \in \mathbb{R}_2 | \pi_1) = \int_{\mathbb{R}_2 = \Omega - \mathbb{R}_1} f_1(\mathbf{x}), d\mathbf{x} \quad (192)$$

and $P(1 | 2)$ represent the likelihood of misclassification from $\mathbb{R}_1$ to $\mathbb{R}_2$ given as

$$P(1 | 2) = P(\mathbf{x} \in \mathbb{R}_1 | \pi_2) = \int_{\mathbb{R}_1} f_2(\mathbf{x}), d\mathbf{x} \quad (193)$$

and vice versa. The prior probabilities of populations π_1 and π_2 are denoted as p_1 and p_2 , respectively. The overall probabilities of correct and incorrect classifications are given by

1. Probability of correct classification into π_1 :

$$P(\text{observation is correctly classified as } \pi_1) = P(1 | 1)p_1$$

2. Probability of correct misclassification into π_1 :

$$P(\text{observation is correctly misclassified as } \pi_1) = P(1 | 2)p_2$$

3. Probability of correct classification into π_2 :

$$P(\text{observation is correctly classified as } \pi_2) = P(2 | 2)p_2$$

4. Probability of correct misclassification into π_2 :

$$P(\text{observation is correctly misclassified as } \pi_2) = P(2 | 1)p_1$$

A confusion matrix (Table 2) summarizes the classification results, considering the costs of misclassification.

Table 2: Confusion Matrix for the Cost of Misclassification

		Classify Into		
		π_1	π_2	
True Population	π_1	0	$c(1 2)$	n_1
	π_2	$c(2 1)$	0	n_2

Source: Onaseso and Olosunde (2017)

Cost of Misclassification

The cost matrix in Table 2 determines the cost of misclassification. For correct classifications, the cost is zero; for misclassifying an observation from π_2 as π_1 , the cost is $C(1 | 2)$; and for misclassifying an observation from π_1 as π_2 , the cost is $C(2 | 1)$. For any given rule, the expected cost of misclassification (ECM) is calculated by multiplying the probabilities of occurrence of the off-diagonal entries, which are derived from the probabilities of misclassification.

ECM is therefore expressed as

$$\text{ECM} = C(2 | 1)P(2 | 1)p_1 + C(1 | 2)P(1 | 2)p_2 \quad (194)$$

By satisfying the inequalities, the values of x determine the regions $\mathbb{R}_1$ and $\mathbb{R}_2$ that minimise ECM

$$\mathbb{R}_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{C(1 | 2)}{C(2 | 1)} \right) \quad (195)$$

and

$$\mathbb{R}_2 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left(\frac{C(1 | 2)}{C(2 | 1)} \right) \quad (196)$$

Under conditions of equal probabilities and equal misclassification costs, these regions simplify to

$$\mathbb{R}_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \quad (197)$$

and

$$\mathbb{R}_2 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < 1 \quad (198)$$

Under circumstances where probabilities are equal, and misclassification costs are the same, these regions can be expressed as $\mathbb{R}_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1$ and $\mathbb{R}_2$ otherwise (Johnson & Wichern, 2013).

Separation and Classification for Two Populations

For a $p \times 1$ random vector in populations Π_1 and Π_2 , consider $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ as the corresponding probability density functions (PDFs). The sample space is indicated by Ω . $\mathbb{R}_1 \subset \Omega$ is the set of values that determine which objects belong in Π_1 . Moreover, let $\mathbb{R}_2 = \Omega - \mathbb{R}_1$ denote the remaining x values for which the objects belong to Π_2 . This partitioning guarantees that the sets $\mathbb{R}_1$ and $\mathbb{R}_2$ are mutually exclusive and cover the whole sample space Ω .

Classification With Two Multivariate Normal Populations

Within the framework of a normal distribution, it is assumed that there exist n_k observations in class k . Each p -dimensional multivariate class is distributed as follows: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K \stackrel{IID}{\sim} N_p(\mu_k, \Sigma_k)$, where μ_k and Σ_k denote the mean vector and covariance matrix of the p -dimensional multivariate normal

distribution for each known class k . $n = n_1 + \dots + n_K$ is the total number of observations. With $\pi_1 + \dots + \pi_K = 1$, let π_k represent the prior probability of observing a class k member.

A new subject $\mathbf{x}$ is assigned to class k in order to minimise the discriminant score, assuming a normal distribution

$$D_k(\mathbf{x}) = (\mathbf{x} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x} - \mu_k) + \ln |\Sigma_k| - 2 \ln \pi_k. \quad (199)$$

The assignment is made to $\hat{k} = \operatorname{argmin}_k D_k(\mathbf{x})$ in $\mathbf{x}$. Since there are quadratic boundaries separating the disjoint regions that belong to each class, this method is called Quadratic Discriminant Analysis (QDA) (Aitchison & Silvey, 1958; Cox, 1966).

When all covariance matrices are equal, that is, $\Sigma_k = \Sigma$ for every k , the discriminant score reduces to

$$d_k(\mathbf{x}) = (\mathbf{x} - \mu_k)^\top \Sigma^{-1} (\mathbf{x} - \mu_k) - 2 \ln \pi_k \quad (200)$$

This is known as linear discriminant analysis (LDA), in which a new subject is assigned using linear boundaries in the following formula: $\hat{k} = \operatorname{argmin}_k d_k(\mathbf{x})$.

When known, the covariance matrices Σ_k and mean vectors μ_k are estimated using their maximum likelihood estimates. Typically, the fraction of each class in the pooled training sample is used to estimate the prior probabilities, as in $\hat{\pi}_k = n_k/n$. The QDA sample version rule is provided by

$$\ell(x) = \operatorname{argmin}_k \hat{D}_k(\mathbf{x}) \quad (201)$$

where

$$\hat{D}_k(\mathbf{x}) = (\mathbf{x} - \hat{\mu}_k)^\top \hat{\Sigma}_k^{-1} (\mathbf{x} - \hat{\mu}_k) + \ln |\hat{\Sigma}_k| - 2 \ln \pi_k \quad (202)$$

Likewise, the LDA sample version rule is

$$\ell = \operatorname{argmin}_k \hat{d}_k(\mathbf{x}) \quad (203)$$

where

$$\hat{d}_k(\mathbf{x}) = (\mathbf{x} - \hat{\mu}_k)^\top \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}_k) - 2 \ln \pi_k \quad (204)$$

In situations with unclear separation boundaries, minimizing the cost of misclassification becomes crucial. Define $c(\cdot | k)$ as the cost of misallocating a new subject $\mathbf{x}$ to class k , then $c(\cdot | k) = 1 - \Pr(\hat{D}_k(\mathbf{x}) | \hat{\mu}_k, \hat{\Sigma}_k)$ (Anderson, 1972; Hands and Henley, 1997). Compared to standard linear discriminators, logistic discrimination is better for a wider class of distributions. This was proposed by Cox (1966) and Day and Kerridge (1967). It is possible to extend the method to situations in which the likelihood ratios in the observations are quadratic. Due to its versatility and efficiency with different population models, the normal population classification method is widely used in statistical practice.

Classification of a Normal Population when $\Sigma_1 = \Sigma_2$

Assuming joint densities for populations π_1 and π_2 given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{-\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_i)^\top \Sigma^{-1} (\mathbf{x} - \mu_i) \right] \quad (205)$$

where $\mathbf{x}^\top = [X_1, X_2 \dots X_p]$, and the population parameters μ_1 , μ_2 , and Σ are known, the minimum ECM regions in Equation (194) can be expressed as follows after canceling terms $(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}$

$$\begin{aligned} \mathbb{R}_1 : \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_1)^\top \Sigma^{-1} (\mathbf{x} - \mu_1) + \frac{1}{2} (\mathbf{x} - \mu_2)^\top \Sigma^{-1} (\mathbf{x} - \mu_2) \right] &\geq \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \\ \mathbb{R}_2 : \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_1)^\top \Sigma^{-1} (\mathbf{x} - \mu_1) + \frac{1}{2} (\mathbf{x} - \mu_2)^\top \Sigma^{-1} (\mathbf{x} - \mu_2) \right] &< \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \end{aligned} \quad (206)$$

These expressions define the regions $\mathbb{R}_1$ and $\mathbb{R}_2$ based on the likelihood ratios, where the cost matrix values and prior probabilities influence the classification boundaries.

Classification of a Normal Population when $\Sigma_1 \neq \Sigma_2$

The consideration of multivariate densities in Equation (42) involves $\Sigma_i, i = 1, 2$, indicating unequal covariance matrices and mean vectors. In contrast to the scenario where $\Sigma_1 = \Sigma_2$, the presence of unequal covariance matrices prevents the cancellation of terms involving $|\Sigma_i|^{\frac{1}{2}}$ in the density ratio. Following Equation (194), substitution of multivariate densities with distinct covariance matrices and natural logarithm computation, the expressions for the minimum ECM regions are obtained as

$$\begin{aligned}\mathbb{R}_1 : \frac{-1}{2} \mathbf{x}^\top (\Sigma_1^{-1} - \Sigma_2^{-1}) \mathbf{x} + (\mu_1^\top \Sigma_1^{-1} - \mu_2^\top \Sigma_2^{-1}) \mathbf{x} - k &\geq \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right] \\ \mathbb{R}_2 : \frac{-1}{2} \mathbf{x}^\top (\Sigma_1^{-1} - \Sigma_2^{-1}) \mathbf{x} + (\mu_1^\top \Sigma_1^{-1} - \mu_2^\top \Sigma_2^{-1}) \mathbf{x} - k &< \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]\end{aligned}\quad (207)$$

where

$$k = \frac{1}{2} \ln \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} (\mu_1^\top \Sigma_1^{-1} \mu_1 - \mu_2^\top \Sigma_2^{-1} \mu_2)$$

To allocate a new observation x_0 to π_1 or π_2 , the following decision rule can be applied:

Allocate x_0 to π_1 if

$$\frac{-1}{2} x_0^\top (\Sigma_1^{-1} - \Sigma_2^{-1}) x_0 + (\mu_1^\top \Sigma_1^{-1} - \mu_2^\top \Sigma_2^{-1}) x_0 - k \geq \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right] \quad (208)$$

Allocate x_0 to π_2 otherwise.

This approach employs the Normal distribution as the underlying distribution for discriminating and classifying multivariate data using the likelihood

ratio method. The subsequent chapter will introduce AMEPD as the underlying distribution for the discrimination and classification of multivariate data through the likelihood ratio method. With the primary objective of lowering the cost of misclassification when assigning an unknown subject to one of k classes based on the multivariate observation $\mathbf{x} = (X_1, \dots, X_p)^\top$, where p is the number of features in each class, this will specifically address multivariate data with tails heavier or lighter than the usual normal distribution.

Evaluating Classifications Functions for Two Populations

If information about the parent populations is seldom available, Onaseso and Olosunde (2017) proposed a methodology for assessing the performance of a classification method by focusing on error rates. The evaluation of straight-forward classifications involves computing the actual error rate (AER), defined as

$$\text{AER} = p_1 \int_{\hat{\mathbb{R}}_2} f_1(\mathbf{x}), d\mathbf{x} + p_2 \int_{\hat{\mathbb{R}}_1} f_2(\mathbf{x}), d\mathbf{x}. \quad (209)$$

Here, $\hat{\mathbb{R}}_1$ and $\hat{\mathbb{R}}_2$ denote the classification regions based on the sizes of n_1 and n_2 , respectively.

The AER is the proportion of misclassifications made by the discriminant function on the entire population or the true distribution of data. The AER is a useful metric for predicting the future performance of the sample classification function. This rate is usually unknown because of not having access to the full population, thus, direct calculation is not feasible since the density functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are unknown. Instead, we work with a sample of data using

some approaches. One of the approach is the use of Lachenbruch's holdout procedure which can yield an almost unbiased estimate of the expected AER, $\hat{E}(\text{AER})$, under the assumption of equal prior probabilities and misclassification costs. This estimate is obtained from the confusion matrix using methods like cross-validation and jackknifing, which eliminate the need for distributional assumptions. Another is the use of Average Performance Error Rate (APER) which is also called the Estimated Error Rate. This refers to the error rate that is estimated using a sample of data. Several methods like cross-validation, resampling techniques (e.g., bootstrapping), or splitting data into training and test sets are used to estimate the error rate. The APER serves as an approximation of the actual error rate. Since we do not have access to the full population, this estimate helps to give a sense of how well the discriminant function will perform on new, unseen data.

Table 3 displays the confusion matrix for n_1 and n_2 observations from populations AMEPD Π_1 and Π_2 , respectively. n_1M represents the number of items from Π_i that were incorrectly classified into Π_j ($i \neq j$), and n_1C represents the number of items from Π_i that were correctly classified into Π_i (for $i = 1, 2$).

The following provides an objective estimate of the APER:

$$APER = \frac{n_1M^{(H)} + n_2M^{(H)}}{n_1 + n_2} \quad (210)$$

where $n_1M^{(H)}$ represents the number of holdout observations misclassified into the i^{th} groups ($i = 1, 2$).

Table 3: Confusion Matrix for Observations from Known Populations

		Predicted Membership		
		π_1	π_2	
Actual Membership	π_1	n_1C	$n_1M = n_1 - n_1C$	n_1
	π_2	$n_2 = n_2 - n_2C$	n_2C	n_2

Source: Onaseso and Olosunde (2017)

If APER is applied using the densities of AMEPD, it will result in the creation of the confusion matrix presented in Table 3.

Chapter Summary

The study utilizes a simulated Bivariate Data generated from a population distributed as the Asymmetrically Modified Empirical Predictive Distributions (AMEPD) for varying sample sizes (150, 200, 500, 1000, 10000). Additionally, three real datasets are employed in the investigation. The first one is a Carsales Dataset which is a sample data in IBM SPSS 27, containing estimates for hypothetical car sales, list prices, and physical specifications. It involves 11 variables with 156 observations, including Sales, Price, Type of Vehicle, Engine Size, Horsepower, Wheelbase, Width, Length, Curb Weight, Fuel Capacity, and Fuel Efficiency. It is considered for small sample size analysis. Vehicle-related variables are expressed in thousands of US dollars, cubic centimeters, centimeters, and kilograms. The second dataset is a Concrete Dataset (Yeh, 2006) comprises 9 variables with 1133 observations, including Cement, Blast-Furnace

Slag (BFS), Fly Ash (F. Ash), Water, Super-Plasticizer (SP), Coarse Aggregate (CA), Fine Aggregate (Fine A.), Age of Testing (AoT), and Concrete Compressive Strength (CCS). It is considered for large sample size analysis. Variables measure density, weight per unit volume, and age in days or compressive strength in megapascals. The final dataset is from Ghana Minerals Commission containing monthly records of heavy metal data (mercury, copper, lead, and arsenic) in the Pra River, associated with illegal mining activities. It involves seven variables with 121 observations, including town, season, year, amount of mercury, amount of copper, amount of lead, and amount of arsenic. It also considered for small sample size analysis. Variables are measured in milligrams per liter (mg/l).

This chapter introduces two essential distributions, the Gamma and Dirichlet distributions, playing a crucial role in developing elliptical and spherical distributions. The gamma function plays a role in the development of univariate and multivariate cases of the Exponential Power distribution. It is a continuous extension of the factorial function, specifically convex for positive real numbers. This function helped to simplify the characteristics of the EPD used in this study, since the distribution depicts the properties of the Gamma function. The Dirichlet distribution, a multivariate extension of the Beta distribution, is foundational in statistical modeling, distribution theory, and Bayesian inference. Derived from the Gamma distribution, it is especially relevant for compositional data and modeling proportions. The Dirichlet distribution exhibits interesting amalgamation properties, allowing for partitioning and combining components with

specific distributions. Additionally, it demonstrates associations between components and has applications in deriving multivariate EPD. The mathematical preliminaries also include the Liouville distribution, and various integral expressions relevant to the derivation and analysis of multivariate EPD.

The chapter also discusses the definition and properties of spherical and elliptical symmetric distributions. The spherical symmetric distribution is said to be a multivariate distribution with independent variables having the same variance. The Covariance matrix is a diagonal matrix with equal elements on the diagonal. Also, PDF depends only on the Euclidean distance from the central point. On the other hand, elliptical symmetric distribution allows correlations between variables and different variances. The covariance matrix is not restricted to a diagonal matrix. Properties of Spherical and Elliptical Symmetric Distribution are presented in Theorem 1 (characterizes spherical distribution based on the characteristic function), Theorem 2 (Schoenberg, 1938) which describes characteristic generators in terms of an integral involving a function related to the unit sphere's surface. Theorems 3 to 8 address various properties including moments, covariance, density generators, and conditions for having a density. The relationships between moments, cumulants, and parameters like multivariate kurtosis are discussed. The again discusses the uniform distribution on the unit sphere surface and within a unit sphere. We have presented the density, probability distribution, and characteristic function for the uniform distribution.

The chapter also contains detailed information about the properties and

derivatives of multivariate normal distributions, as well as the properties of multivariate EPD. The content covers various mathematical expressions, transformations, and properties associated with these distributions. It presents Multivariate Normal Distribution as a linear transformation applied to independent normal random variables. It introduces the transformation function and standardization of independent normal variables. It discusses linear transformations of standard normal variables and their representation. The properties such as the observations related to the rank of the transformation matrix are presented. It discusses cases of linear dependence and independence of component variables. It highlights conditions for singular and non-singular distributions. The Non-Singular property is presented in Theorem 9. Also, the chapter defines the MEPD and its parameters, including location, scale, and shape. It discusses the stochastic representation and density function of MEPD. It provides the CF and various properties of MEPD, including mean, variance, asymmetry, and kurtosis coefficients.

The chapter presents the method of deriving Spherical and Elliptical Symmetric Distribution. It discusses the derivation of spherical and elliptical symmetric distributions. It highlights the process involving density functions, spherical coordinates, and normalization. The four approaches of deriving Spherical and elliptical symmetric distribution are discussed. The approaches discussed are by means density, by means of the characteristic function, by means of linear combination and by means of stochastic decomposition.

The chapter discusses approaches to parameter estimation, focusing on

the maximum likelihood estimation (MLE) method due to its reliability and efficiency. The theoretical justifications for adopting MLE involve assumptions about distinct cumulative distribution functions, common support of probability density functions, and the true parameter being an interior point. The asymptotic properties of MLE, including its convergence to the true value, are discussed.

The text then extends the discussion to univariate and multivariate cases, emphasizing the use of likelihood functions and log-likelihood functions. In the multivariate case, the Newton-Raphson optimization method is introduced for estimating the shape parameter. The challenges associated with numerical methods for estimating the shape parameter in multivariate contexts are acknowledged. The Newton-Raphson method is chosen for its robustness in addressing these challenges. The iterative nature of Newton's method and its application to the multivariate case are detailed. The text further explores Newton's method, providing examples and illustrations of its application. The convergence properties of Newton's method are discussed, emphasizing quadratic convergence.

The discussion also compares Newton's method with the bisection method in terms of convergence and robustness. Again, the text delves into model selection techniques, addressing the trade-off between bias and variance. Information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are introduced as tools for comparing non-nested models. The text explains the calculation of AIC and BIC, emphasizing their roles in balancing model fit and complexity. The importance of model evaluation through Q-Q plots for the multivariate EPD is highlighted.

The chapter discusses the application of multivariate EPD in Multivariate Data Analysis, specifically focusing on discriminant and classification analysis. multivariate EPD is applicable in multivariate statistical techniques such as multivariate multiple regression, multiple regression with unequal variances, and hypothesis testing of means. Emphasis in the study is on its application to discriminant and classification analysis. The discriminant analysis categorizes observations into predefined classes based on predictor variables. Sample means, variances, and covariances are computed for each variable. Linear association between variables is measured by sample covariance and correlation coefficients. Assumptions include the equality of covariance matrices across populations. Box's M-test can be used to test the equality of population covariance matrices. Probability of misclassification is calculated based on conditional probabilities and sample space. Confusion matrix and cost of misclassification are introduced. Cost matrix is defined, and the ECM is computed. Regions minimizing ECM are determined. Two scenarios are considered: equal covariance matrices and unequal covariance matrices. Classification regions are determined based on discriminant scores. Actual error rate (AER) is introduced as an indicator of classification performance. APER is used for estimating expected AER. The chapter introduces AMEPD as an alternative distribution for discrimination and classification. Confusion matrix and Lachenbruch's holdout procedure are discussed in the context of AMEPD. The chapter concludes by mentioning the use of AMEPD in the discrimination and classification of multivariate data to reduce the cost of misclassification.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

This chapter deals with the derivation of the proposed AMEPD and its properties and characteristics. Also, the location, scale and shape parameters of the distribution will be derived and discussed while the techniques for deriving the parameters will be dealt with. Moreover, the statistical techniques for measuring the accuracy of a theoretical distribution to the distribution of real life data, as well as the techniques for goodness-of-fit testing will be discussed. Again, the bivariate distribution of the AMEPD is derived. The maximum likelihood estimation of parameters is derived and it is applied to datasets. Finally, the AMEPD is applied in discriminant analysis.

Derivation of AMEPD

The univariate EPD in Equation (4) may be stated in the form

$$P(x|\mu, \sigma, q) = k\sigma^{-1} \exp \left\{ -c |(x - \mu) \sigma^{-1}|^{q/2} \right\}, \quad (211)$$

with expectation, $E(X) = \mu$ and variance, $Var(X) = v\sigma^2$, where k is the normalizing constant and $v = c^{-2/q} \frac{\Gamma(\frac{3}{q})}{\Gamma(\frac{1}{q})}$. For any spherical distribution, we can extend the univariate to multivariate distribution in a number of ways (see Fang, et al., 1991). One of these approaches which is mainly adopted in the extension is the stochastic decomposition. Now, we define the multivariate EPD as a family of distributions generated by linear transform, or linear vector functions of

independent Exponential Power (EP) random variables of the form

$$\mathbf{y} = \Lambda \mathbf{x} + \mathbf{b} \quad (212)$$

where Λ is $n \times p$ matrix of transform coefficients, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$,

$\mathbf{y} = (y_1, y_2, \dots, y_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$.

Let $\mathbf{X}$ be independent EP variables with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2 v$.

Standardizing the EP gives

$$Z_i = \frac{x_i - \mu_i}{\sqrt{v} \sigma_i}. \quad (213)$$

In terms of standard normal,

$$Z_{(N)i} = Z_i \sqrt{v} = \frac{x_i - \mu_i}{\sigma_i}. \quad (214)$$

Then, $\mathbf{Z}_{(N)} = (Z_{(N)1}, Z_{(N)2}, \dots, Z_{(N)p})$ is a vector of IID standard normal variables. Thus, $X_{(i)} = \sigma_i Z_{(N)i} + \mu_i$. Therefore, we can write

$$\mathbf{X} = \mathbf{S} \mathbf{Z}_{(N)} + \boldsymbol{\mu} \quad (215)$$

where $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)$. Thus, any linear transform in Equation (212) may be expressed as $\mathbf{y} = \mathbf{B} \mathbf{Z}_{(N)} + \mathbf{d}$ where $\mathbf{B} = \Lambda \mathbf{S}$ and $\mathbf{d} = \Lambda \boldsymbol{\mu} + \mathbf{b}$, which is a linear transform of standard normal random variables. In Equation (215), the mean vector and variance-covariance matrix of $\mathbf{X}$ are $E(\mathbf{X}) = \boldsymbol{\mu}$ and $D(\mathbf{X}) = \mathbf{S} \mathbf{S}^\top v$. Since $\mathbf{X}$ is non-singular and distributed as AMEPD($\boldsymbol{\mu}, \Sigma, q$), then there exists a non-singular matrix $\mathbf{S}$ such that

$$\mathbf{S} \mathbf{S}^\top = \Sigma \quad (216)$$

and

$$\mathbf{Z}_{(N)} = \mathbf{S}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad (217)$$

is a vector of independent standard Normal random variables of dimension p . Now every multivariate EP distribution may be determined by some non-singular distribution. With Equation (215) and Equation (217) as the inverse, We define non-singular transformation $\mathbf{Z}_{(N)} \longrightarrow \mathbf{X}$ given by

$$\mathbf{X} = \mathbf{A}^\top \mathbf{Z}_{(N)} + \boldsymbol{\mu} \quad (218)$$

for $\mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$. The Jacobian of the inverse transformation of $\mathbf{Z}_{(N)}$ is given as

$$J_{\mathbf{Z}_{(N)}}(\mathbf{x}) = \det(\mathbf{A}^{-1}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \quad (219)$$

From the distribution in Equation (211), the density of $r = \mathbf{Z}_{(N)}^\top \mathbf{Z}_{(N)}$ is given as

$$f(r) = k \exp \left\{ -c \left| \mathbf{Z}_{(N)}^\top \mathbf{Z}_{(N)} \right|^{q/2} \right\}. \quad (220)$$

According to Fang and Zhang (1990), if $\mathbf{Z}_{(N)}$ has a density (??) for some non-negative function, then

$$\int_{\mathbb{R}^p} f(\mathbf{z}_{(N)}) d\mathbf{z}_{(N)} = \frac{\pi^{p/2}}{\Gamma(\frac{p}{2})} \int_0^\infty y^{\frac{p}{2}-1} f(y) dy = 1. \quad (221)$$

Thus, the function can be used to define the density function of a spherical distribution if and only if $\int_0^\infty y^{\frac{p}{2}-1} f(y) dy < \infty$, where $y = \mathbf{z}_{(N)}^\top \mathbf{z}_{(N)}$. Based on Equation (221), k in Equation (??) can be determined by

$$\int_{\mathbb{R}^p} k \exp \left\{ -c \left| \mathbf{z}_{(N)}^\top \mathbf{z}_{(N)} \right|^{q/2} \right\} d\mathbf{z}_{(N)}. \quad (222)$$

Now, for

$$k \frac{\pi^{p/2}}{\Gamma(\frac{p}{2})} \int_0^\infty y^{\frac{p}{2}-1} \exp \left\{ -c y^{q/2} \right\} dy = 1,$$

we let $t = c y^{q/2}$ so that $y = \frac{t^{2/q}}{c^{2/q}}$, $y^{p/2-1} = \frac{t^{p/q-2/q}}{c^{p/q-2/q}}$ and $dy = \frac{2t^{2/q-1}}{qc^{2/q}} dt$.

Thus,

$$\begin{aligned}\int_0^\infty y^{\frac{p}{q}-1} \exp\{-cy^{q/2}\} dy &= \int_0^\infty \frac{t^{p/q-2/q}}{c^{p/q-2/q}} e^{-t} \frac{2t^{2/q-1}}{qc^{2/q}} dt \\ &= \frac{2}{qc^{p/q}} \int_0^\infty t^{p/q-1} e^{-t} dt \\ &= \frac{2}{qc^{p/q}} \Gamma\left(\frac{p}{q}\right).\end{aligned}$$

Therefore,

$$k \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}\right)} \frac{2}{qc^{p/q}} \Gamma\left(\frac{p}{q}\right) = 1,$$

which gives

$$k = \frac{pc^{p/q} \Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2} \Gamma\left(1 + \frac{p}{q}\right)}. \quad (223)$$

Thus, the density of $\mathbf{X}$ is given as

$$\begin{aligned}\phi(\mathbf{x}) &= f(\mathbf{z}_{(N)}^\top \mathbf{z}_{(N)}) \cdot |J_{\mathbf{z}_{(N)}}(\mathbf{x})| \\ &= f\left(\left(\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)^\top \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \cdot |J_{\mathbf{z}_{(N)}}(\mathbf{x})| \\ &= \frac{k}{|\Sigma|^{1/2}} \exp\left\{-c \left|(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right|^{q/2}\right\} \\ &= \frac{k}{|\Sigma|^{1/2}} \exp\left\{-c |Q(\mathbf{x})|^{q/2}\right\},\end{aligned} \quad (224)$$

where

$$k = \frac{pc^{p/q} \Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2} \Gamma\left(1 + \frac{p}{q}\right)}$$

and

$$Q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}). \quad (225)$$

Thus,

$$\begin{aligned}\phi(\mathbf{x}|\boldsymbol{\mu}, \Sigma, q) &= \frac{pc^{p/q} \Gamma\left(\frac{p}{2}\right)}{2|\Sigma|^{1/2} \pi^{p/2} \Gamma\left(1 + \frac{p}{q}\right)} \\ &\times \exp\left\{-c \left|(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right|^{q/2}\right\}.\end{aligned} \quad (226)$$

Now, it remains to determine an expression for c which requires the determination of the expectation and the variance-covariance of $\mathbf{X}$. These are obtained through the stochastic representation of $\mathbf{X}$.

Properties of the AMEPD

An important information for obtaining the desired expression for c in Equation (226) is the moment of the generating variate, R , seen in the stochastic decomposition of $\mathbf{X}$. This subsection is therefore focused on determining these moments which will also be useful in establishing the characteristic function of $\mathbf{X}$ and its mean and variance-covariance matrix. Other properties such as the skewness and kurtosis will then be determined in this subsection.

The Characteristic Function of $\mathbf{X}$

If the multivariate EPD random vector can be decomposed as in Equation (215), then the variable has a stochastic representation of the form

$$\mathbf{x} = R\mathbf{A}^\top \mathbf{U}^{(p)} + \boldsymbol{\mu}, \quad (227)$$

where, $R \geq 0$ is the generating variate, an absolute continuous non-negative random variable and is independent of $\mathbf{U}^{(p)}$ which is the uniform base of the EPD and $\mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$.

According to Fang and Zhang (1990), the random vector $\mathbf{x}$ is said to have an elliptical contour distribution if and only if R has a density, $h_R(r)$ and the

relationship between $f(\mathbf{x})$ and $h_R(r)$ is such that

$$h_R(r) = \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} r^{p-1} f(r^2), \quad (228)$$

where $\frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})}$ is the surface area of the unit sphere in $\Re^p$.

Assume that for $\mathbf{X}$, there is the density $f(\mathbf{x}^\top \mathbf{x})$ such that if $h(\cdot)$ is any non-negative Borel function, then

$$\begin{aligned} \int_{\Re^p} f(\sum x_i^2) dx_1, \dots, dx_p &= \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} \int_0^\infty y^{p-1} f(y) dy \\ &= \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} I_1(f|p/2) \\ \int_{\Re^p} f(\mathbf{x}^\top \mathbf{x}) d\mathbf{X} &= \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} \int_0^\infty y^{\frac{1}{2}p-1} f(y) dy = 1. \end{aligned}$$

Thus, a function $f(\cdot)$ can be used to define a density of a spherical distribution if and only if $\int_0^\infty y^{\frac{1}{2}p-1} f(y) dy < \infty$. Then,

$$E[h_R(r)] = \int_{\Re^p} h[(\mathbf{x}^\top \mathbf{x})^{1/2}] f(\mathbf{x}^\top \mathbf{x}) d\mathbf{x}.$$

For $\mathbf{x}^\top \mathbf{x} = r$, we have

$$\begin{aligned} E[h_R(r)] &= \int_{\Re^p} h[(\mathbf{x}^\top \mathbf{x})^{1/2}] f(\mathbf{x}^\top \mathbf{x}) d\mathbf{x} \\ &= \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} \int_0^\infty h(r^{1/2}) r^{p-1} f(r^2) dr. \end{aligned}$$

Let $r = y^{1/2}$. We have

$$E[h_R(r)] = \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} \int_0^\infty h(y) y^{1/2p-1} f(y) dy. \quad (229)$$

This shows that R has a density, $h_R(r)$ of the form

$$h_R(r) = \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{p}{2})} r^{p-1} f(r^2).$$

Using the density function of $\mathbf{X}$ given in Equation (226), the density function of the generating variate is obtained as

$$h_R(r) = \frac{2\pi^{\frac{1}{2}p}}{\Gamma\left(\frac{p}{2}\right)} r^{p-1} \frac{pc^{p/q}\Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2}\Gamma\left(1+\frac{p}{q}\right)} \exp\{-cr^q\}. \quad (230)$$

Thus,

$$h_R(r) = \frac{pc^{p/q}}{\Gamma\left(1+\frac{p}{q}\right)} r^{p-1} \exp\{-cr^q\}. \quad (231)$$

The moments of the generating variate, R is given as the s th moment of R given by

$$\begin{aligned} E[R^s] &= \int_{-\infty}^{\infty} r^s h_R(r) dr \\ &= \int_{-\infty}^{\infty} r^s \frac{pc^{p/q}}{\Gamma\left(1+\frac{p}{q}\right)} r^{p-1} \exp\{-cr^q\} dr \\ &= \frac{pc^{p/q}}{\Gamma\left(1+\frac{p}{q}\right)} \int_{-\infty}^{\infty} r^{s+p-1} \exp\{-cr^q\} dr. \end{aligned}$$

Now, we let $y = cr^q$, so that $c^{-1}y = r^q$. Thus, $r = c^{-1/q}y^{1/q}$,

$r^{s+p-1} = c^{-\frac{s+p-1}{q}} y^{\frac{s+p-1}{q}}$ and $\frac{dr}{dy} = \frac{1}{q} c^{-1/q} y^{(1/q)-1}$. The moment's expression becomes

$$\begin{aligned} E[R^s] &= \frac{pc^{p/q} c^{-\frac{s+p-1}{q}} c^{-\frac{1}{q}}}{\Gamma\left(1+\frac{p}{q}\right)} \frac{1}{q} \int_0^{\infty} y^{\frac{s+p-1}{q}} y^{1/q-1} e^{-y} dy \\ &= \frac{pc^{\frac{p}{q}-\frac{s+p-1}{q}-\frac{1}{q}}}{q\Gamma\left(1+\frac{p}{q}\right)} \int_0^{\infty} y^{\frac{s+p-1}{q}-\frac{1}{q}-1} e^{-y} dy \\ &= \frac{p}{q} \frac{c^{-\frac{s}{q}}}{\Gamma\left(1+\frac{p}{q}\right)} \int_0^{\infty} y^{\frac{s+p}{q}-1} e^{-y} dy, \end{aligned} \quad (232)$$

which simplifies as

$$E[R^s] = \frac{p}{q} \frac{c^{-\frac{s}{q}}}{\Gamma\left(1+\frac{p}{q}\right)} \left[\Gamma\left(\frac{s+p}{q}\right) \right].$$

Thus,

$$E(R^s) = \frac{\Gamma\left(\frac{p+s}{q}\right)}{\left[c^{\frac{s}{q}} \Gamma\left(\frac{p}{q}\right)\right]}. \quad (233)$$

Let $\mathbf{U}^{(p)}$ denote a random vector distributed uniformly on a unit sphere in $\Re^p$.

The characteristic function of $\mathbf{U}^{(p)}$ is

$$\psi_{\mathbf{U}^{(p)}}(\mathbf{t}) = \psi_p(\mathbf{t}) = \psi_p(\sqrt{\mathbf{t}^\top \mathbf{t}}). \quad (234)$$

If S_p denotes the surface area of the unit sphere in $\Re^p$, then

$$S_p = \frac{2\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)}. \quad (235)$$

Thus,

$$\phi_{\mathbf{U}^{(p)}}(\mathbf{t}) = \frac{1}{S_p} \int \dots \int_{s: \mathbf{X}^\top \mathbf{X}=1} e^{i\mathbf{t}^\top \mathbf{x}} ds \quad (236)$$

$$= \frac{\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}} \frac{2\pi^{\frac{1}{2}(p-1)}}{\Gamma\left(\frac{1}{2}(p-1)\right)} \int_{-1}^1 e^{i\mathbf{t}^\top \mathbf{x}} (1-x^2)^{\frac{1}{2}(p-3)} dx.$$

Taking $\mathbf{x} = \cos \theta$, we have

$$\phi_{\mathbf{U}^{(p)}}(t) = \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{1}{2}(p-1)\right)} \int_{-1}^1 e^{i\mathbf{t}^\top \cos \theta} \sin^{p-2} \theta d\theta = \psi_p(\sqrt{\mathbf{t}^\top \mathbf{t}}) \quad (237)$$

We can have the characteristic function of $\mathbf{X}$ as

$$\phi_x(\mathbf{t}) = E[\exp\{i\mathbf{t}^\top \mathbf{X}\}] \quad (238)$$

$$= E\left[\exp\left\{i\mathbf{t}^\top \left(\mu + \mathbf{A}^\top R \mathbf{U}^{(p)}\right)\right\}\right] \\ = e^{i\mathbf{t}^\top \mu} \int_0^\infty \psi_p(r\sqrt{\mathbf{t}^\top \Sigma \mathbf{t}}) h_R(r) dr. \quad (239)$$

Using the characteristic function of the uniform distribution of $\mathbf{U}^{(p)}$ and the distribution of R , we have the characteristic function of $\mathbf{X}$ as

$$\phi_x(\mathbf{t}) = e^{i\mathbf{t}^\top \mu} \frac{pc^{p/q}}{\Gamma\left(1 + \frac{p}{q}\right)} \int_0^\infty \psi_p(r\sqrt{\mathbf{t}^\top \Sigma \mathbf{t}}) r^{p-1} \exp\{-cr^q\} dr. \quad (240)$$

The Expectation and Variance-Covariance of $\mathbf{X}$

The expectation of $\mathbf{X}$ is given as

$$\begin{aligned} E(\mathbf{x}) &= E[\mu + R\mathbf{A}^\top \mathbf{U}^{(p)}] \\ &= E(\mu) + E(R)\mathbf{A}^\top E(\mathbf{U}^{(p)}). \end{aligned} \quad (241)$$

Since $E(\mathbf{U}^{(p)}) = 0$,

$$E(\mathbf{x}) = \mu. \quad (242)$$

The variance-covariance of the AMEPD random vector is given as

$$\begin{aligned} D(\mathbf{x}) &= E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^\top] \\ &= E[(\mu + R\mathbf{A}^\top \mathbf{U}^{(p)} - \mu)(\mu + R\mathbf{A}^\top \mathbf{U}^{(p)} - \mu)^\top] \\ &= E(R^2)\mathbf{A}^\top D(\mathbf{U}^{(p)})\mathbf{A} \end{aligned}$$

Since $D(\mathbf{U}^{(p)}) = \frac{1}{p}I_p$,

$$\begin{aligned} D(\mathbf{x}) &= E(R^2)\mathbf{A}^\top \frac{1}{p}I_p\mathbf{A} \\ &= \frac{1}{p}E(R^2)\Sigma \end{aligned}$$

Using the moment of the generating variate, R , we have

$$E(R^2) = \frac{\Gamma\left(\frac{p+2}{q}\right)}{c^{\frac{2}{q}}\Gamma\left(\frac{p}{q}\right)}.$$

Thus,

$$D(\mathbf{x}) = \frac{\Gamma\left(\frac{p+2}{q}\right)}{pc^{\frac{2}{q}}\Gamma\left(\frac{p}{q}\right)}\Sigma \quad (243)$$

From Equation (243) it can be observed that the $D(\mathbf{x})$ is affected by the shape parameter, q , and the constant, c . In order to remove this influence, the

expression for c is determined such that $D(\mathbf{x}) = \Sigma$. Therefore, after some simplification,

$$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[p \Gamma\left(\frac{p}{q}\right) \right]^{-\frac{q}{2}}. \quad (244)$$

Skewness and Kurtosis of $\mathbf{X}$

The measure of lack of symmetry of $\mathbf{X}$ is given by the skewness as

$$\gamma_1(\mathbf{x}) = E \left[\left((\mathbf{x} - \mu)^\top (D(\mathbf{x}))^{-1} (\mathbf{y} - \mu) \right)^3 \right], \quad (245)$$

where $\mathbf{y}$ is a random vector independent and equally distributed as $\mathbf{x}$. Thus,

$$\begin{aligned} \gamma_1(\mathbf{x}) &= E \left[\left((\mu + \mathbf{R}\mathbf{A}^\top \mathbf{U}^{(p)} - \mu)^\top (D(\mathbf{x}))^{-1} (\mu + \mathbf{W}\mathbf{A}^\top \mathbf{V}^{(p)} - \mu) \right)^3 \right] \\ &= E \left[\left((\mathbf{R}\mathbf{A}^\top \mathbf{U}^{(p)})^\top \Sigma^{-1} (\mathbf{W}\mathbf{A}^\top \mathbf{V}^{(p)}) \right)^3 \right] \\ &= E \left[(R\mathbf{W})^3 \right] E \left[\left(\mathbf{U}^{(p)\top} \mathbf{V}^{(p)} \right)^3 \right] \end{aligned} \quad (246)$$

$$\text{Since } E \left[\left(\mathbf{U}^{(p)\top} \mathbf{V}^{(p)} \right)^3 \right] = 0,$$

$$\gamma_1(\mathbf{X}) = 0 : \quad (247)$$

The coefficient of kurtosis is given by

$$\gamma_2(\mathbf{x}) = E \left[(Q(\mathbf{x}))^2 \right] - p(p+2), \quad (248)$$

where $Q(\mathbf{x})$ is as defined in Equation (225), and the term $p(p+2)$ is equivalent to $E \left[(Q(\mathbf{x}))^2 \right]$ when $\mathbf{X}$ is distributed as multivariate Normal. Therefore,

$$\begin{aligned} E \left[(Q(\mathbf{x}))^2 \right] &= E \left[\left((\mathbf{x} - \mu)^\top D(\mathbf{x})^{-1} (\mathbf{x} - \mu) \right)^2 \right] \\ &= E \left[\left((\mathbf{R}\mathbf{A}^\top \mathbf{U}^{(p)})^\top D(\mathbf{x})^{-1} (\mathbf{R}\mathbf{A}^\top \mathbf{U}^{(p)}) \right)^2 \right] \\ &= E \left[(R^4) (\mathbf{U}^{(p)\top} \Sigma \Sigma^{-1} \mathbf{U}^{(p)})^2 \right] \\ &= E \left[(R^4) \right] E \left[(\mathbf{U}^{(p)\top} \mathbf{U}^{(p)})^2 \right]. \end{aligned} \quad (249)$$

Now, $E \left[(\mathbf{U}^{(p)\top} \mathbf{U}^{(p)})^2 \right] = 1$, and from Equation (233), we have

$$E \left[(Q(\mathbf{x}))^2 \right] = \frac{\Gamma\left(\frac{p+4}{q}\right)}{\left[c^{\frac{4}{q}} \Gamma\left(\frac{p}{q}\right)\right]}$$

Therefore, the coefficient of kurtosis of $\mathbf{X}$ is

$$\gamma_2(\mathbf{x}) = \frac{\Gamma\left(\frac{p+4}{q}\right)}{\left[c^{\frac{4}{q}} \Gamma\left(\frac{p}{q}\right)\right]} - p(p+2). \quad (250)$$

Preposition. The density of the AMEPD and its characteristics are summarized in the following proposition.

Let $\mathbf{x} \sim \text{MEPD}(\mu, \Sigma, q)$, then for a non-constant value, c , the symmetric distribution of $\mathbf{x}$ given by

$$\phi(\mathbf{x}|\mu, \Sigma, q) = \frac{k}{|\Sigma|^{1/2}} \exp \left\{ -c \left| (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right|^{q/2} \right\}, \quad (251)$$

is well regulated in the tails such that

$$k = \frac{p c^{p/q} \Gamma\left(\frac{p}{2}\right)}{2 \pi^{p/2} \Gamma\left(1 + \frac{p}{q}\right)},$$

$$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[p \Gamma\left(\frac{p}{q}\right) \right]^{-\frac{q}{2}},$$

with

$$E(\mathbf{x}) = \mu, \quad D(\mathbf{x}) = \Sigma, \quad \gamma_1(\mathbf{x}) = 0, \quad \text{and}$$

$$\gamma_2(\mathbf{x}) = \frac{\Gamma\left(\frac{p+4}{q}\right)}{c^{\frac{4}{q}} \Gamma\left(\frac{p}{q}\right)} - p(p+2).$$

The AMEPD in Equation (251) now provides a generalized expression, c , for the coefficient of the quadratic form which is a function of the shape parameter, q , in a manner that is expected to regulate the tails of the distribution.

Deductions from the AMEPD

1. The AMEPD is a generalization of the MEPD of Gomez et al (1998). It can be deduced that $q = 2\beta$ and if $c = 1/2$, the AMEPD becomes MEPD in Equation (2).
2. The AMEPD is clearly of the form of the EPD of Olosunde (2013). The expression for c may be related to $h(w)$ and $q = w$ in Equation (7).

Bivariate Plots of the AMEPD

For the purposes of generating graphs, we considered the bivariate of AMEPD of the form

$$f(x_1, x_2) = \frac{k}{\sigma_1 \sigma_2 \sqrt{(1 - \rho^2)}} \times \exp \left\{ -c \left[\frac{1}{1 - \rho^2} \left((a_1)^2 - 2\rho(a_1)(a_2) + (a_2)^2 \right) \right]^{\frac{q}{2}} \right\} \quad (252)$$

and bivariate of MEPD of the form in Equation (252) (when $q = 2\beta$ and $c = 1/2$), where $a_1 = \frac{x_1 - \mu_1}{\sigma_1}$ and $a_2 = \frac{x_2 - \mu_2}{\sigma_2}$ for all x_1 and x_2 . The μ_1, μ_2 may be any real numbers, $\sigma_1 > 0, \sigma_2 > 0$ and $-1 \leq \rho \leq 1$. Some bivariate plot for $\mu = (0, 0)^T$ and $\Sigma = \mathbf{I}_2$ for some values of q have been plotted in Figures 14 to 18 along with their respective contours. R software is used to plot the graphs using the code in Appendix D.

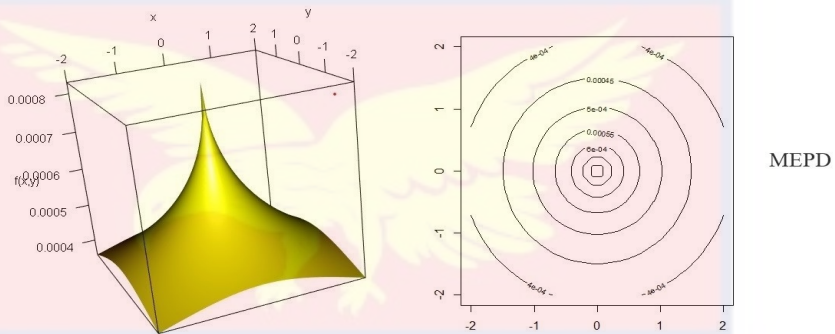
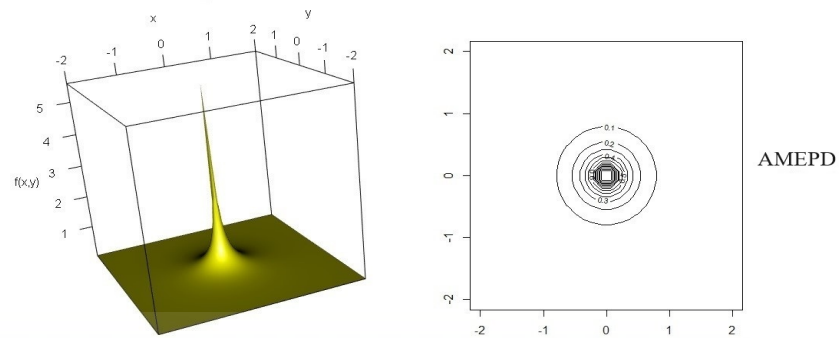


Figure 14: Multivariate EPD for $q = 0.5$.

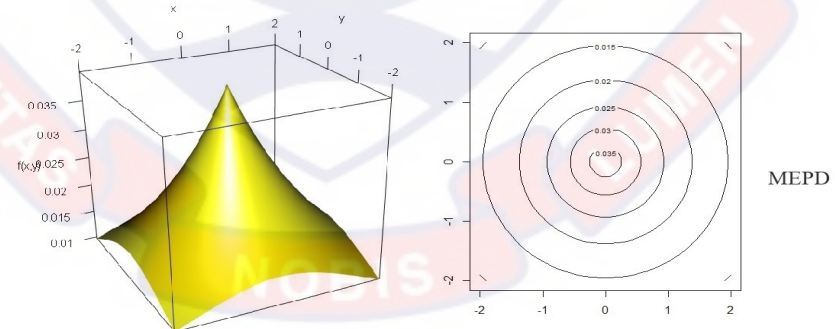
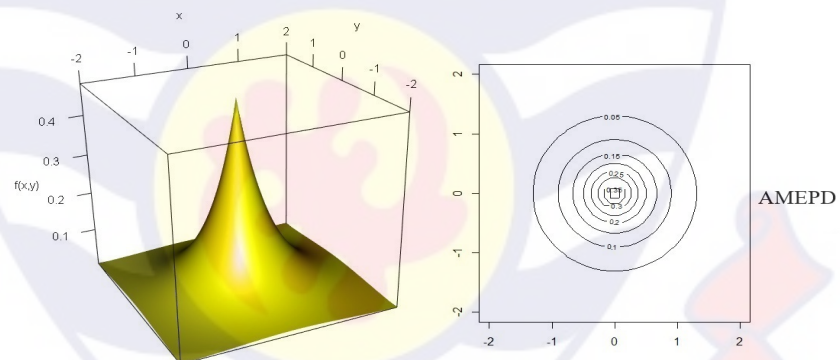


Figure 15: Multivariate EPD for $q = 1$.

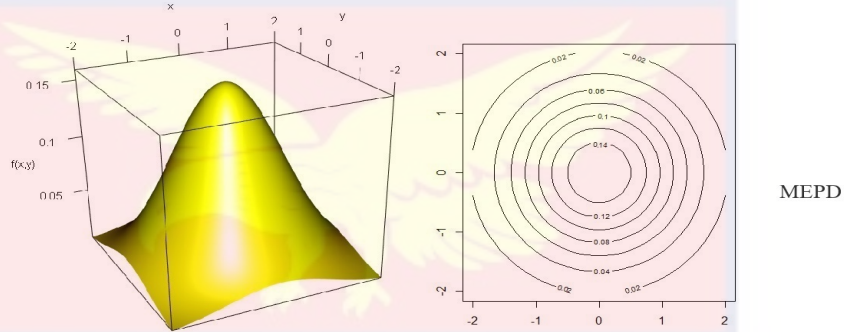
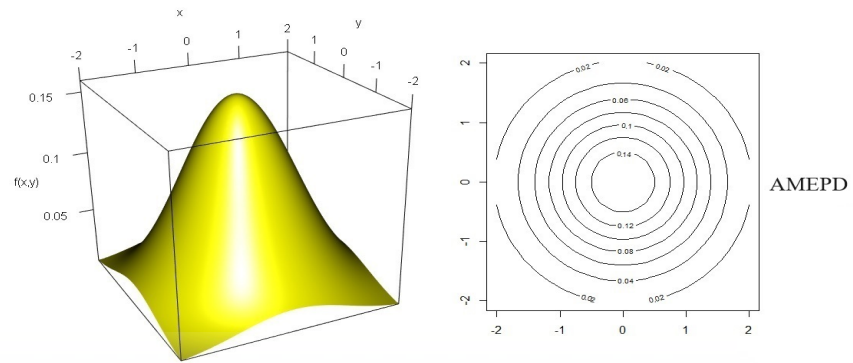


Figure 16: Multivariate EPD for $q = 2$.

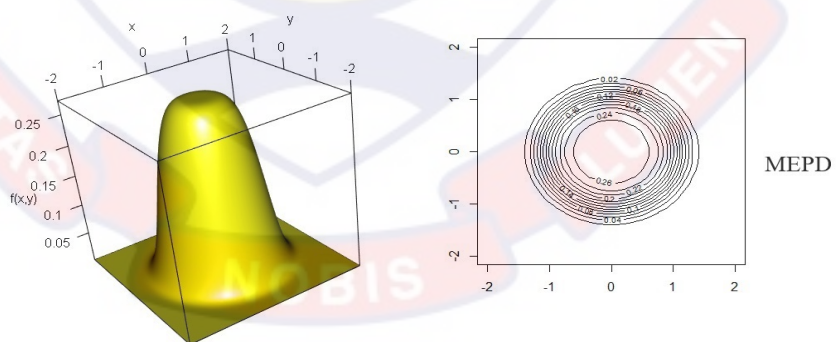
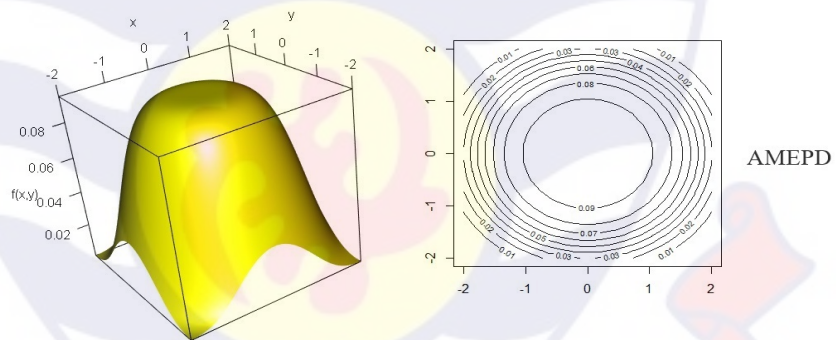


Figure 17: Multivariate EPD for $q = 5$.

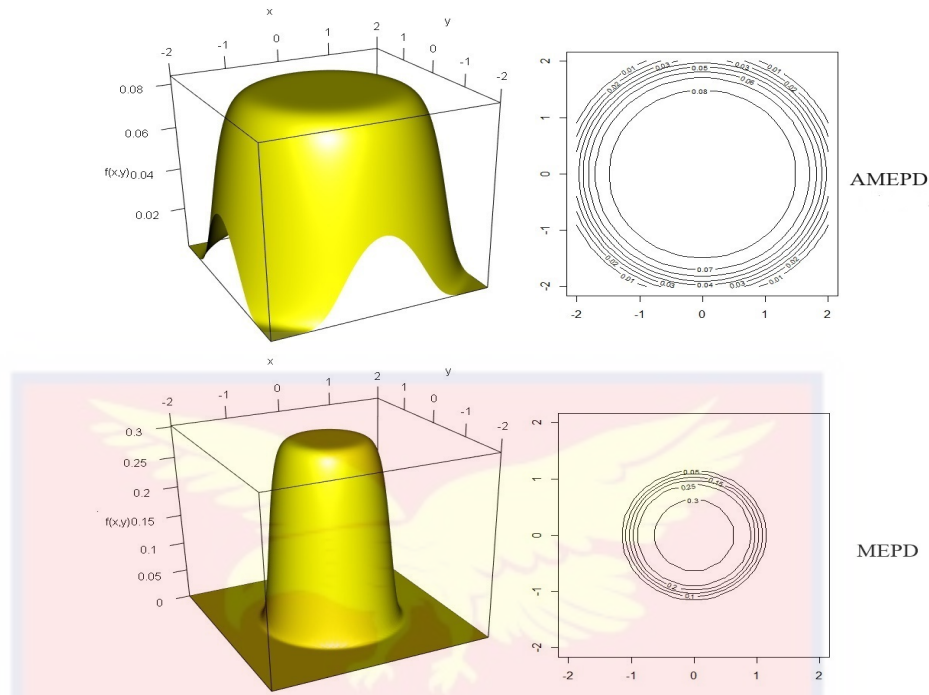


Figure 18: Multivariate EPD for $q = 10$.

The plots show that as the value of q increases, the shape of the function moves from leptokurtic to platykurtic and assumes the Normal shape as that of the MEPD for $q = 2$. In all cases, the AMEPD exhibit clearer flexibility in the tails of the distribution with more appropriate density at the centre of the distribution that reflex a more reasonable peakedness. Thus, the illustrations indicate that the AMEPD provides a more suitable fit.

Simulation Results

We generate random samples of varying sizes ($n=150, 200, 500, 1,000, 10,000, 100,000$) from two distinct populations: one following the AMEPD and the other from the distribution described by Gomez et al. (1998). The purpose is to evaluate the effectiveness of our distribution. In both cases, the random samples are generated with a dispersion matrix of $\Sigma = \mathbf{I}_2$ and a mean vector of

$\mu = (0,0)^T$. The generation process involves utilizing Equation (231), where r^q (of the generating variate, R) follows a distribution of $\Gamma\left(\frac{1}{c}, \frac{p}{q}\right)$. The shape parameter q takes on various values. The computations is carried out using the codes in Appendix B with R software. The comprehensive results of the simulation are presented in Tables 4 to 8. These tables provide detailed insights into the performance of the distribution under different sample sizes and shape parameter values, allowing a thorough assessment of its effectiveness.

Table 4: Summary Statistics for Simulated Data for $q = 0.5$

Sample Size	Mean(X1)	Mean(X2)	Var(X1)	Var(X2)	Cov(X1,X2)
$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[\Gamma\left(\frac{p}{q}\right) p \right]^{-\frac{q}{2}} = 4.527019$					
150	0.008	-0.005	0.772	0.632	-0.031
200	0.070	0.004	0.682	0.767	0.101
500	-0.023	-0.040	0.975	0.862	-0.027
1,000	-0.043	-0.042	0.864	0.933	-0.003
10,000	0.001	0.004	0.995	1.012	0.039
100,000	-0.001	-0.004	0.989	1.009	0.001
$c = 0.5$					
150	0.632	-0.433	5189.212	4244.901	-207.832
200	5.724	0.310	4583.973	5155.501	675.690
500	-1.895	-3.289	6551.130	5791.902	-179.003
1,000	-3.550	-3.413	5807.259	6268.728	-16.813
10,000	0.097	0.355	6685.611	6802.496	265.034
100,000	-0.043	-0.335	6648.954	6781.887	8.281

Source: Field survey, Manford (2023)

Table 5: Summary Statistics for Simulated Data for $q = 1$

Sample Size	Mean(X1)	Mean(X2)	Var(X1)	Var(X2)	Cov(X1,X2)
$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[\Gamma\left(\frac{p}{q}\right) p \right]^{-\frac{q}{2}} = 1.732051$					
150	0.017	0.025	0.893	0.734	-0.022
200	0.024	0.040	0.838	0.844	-0.008
500	-0.023	-0.017	0.956	0.950	-0.089
1,000	-0.075	-0.033	0.996	1.012	0.015
10,000	0.002	0.021	1.040	0.994	0.011
100,000	-0.003	0.000	1.008	0.987	0.000
$c = 0.5$					
150	0.060	0.087	10.716	8.810	-0.267
200	0.083	0.138	10.058	10.123	-0.091
500	-0.078	-0.060	11.466	11.400	-1.068
1,000	-0.261	-0.115	11.946	12.139	0.177
10,000	0.007	0.071	12.485	11.922	0.136
100,000	-0.012	-0.001	12.092	11.840	-0.002

Source: Field survey, Manford (2023)

Table 6: Summary Statistics for Simulated Data for $q = 2$

Sample Size	Mean(X1)	Mean(X2)	Var(X1)	Var(X2)	Cov(X1,X2)
$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[\Gamma\left(\frac{p}{q}\right) p \right]^{-\frac{q}{2}} = 0.5$					
150	-0.014	0.0437	1.109	0.847	0.117
200	0.098	-0.0475	1.041	1.005	-0.108
500	-0.010	0.0057	1.023	0.984	0.008
1,000	-0.027	0.0310	0.930	1.044	-0.022
10,000	-0.004	0.0174	1.026	0.996	0.011
100,000	-0.003	-0.0007	1.003	1.000	0.001
$c = 0.5$					
150	-0.014	0.0437	1.109	0.847	0.117
200	0.098	-0.0475	1.041	1.005	-0.108
500	-0.010	0.0057	1.023	0.984	0.008
1,000	-0.027	0.0310	0.930	1.044	-0.022
10,000	-0.004	0.0174	1.026	0.996	0.011
100,000	-0.003	-0.0007	1.003	1.000	0.001

Source: Field survey, Manford (2023)

Table 7: Summary Statistics for Simulated Data for $q = 5$

Sample Size	Mean(X1)	Mean(X2)	Var(X1)	Var(X2)	Cov(X1,X2)
$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[\Gamma\left(\frac{p}{q}\right) p \right]^{-\frac{q}{2}} = 0.03528091$					
150	0.038	0.053	1.074	0.953	0.064
200	0.156	-0.060	1.001	1.026	-0.037
500	0.000	-0.017	1.044	1.002	0.062
1,000	-0.045	-0.036	1.023	0.998	0.013
10,000	-0.011	-0.004	0.979	1.012	-0.007
100,000	-0.001	0.000	1.002	1.000	0.006
$c = 0.5$					
150	0.022	0.031	0.372	0.330	0.022
200	0.092	-0.035	0.347	0.355	-0.013
500	0.000	-0.010	0.362	0.347	0.021
1,000	-0.027	-0.021	0.354	0.346	0.004
10,000	-0.006	-0.003	0.339	0.350	-0.002
100,000	0.000	0.000	0.347	0.346	0.002

Source: Field survey, Manford (2023)

Table 8: Summary Statistics for Simulated Data for $q = 10$

Sample Size	Mean(X1)	Mean(X2)	Var(X1)	Var(X2)	Cov(X1,X2)
$c = \left[\Gamma\left(\frac{p+2}{q}\right) \right]^{\frac{q}{2}} \left[\Gamma\left(\frac{p}{q}\right) p \right]^{-\frac{q}{2}} = 0.0008229082$					
150	0.038	0.004	1.075	0.981	0.021
200	0.165	0.007	0.980	1.082	0.051
500	0.013	-0.004	1.010	1.054	-0.022
1,000	-0.072	-0.037	0.981	1.044	-0.004
10,000	0.000	0.022	1.024	0.978	0.012
100,000	0.000	0.000	1.000	1.002	-0.002
$c = 0.5$					
150	0.020	0.002	0.298	0.272	0.006
200	0.087	0.004	0.272	0.300	0.014
500	0.007	-0.002	0.280	0.292	-0.006
1,000	-0.038	-0.020	0.272	0.290	-0.001
10,000	0.000	0.011	0.284	0.272	0.003
100,000	0.000	0.000	0.277	0.278	0.000

Source: Field survey, Manford (2023)

The results show that for all specified values of q , and the corresponding derived value of c , the simulated data possess almost the same characteristics as the underlying distribution for the AMEPD for all sample sizes. The performance of the existing MEPD is comparable to that of the new one only when $q = 2$ for which both distributions reduce to the multivariate Normal Distribution and the coefficient $c = 0.5$. The performance of the density function of Gomez, et al. (1998) is rather highly influenced by the value of q and could depict a widely spread-out or dense distribution of the data for the choice of $q < 2$ or $q > 2$, respectively.

Real Data Example

In this subsection, we utilize the carsales sample dataset for the analysis. This example focusses on six (6) distinct variables, which are taken from the dataset: Price, Horsepower (Hpower), Wheelbase (Wbase), Curb Weight (C.Weight), Fuel Capacity (F.Cap), and Fuel Efficiency (F. Eff) (Manford et al., 2024). Similar datasets generated from the AMEPD and MEPD distributions are produced using the data's mean vector and variance-covariance matrix. Summary measures are calculated to assess the data recovery as shown in Tables 9 and 11 for AMEPD and MEPD, respectively. To provide a comprehensive overview, we present relevant summary statistics (SS) of the dataset in the Tables. These tables encapsulates key statistical measures that offer insights into the distribution and characteristics of the selected variables, contributing to a better understanding of the dataset's overall structure.

Table 9: Summary Statistics Data Generated for Price and Specification of Cars Using AMEPD

$q = 0.5$	Price	Hpower	Wbase	C.Weight	F.Cap	F.Eff
Mean	27.09	186.29	106.94	3.32	17.67	24.07
SD	13.72	53.40	7.41	0.57	3.71	4.10
Min.	-21.72	0.97	79.23	1.41	3.14	9.58
1st Qu.	18.84	155.41	102.14	2.97	15.53	21.93
Median	27.52	189.55	106.97	3.36	17.78	23.90
3rd Qu.	35.49	215.74	110.70	3.66	19.82	26.48
Max.	65.88	368.24	126.47	4.92	31.83	35.67
$q = 2$						
Mean	24.74	176.46	107.27	3.29	17.63	24.35
SD	14.83	57.23	7.54	0.66	3.91	4.42
Min.	-8.03	51.82	86.78	2.06	6.42	10.60
1st Qu.	14.39	132.75	101.66	2.81	14.98	21.53
Median	26.08	175.77	107.69	3.27	17.40	24.16
3rd Qu.	34.21	215.74	113.08	3.67	20.04	27.23
Max.	72.34	310.01	126.08	5.29	28.59	34.13
$q = 5$						
Mean	25.72	178.33	107.05	3.29	17.47	24.31
SD	14.79	60.05	7.78	0.66	4.13	4.49
Min.	-11.05	39.52	88.87	1.83	8.57	12.34
1st Qu.	14.90	138.20	101.54	2.84	14.54	21.24
Median	25.24	173.54	106.85	3.27	17.26	24.40
3rd Qu.	35.91	216.87	113.17	3.70	20.13	27.53
Max.	62.57	318.39	126.13	5.04	27.19	35.00
$q = 10$						
Mean	26.84	179.51	106.70	3.32	17.46	24.23
SD	14.33	59.34	7.94	0.63	4.02	4.28
Min.	-9.67	54.96	91.40	1.85	8.83	15.42
1st Qu.	17.58	136.46	100.70	2.91	14.34	20.82
Median	26.07	174.83	106.30	3.33	17.39	24.71
3rd Qu.	36.40	223.15	113.30	3.72	20.22	27.69
Max.	56.20	333.01	125.80	4.72	26.01	33.54

Source: Field survey, Manford (2023)

Table 10: Summary Statistics Data Generated for Price and Specification of Cars Using MEPD

$q = 0.5$	Price	Hpower	Wbase	C.Weight	F.Cap	F.Eff
Mean	-36.99	563.60	-34.20	-15.87	-66.80	87.28
SD	4054.65	15784.28	2190.27	167.40	1096.41	1212.74
Min.	-14465.12	-46229.90	-8225.73	-578.43	-4363.15	-4194.29
1st Qu.	-2476.50	-8562.90	-1452.17	-118.17	-699.59	-543.42
Median	90.21	1526.50	-25.81	-2.08	-36.47	38.90
3rd Qu.	2445.14	9267.90	1076.33	86.60	566.97	801.06
Max.	11427.72	54341.90	5737.63	457.72	4117.28	3515.93
$q = 2$						
Mean	24.74	176.46	107.27	3.29	17.63	24.35
SD	14.83	57.23	7.54	0.66	3.91	4.42
Min.	-8.03	51.82	86.78	2.06	6.42	10.60
1st Qu.	14.39	132.75	101.66	2.81	14.98	21.53
Median	26.08	175.77	107.69	3.27	17.40	24.16
3rd Qu.	34.21	215.74	113.08	3.67	20.04	27.23
Max.	72.34	310.01	126.08	5.29	28.59	34.13
$q = 5$						
Mean	26.57	181.90	107.25	3.34	17.73	24.06
SD	6.84	27.78	3.60	0.30	1.91	2.08
Min.	9.57	117.70	98.84	2.66	13.62	18.53
1st Qu.	21.57	163.40	104.70	3.13	16.38	22.64
Median	26.35	179.70	107.16	3.33	17.63	24.11
3rd Qu.	31.29	199.80	110.08	3.53	18.96	25.55
Max.	43.62	246.70	116.07	4.15	22.23	29.01
$q = 10$						
Mean	27.13	182.90	107.20	3.36	17.76	24.00
SD	5.54	22.96	3.07	0.24	1.55	1.66
Min.	13.00	134.70	101.20	2.79	14.43	20.59
1st Qu.	23.54	166.20	104.80	3.20	16.56	22.68
Median	26.83	181.10	107.00	3.36	17.74	24.18
3rd Qu.	30.83	199.80	109.70	3.51	18.83	25.33
Max.	38.49	242.30	114.50	3.90	21.07	27.60

Source: Field survey, Manford (2023)

Consistent with the earlier assessment, from the Tables, the AMEPD obtains all features of the generated data more closely to that of the original data for all values of the specified shape parameter, q . The results also show that for small values of q (< 1), the existing MEPD provides a very poor fit of the underlining distribution. Clearly, when $q = 2$, the two distributions are the same.

Maximum Likelihood Estimation of the Parameters of the AMEPD

Assume that the random vector $\mathbf{x}$ has an absolutely continuous generating variate R and is elliptically symmetric distributed with a positive definite dispersion matrix Σ . Equation (251))'s density function for $\mathbf{x}$ corresponds to

$$f_{\mathbf{M},m,q}(\mathbf{x}) = |\mathbf{M}|^{-\frac{1}{2}} h_{m,q}(r) \quad (253)$$

where

$$r = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{M}^{-1} (\mathbf{x} - \boldsymbol{\mu}), \quad (254)$$

for any $\mathbf{x} \in \Re^p$, where $\mathbf{M}$ is $p \times p$ symmetric real scatter matrix and $h_{m,q}(r)$ is a density generator defined by

$$h_{m,q}(r) = km^{-\frac{p}{2}} \exp \left\{ -c \frac{r^{\frac{q}{2}}}{m^{\frac{q}{2}}} \right\} \quad (255)$$

where

$$k = \frac{qc^{\frac{p}{q}} \Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}} \Gamma\left(\frac{p}{q}\right)} \quad (256)$$

for $r \in \Re^+$, where m and q are the AMEPD scale and shape parameters, respectively. For any spherical distribution, the scatter or dispersion matrix Σ is given

as (Pascal, et al., 2013)

$$\Sigma = m\mathbf{M}. \quad (257)$$

However, the matrix $\mathbf{M}$ must be normalized such that $tr(\mathbf{M}) = p$ by the likelihood function of X given as

$$L(f_{\mathbf{M},m,q}(\mathbf{x})) = |\mathbf{M}|^{-\frac{n}{2}} \prod_{i=1}^n h_{m,q}((\mathbf{x}_i - \mu)^\top \mathbf{M}^{-1}(\mathbf{x}_i - \mu)) \quad (258)$$

for

$$\prod_{i=1}^n h_{m,q}(r) = k^n m^{-\frac{np}{2}} \exp \left\{ -c \sum_{i=1}^n \frac{r^{\frac{q}{2}}}{m^{\frac{q}{2}}} \right\}. \quad (259)$$

Let $l = \ln L$ denote the log-likelihood function of X . Then

$$l(f_{\mathbf{M},m,q}(\mathbf{x})) = -\frac{n}{2} \ln(|\mathbf{M}|) + \sum_{i=1}^n h_{m,q}((\mathbf{x}_i - \mu)^\top \mathbf{M}^{-1}(\mathbf{x}_i - \mu)) \quad (260)$$

MLE of μ

For maximum likelihood estimator (MLE) of μ , by taken the derivative of Equation (260) with respect to μ and equating to zero, yields

$$\begin{aligned} \frac{\partial}{\partial \mu} [l(f_{\mathbf{M},m,q}(\mathbf{x}))] &= -\mu \sum_{i=1}^n h'_{m,q}((\mathbf{x}_i - \mu)^\top \mathbf{M}^{-1}(\mathbf{x}_i - \mu)) + \\ &\sum_{i=1}^n h'_{m,q}((\mathbf{x}_i - \mu)^\top \mathbf{M}^{-1}(\mathbf{x}_i - \mu)) \mathbf{x}_i = 0 \end{aligned} \quad (261)$$

Thus,

$$\mu \sum_{i=1}^n h'_{m,q}((\mathbf{x}_i - \mu)^\top \mathbf{M}^{-1}(\mathbf{x}_i - \mu)) = \sum_{i=1}^n h'_{m,q}((\mathbf{x}_i - \mu)^\top \mathbf{M}^{-1}(\mathbf{x}_i - \mu)) \mathbf{x}_i$$

and hence,

$$\mu = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \bar{\mathbf{x}} \quad (262)$$

Thus, for the same data feature, the estimate of the mean would be the same for both AMEPD and MEPD since it is not a function of c .

MLE of Σ

To obtain the MLE of Σ as given in Equation (257), we find the MLE of both m and $\mathbf{M}$. The log-likelihood of $h_{m,q}(r)$ in Equation (255) is given as

$$l(h_{m,q}(r)) = n \ln(k) - \frac{np}{2} \ln(m) - cm^{-\frac{q}{2}} \sum_{i=1}^n r^{\frac{q}{2}} \quad (263)$$

For m , obtain the derivative of the log-likelihood function in Equation (263) as

$$\frac{\partial}{\partial m} [l(h_{m,q}(r))] = -\frac{np}{2} m^{-1} + \frac{cq}{2} m^{-\frac{q}{2}-1} \sum_{i=1}^n r^{\frac{q}{2}} = 0$$

Thus,

$$m = \left[\frac{cq}{np} \sum_{i=1}^n (r)^{\frac{q}{2}} \right]^{\frac{2}{q}} \quad (264)$$

Making substitution from Equation (255) will yield

$$m = \left[\frac{cq}{np} \sum_{i=1}^n ((\mathbf{x} - \mu)^{\top} \mathbf{M}^{-1} (\mathbf{x} - \mu))^{\frac{q}{2}} \right]^{\frac{2}{q}} \quad (265)$$

For $\mathbf{M}$, we must maximize the log-likelihood function of $\mathbf{X}$ by finding the derivative of $\ln l_{f_{\mathbf{M},m,q}(x)}$ with respect to $\mathbf{M}$ and solving for $\mathbf{M}$ by equating to zero.

This will lead to

$$\begin{aligned} \frac{\partial}{\partial \mathbf{M}} [l(f_{\mathbf{M},m,q}(\mathbf{x}))] &= -\frac{n}{2} \mathbf{M}^{-1} - \sum_{i=1}^n \frac{h'_{m,q}((\mathbf{x} - \mu)^{\top} \mathbf{M}^{-1} (\mathbf{x} - \mu))}{h_{m,q}((\mathbf{x} - \mu)^{\top} \mathbf{M}^{-1} (\mathbf{x} - \mu))} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^{\top} \mathbf{M}^{-2} \\ &= 0 \end{aligned}$$

Solving gives

$$\mathbf{M} = \frac{2}{n} \sum_{i=1}^n \frac{-h'_{m,q}(r)}{h_{m,q}(r)} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^{\top}. \quad (266)$$

Since

$$h'_{m,q}(r) = \frac{\partial}{\partial r} [h_{m,q}(r)] = km^{-\frac{p}{2}} \left(-c \frac{r^{\frac{q}{2}-1}}{m^{\frac{q}{2}}} \frac{q}{2} \right) \exp \left\{ -c \frac{r^{\frac{q}{2}}}{m^{\frac{q}{2}}} \right\},$$

Therefore,

$$\mathbf{M} = \frac{cq}{nm^{\frac{q}{2}}} \sum_{i=1}^n ((\mathbf{x} - \mu)^{\top} \mathbf{M}^{-1} (\mathbf{x} - \mu))^{\frac{q}{2}-1} (\mathbf{x} - \mu) (\mathbf{x} - \mu)^{\top}. \quad (267)$$

MLE of q

The log-likelihood of $f_{\mathbf{M},m,q}(\mathbf{x})$ as presented in Equation (260) is therefore given as

$$l(f_{\mathbf{M},m,q}(\mathbf{x})) = n \ln(k) - \frac{n}{2} \ln(|\mathbf{M}|) - \frac{np}{2} \ln(m) - cm^{-\frac{q}{2}} \sum_{i=1}^n r^{\frac{q}{2}} \quad (268)$$

For q , consider the derivative equation of the log-likelihood of X as

$$\frac{\partial}{\partial q} [l(f_{\mathbf{M},m,q}(\mathbf{x}))] = n \frac{\partial}{\partial q} [\ln(k)] - \frac{\partial}{\partial q} \left[cm^{-q/2} \sum_{i=1}^n r^{q/2} \right] = 0 \quad (269)$$

Further simplification yields

$$n \frac{\partial}{\partial q} [\ln(k)] - \left[m^{-\frac{q}{2}} \frac{\partial}{\partial q} (c) \right] \sum_{i=1}^n r^{\frac{q}{2}} + \left[c \frac{m^{-\frac{q}{2}}}{2} \ln(m) \right] \sum_{i=1}^n r^{\frac{q}{2}} - \frac{c}{2} m^{-\frac{q}{2}} \sum_{i=1}^n r^{\frac{q}{2}} \ln(r) = 0 \quad (270)$$

and subsequently,

$$\frac{\partial}{\partial q} [\ln(k)] = \frac{\partial}{\partial q} \ln \left[\frac{qc^{p/q} \Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2} \Gamma\left(\frac{p}{q}\right)} \right]$$

which gives

$$\frac{\partial}{\partial q} [\ln(k)] = \frac{1}{q} + \frac{p}{q} \left[\frac{\partial}{\partial q} \ln(c) \right] - \ln(c) \left[\frac{p}{q^2} \right] + \frac{p}{q^2} \left[\Psi\left(\frac{p}{q}\right) \right]. \quad (271)$$

where

$$\psi(\alpha) = \frac{d}{d\alpha} (\ln(\Gamma(\alpha))) \quad (272)$$

is the logarithmic derivative of $\Gamma(\alpha)$, and represent the digamma function $\psi(\alpha)$

and may be written as

$$\psi(\alpha) = \int_0^{\infty} \left(\frac{e^{-x}}{x} - \frac{e^{-\alpha x}}{1 - e^{-x}} \right) dx. \quad (273)$$

Now

$$\ln(c) = \frac{q}{2} \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] - \frac{q}{2} \ln \left[\Gamma \left(\frac{p}{q} \right) \right] - \frac{q}{2} \ln(p). \quad (274)$$

Thus, we have

$$\frac{p}{q^2} \ln(c) = \frac{p}{2q} \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] - \frac{p}{2q} \ln \left[p \Gamma \left(\frac{p}{q} \right) \right]. \quad (275)$$

From Equation (274),

$$\begin{aligned} \frac{\partial}{\partial q} [\ln(c)] &= \frac{1}{2} \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] - \frac{p+2}{2q} \Psi \left(\frac{p+2}{q} \right) - \frac{1}{2} \ln \left[p \Gamma \left(\frac{p}{q} \right) \right] \\ &\quad + \frac{p}{2q} \Psi \left(\frac{p}{q} \right). \end{aligned} \quad (276)$$

Substituting Equation (276) into Equation (271) yields

$$\frac{\partial}{\partial q} [\ln(k)] = \frac{n}{q} - \frac{np(p+2)}{2q^2} \Psi \left(\frac{p+2}{q} \right) + \frac{np^2}{2q^2} \Psi \left(\frac{p}{q} \right) + \frac{np}{q^2} \Psi \left(\frac{p}{q} \right) \quad (277)$$

The differential $\frac{\partial c}{\partial q}$ in Equation (270) will yield

$$\begin{aligned} \frac{\partial c}{\partial q} &= \frac{\partial}{\partial q} \left[\left[\Gamma \left(\frac{p+2}{q} \right) \right]^{\frac{q}{2}} \left[p \Gamma \left(\frac{p}{q} \right) \right]^{-\frac{q}{2}} \right] \\ &= \frac{1}{2} c \left\{ \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] - \frac{(p+2)}{q} \Psi \left(\frac{p+2}{q} \right) - \ln \left[p \Gamma \left(\frac{p}{q} \right) \right] + \frac{p}{q} \Psi \left(\frac{p}{q} \right) \right\} \end{aligned}$$

From Equation (269),

$$m^{-q/2} \left[\sum_{i=1}^n (r)^{\frac{q}{2}} \right] = \frac{np}{cq} \quad (279)$$

Thus, from Equations (278) and (279),

$$\begin{aligned} m^{-q/2} \left[\sum_{i=1}^n (r)^{\frac{q}{2}} \right] \frac{\partial c}{\partial q} &= \frac{np}{2q} \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] - \frac{np}{2q^2} (p+2) \Psi \left(\frac{p+2}{q} \right) \\ &\quad - \frac{np}{2q} \ln \left[p \Gamma \left(\frac{p}{q} \right) \right] + \frac{np^2}{2q^2} \Psi \left(\frac{p}{q} \right) \end{aligned} \quad (280)$$

Again,

$$\begin{aligned} \left[c \frac{m^{-q/2}}{2} \ln(m) \right] \sum_{i=1}^n r^{q/2} &= \frac{c}{2} \left[\sum_{i=1}^n r^{q/2} \right] \frac{np}{cq} \left[\sum_{i=1}^n r^{q/2} \right]^{-1} \frac{2}{q} \ln \left[\frac{cq}{np} \sum_{i=1}^n r^{q/2} \right] \\ &= \frac{np}{q^2} \ln \left[\frac{cq}{np} \sum_{i=1}^n r^{q/2} \right], \end{aligned} \quad (281)$$

and

$$cm^{-q/2} \sum_{i=1}^n r^{q/2} \ln(r) = \frac{np}{q} \left[\sum_{i=1}^n r^{q/2} \right]^{-1} \sum_{i=1}^n \left[r^{q/2} \ln(r) \right]. \quad (282)$$

Now, substituting Equations (277), (280), (281) and (282) into Equation (270),

we have

$$\begin{aligned} \frac{n}{q} + \frac{np}{q^2} \Psi \left(\frac{p}{q} \right) - \frac{np}{2q} \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] + \frac{np}{2q} \ln \left[p \Gamma \left(\frac{p}{q} \right) \right] \\ + \frac{np}{2q} \ln \left[\Gamma \left(\frac{p+2}{q} \right) \right] - \frac{np}{2q} \ln \left[p \Gamma \left(\frac{p}{q} \right) \right] \\ + \frac{np}{q^2} \ln \left(\frac{q}{np} \sum_{i=1}^n r^{q/2} \right) - \frac{np}{q} \left(\sum_{i=1}^n r^{q/2} \right)^{-1} \sum_{i=1}^n \left[r^{q/2} \ln(r) \right] = 0 \end{aligned} \quad (283)$$

Which simplifies as

$$\frac{n}{q} + \frac{np}{q^2} \Psi \left(\frac{p}{q} \right) + \frac{np}{q^2} \ln \left(\frac{q}{np} \sum_{i=1}^n r^{q/2} \right) - \frac{np}{q} \left(\sum_{i=1}^n r^{q/2} \right)^{-1} \sum_{i=1}^n \left[r^{q/2} \ln(r) \right] = 0 \quad (284)$$

Multiplying through Equation (284) by $(-q)$ gives

$$np \left(\sum_{i=1}^n r^{q/2} \right)^{-1} \sum_{i=1}^n \left[r^{q/2} \ln(r) \right] - n - \frac{np}{q} \left[\Psi \left(\frac{p}{q} \right) + \ln \left(\frac{q}{np} \sum_{i=1}^n r^{q/2} \right) \right] = 0$$

Thus, an objective function of q is obtained as

$$\alpha(q) = \frac{np}{2} \left(\sum_{i=1}^n r^{q/2} \right)^{-1} \sum_{i=1}^n \left[r^{q/2} \ln(r) \right] - n - \frac{np}{q} \left[\Psi \left(\frac{p}{q} \right) + \ln \left(\frac{q}{np} \sum_{i=1}^n r^{q/2} \right) \right] \quad (285)$$

From Equation (285), q does not have a unique solution. The estimation of q will be carried out using Newton-Raphson Method of Optimization given by

$$q_{n+1} = q_n - \frac{\alpha(q_n)}{\alpha'(q_n)}. \quad (286)$$

Algorithm (Pseudo-Code) of the MLE

The Algorithm (pseudo-code) of the MLE for the parameters of MEPD is as follows:

Estimate μ using Equation (262).

Initiation of q and $\mathbf{M}$.

for $a = 1, N \text{ iter max}$ **do**

Estimate $\mathbf{M}$ using Equation (267).

Estimate m using Equation (265).

Estimate q using Newton-Raphson iteration combining Equations (285) and (286).

End for

Estimate $\Sigma = m\mathbf{M}$

Implementation of the MLE

In this subsection, the new MEPD is assessed with a simulated and real datasets using information criteria as performance measures. The implementation is carried out using the MLE procedure with R software using the codes in Appendixes A, B and C.

Application to Simulated Data

Random samples are generated based on AMEPD with dispersion matrix $\Sigma = \mathbf{I}_2$ and $\mu = (0,0)^T$, for some specified values of the shape parameter, q , using various sample sizes; 150, 200, 500, 1,000 and 10,000.

Table 11: Summary of Parameter Estimation from Simulated Data for $q=0.5$

S. Size	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_{12}$	$\hat{q}$	LogLikelihood	AIC	BIC
AMEPD									
150	0.008	-0.005	0.754	0.696	0.038	0.611	-173.02	352.04	376.11
200	0.070	0.004	0.758	0.774	0.085	0.566	-180.00	366.01	391.80
500	-0.023	-0.040	0.984	0.973	0.051	0.489	-116.76	239.52	270.81
1000	-0.043	-0.042	0.918	0.929	0.031	0.490	-300.87	607.73	643.18
10000	0.001	0.004	1.023	0.978	0.004	0.502	-2917.31	5840.63	5889.89
MEPD									
150	0.008	-0.005	0.001	0.001	0.000	0.612	-2663.806	5333.611	5357.675
200	0.070	0.004	0.001	0.001	0.000	0.566	-3909.086	7824.173	7849.963
500	-0.023	-0.040	0.000	0.000	0.000	0.489	-11687.57	23381.14	23412.43
1000	-0.043	-0.042	0.000	0.000	0.000	0.490	-23341.07	46688.13	46723.58
10000	0.001	0.004	0.000	0.000	0.000	0.502	-225859.8	451725.7	451774.9

Source: Field survey, Manford (2023)

Table 12: Summary of Parameter Estimation from Simulated Data for $q=1$

S. Size	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_{12}$	$\hat{q}$	LogLikelihood	AIC	BIC
AMEPD									
150	0.017	0.025	0.878	0.750	0.017	1.166	-319.12	644.23	668.30
200	0.024	0.040	0.846	0.858	0.017	1.057	-394.07	794.13	819.92
500	-0.023	-0.017	0.945	0.974	-0.056	0.957	-859.20	1724.40	1755.69
1000	-0.075	-0.033	0.998	1.030	0.029	0.918	-1603.73	3213.46	3248.90
10000	0.002	0.021	1.046	0.992	0.014	0.999	-17189.69	34385.38	34434.64
MEPD									
150	0.017	0.025	0.158	0.135	0.003	1.166	-1140.24	2286.47	2310.53
200	0.024	0.040	0.095	0.096	0.002	1.057	-1721.70	3449.39	3475.18
500	-0.023	-0.017	0.061	0.063	-0.004	0.957	-4843.93	9693.85	9725.14
1000	-0.075	-0.033	0.050	0.052	0.001	0.918	-10187.66	20381.32	20416.76
10000	0.002	0.021	0.087	0.082	0.001	0.999	-90858.88	181723.80	181773.00

Source: Field survey, Manford (2023)

Table 13: Summary of Parameter Estimation from Simulated Data for $q=2$

S. Size	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_{12}$	$\hat{q}$	LogLikelihood	AIC	BIC
AMEPD									
150	-0.014	0.044	1.101	0.843	0.116	2.082	-355.44	716.88	740.95
200	0.098	-0.048	1.031	1.004	-0.106	1.842	-454.21	914.42	940.21
500	-0.010	0.006	1.019	0.985	0.016	1.746	-1129.97	2265.95	2297.24
1000	-0.027	0.031	0.928	1.044	-0.024	1.842	-2306.28	4618.55	4654.00
10000	-0.004	0.017	1.026	0.996	0.010	2.028	-23336.83	46679.65	46728.91
MEPD									
150	-0.014	0.044	1.195	0.915	0.126	2.082	-502.64	1011.27	1035.34
200	0.098	-0.048	0.817	0.801	-0.081	1.809	-807.56	1621.12	1646.91
500	-0.010	0.006	0.741	0.717	0.011	1.746	-2131.16	4268.32	4299.61
1000	-0.027	0.031	0.772	0.868	-0.020	1.842	-3970.41	7946.81	7982.26
10000	-0.004	0.017	1.057	1.026	0.011	2.028	-34510.96	69027.93	69077.19

Source: Field survey, Manford (2023)

Tables 11 to 13 present the results of parameter estimation for the simulated data under different conditions. The conditions are defined by the value of q (0.5, 1, and 2) and the sample sizes (S. Size). The tables also provide the goodness-of-fit measures such as the LogLikelihood, AIC and BIC for the estimated models.

From the tables, it can be observed throughout that both the AIC and BIC are lower for the AMEPD than MEPD, indicating that AMEPD provides a better fit of the data feature. It is further observed that for a specified value of q , parameter estimates are quiet close to the specified values for all sample sizes, especially for extremely large data size. Furthermore, almost the same estimates are obtained with slight deviations for both distributions for each sample size.

MLE Based on Real Datasets

Next is to assess the new MEPD with real dataset using information criteria as performance measures. Three datasets are used. These include the Carsales Data, Chemical Composition of River Data and Concrete Data.

Dataset 1: CarSales Dataset

Here, the data on car sales is used (Manford et al., 2024). Figure 19 is a graphical representation of distributions of the 6 variables.

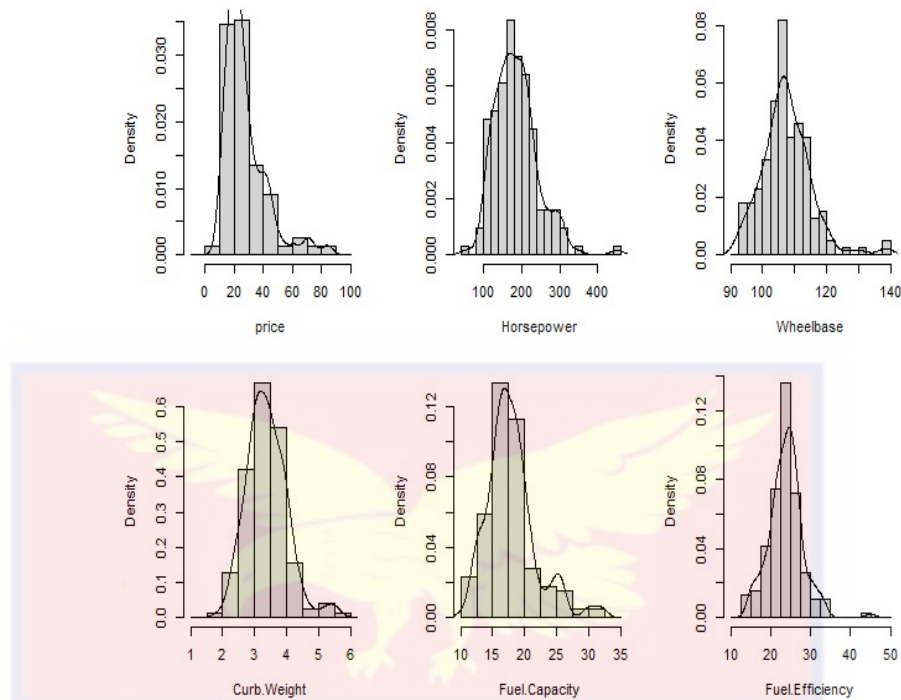


Figure 19: Distribution of Variables in Carsales Dataset

Figure 19 shows that the distributions of all the variables are fairly symmetrical with slight deviations.

The mean vector and variance-covariance matrix of the dataset are as follows:

$$\mu = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 27.31 & 185.01 & 107.42 & 3.38 & 17.96 & 23.85 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 206.60 & 683.93 & 11.96 & 4.81 & 23.91 & -30.32 \\ 683.93 & 3213.59 & 123.24 & 21.96 & 112.36 & -149.48 \\ 11.96 & 123.24 & 59.19 & 3.18 & 19.83 & -16.40 \\ 4.81 & 21.96 & 3.18 & 0.40 & 2.15 & -2.23 \\ 23.91 & 112.36 & 19.83 & 2.15 & 15.40 & -13.50 \\ -30.32 & -149.48 & -16.40 & -2.23 & -13.50 & 18.42 \end{pmatrix}$$

For both distributions the MLE estimate of the mean is given as

$$\hat{\mu} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C.Weight} & \text{F.Capacity} & \text{F.Efficiency} \\ 27.34 & 185.95 & 107.49 & 3.38 & 17.95 & 23.85 \end{pmatrix}$$

The MLE estimate of variance-covariance matrix based on AMEPD and MEPD are given, respectively, as

$$\hat{\Sigma}_{\text{AMEPD}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C.Weight} & \text{F.Capacity} & \text{F.Efficiency} \\ 207.69 & 758.30 & 34.85 & 6.68 & 35.33 & -38.70 \\ 758.30 & 3906.68 & 243.87 & 33.56 & 170.06 & -196.67 \\ 34.85 & 243.87 & 58.91 & 3.89 & 21.17 & -18.25 \\ 6.68 & 33.56 & 3.89 & 0.50 & 2.56 & -2.78 \\ 35.33 & 170.06 & 21.17 & 2.56 & 16.76 & -15.73 \\ -38.70 & -196.67 & -18.25 & -2.78 & -15.73 & 21.46 \end{pmatrix}$$

with $\hat{q} = 0.5659$, Loglikelihood = 4404.264, AIC = -8802.528, BIC = -8778.229;

and

$$\hat{\Sigma}_{\text{MEPD}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C.Weight} & \text{F.Capacity} & \text{F.Efficiency} \\ 0.0174 & 0.0634 & 0.0029 & 0.0006 & 0.0030 & -0.0032 \\ 0.0634 & 0.3264 & 0.0204 & 0.0028 & 0.0142 & -0.0164 \\ 0.0029 & 0.0204 & 0.0049 & 0.0003 & 0.0018 & -0.0015 \\ 0.0006 & 0.0028 & 0.0003 & 0.0000 & 0.0002 & -0.0002 \\ 0.0030 & 0.0142 & 0.0018 & 0.0002 & 0.0014 & -0.0013 \\ -0.0032 & -0.0164 & -0.0015 & -0.0002 & -0.0013 & 0.0018 \end{pmatrix}$$

with $\hat{q} = 0.5659$, loglikelihood = -4829.28 , AIC = 9665.456 , BIC = 9689.755 .

Both distributions yield the same estimate of q . Notably, the AMEPD gives a better fit with smaller performance measures.

Dataset 2: River Pra Dataset

The data is obtain on amount of chemical composition of River Pra (Ghana Minerals Commission, 2022). It is extracted on four variables with the following descriptions: mercury, copper, lead and arsenic. Figure 20 presents a graphical representation of the distributions of the four variables of Dataset 2.

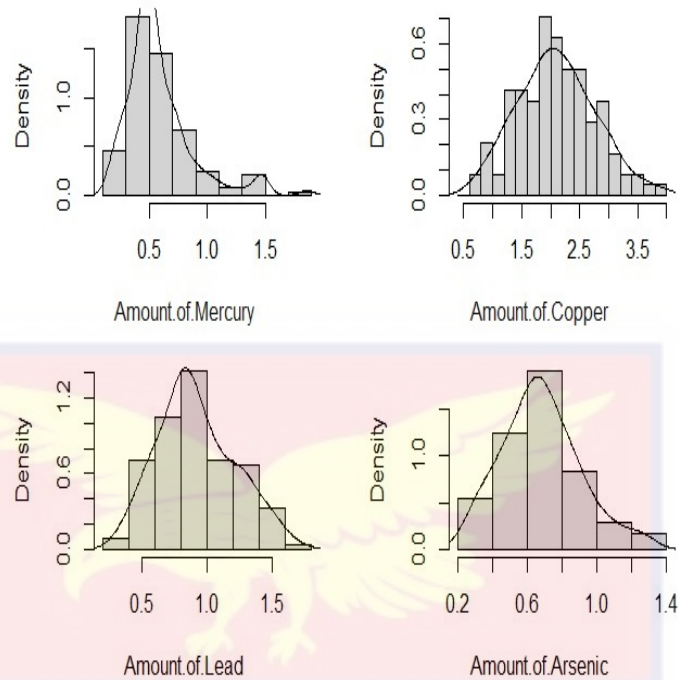


Figure 20: Distribution of Variables in River Pra Data

Figure 20 shows that almost all the variables have distributions that are fairly symmetric.

The mean vector and variance-covariance matrix of Dataset 2 is presented as follows:

$$\mu = \begin{pmatrix} \text{Mercury} & \text{Copper} & \text{Lead} & \text{Arsenic} \\ 0.588 & 2.099 & 0.917 & 0.678 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \text{Mercury} & \text{Copper} & \text{Lead} & \text{Arsenic} \\ 0.088 & 0.047 & 0.043 & 0.001 \\ 0.047 & 0.449 & 0.161 & 0.066 \\ 0.043 & 0.161 & 0.088 & 0.026 \\ 0.001 & 0.066 & 0.026 & 0.051 \end{pmatrix}$$

The estimate of the parameters for Dataset 2 using the AMEPD are as shown.

For both distributions, the estimate of the mean is obtained as

$$\hat{\mu} = \begin{pmatrix} \text{Mercury} & \text{Copper} & \text{Lead} & \text{Arsenic} \\ 0.588 & 2.099 & 0.917 & 0.678 \end{pmatrix}$$

The MLE estimate of the variance-covariance matrix based on AMEPD and MEPD are obtained, respectively, as

$$\hat{\Sigma} = \begin{pmatrix} \text{Mercury} & \text{Copper} & \text{Lead} & \text{Arsenic} \\ 0.117 & 0.227 & 0.114 & 0.029 \\ 0.227 & 1.202 & 0.494 & 0.175 \\ 0.114 & 0.494 & 0.219 & 0.074 \\ 0.029 & 0.175 & 0.074 & 0.116 \end{pmatrix}$$

with $\hat{q} = 0.4174$, loglikelihood = 141.68418, AIC = -277.368, BIC = -254.643;

and

$$\hat{\Sigma} = \begin{pmatrix} \text{Mercury} & \text{Copper} & \text{Lead} & \text{Arsenic} \\ 0.00000014 & 0.00000028 & 0.00000014 & 0.00000004 \\ 0.00000028 & 0.00000148 & 0.00000061 & 0.00000022 \\ 0.00000014 & 0.00000061 & 0.00000027 & 0.00000009 \\ 0.00000004 & 0.00000022 & 0.00000009 & 0.00000014 \end{pmatrix}$$

with $\hat{q} = 0.4174$, loglikelihood = -6964.181, AIC = 13934.36, BIC = 13957.09.

The results show that the parameter estimate for shape parameter q is the same for both the AMEPD and the MEPD. It is clear that AMEPD provides a much better estimates of the parameters.

Dataset 3: Concrete Dataset

Figure 21 presents a graphical representation of the distribution of the nine (9) variables that make Concrete dataset.

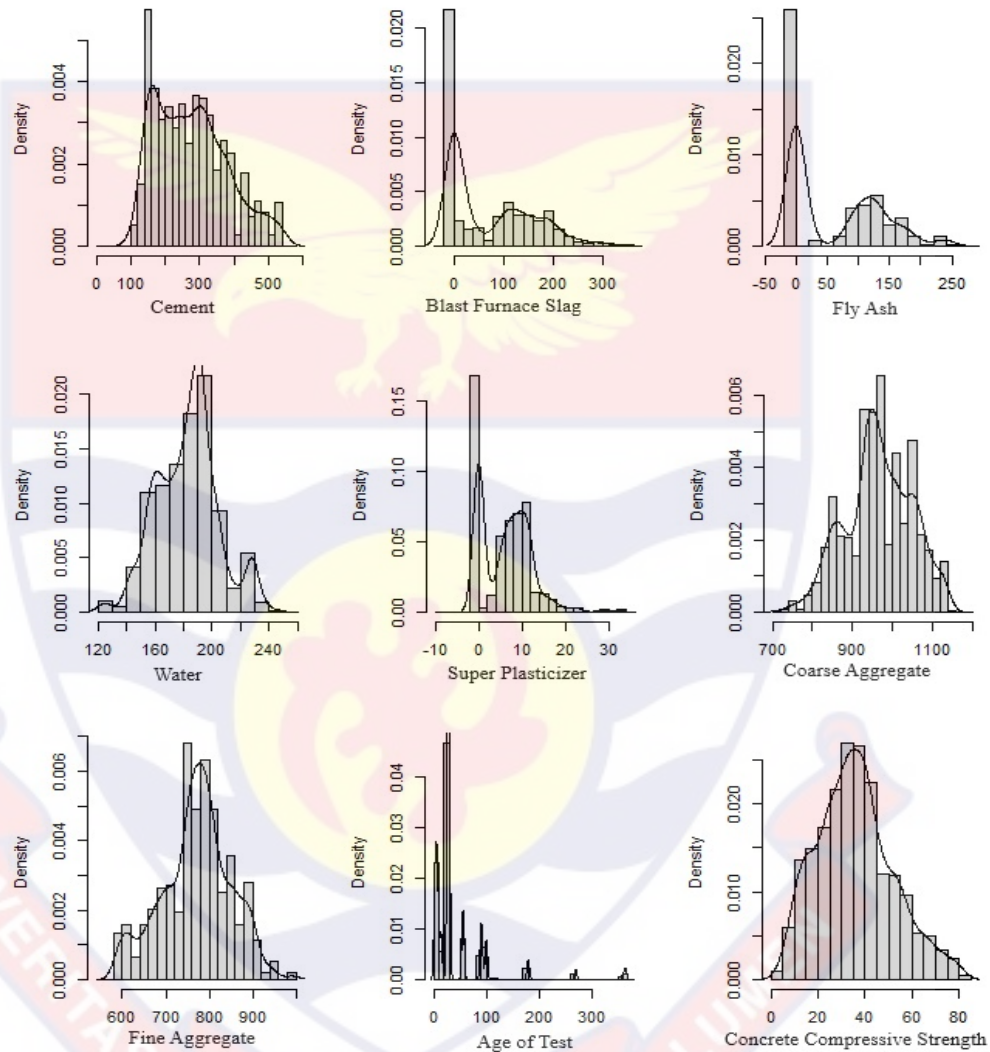


Figure 21: Distribution of Variables in Concrete Dataset

Figure 21 shows that the distribution is rather a complex one for all the variables. There are clear instances of multimodal distributions and skewness of the distributions. Thus, in all cases there is extreme departure from symmetry of the distributions of all the variables.

The mean vector and variance-covariance matrix of Dataset 3 is presented

as follows:

$$\mu = \begin{pmatrix} \text{Cement} & \text{BFS} & \text{F. Ash} & \text{Water} & \text{SP} & \text{CA} & \text{Fine A.} & \text{AoT} & \text{CCS} \\ 276.5 & 74.27 & 62.81 & 182.98 & 6.416 & 964.83 & 770.49 & 44.06 & 35.838 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \text{Cement} & \text{BFS} & \text{F. Ash} & \text{Water} & \text{SP} & \text{CA} & \text{Fine A.} & \text{AoT} & \text{CCS} \\ 10706.03 & -2377.54 & -3114.01 & -199.84 & 40.61 & -625.22 & -1526.62 & 566.67 & 813.96 \\ -2377.54 & 7097.52 & -1742.19 & 182.01 & 25.80 & -1869.63 & -1845.46 & -225.17 & 162.56 \\ -3114.01 & -1742.19 & 5124.15 & -234.49 & 146.35 & -625.33 & -35.57 & -705.82 & -74.23 \\ -199.84 & 182.01 & -234.49 & 471.49 & -74.02 & -486.88 & -732.00 & 317.63 & -97.26 \\ 40.61 & 25.80 & 146.35 & -74.02 & 33.60 & -131.95 & 91.23 & -69.52 & 33.18 \\ -625.22 & -1869.63 & -625.33 & -486.88 & -131.95 & 6853.89 & -1008.06 & 116.48 & -206.40 \\ -1526.62 & -1845.46 & -35.57 & -732.00 & 91.23 & -1008.06 & 6300.21 & -669.00 & -211.16 \\ 566.67 & -225.17 & -705.82 & 317.63 & -69.52 & 116.48 & -669.00 & 3653.15 & 315.16 \\ 813.96 & 162.56 & -74.23 & -97.26 & 33.18 & -206.40 & -211.16 & 315.16 & 259.23 \end{pmatrix}$$

The estimate of the parameters for Dataset 3 using the AMEPD are as shown. For both distributions, the estimate of the mean is obtained as

$$\hat{\mu} = \begin{pmatrix} \text{Cement} & \text{BFS} & \text{F. Ash} & \text{Water} & \text{SP} & \text{CA} & \text{Fine A.} & \text{AoT} & \text{CCS} \\ 276.50 & 74.27 & 62.81 & 182.98 & 6.42 & 964.83 & 770.49 & 44.06 & 35.84 \end{pmatrix}$$

The MLE estimate of the variance-covariance matrix based on AMEPD and MEPD are obtained, respectively, as

$$\hat{\Sigma} = \begin{pmatrix} \text{Cement} & \text{BFS} & \text{F. Ash} & \text{Water} & \text{SP} & \text{CA} & \text{Fine A.} & \text{AoT} & \text{CCS} \\ 11308.21 & -2365.52 & -3516.56 & -158.79 & 24.41 & -974.98 & -1464.38 & 424.49 & 850.36 \\ -2365.52 & 8353.77 & -2082.94 & 274.07 & 25.01 & -2707.75 & -2372.74 & -131.95 & 236.09 \\ -3516.56 & -2082.94 & 6026.95 & -355.20 & 198.92 & -462.55 & -102.44 & -379.27 & -18.65 \\ -158.79 & 274.07 & -355.20 & 475.84 & -76.59 & -573.25 & -705.57 & 170.03 & -114.19 \\ 24.41 & 25.01 & 198.92 & -76.59 & 35.25 & -141.45 & 68.57 & -32.19 & 41.14 \\ -974.98 & -2707.75 & -462.55 & -573.25 & -141.45 & 8021.89 & -681.88 & 35.29 & -304.70 \\ -1464.38 & -2372.74 & -102.44 & -705.57 & 68.57 & -681.88 & 6590.81 & -382.53 & -235.43 \\ 424.49 & -131.95 & -379.27 & 170.03 & -32.19 & 35.29 & -382.53 & 2433.55 & 335.64 \\ 850.36 & 236.09 & -18.65 & -114.19 & 41.14 & -304.70 & -235.43 & 335.64 & 297.00 \end{pmatrix}$$

with $q = 0.889$, loglikelihood = 42363.7, AIC = -84721.4, BIC = -84685.21, and

$$\hat{\Sigma} = \begin{pmatrix} 102.629 & -21.468 & -31.915 & -1.441 & 0.222 & -8.849 & -13.290 & 3.852 & 7.718 \\ -21.468 & 75.817 & -18.904 & 2.487 & 0.227 & -24.576 & -21.535 & -1.197 & 2.143 \\ -31.915 & -18.904 & 54.699 & -3.224 & 1.805 & -4.198 & -0.930 & -3.442 & -0.169 \\ -1.441 & 2.487 & -3.224 & 4.319 & -0.695 & -5.203 & -6.403 & 1.543 & -1.036 \\ 0.222 & 0.227 & 1.805 & -0.695 & 0.320 & -1.284 & 0.622 & -0.292 & 0.373 \\ -8.849 & -24.576 & -4.198 & -5.203 & -1.284 & 72.805 & -6.188 & 0.320 & -2.766 \\ -13.290 & -21.535 & -0.930 & -6.403 & 0.622 & -6.188 & 59.816 & -3.471 & -2.137 \\ 3.852 & -1.197 & -3.442 & 1.543 & -0.292 & 0.320 & -3.471 & 22.084 & 3.046 \\ 7.718 & 2.143 & -0.169 & -1.036 & 0.373 & -2.766 & -2.137 & 3.046 & 2.696 \end{pmatrix}$$

with $q = 0.889$, loglikelihood = -6902.039, AIC = 13810.08, BIC = 13846.279.

The parameter estimate for the concrete dataset gives the same value of q as 0.889 for both AMEPD and MEPD. However, the information criteria measures indicate that the performance of the AMEPD is no better than the MEPD, and this is to be expected. Since the coefficient c is fixed for the MEPD, it is not constrained by the value of the shape parameter allowing extremes in the data to be depicted and in an exaggerated manner. However, this may not provide the desired fit for the data. For the AMEPD on the other hand, the coefficient c is constrained by the value of the shape q . As a result, since the data contain extreme values that constitute severe deviation from symmetry, it does not appeal to provide a better fit.

Application in Discriminant Analysis

When there are equal prior probabilities and misclassification costs, the AMEPD is can be used for a discrimination problem that minimises the ex-

pected cost of misclassification. The procedure that is used is the same as that based on the multivariate normal distribution. However, parameter estimates are obtained numerically using the MLE procedure with R software using the codes in Appendixes A, C and E.

The Discriminant Function

Maximum likelihood method has been used to estimate AMEPD parameters in Equation (251). The resulting equations are not in close form, so to estimate the parameters for any given data, codes are developed in the R environment. Areas where normal distribution assumptions have not been reliable, they are replaced by AMEPD for such elliptically contoured data. The discriminant function between two AMEPD classes i and j from among classes $g = 1, 2, \dots, i, \dots, j, \dots, G$ may be derived as

$$\ln[k_j] - \ln[k_i] - \frac{1}{2} \ln \left(\frac{|\Sigma_j|}{|\Sigma_i|} \right) - \left[c_j (M_D(\mathbf{x}_j))^{q_j/2} - c_i (M_D(\mathbf{x}_i))^{q_i/2} \right] \quad (287)$$

where

$$M_D(\mathbf{x}_j) = \left| \mathbf{x}_j^\top \Sigma_j^{-1} \mathbf{x}_j - 2\mu_j^\top \Sigma_j^{-1} \mathbf{x}_j + \mu_j^\top \Sigma_j^{-1} \mu_j \right|$$

and

$$M_D(\mathbf{x}_i) = \left| \mathbf{x}_i^\top \Sigma_i^{-1} \mathbf{x}_i - 2\mu_i^\top \Sigma_i^{-1} \mathbf{x}_i + \mu_i^\top \Sigma_i^{-1} \mu_i \right|$$

To proof this, suppose $\mathbf{X}_i$ and $\mathbf{X}_j$ are multivariate observations from classes i and j , respectively, with density in Equation (??). Then the ratio $\frac{f_j(\mathbf{x}_j)}{f_i(\mathbf{x}_i)}$ can be obtained as

$$\frac{k_j}{k_i} \frac{|\Sigma_i|^{\frac{1}{2}}}{|\Sigma_j|^{\frac{1}{2}}} \exp \left(- \left\{ c_j \left| (\mathbf{x} - \mu_j)^\top \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right|^{\frac{q_j}{2}} - c_i \left| (\mathbf{x} - \mu_i)^\top \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right|^{\frac{q_i}{2}} \right\} \right) \quad (288)$$

Taking the natural logarithm of Equation (288) gives

$$\ln \left[\frac{f_j(\mathbf{x}_j)}{f_i(\mathbf{x}_i)} \right] = \ln \left(\frac{k_j}{k_i} \right) - \frac{1}{2} \ln \left(\frac{|\Sigma_j|}{|\Sigma_i|} \right) - \left[c_j \left| (\mathbf{x} - \mu_j)^\top \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right|^{\frac{q_j}{2}} - c_i \left| (\mathbf{x} - \mu_i)^\top \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right|^{\frac{q_i}{2}} \right] \quad (289)$$

If we let

$$M_D(\mathbf{x}_j) = \left| (\mathbf{x} - \mu_j)^\top \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right|$$

and

$$M_D(\mathbf{x}_i) = \left| (\mathbf{x} - \mu_i)^\top \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right|,$$

then the result in Equation (287) follows the AMEPD discriminant function between two classes, as stated. In terms of classification of two Populations, two cases are considered which are equal and unequal variance-covariance matrices.

Classification of Two Populations when Covariance Matrices are Equal

Suppose $\{X_{i1}, X_{i2}, \dots, X_{in_1}\}$ and $\{X_{j1}, X_{j2}, \dots, X_{jn_2}\}$ are two sets of random variables from the AMEPD populations, π_i and π_j with known parameters μ_i and μ_j and a common covariance matrix, Σ . Hence the overall discrimination and classification function in favour of group j for the AMEPD two group case is given as

$$- \left\{ c_j \left[(\mathbf{x} - \mu_j)^\top \Sigma^{-1} (\mathbf{x} - \mu_j) \right]^{\frac{q_j}{2}} - c_i \left[(\mathbf{x} - \mu_i)^\top \Sigma^{-1} (\mathbf{x} - \mu_i) \right]^{\frac{q_i}{2}} \right\} \geq \ln \left[\left(\frac{c(i|j)}{c(j|i)} \right) \left(\frac{p_j}{p_i} \right) \right] \quad (290)$$

where the pooled variance-covariance matrix given as

$$\Sigma = \frac{(n_j - 1) \Sigma_j + (n_i - 1) \Sigma_i}{n_j + n_i - 2}. \quad (291)$$

Hence, Equation (290) determines the minimum ECM regions from samples n_i and n_j . Thus, the classification rule that minimizes the ECM implies that $\mathbf{X}_0$ is allocated to group j if Equation (290) is satisfied, otherwise to group i .

Classification of Two Populations when Covariance Matrices are not Equal

Suppose that $f_i(\mathbf{x})$ and $f_j(\mathbf{x})$ are AMEPDs, and suppose also that the population parameters μ_i, Σ_i and μ_j, Σ_j are known. The allocation rule that minimizes the expected ECM is given by

$$H - \left\{ c_j \left[(\mathbf{x} - \mu_j)^\top \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right]^{\frac{q_j}{2}} - c_i \left[(\mathbf{x} - \mu_i)^\top \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right]^{\frac{q_i}{2}} \right\} \geq \ln \left[\left(\frac{c(i|j)}{c(j|i)} \right) \left(\frac{p_j}{p_i} \right) \right] \quad (292)$$

where $H = \frac{1}{2} \ln \left| \frac{\Sigma_j}{\Sigma_i} \right| - \ln \left[\frac{k_j}{k_i} \right]$. So $\mathbf{X}_0$ is allocated to group j if Equation (292) is satisfied, otherwise to group i .

Demonstration of Discriminant Analysis to Multivariate Dataset using MEPD and AMEPD

In this subsection, we apply the discriminant function obtained and the classification rule from Equations (290) and (292) to the Carsales dataset. Figure 22 presents the distributions of pairs of variables in two selected groups of automobile and truck that constitute the same dataset.

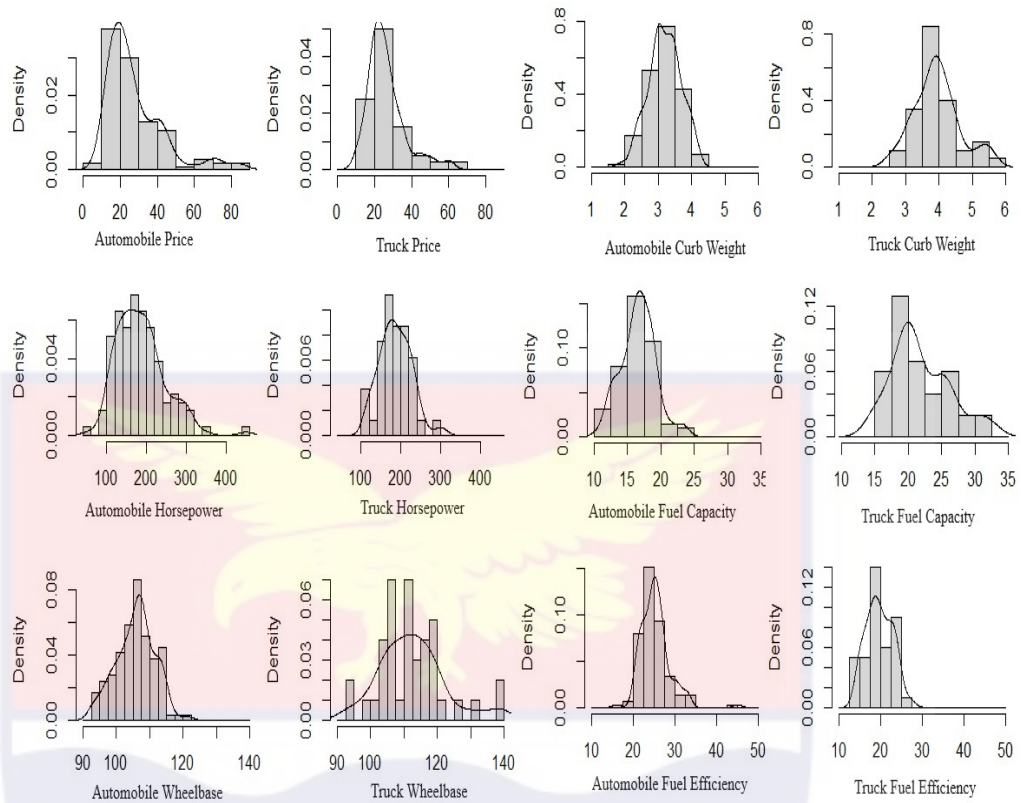


Figure 22: Distribution of Six Car Descriptors by Type of Car

It can be seen that the distributions of variables in one group do not have quite distinguishable features in general from those of the other group. However, the distribution on automobile variables appears to be a bit more clustered on most of the variables. The dataset thus possesses interesting features that could provide a good test of performance for the proposed distribution.

We estimate the parameters of MEPD and AMEPD for the Carsales Dataset for the truck group. For AMEPD, the result is presented as follows

$$\hat{\mu}_{\text{truck}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 26.32 & 186.40 & 190.43 & 3.93 & 21.90 & 19.70 \end{pmatrix}$$

$$\hat{\Sigma}_{\text{truck}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 93.789 & 313.677 & 10.009 & 5.835 & 18.654 & -19.357 \\ 313.677 & 1695.661 & 122.390 & 24.411 & 99.545 & -91.488 \\ 10.009 & 122.390 & 94.534 & 3.233 & 28.771 & -9.827 \\ 5.835 & 24.411 & 3.233 & 0.502 & 2.375 & -1.731 \\ 18.654 & 99.545 & 28.771 & 2.375 & 20.025 & -10.123 \\ -19.357 & -91.488 & -9.827 & -1.731 & -10.123 & 10.925 \end{pmatrix},$$

and $\hat{q} = 1.012162$. We then estimated the parameters of AMEPD for the Auto-mobile Group of Carsales Dataset. The result is presented as follows

$$\hat{\mu}_{\text{auto}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 27.69 & 185.79 & 186.280172 & 3.20 & 16.59 & 25.28 \end{pmatrix}$$

$$\hat{\Sigma}_{\text{auto}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 213.256 & 769.959 & 25.345 & 5.263 & 29.096 & -33.239 \\ 769.959 & 3897.725 & 191.039 & 27.236 & 135.828 & -167.436 \\ 25.345 & 191.039 & 36.244 & 2.378 & 9.783 & -9.873 \\ 5.263 & 27.236 & 2.378 & 0.286 & 1.273 & -1.528 \\ 29.096 & 135.828 & 9.783 & 1.273 & 7.548 & -7.598 \\ -33.239 & -167.436 & -9.873 & -1.528 & -7.598 & 12.424 \end{pmatrix},$$

with $\hat{q} = 0.6914138$.

For MEPD,

$$\hat{\Sigma}_{\text{truck}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 3.6584 & 12.2349 & 0.3903 & 0.2276 & 0.7276 & -0.7550 \\ 12.2349 & 66.1376 & 4.7738 & 0.9521 & 3.8826 & -3.5683 \\ 0.3903 & 4.7738 & 3.6874 & 0.1261 & 1.1222 & -0.3833 \\ 0.2276 & 0.9521 & 0.1261 & 0.0196 & 0.0926 & -0.0675 \\ 0.7276 & 3.8826 & 1.1222 & 0.0926 & 0.7811 & -0.3948 \\ -0.7550 & -3.5683 & -0.3833 & -0.0675 & -0.3948 & 0.4261 \end{pmatrix},$$

and

$$\hat{\Sigma}_{\text{auto}} = \begin{pmatrix} \text{Price} & \text{Hpower} & \text{Wbase} & \text{C. Weight} & \text{F. Capacity} & \text{F. Efficiency} \\ 0.2487 & 0.8980 & 0.0296 & 0.0061 & 0.0339 & -0.0388 \\ 0.8980 & 4.5460 & 0.2228 & 0.0318 & 0.1584 & -0.1953 \\ 0.0296 & 0.2228 & 0.0423 & 0.0028 & 0.0114 & -0.0115 \\ 0.0061 & 0.0318 & 0.0028 & 0.0003 & 0.0015 & -0.0018 \\ 0.0339 & 0.1584 & 0.0114 & 0.0015 & 0.0088 & -0.0089 \\ -0.0388 & -0.1953 & -0.0115 & -0.0018 & -0.0089 & 0.0145 \end{pmatrix}.$$

with the same mean vectors and values of $\hat{q}$ as the AMEPD in each case, respectively.

If we assume equal costs of misclassification, we have

$C(\text{Automobile} | \text{Truck}) = C(\text{Truck} | \text{Automobile})$, and prior probabilities:

$P(\text{Automobile}) = 0.7436$, $P(\text{Truck}) = 0.2564$, the following classification rule

is used: Classify as Automobile if

$$\frac{f_{\text{Automobile}}(\mathbf{x})}{f_{\text{Truck}}(\mathbf{x})} \geq \left(\frac{C(\text{Automobile} | \text{Truck})}{C(\text{Truck} | \text{Automobile})} \right) \left(\frac{P(\text{Truck})}{P(\text{Automobile})} \right) = 0.3448, \text{ and}$$

Truck otherwise.

The results (Tables 14 and 15) show that the performance of the AMEPD is much better in the determination of the discrimination problem. Particularly, the performance of AMPED is optimal under condition of unequal variances with much smaller APER.

Table 14: Confusion Matrix for Type of Car from AMEPD Populations for**Equal Variance-Covariance Matrix**

		Predicted Type of Car			
		AMEPD	Automobile	Truck	% CC
			(π_1)	(π_2)	
Actual Type of Car	Automobile (π_1)		99	17	116
	Truck (π_2)		3	37	40
					85.3
					92.5
					87.2
Actual Type of Car	Automobile (π_1)		0	116	116
	Truck (π_2)		0	40	40
					0.0
					100.0
					25.6

Source: Field survey, Manford (2023)

Table 15: Confusion Matrix for Type of Car from AMEPD Populations for**Unequal Variance-Covariance Matrix**

		Predicted Type of Car			
		AMEPD	Automobile	Truck	% CC
			(π_1)	(π_2)	
Actual Type of Car	Automobile (π_1)		106	10	116
	Truck (π_2)		3	37	40
					91.4
					92.5
Actual Type of Car	Automobile (π_1)		64	52	116
	Truck (π_2)		0	40	40
					55.2
					100.0
					66.7

Source: Field survey, Manford (2023)

Discussion

The core innovation proposed in this study revolves around the introduc-

tion of a non-constant coefficient in the quadratic form, thereby giving rise to the AMEPD. This modification is strategically designed to mitigate the influence of the shape parameter on the scale parameter, offering a solution to limitations observed in existing distribution by Gomez et al. (1998). The envisioned outcome is the creation of a family of elliptical contour distributions, blending characteristics from Kotz-type distributions and the distribution presented by Gómez et al. (1998). The research carries substantial significance as it contributes to the evolution of a more explicit and flexible multivariate EPD that can adeptly fit symmetrically to multivariate data. By deriving the density and characteristics of the AMEPD, provide estimators through maximum likelihood estimation, compare its efficiency with related distributions using both simulated and real data, and showcase its application in various multivariate statistical techniques. In essence, this research has the potential to redefine the landscape of multivariate statistical analysis and contribute to the ongoing evolution of sophisticated methodologies for handling complex and diverse datasets.

The demonstrated effectiveness of the EPD in discriminant and classification analyses, beckon a deeper inquiry into its potential applications in hypothesis testing, multivariate regression analysis, and Bayesian analysis. The study's key takeaway, emphasizing that symmetrical data do not invariably conform to a Normal distribution, highlights the importance of a paradigm shift in distribution selection. It calls for a more nuanced approach, urging researchers to consider both skewness and kurtosis thoroughly before settling on a distribution for data analysis. This call for a broader perspective aligns with the

evolving landscape of statistical methodologies, advocating for models that better capture the complexities inherent in diverse datasets. A promising avenue for further investigation lies in exploring the mathematical derivatives of the relationship between the multivariate EPD and other statistical distributions. Given the EPD's diverse properties shared with several statistical distributions, unraveling these mathematical connections could deepen our comprehension of its behavior and potentially unveil novel applications and insights.

The application to real data, in particular, has highlighted the specific data type for which the proposed distribution is most suitable. The desired data type is one that has no severe departure from symmetry. Thus, the study has achieved a distribution that has features that were initially envisaged. This means that for data that show marked deviations from symmetry, it would require distributions that are specifically designed for them. The Concrete data (Yeh, 2006) is a typical data that has been useful for highlighting complex features for which the proposed distribution may not be suitable. It is observed that the relevant complexity is one that is much more than the usual skewness and kurtosis for which notable research (Arellano-Valle & Richter, 2012; Anzallini, 2001) in the area have been engaged. It is envisaged that the AMEPD could be applied quite effectively to data with multiple distributions as that of the Concrete data if the data is appropriately segmented based on well-determined change-points in the symmetry. This process will require piece-wise application of the distribution over various intervals. Such an application of the AMEPD will also serve as an innovative approach to the treatment of data problems that consist of mul-

multiple distributions. As the new AMEPD has shown more flexibility to involve the existing MEPD, it is envisaged that by appropriately introducing a skewed parameter, a new distribution could be obtained that will further provide more flexibility into the existing family of distributions that are designed for skewed data.

Chapter Summary

This chapter have presented some characteristics of the AMEPD using the properties of Gamma function. We derived the mean, variance, skewness and kurtosis of the distribution. The normalising constant, k , was also derived. The constant, c , was also derived such that the variance-covariance matrix becomes the scale parameter, Σ , (that is, $D(x) = \Sigma$). Some values of the coefficient of kurtosis of the EPD was estimated with respect to some values of q . It was found that for $q = 2$, the distribution is the same as the multivariate Normal distribution. Also, for values of $q < 2$, the distribution is platykurtic, and for values of $q > 2$, the distribution is leptokurtic. Again, for $q \rightarrow \infty$, the distribution is rectangular, that is multivariate uniform distribution. Also, it was found that for the coefficient of kurtosis, $q = 1$, the distribution depict a multivariate double exponential or Laplace distribution. In line with the prior evaluation, the AMEPD accurately replicates all characteristics of the generated data in closer alignment with the original dataset across various values of the designated shape parameter, denoted as q . The findings also indicate that, particularly for low values of q (less than 1), the established MEPD demonstrates a notably inadequate fit to the

underlying distribution. It is evident that when q equals 2, the two distributions become identical.

Since AMEPD is a parametric distribution, the maximum likelihood estimation method was used to derive an expression for the location, μ , the scale, Σ and the shape, q parameters. In order to carry out the data analysis, this is fitting the AMEPD to a data set, the arithmetic mean was used as the estimator for the location parameter, since the expression for the maximum likelihood for the location parameter does not have an explicit solution. The use of the arithmetic mean was due to the symmetrical nature of the Exponential Power Distribution. For the scale parameter, the maximum likelihood estimator of Σ was used. Also, the estimation of the shape parameter by means of maximum likelihood has some computational difficulty due to the complexity of the expression derived by means of maximum likelihood. Thus, Newton Raphson optimization technique was adopted. Here, the location parameter, μ , the scale parameter, Σ and the shape parameter, q were estimated for a simulated data generated from AMEPD population for sample sizes of 150, 200, 500, 1000 and 10000. The estimation was done for both AMEPD and MEPD. It was also revealed that AMEPD performs better than MEPD. Again, the estimation was done for the three datasets used in this study. However, using the AIC and BIC, we observed that the AMEPD performed better for all the datasets except for the concrete data. The results also show that both AMEPD and MEPD will be inappropriate to use in fitting multivariate asymmetrical (skewed) distributed data set, since the variables which were not significant were highly skewed data.

The AMEPD was then used in discriminant analysis. The confusion matrices for the Type of Car observations from AMEPD populations, comparing predictions under Equal and Unequal Variance-Covariance Matrix assumptions were presented. The findings also revealed that, for equal variance-covariance matrix, we have APER of 0.12821 (12.82% average prediction error rate). Again, 99 out of 116 actual Automobiles were correctly predicted, and 17 were incorrectly predicted as Trucks. Also, 37 out of 40 actual Trucks were correctly predicted, and 3 were incorrectly predicted as Automobiles. For unequal variance-covariance matrix, it yield APER fo 0.08333 (8.33% average prediction error rate), indicating an improvement from the equal covariance matrix case. Also, 106 out of 116 actual Automobiles were correctly predicted, and 10 were incorrectly predicted as Trucks. The prediction for Trucks remains the same as in equal variance-covariance assumption. Thus, the model with Unequal Variance-Covariance Matrix shows a reduced average prediction error rate compared to the model with Equal Variance-Covariance Matrix, suggesting improved overall performance. The improvement is particularly evident in the prediction of Automobiles, with a decrease in misclassifications. The performance in predicting Trucks remains consistent between the two models. The results imply that the choice of variance-covariance matrix assumption has a notable impact on the model's ability to predict certain types of cars, specifically Automobiles.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATION

Overview

This chapter focuses on the summary of the study, the findings resulting from analysis of the dataset used, conclusions based on the findings and possible recommendations.

Summary

In various fields, the advancement of research has generated real-life data that often defies modeling using classical probability distributions. Consequently, there is a crucial need for developing new classes of distributions, generalizations, or extensions of existing models to enhance the flexibility of probability distributions. While the Normal distribution is frequently assumed in hypothesis testing for both univariate and multivariate techniques, the multivariate Normal distribution may prove inadequate in practical situations where data exhibits non-Gaussian characteristics. The univariate EPD, introduced by Subbotin in 1923, offers an alternative to the Normal distribution, particularly in terms of kurtosis. The EPD, also known as the GED, is characterized by varying kurtosis and symmetry. Gómez (2007) has significantly contributed to the field by incorporating heavy-tailed characteristics into linear regression and time series models, assuming the underlying distribution to be GED. The versatility of univariate GED has been explored across diverse domains, emphasizing its

adaptability and relevance.

In the field of statistical tools, Mineo and Ruggieri (2005) introduced the "normalp package" tailored for the statistical environment R, providing tools for handling univariate GED. The GED's applications extend across various domains, highlighting its utility in capturing diverse data patterns. The multivariate EPD, a subset of Elliptical Contoured Distributions, has gained prominence for its ability to model complex data structures in fields like finance, biology, and engineering. The family of Elliptical Contoured Distributions, traced back to foundational works by various scholars, offers a rich framework for multivariate data modeling. The multivariate EPD inherits the properties of Elliptical Contoured Distributions and is characterized by an exponential power form, making it flexible and suitable for various applications.

The study proposes an alternative multivariate EPD (AMEPD) by introducing a non-constant coefficient in the quadratic form. This modification aims to eliminate the influence of the shape parameter on the scale parameter, addressing limitations observed in existing formulations. The AMEPD is expected to be a family of elliptical contour distributions, combining characteristics of Kotz-type distributions and the distribution presented by Gómez et al. (1998).

The significance of this research lies in its contribution to the development of a more explicit and flexible multivariate EPD that can better fit symmetrically to multivariate data. The study aims to derive the density and characteristics of AMEPD, provide estimators using maximum likelihood estimation, compare efficiency with related distributions using simulated and real data, and demon-

strate its application in multivariate statistical techniques. The purpose of the study is to contribute to advancements in statistical modeling and analysis by introducing a novel perspective on multivariate EPD. The objectives include deriving the density and characteristics of AMEPD, developing estimators, comparing efficiency, and demonstrating practical applications. The significance of this study is underscored by its potential to enhance our understanding of multivariate distributions and provide a valuable tool for statisticians and researchers working with complex data structures. It fills a gap in the literature by addressing limitations in existing multivariate EPD formulations and opens avenues for further research in the field of statistical distributions and data modeling.

The study employed simulated Bivariate Data generated from an AMEPD for various sample sizes. Additionally, three real datasets, namely Carsales, Concrete, and River pra datasets, are utilized for analysis. It introduced the Gamma and Dirichlet distributions, integral to developing elliptical and spherical distributions. The Gamma function simplifies the characteristics of the AMEPD, while the Dirichlet distribution is foundational for the distribution. We covered the Liouville distribution and integral expressions relevant to the derivation of multivariate EPD. We also delved into the definition and properties of spherical and elliptical symmetric distributions. The study established theorems characterizing these distributions based on their characteristic functions and presents properties, including moments, covariance, and density generators. The study provided detailed information on multivariate normal distributions and the multivariate EPD. It covers mathematical expressions, trans-

formations, and properties associated with these distributions. The method of deriving spherical and elliptical symmetric distributions is discussed, including approaches such as density, characteristic function, linear combination, and stochastic decomposition. Approaches to parameter estimation are explored, with a focus on maximum likelihood estimation (MLE) and its application in univariate and multivariate cases. The Newton-Raphson optimization method is introduced for estimating the shape parameter in the multivariate case, considering its robustness and convergence properties. Model selection techniques are discussed, emphasizing the trade-off between bias and variance. Information criteria such as AIC and BIC are introduced for comparing non-nested models. The study highlights the importance of model evaluation through Q-Q plots for the multivariate EPD. The application of multivariate EPD in Multivariate Data Analysis, particularly in discriminant and classification analysis, is discussed. The study explores its use in various statistical techniques and concludes by emphasizing the role of AMEPD in reducing the cost of misclassification in the discrimination and classification of multivariate data.

The coefficient of kurtosis for the AMEPD was examined concerning various values of q . Notably, for $q = 2$, the distribution aligns with the multivariate Normal distribution as well as MEPD. For $q < 2$, the distribution is platykurtic, while for $q > 2$, it is leptokurtic. As $q \rightarrow \infty$, the distribution approaches a rectangular shape, resembling a multivariate uniform distribution. Additionally, $q = 1$ results in a distribution resembling a multivariate double exponential or Laplace distribution. Given that AMEPD is a parametric distribution, the

maximum likelihood estimation (MLE) method is employed to derive expressions for the location (μ), the scale (Σ), and shape (q) parameters. For practical data analysis and fitting AMEPD to a dataset, the arithmetic mean serves as the estimator for the location parameter due to the symmetrical nature of the Exponential Power Distribution. The scale parameter (Σ) is estimated using the MLE, and the shape parameter (q) is computed through the Newton-Raphson optimization technique.

Estimations are carried out for both AMEPD and MEPD using simulated data and three real datasets. The results reveal that AMEPD outperforms MEPD, except for the concrete data, as indicated by AIC and BIC. The study concludes that both distributions are inappropriate for fitting multivariate asymmetrical (skewed) datasets, particularly when dealing with highly skewed variables. The study proceeds to apply AMEPD in discriminant analysis. Confusion matrices are presented for Type of Car observations, comparing predictions under Equal and Unequal Variance-Covariance Matrix assumptions. The findings indicate that the model with Unequal Variance-Covariance Matrix performs better, showing a reduced average prediction error rate and improved overall performance, particularly in predicting Automobiles. The choice of variance-covariance matrix assumption significantly influences the model's ability to predict specific car types, underscoring the importance of this consideration in discriminant analysis.

Conclusions

The study has examined extensively the AMEPD as given in Equation (251). In particular, the study has examined the characteristics and properties of AMEPD. It has also assessed the AMEPD and the MEPD, particularly, its deviation from normality using the shape parameter. The derivation and estimation of the parameters of the AMEPD with respect to a simulated and real dataset have been done and the examination of the fitness of AMEPD to the dataset has been conducted.

The examination of the coefficient of kurtosis for the AMEPD considered various values of q . It can be concluded that when $q = 2$, the distribution aligns with both the multivariate Normal distribution and MEPD. For $q < 2$, the distribution exhibits platykurtic characteristics, while for $q > 2$, it displays leptokurtic behavior. As $q \rightarrow \infty$, the distribution transforms into a rectangular shape, resembling a multivariate uniform distribution. Moreover, when $q = 1$, the distribution takes on features reminiscent of a multivariate double exponential or Laplace distribution.

The maximum likelihood estimation (MLE) method was employed to derive expressions for the location, scale, and shape parameters. For practical data analysis and the fitting of AMEPD to a dataset, the arithmetic mean serves as the estimator for the location parameter due to the symmetrical nature of the Exponential Power Distribution. The scale parameter is estimated using MLE, and the shape parameter is determined through the Newton-Raphson optimization technique. Estimations are conducted for both AMEPD and MEPD using simu-

lated data and three real datasets. AIC and BIC metrics indicate potential advantages of AMEPD over MEPD. However, the adequacy of the concrete data for applying AMEPD is questionable. More rigorous data assessment is required to establish a definitive conclusion. Additionally, both distributions are unsuitable for fitting multivariate asymmetrical (skewed) datasets, especially when dealing with highly skewed variables.

In the application to multivariate data analysis, the AMEPD was applied in discriminant analysis, presenting confusion matrices for Type of Car observations and comparing predictions under Equal and Unequal Variance-Covariance Matrix assumptions. It can be concluded that AMEPD exhibit a superior performance with unequal variance-covariance Matrix, demonstrating a reduced average prediction error rate and improved overall accuracy, particularly in predicting Automobiles. Thus, it underscores the crucial impact of the variance-covariance matrix assumption on the model's ability to predict specific car types, emphasizing its significance in discriminant analysis.

Recommendations

Based on the insights derived from this study, the adoption of the multivariate EPD presents itself as a viable alternative model when the assumption of normality is in question. This distribution can effectively model random phenomena with distributions characterized by tails that are either heavier or lighter than those of the Normal distribution. As a result, it is proposed that the multivariate EPD offers considerable flexibility in fitting real data, striking a balance

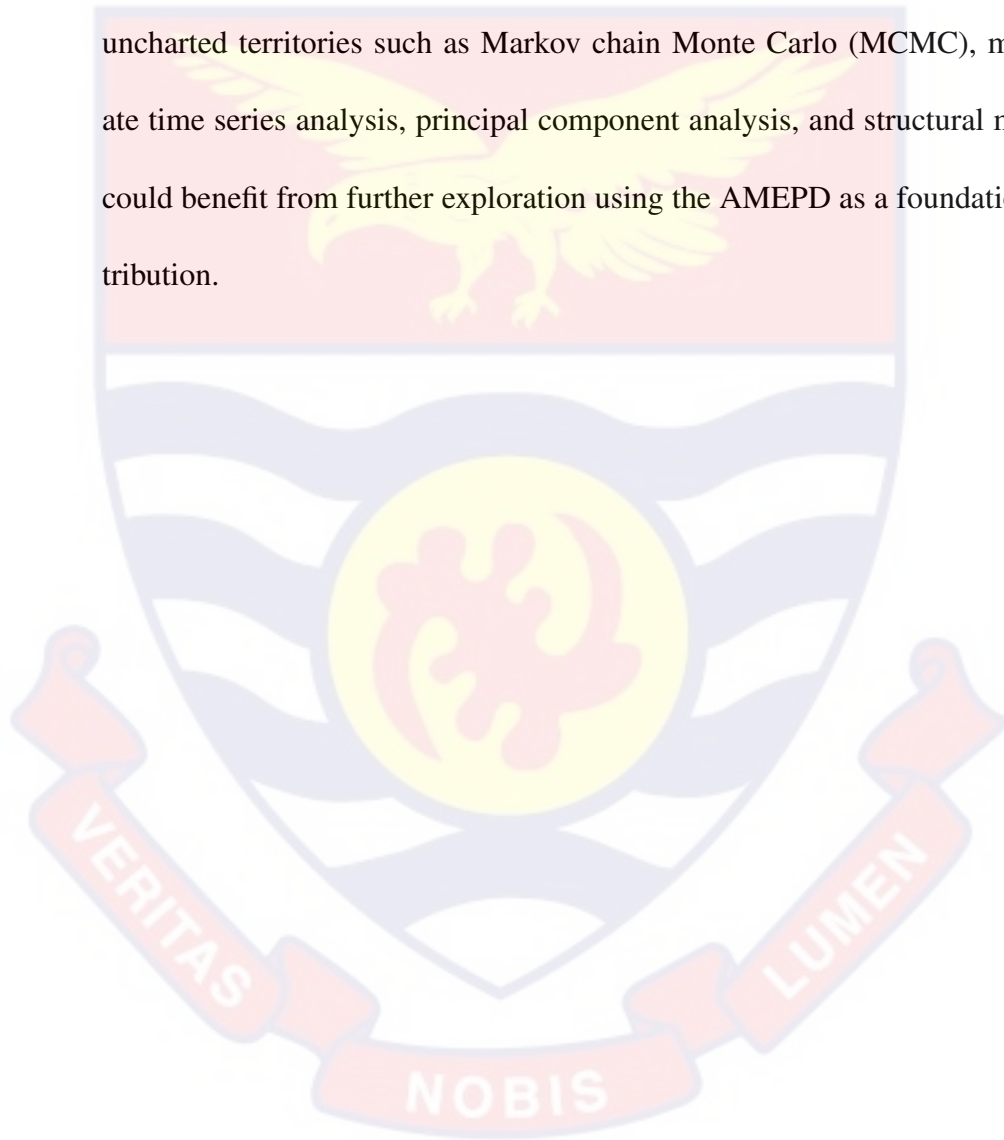
between adaptability and robustness, although it may not be suitable for skewed data.

Further exploration of the relevance of this distribution in various field of statistical data analysis is justified. Specifically, its utility in hypothesis testing, multivariate regression analysis, and Bayesian analysis should be investigated, given its demonstrated effectiveness in discriminant and classification analyses. The findings of this study underscore the notion that symmetrical data do not always conform to a Normal distribution. Therefore, a thorough consideration of both skewness and kurtosis should precede the selection of a distribution for data analysis. The study advocates for the incorporation of more advanced statistical procedures in tests of differences and tests of relationships, utilizing the multivariate EPD as the foundational assumption due to its robustness in accommodating various kurtosis levels in datasets.

It is imperative to acknowledge that the AMEPD is unsuitable for skewed data. Consequently, there is a pressing need to extend this work to address the challenges associated with the asymmetric multivariate case, as many multivariate datasets may exhibit skewness, either positive or negative. While some existing studies, such as those on the Anzalini (1985) type of distribution, touch upon the skewness aspect, further research is warranted. Exploring the mathematical derivatives of the relationship between the multivariate EPD and other statistical distributions is also a promising avenue, considering the diverse properties it shares with several statistical distributions.

Moreover, the application of alternative mathematical optimization tech-

niques and algorithms, such as genetic algorithms, minimax optimization, nonlinear least squares, simulated annealing algorithms, nonlinear maximization, and others, could enhance the precision and accuracy of parameter estimation for the multivariate EPD. This is particularly relevant since the maximum likelihood method lacks explicit solutions for the shape parameter. Additionally, uncharted territories such as Markov chain Monte Carlo (MCMC), multivariate time series analysis, principal component analysis, and structural modeling could benefit from further exploration using the AMEPD as a foundational distribution.



REFERENCES

- Adèr, H. J. (2008). *Advising on research methods: A Consultant's Companion*. Huizen, Netherlands: Johannes van Kessel Publishing.
- Adler, R. J., Feldman, R. E., & Taqqu, M. S. (1999). A Practical Guide to heavy tails: Statistical techniques and applications. *Journal of the American Statistical Association*. 94(446). 653. <https://doi.org/10.2307/2670194>.
- Agrò, G. (1995). Maximum likelihood estimation for the exponential power function parameters. *Communications in Statistics: Simulation and Computation*. 24(2). 523–536. <https://doi.org/10.1080/03610919508813255>.
- Aitchison, J. (1986). *The statistical analysis of compositional data*. Dordrecht, Netherlands: Springer.
- Aitchison, J., & Silvey, S. D. (1958). Maximum-Likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*. 29(3). 813–828. <https://doi.org/10.1214/aoms/1177706538>.
- Ali, H. R., Dariush, A., Provenzano, E., Bardwell, H., Abraham, J., Iddawela, M., Vallier, A., Hiller, L., Dunn, J. A., Bowden, S., Hickish, T., McAdam, K., Houston, S., Irwin, M. J., Pharoah, P. D., Brenton, J. D., Walton, N. A., Earl, H., & Caldas, C. (2016). Computational pathology of pre-treatment biopsies identifies lymphocyte density as a predictor of response to neoadjuvant chemotherapy in breast cancer. *Breast Cancer Research*. 18(1). <https://doi.org/10.1186/s13058-016-0682-8>.
- Althuwaynee, O. F., Pradhan, B., Park, H., & Lee, J. H. (2014). A novel ensemble bivariate statistical evidential belief function with knowledge-

based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *CATENA*. 114. 21–36. <https://doi.org/10.1016/j.catena.2013.10.011>.

Álvarez, I., Niemi, J., & Simpson, M. (2014). Bayesian inference for a covariance matrix. *Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1004>.

Anderson, D. N. (1992). A multivariate Linnik distribution. *Statistics & Probability Letters*. 14(4). 333–336. [https://doi.org/10.1016/0167-7152\(92\)90067-f](https://doi.org/10.1016/0167-7152(92)90067-f).

Anderson, J. (1969). Constrained Discrimination between K Populations. *Journal of the Royal Statistical Society Series B-methodological*. 31(1). 123–139. <https://doi.org/10.1111/j.2517-6161.1969.tb00774.x>.

Anderson, M., Fu, G., Phlypo, R., & Adalı, T. (2013). Independent vector analysis, the Kotz distribution, and performance bounds. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. <https://doi.org/10.1109/icassp.2013.6638257>.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Hoboken, NJ, USA: Wiley-Interscience.

Anderson, T. W., & Fang, K. (1982). Distributions of quadratic forms and Cochran's theorem for elliptically contoured distributions and their applications. In *Technical Report NO. 53, ONR Contract N00014-75-C-0442* (No. ADA115355). Department of Statistics, Stanford University California. Retrieved December 14, 2022, from <https://apps.dtic.mil/sti/>

pdfs/ADA115355.pdf.

Anderson, T. W., & Fang, K. (1990). *Theory and applications of elliptically contoured and related distributions*. Allerton, New York. <https://doi.org/10.21236/ada230672>.

Arellano-Valle, R. B. (2001). On some characterizations of spherical distributions. *Statistics & Probability Letters*. 54(3). 227–232. [https://doi.org/10.1016/s0167-7152\(00\)00191-7](https://doi.org/10.1016/s0167-7152(00)00191-7).

Arellano-Valle, R. B., Del Pino, G., & MartiN, E. S. (2002). Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters*. 58(2). 111–121. [https://doi.org/10.1016/s0167-7152\(02\)00088-3](https://doi.org/10.1016/s0167-7152(02)00088-3).

Arellano-Valle, R. B., & Richter, W.-D. (2012). On skewed continuous In, p-symmetric distributions. *Chilean Journal of Statistics*. 3(2). 193–212. <http://www.math.uni-rostock.de/richter/W-DR2012-2.pdf>.

Arias-Castro, E. (2022). *Principles of statistical analysis: Learning from Randomized Experiments*. Cambridge, United Kingdom: Cambridge University Press.

Arslan, G. (2014). A new characterization of the power distribution. *Journal of Computational and Applied Mathematics*. 260. 99–102. <https://doi.org/10.1016/j.cam.2013.09.068>.

Ashby, F. G. (2014). *Multidimensional models of perception and cognition*. New York, NY, USA: Psychology Press.

Ateya, S. F., & Madhagi, E. A. (2012). On multivariate truncated generalized

Cauchy distribution. *Statistical Papers*. 54(3). 879–897. <https://doi.org/10.1007/s00362-012-0467-9>.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*. 12(2). 171–178. <https://ci.nii.ac.jp/naid/10020495032>.

Balakrishnan, N. (2013). *Handbook of the Logistic Distribution*. Boca Raton, FL, USA: CRC Press.

Banasaz, M. M. (2022). *Volatility Modeling and Risk Measurement using Statistical Models based on the Multivariate Student's t Distribution* [Doctoral dissertation, Virginia Polytechnic Institute and State University]. Blacksburg, Virginia, USA: Virginia Tech.

Barnarda, R. W., Trindade, A. A., & Wickramasinghe, R. I. P. (2014). Autoregressive moving average models under exponential power distributions. *ProbStatForum*. 7. 65–77.

Bartlett, M. S., & White, F. N. (1934). The vector representation of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*. 30(3). 327–340. <https://doi.org/10.1017/s0305004100012512>.

Bazi, Y., Bruzzone, L., & Melgani, F. (2007). Image thresholding based on the EM algorithm and the generalized Gaussian distribution. *Pattern Recognition*. 40(2). 619–634. <https://doi.org/10.1016/j.patcog.2006.05.006>.

Bennett, B. M., & Lindley, D. V. (1961). On a certain multivariate non-normal distribution. *Mathematical Proceedings of the Cambridge Philosophical Society*. 57(2). 434–436. <https://doi.org/10.1017/s030500410003543x>.

- Berkane, M., & Bentler, P. M. (1986). Moments of elliptically distributed random variates. *Statistics & Probability Letters*. 4(6). 333–335. [https://doi.org/10.1016/0167-7152\(86\)90054-4](https://doi.org/10.1016/0167-7152(86)90054-4).
- Berkane, M., & Bentler, P. M. (1987). Characterizing parameters of multivariate elliptical distributions. *Communications in Statistics: Simulation and Computation*. 16(1). 193–198. <https://doi.org/10.1080/03610918708812584>.
- Bicego, M., González-Jiménez, D., Grosso, E., & Alba-Castro, J. L. (2008). Generalized Gaussian distributions for sequential data classification. In *2008 19th International Conference on Pattern Recognition*. IEEE, Tampa, FL, USA. <https://doi.org/10.1109/icpr.2008.4761771>.
- Bingham, . H., & Kiesel, R. (2002). Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance*, 2(4), 241–250. <https://doi.org/10.1088/1469-7688/2/4/201>.
- Bishara, A. J., & Hittner, J. B. (2014). Reducing bias and error in the correlation coefficient due to nonnormality. *Educational and Psychological Measurement*. 75(5). 785–804. <https://doi.org/10.1177/0013164414557639>.
- Blake, I. F., & Thomas, J. B. (1968). On a class of processes arising in linear estimation theory. *IEEE Transactions on Information Theory*. 14(1). 12–16. <https://doi.org/10.1109/tit.1968.1054100>.
- Boedeker, P., & Kearns, N. T. (2019). Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer. *Advances in Methods and Practices in Psychological Science*. 2(3). 250–263. <https://doi.org/10.1177/2515245919854100>.

org/10.1177/2515245919849378.

Boisbunon, A. (2012). The class of multivariate spherically symmetric distributions. *Université de Rouen, Technical Report*. 5. 2012.

Bombrun, L., Pascal, F., Tournet, J., & Berthoumieu, Y. (2012). Performance of the maximum likelihood estimators for the parameters of multivariate generalized Gaussian distributions. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Kyoto, Jaapan. <https://doi.org/10.1109/icassp.2012.6288677>.

Boubchir, L., & Fadili, J. M. (2006). Multivariate statistical modeling of images with the curvelet transform. *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*. <https://doi.org/10.1109/isspa.2005.1581046>.

Bouhlel, N., & Dziri, A. (2019). Kullback–Leibler divergence between multivariate generalized Gaussian distributions. *IEEE Signal Processing Letters*. 26(7). 1021–1025. <https://doi.org/10.1109/lsp.2019.2915000>.

Bouhlel, N., & Rousseau, D. (2022). A Generic Formula and Some Special Cases for the Kullback–Leibler Divergence between Central Multivariate Cauchy Distributions. *Entropy*. 24(6). 838. <https://doi.org/10.3390/e24060838>.

Boukouvalas, Z., Fu, G., & Adali, T. (2015). An efficient multivariate generalized Gaussian distribution estimator: Application to IVA. *49th Annual Conference on Information Sciences and Systems (CISS)*. <https://doi.org/10.1109/ciss.2015.7086828>.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA, USA: Addison-Wesley Publishing.

Branco, M. D., & Dey, D. K. (2001). A general class of multivariate Skew-Elliptical distributions. *Journal of Multivariate Analysis*, 79(1), 99–113. <https://doi.org/10.1006/jmva.2000.1960>.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9), 2812–2831. <https://doi.org/10.1039/c3ay41907j>.

Brungard, C., Boettinger, J. L., Duniway, M. C., Wills, S., & Edwards, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239–240, 68–83. <https://doi.org/10.1016/j.geoderma.2014.09.019>.

Cacoullos, T. (2014). *Discriminant analysis and applications*. London, United Kingdom: Academic Press.

Cam, S. L., Belghith, A., Collet, C., & Salzenstein, F. (2009). Wheezing sounds detection using multivariate generalized gaussian distributions. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, Taipei, Taiwan. <https://doi.org/10.1109/icassp.2009.4959640>.

Cambanis, S., Huang, S. T., & Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3), 368–385. [https://doi.org/10.1016/0047-259x\(81\)90082-8](https://doi.org/10.1016/0047-259x(81)90082-8).

Campbell, T., & Broderick, T. (2019). Automated Scalable Bayesian Inference via Hilbert Coresets. *Journal of Machine Learning Research*, 20(15).

1–38. <http://jmlr.org/papers/volume20/17-613/17-613.pdf>.

Cannon, A. J. (2017). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*. 50(1–2). 31–49. <https://doi.org/10.1007/s00382-017-3580-6>.

Castro, L. M., Costa, D. R., Prates, M. O., & Lachos, V. H. (2014). Likelihood-based inference for Tobit confirmatory factor analysis using the multivariate Student-*t* distribution. *Statistics and Computing*. 25(6). 1163–1183. <https://doi.org/10.1007/s11222-014-9502-0>.

Chang, S., Yu, B., & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*. 9(9). 1532–1546. <https://doi.org/10.1109/83.862633>.

Chatfield, C. (2018). *Introduction to multivariate analysis*. New York, NY, USA: Routledge.

Chen, Y., Zhang, H., Zheng, Y., Jeon, B., & Wu, Q. M. J. (2016). An improved anisotropic hierarchical fuzzy c-means method based on multivariate student t-distribution for brain MRI segmentation. *Pattern Recognition*. 60. 778–792. <https://doi.org/10.1016/j.patcog.2016.06.020>.

Chen, Z., Wang, B., & Gorban, A. N. (2019). Multivariate Gaussian and Student-*t* process regression for multi-output prediction. *Neural Computing and Applications*. 32(8). 3005–3028. <https://doi.org/10.1007/s00521-019-04687-8>.

Cheng, P. (1987). The limit distribution of empirical distribution of random.

Chinese Journal of Applied Probability and Statistics. 3. 8–20.

Cheng, S. W., & Spiring, F. (1987). A Test to Identify the Uniform Distribution, with Applications to Probability Plotting and Other Distributions. *IEEE Transactions on Reliability*. 36(1). 98–105. <https://doi.org/10.1109/tr.1987.5222309>.

Chitour, Y., & Pascal, F. (2008). Exact maximum likelihood estimates for SIRV Covariance Matrix: existence and Algorithm analysis. *IEEE Transactions on Signal Processing*. 56(10). 4563–4573. <https://doi.org/10.1109/tsp.2008.927464>.

Chmielewski, M. A. (1981). Elliptically Symmetric Distributions: A Review and Bibliography. *International Statistical Review*. 49(1). 67. <https://doi.org/10.2307/1403038>.

Cho, D., & Bui, T. D. (2005). Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Processing: Image Communication*. 20(1). 77–89. <https://doi.org/10.1016/j.image.2004.10.003>.

Chu, K. (1973). Estimation and decision for linear systems with elliptical random processes. *IEEE Transactions on Automatic Control*. 18(5). 499–505. <https://doi.org/10.1109/tac.1973.1100374>.

Cleff, T. (2019). *Applied statistics and multivariate data analysis for business and economics: A Modern Approach Using SPSS, Stata, and Excel*. Cham, Switzerland: Springer.

Coban, M., & Mersereau, R. (1996). Adaptive subband video coding using bivariate generalized Gaussian distribution model. In *1996 IEEE Interna-*

tional Conference on Acoustics, Speech, and Signal Processing Conference Proceedings. IEEE, Atlanta, GA, USA. <https://doi.org/10.1109/icaassp.1996.544844>.

Coin, D. (2017). A goodness-of-fit test for generalized error distribution. *Communications in Statistics*. 46(23). 11485–11499. <https://doi.org/10.1080/03610926.2016.1271426>.

Cox, D. (1966). A simple example of a comparison involving quanta. *Biometrika*. 53(1–2). 213–220. <https://doi.org/10.1093/biomet/53.1-2.213>.

Ct  roux, R., & Ducharme, G. R. (1989). Vector correlation for elliptical distributions. *Communications in Statistics - Theory and Methods*. 18(4). 1441–1454. <https://doi.org/10.1080/03610928908829977>.

Dadi, M. I., & Marks, R. J. (1987). Detector relative efficiencies in the presence of laplace noise. *IEEE Transactions on Aerospace and Electronic Systems (AES)*. 23(4). 568–582. <https://doi.org/10.1109/taes.1987.310890>.

Dang, U. (2014). *Mixtures of power exponential distributions and topics in regression-based mixture models* [Doctoral thesis, University of Guelph]. Guelph, Ontario, Canada: University of Guelph.

Dang, U. J., Browne, R. P., & McNicholas, P. D. (2015). Mixtures of multivariate power exponential distributions. *Biometrics*. 71(4). 1081–1089. <https://doi.org/10.1111/biom.12351>.

Day, N. E., & Kerridge, D. F. (1967). A general maximum likelihood discriminant. *Biometrics*. 23(2). 313–323. <https://doi.org/10.2307/2528164>.

De Andrade, L. R., Ferreira, D. F., S  fadi, T., & Barroso, L. P. (2018). Bayesian

analysis of dynamic factor models using multivariate T distribution.

Brazilian Journal of Biometrics. 36(1). 140–156. <https://doi.org/10.28951/rbb.v36i1.155>.

De Figueiredo, L. P., Grana, D., Santos, M., Figueiredo, W., Roisenberg, M., & Neto, G. S. (2017). Bayesian seismic inversion based on rock-physics

prior modeling for the joint estimation of acoustic impedance, porosity and lithofacies. *Journal of Computational Physics*. 336. 128–142. <https://doi.org/10.1016/j.jcp.2017.02.013>.

Dehghani, H., & Fadaee, M. J. (2020). Probabilistic prediction of earthquake by bivariate distribution. *Asian Journal of Civil Engineering*. 21(6). 977–983. <https://doi.org/10.1007/s42107-020-00254-y>.

Dempster, A. P. (1969). *Elements of continuous multivariate analysis*. Reading, MA, USA: Addison-Wesley.

Desai, M. A., & Mangoubi, R. (2003). Robust gaussian and non-gaussian matched subspace detection. *IEEE Transactions on Signal Processing*. 51(12). 3115–3127. <https://doi.org/10.1109/tsp.2003.818907>.

Di Crescenzo, A., Paraggio, P., Román-Román, P., & Torres-Ruiz, F. (2022). Statistical analysis and first-passage-time applications of a lognormal diffusion process with multi-sigmoidal logistic mean. *Statistical Papers*. 64(5). 1391–1438. <https://doi.org/10.1007/s00362-022-01349-1>.

Diaconis, P., & Freedman, D. A. (1986). On the Consistency of Bayes Estimates. *Annals of Statistics*. 14(1). <https://doi.org/10.1214/aos/1176349830>.

- Diananda, P. H. (1949). Note on some properties of maximum likelihood estimates. *Mathematical Proceedings of the Cambridge Philosophical Society*. 45(4). <https://doi.org/10.1017/S0305004100025238>.
- DiBmann, J., Brechmann, E. C., Czado, C., & Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*. 59. 52–69. <https://doi.org/10.1016/j.csda.2012.08.010>.
- Dickey, J. M., & Chen, C. H. (1985). Direct subjective-probability modelling using ellipsoidal distributions. *Bayesian Statistics*. 2. 157–182.
- Dickey, J. M., & Chen, C. H. (1987). Representations of elliptical distributions and quadratic forms. *J. Chinese Statist. Assoc.* 25. 33–42.
- Dimitrov, N. B., & Meyers, L. A. (2010). Mathematical Approaches to Infectious Disease Prediction and Control. In *INFORMS TutORials in Operations Research*. 1–25. <https://doi.org/10.1287/educ.1100.0075>.
- Diaz-Garcia, J. A., & Jáimez, R. G. (2012). Matricvariate and matrix multivariate Pearson type II distributions and related distributions: theory and methods. *South African Statistical Journal*. 46(1). 31–52. <https://journals.co.za/content/sasj/46/1/EJC99146>.
- Do, M., & Vetterli, M. (2002). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*. 11(2). 146–158. <https://doi.org/10.1109/83.982822>.
- Dray, S., & Josse, J. (2014). Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology*. 216(5). 657–667.

<https://doi.org/10.1007/s11258-014-0406-z>.

Duttweiler, D., & Messerschmitt, D. G. (1976). Nearly instantaneous companding for nonuniformly quantized PCM. *IRE Transactions on Communications Systems*. 24(8). 864–873. <https://doi.org/10.1109/tcom.1976.1093389>.

Dytso, A., Bustin, R., Poor, H. V., & Shamai, S. (2018). Analytical properties of generalized Gaussian distributions. *Journal of Statistical Distributions and Applications*. 5(1). <https://doi.org/10.1186/s40488-018-0088-5>.

Eaton, M. L. (1986). A characterization of spherical distributions. *Journal of Multivariate Analysis*. 20(2). 272–276. [https://doi.org/10.1016/0047-259x\(86\)90083-7](https://doi.org/10.1016/0047-259x(86)90083-7).

Elguebaly, T., & Bouguila, N. (2010). Bayesian learning of generalized Gaussian mixture models on biomedical images. In *Lecture Notes in Computer Science* (pp. 207–218). https://doi.org/10.1007/978-3-642-12159-3_19.

Eltoft, T., Kim, T., & Lee, T. (2006a). Multivariate scale mixture of Gaussians modeling. In Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds) *Independent Component Analysis and Blind Signal Separation* (ICA 2006. Lecture Notes in Computer Science, Vol. 3889). Springer. https://doi.org/10.1007/11679363_99.

Eltoft, T., Kim, T., & Lee, T. (2006b). On the multivariate Laplace distribution. *IEEE Signal Processing Letters*. 13(5). 300–303. <https://doi.org/10.1109/lsp.2006.870353>.

- Epstein, B. (1948). Statistical aspects of fracture problems. *Journal of Applied Physics*. 19(2). 140–147. <https://doi.org/10.1063/1.1698380>.
- Everitt, B., & Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. New York, NY, USA: Springer Science & Business Media.
- Fang, K., & Anderson, T. W. (1992). Statistical inference in elliptically contoured and related distributions. *Journal of the American Statistical Association*. 87(417). 253. <https://doi.org/10.2307/2290483>.
- Fang, K., Kotz, S., & Ng, K. W. (1991). Symmetric multivariate and related distributions. *Journal of the American Statistical Association*. 86(416). 1144. <https://doi.org/10.2307/2290544>.
- Fang, K., & Zhang, Y. (1990). *Generalized Multivariate analysis*. Dordrecht, Netherlands: Springer. <https://ci.nii.ac.jp/ncid/BA11294315>.
- Felícíimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E. (2012). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides*. 10(2). 175–189. <https://doi.org/10.1007/s10346-012-0320-1>.
- Ferguson, T. S. (1962). A representation of the symmetric bivariate Cauchy distribution. *Annals of Mathematical Statistics*. 33(4). 1256–1266. <https://doi.org/10.1214/aoms/1177704357>.
- Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical distributions*. Hoboken, NJ, USA: John Wiley & Sons.
- Fragiadakis, K., & Meintanis, S. G. (2011). Goodness-of-fit tests for multi-

variate Laplace distributions. *Mathematical and Computer Modelling*. 53(5–6). 769–779. <https://doi.org/10.1016/j.mcm.2010.10.014>.

Frahm, G., Junker, M., & Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*. 63(3). 275–286. [https://doi.org/10.1016/s0167-7152\(03\)00092-0](https://doi.org/10.1016/s0167-7152(03)00092-0).

Fyodorov, Y. V., & Keating, J. P. (2014). Freezing transitions and extreme values: random matrix theory, and disordered landscapes. *Philosophical Transactions of the Royal Society A*. 372(2007). 20120503. <https://doi.org/10.1098/rsta.2012.0503>.

Garay, A. M., Lachos, V. H., Bolfarine, H., & Cabral, C. R. B. (2015). Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*. 58(1). 247–278. <https://doi.org/10.1007/s00362-015-0696-9>.

Genç, A. İ. (2007). A generalization of the univariate slash by a Scale-Mixed exponential power distribution. *Communications in Statistics - Simulation and Computation*. 36(5). 937–947. <https://doi.org/10.1080/03610910701539161>.

Giller, G. (2024). Optimal asset allocation with a multivariate LAPlace distribution: an analytical result 24 years in the making. *Adventures in Data Science*. Retrieved June 19, 2024, from <https://medium.gillerinvestments.com/optimal-asset-allocation-with-a-multivariate-laplace-distribution-CBB207b05d81>.

Giller, G. L. (2005). A generalized error distribution. *Social Science Research*

Network. <https://doi.org/10.2139/ssrn.2265027>.

Givens, G. H., & Hoeting, J. A. (2012). *Computational statistics*. Hoboken, NJ, USA: John Wiley & Sons.

Glonek, G., & McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society Series B-methodological*. 57(3). 533–546. <https://doi.org/10.1111/j.2517-6161.1995.tb02046.x>.

Godwin, A. (2016). The Development of a Measure of Engineering Identity. *ASEE Annual Conference & Exposition*. <https://doi.org/10.18260/p.26122>.

Gómez, E., Gomez-Villegas, M., & Marín, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics*, 27(3), 589–600. <https://doi.org/10.1080/03610929808832115>.

Gómez-Sánchez-Manzano, E., Gómez-Villegas, M. A., & Marín, J. M. (2008). Multivariate Exponential Power Distributions as Mixtures of Normal Distributions with Bayesian Applications. *Communications in Statistics*. 37(6). 972–985. <https://doi.org/10.1080/03610920701762754>.

Gradshteyn, I. S., & Ryzhik, I. M. (2014). *Table of integrals, series, and products*. London, United Kingdom: Academic Press.

Guardiola, J. (2020). The spherical-Dirichlet distribution. *Journal of Statistical Distributions and Applications*. 7(1). <https://doi.org/10.1186/s40488-020-00106-9>.

Gumbel, E. J. (1961). Bivariate Logistic distributions. *Journal of the American*

Statistical Association. 56(294). 335–349. <https://doi.org/10.1080/01621459.1961.10482117>.

Guo, Y., & Bondell, H. (2023). On robust probabilistic principal component analysis using multivariate t-distributions. *Communications in Statistics - Theory and Methods*. 52(23). 8261–8279. <https://doi.org/10.1080/03610926.2022.2060512>.

Gupta, A. K., & Varga, T. (2012). *Elliptically contoured models in statistics*. New York, NY, USA: Springer Science & Business Media.

Hartman, P., & Wintner, A. (1940). On the Spherical Approach to the Normal Distribution Law. *American Journal of Mathematics*. 62(1/4). 759. <https://doi.org/10.2307/2371486>.

Hristova, H., Le Meur, O., Cozot, R., & Bouatouch, K. (2017). Transformation of the multivariate generalized Gaussian distribution for image editing. *IEEE Transactions on Visualization and Computer Graphics*. 24(10). 2813–2826. <https://doi.org/10.1109/tvcg.2017.2769050>.

Ikeshita, R., & Kawaguchi, Y. (2018). Independent Low-Rank Matrix Analysis Based on Multivariate Complex Exponential Power Distribution. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Calgary, AB, Canada. <https://doi.org/10.1109/icassp.2018.8461387>.

Iliopoulos, G., & Karlis, D. (2003). Simulation from the bessel distribution with applications. *Journal of Statistical Computation and Simulation*. 73(7). 491–506. <https://doi.org/10.1080/0094965031000123124>.

- Itano, K., Ueki, K., Iizuka, T., & Kuwatani, T. (2020). Geochemical discrimination of monazite source Rock based on machine learning techniques and multinomial logistic regression analysis. *Geosciences*. 10(2). 63. <https://doi.org/10.3390/geosciences10020063>.
- Jammalamadaka, S. R., Sengupta, A., & Sengupta, A. (2001). *Topics in circular statistics*. Singapore: World Scientific.
- Jayakumar, G. D. S., Sulthan, A., & Samuel, W. (2022). An entropy divergence, D-distribution and its selected application in multivariate financial time series. *International Journal of Financial Services Management*. 11(2). 134–148. <https://doi.org/10.1504/IJFSM.2021.120359>.
- Jiang, J. (1988). Starlike functions and linear functions of a dirichlet distributed vector. *Siam Journal on Mathematical Analysis*. 19(2). 390–397. <https://doi.org/10.1137/0519029>.
- Joarder, A. H. (1997). On the characteristic function of the multivariate Pearson type II distribution. *Journal of Information and Optimization Sciences*. 18(1). 177–182. <https://doi.org/10.1080/02522667.1997.10699322>.
- Johnson, M. E. (1987). *Multivariate Statistical simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions*. Hoboken, NJ, USA: John Wiley & Sons.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2*. Hoboken, NJ, USA: John Wiley & Sons.
- Johnson, R. A., & Wichern, D. W. (2013). *Applied Multivariate Statistical analysis: Pearson New International Edition*. Boston, MA, USA: Pearson

Higher Ed.

Kalke, S., & Richter, W. (2013). Simulation of the p-generalized Gaussian distribution. *Journal of Statistical Computation and Simulation*. 83(4). 641–667. <https://doi.org/10.1080/00949655.2011.631187>.

Kamdem, J. S. (2005). Value-at-risk and expected shortfall for linear portfolios with elliptically distributed risk factors. *International Journal of Theoretical and Applied Finance*. 08(05). 537–551. <https://doi.org/10.1142/s0219024905003104>.

Kamdem, J. S. (2009). -VaR and -TVaR for portfolios with mixture of elliptic distributions risk factors and DCC. *Insurance: Mathematics and Economics*. 44(3). 325–336. <https://doi.org/10.1016/j.insmatheco.2008.11.013>.

Kan, R., & Zhou, G. (2017). Modeling non-normality using multivariate t: implications for asset pricing. *China Finance Review International*. 7(1). 2–32. <https://doi.org/10.1108/cfri-10-2016-0114>.

Karamikabir, H., & Afshari, M. (2019). Wavelet shrinkage generalized Bayes estimation for elliptical distribution parameter's under LINEX loss. *International Journal of Wavelets, Multiresolution and Information Processing*. 17(03). 1950009. <https://doi.org/10.1142/s0219691319500097>.

Karian, Z. A., & Dudewicz, E. J. (2016). *Handbook of Fitting Statistical Distributions with R*. Boca Raton, FL, USA: CRC Press. <https://doi.org/10.1201/b10159>.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-

- scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*. 32(4). 419–430. <https://www.jstor.org/stable/25049690>.
- Kent, J. T., Ganeiber, A. M., & Mardia, K. V. (2018). A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics*. 27(2). 291–301. <https://doi.org/10.1080/10618600.2017.1390468>.
- Khalafi, M., & Azimmohseni, M. (2014). Multivariate Pearson type II distribution: statistical and mathematical features. *Probability and Mathematical Statistics*. 34. 119–126. <http://math.uni.wroc.pl/pms/files/34.1/Article/34.1.7.pdf>.
- Khelil-Cherif, N., & Benazza-Benyahia, A. (2004). Wavelet-based multivariate approach for multispectral image indexing. In *SPIE Conference on Wavelet Applications in Industrial Processing* (Vol. 19). Philadelphia, PE, USA.
- Kleiber, C., & Stoyanov, J. (2013). Multivariate distributions and the moment problem. *Journal of Multivariate Analysis*. 113. 7–18. <https://doi.org/10.1016/j.jmva.2011.06.001>.
- Klein, G. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision Making in Action: Models and Methods*. 5(4). 138–147. <http://psycnet.apa.org/record/1993-97634-006>.
- Kollo, T., & Srivastava, M. S. (2005). Estimation and testing of parameters in multivariate Laplace distribution. *Communications in Statistics - Theory and Methods*. 33(10). 2363–2387. <https://doi.org/10.1081/sta-200031>

408.

Korolev, V. Y. (2020). Some properties of univariate and multivariate exponential power distributions and related topics. *Mathematics*. 8(11). 1918. <https://doi.org/10.3390/math8111918>.

Korotcov, A., Tkachenko, V., Russo, D. P., & Ekins, S. (2017). Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Molecular Pharmaceutics*. 14(12). 4462–4475. <https://doi.org/10.1021/acs.molpharmaceut.7b00578>.

Kotz, S. (1975). Multivariate distributions at a cross road. In *Springer eBooks* (pp. 247–270). https://doi.org/10.1007/978-94-010-1842-5_20.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). *Continuous Multivariate Distributions: Models and Applications* (2nd ed., Vol. 1). Hoboken, NJ, USA: John Wiley & Sons.

Kotz, S., Kozubowski, T. J., & Podgórski, K. (2001). Asymmetric multivariate laplace distribution. In *Birkhäuser Boston eBooks* (pp. 239–272). https://doi.org/10.1007/978-1-4612-0173-1_7.

Kotz, S., Kozubowski, T., & Podgorski, K. (2012). *The LAPlace Distribution and generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. New York, NY, USA: Springer Science & Business Media.

Kotz, S., & Nadarajah, S. (2002). Letter to the Editor of Biometrics: A novel application of the multivariate power exponential distribution to repeated measurements from a crossover trial on insulin applied to rabbits. *Bio-*

metrics. 58(3). 689–690. <https://doi.org/10.1111/j.0006-341x.2002.00689.x>.

Kotz, S., & Nadarajah, S. (2004). *Multivariate T-Distributions and their applications*. Cambridge, United Kingdom: Cambridge University Press.

Koutras, M. V. (1986). On the generalized noncentral chi-squared distribution induced by an elliptical gamma law. *Biometrika*. 73(2). 528–532. <https://doi.org/10.1093/biomet/73.2.528>.

Kozubowski, T. J. (2002). On the vertical density of the multivariate exponential power distribution. *Statistics*. 36(3). 219–221. <https://doi.org/10.1080/02331880212858>.

Kozubowski, T. J., Podgórski, K., & Rychlik, I. (2013). Multivariate generalized Laplace distribution and related random fields. *Journal of Multivariate Analysis*. 113. 59–72. <https://doi.org/10.1016/j.jmva.2012.02.010>.

Krishnamoorthy, K. (2006). Handbook of Statistical Distributions with Applications. In *Chapman and Hall/CRC eBooks*. <https://doi.org/10.1201/9781420011371>.

Kwitt, R., Meerwald, P., & Uhl, A. (2009). Color-image watermarking using multivariate power-exponential distribution. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, Cairo, Egypt. <https://doi.org/10.1109/icip.2009.5413715>.

Lahcene, B. (2013). On Pearson families of distributions and its applications. *African Journal of Mathematics and Computer Science Research*. 6(5). 108–117. <https://doi.org/10.5897/ajmcsr2013.0465>.

- Lan, S., Holbrook, A. J., Elias, G. A., Fortin, N. J., Ombao, H., & Shahbaba, B. (2020). Flexible Bayesian dynamic modeling of correlation and covariance matrices. *Bayesian Analysis*. 15(4). <https://doi.org/10.1214/19-ba1173>
- Landsman, Z., & Valdez, F. E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4), 55–71. <https://doi.org/10.1080/10920277.2003.10596118>.
- Lee, C. (2015). Fast simulated annealing with a multivariate Cauchy distribution and the configuration's initial temperature. *Journal of the Korean Physical Society*. 66(10). 1457–1466. <https://doi.org/10.3938/jkps.66.1457>.
- Lee, H., Park, H., & Kim, H. (2014). A clarification of the cauchy distribution. *Communications for Statistical Applications and Methods*. 21(2). 183–191. <https://doi.org/10.5351/csam.2014.21.2.183>.
- Leemis, L. M., & McQueston, J. (2008). Univariate distribution relationships. *The American Statistician*. 62(1). 45–53. <https://doi.org/10.1198/000313008x270448>.
- Lhéritier, A., Bocamazo, M., Delahaye, T., & Acuña-Agost, R. (2019). Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling*. 31. 198–209. <https://doi.org/10.1016/j.jocm.2018.02.002>.
- Li, D., Chen, D., Goh, J., & Ng, S. (2018). Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series. *arXiv (Cornell University)*. <http://export.arxiv.org/pdf/1809.04758>.
- Li, G., & Peng, X. (2020). A review of Prof. Kai-Tai Fang's contribution

to the education, promotion, and advancement of statistics in China. In *Springer eBooks* (pp. 49–65). https://doi.org/10.1007/978-3-030-46161-4_4.

Li, Y., Cao, W., Hu, W., Gan, C., & Wu, M. (2020). Fault detection for geological drilling processes using multivariate generalized Gaussian distribution and Kullback Leibler divergence. *IFAC-PapersOnLine*. 53(2). 164–169. <https://doi.org/10.1016/j.ifacol.2020.12.115>.

Li, Y., Cao, W., Hu, W., Xiong, Y., & Wu, M. (2021). Incipient fault detection for geological drilling processes using multivariate generalized Gaussian distributions and Kullback–Leibler divergence. *Control Engineering Practice*. 117. 104937. <https://doi.org/10.1016/j.conengprac.2021.104937>.

Liang, J., & Bentler, P. M. (1998). Characterizations of some subclasses of spherical distributions. *Statistics & Probability Letters*. 40(2). 155–164. [https://doi.org/10.1016/s0167-7152\(98\)00084-4](https://doi.org/10.1016/s0167-7152(98)00084-4).

Lindholm, A., Wahlström, N., Lindsten, F., & Schön, T. B. (2022). *Machine learning: A First Course for Engineers and Scientists*. Cambridge, United Kingdom: Cambridge University Press.

Lindsey, J. K. (1999). Multivariate elliptically contoured distributions for repeated measurements. *Biometrics*. 55(4). 1277–1280. <https://doi.org/10.1111/j.0006-341x.1999.01277.x>.

Liu, M., & Bozdogan, H. (2008). Multivariate Regression Models with Power Exponential Random Errors and Subset Selection Using Genetic Algo-

rithms With Information Complexity. *European Journal of Pure and Applied Mathematics*. 1(1). 4–37. <https://www.ejpam.com/index.php/ejpam/article/view/86>.

Madan, D. B. (2020). Multivariate distributions for financial returns. *International Journal of Theoretical and Applied Finance*. 23(06). 2050041. <https://doi.org/10.1142/s0219024920500417>.

Maghrebi, H., Portigliatti, T., & Prouff, E. (2016). Breaking cryptographic implementations using deep learning techniques. In *Lecture Notes in Computer Science* (pp. 3–26). https://doi.org/10.1007/978-3-319-49445-6_1.

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. *The European Symposium on Artificial Neural Networks*. <http://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2015-56.pdf>.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 57(3). 519–530. <https://doi.org/10.1093/biomet/57.3.519>.

Marks, R. J., Wise, G. L., Haldeman, D., & Whited, J. L. (1978). Detection in Laplace noise. *IEEE Transactions on Aerospace and Electronic Systems, AES*. 14(6). 866–872. <https://doi.org/10.1109/taes.1978.308550>.

Marnissi, Y., Benazza-Benyahia, A., Chouzenoux, É., & Pesquet, J. (2013). Generalized multivariate exponential power prior for wavelet-based multichannel image restoration. In *2013 IEEE International Conference on*

Image Processing. IEEE, Melbourne, VIC, Australia. <https://doi.org/10.1109/icip.2013.6738495>.

Maxwell, J. C. (1986). *Maxwell on Molecules and Gases*. Cambridge, MA, USA: MIT Press.

McDonald, J. B., & Newey, W. K. (1988). Partially Adaptive Estimation of Regression Models via the Generalized t Distribution. *Econometric Theory*. 4(3). 428–457. <https://doi.org/10.1017/S0266466600013384>.

McGraw, D., & Wagner, J. (1968). Elliptically symmetric distributions. *IEEE Transactions on Information Theory*. 14(1). 110–120. <https://doi.org/10.1109/tit.1968.1054081>.

Mecklin, C. J., & Mundfrom, D. J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*. 75(2). 93–107. <https://doi.org/10.1080/0094965042000193233>.

Mehar, M., Mittal, S., & Prasad, N. (2016). Farmers coping strategies for climate shock: Is it differentiated by gender? *Journal of Rural Studies*. 44. 123–131. <https://doi.org/10.1016/j.jrurstud.2016.01.001>.

Mineo, A., & Ruggieri, M. (2005). A Software Tool for the Exponential Power Distribution: Thenormalppackage. *Journal of Statistical Software*. 12(4). <https://doi.org/10.18637/jss.v012.i04>.

Mitropoulos, P., Talias, . A., & Mitropoulos, I. (2015). Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals. *European Jour-*

Journal of Operational Research. 243(1). 302–311. <https://doi.org/10.1016/j.ejor.2014.11.012>.

Mittal, S., & Mehar, M. (2015). Socio-economic factors affecting adoption of modern information and communication technology by farmers in India: analysis using multivariate Probit model. *The Journal of Agricultural Education and Extension*. 22(2). 199–212. <https://doi.org/10.1080/1389224x.2014.997255>.

Morgan, J. P. (1996). *Risk Metrics technical document* (4th ed.). New York, NY, USA: Morgan Guaranty Trust Company.

Mott, J. H. (2018). Estimation of aircraft distances using transponder signal strength information. *Cogent Engineering*. 5(1). 1466619. <https://doi.org/10.1080/23311916.2018.1466619>.

Moulin, P., & Liu, J. (1999). Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Transactions on Information Theory*. 45(3). 909–919. <https://doi.org/10.1109/18.761332>.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. Hoboken, NJ, USA: John Wiley & Sons.

Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. Cambridge, MA, USA: MIT Press.

Muzellec, B., & Cuturi, M. (2018). Generalizing Point Embeddings using the Wasserstein Space of Elliptical Distributions. *arXiv (Cornell University)*. 31. 10237–10248. <https://arxiv.org/pdf/1805.07594.pdf>.

- Nadarajah, S. (2003). The Kotz-type distribution with applications. *Statistics*. 37(4). 341–358. <https://doi.org/10.1080/0233188031000078060>.
- Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*. 32(7). 685–694. <https://doi.org/10.1080/02664760500079464>.
- Nadarajah, S. (2008). A new model for symmetric and skewed data. *Probability in the Engineering and Informational Sciences*. 22(2). 261–271. <https://doi.org/10.1017/s0269964808000156>.
- Najar, F., Bourouis, S., Al-Azawi, R., & Al-Badi, A. H. (2019). Online recognition via a finite mixture of multivariate generalized Gaussian distributions. In *Unsupervised and semi-supervised learning*. https://doi.org/10.1007/978-3-030-23876-6_5.
- Nayebi, A., & Ganguli, S. (2017). Biologically inspired protection of deep networks from adversarial attacks. *arXiv (Cornell University)*. <http://export.arxiv.org/pdf/1703.09202>.
- Ng, K. W., Tian, G., & Tang, M. (2011). *Dirichlet and related distributions: Theory, Methods and Applications*. Hoboken, NJ, USA: John Wiley & Sons.
- Nguyen, T. H. A., Ruiz-Gazen, A., Thomas-Agnan, C., & Laurent, T. (2019). Multivariate Student versus Multivariate Gaussian Regression Models with Application to Finance. *Journal of Risk and Financial Management*. 12(1). 28. <https://doi.org/10.3390/jrfm12010028>.
- Nkansah, B. K., Manford, M., & Howard, N. (2018). On the Exponential Power Distribution. *Journal of Mathematical Theory and Modeling*. 8(8).

121–140. <https://core.ac.uk/download/pdf/234680621.pdf>.

Nkemnole, E. B., & Abass, O. (2015). A t-distribution based particle filter for univariate and multivariate stochastic volatility models. *Journal of the Nigerian Mathematical Society*. 34(2). 227–242. <https://doi.org/10.1016/j.jnnms.2014.11.002>.

Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*. 28(5). 2067–2089. <https://doi.org/10.1007/s00180-013-0396-7>.

Noorae, N., Abegaz, F., Ormel, J., Wit, E., & Van Den Heuvel, E. R. (2015). An approximate marginal logistic distribution for the analysis of longitudinal ordinal data. *Biometrics*. 72(1). 253–261. <https://doi.org/10.1111/biom.12414>.

Novey, M., Adali, T., & Roy, A. (2010). A complex generalized Gaussian distribution—characterization, generation, and estimation. *IEEE Transactions on Signal Processing*. 58(3). 1427–1433. <https://doi.org/10.1109/tsp.2009.2036049>.

Olkin, I., & Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *The Annals of Mathematical Statistics*. 35(1). 261–269. <https://doi.org/10.1214/aoms/1177703748>.

Ollila, E., Tyler, D. E., Koivunen, V., & Poor, H. V. (2012). Complex elliptically symmetric distributions: survey, new results and applications. *IEEE Transactions on Signal Processing*. 60(11). 5597–5625. <https://doi.org/>

10.1109/tsp.2012.2212433.

Olósundé, A. A. (2013). On Exponential Power Distribution And Poultry Feeds

Data: A Case Study. *Journal of the Iranian Statistical Society*. 12(2).

253–270. <http://jirss.irstat.ir/article-1-232-en.pdf>.

Olosunde, A. A., & Soyinka, A. T. (2013). Discrimination and Classification

of Poultry Feeds data. *International Journal of Mathematical Research*.

2(5). 37–44. <https://doi.org/10.18488/journal.24/2013.2.5/24.5.37.44>.

Olosunde, A. A., & Soyinka, A. T. (2020). Discrimination and classification

model from multivariate exponential power distribution. *Electronic Jour-*

nal of Applied Statistical Analysis. 13(2). 284–292. [https://doi.org/10.](https://doi.org/10.1285/i20705948v13n2p284)

1285/i20705948v13n2p284.

Omari, M., Abdelouahad, A. A., Hassouni, M. E., & Cherifi, H. (2013). Color

Image Quality Assessment Measure Using Multivariate Generalized

Gaussian Distribution. In *2013 International Conference on Signal-*

Image Technology & Internet-Based Systems. IEEE, Kyoto, Japan. [https:](https://doi.org/10.1109/sitis.2013.42)

[//doi.org/10.1109/sitis.2013.42](https://doi.org/10.1109/sitis.2013.42).

Onaseso, O., & Olosunde, A. (2017). *Discriminant analysis of data under ex-*

ponential power distribution. Saarbrücken, Germany: LAP LAMBERT

Academic Publishing. 978-620-2-02785-4.

Osnas, J. L. D., Lichstein, J. W., Reich, P. B., & Pacala, S. W. (2013). Global

leaf trait relationships: mass, area, and the leaf Economics spectrum.

Science. 340(6133). 741–744. <https://doi.org/10.1126/science.1231574>.

Ouzir, N., Pesquet, J.-C., & Pascal, F. (2022). A Convex Formulation for the

Robust Estimation of Multivariate Exponential Power Models. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://doi.org/10.1109/ICASSP43922.2022.9747354>.

Park, S. Y., & Bera, A. K. (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*. 150(2). 219–230. <https://doi.org/10.1016/j.jeconom.2008.12.014>

Pascal, F., Bombrun, L., Tournet, J., & Berthoumieu, Y. (2013). Parameter Estimation For Multivariate Generalized Gaussian Distributions. *IEEE Transactions on Signal Processing*. 61(23). 5960–5971. <https://doi.org/10.1109/tsp.2013.2282909>.

Pascal, F., Chitour, Y., Ovarlez, J., Forster, P., & Larzabal, P. (2008). Covariance Structure Maximum-Likelihood Estimates in Compound Gaussian Noise: Existence and Algorithm Analysis. *IEEE Transactions on Signal Processing*. 56(1). 34–48. <https://doi.org/10.1109/tsp.2007.901652>.

Pearson, E. S. (1963). Some problems arising in approximating to probability distributions, using moments. *Biometrika*. 50(1–2). 95–112. <https://doi.org/10.1093/biomet/50.1-2.95>.

Pearson, K. (1916). On the Application of Goodness-of-Fit Tables to Test Regression Curves and Theoretical Curves Used to Describe Observational or Experimental Data. *Biometrika*. 11(3). 239–261. <https://doi.org/10.1093/biomet/11.3.239>.

Pei, W., Dibeklioglu, H., Tax, D. M. J., & Van Der Maaten, L. (2018). Multi-

variate Time-Series classification using the Hidden-Unit Logistic model. *IEEE Transactions on Neural Networks and Learning Systems*. 29(4). 920–931. <https://doi.org/10.1109/tnnls.2017.2651018>.

Pesaran, M. H. (2015). *Time series and panel data econometrics*. New York, NY, USA: Oxford University Press.

Peterson, C. B., Stingo, F. C., & Vannucci, M. (2015). Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*. 110(509). 159–174. <https://doi.org/10.1080/01621459.2014.896806>.

Phoon, K. K. (2019). The story of statistics in geotechnical engineering. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. 14(1). 3–25. <https://doi.org/10.1080/17499518.2019.1700423>.

Plungpongpun, K. (2003). *Analysis of multivariate data using KOTZ type distribution* [Doctoral dissertation, Old Dominion University]. Norfolk, Virginia, USA: Old Dominion University.

Pogány, T. K., & Nadarajah, S. (2009). On the characteristic function of the generalized normal distribution. *Comptes Rendus Mathématique*. 348(3–4). 203–206. <https://doi.org/10.1016/j.crma.2009.12.010>.

Poincaré, H. (1912). *Calcul des probabilités*. Paris, France: Gauthier-Villars.

Provis, J. L., Palomo, A., & Shi, C. (2015). Advances in understanding alkali-activated materials. *Cement and Concrete Research*. 78. 110–125. <https://doi.org/10.1016/j.cemconres.2015.04.013>.

Provost, S. B., & Cheong, Y. (2002). The distribution of Hermitian quadratic

forms in elliptically contoured random vectors. *Journal of Statistical Planning and Inference*. 102(2). 303–316. [https://doi.org/10.1016/s0378-3758\(01\)00096-9](https://doi.org/10.1016/s0378-3758(01)00096-9).

Provosta, S. B., Zareamoghaddam, H., Ahmed, S. E., & Ha, H.-T. (2022). The generalized Pearson family of distributions and explicit representation of the associated density functions. *Communications in Statistics-Theory and Methods*. 51(16). 5590–5606. <https://doi.org/10.1080/03610926.2020.1843680>.

Punzo, A., & Bagnato, L. (2020). The multivariate tail-inflated normal distribution and its application in finance. *Journal of Statistical Computation and Simulation*. 91(1). 1–36. <https://doi.org/10.1080/00949655.2020.1805451>.

Purczyński, J., & Bednarz-Okrzyńska, K. (2014). Estimation of the shape parameter of GED distribution for a small sample size. *Folia Oeconomica Stetinensia*. 14(1). 35–46. <https://doi.org/10.2478/fole-2014-0103>.

Rabbani, H., Vafadust, M., Gazor, S., & Selesnick, I. (2006). Image Denoising Employing a Bivariate Cauchy Distribution with Local Variance in Complex Wavelet Domain. In *2006 IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop*. IEEE. <https://doi.org/10.1109/dspws.2006.265407>.

Rahnamaei, Z., Nematollahi, N., & Farnoosh, R. (2012). The Location-Scale Mixture Exponential Power Distribution: A Bayesian and Maximum Likelihood Approach. *Journal of Applied Mathematics*. 2012. 978282.

<https://doi.org/10.1155/2012/978282>.

Rangaswamy, M., Weiner, D., & Öztürk, A. (1993). Non-Gaussian random vector identification using spherically invariant random processes. *IEEE Transactions on Aerospace and Electronic Systems*. 29(1). 111–124. <https://doi.org/10.1109/7.249117>.

Rangaswamy, M., Weiner, D., & Öztürk, A. (1995). Computer generation of correlated non-Gaussian radar clutter. *IEEE Transactions on Aerospace and Electronic Systems*. 31(1). 106–116. <https://doi.org/10.1109/7.366297>.

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis*. Hoboken, NJ, USA: John Wiley & Sons. <https://doi.org/10.1002/9781118391686>.

Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., & De Bastiani, F. (2019). *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. Boca Raton, FL, USA: CRC Press.

Ritzema, H. P. (1994). *Drainage principles and applications* (2nd ed., Vol. 16). Wageningen, Netherlands: International Institute for Land Reclamation and Improvement (ILRI).

Rudolph, M. (2001). Digital simulation with the fast implicit finite difference (FIFD) algorithm. *Journal of Electroanalytical Chemistry*. 503(1–2). 15–27. [https://doi.org/10.1016/s0022-0728\(00\)00542-8](https://doi.org/10.1016/s0022-0728(00)00542-8).

Sadrezami, H., Ahmad, M. O., & Swamy, M. (2016). Color image denoising using multivariate cauchy PDF in the contourlet domain. In *2016*

IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, Vancouver, BC, Canada, 1-4. <https://doi.org/10.1109/ccece.2016.7726708>.

Safo, S. E., & Ahn, J. (2016). General sparse multi-class linear discriminant analysis. *Computational Statistics & Data Analysis*. 99. 81–90. <https://doi.org/10.1016/j.csda.2016.01.011>.

Sailaja, Rao, K. S., & Reddy, K. V. K. (2010). Text Independent Speaker Identification with Finite Multivariate Generalized Gaussian Mixture Model and Hierarchical Clustering Algorithm. *International Journal of Computer Applications*. 11(11). 25–31. <https://doi.org/10.5120/1626-2187>.

Sánchez-Manzano, E. G., Gómez-Villegas, M. A., & Marín-Díazaraque, J. (2002). A matrix variate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods*. 31(12). 2167–2182. <https://doi.org/10.1081/sta-120017219>.

Scharcanski, J. (2007). A Wavelet-Based approach for analyzing industrial stochastic textures with applications. *IEEE Transactions on Systems, Man, and Cybernetics*. 37(1). 10–22. <https://doi.org/10.1109/tsmca.2006.886345>.

Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*. 44(3). 522–536. <https://doi.org/10.1090/s0002-9947-1938-1501980-0>.

Schwer, D. A., & Kailasanath, K. (2011). Numerical investigation of the physics of rotating-detonation-engines. *Proceedings of the Combustion Institute*.

33(2). 2195–2202. <https://doi.org/10.1016/j.proci.2010.07.050>.

Sedighi, V., Fridrich, J., & Coganne, R. (2015). Content-adaptive pentary steganography using the multivariate generalized Gaussian cover model. *Proceedings of SPIE*. <https://doi.org/10.1117/12.2080272>.

Selig, K., Shaw, P. A., & Ankerst, D. P. (2020). Bayesian information criterion approximations to Bayes factors for univariate and multivariate logistic regression models. *The International Journal of Biostatistics*. 17(2). 241–266. <https://doi.org/10.1515/ijb-2020-0045>.

Shatskikh, S. Y. (1996). Multivariate cauchy distributions as locally gaussian distributions. *Journal of Mathematical Sciences*. 78(1). 102–108. <https://doi.org/10.1007/bf02367960>.

Shimizu, K., & Iida, K. (2002). Pearson type VII distributions on spheres. *Communications in Statistics - Theory and Methods*. 31(4). 513–526. <https://doi.org/10.1081/sta-120003131>.

Sisodia, P. S., Tiwari, V., & Kumar, A. (2014). Analysis of Supervised Maximum Likelihood Classification for remote sensing image. *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*. <https://doi.org/10.1109/icraie.2014.6909319>.

Solaro, N. (2004). Random variate generation from Multivariate Exponential Power distribution. *Statistica & Applicazioni*. 2(2). 25–44. <https://boa.unimib.it/handle/10281/2595>.

Sóyínká, A. T., & Olósundé, A. A. (2021). Inferences from Asymmetric Multivariate Exponential Power Distribution. *Sankhya B*. 83(S2). 350–370.

<https://doi.org/10.1007/s13571-020-00235-w>.

Srivástava, H. M., & Nadarajah, S. (2006). Some families of Bessel distributions and their applications. *Integral Transforms and Special Functions*. 17(1). 65–71. <https://doi.org/10.1080/10652460500431891>.

Stigler, S. M. (1986). Laplace's 1774 Memoir on Inverse Probability. *Statistical Science*. 1(3). <https://doi.org/10.1214/ss/1177013620>.

Subbotin, M. T. (1923). On the Law of Frequency of Errors. *Matematicheskii Sbornik*. 31. 296–301. <https://www.mathnet.ru/php/getFT.phtml?jrnid=sm&paperid=6854&what=fullt>.

Sun, X., Yang, Z., Zhang, C., Ling, K. V., & Peng, G. (2020). Conditional Gaussian Distribution Learning for Open Set Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13480–13489. <https://doi.org/10.1109/cvpr42600.2020.01349>.

Sur, P., & Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences of the United States of America*. 116(29). 14516–14525. <https://doi.org/10.1073/pnas.1810420116>.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2018). *Using multivariate statistics* (7th ed.). Boston, MA, USA: Pearson.

Tahir, M. H., Cordeiro, G. M., Alizadeh, M., Mansoor, M., Zubair, M., & Hamedani, G. G. (2015). The odd generalized exponential family of distributions with applications. *Journal of Statistical Distributions and Applications*. 2(1). <https://doi.org/10.1186/s40488-014-0024-2>

Traat, I. (1989). Moments and cumulants of multivariate elliptical distributions. In *Acta et Commentationes Universitatis Tartuensis da Mathematica*. Estonia.

Turner, M. E. (1960). On heuristic estimation methods. *Biometrics*. 16(2). 299–301. <https://doi.org/10.2307/2527560>.

Varanasi, M. K., & Aazhang, B. (1989). Parametric generalized Gaussian density estimation. *Journal of the Acoustical Society of America*. 86(4). 1404–1415. <https://doi.org/10.1121/1.398700>.

Vasudevay, R., & Kumari, J. V. (2013). On general error distributions. *ProbStat Forum*. 6. 89–95. <https://www.probstat.org.in/PSF-2013-10.pdf>.

Verdoolaege, G., De Backer, S., & Scheunders, P. (2008). Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models. In *2008 15th IEEE International Conference on Image Processing*. IEEE, San Diego, CA, USA. <https://doi.org/10.1109/icip.2008.4711718>.

Verdoolaege, G., & Scheunders, P. (2011). Geodesics on the Manifold of Multivariate Generalized Gaussian Distributions with an Application to Multicomponent Texture Discrimination. *International Journal of Computer Vision*. 95(3). 265–286. <https://doi.org/10.1007/s11263-011-0448-9>.

Verdoolaege, G., & Scheunders, P. (2012). On the Geometry of Multivariate Generalized Gaussian Models. *Journal of Mathematical Imaging and Vision*. 43(3). 180–193. <https://doi.org/10.1007/s10851-011-0297-8>.

Vianelli, S. (1963). La Misura di Variabilita Condizionata in uno Schema Gen-

erale delle Curve Normalli di Frequenza. *Statistica*. 23. 447–473.

Wang, B., Zhang, H., Zhao, Z., & Sun, Y. (2021). Globally Convergent Algorithms for Learning Multivariate Generalized Gaussian Distributions. In *2021 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, Rio de Janeiro, Brazil. <https://doi.org/10.1109/ssp49050.2021.9513857>.

Wang, J. (2009). A family of kurtosis orderings for multivariate distributions. *Journal of Multivariate Analysis*. 100(3). 509–517. <https://doi.org/10.1016/j.jmva.2008.06.001>.

Watson, G. S. (1984). Statistics on spheres. *Journal of the American Statistical Association*. 79(387). 733. <https://doi.org/10.2307/2288434>.

Watson, G. S. (1985). *Statistics on Spheres*. Hoboken, NJ, USA: Wiley-Interscience.

Wu, X., Chou, P. A., & Xue, X. (2000). Minimum conditional entropy context quantization. *IEEE International Symposium on Information Theory*. <https://doi.org/10.1109/isit.2000.866333>.

Yan, B., Ma, X., Huang, G., & Zhao, Y. (2021). Two-stage physics-based Wiener process models for online RUL prediction in field vibration data. *Mechanical Systems and Signal Processing*. 152. 107378. <https://doi.org/10.1016/j.ymssp.2020.107378>.

Yan, D., Zhao, T., & Xu, L. (2023). Statistical Estimation of Loess Landslide Impact by Multivariate Normal Distribution Models with Consideration of Transformation Methods. *Geo-Risk* 2023. 248–258. <https://doi.org/10.1061/9780784484999.026>.

- Yan, J. (2023). Multivariate Modeling with Copulas and Engineering Applications. In *Springer handbook of engineering statistics* (pp. 931–945). https://doi.org/10.1007/978-1-4471-7503-2_46.
- Yang, J., Wang, Y., Wang, X., & Dai, Q. (2009). Image and video denoising using adaptive Dual-Tree discrete Wavelet packets. *IEEE Transactions on Circuits and Systems for Video Technology*. 19(5). 642–655. <https://doi.org/10.1109/tcsvt.2009.2017402>.
- Yeh, I. (2006). Analysis of strength of concrete using design of experiments and neural networks. *Journal of Materials in Civil Engineering*. 18(4). 597–604. [https://doi.org/10.1061/\(asce\)0899-1561\(2006\)18:4\(597](https://doi.org/10.1061/(asce)0899-1561(2006)18:4(597).
- Yin, C. (2021). Stochastic orderings of multivariate elliptical distributions. *Journal of Applied Probability*, 58(2), 551–568. <https://doi.org/10.1017/jpr.2020.104>.
- Yin, C., & Dong, H. (2023). The Bessel function expression of characteristic function. *Communications in Statistics - Theory and Methods*. 1–17. <https://doi.org/10.1080/03610926.2023.2278426>.
- Yin, C., Wang, Y., & Sha, X. (2022). A new class of symmetric distributions including the elliptically symmetric logistic. *Communications in Statistics - Theory and Methods*, 51(13), 4537–4558. <https://doi.org/10.1080/03610926.2020.1818098>.
- Yin, J., Wu, L., Lu, H., & Dai, L. (2019). New estimation in mixture of experts models using the Pearson type VII distribution. *Communications in Statistics: Simulation and Computation*. 49(2). 472–483. <https://doi.org/>

10.1080/03610918.2018.1485943.

Yu, X., & Yin, C. (2023). Some results on multivariate measures of elliptical and skew-elliptical distributions: higher-order moments, skewness and kurtosis. *AIMS Mathematics*. 8(3). 7346–7376. <https://doi.org/10.3934/math.2023370>.

Yue, J., Zhao, W., Mao, S., & Liu, H. (2015). Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*. 6(6). 468–477. <https://doi.org/10.1080/2150704x.2015.1047045>.

Zeckhauser, R., & Thompson, M. (1970). Linear Regression with Non-Normal Error Terms. *The Review of Economics and Statistics*. 52(3). 280–286. <https://doi.org/10.2307/1926296>.

Zhang, H., Wang, N., Li, Y., & Zhan, Y. (2019). Bayesian asset pricing testing under multivariate t-distribution. *Applied Economics Letters*. 26(11). 898–901. <https://doi.org/10.1080/13504851.2018.1512740>.

Zhang, J. (2010). A highly efficient L-estimator for the location parameter of the Cauchy distribution. *Computational Statistics*. 25(1). 97–105. <https://doi.org/10.1007/s00180-009-0163-y>.

Zhang, L., Bioucas-Dias, J. M., & Plaza, A. (2013). Spectral–Spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Transactions on Geoscience and Remote Sensing*. 51(2). 844–856. <https://doi.org/10.1109/tgrs.2012.2205263>.

Zhang, T., Wiesel, A., & Greco, M. (2013). Multivariate generalized Gaussian

- distribution: convexity and graphical models. *IEEE Transactions on Signal Processing*. 61(16). 4141–4148. <https://doi.org/10.1109/tsp.2013.2267740>.
- Zhou, Z., Li, X., & Zare, R. N. (2017). Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Central Science*. 3(12). 1337–1344. <https://doi.org/10.1021/acscentsci.7b00492>.
- Zhu, F., Gao, J., Yang, J., & Ye, N. (2022). Neighborhood linear discriminant analysis. *Pattern Recognition*. 123. 108422. <https://doi.org/10.1016/j.patcog.2021.108422>.
- Zografos, K. (1999). On maximum entropy characterization of Pearson's Type II and VII multivariate distributions. *Journal of Multivariate Analysis*. 71(1). 67–75. <https://doi.org/10.1006/jmva.1999.1824>.
- Zografos, K., & Nadarajah, S. (2005). Expressions for Rényi and Shannon entropies for multivariate distributions. *Statistics & Probability Letters*. 71(1). 71–84. <https://doi.org/10.1016/j.spl.2004.10.023>.
- Zozor, S., Ren, C., & Renaux, A. (2018). On the Maximum Likelihood Estimator Statistics for Unimodal Elliptical Distributions in the High Signal-to-Noise Ratio Regime. *IEEE Signal Processing Letters*. 25(6). 883–887. <https://doi.org/10.1109/lsp.2018.2831178>.
- Zuo, B., Yin, C., & Balakrishnan, N. (2021). Expressions for joint moments of elliptical distributions. *Journal of Computational and Applied Mathematics*. 391. 113418. <https://doi.org/10.1016/j.cam.2021.113418>.

APPENDICES

APPENDIX A

R STUDIO CODE FOR AMEPD DENSITY COMPUTATION

```

ProbMEPD <- function (x, mu, Sigma, q, tol = 1e-06) {
  p <- length(mu)
  if (is.numeric(Sigma) & !is.matrix(Sigma))
    Sigma <- as.matrix(Sigma)
  if (length(x) != p)
    stop(paste("x does not have", p, "elements.\n x
               and mu must have the same length."))
  if (nrow(Sigma) != p | ncol(Sigma) != p)
    stop("Sigma must be a square matrix with size
         equal to length(mu).")
  if (!isSymmetric(Sigma))
    stop("Sigma must be a symmetric, positive-
         definite matrix.")
  eig <- eigen(Sigma, symmetric = TRUE); lambda <-
    eig$values
  if (any(lambda < tol * max(abs(lambda))))
    stop("Sigma must be a symmetric, positive-
         definite matrix.")
  if (q < .Machine$double.eps)

```

```
stop("q must be positive.")

invSigma <- solve(Sigma); xcent <- cbind(x - mu)

c<-((gamma((p+2)/q))^(q/2))*((gamma(p/q)*p)^(-q/2))

k <- (p*(c^(p/q))*gamma(p/2))/(2*(pi^(p/2))*

(gamma(1+(p/q))))

result <- k /sqrt(det(Sigma))

result <- result * exp(-c * abs(t(xcent) %*%

invSigma %*% xcent)^(q/2))

return(as.numeric(result)) }
```



APPENDIX B

R STUDIO CODE FOR SIMULATION FROM AMEPD

```

SimuMEPD <- function (n = 1, mu, Sigma, q, tol =
1e-06){
  p <- length(mu); if (is.numeric(Sigma) & !is.
matrix(Sigma))
    Sigma <- as.matrix(Sigma)
  if (nrow(Sigma) != p | ncol(Sigma) != p)
    stop("Sigma must be a square matrix with size
equal to length(mu).")
  if (!isSymmetric(Sigma))
    stop("Sigma must be a symmetric, positive-
definite matrix.")
  eig <- eigen(Sigma, symmetric = TRUE)
  lambda <- eig$values
  v <- eig$vectors; A <- v %*% sqrt(diag(lambda))
  %*% t(v)
  invSigma<-solve(Sigma); if(any(lambda <tol*max(
abs(lambda))))
    stop("Sigma must be a symmetric, positive-
definite matrix.")
  if (q < .Machine$double.eps); stop("q must be

```

```

positive.")

c<-((gamma((p+2)/q))^(q/2))*((gamma(p/q)*p)^(-q/2))

r <- rgamma(n, shape = p/q, scale = 1/c)^(1/q)

U <- mvrnorm(n, mu = rep(0, p), Sigma = diag(1,

nrow = p), tol = tol)

if (n == 1); U <- rbind(U)

normU <- sqrt(apply(U^2, 1, sum))

U <- U/matrix(normU, nrow = n, ncol = p, byrow =

FALSE)

result <- matrix(r, nrow = n, ncol = p, byrow =

FALSE) * (U %*% t(A))

result <- result + matrix(mu, nrow = n, ncol = p,

byrow = TRUE)

return(result)      }

```

APPENDIX C

R STUDIO CODE FOR AMEPD PARAMETER ESTIMATION

```

.estq <- function(u, q0, p, eps=.Machine$double.eps)
{
  N <- length(u)
  # Define a sequence of initial values to try
  q0_sequence <- seq(0.1, 10, by = 0.1)
  for (initial_q in q0_sequence) {
    result <- tryCatch({
      uniroot(
        function(z) {
          ((p * N / (z^2)) * (((digamma(p / z))) + (log((z /
            (p * N)) * sum(u^(z / 2)))))) +
          N / z - (((p * N) / (2 * z * sum(u^(z / 2)))) *
            (sum(log(u) * u^(z / 2))))
        },
        c(eps, 2 * ceiling(initial_q))
      )$root
    }, error = function(e) NULL)

    if (!is.null(result)) {
      return(result)
    }
  }
}

```

```
}  
  
# If none of the initial values work, return an  
error or a default value stop("Unable to find  
suitable q using the provided initial values.")  
}
```

```
estpaMEPD <- function(x, eps = 1e-4, display = FALSE,  
plot = display) {  
  if (is.vector(x))  
    x <- cbind(x)  
  if (is.data.frame(x))  
    x <- as.matrix(x)  
  
  # Number of variables  
  p <- ncol(x)  
  # Number of observations  
  n <- nrow(x)  
  # Mean vector  
  mu <- colMeans(x)  
  # Centering the data  
  x <- scale(x, center = TRUE, scale = FALSE)  
  
  Sigma <- diag(1, nrow = p, ncol = p)
```

```
q <- 0.1

q_1 <- Inf

if (plot) qseq <- numeric(0)

invSigma <- matrix(0, nrow = n, ncol = p)
```

```
if (display)
  cat(q, " ")

s <- 0

while (abs(q - q_1) > eps) {
  s <- s + 1

  sigmas <- matrix(0, nrow = p, ncol = p)
  invSigma <- x %*% solve(Sigma)
  # u <- numeric(length = n)
  u <- apply(invSigma * x, 1, sum)
  for (i in 1:n) {
    sigmas <- sigmas + u[i]^((q/2)-1) * t(x[i, ,
drop = FALSE]) %*% x[i, , drop = FALSE]
  }
  Sigma <- sigmas

  Sigma <- ((p)/sum(diag(Sigma)))*Sigma

  q_1 <- q
```

```

q <- .estq(u, q, p, eps = eps)

if (plot) qseq <- c(qseq, q)

if (display)

  cat(q, " ")

}

if (plot)

  plot(qseq, type = "l")

if (display)

  cat("\n")

m <- (((gamma((p+2)/q))^(q/2)) * (gamma(p/q)*p)^(
(-q/2))) * q) / (p*n) * sum(u^(q/2))^(2/q)

cc <- ((gamma((p+2)/q))^(q/2)) * (gamma(p/q)*p)^(
(-q/2))

kk <- (p*(cc^(p/q))*gamma(p/2)) / (2*(pi^(p/2)) *
(gamma(1+(p/q))))

loglikef <- (n*(log(kk))) - ((n/2)*log(det(Sigma))) -

```

```

-((n * p / 2) * log(m)) - ((cc*m^(-q/2) / 2) *
  sum(u^(q/2)))

Sigma <- m*Sigma

AIC_f <- 2*3 - 2*loglikef

BIC_f <- -2*loglikef + 2*3*log(n)

# Returned result
result <- list(mu = mu, Sigma = Sigma, q = q,
loglikef =
loglikef, AIC_f = AIC_f, BIC_f = BIC_f)
# Attributes:
# precision of the result
attr(result, "epsilon") <- eps
# number of iterations
attr(result, "s") <- s
return(result)
}
```



APPENDIX D

R STUDIO CODE FOR PLOTTING 3D DENSITY OF BIVARIATE AMEPD

```
## 3D Plot of the density of a AMEPD with 2 variables

plotMEPD <- function (mu, Sigma, q, xlim = c(mu[1] +
c(-2, 2) * Sigma[1,1]), ylim = c(mu[2] + c(-2, 2) *
Sigma[2, 2]), n = 101,
  xvals = NULL, yvals = NULL, xlab = "x", ylab = "y",
  zlab = "f(x,y)",
  col = "yellow", tol = 1e-06, ...)
{if (length(mu) != 2 | nrow(Sigma) != 2 |
ncol(Sigma) !=2)
  stop(paste("plotMEPD only allows plotting a AMEPD
with 2 variables.", "mu must be a length 2 numeric
vector and Sigma must be a 2*2 square matrix.",
sep = "\n"))
f <- function(x) ProbMEPD(x, mu = mu, Sigma =
Sigma, q = q, tol = tol)
ff <- function(x, y) sapply(1:length(x),
function(i) as.numeric(f(c(x[i], y[i]))))
  if (length(n) == 1); n <- rep(n, 2)
  if (is.null(xvals))
    xvals = seq.int(min(xlim), max(xlim), length.out
```

```

    = n[1])

    if (is.null(yvals))

        yvals = seq.int(min(ylim), max(ylim), length.out

        = n[2])

    plot3d(ff, xlim = xlim, ylim = ylim, n = n,

        xvals = xvals, yvals = yvals, xlab = xlab,

        ylab = ylab, zlab = zlab, col = col, ...)

    return(invisible(f)) }

## Contour Plot of the density of a MEPD with 2
variables

contourMEPD <- function (mu, Sigma, q, xlim = c(mu[1]
+ c(-2, 2) * Sigma[1, 1]), ylim = c(mu[2] + c(-2, 2)
* Sigma[2, 2]), zlim = NULL,

    npt = 30, nx = npt, ny = npt, main =

    "AMEPD density",

    sub = NULL, nlevels = 10, levels = pretty(zlim,

    nlevels), tol = 1e-06, ...)

{  if (length(mu) != 2 | nrow(Sigma) != 2 |

    ncol(Sigma) !=2)

    stop(paste("contourMEPD only allows plotting a

    AMEPD with 2 variables.", "mu must be a

```

```

length 2 numeric vector and Sigma must
be a 2*2 square matrix.", sep = "\n"))

f <- function(x) ProbMEPD(x, mu = mu, Sigma = Sigma
, q = q, tol = tol)

ff <- function(x, y) sapply(1:length(x),
function(i) as.numeric(f(c(x[i], y[i]))))

x <- seq(xlim[1], xlim[2], length = nx)
y <- seq(ylim[1], ylim[2], length = ny)
z <- outer(x, y, ff)

if (is.null(zlim))
  zlim <- range(z)

contour(x, y, z, nlevels = nlevels, levels =
levels,
  labels = NULL,
  xlim = xlim, ylim=ylim, zlim=zlim, labcex=0.6,
drawlabels =TRUE, method="flatteest", vfont=NULL,
axes = TRUE, frame.plot=TRUE, col=par("fg"),
lty = par("lty"),
lwd = par("lwd"), add = FALSE, main = main, ...)
return(invisible(f)) }

```

APPENDIX E

R STUDIO CODE FOR DISCRIMINANT ANALYSIS FOR AMEPD

POPULATION

```
Carsales <- read.csv("C:/Users/USER/Desktop/
Car_Sales.csv")
attach(Car_Data_.for.the.Article)
head(Car_Data_.for.the.Article)
summary(Car_Data_.for.the.Article)

(n.Truck <- length(Type_of_Car[Type_of_Car==2]))
(n.Automobile <- length(Type_of_Car[Type_of_Car==1]))
(n.Total <- n.Truck + n.Automobile)
(p.Truck <- n.Truck / n.Total)
(p.Automobile <- 1 - p.Truck)
CAutomobile.Truck <- 1; CTruck.Automobile <- 1

##DISCRIMINANTING FROM AMEPD

## Generate 6 x N matrix, Row 1 =Price,
Row 2 = Horsepower, Row 3 = Wheelbase,
## Row 4 = Curb Weight, Row 5 = Fuel Capacity,
Row 10 = Fuel Efficiency, Columns=Carsales

## "Short/Fat" as opposed to "Tall/Thin" X-matrix
## makes density evaluations easier below
```

```
X <- matrix(rep(0,6*n.Total),ncol=n.Total)

for (i in 1:(n.Automobile+n.Truck)) {

  X[1,i] <- price[i]; X[2,i] <- Horsepower[i];

  X[3,i] <- Wheelbase[i]; X[4,i] <- Curb.Weight[i];

  X[5,i] <- Fuel.Capacity[i];

  X[6,i] <- Fuel.Efficiency[i]
}

## Generate mu vectors and Sigma matrices by Gender
mu.Automobile <- matrix(c(27.687069, 185.793103,
105.751724, 3.197414, 16.592241, 25.275000), ncol=1)
mu.Truck <- matrix(c(26.3225, 186.4000, 112.5200,
3.9325, 21.8950, 19.7000), ncol=1)
Sigma.Automobile <- matrix(c(213.25630, 769.95889,
25.345441, 5.2633104, 29.096163, -33.239316,
769.95889, 3897.72496, 191.038528, 27.2355223,
135.827876, -167.436297, 25.345441, 191.038528,
36.243706, 2.3776747, 9.782577, -9.872663,
5.2633104, 27.2355223, 2.3776747, 0.2855665,
1.2727385, -1.5277836, 29.096163, 135.827876,
9.782577, 1.2727385, 7.548311, -7.597572,
-33.239316, -167.436297, -9.872663,
-1.5277836, -7.597572, 12.424340), ncol=6)
```

```

Sigma.Truck <- matrix(c( 93.789209, 313.677146,
10.008807, 5.8347313, 18.653635, -19.357034,
313.677146, 1695.66068, 122.389590, 24.4107460,
99.544504, -91.487679, 10.008807, 122.38959,
94.534491, 3.2328452, 28.770697, -9.827492,
5.8347313 , 24.41075, 3.2328452, 0.5015361,
2.374828, -1.7310007, 18.653635, 99.54450,
28.770697, 2.3748284, 20.024861, -10.122568,
-19.357034, -91.48768, -9.827492, -1.7310007,
-10.122568, 10.925333), ncol=6)

qAuto <- 0.6914138
qtruck <- 1.012162

# Estimate for Automobile loglikef 3281.292
AIC_f -6556.584 BIC_f -6534.063

# Estimate for Truck loglikef 855.7057
AIC_f -1705.411 BIC_f -1689.278

## Pooled Variance-Covariance
Matrix (Common Variance-Covariance Matrices)
Sigma.TruckAuto <- ((n.Automobile-1)*Sigma.Automobile
+ (n.Truck-1)*Sigma.Truck) / (n.Automobile+n.Truck-2)
Sigma.TruckAuto.inv <- solve(Sigma.TruckAuto)

## Set up AMEPD Densities (Automobile and Truck)

```

evaluated

at x for each Car

```
fTruck.X <- rep(0,n.Total);
```

```
fAutomobile.X <- rep(0,n.Total)
```

```
p <- nrow(X)
```

```
c.truck <- ((gamma((p+2)/qtruck))^(qtruck/2))*
```

```
((gamma(p/qtruck)*p)^(-qtruck/2))
```

```
c.Auto <- ((gamma((p+2)/qAuto))^(qAuto/2))*
```

```
((gamma(p/qAuto)*p)^(-qAuto/2))
```

```
ktruck <- (p*(c.truck^(p/qtruck))*
```

```
gamma(p/2))/(2*(pi^(p/2))*(gamma(1+(p/qtruck))))
```

```
kAuto <- (p*(c.Auto^(p/qAuto))*gamma(p/2))/(2*
```

```
(pi^(p/2))*(gamma(1+(p/qAuto))))
```

```
## Equal Variance
```

```
norm.truck <- ktruck/(sqrt(det(Sigma.TruckAuto)))
```

```
norm.Auto <- kAuto/(sqrt(det(Sigma.TruckAuto)))
```

```
for (i in 1:n.Total) {
```

```
fAutomobile.X[i] <- norm.Auto*
```

```
exp(-c.Auto*(t(X[,i]-mu.Automobile) %*% Sigma.
```

```
TruckAuto.inv %*% (X[,i]-mu.Automobile))^(qAuto/2))
```

```
fTruck.X[i] <- norm.truck*
```

```
exp(-c.truck*(t(X[,i]-mu.Truck) %*% Sigma.
```

```

    TruckAuto.inv %*% (X[,i]-mu.Truck))^(qAuto/2))
}

## Classify Type of Car as Automobile(1) or Truck(2)
based on ratio of Automobile/Truck Densities

classify <- ifelse((fAutomobile.X/fTruck.X) >=
  (CAutomobile.Truck/CTruck.Automobile)*(p.Truck/
  p.Automobile),1,2)

## Confusion Matrix
(CAuto.mat <- table(Type_of_Car,classify))

## Apparent Error Rate
(APER <- (CAuto.mat[1,2]+CAuto.mat[2,1])/n.Total)

## Unequal Variance
norm.truck <- ktruck/(sqrt(det(Sigma.Truck)))
norm.Auto <- kAuto/(sqrt(det(Sigma.Automobile)))

for (i in 1:n.Total) {
  fAutomobile.X[i] <- norm.Auto*
    exp(-c.Auto*(t(X[,i]- mu.Automobile) %*%
      solve(Sigma.Automobile) %*% (X[,i]-
        mu.Automobile))^(qAuto/2))
  fTruck.X[i] <- norm.truck*
    exp(-c.truck*(t(X[,i]-mu.Truck) %*%
      solve(Sigma.Truck)
        %*% (X[,i]-mu.Truck))^(qAuto/2)) }

```

```
## Classify type of car as automobile(1) or truck(2)
based on ratio of A/T Densities

classify <- ifelse((fAutomobile.X/fTruck.X) >=
  (CAutomobile.Truck/CTruck.Automobile)*
  (p.Truck/p.Automobile),1,2)

## Confusion Matrix
(CAuto.mat <- table(Type_of_Car,classify))

## Apparent Error Rate
(APER <- (CAuto.mat[1,2]+CAuto.mat[2,1])/n.Total)
```

