UNIVERSITY FOR DEVELOPMENT STUDIES

SOME CONTRIBUTIONS TO ODD FAMILY OF DISTRIBUTIONS WITH APPLICATIONS TO CANCER DATASETS

YAKUBU AMADU

2021

UNIVERSITY FOR DEVELOPMENT STUDIES, TAMALE

SOME CONTRIBUTIONS TO ODD FAMILY OF DISTRIBUTIONS WITH APPLICATIONS TO CANCER DATASETS

YAKUBU AMADU (BSc Agric Tech., MSc Biometry)

(UDS/DAS/0003/18)

THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, FACULTY OF MATHEMATICAL SCIENCES, UNIVERSITY FOR DEVELOPMENT STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY DEGREE IN BIOMETRY

September, 2021

DECLARATION

Student

I hereby declare that this results is my own work and to the best of my knowledge, it contains no material previously presented for the award of any other degree in this university or elsewhere except where due acknowledgement has been made in the text.

Candidate's Signature:…….. Date:…….

Name: Yakubu Amadu

Supervisor

We hereby declare that the preparation and presentation of the thesis was duly supervised in accordance with the guidelines on supervision of thesis laid down by the University for Development Studies:

Supervisor's Signature:…….. Date:…….

Name: Prof. Albert Luguterah

Co-Supervisor's Signature:…….. Date:…….

Name: Dr. Suleman Nasiru

ABSTRACT

A new family of distributions, by name generalized odd inverse exponential distribution was developed in this study. The new family of distributions were developed using the concepts of relative odds. The statistical properties such as quantile function, characteristic function, moments, moment generating function, mean residual life, inequality measures and order statistics for the new family of distributions were obtained. The maximum likelihood, ordinary least squares and Cramér-von Mises procedures were employed to develop the estimators for the parameters of the new family of distributions. The study also derive some special distributions from the generalized odd inverse exponential family of distributions and these are; Generalized odd inverse exponential Weibull and generalized odd inverse exponential Lomax distributions. The hazard rates of these special distributions indicates that they can handle datasets that exhibit different kinds of non-monotonic failure rates. Again, regression models with cure fraction were developed using the special distributions. Monte Carlo simulations were performed to examine the behavior of the estimators and the results indicates that the estimators were consistent and that the maximum likelihood estimator was the best. The applications of the special distributions were demonstrated using eight cancer datasets and their performances were compared to other well-known existing distributions. The results showed that the special distributions perform better than the other existing distributions in terms of modeling cancer datasets.

ACKNOWLEDGEMENTS

First and foremost, my deepest gratitude goes to the Almighty God for seeing me through this work successfully. Furthermore, I am highly indebted to my supervisor, Prof. Albert Luguterah for his love, care, relentless effort and professional guide and also monitoring this work keenly from the very beginning to the last end of it to make it a worthy one. I owe the gratitude to his supervision, constructive comments and discussions he readily offered me. My profound gratitude also goes to Dr. Suleman Nasiru, my co-supervisor and Head of the Department of Statistics, for his brotherly love, care, courage and advice, help and friendship during this period of my study. His contribution to the success of this academic document is immeasurable and may the Almighty God reward him in thousand folds. Again, I wish to express my sincere gratitude to Mr. Abudu Ballu Duwiejuah of the Faculty of Biosciences at the Nyankpala campus of UDS for taking time out of his busy schedule and offered invaluable contributions to the successful completion of this work.

Moreover, I wish to express my heartfelt gratitude to my wife, father and mother Madam Hadija Siibaway, Mr. Amadu Sayibu and Madam Mariama Haruna respectively for their endless love, moral support and encouragement even at the most difficult times. My deepest appreciation also goes to my siblings especially Mr. Abdul-Latif Amadu for his love and support in the completion of this work.

Again, I owe the gratitude to Prof. Abdul-Halim Abubakari, Head, Department of Horticulture whose care and support have brought me this far. Lastly, I would like to thank the Ghana scholarship secretariat for supporting me pay my final year school fees.

DEDICATION

This work is dedicated to my wife, Madam Hadija Siibaway and my dear parents, Mr. Amadu Sayibu and Madam Mariama Haruna.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AB	Average Bias
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
K-S	Kolmogorov-Smirnov
LRT	Likelihood Ratio Test
MGF	Moment Generating Function
PDF	Probability Density Function
TTT	Total Time on Test
MSE	Mean Square Error
GOIE	Generalized Odd Inverse Exponential
GOIEW	Generalized Odd Inverse Exponential Weibull
GOIEL	Generalized Odd Inverse Exponential Lomax
E-Lx	Exponentiated Lomax
GIW	Generalized Inverse Weibull

IW	Inverse Weibull
KIE	Kumaraswamy Inverse Exponential
OGEW	Odd Generalized Exponential Weibull
EOFW	Extended Odd Fréchet Weibull
GIE	Generalized Inverse Exponential
Kum-BIII	Kumaraswamy Burr Three
PWLx	Poisson Weibull-Lomax
PWB	Poisson Weibull Burr
PWLL	Poisson-Weibull-Log-Logistic
PW	Poisson Weibull
PRB	Poisson Rayleigh Burr
PB	Poisson Burr
PEB	Poisson Exponential Burr
PGGHN	Poisson-Gamma Generalized Half-Normal
PGBS	Poisson-Gamma Birnbaum-Saunders
PGLL	Poisson-Gamma Log-Logistic
PGW	Poisson-Gamma Weibull

CHAPTER ONE

INTRODUCTION

1.1 Background

The devastating nature of chronic diseases has made it the leading cause of mortality among human population across the globe, and account for about 60% of deaths in the world (Jennings, 2014). Deaths associated with chronic diseases was estimated as double of that of infectious diseases, with low and middle-income countries recording about 80% of chronic disease deaths, half of which were women. Without any remedy actions on the prevelance of chronic diseases, the death toll from chronic diseases was therefore estimated to increase by 17% between 2005 and 2015 (WHO, 2005).

In technical terms, a chronic disease is commonly considered an illness that lasts one year or more, necessitates ongoing medical attention and/or limits a person's daily activities (Megan, 2013). The most common chronic diseases across the globe include; chronic respiratory diseases, heart diseases, stroke, diabetes, arthritis, asthma, cancer and hypertension among others (Megan, 2013; Wullianallur and Viju, 2018).

Among the global recognized chronic diseases, cancer is regarded as the leading cause of deaths associated with chronic diseases in the global community. The WHO (2005) projected deaths associated with cancer related diseases to be 13% of total cause of deaths.This striking projection is mainly attributed to the larger geographical diversity of cancer occurrence within countries across the globe. Also, WHO (2018) estimated an increase in the global burden of cancer disease to be 18.1 million new cases, with the death toll increasing to 9.6 million.

Due to the increasing awareness on cancer disease, its risks factors and associated burden on human population, it is of great importance to conduct statistical analysis on the incidence, prevelance, mortality and survival of cancer disease over a period of time. Several researchers therefore employ different statistical techniques to model the behaviour of cancer disease incidence, prevelance, mortality and survival across various regions of the world.

The most widely used statistical approach in modeling cancer data is the parametric statistical distributions with Exponential, Weibull, Gompertz, Lognormal and Loglogistic as the commonly used models (Vallinayagam *et al.*, 2014). However, barrage of these statistical models depends on certain distributional assumptions since they are parametric in nature. Therefore to provide an accurate fit to a cancer data that have varrying degree of skewness and kurtosis or that are characterized by non-monotonic failure rates, the classical distributions may not be suitable. Hence, the classical distributions which have essential limitations in data modeling, has led statistical researchers to develop new flexible distributions by adding one or more parameters to the existing distributions. This current study is focus on developing a new odd family of distributions for modeling cancer datasets.

1.2 Statement of the Problem

The increasing incidence of cancer diseases that has claim a lot of lives across various regions of the world has triggered myriad of statistical researchers to conduct statistical analysis on the incidence, prevalence, mortality and survival of cancer disease over a period of time. As a result, different statistical techniques have been employed to model the

incidence, prevalence, mortality and survival of the cancer diseases throughout the world. The commonly used statistical models are; Exponential, Weibull, Gompertz, Lognormal and Log-logistic which are all parametric distributions (Vallinayagam *et al.*, 2014).

Meanwhile, almost all of these statistical models rely on certain distributional assumptions as they are parametric in nature. However, most of these cancer data have varying degree of skewness and kurtosis and sometimes characterized by non-monotonic failure rates. Therefore modeling with these classical distributions may not be suitable. As a result, statistical researchers have developed more flexible distributions by adding more parameters to the existing distributions. These flexible distributions have the tendency to handle data with non-monotonic failure rates and be able to accurately model datasets that are highly skewed or have fat tails (Nasiru, 2018). However, the flexibility of these modified distributions depend on the parameters added to the baseline distribution. The effect of adding one shape parameter will not be the same as adding two parameters. So to develop a generator that will produce both light-tail and heavy-tail distributions and also handle skewness and kurtosis at the same time we need to add more than one shape parameter to the baseline distribution (Nasiru, 2018). Hence, this current study is focus on developing a new odd family of distributions by adding two extra shape parameters to baseline distribution to enhance flexibility when modeling cancer datasets.

1.3 General Objective of the Study

The general objective of the study is to develop and investigate the statistical properties of a new odd family of statistical distributions.

1.4 Specific Objectives

The specific objectives of the study are:

i. To develop the generalized odd family of distributions.

ii. To derive the statistical properties of the generator.

iii. To come up with estimators for the parameters of the generator.

iv. Perform simulation studies to examine the properties of the estimators.

v. To demonstrate the applications of the new distributions using cancer data.

1.5 Significance of the Study

The development of generalized class of distributions from existing distributions have received an increasing attention in the statistical literature, due to their wider application in different fields of studies. These modified distributions have the tendency of improving the flexibility as well as the goodness-of-fit when modeling lifetime dataset. Thus, in this study a new generator called generalized odd inverse exponential family of distributions was developed and studied.

1.6 Scope of the Study

This study mainly focuses on developing a generator with inverse exponential distribution as the baseline model using the concept of relative odd. The study also examined the properties of the proposed generator and demonstrating the usefulness of this generator using cancer datasets.

1.7 Thesis outline

This thesis is consist of six chapters including this one. Chapter two is focused on literature on related works of this study. Chapter three presents the methodology of the study. Chapter four presents the theoretical results of the study. Chapter five deals with the simulations and applications of the GOIE family of distributions. Finally, chapter six presents the summary, conclusions and recommendations of the study.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter reviews related works done on odd family of distributions, generalized class of distributions, cure rate models and inverse exponential distributions. The chapter is divided into four main headings namely; review on existing odd family of distributions, review on existing generalized class of distributions, empirical researches on cure rate models and empirical researches on inverse exponential distributions.

2.1 Review on Existing Odd Family of Distributions

Some studies that have been done on the odd family of distributions are discussed in this section.

Cordeiro *et al.* (2019) introduced and studied the mathematical properties of a new generator called odd Lomax-G family based on the Lomax distribution. This distribution has special cases which includes: odd Lomax Weibull distribution, odd Lomax-Lomax distribution, odd Lomax-log-logistic distribution and odd Lomax-Lindley distribution. The plots of the PDF of each of the models were generated to illustrate the shapes of these distributions. Shapes such as; decreasing, unimodal, right skewed and upside down bathtub modified were assumed. The nature of the shapes of the plots of the density function for each of the distributions indicates that the proposed distribution was flexible.

Also, Nasiru (2018) proposed a new class of distribution known as the extended odd Fréchet-G family of distributions. He developed two special cases of this distribution and

these include; extended odd Fréchet Nadarajah-Haghighi distribution and extended odd Fréchet weibull distribution. Based on these two distributions, he generated plots for the PDF and the hazard rate function using some selected parameter values in order to examine the flexibility of the proposed distribution. The plots for the PDF of the extended odd Fréchet Nadarajah-Haghighi distribution and the extended odd Fréchet Weibull distribution assumed different shapes. These shapes include; upside down bathtub, unimodal, reverse J-shape, bathtub modified and decreasing. Also, the plot of the hazard rate function of the two distribution for some selected parameter values assumed different shapes. These include; decreasing, bathtub and upside down. These distributions were again compared to other competing distributions in terms of modeling the fatigue time of 1016061-T6 in order to justify the performance of the proposed distribution. The results revealed that EOFNH distribution provides a perfect fit for the data than the other distributions.

More so, Ahsan and Elgarphy (2018) proposed a new generator from Fréchet random variable called the odd Fréchet-G family of distributions. Special distributions of the odd Frèchet-G family of distributions were also derived and these include; odd Fréchet-weibull distribution, odd Fréchet–Lomax distribution, odd Fréchet-pareto distribution and odd Fréchet-Gamma distribution. They considered some scale and shape parameters to drive the special distributions of the proposed odd Fréchet-G family of distributions and then generate the plots of PDF and hazard rate function for two of the special models; odd Fréchet-weibull distribution and odd Fréchet–Lomax distribution. Given the plots generated, the PDF of the two distributions exhibited different kinds of shape, thus, decreasing, right skewed and unimodal shapes, whereas the hazard rate functions of those

distributions exhibited an increasing, decreasing and J- shape. In addition, the odd Frèchet-weibull distribution was compared with some existing distributions from the odd family to examine the flexibility of the new distribution in terms of modeling life time dataset. The results indicates that the new distribution was better in terms of modeling a lifetime data than the existing distributions.

Hosseini *et al.* (2018) introduced the generalized odd Gamma-G family of distributions. The generalized odd Gamma-Uniform distribution and the generalized odd Gamma-Weibull distribution were also developed as special cases to the generalized odd Gamma-G family of distributions. When the generalized odd Gamma-Uniform distribution was compared to other candidate models in terms of modeling the AG negative data, it did better than the existing models. Also, the hazard rate function exhibited and increasing and a bathtub shapes when it was plotted. The density function was also found to be decreasing, left skewed and symmetric when plotted.

Further, Alizadeh *et al.* (2017) propose and derived the mathematical properties of a new generator of continuous distributions with three extra parameters called the odd log-logistic logarithmic generalized family of distributions. This distribution has special cases which includes: odd log-logistic logarithmic Weibull distribution and odd log-logarithmic normal distribution. The plot of the odd log-logistic logarithmic Weibull distribution density produced a decreasing unimodal and bimodal shapes. On the other hand, the plot of hazard rate function of the odd log-logistic logarithmic Weibull distribution generates a decreasing, increasing, in, bathtub shaped and upside down bathtub shaped.

Again, Alizadeh *et al.* (2017) proposed the Odd log-logistic Marshall-Olkin Lindley (OLLMO-L) distribution, using the Lindley distribution as the baseline distribution. They premised that the OLLMO-L distribution is highly flexible than the Lindley distribution, and thus allows for greater flexibility of the tails. Also, they generate a plots to examine the nature of the density function, as well the survival and hazard rate function of the distribution. The plot of the density function displayed several shapes such as unimodal, symmetric, right skewed, and monotonically decreasing shapes. In addition, the plot of the hazard rate function of the OLLMO-L distribution showed a highly flexible shapes, such as increasing, decreasing, and upside-down bathtub. They further illustrate the flexibility of the model by comparing its performance to other existing models in terms of modeling a real lifetime dataset. The results indicates that the proposed model gives a better fit to the dataset than the existing models.

Cordeiro *et al.* (2017) introduced and derived the general statistical properties of a new generator of continuous distributions with one extra parameter called the generalized odd half –Cauchy family. Generalized odd half-Cauchy-Weibull distribution, generalized odd half-Cauchy-normal distribution, generalized odd half-Cauchy-log-logistic distribution and generalized odd half-Cauchy-Gumbel distribution are special distributions to the generalized odd half –Cauchy family.

Also, Korkmaz *et al.* (2017) proposed and studied the statistical properties of a new generator called the exponential Lindley odd log-logistic-G family. He also developed the special distributions of this family and this includes exponential Lindley odd log-logistic-normal distribution, exponential Lindley odd log-logistic Weibull distribution and exponential Lindley odd Lindley-log-Logistic-Lomax distribution.

Furthermore, Alizadeh *et al.* (2017) introduced a new class of distributions called the generalized odd generalized exponential family. In their work, they also derive generalized odd generalized exponential Weibull distribution, generalized odd generalized Exponential Normal distribution and generalized odd generalized Exponential Kumaraswamy distribution as special cases to the generalized odd generalized exponential family. They went further to provide a plots of the PDF and the hazard rate function of some of the distributions. The results of the plots produced different kinds of shapes. For the density function, decreasing, bathtub, left skewed and right skewed shapes were recorded. The shapes produced by the hazard rate function plots include, bathtub, J-shape and an increasing shape.

Again, Brito *et al.* (2017) proposed and studied the statistical properties of a new class of continuous distributions called the Topp-Leone Odd Log-logistic family. This distribution extends the one-parameter distribution introduced by Topp and Leone. The special cases of the Topp-Leone Odd Log-logistic family are as follows; Topp-Leone odd log-logistic normal distribution, Topp-Leone odd log-logistic Weibull distribution and Topp-Leone odd log-logistic generalized half-normal distribution.

Yousef *et al.* (2017) worked on a new model for analysis of lifetime data referred to as the odd Lindley NH distribution. The strength of the distribution was demonstrated to be good. Certain characteristics were also developed and the model was shown to exhibit various shapes.

Tahir *et al.* (2015) proposed a new family of continuous distributions called the odd generalized exponential family whose hazard rate could be increasing, decreasing J,

reversed-J, bathtub and upside-down bathtub. This distribution includes as a special case of the widely known exponentiated Weibull distribution. The odd generalized exponentiated Weibull distribution, odd generalized exponentiated Frèchet distribution and the odd generalized exponentiated normal distributions are the special cases of the odd generalized exponential family of distributions.

2.2 Review on Existing Generalized Class of Distributions

This section presents some current works on generalized class of distributions. Alizadeh *et al.* (2020) introduced and studied a four-parameter lifetime distribution known as the odd log-logistic generalized Gompertz model. This model was used to generalize some existing distributions and it includes; the generalized Gompertz, exponential, generalized exponential distributions and, among others. The method of maximum likelihood estimation of parameters is compared by six different methods of estimations with simulation study. After further studies, the new model was deemed to be a better fit to the real dataset that was provided.

Ahmad *et al.* (2020) studied a new family of distributions called the odd generalized N-H. Characterizations based on the truncated moments, hazard function and conditional expectations are presented for the generated family. Parameter estimates of the family are obtained based on maximum likelihood procedure. Two real datasets are employed to show the usefulness of the new family.

Elsayed *et al.* (2020) introduced a new univariate extension of the Fréchet distribution. A simple type Copula based construction using Morgenstern family and via Clayton Copula

is employed to derive many bivariate and multivariate extensions of the new model. They assessed the performance of the maximum likelihood estimators using a simulation study. The importance of the new model is shown by means of two applications to real datasets.

Oluyede *et al.* (2020) proposed a new generalized family of distributions called the exponentiated generalized power series (EGPS) family of distributions. Some sub-models of the proposed family were derived and studied. Also, Rényi entropy and some other mathematical properties as well as maximum likelihood estimates of the proposed family were derived. Simulation study was carried out to examine the bias and the mean square error of the maximum likelihood estimators for each of the model's parameters. The usefulness, applicability and flexibility of the proposed distribution were illustrated by means of real life datasets.

Raheem (2019) uses three optimality criteria to conclude the optimal allocation of multiple accelerated life testing for the generalized half-normal model under type-I censoring. They derive the maximum likelihood estimates of the parameters and their Fisher information matrix. Numerical and simulations examples are used to demonstrate the effectiveness of the optimal allocation. A sensitivity analysis of the optimal allocation to misspecification of the model parameters is conducted.

Alizadeh *et al.* (2019) came up with a three-parameter lifetime model, called the new odd log-logistic Lindley distribution. Some structural properties of the new distribution including ordinary and incomplete moments, quantile and generating functions and order

statistics were obtained. The new density function were expressed as a linear mixture of exponentiated Lindley densities. Different methods for the estimation of model parameters were discussed. It was shown that the model is quite good and performs better than some existing ones.

Hassan *et al.* (2019) in this article introduced and studied a new four-parameter distribution, called the odd generalized exponential power function distribution. The proposed model is a particular case from the odd generalized exponential family. The characteristics of this new model were expressed. The model parameters were estimated via the maximum likelihood and percentiles methods of estimation. A simulation study was carried out to evaluate and compare the performance of estimates in terms of their biases, standard errors and mean square errors. Eventually, the practical importance and flexibility of the proposed distribution in modeling real data application was checked. It can be concluded that the new distribution works better than some other known distributions.

Zubair *et al.* (2019) worked on a new model by name log-odd normal generalized family of distributions based on log-odds for the analysis of lifetime data. This new model has sub-models including the log-odd normal power-Cauchy distribution. Also, certain characteristics of this new distribution were expressed. The usefulness of the proposed family is proved empirically by means of a real air pollution dataset.

Afify *et al.* (2019) introduced a new class of continuous distributions called the generalized odd Lindley-G family. Four special models of the new family are provided. The study

elaborated and explained expressions for some structural properties and behaviors of the new family. The maximum likelihood method is used for estimating the model parameters. The flexibility of the generated family is illustrated by means of two applications to real datasets.

Khalil *et al.* (2019) pioneered a new three-parameter lifetime model called the Burr X exponentiated Weibull model. The major justification for the practicality of the new lifetime model is based on the wider use of the exponentiated Weibull and Weibull models. They are motivated to propose this new lifetime model because it exhibits different kinds of failure rates. And they prove empirically the importance and flexibility of the new model in modeling two types of lifetime data.

Anwar and Bibi (2018) introduced a model named half-logic generalized Weibull distribution. This new distribution has sub-models and certain characteristics were expressed. Its usefulness and potentiality were demonstrated on two datasets. Their study revealed that the recent one out-classed the models it was compared with.

Again, Muhammad *et al.* (2018) engineered and study a new family of distributions called the Poisson-odd generalized exponential distribution. They derived certain mathematical characteristics of the new distribution and some of its behaviors were displayed. They presented two special cases of the new family, namely the Poisson odd generalized exponential-half logistic and the Poisson odd generalized exponential-uniform distributions. Applications to two real datasets showed that the new model out-classed several models in terms of performance.

Hosseini *et al.* (2018) proposed a current family of distribution known as the Generalized Odd Gamma-G family of distribution. Special cases of this current family of distribution were developed and studied. The characteristics of the new distribution were investigated. The study also highlighted various shapes of the distribution and the applicability also displayed.

Prataviera *et al.* (2018) studied a four-parameter model called the generalized odd log-logistic flexible Weibull distribution. The proposed distribution can handle various forms of hazard rate and as a result can be used effectively in reliability analysis. Also, a parametric regression model based on the new distribution as an alternative to the location-scale regression model is presented. Applications in real engineering datasets illustrate the flexibility of the proposed models.

Alizadeh *et al.* (2018) worked on a modern family of continues models by name the complementary generalized transmuted Poisson -G family, an extension of the transmuted family pioneered by Shaw and Buckley (2007). Special models were provided and some general characteristics were expressed. The new model outranked other ones mentioned in literature.

Rahmouni and Orabi (2018) introduced a current distribution by name the exponential-generalized truncated geometric distribution. Some characteristics of this new model and certain shapes were exhibited. The potency of the new model was shown to be very good and it fits well to real datasets.

Elgarhy *et al.* (2018) developed and examined the exponential generalized Kumaraswamy model. In this study, some mathematical characteristic such as moments were expressed. The model parameter estimation was done and the applicability of this model was exhibited with real datasets and the results revealed were positive.

Korkmaz *et al.* (2018) proposed a new class of lifetime distributions called the generalized odd Weibull generated family. Some mathematical properties of the new family are derived. The maximum likelihood method was used for estimating the model parameters. They study the behaviour of the estimators by means of two Monte Carlo simulations. The importance of the family illustrated by means of two applications to real datasets.

Aryal and Yousof (2017) engineered and investigated a current family of distribution by the name the exponentiated generalized -G Poisson class of models. They derived certain mathematical characteristics of the current distribution and certain behaviors were displayed. The model was shown to be very acceptable in terms of performance.

Nasiru *et al.* (2017) came up with a new distribution called the exponentiated generalized exponential Dagum distribution. Burr III, exponentiated generalized Dagum distribution, among others, were the special cases of the proposed distribution. Different kinds of failure rates were exhibited by the proposed distribution. Some structural properties of the proposed family were also examined. Maximum likelihood estimators of the parameters of the distribution were developed and simulation studies performed to assess the properties of the estimators. The applications to lifetime datasets indicates that the new family can provide better fits than other well-known classes of distributions.

Haghbin *et al.* (2017) developed a new generator of continuous distribution with three extra parameters called the new generalized odd log-logistic family of distributions. The study presented two special cases of the new generator and discussed some mathematical properties of the proposed family. They also present certain characterization of the proposed distribution and derive a power series for the quantile function. The importance of the new family is illustrated by means of two real datasets. These applications indicate that the new family can provide better fits than other well-known classes of distributions.

Alkarni (2016) proposed a modern class of distribution known as generalized extended Weibull power series class of models. He followed the same procedure adopted by Adamidis and Loukas (1998). He also worked on some sub-models and established a number of the derived model.

Also, Vatto *et al.* (2016) pioneered a new distribution called the exponential generalized NH model. They studied and derived some of its characteristics exhibited in the shapes of its hazard function. They also provided a maximum likelihood procedure for estimating the exponential generalized NH distribution parameters.

Nwezza *et al.* (2016) introduced a new flexible five parameter lifetime distribution called Marshall-Olkin Gumbel-Lomax distribution. Some characterizations of the distribution such as the Trimmed L-moments, moment generating function, and order statistics are derived. The unknown parameters of the new distribution are estimated using the maximum likelihood approach. The potentials of the new distribution are illustrated using two real life datasets.

Okashaa *et al.* (2015) introduced a new family of Marshall–Olkin extended generalized linear exponential distribution. This new family has the advantage of modeling various

shapes of aging and failure criteria. Several special cases of the proposed family were also discussed. In addition, the asymptotic confidence intervals for the parameters are derived from the Fisher information matrix. Finally, the obtained results are validated using some real datasets and it is shown that the new family provides a better fit than some other known distributions.

Tahmasebi and Jafari (2015) introduced the generalized Gompertz-power series family of models a compound of the power series and generalized Gompertz models. The distribution comprised of sub-models which were duly discussed in the study to great effect. Advanced studies revealed the potentiality and usefulness of the model.

Cordeiro *et al.* (2014) propose a new class of distributions called the Lomax generator with two extra positive parameters to generalize any continuous baseline distribution. Some special models as well as certain mathematical properties of the new generator were also examined. Maximum likelihood estimation procedure was used to estimate the parameters of the model. They define a log- Lomax–Weibull regression model for censored data. The importance of the new generator is illustrated by means of three real datasets.

Muhammed (2007) introduced another model known as the generalized half-logistic Poisson model which exhibits favorable behaviors. The practical importance, applicability and tractability were demonstrated using real data and this showed that the generalized half-logistic Poisson distribution out-performed certain models.

2.3 Empirical Researches on Cure Rate Models

A barrage of researches have been carried out on cure rate models all over the world. Borges (2020) proposed a novel model known as a mixture cure rate model, with the help of log-logistic distribution. The model aimed to overcome the challenge of estimating the mode of the hazard function of patients in different stages of uterine cervical cancer in the midst of longstanding survivors. Thus, parameterization of the new model was conducted through the hazard function mode. With the hazard function, it was estimated that cancer stages can affect the cured fraction and the mode. Inference of the new model was done through the maximum likelihood estimation methods. Also, the properties of the maximum likelihood estimators was verified by means of Monte Carlo simulation methods. Finally, the effectiveness of this model was assessed by subjecting it to real time uterine cervical cancer data.

Boussari *et al.* (2020) introduced a new model known as excess hazard regression model, with the "time-to-cure" as a covariate dependent parameter. The parameters of this model was estimated by means of maximum likelihood method. Simulation studies was also conducted to assess the performance of the model. To carry out effective simulation, age at diagnosis as the only factor was considered as covariate in order to reduce the complexity of the model while not compromising generality. Thus, the model; beta-Time-to-Null-Excess-Hazard was employed. Furthermore, the applicability of the model was assessed by subjecting it to three different real time datasets namely; testicular cancer data, women pancreatic cancer data and women colon cancer data.

Oliveira *et al.* (2019) studied the mixture and non-mixture cure fraction models through a discrete lifetime distribution as a surrogate to the conventional lifetime continuous distribution for lifetime data in the midst of cured fraction, censored data and covariate. Thus, the objective of the study is to introduce a new model; Weibull cure fraction models to examine longstanding lifetimes and risks factors with application priority given to lifetimes of patients on treatment for pelvic sarcomas. The inference and residual of the new model was determined using a maximum likelihood method and randomized quantile residuals method. Also, properties of the model estimates was assessed through an extensive simulation procedures. Finally, the application of the discrete Weibull to lifetime data indicated that, the new model offers the best simplicity of the likelihood functions as compared to the continuous Weibull model, largely used in analysing lifetime data with cure fraction, censored data and covariate.

Calsavara *et al.* (2019) proposed a new cure rate model known as the generalized time-dependent logistic model. This model has a power variance function frailty term, introduced in the hazard function to cater for the unobservable heterogeneity amongst patient populations. The model give room for non-proportional hazards as well as survival data with longstanding survivors. In addition, the parameter estimates of the model was done through the maximum likelihood approach, with Monte Carlo techniques employed to conduct an evaluation of the performance of the model. The applicability of the model over its competing models in practice, was examined by subjecting the model to a real time data from a particular population-based study of incidents cased of melanoma diagnosed. The model is flexible and applicable to situation with or without cure fractions. This

attribute of the model is as a result of the fact that, it does not make assumptions about any existence of the cure rate, as far as the parameter value has led to proper or improper distribution. Based on these, the model is adjudged more flexible over other alternative models.

Barriga *et al.* (2018) introduced a new survival model for lifetime data with surviving fraction and obtained some of its special properties. The new model is an extension of the promotion time cure model with an extra parameter having some control on the heterogeneity or dependence of an unobserved number of lifetimes. The effects of the covariate in the cure fraction is evaluated by constructing a regression model. Inference of the new model was drawn through maximum likelihood approaches. Also, a special algorithm, known as expectation maximization algorithm was established to determine the maximum likelihood estimates of the model parameters. Further, likelihood ratio test was conducted through an empirical analysis in order to compare the promotion time cure and the proposed model. The applicability and the flexibility of the model was illustrated through a real time colorectal cancer data.

Leão *et al.* (2018) incorporated frailties into a cure rate regression model, based on the Birnbaum-Saunders distribution, to generate an alternative model to the existing models. Likelihood-based approaches were used to estimate the new model parameters, and as well derived its influence diagnostics. An assessment of some local influence on the parameter estimates under different trepidation schemes was carried out. The proposed model offers some advantage over its competing models, and this include the possibility to jointly study

the heterogeneity among patients by their frailties and the existence of a cured fraction of them. The performance and application flexibility of the new was evaluated through Monte Carlo simulation techniques and by illustration using melanoma dataset. Finally, as both Monte Carlo simulation and the illustration shown better performance and possible applications of the new model, the illustration in particular established the significance of statistical diagnostics in the modeling.

A non-mixture cure rate model for right-censored data with Fréchet distribution was proposed by Kutal and Qian (2018). Maximum likelihood estimators of the parameters of the model were developed and simulation studies performed to assess the properties of the estimators. The applications to melanoma and leukemia datasets indicates that the proposed model provide a better fits than the other candidate models.

Martinez and Achcar (2018) introduced a new distribution known as the defective Dagum distribution, as an extension of the Type I Dagum distribution. This new distribution has a unique advantage such that the cure fraction can be expressed as a function of a single parameter. In order to estimate the model parameters, maximum likelihood and Bayesian approaches were employed. The appropriateness of the maximum likelihood and Bayesian approaches to estimating the model parameters were evaluated by means of simulation approaches. To this end, posterior distributions of the parameters were estimated with the help of Markov chain Monte Carlo method, for the Bayesian analysis. An illustration involving a real time dataset was made to ascertain the performance and flexibility of the new distribution. Based on the evaluation of the parameters of the new distribution, and

the illustration with the real time dataset, the new distribution is adjudged flexible and it is better alternative for the analysis of real time-to-event data involving censored information and a cure fraction.

Alizadeh *et al.* (2017) proposed the heteroscedastic cure rate regression models based on the odd log-logistic Topp-Leone G family of distributions. These cure rate models includes the log-generalized odd log-logistic Topp-Leone Weibull cure rate heteroscedastic regression model. The maximum likelihood estimators of the cure rate models parameters were also established. The potentiality of the developed cure rate models were shown by means of gastric cancer datasets.

Ortega *et al.* (2015) studied a new flexible models, generated by gamma random variables for life time modeling known as; Gamma Nadarajah–Haghighi distribution. This new distribution introduced, is three-parameter exponential-type distribution, which is better flexible and considered effective in modeling survival data and reliability problems. The plots of the hazard rate function of the distribution exhibits some special properties such as constant, decreasing, increasing, upside-down bathtub and bathtub-shapes. The parameters of the distribution are estimated by means of maximum likelihood methods. The model was subjected to AIDS data and melanoma data to test its applicability in practice.

Fachini *et al.* (2014) worked on a new model known as bivariate regression model with cure fraction. In this study, a location-scale model was introduced for bivariate survival time's base on the copula and used to model the dependence of bivariate survival data with cure fraction. Inferential procedures with constrained parameters using maximum

likelihood was considered. Finally, the importance of this model was displayed using diabetic retinopathy dataset.

Martinez and Achcar (2014) proposed a new model known as bivariate model for survival data with a cure fraction, grounded on the three-parameter generalized Lindley distribution. Copula functions were employed to obtain a joint distribution of the survival times in the new model. In this regard, three type of copula functions were employed, namely; Farlie–Gumbel–Morgenstern, Gumbel–Barnett and Clayton copulas. The parameters of the model were estimated using Markov Chain Monte Carlo methods, under the Bayesian framework. The suitability of the application of the model was done by considering a real time dataset that concerns invasive cervical cancer. The proposed model provided a much better fit to the real time data of invasive cervical cancer than the other alternative models.

Martinez *et al.* (2013) studied a mixture and non-mixture cure fraction models based on the Bayesian analysis of the four-parameter generalized Weibull distribution with cure fraction, censored data and covariate. Based on that study, a cure fraction regression model was proposed to model censored and uncensored real time dataset. The new model is an extension of numerous distributions used extensively in lifetime data analysis to model monotone and non-monotone shape hazard rates and it serves as a good alternative for the analysis of real datasets with flexibility. Markov Chain Monte Carlo techniques was used to evaluate the performance of the model in order to draw inferences. The new regression model was subjected to practical application using gastric cancer lifetime data to test its flexibility, practical relevance and applicability.

Rodrigues *et al.* (2010) introduced new models by name destructive weighted Poisson cure rate models. These models are destructive length-biased Poisson, destructive exponentially weighted Poisson, destructive negative binomial and destructive Conway-Maxwell Poisson models. The model parameters were estimated using maximum likelihood estimation procedure. The potency of the models were also provided through application to malenoma dataset.

2.4 Empirical Researches on Inverse Exponential Distributions

Al Sobhi (2020) proposed a new distribution known as the Inverse Power Logistic Exponential (IPLE) distribution. This distribution serves as an extension to well-known distributions, namely; the inverse Weibull, inverse logistic exponential, inverse Rayleigh, and inverse exponential distributions. The density plots of the distribution depicts shapes namely; symmetrical, right-skewed, left-skewed, reversed-J-shape, and J-shape. Also with the hazard function, the distribution shows an increasing, decreasing, unimodal, reversed-J-shape and J-shape plots. Further, the parameters of the new distribution was estimated using five estimation approaches, namely; the maximum likelihood, Anderson–Darling, least-squares, Cramér–Von Mises, and weighted least-squares estimation approaches. The applicability and flexibility of the proposed model was assessed with the help of an insurance dataset. The data application indicated the new model as better fit to insurance data as compared to the other competing models.

Kumar *et al.* (2018) introduced a new distribution known as a Poisson-inverse exponential distribution to model a life time data with increasing and decreasing failure rates. The statistical properties of the new distribution include; quantile, moments, mean, variance

and reliability. The plots of both the probability density function and the hazard rate function depict an initially increasing shape, which attained maximum point and then decreased. The model parameters were estimated using the maximum likelihood and the Bayes estimators. Also, the expectation-maximization algorithm was introduced to handle the incomplete data obtained under the progressive type-II censoring with binomial removal sample. Finally, the Bayesian procedure is considered as the flexible approach to Poisson-inverse exponential distribution parameter estimation under progressive type-II censoring with binomial removal sample.

Oguntunde *et al.* (2017) pointed out the inability of the exponential distribution to better model real lifetime phenomena whose failure rate are not constant. Based on the shortcomings of the exponential distribution, the Exponential Inverse Exponential (EIE) distribution was proposed, with the help of the exponential generalized family of distributions. The reliability analysis of the proposed distribution expressed key statistical functions such as; survival function, hazard function, reversed hazard function and the odds function. Maximum likelihood approach was employed for the parameter estimation of the proposed distribution. The shapes of the distribution is unimodal and therefore proved as better improvement over the other competing distribution in modeling real lifetime data.

Kan and Han (2015) conducted a goodness of fit test of the inverse Weibull distribution. The maximum likelihood estimators of some specific parameters such as the scale and shape of the inverse Weibull distribution were derived, using multiply type-II censoring samples. The inverse Weibull distribution has two special case namely; inverse Rayleigh

distribution and inverse exponential distribution. The goodness of fit test was conducted using statistics based on the empirical distribution function, and then applied the modified normalized sample Lorenz curve to fit the data.

Oguntunde *et al.* (2014) proposed an alternative distribution to the generalized inverse exponential distribution and the inverse exponential distribution. This new distribution is known as a generalization of the inverted exponential distributions. The proposed distribution is characterized by some special cases and properties. The plots for the probability density function of the new distribution depicts a decreasing and unimodal shapes at various parameter values. Further, the hazard rate function of the proposed distribution indicates that, the hazard function initially increased and then start decreasing at certain points. Thus, the hazard function has an inverted bathtub shape. Maximum likelihood approach was used to estimate the model parameters. The applications to breast and bladder cancer datasets indicates that the proposed model provide a better fits than the other candidate models.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter focused on the statistical techniques that were used to achieve the objective of the study. It was subdivided into eight headings and this include: Inverse exponential, Parameter estimation, Confidence interval for parameters, Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, Goodness-of-fit tests, information criteria, Total time on test and Data and source.

3.1 Inverse Exponential Distribution

If a random variable Y, has a PDF $g(y)$ and a CDF $G(y)$, then PDF of the inverse exponential distribution is represented as:

$$g(y) = \frac{1}{y^2}\exp(-\frac{1}{y}), y > 0, \qquad (3.1)$$

and the corresponding CDF is

$$G(y) = \exp(-\frac{1}{y}), \text{ y} > 0. \qquad (3.2)$$

3.2 Parameter Estimation

When estimating parameters, several methods can be used. However, this study have employed three different estimation methods to obtain the unknown parameters of the proposed generator and this include; ordinary least squares estimation (OLS), maximum likelihood estimation (MLE) and Cramér-Von Mises minimum distance estimation (CVM).

3.2.1 Maximum Likelihood Estimation

The maximum likelihood estimation method is the most widely used estimation method in literature. In this study, the maximum likelihood estimation was employed to estimate the unknown parameters of the proposed distribution for both complete and censored datasets.

3.2.1.1 Maximum Likelihood Estimation for Complete Dataset

If a set of identically and independently distributed random variables comprises $Y_1, Y_2, ..., Y_n$ with PDF $g(y;\xi)$, then the likelihood function is represented as:

$$L(\xi \mid y) = \prod_{i=1}^{n} g(y_i;\xi). \tag{3.3}$$

The log-likelihood function (denoted as ℓ) is expressed as :

$$\ell(\xi \mid y_1, y_2, ..., y_n) = \sum_{i=1}^{n} \log g(y_i;\xi). \tag{3.4}$$

The values of ξ that maximize the probability of obtaining the random sample is determined by:

$$\frac{\partial \ell(\xi \mid y_1, y_2, ..., y_n)}{\partial \xi_i} = 0, i = 1, 2, ..., k. \tag{3.5}$$

3.2.1.2 Maximum Likelihood Estimation for Censored Data

Given a dataset $U = (x_i, \delta_i)$, where x_i is the observations that corresponds to the censored failure times and δ_i is the censoring indicator. If failure is observed, $\delta_i = 1$ and $\delta_i = 1$ if censored. Given that $U = (x_i, \delta_i)$ is independent and identically distributed and $\boldsymbol{\theta}$ is the vector of parameters from the GOIE family of distributions, the likelihood of $\boldsymbol{\theta}$ is given by

$$L=\prod_{i=1}^{n}\left[f\left(x_i;\theta\right)\right]^{\delta_i}\left[1-F\left(x_i;\theta\right)\right]^{1-\delta_i}$$

$$=\prod_{i=1}^{n}\left[f\left(x_i;\theta\right)\right]^{\delta_i}\left[S\left(x_i;\boldsymbol{\theta}\right)\right]^{1-\delta_i},$$

where $S\left(x_i;\boldsymbol{\theta}\right)=1-F\left(x_i;\boldsymbol{\theta}\right)$ is the survival function. The log likelihood function is

$$\ell=\sum_{i=1}^{n}\log\left[\left(f\left(x_i;\theta\right)\right)^{\delta_i}\left(S\left(x_i;\theta\right)\right)^{1-\delta_i}\right]$$

$$=\sum_{i=1}^{n}\left[\delta_i\log\left(f\left(x_i;\theta\right)\right)+\left(1-\delta_i\right)\log\left(S\left(x_i;\theta\right)\right)\right]. \qquad (3.6)$$

Also, the parameters of a regression model with cure fraction can be estimated using this estimation method. This is achieved by substituting the population survival function, $S_{pop}\left(x_i;\theta\right)$ and its density function, $f_{pop}\left(x_i;\theta\right)$ of the distribution into equation (3.6). Hence, the total log-likelihood function with non-informative censoring is given as

$$\ell=\sum_{i=1}^{n}\left[\delta_i\log\left(f_{pop}\left(x_i;\theta\right)\right)+\left(1-\delta_i\right)\log\left(S_{pop}\left(x_i;\theta\right)\right)\right]. \qquad (3.7)$$

The maximum likelihood estimation is the most widely used method of estimation method because of its special properties that are explained below.

3.2.1.3 Consistency

If a population Y has a density function $g(y,\boldsymbol{\xi})$ and consists of a set of random samples which are identical and independently distributed, then $\boldsymbol{\xi}_n$ is a consistent estimator if its mean squared error goes to zero as n approaches infinity.

Therefore, for any $\in > 0$

$$\lim_{n\to\infty} P\left(\left|\xi_n - \xi\right| \geq \in\right) = 0. \qquad (3.8)$$

3.2.1.4 Asymptotic Normality

As sample size increases $(n \to \infty)$, there is a corresponding convergence of the probability distribution of the maximum likelihood estimators to multivariate normal distributions. That is,

$$\sqrt{n}\left(\xi - \xi\right) \xrightarrow{Dist} N\left(\mathbf{0}, I^{-1}\left(\xi\right)\right),$$

where $\mathbf{0}$ is defined as a C-dimensional mean zero vector and $I^{-1}\left(\xi\right)$ is a $C \times C$ dimensional information matrix of Fisher. Which is given as:

$$I\left(\psi\right) = -E\left[\frac{\partial^2 g\left(y \mid \xi\right)}{\partial\xi\partial\xi'}\right] = \int_{-\infty}^{\infty}\left[\frac{\partial^2 g\left(y \mid \xi\right)}{\partial\xi\partial\xi'}\right] g\left(y\right) dy. \qquad (3.9)$$

3.2.1.5 Asymptotic Efficiency

Maximum likelihood estimators are considered asymptotically most efficient since they are the estimators with the least variance when compared to other unbiased estimators within the same class.

Mathematically, if there exist an alternative unbiased estimator $\overline{\xi}$, such that

$$\sqrt{n}\left(\xi - \xi\right) \xrightarrow{Dist} N\left(\mathbf{0}, I^{-1}\left(\phi\right)\right), \qquad (3.10)$$

then $I^{-1}(\phi)$ is greater than or equal to $I^{-1}(\xi)$ always.

3.2.1.6 Invariance Property

If $\hat{\xi}$ is the maximum likelihood estimator of ξ, then $g(\hat{\xi})$ is the maximum likelihood estimator of $g(\xi)$.

3.2.2 Ordinary Least Squares Estimation

The ordinary least square (OLS) estimation method is one of the estimation methods that estimates the parameters by minimizing the objective function. This method was introduced by Swain *et al.* (1988). If $x_{(1)}, x_{(2)}, ..., x_{(n)}$ are order statistics of a random sample of size n obtained from the GOIE distribution. The OLS estimates, $\alpha_{OLS}, \beta_{OLS}, \xi_{OLS}$ for the GOIE distribution parameters can be derived by minimizing the function

$$\varphi(\alpha,\beta,\xi)=\sum_{i=1}^{n}\left[F\left(x_{(i)}\mid\alpha,\beta,\xi\right)-\frac{i}{n+1}\right]^2, \qquad (3.11)$$

with respect to α, β and ξ. Similarly, the OLS can be derived by solving the non-linear equation numerically.

$$\sum_{i=1}^{n}\left[F\left(x_{(i)}\mid\alpha,\beta,\xi\right)-\frac{i}{n+1}\right]^2\Delta_k\left(x_{(i)};\alpha,\beta,\xi\right)=0, \qquad k=1,2,3,$$

where

$$\Delta_1\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\alpha}F\left(x_{(i)}\mid\alpha,\beta,\xi\right),\ \Delta_2\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\beta}F\left(x_{(i)}\mid\alpha,\beta,\xi\right), \text{ and}$$

$$\Delta_3\left(x_{(i)};\alpha,\beta,\boldsymbol{\xi}\right)=\frac{\partial}{\partial\boldsymbol{\xi}}F\left(x_{(i)}\mid\alpha,\beta,\xi\right).$$

3.2.3 Cramér-Von Mises Minimum Distance Estimation

The Cramér-Von Mises minimum **(**CVM) estimation method, also known as goodness-of-fit estimation method has minimum bias compared to other minimum distance estimation methods (MacDonald, 1971). It rely on the difference between the estimates of the CDF and the empirical distribution (Luceno, 2006). Let $x_{(1)},x_{(2)},...,x_{(n)}$ be order statistics of a random sample of size n obtained from the GOIE distribution. The estimators of the CVM for the GOIE distribution parameters can be derived by minimizing the function

$$V\left(\alpha,\beta,\boldsymbol{\xi}\right)=\frac{1}{12n}+\sum_{i=1}^{n}\left[F\left(x_{(i)}\mid\alpha,\beta,\boldsymbol{\xi}\right)-\frac{2i-1}{2n}\right]^2, \qquad (3.12)$$

with respect to α,β and $\boldsymbol{\xi}$. Also, the nonlinear equations can be solved numerically to obtain the estimates of the CVM. That is

$$\sum_{i=1}^{n}\left[F\left(x_{(i)}\mid\alpha,\beta,\boldsymbol{\xi}\right)-\frac{2i-1}{2n}\right]^2\Delta_k\left(x_{(i)};\alpha,\beta,\boldsymbol{\xi}\right)=0, \qquad k=1,2,3,...$$

where

$$\Delta_1\left(x_{(i)};\alpha,\beta,\boldsymbol{\xi}\right)=\frac{\partial}{\partial\alpha}F\left(x_{(i)}\mid\alpha,\beta,\boldsymbol{\xi}\right),\ \Delta_2\left(x_{(i)};\alpha,\beta,\boldsymbol{\xi}\right)=\frac{\partial}{\partial\beta}F\left(x_{(i)}\mid\alpha,\beta,\boldsymbol{\xi}\right),\text{ and}$$

$$\Delta_3\left(x_{(i)};\alpha,\beta,\boldsymbol{\xi}\right)=\frac{\partial}{\partial\boldsymbol{\xi}}F\left(x_{(i)}\mid\alpha,\beta,\boldsymbol{\xi}\right).$$

3.3 Confidence Intervals for Parameters

If a given distribution consists of the parameters $\gamma_1, \gamma_2, ..., \gamma_n$ and their variances $\sum_{11}, \sum_{22}, ..., \sum_{nn}$ respectively, then their $100(1-\eta)\%$ confidence intervals can be estimated by employing the multivariate normal approximation as seen in the equation below:

$\gamma_1 \in \gamma_1 \mp z_\eta / 2\sqrt{\sum_{11}}, \gamma_2 \in \gamma_2 \mp z_\eta / 2\sqrt{\sum_{22}}, ..., \gamma_1 \in \gamma_n \mp z_\eta / 2\sqrt{\sum_{nn}}$, and $z_\eta / 2$ is the upper η^{th} percentile of the standard normal distribution.

3.4 Broyden-Fletcher-Goldfarb-Shanno Algorithm

The Broyden-Fletcher-Goldfarb-Shanno Algorithm (BFGS) method was proposed by Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970). This was to find solutions to a class of equations that have no closed forms. The algorithm for the BFGS is an iterative technique for solving unconstrained non-linear optimization problem (Nasiru, 2018). To optimize a given function, the following steps are express as $\boldsymbol{\xi}_i$ converges to the solution with an initial presumption of $\boldsymbol{\xi}_0$ and estimated Hessian matrix $\boldsymbol{H}_0$.

The first step is to solve $\boldsymbol{H}_i \boldsymbol{u}_i + \nabla \ell(\boldsymbol{\xi}_i) = 0$ so as to get a direction $\boldsymbol{u}_i$.

1. A one dimensional optimisation is then performed to look for a recommended step size β_i in the direction found in step 1.
2. Set $\boldsymbol{c}_i = \beta_i \boldsymbol{u}_i$ and update $\boldsymbol{\xi}_{i+1} = \boldsymbol{\xi}_i + \boldsymbol{c}_i$

3. Let $m_i = \nabla\ell\left(\boldsymbol{\xi}_{i+1}\right) - \nabla\ell\left(\boldsymbol{\xi}_i\right)$

4. $\boldsymbol{H}_{i+1} = \boldsymbol{H}_0 + \frac{m_i m_i'}{m' c_i} - \frac{H_i c_i c_i' H_i}{c_i' H_i c_i}$

The algorithm's convergence is determined by observing the norm of the gradient, $\left|\nabla\ell\left(\xi_i\right)\right|$.

Technically, H_0 can be set with the unit matrix , $\boldsymbol{H}_0 = \boldsymbol{I}$, to make the first step equivalent to a gradient descent, but additional procedures are enhanced, using the estimation of the Hessian, $\boldsymbol{H}_i$ First step of the algorithm is performed using the inverse of $\boldsymbol{H}_i$, which can be efficiently obtained by introducing the Sherman-Morrison formula in the fifth step of the algorithm. Hence,

$$\boldsymbol{H}_{i+1}^{-1} = \left(\boldsymbol{I} - \frac{\boldsymbol{c}_i \boldsymbol{m}_i'}{\boldsymbol{m}_i' \boldsymbol{c}_i}\right) \boldsymbol{H}_i^{-1} \left(\boldsymbol{I} - \frac{\boldsymbol{m}_i' \boldsymbol{c}_i}{\boldsymbol{m}_i' \boldsymbol{c}_i}\right) + \frac{\boldsymbol{c}_i \boldsymbol{c}_i'}{\boldsymbol{m}_i' \boldsymbol{c}_i} \qquad (3.13)$$

Since $\boldsymbol{H}_{i+1}^{-1}$ is symmetric and the terms $\boldsymbol{m}_i' \boldsymbol{H}_i^{-1} \boldsymbol{m}_i$ and $\boldsymbol{c}_i' \boldsymbol{m}_i$ are scalar, equation (3.13) can be estimated more efficiently using the expression

$$\boldsymbol{H}_{i+1}^{-1} = \boldsymbol{H}_i^{-1} + \frac{\left(c_i' m_i + m_i' \boldsymbol{H}_i^{-1} m_i\right)\left(c_i c_i'\right)}{\left(c_i' m_i\right)^2} - \frac{\boldsymbol{H}_i^{-1} m_i c_i' + c_i m_i' \boldsymbol{H}_i^{-1}}{c_i' m_i} \qquad (3.14)$$

3.5 Goodness of Fit Test

The goodness of fit test, which is a statistical hypothesis test, can be used to check how well a distribution fits a given data. It enable us determine whether or not our sample data is representative of what we expect to see in the real population. In the current study, three goodness of fit test will be employed. These are likelihood ratio test, Kolmogorov-Smirnov test and Cramér-Von Mises test.

3.5.1 Likelihood Ratio Test (LRT)

For any two competing nested statistical models, the LRT utilizes the ratio of their likelihoods to determine which of them provides the best fit. Suppose that a random variable has a PDF expressed as $f\left(y;\xi\right)$ with unknown parameter θ. Our interest is to test the null and alternative hypotheses; $H_0:\theta\in\xi_0$ and $H_1:\theta\in\xi_1$. Where ξ_0 and ξ_1 are the parameters of the reduced model and full model respectively. The test statistic for the LRT test is given by

$$\omega=-2\ln\left(\frac{L_0\left(\theta\right)}{L_1\left(\theta\right)}\right), \tag{3.15}$$

where L_0 and L_1 denote the likelihood function for a reduced and full model respectively. Under, H_0,ω will be asymptotically distributed as a Chi-square random variable with degrees of freedom equivalent to the difference amongst the number of parameters of the reduced and full model.

3.5.2 Kolmogorov-Smirnov Test (K-S)

The Kolmogorov-Smirnov test is a nonparametric test used to determine if a sample comes from a given population. The Kolmogorov-Smirnov test calculates the distance from the observed distribution function of the given sample to the estimated CDF of the candidate distribution. The two hypotheses; null and the alternative hypotheses are expressed as;

H_0 : The sample follows the specific distribution versus

H_1 : The sample does not follow the specific distribution

Supposed that $G(y_i)$ is the CDF of the competing distribution at y_i and $G(y_i)$ represents the empirical distribution at y_i. Then the K-S test statistic is given by

$$K-S = \max\left\{\left|G(y_i)-G(y_i)\right|, \left|G(y_i)-G(y_{i-1})\right|\right\}, \qquad (3.16)$$

where

$$G(y_i) = \frac{\#\{y_j : y_j \le y_i\}}{n}.$$

The computed test statistic is then compared with a tabulated K-S at a significance level to judge whether the null hypothesis will be accepted or rejected. If there are more than one distribution to be compared the distribution with the smaller K-S value is the most appropriate to fit the given sample.

3.5.3 Cramér-von Mises Test Statistic

Cramer-von mises test is one of the goodness of fit test that rely on the empirical distribution. It makes use of the summed squared differences between observed and expected cumulative proportions as test statistic. Suppose $G\left(y_i;\boldsymbol{\xi}\right)$ is the CDF such that the procedure of G is known but the k-dimensional parameter vector $\boldsymbol{\xi}$ is not known. The test statistic, W^* is obtained with the following procedure:

1. The set of y_i's should first be organized in increasing order and $G\left(y_i;\boldsymbol{\xi}\right)=u_i$ should be estimated.
2. Evaluate $z_i=\varphi^{-1}\left(u_i\right)$, where $\varphi^{-1}\left(.\right)$, stands for the quantile of the standard normal distribution and $\varphi\left(.\right)$, being CDF.
3. Calculate $W^2=\sum_{i=1}^{n}\left(z_i-\frac{\left(2i-1\right)}{2n}\right)^2+\frac{1}{12n}$
4. Transform W^2 into $W^*=W^2\left(1+\frac{0.5}{n}\right)$ to get the test statistic.

The test statistic W^* with the smallest value is chosen during model selection.

3.6 Information Criteria

The information criteria are considered model selection tools that enable us to compare non-nested models. They are basically likelihood-based measures of model fit that include a penalty for complexity. The mostly widely used ones are; the Corrected Akaike

Information Criterion (AICc), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC).

3.6.1 Akaike Information Criterion (AIC)

Akaike information criterion serves as a guide for selection of model, that is, it is employed to estimate the comparative quality of models for a certain data. The AIC is still the most used tool by researchers in selecting and choosing statistical models. To apply AIC, we start with some optional models, which are regarded as proper models for certain data. The test statistic is represented as:

$$AIC = -2\log L\left(\theta\right) + 2k \ . \qquad (3.17)$$

where k stands for the number of estimated parameters for the model. Model with the least AIC relative to the others is the most desirable. For large samples, the AIC introduces good model selection. However, there are issues of bias associated with the AIC. The AICc was therefore developed to overcome this problem (Sugiura, 1978). Hurvich and Tsai (1989) proved that the AICc improved model selections also in small samples. Also, when the model parameters are large, then the AICc is preferred. The test statistic of the AICc is given by

$$AICc = AIC + \frac{2k\left(k+1\right)}{n-k-1}. \qquad (3.18)$$

3.6.2 Bayesian Information Criterion (BIC)

Bayesian Information Criterion also referred to as the Schwarz Information Criterion (SIC) is partly based on the likelihood function. It is a standard used to select the best models among a limited set of models and was introduced by Schwarz (1978). The central idea of BIC originates from approximating the Bayes factor with the notion that the data is independent and identically distributed. The test statistics for the BIC is presented as:

$$BIC = -2\log L\left(\theta\right) + k\log\left(n\right), \qquad (3.19)$$

where n denotes the sample size as well as $\log L\left(\theta\right)$ given as the natural logarithm of the likelihood function. Like the AIC, the suitable model is the one with the minimum BIC compared to others.

3.7 Total Time on Test

The total test on time (TTT) is a concept introduced by Barlow and Doksum (1972). It has since been developed to identify the reliability of mathematical models and to characterize failure rate. The technique was employed by Aarset (1987) to check if a random sample is from a family of life distributions with bathtub shaped hazard rate. If G is the CDF of a distribution, the TTT-transform is then given as

$$H^{-1}\left(p\right) = \int_0^{G^{-1}(p)} S\left(u\right)du,\ p \in \left[0,1\right], \qquad (3.20)$$

where $S\left(u\right) = 1 - G\left(u\right)$ is the survival function. Also, the scaled TTT-transform is computed using

$$\psi G(p) = \frac{H^{-1}(p)}{H^{-1}(1)}. \qquad (3.21)$$

The curve of $\psi G(p)$ versus $0 \le p \le 1$ is the scaled TTT-transform curve. Barlow and Doksum (1972), classified the shape of the hazard rate function using the scaled TTT-transform curve;

1. A concave curve of the scaled TTT-transform above the 45^0 line signifies an increasing hazard rate function.
2. A convex curve of the scaled TTT-transform below the 45^0 line signifies an increasing hazard rate function.
3. A convex curve of the scaled TTT-transform below the 45^0 line and a subsequent concave curve above the 45^0 line forms a bathtub shape for the hazard rate function.
4. A concave curve of the scaled TTT-transform above the 45^0 line and a subsequent convex curve below the 45^0 line forms an upside down bathtub shape for the hazard rate function.

Given an ordered sample $Y_{1:n}, Y_{2:n}, ..., Y_{n:n}$, the total test statistics are calculated using

$$TTT_i = \sum_{j=1}^{j} (n - j + 1)\left(y_{j:n} - y_{j-1:n}\right), i = 1, 2, ..., n. \qquad (3.22)$$

The empirical scaled TTT-transform is represented as:

$$TTT^* = \frac{TTT_i}{TTT'_n}, \qquad (2.23)$$

where $0 \leq TTT_n \leq 1$. The empirical scaled TTT-transform curve obtained by plotting $\frac{i}{n}$ against TTT^*.

3.8 Data and Source

The study employed eight secondary cancer datasets (four complete and four censored) to demonstrate the applications of the GOIE family of distributions. The first dataset is made up of 40 patients suffering from leukemia. The dataset is presented in Table 3.1 and can be found in Abouammoh *et al.* (1994).

Table 3.1: Data on 40 patients suffering from leukemia

115	461	807	1062	1251	1408	1578	1696
181	516	865	1063	1277	1455	1578	1735
255	739	924	1165	1290	1478	1599	1799
418	743	983	1191	1357	1222	1603	1815
441	789	1024	1222	1369	1549	1605	1852

The second dataset consist of the remission time (in months) of a random sample of 128 bladder cancer patients and this is shown in Table 3.2. The dataset can be found in Lee and Wang (2003).

Table 3.2: Data on the remission times of 128 bladder cancer patients

0.08	6.97	2.46	9.74	3.88	15.96	4.26	79.05	11.79	8.37	12.07
2.09	9.02	3.64	14.76	5.32	36.66	5.41	1.35	18.1	12.02	21.73
3.48	13.29	5.09	26.31	7.39	1.05	7.63	2.87	1.46	2.02	2.07
4.87	0.4	7.26	0.81	10.34	2.69	17.12	5.62	4.4	3.31	3.36
6.94	2.26	9.47	2.62	14.83	4.23	46.12	7.87	5.85	4.51	6.93
8.66	3.57	14.24	3.82	34.26	5.41	1.26	11.64	8.26	6.54	8.65
13.11	5.06	25.82	5.32	0.9	7.62	2.83	17.36	11.98	8.53	12.63
23.63	7.09	0.51	7.32	2.69	10.75	4.33	1.4	19.13	12.03	22.69
0.2	9.22	2.54	10.06	4.18	16.62	5.49	3.02	1.76	20.28	
2.23	13.8	3.7	14.77	5.34	43.01	7.66	4.34	3.25	2.02	
3.52	25.74	5.17	32.15	7.59	1.19	11.25	5.71	4.5	3.36	
4.98	0.5	7.28	2.64	10.66	2.75	17.14	7.93	6.25	6.76	

The third dataset is presented in Table 3.3 and it shows the data on the death times (in weeks) of patients with cancer of the tongue with aneuploidy DNA profile. This dataset can be found in Oguntunde *et al.* (2016).

Table 3.3: Cancer of the tongue with aneuploidy DNA profile

1	16	41	74*	89*	104*	231*
3	24	51	77	91	108*	240*
3	26	61*	79*	93*	109*	400*
4	27	65	80*	96	120*	
10	28	67	81*	97*	131*	
13	30	70	87*	100	150*	
13	30	72	87*	101*	157	
16	32	73	88*	104	167	

Asterisks represent censored observation

The fourth dataset is focus on the survival times of a group of patients suffering from Head and Neck cancer diseases treated using a combination of radiotherapy and chemotherapy (RT + CT). This data can be found in Shanker *et al.* (2015).

Table 3.4: Head and Neck cancer dataset

12.2	47.38	81.43	127	173	319	817
23.56	55.46	84	130	179	339	1776
23.74	58.36	92	133	194	432	
25.87	63.47	94	140	195	469	
31.98	68.46	110	146	209	519	
37	78.26	112	155	249	633	
41.35	74.47	119	159	281	725	

The fifth dataset is on the survival times of 121 patients with breast cancer obtained from a large hospital from 1929 to 1938 is given in Table 3.5. This data can be found in Oguntunde *et al.* (2016) and Ramos *et al.* (2013).

Table 3.5: Data on Survival times of 121 patients with Breast cancer

0.3	11	17.2	24	38	44	56	69	115
0.3	11.8	17.3	27.9	38	45	57	78	117
4.0	12.2	17.5	28.2	39	45	58	80	125
5.0	12.3	17.9	29.1	39	46	59	83	126
5.6	13.5	19.8	30	40	46	60	88	127
6.2	14.4	20.4	31	40	47	60	89	129
6.3	14.4	20.9	31	40	48	60	90	129
6.6	14.8	21.0	32	41	49	61	93	139
6.8	15.5	21.0	35	41	51	62	96	154
7.4	15.7	21.1	35	41	51	65	103	
7.5	16.2	23.0	37	42	51	65	105	
8.4	16.3	23.4	37	43	52	67	109	
8.4	16.5	23.6	37	43	54	67	109	
10.3	16.8	24.0	38	43	55	68	111	

The sixth dataset is focus on Head and neck cancer data with censored observations as presented in Table 3.6. This data was first presented by Efron (1988).

Table 3.6: Data on head and neck cancer with censored observations

7	91	140	173	277	523	1349*
34	108	146	176	279*	523*	1412*
42	112	149	185*	297	583	1417
63	129	154	218	319*	594	
64	133	157	225	405	1101	
74*	133	160	241	417	1116*	
83	139	160	248	420	1146	
84	140	165	273	440	1226*	

Asterisks represent censored observation

The seventh dataset is shown in Table 3.7, which refers to the remission times (in weeks) for a group of 30 patients with leukemia receiving similar treatments. This data can be found in Lawless (2003).

Table 3.7: Data on leukemia with censored observations

1	1	2	4	4	6	6	6	7
8	9	9	10	12	13	14	18	19
24	26	29	31*	42	45*	50*	57	
60	71*	85*	91					

Asterisks represent censored observation

Table 3.8 illustrate the eighth dataset and it represent 201 patients with gastric adenocarcinoma. This data can be found in Arslan *et al.* (2018), Ortega *et al.* (2017) and Martinez *et al.* (2013).

Table 3.8: Data on 76 patients receiving adjuvant chemoradiotherapy and 125 receiving resection alone

Surgery Alone								
0.1	1.61	6.15	10.2	16.09	21.48	25.33*	31.38*	35.3*
0.2	1.78	6.55*	10.53	16.18	21.84*	25.36*	31.68*	35.59*
0.23	2.63*	6.91	10.76	16.94	21.88*	26.15*	32.5*	36*
0.26	2.73	7.17	11.41	16.94*	22.14	26.32	32.8*	36*
0.3	2.8	7.7	11.88	17.14	22.99*	26.32*	33.09*	36*
0.33	2.89	7.93*	12.5	17.24	23.39*	26.78*	33.36*	36*
0.49	2.96	8.32	13.13*	17.43	23.52*	27.37*	33.65*	36*
0.53	3.32	8.36	13.95	18.62	23.55*	28.93*	33.91*	36*
0.56	3.49*	8.39	14.01	19.14	23.85*	29.28*	34.08*	36*
0.63	4.01	8.78	14.05	19.44	24.01*	29.31*	34.21*	36*
0.66	4.54	8.91	14.34	19.84*	24.57*	29.97*	34.41*	36*
0.66	4.67	9.28	14.38*	19.93	24.8*	30.16*	35*	36*
1.18	4.67	9.7	15.43	20.49*	25.26*	30.49*	35.03*	36*
1.45*	4.93	10.03	15.76*	21.38*	25.3	30.63*	35.2*	
Adjuvant Chemoradiotherapy								
5.76	12.83	17.07	21.02	25.23	28.22*	31.84*	36*	36*
7.89	13.09	17.14	22.86	25.33*	28.59	32.4*	36*	36*
8.85	13.49	17.34*	23.39*	25.56*	28.65*	32.47*	36*	36*
8.95	13.78	17.7	23.82*	25.59*	29.08*	32.53*	36*	36*
9.05	13.82	18.39	24.21	25.76*	29.31*	33.88*	36*	
9.47	14.7	19.21	24.21*	25.79	30.26	34.74*	36*	
10.72	14.77	19.38*	24.31	25.79*	30.69*	34.9*	36*	
11.97	16.38	20.49*	24.28*	26.05	30.95*	35.72*	36*	
12.5	16.51	20.76*	24.97*	27.89*	31.64*	35.89*	36*	

Asterisks represent censored observation

CHAPTER FOUR

THEORETICAL RESULTS

4.1 Introduction

This chapter presents the theoretical results obtained in this study. This includes derivation of the proposed family of distributions, the statistical properties of the new family, estimators for estimating the parameters of the proposed distribution and the derivation of regression models with cure fraction.

4.2 Generalized Odd Inverse Exponential Family of Distributions

The generalized odd inverse exponential (GOIE) family of distributions is developed in this section. Let $g(x;\boldsymbol{\xi})$ and $G(x;\boldsymbol{\xi})$ be the PDF and CDF of a random variable X respectively, and $\boldsymbol{\xi}$ be a $p\times 1$ vector of parameters. The CDF of the GOIE family of distribution is defined as:

$$F(x) = \int_0^{\frac{G(x;\boldsymbol{\xi})^{\alpha}}{1-G(x;\boldsymbol{\xi})^{\beta}}} x^{-2}e^{-x^{-1}}dx$$
$$= \exp\left[-\frac{1-G(x;\boldsymbol{\xi})^{\beta}}{G(x;\boldsymbol{\xi})^{\alpha}}\right], \quad \alpha = 0, \beta > 0, \ x \in \mathbb{R}. \qquad (4.1)$$

The GOIE family generalizes other odd families of distributions arising from the inverse exponential distribution. For instance, when $\alpha = \beta = 1$, the odd inverse exponential family is obtained and its CDF is given by

$$F(x) = \exp\left[-\frac{1-G(x;\boldsymbol{\xi})}{G(x;\boldsymbol{\xi})}\right], \quad x \in \mathbb{R}.$$

The CDF of the exponentiated odd inverse exponential family is obtained when $\alpha = \beta \neq 1$ and is defined as

$$F(x)=\exp\left[-\frac{1-G(x;\xi)^{\alpha}}{G(x;\xi)^{\alpha}}\right],\quad x\in\mathbb{R}.$$

The CDF of the GOIE family of distributions can be interpreted as follows: Suppose that the odd that a cancer patient will die at time x is $\frac{G(x;\xi)^{\alpha}}{1-G(x;\xi)^{\beta}}$. Then the probability of these odds of death represented by a random variable X that follows the inverse exponential distribution is $P\left(X\le\frac{G(x;\xi)^{\alpha}}{1-G(x;\xi)^{\beta}}\right)=F(x)$, which is given in equation (4.1). Hence, a random variable X that follows the GOIE distribution is denoted by $X\sim GOIE(x;\alpha,\beta,\xi)$ and for the sake of simplicity $G(x;\xi)$ may be represented by $G(x)$.

However, if $\alpha\neq\beta$, then we have the concept of relative odd. The corresponding PDF is obtained by differentiating the CDF and is given by

$$f(x)=g(x;\xi)\left[(\beta-\alpha)G(x;\xi)^{\beta-\alpha-1}+\alpha G(x;\xi)^{-\alpha-1}\right]\exp\left[\frac{G(x;\xi)^{\beta}-1}{G(x;\xi)^{\alpha}}\right],\ \alpha>0,\ \beta>0,\ x\in\mathbb{R},\quad (4.2)$$

where α and β are positive shape parameters and ξ is vector of parameters for the baseline $G(x;\xi)$.

Proposition 4.1. The CDF of the GOIE family of distribution is well defined.

Proof. For the CDF to be well defined, $F(x)\in[0,1]$.

As $x\to-\infty$, $G(x;\xi)\to 0$. Hence, $F(x)\to 0$. Also, as $x\to\infty$, $G(x;\xi)\to 1$. Thus, $F(x)\to 1$. Since $F(x)\in[0,1]$, the CDF is well defined.

Proposition 4.2. The PDF of the GOIE family of distribution is a legitimate density function.

Proof. For the PDF to be legitimate, $\int_{-\infty}^{\infty} f(x)dx = 1.$

Hence, we need to show that when the PDF of the GOIE family is integrated over its support we will get 1.

$$\int_{-\infty}^{\infty} g(x)\left[\frac{(\alpha-\beta)G(x)^{\beta}}{G(x)^{\alpha+1}}\right] e^{-\left[\frac{1-G(x)^{\beta}}{G(x)^{\alpha}}\right]} dx.$$

Let

$$y = \left[\frac{1-G(x)^{\beta}}{G(x)^{\alpha}}\right] \quad .$$

As $x \to -\infty$, $G(x) \to 0$ and $y \to \infty$, as $x \to \infty$, $G(x) \to 1$ and $y \to 0$. Also,

$$\frac{dy}{dx} = -g(x)\left[\frac{(\alpha-\beta)G(x)^{\beta}+\alpha}{G(x)^{\alpha+1}}\right].$$

This implies that

$$dy = -g(x)\left[\frac{(\alpha-\beta)G(x)^{\beta}+\alpha}{G(x)^{\alpha+1}}\right] dx.$$

Hence, $-\int_{\infty}^{0} e^{-y}dy = \int_{0}^{\infty} e^{-y}dy = \left[e^{-y}\right]_0^{\infty} = 0-(-1) = 1$. Thus, the proof is complete.

The survival function is very useful when studying the average time to death or recovery of a cancer patient. Hence, the survival function of the GOIE family is given by

$$S(x) = 1 - \exp\left[-\frac{1-G(x;\xi)^{\beta}}{G(x;\xi)^{\alpha}}\right],\ \alpha>0,\ \beta>0,\ x \in \mathbb{R}. \qquad (4.3)$$

The hazard rate function of the GOIE family is therefore given by

$$h(x)=\frac{g(x;\xi)\left[(\beta-\alpha)G(x;\xi)^{\beta-\alpha-1}+\alpha G(x;\xi)^{-\alpha-1}\right]\exp\left[\frac{G(x;\xi)^{\beta}-1}{G(x;\xi)^{\alpha}}\right]}{1-\exp\left[\frac{G(x;\xi)^{\beta}-1}{G(x;\xi)^{\alpha}}\right]},\alpha>0,\beta>0,x\in\mathbb{R}. \quad (4.4)$$

Lemma 4.1. The density function of the GOIE family can be expressed in a mixture form as

$$f(x)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}g(x)G(x)^{m}, \quad (4.5)$$

where

$$\omega_{ijkm}=\frac{(-1)^{i+j+k+m}}{i!}\binom{i}{j}\binom{k}{m}\left[(\beta-\alpha)\binom{\beta(j+1)-\alpha(i+1)-1}{k}+\alpha\binom{\beta j-\alpha(i+1)-1}{k}\right].$$

Proof. Using the Taylor series expansion $e^{-z}=\sum_{i=0}^{\infty}\frac{(-1)^i Z^i}{i!}$, the PDF can be written as

$$f(x)=\left[(\beta-\alpha)g(x)G(x)^{\beta-\alpha-1}+\alpha g(x)G(x)^{-\alpha-1}\right]\sum_{i=0}^{\infty}\frac{(-1)^i\left[\frac{1-G(x)^{\beta}}{G(x)^{\alpha}}\right]^i}{i!}.$$

Thus,

$$f(x)=(\beta-\alpha)g(x)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}G(x)^{\beta-\alpha(i+1)-1}\left[1-G(x)^{\beta}\right]^i+\alpha g(x)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}G(x)^{-\alpha(i+1)-1}\left[1-G(x)^{\beta}\right]^i.$$

Using the binomial expansion $(1-x)^n=\sum_{j=0}^{n}(-1)^j\binom{n}{j}x^j$, $|x|<1$,

$$f(x)=(\beta-\alpha)g(x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\frac{(-1)^{i+j}}{i!}\binom{i}{j}G(x)^{\beta(j+1)-\alpha(i+1)-1}+\alpha g(x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\frac{(-1)^{i+j}}{i!}\binom{i}{j}G(x)^{\beta j-\alpha(i+1)-1}.$$

But

$$\left[1-(1-G(x)\right]=G(x).$$

This implies

$$f(x)=(\beta-\alpha)g(x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\frac{(-1)^{i+j}}{i!}\binom{i}{j}\left[1-(1-G(x))\right]^{\beta(j+1)-\alpha(i+1)-1}+\alpha g(x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\frac{(-1)^{i+j}}{i!}\binom{i}{j}\left[1-(1-G(x))\right]^{\beta j-\alpha(i+1)-1}.$$

Applying the binomial series expansion twice, yields

$$f(x)=(\beta-\alpha)g(x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{(-1)^{i+j+k+m}}{i!}\binom{i}{j}\binom{\beta(j+1)-\alpha(i+1)-1}{k}\binom{k}{m}G(x)^{m}+$$

$$\alpha g(x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{(-1)^{i+j+k+m}}{i!}\binom{i}{j}\binom{\beta j-\alpha(i+1)-1}{k}\binom{k}{m}G(x)^{m}.$$

Thus

$$f(x)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}g(x)G(x)^{m}.$$

The proof is therefore complete.

Alternatively, the density function can be expressed in terms of the exponentiated-G (exp-G) density function as

$$f(x)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega^{*}_{ijkm}\ \Delta_{m+1}(x), \qquad (4.7)$$

where $\omega^{*}_{ijkm}=\frac{\omega_{ijkm}}{m+1}$, $\Delta_{m+1}(x)=(m+1)g(x)G(x)^{m}$ is the $\exp-G$ density function with power parameter $m+1$. By integrating equation (4.7), the CDF of the mixture representation becomes

$$F(x)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega^{*}_{ijkm}\ \Omega_{m+1}(x), \qquad (4.8)$$

where $\Omega_{m+1}(x) = G(x)^{m+1}$ is the CDF of the exp-G family with power parameter $m+1$.

4.3 Statistical Properties of the GOIE family

This section presents the statistical properties of the GOIE family of distribution.

4.3.1 Quantile Function

Quantile functions are very useful when simulating random samples from a given distribution. The median, skewness and kurtosis of a distribution can also be derived using the quantile functions.

Proposition 4.3. The quantile function of the GOIE family for $u \in [0,1]$ is given by

$$G(x;\xi)^{\alpha} \log(u) - G(x;\xi)^{\beta} + 1 = 0\ ,\ u \in [0,1]. \qquad (4.9)$$

Proof. Using the CDF of the GOIE family defined in equation (4.1), the quantile function is obtained as follows;

$$F(x) = \exp\left[\frac{G(x;\xi)^{\beta} - 1}{G(x;\xi)^{\alpha}}\right] = u.$$

This implies

$$\exp\left[\frac{G(x;\xi)^{\beta} - 1}{G(x;\xi)^{\alpha}}\right] = u.$$

Taking logarithm of both sides and simplifying yields

$$G(x;\xi)^{\alpha} \log(u) = G(x;\xi)^{\beta} - 1.$$

Hence,

$$G(x;\xi)^{\alpha} \log(u) - G(x;\xi)^{\beta} + 1 = 0,\ \ u \in [0,1].$$

Since the quantile function does not have a closed form, it is therefore solved numerically to obtain the quantile values. The first quartile, median and upper quartile of the GOIE distribution are obtained by substituting $u = 0.25, 0.5$ and 0.75 respectively.

4.3.2 Moment

In any statistical analysis, the moment plays an important role. It is very useful when computing measures of shapes, central tendencies and dispersion.

Proposition 4.4. The r^{th} non-central moment of the GOIE family of distributions is

$$\mu_r^{'} = \sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k} \omega_{ijkm} \int_{-\infty}^{\infty} x^r g(x) G(x)^m \, dx, \quad r = 1, 2, ... \tag{4.10}$$

Proof. The r^{th} non-central moment is defined as

$$\mu_r^{'} = \int_{-\infty}^{\infty} x^r f(x) dx.$$

Substituting the mixture representation of the density function into the definition gives

$$\begin{aligned} \mu_r^{'} &= \int_{-\infty}^{\infty} x^r \sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k} \omega_{ijkm} g(x) G(x)^m dx \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k} \omega_{ijkm} \int_{-\infty}^{\infty} x^r g(x) G(x)^m \, dx, \quad r = 1, 2, ... \end{aligned}$$

This complete the proof.

4.3.3 Moment Generating Function

The moment generating function (MGF) are useful functions when estimating the moments of a random variable. The MGF of a random variable X having the GOIE distribution if it exist is given by the following proposition.

Proposition 4.5. The MGF of the GOIE family of distribution is given by

$$M_X(t)=\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{\omega_{ijkm}}{r!}\int_{-\infty}^{\infty}x^r g(x)G(x)^m dx,\quad r=1,2,.... \tag{4.11}$$

Proof. Using the identity $e^{tX}=\sum_{r=0}^{\infty}\frac{t^r X^r}{r!}$.

The MGF of a random variable is

$$M_X(t)=E\left[e^{tX}\right]$$

$$=\sum_{r=0}^{\infty}E\left[\frac{t^r X^r}{r!}\right]$$

$$=\sum_{r=0}^{\infty}\frac{t^r}{r!}E\left[X^r\right].$$

But $E\left[X^r\right]$ is the r^{th} non-central moment. This implies

$$M_X(t)=\sum_{r=0}^{\infty}\frac{t^r}{r!}\mu_r^{'}.$$

Hence,

$$M_X(t)=\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{\omega_{ijkm}}{r!}\int_{-\infty}^{\infty}x^r g(x)G(x)^m dx.$$

The proof is complete.

4.3.4 Incomplete Moment

The incomplete moments is very useful when estimating the median deviation, mean deviation and measures of inequalities such as Lorenz and Bonferroni curves.

Proposition 4.6. The r^{th} incomplete moment of the GOIE distribution is given by

$$M_r(y)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}x^r g(x)G(x)^m\,dx,\qquad r=1,2,.... \tag{4.12}$$

Proof. The r^{th} incomplete moment is defined as

$$M_r(y)=E\left[X\mid X<Y\right].$$

For continuous random variable,

$$E\left[X\mid X<Y\right]=\int_{-\infty}^{y}x^r f(x)dx.$$

Substituting the mixture representation of the density function into the definition yields

$$M_r(y)=\int_{-\infty}^{y}x^r\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}g(x)G(x)^m\,dx.$$

Hence,

$$M_r(y)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}x^r g(x)G(x)^m\,dx.$$

The incomplete moment can also be expressed in terms of the quantile function. Let $G(x)=u$, then $x=Q_{G(u)}$, where $Q_{G(.)}=G^{-1}$. $\frac{du}{dx}=g(x)$, it implies that $du=g(x)dx$. As $x\to-\infty,\ G(x)\to 0$ and $x\to y,\ G(x)\to G(y)$. Hence,

$$M_r(y)=\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{0}^{G(y)}Q_G(u)^r u^m du.$$

4.3.5 Characteristic Function

The characteristic function play an important role in statistics and probability theory. When the MGF of a random variable does not exist, the characteristic function becomes more useful.

Proposition 4.7. The characteristic function of the GOIE distribution is given by

$$\phi_X(t) = \sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{p}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{\omega_{ijkm}}{r!}\int_{-\infty}^{\infty}(it)^r\, g(x)G(x)^m dx, \quad r=1,2,\ldots., \; i=\sqrt{-1}. \tag{4.13}$$

Proof. Using the identity

$$e^{itX} = \sum_{r=0}^{\infty}\frac{i^r t^r X^r}{r!},$$

the characteristic function of the GOIE distribution is

$$\phi_X(t) = E\left[e^{itX}\right], \; i=\sqrt{-1}.$$

Using Taylor series expansion

$$E\left[e^{itX}\right] = \sum_{r=0}^{\infty}E\left[\frac{i^r t^r X^r}{r!}\right]$$

$$= \sum_{r=0}^{\infty}\frac{i^r t^r}{r!}E\left[X^r\right].$$

This implies

$$\phi_X(t) = \sum_{r=0}^{\infty}\frac{i^r t^r}{r!}\mu_r'.$$

Hence,

$$\phi_X(t) = \sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{\omega_{ijkm}}{r!}\int_{-\infty}^{\infty}(it)^r\, g(x)G(x)^m dx.$$

4.3.6 Inequality Measures

Lorenz and Bonferroni curves are the most widely used methods for estimating the income inequality of a given population. It can also be used to assess the inequality in terms of the survival times of cancer patients in a given population.

Proposition 4.8. If $Y\sim$ GOIE (α,β,ξ), then the Lorenz curve $L_F(y)$ is given by

$$L_F(y)=\frac{1}{\mu}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}xg(x)G(x)^m\,dx. \qquad (4.14)$$

Proof. The Lorenz curve is defined as

$$L_F(y)=\frac{1}{\mu}\int_{-\infty}^{y}xf(x)\,dx.$$

But $\int_{-\infty}^{y}xf(x)\,dx$ is the first incomplete moment. Hence,

$$L_F(y)=\frac{1}{\mu}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}xg(x)G(x)^m\,dx.$$

Proposition 4.9. If $Y\sim$ GOIE (α,β,ξ), then the Bonferroni curve $B_F(y)$ is given by

$$B_F(y)=\frac{1}{\mu\exp\left[-\frac{1-G(y)^{\beta}}{G(y)^{\alpha}}\right]}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}xg(x)G(x)^m\,dx. \qquad (4.15)$$

Proof. By definition

$$B_F(y)=\frac{L_F(y)}{F(y)}.$$

This implies

$$B_F(y)=\frac{1}{\mu\exp\left[-\frac{1-G(y)^{\beta}}{G(y)^{\alpha}}\right]}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}xg(x)G(x)^m\,dx.$$

4.3.7 Mean Residual Life

The Mean Residual Life (MRL) function plays a vital role in survival analysis. The Mean Residual Life function, $m(y)$ characterizes the distribution function, $F(y)$ uniquely (Kotz and Shanbhag, 1980).

Proposition 4.10. If Y is a random variable which represents the life of a component with distribution function $F(y)$, then the MRL is defined as

$$m(y)=\frac{1}{\left[1-\exp\left[-\frac{1-G(y)^{\beta}}{G(y)^{\alpha}}\right]\right]}\left[\mu-\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}xg(x)G(x)^{m}\,dx\right]-y, \tag{4.16}$$

where $\mu=\mu_1^{'}$.

Proof. By definition $m(y)=E[X-y\,|\,X>y]$.

$$m(y)=\frac{\int_{y}^{\infty}(x-y)f(x)dx}{1-F(y)}$$

$$=\frac{\mu_1^{'}-\int_{-\infty}^{y}xf(x)dx}{1-F(y)}-y.$$

But $\int_{-\infty}^{y}xf(x)dx$ is the first incomplete moment and $1-F(y)$ is the survival function. Substituting the first incomplete moment yields

$$m(y)=\frac{1}{\left[1-\exp\left[-\frac{1-G(y)^{\beta}}{G(y)^{\alpha}}\right]\right]}\left[\mu-\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{y}xg(x)G(x)^{m}\,dx\right]-y.$$

4.3.8 Mean and Median Deviation

The combined deviations of the mean and median to some extent can be used to measure the variation in a population. If the random variable X follows the GOIE distribution, then the following propositions represents the mean and median deviations respectively.

Proposition 4.11. The mean deviation of a random variable X having the GOIE distribution is given by

$$\delta_1 = 2\mu F(\mu) - 2\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{\mu} xg(x)G(x)^m\,dx. \tag{4.17}$$

Proof. By definition

$$\begin{aligned}\delta_1 &= \int_{-\infty}^{\infty}|x-\mu|\,f(x)dx \\ &= 2\mu F(\mu) - 2\int_{-\infty}^{\mu} xf(x)dx.\end{aligned}$$

But $\int_{-\infty}^{y} xf(x)\,dx$ is the first incomplete moment. Substituting the first incomplete moment gives

$$\delta_1 = 2\mu F(\mu) - 2\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{\mu} xg(x)G(x)^m\,dx.$$

Proposition 4.12. The median deviation of a random variable X having the GOIE distribution is given by

$$\delta_2 = \mu - 2\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{m} xg(x)G(x)^m\,dx. \tag{4.18}$$

Proof. By definition

$$\delta_2 = \int_{-\infty}^{\infty} |x-m| f(x)dx$$

$$= \mu - 2\int_{-\infty}^{m} xf(x)dx.$$

But $\int_{-\infty}^{y} xf(x)dx$ is the first incomplete moment. Hence,

$$\delta_2 = \mu - 2\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{k}\omega_{ijkm}\int_{-\infty}^{m} xg(x)G(x)^m dx.$$

4.3.9 Order Statistics

Order statistics plays a vital role in the field of biostatistics. For example the least order statistics, average order statistics and the highest order statistics can be used respectively to estimate the minimum, average and maximum time it will take for a patient to die. Therefore, this subsection is focused on deriving the p^{th} order statistics of the GOIE distribution. Let $X_1, X_2, ..., X_n$ be a random sample from the GOIE distribution and $X_{1:n} < X_{2:n} < ... < X_{n:n}$ are order statistics obtained from the sample, then the PDF, $f_{p:n}(x)$, of the p^{th} order statistic $X_{n:p}$ is given by

$$f_{p:n}(x) = \frac{n!}{(p-1)!(n-p)!}\sum_{i=0}^{n-p}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\infty}\sum_{q=0}^{m}\bar{\omega}_{ijkmq}g(x)G(x)^q, \qquad (4.19)$$

where

$$\bar{\omega}_{ijkmq} = \frac{(-1)^{i+j+k+m+q}(p+i)^j}{j!}\binom{n-p}{i}\binom{j}{k}\binom{m}{q}\left[(\beta-\alpha)\binom{\beta(k+1)-\alpha(j+1)-1}{m}+\alpha\binom{\beta k-\alpha(j+1)-1}{m}\right].$$

Proof. By definition, the density function of the p^{th} order statistics is

$$f_{p:n}(x)=\frac{n!}{(p-1)!(n-p)!}\left[F(x)\right]^{p-1}\left[1-F(x)\right]^{n-p}\ f(x).$$

Using the binomial expansion

$$f_{p:n}(x)=\frac{n!}{(p-1)!(n-p)!}\left[F(x)\right]^{p-1}\sum_{i=0}^{n-p}(-1)^{i}\binom{n-p}{i}\left[F(x)\right]^{i}f(x).$$

This implies

$$f_{p:n}(x)=\frac{n!}{(p-1)!(n-p)!}\sum_{i=0}^{n-p}(-1)^{i}\binom{n-p}{i}\left[F(x)\right]^{p+i-1}f(x).$$

But

$$F(x)=\exp\left[\frac{G(x)^{\beta}-1}{G(x)^{\alpha}}\right]$$

and

$$f(x)=\left[(\beta-\alpha)g(x)G(x)^{\beta-\alpha-1}+\alpha g(x)G(x)^{-\alpha-1}\right]\exp\left[\frac{G(x)^{\beta}-1}{G(x)^{\alpha}}\right].$$

Hence,

$$f_{p:n}(x)=\frac{n!}{(p-1)!(n-p)!}\sum_{i=0}^{n-p}(-1)^{i}\binom{n-p}{i}\left[(\beta-\alpha)g(x)G(x)^{\beta-\alpha-1}+\alpha g(x)G(x)^{-\alpha-1}\right]\exp\left[(p+i)\left[\frac{G(x)^{\beta}-1}{G(x)^{\alpha}}\right]\right].$$

Using Taylor series expansion,

$$f_{p:n}(x)=\frac{n!}{(p-1)!(n-p)!}\sum_{i=0}^{n-p}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}(p+i)^{j}}{j!}\binom{n-p}{i}\left[(\beta-\alpha)g(x)G(x)^{\beta-\alpha(j+1)-1}+\alpha g(x)G(x)^{-\alpha(j+1)-1}\right]\left[1-G(x)^{\beta}\right]^{j}.$$

Applying the binomial expansion twice, yields

$$f_{p:n}(x)=\frac{n!}{(p-1)!(n-p)!}\sum_{i=0}^{n-p}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\infty}\sum_{q=0}^{m}\overline{\omega}_{ijkmq}g(x)G(x)^{q}.$$

4.3.10 Moment of the p^{th} Order Statistic

Proposition 4.13. The r^{th} non-central of the p^{th} order statistic of a random variable X having the GOIE distribution is given by

$$\mu_r^{'(p:n)} = \sum_{i=0}^{n-p}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\infty}\sum_{q=0}^{m}\overline{\omega}_{ijkmq}\int_{-\infty}^{\infty} x^r g(x)G(x)^q\,dx. \tag{4.20}$$

Proof. By definition

$$\mu_r^{'(p:x)} = \int_{-\infty}^{\infty} x^r f_{p:x}(x)\,dx.$$

But $f_{p:n}(x)$ is the PDF of the p^{th} order statistic. This implies

$$\mu_r^{'(p:n)} = \int_{-\infty}^{\infty} x^r \sum_{i=0}^{n-p}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\infty}\sum_{q=0}^{m}\overline{\omega}_{ijkmq} g(x)G(x)^q\,dx.$$

Hence,

$$\mu_r^{'(p:n)} = \sum_{i=0}^{n-p}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\infty}\sum_{q=0}^{m}\overline{\omega}_{ijkmq}\int_{-\infty}^{\infty} x^r g(x)G(x)^q\,dx.$$

4.4 Parameter Estimation

In this section, we employed three estimation methods to obtain the unknown parameters of the GOIE distribution. These estimation methods are; maximum likelihood estimation, ordinary least squares estimation and Cramér-Von Mises minimum distance estimation.

4.4.1 Maximum likelihood Estimation for Complete Dataset

Given that $X_1, X_2, ..., X_n$ is a random sample of n independently and identically distributed random variables from the GOIE distribution, then the total log-likelihood function is given by;

$$\ell = \sum_{i=1}^{n} \log g(x_i;\xi) + \sum_{i=1}^{n} \log\left[(\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1} + \alpha G(x_i;\xi)^{-\alpha-1}\right] - \sum_{n-1}^{n}\left[G(x_i;\xi)^{-\alpha} + G(x_i;\xi)^{\beta-\alpha}\right]. \quad (4.22)$$

Differentiating the total log-likelihood function with respect to the parameters yield the following score functions;

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n}\left[\frac{G(x;\xi)^{-\alpha-1} - G(x_i;\xi)^{\beta-\alpha-1} - \alpha G(x_i;\xi)^{-\alpha-1}\log G(x_i;\xi) - (\beta-\alpha)G(x_i;\xi)\log G(x_i;\xi)}{(\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1} + \alpha G(x_i;\xi)^{-\alpha-1}}\right] -$$
$$\sum_{i=1}^{n}\left[G(x_i;\xi)^{-\alpha}\log G(x_i;\xi) + G(x_i;\xi)^{\beta-\alpha}\log G(x_i;\xi)\right], \quad (4.23)$$

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n}\left[\frac{G(x_i;\xi)^{\beta-\alpha-1} - (\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1}\log G(x_i;\xi)}{\alpha G(x_i;\xi)^{\alpha-1} + (\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1}}\right] + \sum_{i=1}^{n}\left[G(x_i;\xi)^{\beta-\alpha}\log G(x_i;\xi)\right], \quad (4.24)$$

and

$$\frac{\partial \ell}{\partial \xi} = \sum_{i=1}^{n}\left[\frac{g'(x_i;\xi)}{g(x_i;\xi)}\right] + \sum_{i=1}^{n}\left[\frac{(\beta-\alpha)(\beta-\alpha-1)G'(x_i;\xi)G(x_i;\xi)^{\beta-\alpha-2} + \alpha(-\alpha-1)G'(x_i;\xi)G(x_i;\xi)^{-\alpha-2}}{(\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1} + \alpha G(x_i;\xi)^{\alpha-1}}\right] -$$
$$\sum_{i=1}^{n}\left[-\alpha G'(x_i;\xi)G(x_i;\xi)^{-\alpha-1} - (\beta-\alpha)G'(x_i;\xi)G(x_i;\xi)^{\beta-\alpha-1}\right], \quad (4.25)$$

where $g'(x;\xi) = \frac{\partial g(x;\xi)}{\partial \xi}$ and $G'(x;\xi) = \frac{\partial G(x;\xi)}{\partial \xi}$.

Equating the scores functions to zero and solving the system of equations numerically, gives the maximum likelihood estimates of the parameters.

4.4.2 Maximum likelihood Estimation for Censored Case

Given a dataset $U=(x_i,\delta_i)$, where x_i is the observation that corresponds to the censored failure times and δ_i is the censoring indicator. If failure is observed, $\delta_i=1$ and $\delta_i=0$ if censored. Given that $U=(x_i,\delta_i)$ is independent and identically distributed and $\boldsymbol{\theta}$ is the vector of parameters from the GOIE family of distributions, the likelihood of $\boldsymbol{\theta}$ is given by

$$L=\prod_{i=1}^{n}\left[f(x_i;\theta)\right]^{\delta_i}\left[1-F(x_i;\theta)\right]^{1-\delta_i}$$

$$=\prod_{i=1}^{n}\left[f(x_i;\theta)\right]^{\delta_i}\left[S(x_i;\boldsymbol{\theta})\right]^{1-\delta_i}.$$

where $S(x_i;\boldsymbol{\theta})=1-F(x_i;\boldsymbol{\theta})$. The log likelihood function is

$$\ell=\sum_{i=1}^{n}\log\left[\left(f(x_i;\theta)\right)^{\delta_i}\left(S(x_i;\theta)\right)^{1-\delta_i}\right]$$

$$=\sum_{i=1}^{n}\left[\delta_i\log\left(f(x_i;\theta)\right)+(1-\delta_i)\log\left(S(x_i;\theta)\right)\right]. \qquad (4.26)$$

Substituting the PDF and survival function of the GOIE family of distributions into equation (4.26) yields

$$\ell=\sum_{i=1}^{n}\delta_i\log\left[g(x_i;\xi)\left((\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1}+\alpha G(x_i;\xi)^{-\alpha-1}\right)\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]\right]+$$

$$\sum_{i=1}^{n}(1-\delta_i)\log\left[1-\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]\right]. \qquad (4.27)$$

Differentiating ℓ in equation (4.27) with respect to α, β and ξ respectively yields

$$\frac{\partial \ell}{\partial \alpha}=\sum_{i=1}^{n}\delta_i\left[\frac{G(x_i;\xi)^{-\alpha-1}-G(x_i;\xi)^{\beta-\alpha-1}-\alpha G(x_i;\xi)^{-\alpha-1}\log G(x_i;\xi)-(\beta-\alpha)G(x_i;\xi)\log G(x_i;\xi)}{(\beta-\alpha)G(x_i;\xi)^{\beta-\alpha-1}+\alpha G(x_i;\xi)^{-\alpha-1}}\right]-$$
$$\sum_{i=1}^{n}\delta_i\left[G(x_i;\xi)^{-\alpha}\log G(x_i;\xi)+G(x_i;\xi)^{\beta-\alpha}\log G(x_i;\xi)\right]\quad -$$
$$\sum_{i=1}^{n}\left[\frac{(1-\delta_i)G(x_i;\xi)^{-\alpha}\left(1-G(x_i;\xi)^{\beta}\right)\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]\log\left(G(x_i;\xi)\right)}{1-\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]}\right], \qquad (4.28)$$

$$\frac{\partial \ell}{\partial \beta}=\sum_{i=1}^{n}\delta_i\left[\frac{G(x;\xi)^{\beta-\alpha-1}-(\beta-\alpha)G(x;\xi)^{\beta-\alpha-1}\log G(x;\xi)}{\alpha G(x;\xi)^{\alpha-1}+(\beta-\alpha)G(x;\xi)^{\beta-\alpha-1}}\right]+\sum_{i=1}^{n}\delta_i\left[G(x;\xi)^{\beta-\alpha}\log G(x;\xi)\right]-$$
$$\sum_{i=1}^{n}\left[\frac{(1-\delta_i)G(x_i;\xi)^{-\alpha+\beta}\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]\log\left(G(x_i;\xi)\right)}{1-\exp\left[\frac{1-G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]}\right], \qquad (4.29)$$

$$\frac{\partial \ell}{\partial \xi}=\left[\sum_{i=1}^{n}\delta_i\left[\frac{g'(x_i;\xi)}{g(x_i;\xi)}\right]+\sum_{i=1}^{n}\left[\frac{(\beta\text{-}\alpha)(\beta\text{-}\alpha\text{-}1)G'(x_i;\xi)G(x_i;\xi)^{\beta-\alpha-2}+\alpha(-\alpha\text{-}1)G'(x_i;\xi)G(x_i;\xi)^{-\alpha-2}}{(\beta\text{-}\alpha)G(x_i;\xi)^{\beta-\alpha-1}+\alpha G(x_i;\xi)^{\alpha-1}}\right]-\right.$$
$$\left.\sum_{i=1}^{n}\delta_i\left[-\alpha G'(x_i;\xi)G(x_i;\xi)^{-\alpha-1}-(\beta-\alpha)G'(x_i;\xi)G(x_i;\xi)^{\beta-\alpha-1}\right]\right]-$$
$$\sum_{i=1}^{n}\left[\frac{(1-\delta_i)\left\{G'(x_i;\xi)G(x_i;\xi)^{-\alpha}(\beta-\alpha)\right\}G'(x_i;\xi)G(x_i;\xi)^{\beta-\alpha}\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]}{1-\exp\left[\frac{G(x_i;\xi)^{\beta}-1}{G(x_i;\xi)^{\alpha}}\right]}\right], \qquad (4.30)$$

where $g'(x;\xi)=\frac{\partial g(x;\xi)}{\partial \xi}$ and $G'(x;\xi)=\frac{\partial G(x;\xi)}{\partial \xi}$.

Solving $\frac{\partial \ell}{\partial \alpha}=0, \frac{\partial \ell}{\partial \beta}=0$ and $\frac{\partial \ell}{\partial \xi}=0$ simultaneously using numerical methods produces the Maximum likelihood estimates of parameters α, β and ξ respectively for the censored datasets.

4.4.3 Ordinary Least Squares Estimation Method

The ordinary least squares (OLS) estimation method is one of the estimation methods that estimates the parameters by minimizing the objective function. This method was introduced by Swain *et al.* (1988). If $x_{(1)}, x_{(2)}, ..., x_{(n)}$ are order statistics of a random sample of size n obtained from the GOIE distribution. The OLS estimates, $\alpha_{OLS}, \beta_{OLS}, \xi_{OLS}$ for the GOIE distribution parameters can be derived by minimizing the function

$$\varphi(\alpha,\beta,\xi)=\sum_{i=1}^{n}\left[F\left(x_{(i)}\mid\alpha,\beta,\xi\right)-\frac{i}{n+1}\right]^2, \tag{4.31}$$

with respect to α, β and ξ . But

$$F(x)=\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right], \alpha,\beta>0, x\in R.$$

Hence,

$$\varphi(\alpha,\beta,\xi)=\sum_{i=1}^{n}\left[\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right]-\frac{i}{n+1}\right]^2. \tag{4.32}$$

Differentiating equation (4.32) with respect to the parameters yields

$$\sum_{i=1}^{n}\left[\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right]-\frac{i}{n+1}\right]^{2}\Delta_{k}\left(x_{(i)};\alpha,\beta,\xi\right)=0, \qquad k=1,2,3,$$

where

$$\Delta_{1}\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\alpha}\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right],$$

$$\Delta_{2}\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\beta}\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right],$$

and

$$\Delta_{3}\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\xi}\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right].$$

4.4.4 Cramér-Von Mises Minimum Distance Estimation Method

The Cramér-Von Mises (CVM) distance estimation method, also known as maximum goodness of fit estimation method has minimum bias compared to other minimum distance estimation methods (MacDonald, 1971). It rely on the difference between the estimates of the CDF and the empirical distribution (Luceno, 2006). Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be order statistics of a random sample of size n obtained from the GOIE distribution. The estimators of the CVM for the GOIE distribution parameters can be derived by minimizing the function.

$$V(\alpha,\beta,\xi)=\frac{1}{12n}+\sum_{i=1}^{n}\left[F\left(x_{(i)}\mid\alpha,\beta,\xi\right)-\frac{2i-1}{2n}\right]^2, \tag{4.34}$$

with respect to α,β and ξ. Putting the GOIE CDF into (4.34) yields

$$V(\alpha,\beta,\xi)=\frac{1}{12n}+\sum_{i=1}^{n}\left[\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right]-\frac{2i-1}{2n}\right]^2.$$

Also, the nonlinear equations can be solved numerically to obtain the estimates of the CVM. That is

$$\sum_{i=1}^{n}\left[\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right]-\frac{2i-1}{2n}\right]^2\Delta_k\left(x_{(i)};\alpha,\beta,\xi\right)=0,\qquad k=1,2,3,\ldots$$

where

$$\Delta_1\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\alpha}\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right],$$

$$\Delta_2\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\beta}\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right],$$

and

$$\Delta_3\left(x_{(i)};\alpha,\beta,\xi\right)=\frac{\partial}{\partial\xi}\exp\left[\frac{G\left(x_{(i)};\xi\right)^{\beta}-1}{G\left(x_{(i)};\xi\right)^{\alpha}}\right].$$

4.5 Special Distributions

In this section, two special distributions of the GOIE family of distribution are discussed. These includes: generalized odd inverse exponential Weibull distribution (GOIEW) and generalized odd inverse exponential Lomax distribution (GOIEL).

4.5.1 Generalized Odd Inverse Exponential Weibull Distribution

Given the baseline CDF is that of the Weibull distribution, that is $G(x;\gamma,\theta)=1-\exp\left[-\gamma x^{\theta}\right]$, with corresponding PDF $g(x)=\theta\gamma x^{\theta-1}\exp\left[-\gamma x^{\theta}\right]$, and positive parameters $\gamma,\theta>0$. The CDF of the GOIEW distribution is given by

$$F(x)=\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right], x>0. \qquad (4.35)$$

4.5.1.1 Sub-models from GOIEW

The GOIEW have a number of sub-models that are used in modeling lifetime data. These are:

1. Odd Inverse Exponential Weibull Distribution

When $\alpha=\beta=1$, the GOIEW reduces to the odd inverse exponential Weibull distribution (OIEW) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)}\right],$$

for $\gamma,\theta>0$ and $x>0$.

2. Exponentiated Odd Inverse Exponential Weibull Distribution

When $\alpha=\beta\neq 1$, the GOIEW reduces to the exponentiated odd inverse exponential Weibull distribution (EOIEW) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right],$$

for $\alpha,\gamma,\theta>0$ and $x>0$.

3. Generalized Odd Inverse Exponential Exponential Distribution

When $\theta=1$, the GOIEW reduces to the generalized odd inverse exponential exponential distribution (GOIEE) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\exp\left(-\gamma x\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x\right)^{\alpha}\right)}\right],$$

for $\alpha,\beta,\gamma>0$ and $x>0$.

4. Generalized Odd Inverse Exponential Rayleigh Distribution

When $\theta=2$, the GOIEW reduces to the generalized odd inverse exponential Rayleigh distribution (GOIER) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\exp\left(-\gamma x^{2}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{2}\right)^{\alpha}\right)}\right],$$

for $\alpha,\beta,\gamma>0$ and $x>0$.

Table 4.1: Summary of sub-models from the GOIEW distribution

Distribution	α	β	γ	θ
OIEW	1	1	γ	θ
EOIEW	α	α	γ	θ
GOIEE	α	β	γ	1
GOIER	α	β	γ	2

Figure 4.1 displays the plot of the CDF of the GOIEW distribution for some selected parameter values. It is observed from Figure 4.1 that when the scale parameter is equal to one or approximately equal to one, the curves shows fast convergence to one. However, the curves shows a slow convergence to one when the scale parameter is less than one.

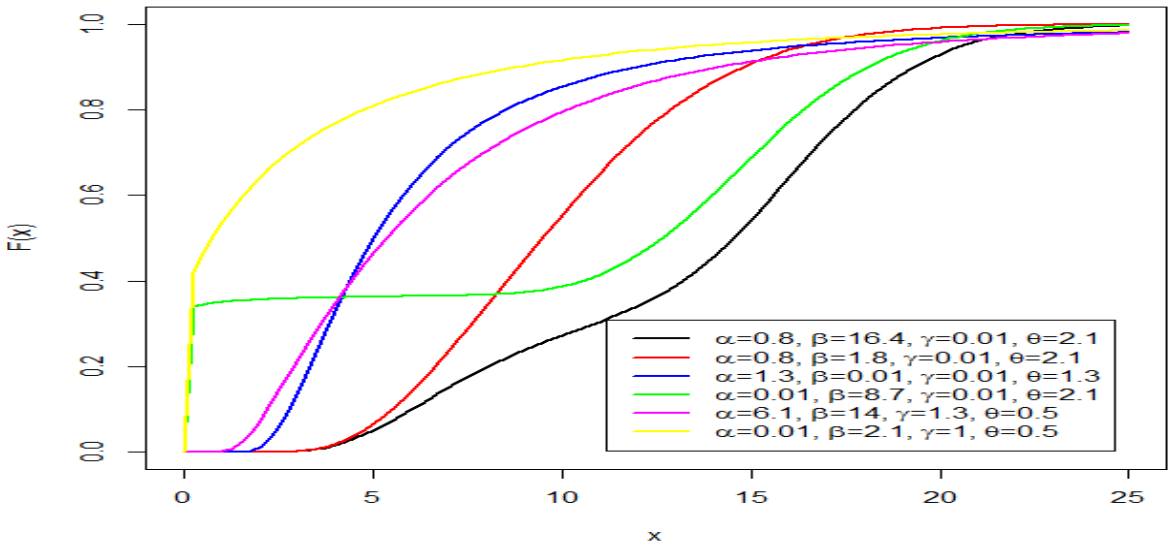

Figure 4.1: Plot of the CDF of the GOIEW distribution

The PDF of the GOIEW distribution is represented in equation (4.36).

$$f(x)=\theta\gamma x^{\theta-1}\exp\left(-\gamma x^{\theta}\right)\left[(\beta-\alpha)\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta-\alpha-1}+\alpha\left(1-\exp\left(-\gamma x^{\theta}\right)^{-\alpha-1}\right)\right]\times$$
$$\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right], x>0. \quad (4.36)$$

Figure 4.2 shows the plot of the PDF of the GOIEW distribution for some chosen parameter values. The density function exhibit bimodal shape, reverse J-shape, left skewed, right skewed and approximately symmetric shape. This means that the GOIEW family would be able to handle datasets under different shapes.

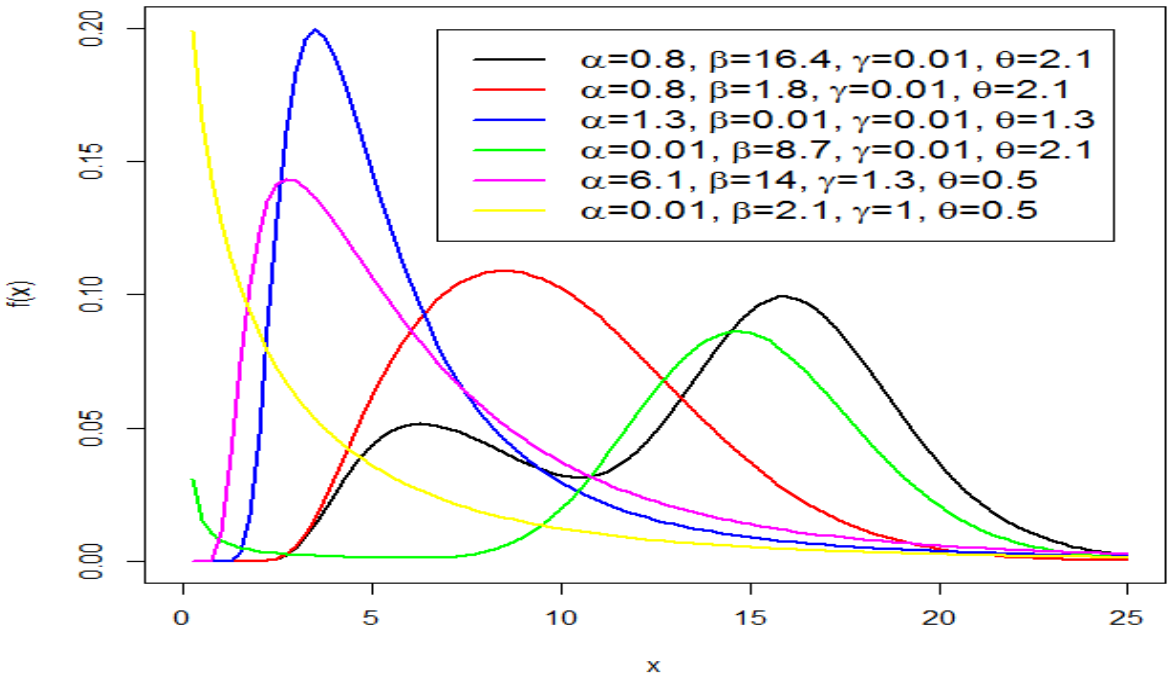

Figure 4.2: Plot of the GOIEW distribution density function

The hazard rate function of the GOIEW distribution is given by:

$$h(x)=\frac{\theta\gamma x^{\theta-1}\exp\left(-\gamma x^{\theta}\right)\left[(\beta-\alpha)\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta-\alpha-1}+\alpha\left(1-\exp\left(-\gamma x^{\theta}\right)^{-\alpha-1}\right)\right]\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right]}{1-\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right]},\ x>0. \qquad (4.37)$$

The plot of the hazard rate function of the GOIEW distribution for some selected parameter values is exhibited in Figure 4.3. It is clear that the GOIE family can produce a hazard rate shapes which include; decreasing, bathtub, reversed-J, upside-down bathtub, modified bathtub and modified upside-down bathtub.

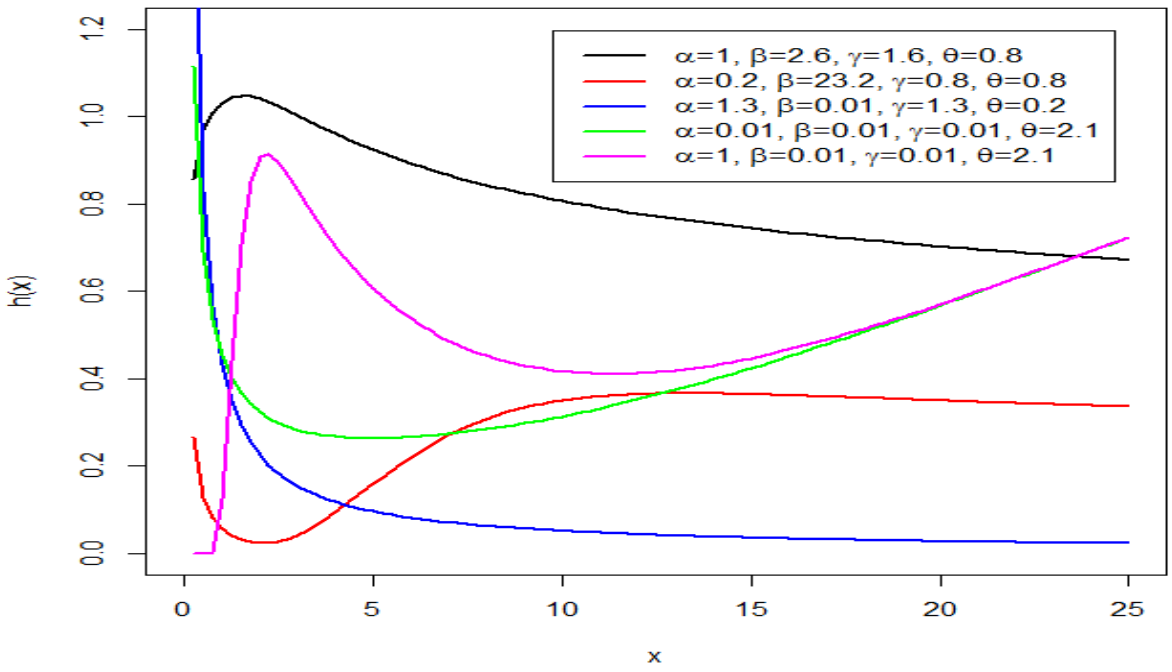

Figure 4.3: Plot of the GOIEW distribution hazard rate function

The associated survival function of the GOIEW distribution is given by:

$$S(x)=1-\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right], x>0. \qquad (4.38)$$

The plot of the survival function of the GOIEW distribution for different combination of the values of the parameters is shown in Figure 4.4. The scale parameter values of the GOIEW distributions that are equal to one or approximately equal to one indicates a fast convergence to zero than those whose scale parameter values are less than one.

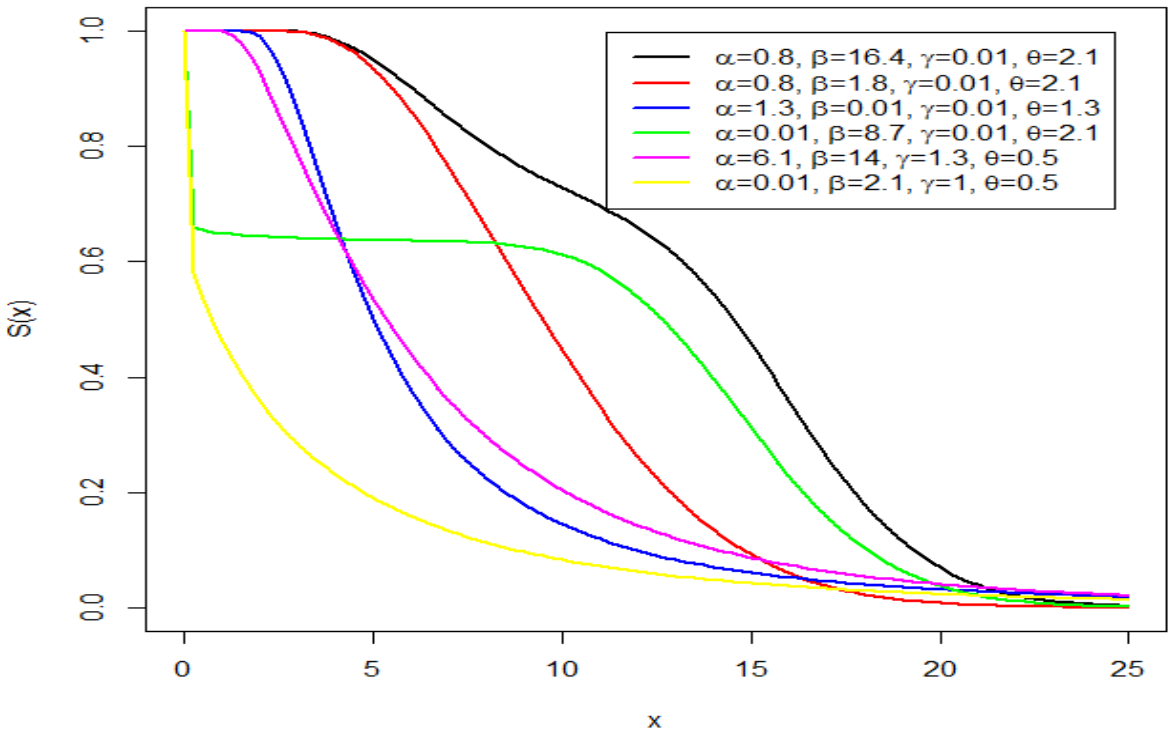

Figure 4.4: Plot of the survival function of the GOIEW distribution

The quantile function can be used to compute the median, skewness and kurtosis of a distribution. It is also useful when simulating random samples from a given distribution. The quantile function of the GOIEW distribution is

$$\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}\log u-\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}+1=0,\ u\in[0,1]. \qquad (4.39)$$

The quantile values of the GOIEW distribution for different parameter values are presented in Table 4.2. With the help of numerical technique in R, these values were generated. The different parameter values that were used for the computation are as follows; I: $\alpha=0.1$, $\beta=0.3$, $\gamma=0.7$, $\theta=0.2$; II: $\alpha=10$, $\beta=20$, $\gamma=40$, $\theta=0.5$; III: $\alpha=1.0$, $\beta=7$, $\gamma=0.8$, $\theta=0.5$ and IV: $\alpha=15$, $\beta=24$, $\gamma=33$, $\theta=12$.

Table 4.2: Quantile values of GOIEW distribution for some chosen parameter values

u	I	II	III	IV
0.1	9.7769×10^{-10}	0.0035	0.5032	0.8117
0.2	5.2442×10^{-10}	0.0044	1.3354	0.8179
0.3	5.7922×10^{-10}	0.0052	2.7503	0.8229
0.4	1.8372×10^{-7}	0.0062	4.5883	0.8277
0.5	1.9772×10^{-5}	0.0072	6.7672	0.8324
0.6	9.0276×10^{-4}	0.0085	9.4095	0.8375
0.7	2.4778×10^{-2}	0.0102	12.8391	0.8432
0.8	5.4786×10^{-1}	0.0127	17.8390	0.8503
0.9	1.4952×10^{1}	0.0172	27.1113	0.8606

The values of the first six (6) moments and other related measures such as skewness, kurtosis, standard deviation and coefficient of variation of the GOIEW distribution for some chosen parameter values are shown in Table 4.3. Numerical integration was used to derive the values of the first six moments. The coefficients of skewness and kurtosis of the GOIEW distribution from Table 4.3 are all positive. This means that the GOIEW distribution can handled datasets that are positively skewed and leptokurtic in nature. The following values of the parameters were used for the calculation. I: $\alpha = 10.5$, $\beta = 10.8$, $\gamma = 5, \theta = 1.5$; II: $\alpha = 0.5$, $\beta = 0.8$, $\gamma = 5, \theta = 1.5$; III: $\alpha = 1.5$, $\beta = 4.5$, $\gamma = 2.5, \theta = 3.5$ and IV: $\alpha = 11.5$, $\beta = 8.5$, $\gamma = 1.5, \theta = 0.5$.

Table 4.3: First six moments of GOIEW distribution

μ_r'	I	II	III	IV
μ_1'	0.7493	0.2921	0.9211	4.9782
μ_2'	0.5855	0.1236	0.8720	39.1660
μ_3'	0.4797	0.0681	0.8474	547.3385
μ_4'	0.4141	0.0456	0.8442	13499.68
μ_5'	0.3783	0.0354	0.8610	536497.2
μ_6'	0.3671	0.0311	0.8777	31406060.0000
SD	0.0240	0.0383	0.0236	14.3835
CV	0.2070	0.6698	0.1667	0.7618
CS	1.3243	1.2800	0.2127	3.8341
CK	5.2414	5.0816	0.0322	0.6733

CK=coefficient of kurtosis, CS=coefficient of skewness, CV=coefficient of variation and SD=standard deviation.

4.5.2 Generalized Odd Inverse Exponential Lomax Distribution

Consider the Lomax distribution with shape parameter $\theta > 0$ and scale parameter $\gamma > 0$, where the CDF and PDF for $x > 0$ are given by $G(x) = 1-(1+\gamma x)^{-\theta}$ and $g(x) = \theta\gamma(1+\gamma x)^{-\theta-1}$ respectively. Inserting the PDF and CDF of the Lomax distribution into equations (4.1) and (4.2), the CDF and PDF of the GOIEL distribution is given by

$$F(x) = \exp\left[\frac{\left(1-(1+\gamma x)^{-\theta}\right)^{\beta}-1}{\left(1-(1+\gamma x)^{-\theta}\right)^{\alpha}}\right], x > 0. \qquad (4.40)$$

4.5.2.1 Sub-models from GOIEL

The GOIEW have a number of sub-models that are used in modeling lifetime data. These are:

1. Odd Inverse Exponential Lomax Distribution

When $\alpha = \beta = 1$, the GOIEL reduces to the odd inverse exponential Lomax distribution (OIEL) with the following CDF:

$$G(x) = \exp\left[\frac{\left(1-(1+\gamma x)^{-\theta}\right)-1}{\left(1-(1+\gamma x)^{-\theta}\right)}\right],$$

for $\gamma, \theta > 0$ and $x > 0$.

2. Exponentiated Odd Inverse Exponential Lomax Distribution

When $\alpha = \beta \neq 1$, the GOIEL reduces to the exponentiated odd inverse exponential Lomax distribution (EOIEL) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}-1}{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}}\right],$$

for $\alpha,\gamma,\theta>0$ and $x>0$.

3. Generalized Odd Inverse Exponential Exponential Distribution

When $\theta=1$, the GOIEL reduces to the generalized odd inverse exponential exponential distribution (GOIEE) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-1}\right)^{\beta}-1}{\left(1-\left(1+\gamma x\right)^{-1}\right)^{\alpha}}\right],$$

for $\alpha,\beta,\gamma>0$ and $x>0$.

4. Generalized Odd Inverse Exponential Rayleigh Distribution

When $\theta=2$, the GOIEL reduces to the generalized odd inverse exponential Rayleigh distribution (GOIER) with the following CDF:

$$G(x)=\exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-2}\right)^{\beta}-1}{\left(1-\left(1+\gamma x\right)^{-2}\right)^{\alpha}}\right],$$

for $\alpha,\beta,\gamma>0$ and $x>0$.

Table 4.4: Summary of sub-models from GOIEL distribution

Distribution	α	β	γ	θ
OIEL	1	1	γ	θ
EOIEL	α	α	γ	θ
GOIEE	α	β	γ	1
GOIER	α	β	γ	2

Figure 4.6 shows the plot of the CDF of the GOIEL distribution for some given parameter values. The convergence of the CDF to one is dependent on the shape parameter β. When the values of β is less than one, we experience a delay convergence and a quick convergence is realized when the values of β is far greater than one.

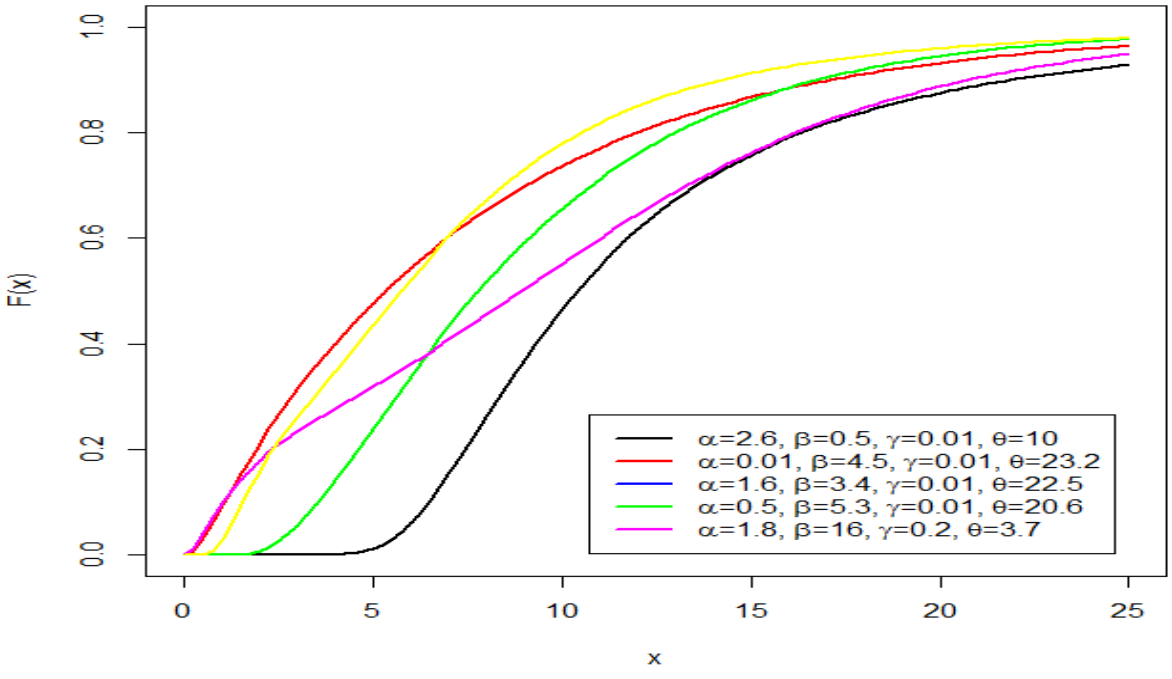

Figure 4.5: Plot of the CDF of the GOIEL distribution

The corresponding PDF of the GOIEL distribution is given in equation (4.4.41)

$$f(x)=\theta\gamma(1+\gamma x)^{-\theta-1}\left[(\beta-\alpha)\left(1-(1+\gamma x)^{-\theta}\right)^{\beta-\alpha-1}+\alpha\left(1-(1+\gamma x)^{-\theta}\right)^{-\alpha-1}\right]\exp\left[\frac{\left(1-(1+\gamma x)^{-\theta}\right)^{\beta}-1}{\left(1-(1+\gamma x)^{-\theta}\right)^{\alpha}}\right], x>0. \quad (4.41)$$

Figure 4.3 shows some of the possible shapes of the density function of the GOIEL distribution for some selected parameter values. From Figure 4.5, the density function of the GOIEL distribution can have shapes such as right skewed, decreasing and reversed-J for some selected parameter values, the PDF also exhibit bimodality.

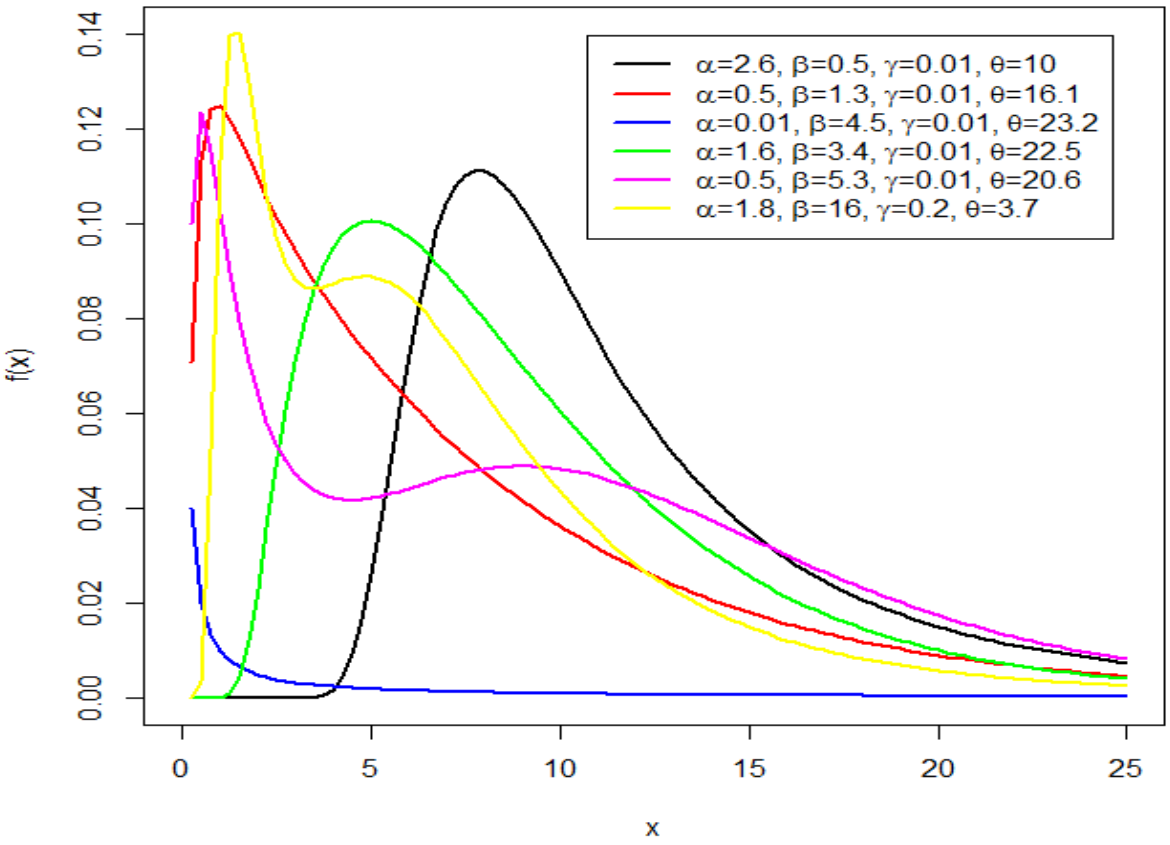

Figure 4.6: Plot of the GOIEL distribution density function

The hazard rate function of the GOIEL distribution is given by

$$h(x)=\frac{\theta\gamma(1+\gamma x)^{-\theta-1}\left[(\beta-\alpha)\left(1-(1+\gamma x)^{-\theta}\right)^{\beta-\alpha-1}+\alpha\left(1-(1+\gamma x)^{-\theta}\right)^{-\alpha-1}\right]\exp\left[\frac{\left(1-(1+\gamma x)^{-\theta}-1\right)^{\beta}}{\left(1-(1+\gamma x)^{-\theta}\right)^{\alpha}}\right]}{1-\exp\left[\frac{\left(1-(1+\gamma x)^{-\theta}\right)^{\beta}-1}{\left(1-(1+\gamma x)^{-\theta}\right)^{\alpha}}\right]},\ x>0. \quad (4.42)$$

The plot of the hazard rate function of the GOIEL distribution for different combination of parameter values are illustrated in Figure 4.7. The shapes of the hazard function can be increasing, upside-down bathtub and modified upside-down bathtub.

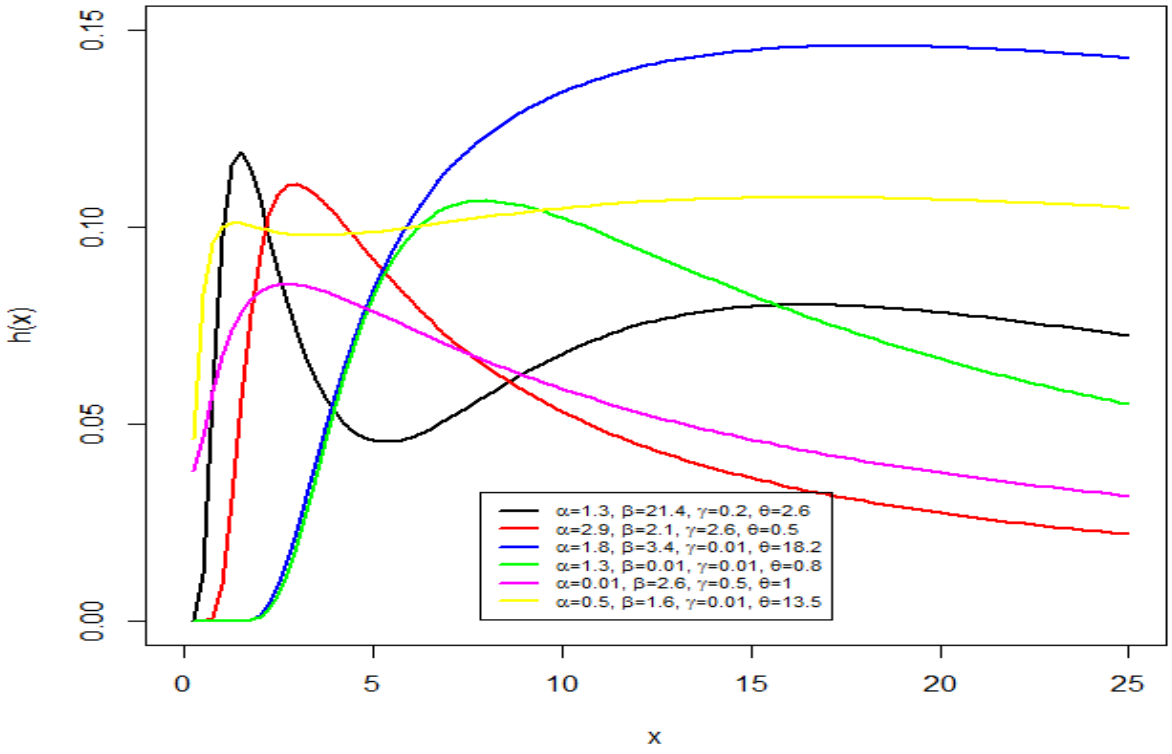

Figure 4.7: plot of the GOIEL distribution hazard rate function

The associated survival function of the GOIEL distribution is given by

$$S(x)=1-\exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta}-1}{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}}\right],x>0. \qquad (4.43)$$

The plot of the survival function of the GOIEL distribution for different combination of the parameter values is represented in Figure 4.8. From Figure 4.8, it is clear that the shape parameter β determines the rate at which the survival curve converges to zero. Values of β less than one, shows a slow convergence compared to values of β greater than one.

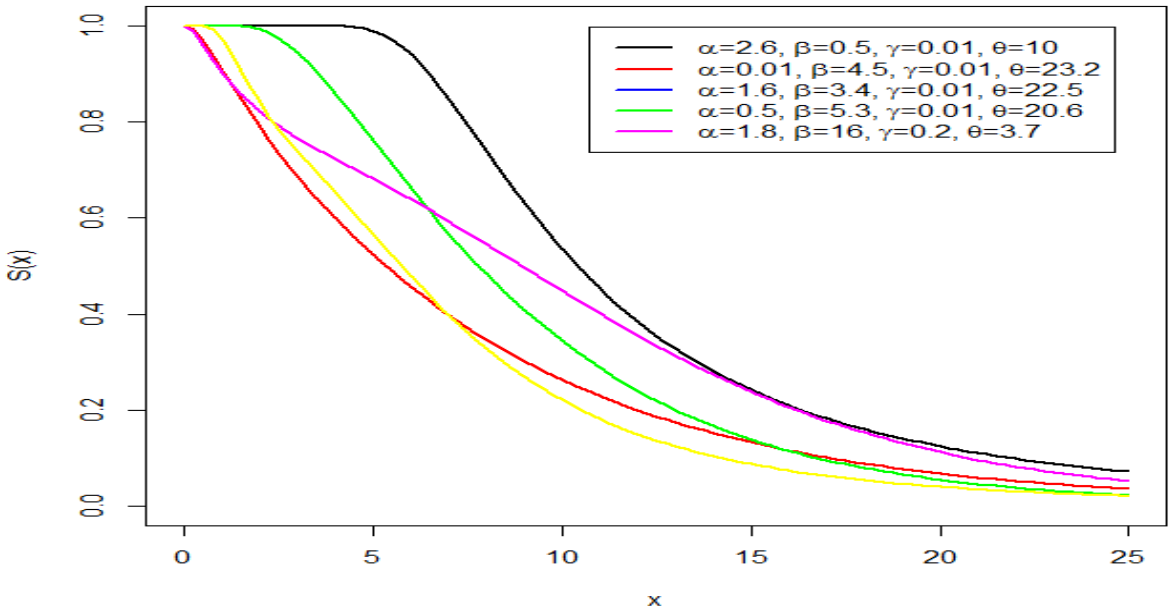

Figure 4.8: Plot of the survival function of the GOIEL distribution

The quantile function of the GOIEL distribution is

$$\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}\log u-\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta}+1=0,\ u\in[0,1]. \qquad (4.44)$$

Generation of random numbers from the GOIEL distribution can be done using equation (4.44). The first quartile, median and upper quartile of the GOIEL distribution is obtained by replacing $u = 0.25, 0.5$ and 0.75 respectively.

Table 4.5 indicates the quantile values of the GOIEL distribution for different parameter values. Using numerical approach in R, the quantile values are obtained. The different parameter values that were used for the computation are as follows; I: $\alpha = 0.8$, $\beta = 0.1$, $\gamma = 0.3, \theta = 0.9$; II: $\alpha = 1.2$, $\beta = 1.0$, $\gamma = 3.0, \theta = 0.1$; III: $\alpha = 3.0$, $\beta = 8$, $\gamma = 0.4, \theta = 0.6$ and IV: $\alpha = 14$, $\beta = 27$, $\gamma = 32, \theta = 16$.

Table 4.5: Quantile values of GOIEL distribution for some chosen parameter values

u	I	II	III	IV
0.1	0.1751	3.8001×10^{-9}	7.3930	0.8564
0.2	0.2523	4.5082×10^{-8}	10.2694	0.8617
0.3	0.3366	3.1805×10^{-7}	13.2090	0.8660
0.4	0.4386	1.8574×10^{-6}	16.4286	0.8700
0.5	0.5709	1.0341×10^{-5}	20.1159	0.8739
0.6	0.7552	6.0773×10^{-5}	24.5470	0.8781
0.7	1.0371	4.2132×10^{-4}	30.2335	0.8827
0.8	1.5365	4.1443×10^{-3}	38.3515	0.8883
0.9	2.7289	9.4023×10^{-2}	52.8417	0.8963

The values of the first six moments, skewness, kurtosis, standard deviation and coefficient of variation of the GOIEL distribution for some chosen parameter values are shown in Table 4.6. Numerical integration was used to derive the values of the first six moments. It

is observed that, the coefficient of skewness of the GOIEL distribution can either positive or negative. This means that datasets that are positively or negatively skewed can easily be modeled by this distribution. Again, datasets that are leptokurtic in nature can be modeled by the GOIEL distribution. The following values of the parameter were used for the computation. I: $\alpha = 12.5$, $\beta = 11.8$, $\gamma = 10.5$, $\theta = 11.5$; II: $\alpha = 10.5$, $\beta = 13.8$, $\gamma = 17.5$, $\theta = 12.5$; III: $\alpha = 16.5$, $\beta = 11.8$, $\gamma = 14.1$, $\theta = 13.5$ and IV: $\alpha = 14.7$, $\beta = 12.1$, $\gamma = 18.2$, $\theta = 10.5$.

Table 4.6: First six moments of GOIEL distribution

μ_r'	I	II	III	IV
μ_1'	3.348×10^{-2}	1.8415×10^{-2}	2.1432×10^{-2}	2.2074×10^{-2}
μ_2'	1.2873×10^{-3}	3.9311×10^{-4}	5.1416×10^{-4}	5.5678×10^{-4}
μ_3'	6.188×10^{-5}	9.5693×10^{-6}	1.0794×10^{-5}	1.3858×10^{-5}
μ_4'	3.535×10^{-6}	3.0870×10^{-7}	4.2988×10^{-7}	6.0064×10^{-7}
μ_5'	2.6703×10^{-7}	1.3040×10^{-8}	2.025×10^{-8}	3.2926×10^{-8}
μ_6'	2.789×10^{-8}	7.1664×10^{-10}	1.1861×10^{-9}	2.4617×10^{-9}
SD	1.6639×10^{-4}	5.3998×10^{-5}	5.48294×10^{-5}	6.95185×10^{-5}
CV	0.3853	0.3990	0.3455	0.3777
CS	3.5595	0.8605	-6.3441	-2.5904
CK	4.9267	20.1267	95.9979	60.5334

CK=coefficient of kurtosis, CS=coefficient of skewness, CV=coefficient of variation and SD=standard deviation.

4.8 The GOIEW and GOIEL Regression Models with Cure Fraction

Long-term survival models, also known as cure rate models play a key role in survival and reliability analysis. Models to handle the cure fraction have been widely developed. In the literature, two main models have been identified to fit survival data with a cure fraction. The first one in the standard cure rate model, also known as the mixture cure rate model (MCRM) and was pioneered by Boag (1949) and Berkson and Gage (1952). The assumption with this model is that, certain proportion of the population is cured whiles the remaining is not. The other type of cure rate models is the non-mixture cure rate model (NMCRM), also known as the promotion time model or bounded cumulative hazard model. This model was first developed by Yakovlev *et al.* (1993) and later studied by Chen *et al.* (1999). According to Tsodikov *et al.* (2003), the non-mixture cure rate model have the following advantages: the NMCRM has proportional hazard model structure, it presents a clearer biological meaningful interpretation of the data analysis and finally, the simple structure of the survival function provides an easy computations especially when developing the maximum likelihood estimates. In this section, we derived the GOIEW and GOIEL cure rate models and compared them with other competing models in terms of modeling a gastric cancer dataset. The GOIEW and GOIEL cure rate models are derived as follows. Given that N represent the unobservable number of causes of the event of interest for an individual in a population, and that N follows the Poisson distribution with mean λ. The time for the kth cause to give the event of interest is represented by $Z_k,\ k=1,...,N.$ It was also assume that, the Z_k are identically and independent random variables with CDF defined in equation (1), where $Z_1, Z_2 ...$ are independent on $N.$ The

observable time to the event of interest is defined as $X = \min\{Z_1,...,Z_N\}$, and $T = \infty$ when $N = 0$ with $P(X = \infty \mid N = 0) = 1$.

Based on this setup, the survival function for the population is given as

$$S_{pop} = P(N = 0) + P(Z_1 > x,...,Z_N \mid N \geq 1)P(N \geq 1).$$

This implies that, $S_{pop}(t) = A[S(t)]$ (Tsodikov *et al.*, 2003; Rodrigues *et al.*, 2009), where $A(.)$ is the probability generating function of the number of competing causes (N). Hence, the survival function for the population of the GOIEW distribution is given by

$$S_{pop}(x) = \exp\left(-\lambda \exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right]\right), x > 0 \qquad (4.45)$$

with the cured fraction as $S_{pop}(\infty) = \pi_0 = e^{-\lambda}$. The corresponding PDF for the population of the GOIEW distribution is given by

$$f_{pop}(x) = \lambda\theta\gamma x^{\theta-1}\exp(-\gamma x^{\theta})(1-\exp(-\gamma x^{\theta}))^{-1-\alpha}\left[(\beta-\alpha)\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}+\alpha\right]\times\psi, \quad x > 0, \qquad (4.46)$$

where $\psi = \exp\left(-\lambda\left(\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right]-\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right]\right)\right)$.

The associated hazard rate function for the population of the GOIEW distribution is given by

$$h_{pop}(x) = \frac{\lambda\theta\gamma x^{\theta-1}\exp(-\gamma x^{\theta})(1-\exp(-\gamma x^{\theta}))^{-1-\alpha}\left[(\beta-\alpha)\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}+\alpha\right]\times\psi}{\exp\left(-\lambda\exp\left[\frac{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\beta}-1}{\left(1-\exp\left(-\gamma x^{\theta}\right)\right)^{\alpha}}\right]\right)}, x > 0. \qquad (4.47)$$

Equation (4.45), (4.46) and (4.47) are termed as GOIEW model with cure fraction in competitive-risk structure.

Similarly, the GOIEL model with its cure fraction were also derived. The survival function for the population of the GOIEL distribution is given by

$$S_{pop}(x) = \exp\left(-\lambda \exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta} - 1}{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}} \right] \right), x > 0. \qquad (4.48)$$

The corresponding PDF for the population of the GOIEL distribution is given by

$$f_{pop}(x) = \lambda\theta\gamma\left(1+\gamma x\right)^{-\theta-1}\left[\left(\beta-\alpha\right)\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta-\alpha-1} + \alpha\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{-\alpha-1} \right] \times \vartheta, \quad x > 0, \qquad (4.49)$$

where $\vartheta = \exp\left(-\lambda \left(\left[\frac{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta} - 1}{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}} \right] - \exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta} - 1}{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}} \right] \right) \right).$

The hazard rate function for the population of the GOIEL distribution was also derived and is given b

$$h_{pop}(x) = \frac{\lambda\theta\gamma\left(1+\gamma x\right)^{-\theta-1}\left[\left(\beta-\alpha\right)\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta-\alpha-1} + \alpha\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{-\alpha-1} \right] \times \vartheta}{\exp\left(-\lambda \exp\left[\frac{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\beta} - 1}{\left(1-\left(1+\gamma x\right)^{-\theta}\right)^{\alpha}} \right] \right)}, x > 0. \qquad (4.50)$$

Equations (4.48), (4.49) and (4.50) are referred to us the GOIEL model with cure fraction.

4.8.1 Estimation of Parameters

In order to estimate the parameters of the regression models with cure fractions, the maximum likelihood estimation method with censored observations is employed. In this subsection, we consider a lifetime data that is right censored and assumed that there are n patients undergoing cancer study. Let Q_i represent the censoring time. We observe $x_i = \min\{X_i, Q_i\}$ and $\delta_i = I(X_i \leq Q_i)$, where $\delta_i = 1$ if x_i is a time-to-event and $\delta_i = 0$ given that it is right censored, for $i = 1,...,n.$ For censoring indicators $(x_1,\delta_1),...,(x_n,\delta_n)$, the total log-likelihood function with non-informative censoring can be expressed as

$$\ell = \sum_{i=1}^{n}\left[\delta_i \log\left(f_{pop}(x_i;\theta)\right) + (1-\delta_i)\log\left(S_{pop}(x_i;\theta)\right)\right], \qquad (4.51)$$

where $\boldsymbol{\theta}$ represents the vector of parameters. Therefore, the total log-likelihood of both GOIEW and GOIEL regression models with cure fraction are obtain by substituting $f_{pop}(x_i,\theta)$ and $S_{pop}(x_i,\theta)$ respectively into equation (4.51). The maximum likelihood estimates of $\boldsymbol{\theta}$ are obtained by maximizing their total log-likelihood functions directly.

CHAPTER FIVE

SIMULATIONS AND EMPIRICAL APPLICATIONS

5.0 Introduction

This chapter presents the simulations and empirical applications of the GOIE family of distributions. It was subdivided into three headings and this include: Monte Carlo simulation, applications of the GOIE family of distributions to cancer datasets and conclusion.

5.1 Monte Carlo Simulation

This section examines the properties of the estimators for the parameters of the GOIEW distribution using Monte Carlo simulation. Random samples from the GOIEW distribution where generated using the quantile function in equation (4.39). The average bias (AB) and the mean square error (MSE) of the MLE, OLS and CVM estimators for the parameters are presented in Table 5.1, 5.2, 5.3 and 5.4 for some chosen parameter values. Five thousand replications were used in the simulation experiment ($N = 5,000$) with sample sizes $n =$ 30, 50, 80, 120, 200 and 250 with parameter values $(\alpha, \beta, \gamma, \theta) = (0.8, 0.4, 1.8, 0.3)$, $(0.5, 0.9, 2.5, 0.6)$, $(0.4, 0.7, 2.6, 0.4)$, and (0.6, 0.5,2.4, 0.8).

Table 5.1 shows the simulation results of $(\alpha, \beta, \gamma, \theta) = (0.8, 0.4, 1.8, 0.3)$ for three estimators (MLE, OLS, and CVM). From Table 5.1, the ABs for the parameters were all positive and decreases towards zero as the sample size increases. Also, as the sample size increases the MSEs of the estimators of the parameters decreases indicating that the

estimators of the parameters are consistent. However, the maximum likelihood estimator recorded the least values of the ABs and the MSEs making it the best estimator.

Table 5.1: Simulation results for $(\alpha,\beta,\gamma,\theta)=(0.8,0.4,1.8,0.3)$

Parameter	M	AB			MSE		
		MLE	OLS	CVM	MLE	OLS	CVM
α	30	0.5138	0.6857	0.7186	0.2815	0.5121	0.5407
	50	0.5025	0.6262	0.676	0.2702	0.4637	0.5086
	80	0.4857	0.5989	0.5257	0.2520	0.4384	0.3736
	120	0.4579	0.4837	0.4764	0.2290	0.3378	0.3352
	200	0.4377	0.4518	0.4409	0.2122	0.3248	0.3243
	250	0.4138	0.5369	0.4871	0.1929	0.3940	0.3183
β	30	0.3551	0.4105	0.3887	0.1147	0.2050	0.2323
	50	0.3427	0.4281	0.4248	0.1139	0.2243	0.2201
	80	0.3397	0.4325	0.3857	0.1121	0.2210	0.1923
	120	0.3353	0.3929	0.3816	0.1111	0.2050	0.1885
	200	0.3286	0.3633	0.3878	0.1100	0.1842	0.1877
	250	0.3229	0.4094	0.4144	0.0073	0.2158	0.1745
γ	30	1.0206	1.2199	1.2152	1.6223	1.6827	1.6948
	50	1.0141	1.2643	1.3282	1.5524	1.6768	1.6545
	80	1.0121	1.2267	1.2455	1.4440	1.5638	1.6092
	120	1.0111	1.2303	1.2641	1.3034	1.5390	1.6021
	200	1.0109	1.2637	1.2404	1.2516	1.5159	1.5523
	250	1.0044	1.2981	1.2656	1.2273	1.4038	1.5136
θ	30	0.1852	0.3037	0.3175	0.1298	0.1506	0.1663
	50	0.1696	0.2647	0.3197	0.1288	0.1452	0.1536
	80	0.1321	0.2410	0.2827	0.1184	0.1399	0.1464
	120	0.1812	0.2651	0.2800	0.1042	0.1287	0.1347
	200	0.1423	0.2760	0.2720	0.1015	0.1135	0.1272
	250	0.1001	0.2786	0.2685	0.0058	0.0822	0.0729

The simulation results of $(\alpha,\beta,\gamma,\theta)=(0.6,0.5,2.8,0.8)$ for three estimators (MLE, OLS, and CVM) are shown in Table 5.2. The results from Table 5.2 shows that the three estimators are consistent since MSEs of the estimators of the parameters decreases as the

sample size increases. The maximum likelihood estimates remains the best estimator since it recorded the least values of ABs and MSEs compared to the other estimators.

Table 5.2: Simulation results for $(\alpha, \beta, \gamma, \theta) = (0.6, 0.5, 2.8, 0.8)$

Parameter	m	AB			MSE		
		MLE	OLS	CVM	MLE	OLS	CVM
	30	0.1069	0.5397	0.5393	0.0166	0.3032	0.3057
	50	0.0923	0.5435	0.5338	0.0121	0.3082	0.2989
	80	0.0867	0.5419	0.543	0.0102	0.3061	0.3062
	120	0.0856	0.552	0.5503	0.0100	0.3148	0.3136
	200	0.0844	0.5275	0.5206	0.0092	0.2906	0.2888
α	250	0.0839	0.5177	0.5395	0.0089	0.2802	0.3022
	30	0.2010	0.3252	0.2772	0.1027	0.1339	0.1030
	50	0.1980	0.3609	0.3634	0.1026	0.1522	0.1545
	80	0.1893	0.368	0.3506	0.1024	0.1578	0.1472
	120	0.1781	0.3774	0.4048	0.1019	0.1601	0.1818
	200	0.1678	0.3574	0.3805	0.1017	0.1480	0.1655
β	250	0.1580	0.3475	0.3633	0.1015	0.1419	0.1501
	30	1.8054	1.8163	1.8369	3.2616	3.3904	3.4793
	50	1.8047	1.9358	1.9166	3.2581	3.8153	3.7283
	80	1.8027	1.9736	1.9740	3.2507	3.9471	3.9511
	120	1.8006	2.0177	1.9842	3.2422	4.1013	3.9616
	200	1.8001	1.9990	1.9920	3.2404	4.0190	3.9906
γ	250	1.8000	1.9890	2.0085	3.2401	3.9756	4.0493
	30	0.1988	0.4787	0.4427	0.0399	0.2990	0.2713
	50	0.1984	0.5999	0.6396	0.0398	0.4173	0.4537
	80	0.1978	0.6264	0.6383	0.0396	0.4369	0.4502
	120	0.1968	0.7146	0.7385	0.0394	0.5347	0.5562
	200	0.1000	0.7202	0.7319	0.0201	0.5369	0.5512
θ	250	0.1000	0.7284	0.7549	0.0200	0.5453	0.5766

Table 5.3 displays the simulation results of $(\alpha, \beta, \gamma, \theta) = (0.5, 0.9, 2.5, 0.6)$ for MLE, OLS CVM estimators. Generally, the ABs and MSEs of all the estimators were found to be decreasing as the sample size increases indicating that they were consistent estimators. The MLE from Table 5.3 was the best estimator since it recorded the smallest values of ABs and MSEs.

Table 5.3: Simulation results for $(\alpha, \beta, \gamma, \theta) = (0.5, 0.9, 2.5, 0.6)$

Parameter	n	AB			MSE		
		MLS	OLS	CVM	MLE	OLS	CVM
α	30	0.1845	0.4439	0.4537	0.0586	0.2086	0.2158
	50	0.1777	0.4537	0.4619	0.0459	0.2179	0.2220
	80	0.1092	0.4494	0.4567	0.0432	0.2143	0.2159
	120	0.1091	0.4831	0.4820	0.0419	0.2357	0.2347
	200	0.1024	0.4858	0.4710	0.0416	0.2288	0.2280
	250	0.1005	0.4849	0.4893	0.0408	0.2382	0.2403
β	30	0.1240	0.2897	0.2483	0.0529	0.1472	0.1109
	50	0.1137	0.2365	0.1964	0.0512	0.1003	0.0717
	80	0.1082	0.2484	0.2331	0.0503	0.1044	0.0954
	120	0.1073	0.1979	0.1660	0.0501	0.0713	0.0526
	200	0.1048	0.2083	0.2053	0.0497	0.0712	0.0723
	250	0.1042	0.2273	0.1969	0.0496	0.0809	0.0582
γ	30	1.5563	2.0620	1.9790	2.4422	4.3700	4.0164
	50	1.5350	2.1246	2.0828	2.3681	4.5696	4.4036
	80	1.5243	2.1382	2.0975	2.3314	4.6266	4.4642
	120	1.5150	2.1577	2.1575	2.2988	4.6836	4.6308
	200	1.5042	2.1826	2.1629	2.2632	4.7856	4.7030
	250	1.5041	2.1963	2.2166	2.2626	4.8458	4.9300
θ	30	0.3896	0.4593	0.4317	0.1539	0.2350	0.2118
	50	0.3839	0.5132	0.4990	0.1501	0.2785	0.2664
	80	0.3820	0.5228	0.5022	0.1488	0.2853	0.2661
	120	0.3813	0.5626	0.5647	0.1415	0.3245	0.3252
	200	0.3810	0.5758	0.5640	0.1410	0.3334	0.3212
	250	0.3800	0.5759	0.5838	0.1401	0.3335	0.3418

The simulation results of $(\alpha,\beta,\gamma,\theta)=(0.4,0.7,2.6,0.4)$ for the three estimators are presented in Table 5.4. Again, the MLE emerges as the best estimator since it recorded the least values of ABs and MSEs. It was also observed from Table 5.4 that all the three estimators were consistent as their ABs and MSEs values decreases with increasing sample size.

Table 5.4: Simulation results for $(\alpha,\beta,\gamma,\theta)=(0.4,0.7,2.6,0.4)$

parameter	M	AB			MSE		
		MLE	OLS	CVM	MLE	OLS	CVM
	30	0.2173	0.3506	0.3539	0.0498	0.134	0.1372
	50	0.2166	0.376	0.3509	0.0495	0.1546	0.1327
	80	0.2119	0.3653	0.3757	0.0475	0.148	0.1509
	120	0.2084	0.3829	0.3802	0.0456	0.1611	0.1581
	200	0.2077	0.4216	0.3791	0.0455	0.1879	0.162
α	250	0.207	0.3911	0.4142	0.045	0.1687	0.1812
	30	0.059	0.3026	0.3127	0.0351	0.1056	0.112
	50	0.0587	0.3038	0.2862	0.0346	0.1042	0.0909
	80	0.058	0.3006	0.2929	0.0338	0.1013	0.0951
	120	0.0573	0.2744	0.2897	0.033	0.0839	0.0948
	200	0.0571	0.2636	0.266	0.0326	0.0797	0.0809
β	250	0.0568	0.2771	0.2833	0.0324	0.0822	0.0889
	30	1.732	2.0237	2.023	3.0577	4.1828	4.1793
	50	1.7142	2.0904	2.0898	2.9831	4.4295	4.421
	80	1.6682	2.1037	2.1334	2.8081	4.4731	4.5867
	120	1.6517	2.1507	2.1625	2.746	4.6461	4.7011
	200	1.6313	2.1556	2.1362	2.6707	4.6657	4.5809
γ	250	1.6235	2.1443	2.188	2.6427	4.6099	4.8049
	30	0.0546	0.338	0.3446	0.0314	0.1332	0.1411
	50	0.0523	0.3474	0.352	0.0293	0.1338	0.134
	80	0.0489	0.3481	0.3649	0.0263	0.1322	0.1401
	120	0.0456	0.3699	0.3767	0.023	0.1399	0.1462
	200	0.0445	0.3734	0.3715	0.0216	0.1414	0.1393
θ	250	0.0433	0.3795	0.3833	0.0204	0.1446	0.1484

5.2 Applications of the GOIE-G Family

In this section, the applications of the special distributions (GOIEW and GOIEL) of the GOIE distribution were examined using eight cancer datasets. These datasets were categorized into complete data and censored datasets. Tests such as A^*, K-S and W^*, as well as AIC, AICc and BIC were employed to compare the goodness-of-fit of the GOIEW and GOIEL distributions with other models.

5.2.1 Complete Datasets

In this subsection, the GOIEW and GOIEL distributions were applied to four complete cancer datasets. The parameters of the fitted models were obtained using the maximum likelihood estimation method. The PDFs and the CDFs of the fitted models were also plotted for visual comparisons. All this computations were done using the R-language.

5.2.1.1 Leukemia Dataset

The descriptive statistics of the leukemia dataset is explained in Table 5.5. The minimum, maximum and the average values of the dataset were observed to be 115, 1965 and 1192.3 respectively. The results from Table 5.5 also suggest that the data is negatively skewed and platykurtic in nature. Thus, less peaked compared to the normal distribution.

Table 5.5: Descriptive statistics for leukemia dataset

Minimum	Maximum	Mean	Skewness	Excess kurtosis
115.0000	1965.0000	1192.3000	-0.4600	-0.6600

The behavior of the failure rate of the dataset is explored using the TTT plot. The TTT transform curve of the leukemia data exhibit an increasing failure rate as shown in Figure 5.1.

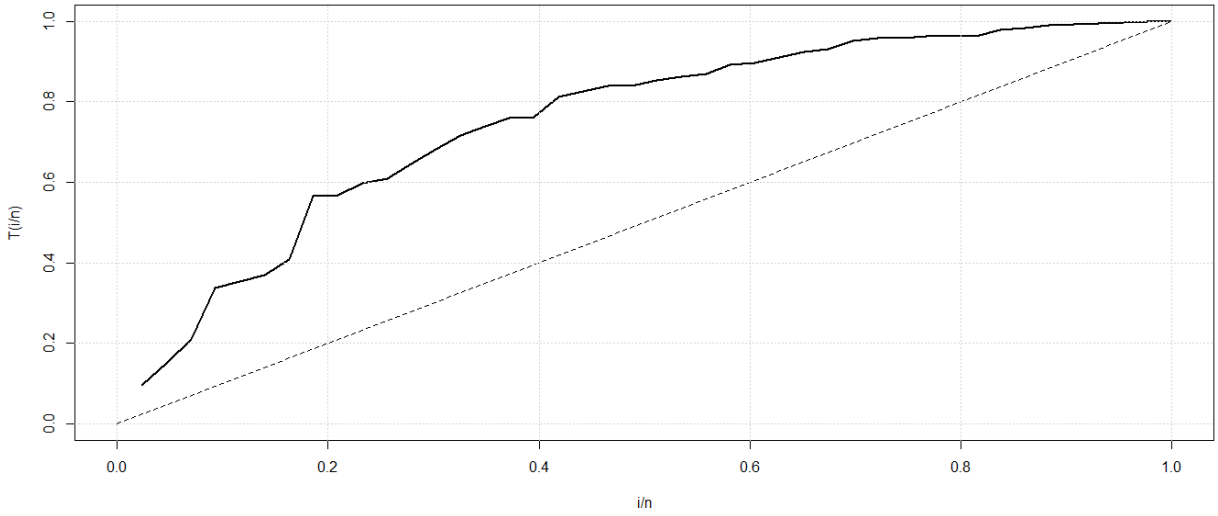

Figure 5.1: TTT transform plot for the leukemia data

Table 5.6 presents the maximum likelihood estimates of the GOIEW, GOIEL, generalized inverse Weibull (GIW) and exponentiated Lomax (E-Lx)) distributions fitted to the leukemia dataset. It was evident from Table 5.6 that all the parameters of the fitted distributions were significant at the 5% level of significance except γ in GOIEW distribution.

Table 5.6: Maximum likelihood estimates for leukemia dataset

Distribution	Parameter	Standard Error	Z-Value	*P-Value*
GOIEW	$\hat{\alpha}$ = 1.7676	0.5999	2.9467	0.0032*
	$\hat{\beta}$ = 42.2469	0.0143	2957.5338	2.2000×10^{-16}***
	$\hat{\gamma}$ = 0.0079	0.0078	1.0162	0.3095
	$\hat{\theta}$= 0.8782	0.1352	6.4966	8.2140×10^{-11} ***
GIW	$\hat{\lambda}$ = 20.4052	6.3882	3.1942	0.0014*
	$\hat{\theta}$=1.1893	0.1184	10.0464	2.2000×10^{-16}***
	$\hat{b}$ =69.0472	1.5874	43.4970	2.2000×10^{-16}***
GOIEL	$\hat{\alpha}$ = 6.7875	1.3537×10^{-1}	5.0139×10^{1}	2.2000×10^{-16}***
	$\hat{\beta}$ = 5.5062×10^{2}	6.6430×10^{-4}	8.2888×10^{5}	2.2000×10^{-16}***
	$\hat{\gamma}$ = 5.2442×10^{-3}	1.6279×10^{-3}	3.2213	0.0013
	$\hat{\theta}$= 3.3683	4.6765×10^{-1}	7.2026	5.9070×10^{-13}***
E-Lx	$\hat{\alpha}$ = 2.0759×10^{2}	5.0510×10^{-5}	4.1099×10^{6}	2.2000×10^{-16}***
	$\hat{\lambda}$ = 1.0156×10^{-1}	4.7377×10^{-2}	2.1437	0.0321
	$\hat{\theta}$=1.2399	1.3932×10^{-1}	8.8998	2.2000×10^{-16}***

*means significant at 5% level of significance

The goodness-of-fit statistics, log-likelihood and the information criteria for the fitted models were also examined as shown in Table 5.7. Results from Table 5.7 shows that GOIEW distribution recorded the highest value of the log-likelihood and the least values

of A^*, W^*, K-S, AIC, AICc and BIC suggesting that GOIEW distribution gives a better fit to the leukemia dataset than the other fitted models.

Table 5.7: Goodness-of-fit statistics and information criteria for leukemia dataset

Model	$-\ell$	AIC	AICc	*BIC*	A^*	W^*	*K-S*	*P-Value*
GOIEW	**-330.480**	**668.9273**	**669.9800**	**675.9721**	**0.4933**	**0.0576**	**0.1703**	**0.1651**
GIW	-352.3200	710.6453	711.2607	715.9289	4.2362	0.7650	0.2607	0.0058
GOIEL	-338.7500	685.5067	686.5593	692.5515	0.8518	0.1092	0.2477	0.0102
ELx	-351.9400	709.8701	710.4855	715.1537	4.2210	0.7621	0.2520	0.0085

*Bolded means best based on selection criteria

The plots of the histogram with the densities of the fitted models and the empirical CDF with the CDFs of the fitted models of the leukemia data are respectively represented in Figure 5.2. It can be seen that the fitted distributions mimic the empirical density and CDF of the leukemia data.

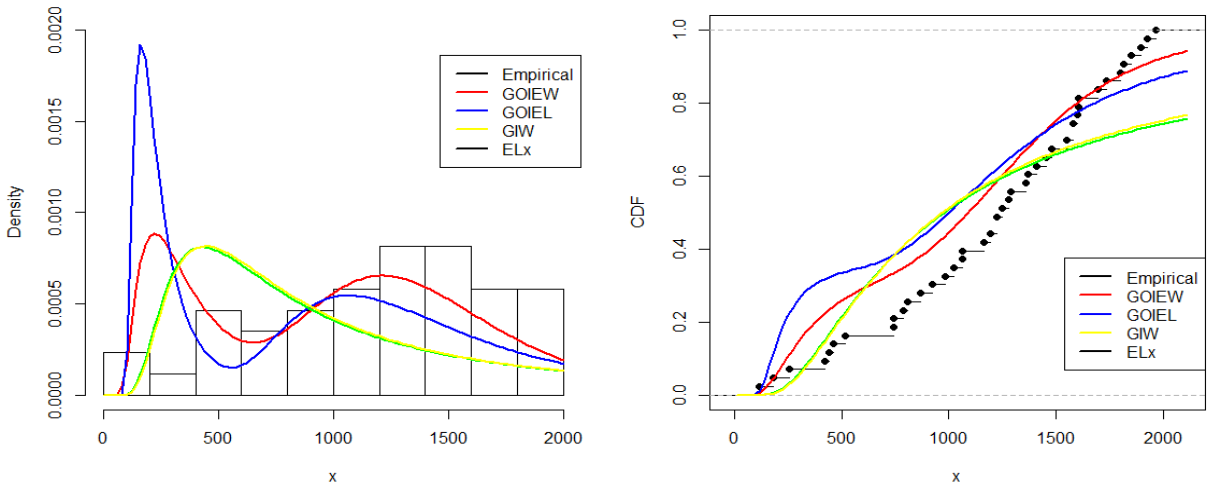

Figure 5.2: Plots of fitted densities and CDFs of leukemia data

Again, the variance-covariance matrix for the parameter estimates of the GOIEW distribution for the leukemia dataset is given by

$$J^{-1} = \begin{bmatrix} 0.3592 & -8.5666 & 3.8425\times10^{-3} & -0.0667 \\ -0.0086 & 2.0405 & -9.2834\times10^{-5} & 0.0016 \\ 0.0038 & -9.2834 & 6.0335\times10^{-5} & -0.0010 \\ -0.0667 & 1.6107 & -1.0484\times10^{-3} & 0.0183 \end{bmatrix}.$$

The 95% confidence intervals for the parameters α, β, γ and θ of the GOIEW distribution were also estimated and these are $[0,\ 1.4839]$, $[1.1847,\ 1.2407]$, $[0,\ 0.0218]$ and $[1.0374,\ 1.5674]$ respectively.

5.2.1.2 Bladder Cancer Data

Table 5.8 represents the descriptive statistics for the bladder cancer remission time's data. The least and the highest remission times of the patients were observed to be 0.08 and 79.05 months respectively. The average time of the patients was 0.929 months. The values of coefficient of skewness of 3.33 and excess kurtosis of 16.15 indicates that the remission times was right skewed and more peak than the normal curve.

Table 5.8: Descriptive statistics for remission times (in months) of Bladder cancer patients

Minimum	Maximum	Mean	Skewness	Excess kurtosis
0.08	79.05	0.929	3.33	16.15

The TTT transform plot for the bladder cancer remission time's data is shown in Figure 5.3. It is observed from Figure 5.3 that the bladder cancer remission time's data exhibit an

upside down bathtub failure rate since the plot of the data first shows a concave shape above the 45^0 line and a subsequent convex curve below the 45^0 line.

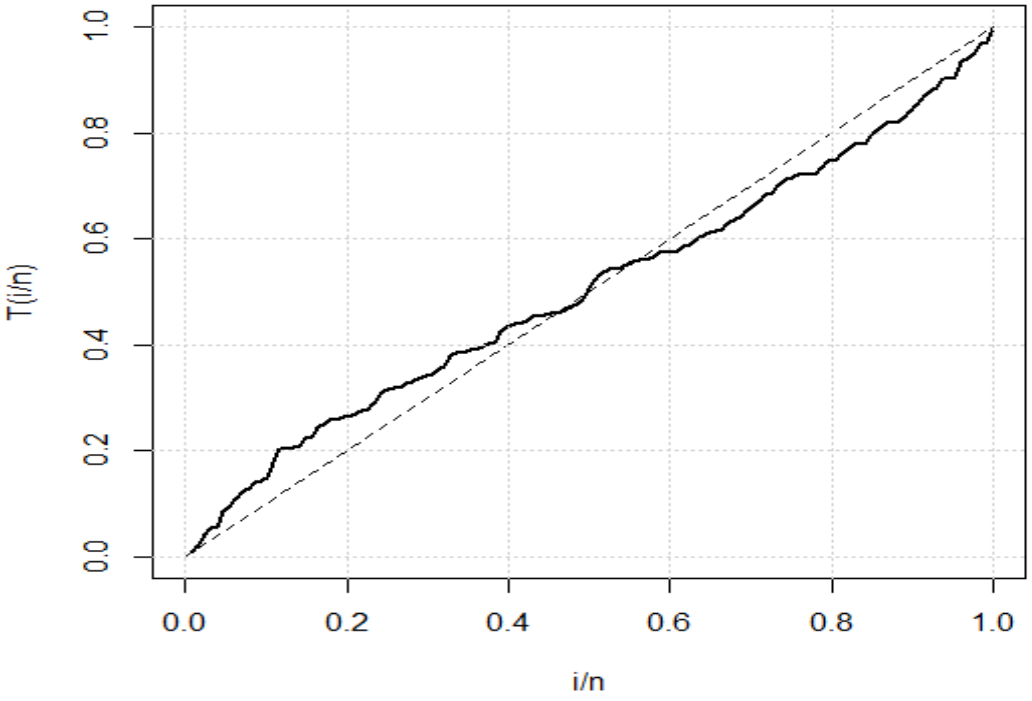

Figure 5.3: TTT transform plot for the bladder cancer remission time's data

The remission times of the bladder cancer patients were modeled using GOIEW, GOIEL, GIW, OIEW, GOIEE, EOIEW, GOIER and IW distributions. The maximum likelihood estimates for the fitted models and its standard errors are show in Table 5.9. The parameters β and γ for the GOIEW and GOIEE distributions as well θ and γ in GOIEL distribution were not significant at the 5% level of significance. However, the parameter values for the GIW, EOIEL, GOIER, OIEL and IW distributions were all significant at the 5% level of significance.

Table 5.9: Maximum likelihood estimates for bladder cancer remission time's data

Distribution	Parameter	Standard Error	Z-Value	*P-Value*
GOIEW	$\hat{\alpha}$ = 0.5624	0.2289	2.4568	0.0140*
	$\hat{\beta}$ = 1.4505	1.1441	1.2677	0.2049
	$\hat{\gamma}$ = 0.1927	0.1857	1.0374	0.2996
	$\hat{\theta}$ = 0.8556	0.2191	3.9059	9.3880×10^{-5} ***
GIW	$\hat{\lambda}$ = 1.2266×10^{-2}	3.6389×10^{-3}	3.3788	7.4950×10^{-4}***
	$\hat{\theta}$=7.5375×10^{-1}	4.2548×10^{-2}	1.7715×10^{1}	2.2000×10^{-16}***
	$\hat{b}$ =6.7195×10^{1}	1.8562×10^{-6}	3.6200×10^{7}	2.2000×10^{-16}***
GOIEL	$\hat{\alpha}$ = 0.4963	0.0630	7.8792	3.2940×10^{-15}***
	$\hat{\beta}$ = 1.6500	0.7664	2.1531	0.0313*
	$\hat{\gamma}$ = 0.0291	0.0277	1.0474	0.2949
	$\hat{\theta}$ = 5.9006	3.7475	1.5746	0.1154
IW	$\hat{\alpha}$ = 2.4310	0.2193	11.0870	2.2000×10^{-16}***
	$\hat{\beta}$ = 0.7521	0.0424	17.7270	2.2000×10^{-16}***
OIEL	$\hat{\gamma}$ = 0.5412	0.1597	3.3891	7.0120×10^{-4}***
	$\hat{\theta}$=0.8108	0.1580	5.1330	2.852×10^{-7}***
GOIER	$\hat{\alpha}$ = 0.5919	0.0800	7.3989	1.3740×10^{-13}***
	$\hat{\beta}$ =2.4962	1.1523	2.1663	0.0303*
	$\hat{\gamma}$ =0.1551	0.0561	2.7663	0.0057**
GOIEE	$\hat{\alpha}$ = 0.5232	0.0651	8.0356	9.3130×10^{-16}***
	$\hat{\beta}$ =0.7194	0.4392	1.6382	0.1014
	$\hat{\gamma}$ =0.1328	0.0830	1.6005	0.1095
EOIEL	$\hat{\alpha}$ = 0.4860	3.6652×10^{-2}	13.2600	2.2000×10^{-16}***
	$\hat{\gamma}$ =0.0082	1.2231×10^{-3}	6.6956	2.1470×10^{-11}***
	$\hat{\theta}$ =9.3761	5.1220×10^{-5}	1.8306×10^{-5}	2.2000×10^{-16}***

*means significant at 5% level of significance

The GOIEL gives a better fit to the bladder cancer remission time's data than the other competing models. From Table 5.10, GOIEL had the highest log-likelihood and the least values of A^*, W^*, K-S, AIC, AICc and BIC when compared to the other models.

Table 5.10: Goodness-of-fit statistics and information criteria for bladder cancer remission time data

Model	$-\ell$	AIC	AICc	*BIC*	A^*	W^*	*K-S*	*P-Value*
GOIEW	-414.5700	837.1456	837.4708	848.5535	0.7809	0.1223	0.0842	0.3237
GIW	-444.000	894.0033	894.1968	902.5595	4.5607	0.7468	0.1402	0.0130
GOIEL	**-413.6500**	**835.3095**	**835.6347**	**846.7176**	**0.5897**	**0.0904**	**0.0904**	**0.2838**
IW	-444.000	892.0015	892.0975	897.7056	4.5463	0.7443	0.1408	0.0125
EOIEL	-418.0600	842.1264	842.3200	850.6825	1.1416	0.1749	0.1012	0.1454
OIEL	-459.8300	923.6685	923.7645	929.3726	7.0392	1.1938	0.2226	6.1860×10^{-5}
GOIER	-418.1800	842.3567	842.5503	850.9128	0.6302	0.0853	0.1186	0.0548
GOIEE	-430.9100	867.8152	868.0087	876.3713	2.1345	0.3314	0.1253	0.0359

*bolded means best based on selection criteria

Likelihood ratio test (LRT) was carried out to compare the GOIEL distribution with its sub-models in Table 5.11. The LRT results indicates that the GOIEL distribution gives a better fit to the bladder cancer data than its sub-models at the 5% level of significance.

Table 5.11: Likelihood ratio test statistic for bladder cancer data

Model	Hypothesis	LRT	P-Value
OIEL	$H_0:\alpha=\beta=1\ Vs\ H_1:\ H_0$ is false	8.8170	0.0030**
GOIEE	$H_0:\theta=1\ Vs\ H_1:\ H_0$ is false	92.3590	2.2000×10^{-16}***
EOIEL	$H_0:\alpha=\beta\ Vs\ H_1:\ H_0$ is false	9.0473	0.0026**
GOIER	$H_0:\theta=2\ Vs\ H_1:\ H_0$ is false	34.506	4.2500×10^{-9}***

The plots of the histogram with the densities of the fitted models and the empirical CDF with the CDFs of the fitted models of the bladder cancer remission time's data are respectively represented in Figure 5.4. It is clear that the fitted distributions have taken the shape of the empirical density and CDF of the bladder remission times data.

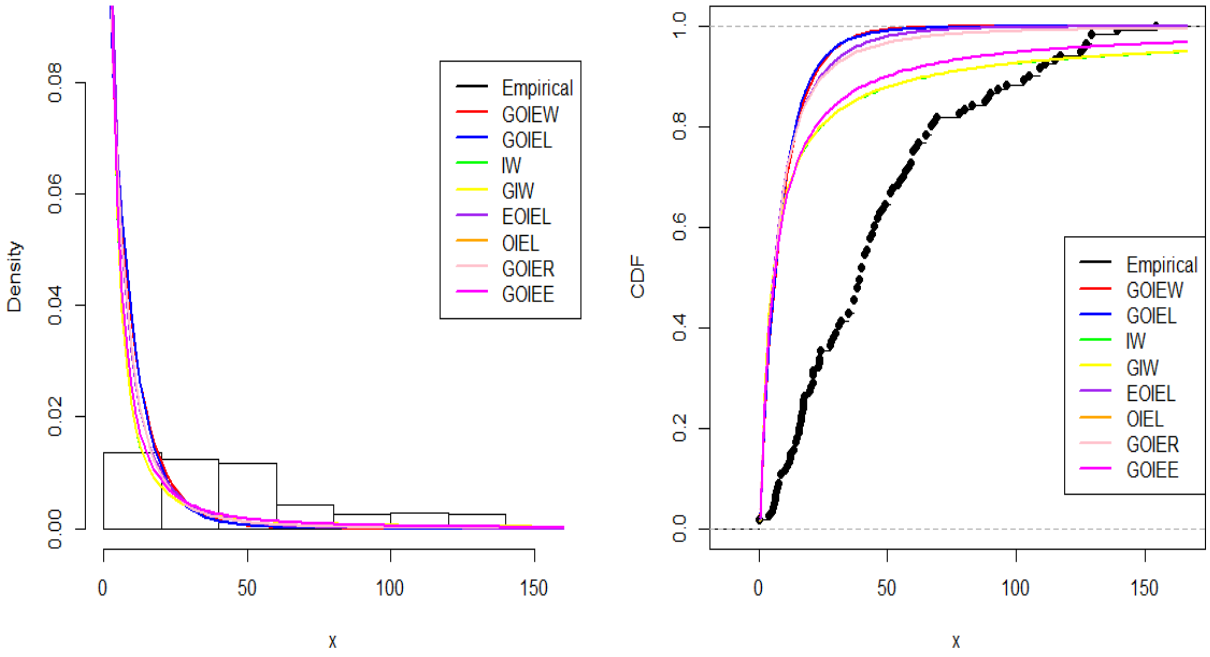

Figure 5.4: Plots of fitted densities and CDFs of bladder cancer remission time's data

Figure 5.5 displays the P-P plots of the fitted models. They were plotted to assess how well the models fit the given dataset. From the Figure 5.5, it was observed that the GOIEL gives a better fit than the other competing models as the plots in the GOIEW and GOIEL cluster along the diagonal than the other models.

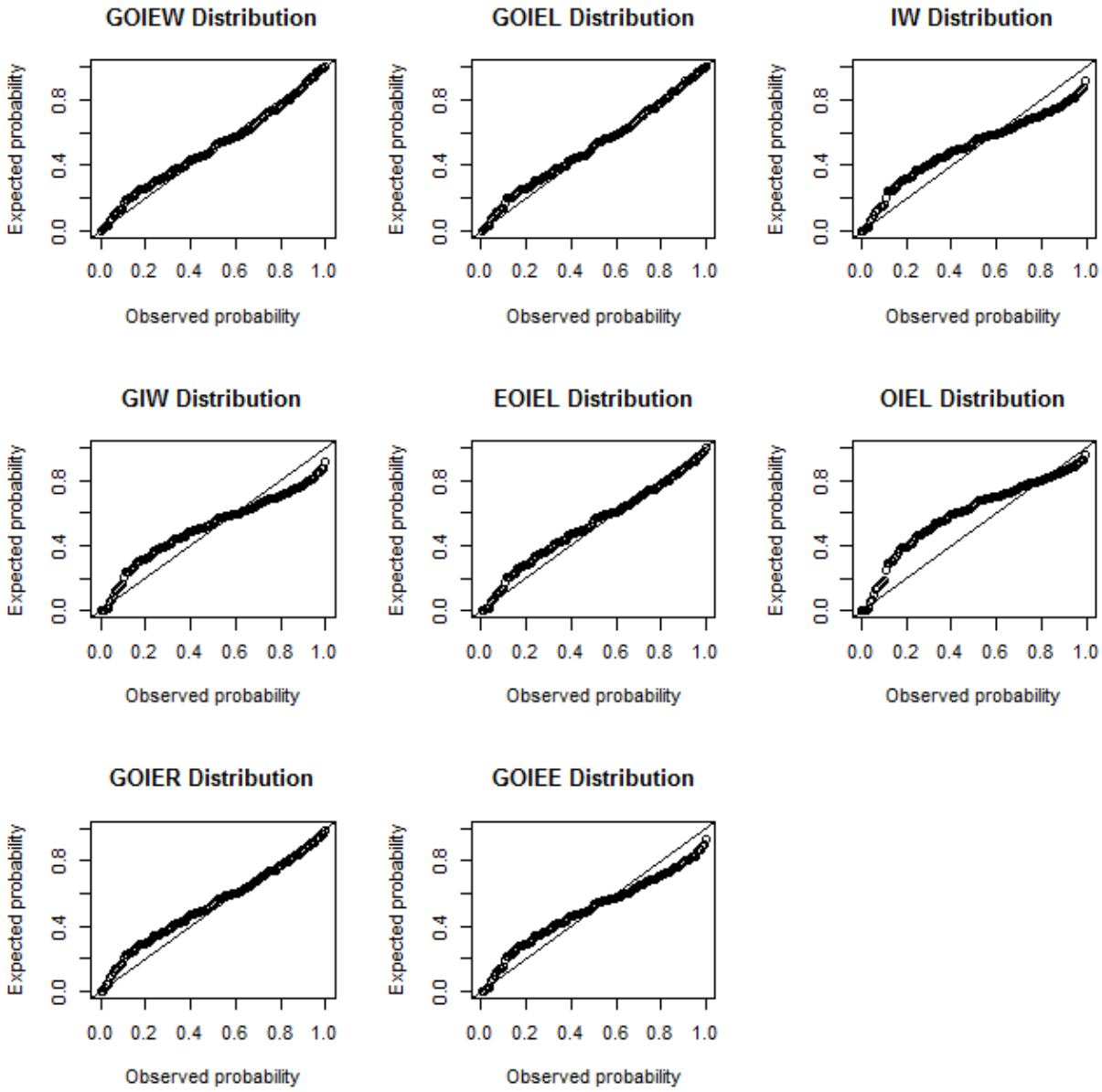

Figure 5.5: P-P plots of fitted distributions for the bladder cancer remission time's data

In addition, the parameter estimates of the GOIEL distribution for the bladder cancer data were used to obtain the following variance-covariance matrix

$$J^{-1} = \begin{bmatrix} 0.0041 & 0.0318 & 0.0012 & -0.1460 \\ 0.0318 & 0.6048 & 0.0178 & -1.9983 \\ 0.0012 & 0.0178 & 0.0008 & -0.1072 \\ -0.1460 & -1.9983 & -0.1072 & 14.8829 \end{bmatrix}.$$

The estimated 95% confidence intervals for the parameters α, β, γ and θ of the GOIEL distribution are respectively $[0.3728,\ 0.6198]$, $[0.1479,\ 3.1521]$, $[0,\ 0.0834]$ and $[0,\ 13.2457]$.

5.2.1.3 Breast Cancer Data

Table 5.12 summarizes the descriptive statistics of the breast cancer data. The maximum, minimum and the average values of the data are respectively 154, 0.3 and 43.33. Table 4.9 also indicates that the breast cancer dataset is positively skewed and leptokurtic in nature.

Table 5.12: Descriptive statistics for breast cancer dataset

Minimum	Maximum	Mean	Skewness	Excess kurtosis
0.3000	154.0000	43.3300	1.0600	0.4700

The TTT transform plot was used to analyze the failure rate of the breast cancer data as indicated in Figure 5.6. It was clear that the breast cancer data exhibit an increasing failure rate since the plot of the data shows a concave shape above the 45^0 line.

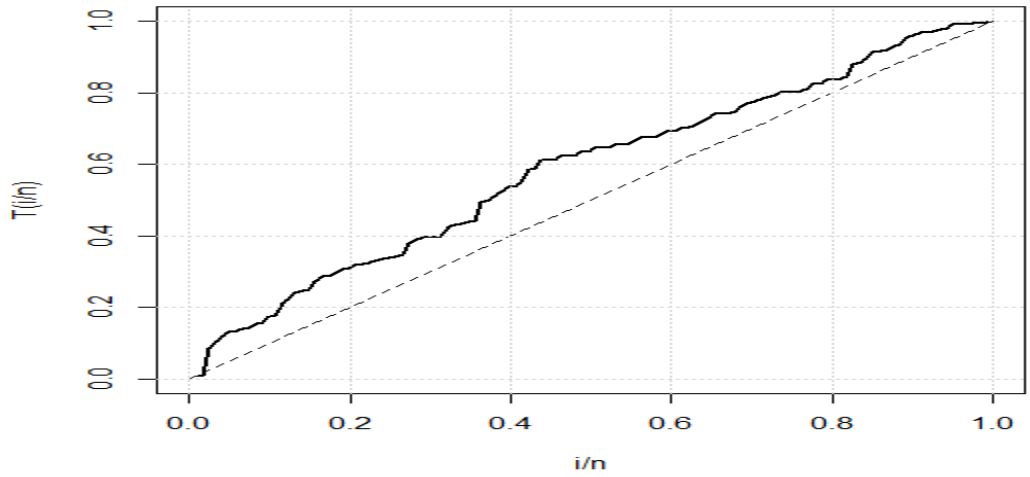

Figure 5.6: TTT transform plot for the breast cancer data

The maximum likelihood estimates of the GOIEW, GOIEL, OIEW, GOIEE, EOIEW, odd generalized exponential Weibull (OGEW), and Kumaraswamy inverse exponential (KIE) distributions fitted to the breast cancer data are shown in Table 5.13. It is clear from Table 5.13 that all the parameters of the distributions were significant at the 5% level of significance except β and γ in GOIEW distribution, α and λ in EOFW distribution as well as γ in EOIEW.

Table 5.13: Maximum likelihood estimates for breast cancer dataset

Distribution	Estimates	Std. Error	Z-Value	P-Value
GOIEW	$\hat{\alpha}$ = 0.3081	0.1124	2.7421	0.0061*
	$\hat{\beta}$ = 1.2127	0.8184	1.4817	0.1384
	$\hat{\gamma}$ = 0.0065	0.0122	0.5332	0.5939
	$\hat{\theta}$ = 1.3024	0.3565	3.6533	0.0003*
*GOIEL	$\hat{\alpha}$ = 0.4650	0.0509	9.1407	$<2.0000\times10^{-16}$***
	$\hat{\beta}$ = 2.6799	0.9127	2.9364	0.0033*
	$\hat{\gamma}$ = 0.0060	0.00219	2.0735	0.0381*
	$\hat{\theta}$ = 6.8388	2.3970	2.8531	0.0043*
OGEW	$\hat{\alpha}$ = 2.1570×10^{1}	9.1458×10^{-5}	2.3585×10^{5}	$<2.2000\times10^{-16}$***
	$\hat{\lambda}$ = 2.9678×10^{2}	4.3273×10^{-7}	6.8584×10^{8}	$<2.2000\times10^{-16}$***
	$\hat{\beta}$ = 2.928×10^{-1}	2.0080×10^{-2}	1.4582×10^{1}	$<2.2000\times10^{-16}$***
	$\hat{\theta}$=4.2040×10^{-3}	3.1483×10^{-4}	1.3353×10^{1}	$<2.2000\times10^{-16}$***
KIE	$\hat{a}$ = 4.0809×10^{1}	9.4117×10^{-5}	4.3360×10^{5}	$<2.2000\times10^{-16}$***
	$\hat{b}$ = 5.5951×10^{-1}	6.6034×10^{-2}	8.4731	$<2.2000\times10^{-16}$***
	$\hat{\theta}$ = 1.5688×10^{-1}	2.4480×10^{-2}	6.4087	$<1.4670\times10^{-16}$***
EOFW	$\hat{\alpha}$ = 0.7029	0.3778	1.8605	0.0628
	$\hat{\beta}$ = 1.5407	0.3651	4.2197	2.4460×10^{-5}***
	$\hat{\lambda}$ = 0.0043	0.0085	0.5126	0.6082
	$\hat{\theta}$ = 0.4149	0.1143	3.6310	0.0003*
OIEW	$\hat{\gamma}$ = 0.1344	0.01452	9.2572	2.2000×10^{-16}***
	$\hat{\theta}$=0.5734	0.03015	19.0197	2.2000×10^{-16}***
GOIEE	$\hat{\alpha}$ = 0.6952	3.6668×10^{-2}	1.8959×10^{1}	2.2000×10^{-16}***
	$\hat{\beta}$ =711.0300	1.6228×10^{-6}	4.3546×10^{8}	2.2000×10^{-16}***
	$\hat{\gamma}$=0.0612	2.1919×10^{-3}	2.7920×10^{1}	2.2000×10^{-16}***
EOIEW	$\hat{\alpha}$ = 0.3217	0.1383	2.3257	0.0200*
	$\hat{\gamma}$=0.0021	0.0055	0.3882	0.6979
	$\hat{\theta}$=1.3725	0.5139	2.6707	0.0076**

*means significant at 5% level of significance

The goodness-of-fit statistics, log-likelihood and information criteria for the breast cancer dataset are shown in Table 5.14. It was evident from Table 5.14 that GOIEW distribution provided a better fit to the breast cancer data than the other distributions as it recorded the highest log-likelihood value and minimum values of A^*, W^*, K-S, AIC, AICc and BIC.

Table 5.14: Goodness-of-fit statistics and information criteria for breast cancer dataset

Model	$-\ell$	AIC	AICc	*BIC*	A^*	W^*	*K-S*	*P-Value*
GOIEW	**-583.4200**	**1174.8400**	**1175.1850**	**1186.0230**	**0.7748**	**0.0958**	**0.0779**	**0.4555**
GOIEL	-588.0100	1184.0160	1184.3610	1195.1990	0.6971	0.0695	0.1111	0.1008
OGEW	-597.2400	1202.4790	1202.8240	1213.6620	2.2858	0.3608	0.1248	0.0462
KIE	-664.1200	1334.2480	1334.453	1342.6350	12.1133	2.0453	0.2793	1.2620×10^{-8}
EOFW	-587.1200	1182.2490	1182.2490	1193.4320	1.0427	0.1430	0.1231	0.0512
OIEW	-610.9700	1225.9440	1226.0460	1231.5360	3.8873	0.5916	0.1620	0.0035
GOIEE	-714.6800	1435.3600	1435.5650	1443.7480	13.3848	2.6303	0.4799	2.2000×10^{-16}
EOIEW	-593.8000	1193.5900	1193.7950	1201.9780	1.6498	0.2388	0.1522	0.0073

*bolded means best based on selection criteria

Table 5.15 shows the comparison of GOIEW distribution with its sub-models using LRT. The LRT statistics and their corresponding *P*-values indicates that GOIEW distribution provided a better fit than its sub-models.

Table 5.15: Likelihood ratio test statistic for breast cancer

Model	Hypothesis	LRT	P-Value
OIEW	$H_0 : \alpha = \beta = 1 \; Vs \; H_1 : \; H_0$ is false	55.1040	1.0820×10^{-12}***
GOIEE	$H_0 : \theta = 1 \; Vs \; H_1 : \; H_0$ is false	262.5200	2.2000×10^{-16}***
EOIEW	$H_0 : \alpha = \beta \; Vs \; H_1 : \; H_0$ is false	20.7500	5.2330×10^{-6}***

The plots of the histogram with the densities of the fitted models and the empirical CDF with the CDFs of the fitted models of the breast cancer data are respectively presented in Figure 5.7. It is clear that the fitted distributions mimic the empirical density and CDF of the breast cancer dataset.

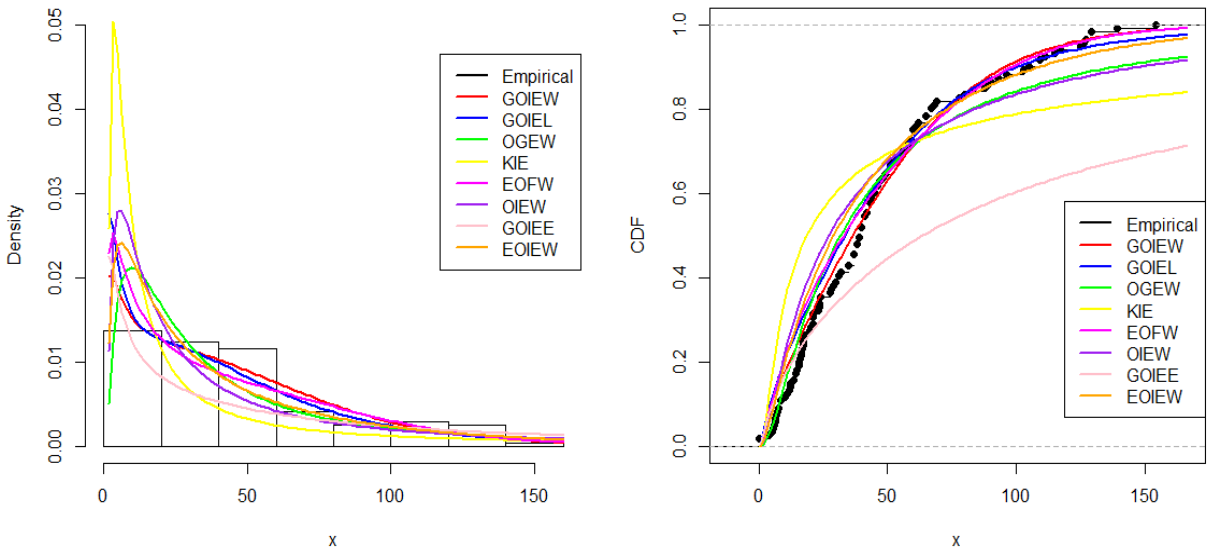

Figure 5.7: Plots of fitted densities and CDFs of breast cancer data

Also, the P-P plots in Figure 5.8 indicates that the GOIEW distribution gives better fit to the data than GOIEL, KIE, EOFW, OIEW, GOIEE, EOIEW and OGEW distributions.

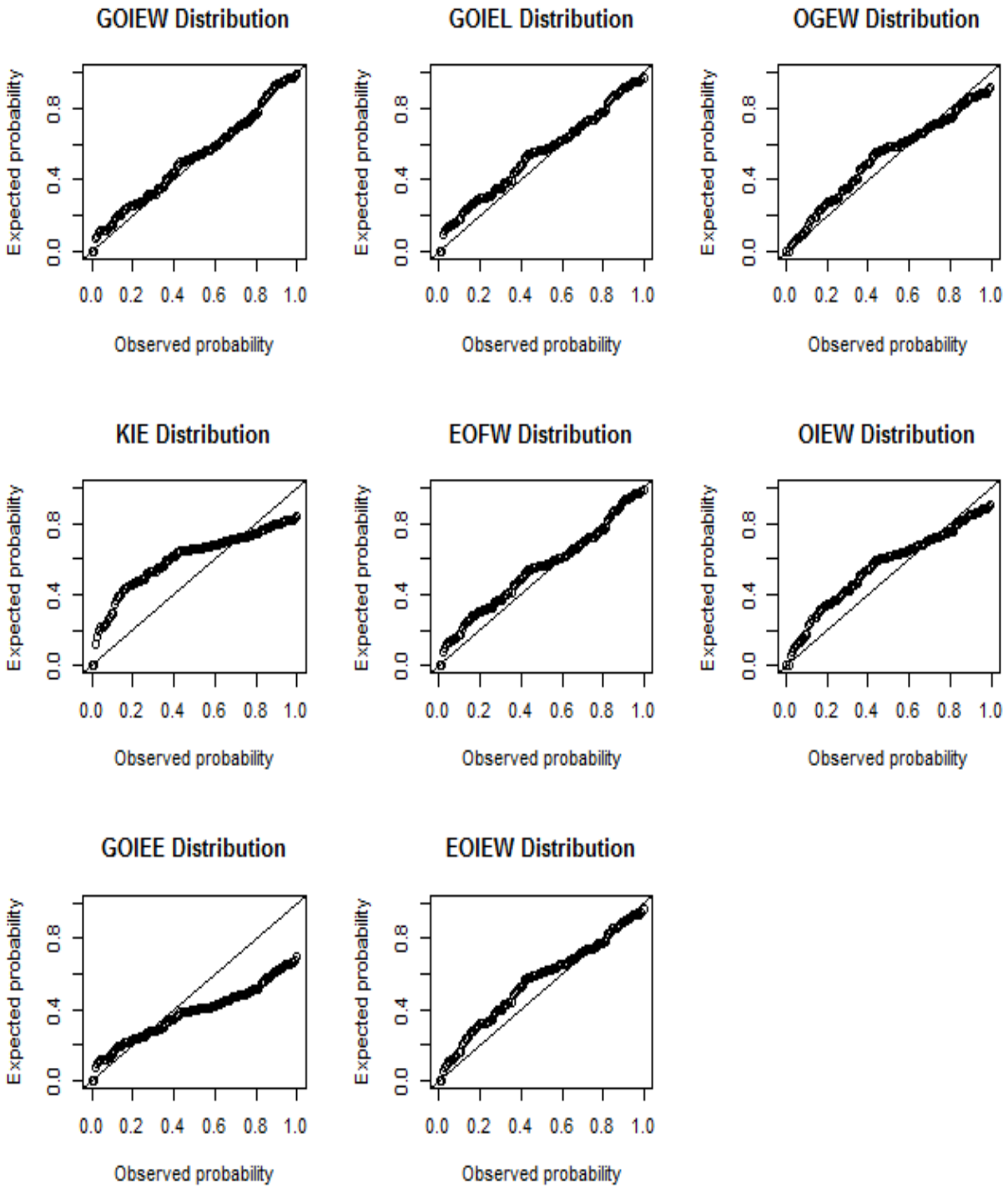

Figure 5.8: Probability-probability plots of the fitted distributions

The variance-covariance matrix of the parameters of the GOIEW distribution for the breast cancer data was also estimated and it is given by

$$J^{-1} = \begin{bmatrix} 0.0141 & 0.0923 & 0.0015 & -0.0430 \\ 0.0923 & 0.7237 & 0.0102 & -0.2957 \\ 0.0015 & 0.0102 & 0.0002 & 0.0047 \\ -0.0430 & -0.2957 & -0.0047 & 0.1401 \end{bmatrix}.$$

The 95% confidence intervals for the parameters α, β, γ and θ of the GOIEW distribution were also estimated and these are $[0.0878, 0.5284]$, $[0, 2.8168]$, $[0, 0.0304]$ and $[0.6037, 2.0011]$ respectively.

5.2.1.4 Head and Neck Cancer Data

The descriptive statistics of the head and neck cancer data are shown Table 5.16. The highest and the least observation in the dataset are 1776 and 12.2 respectively. The average value of the data was observed to be 233.5. In addition, Table 5.16 suggest that the dataset is positively skewed and leptokurtic in nature.

Table 5.16: Descriptive statistics for head and neck cancer dataset

Minimum	Maximum	Mean	Skewness	Excess kurtosis
12.2000	1776.0000	233.5000	3.5000	15.3900

The TTT transform plot for the head and neck cancer data is shown in Figure 5.9. It is observed from Figure 5.9 that the head and neck cancer data exhibit an upside down bathtub failure rate since the plot of the data shows a concave curve shape above the 45^0 line and a subsequent convex shape below the 45^0 line.

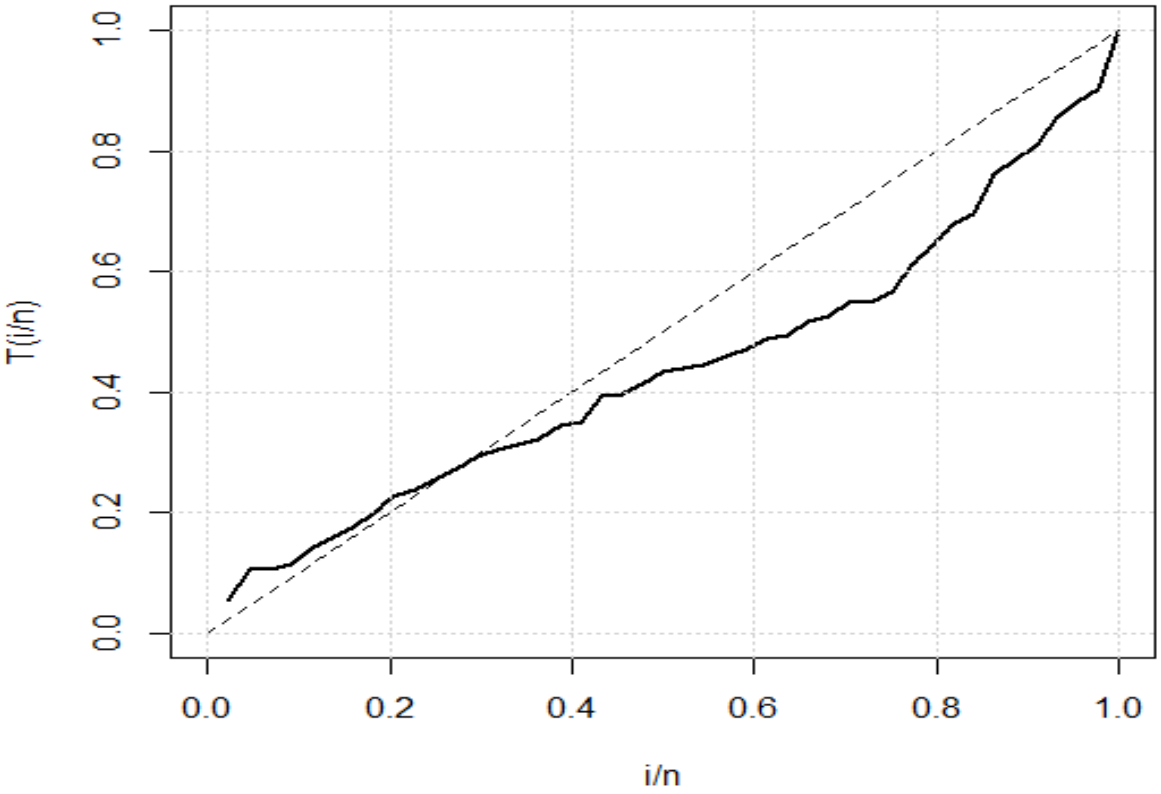

Figure 5.9: TTT transform plot for the head and neck cancer data

The maximum likelihood estimates and its associated standard errors for the parameters of the distributions that were used to modeled the head and neck cancer data are displayed in Table 5.17. The results from Table 5.17 shows that most of the parameters for the various distributions were significant at the 5% level of significance except that of EOFW distribution.

Table 5.17: Maximum likelihood estimates for head and neck cancer dataset

Distribution	Parameter	Standard Error	Z-Value	*P-Value*
GOIEW	$\hat{\alpha}$ = 77.3923	36.1128	2.1431	0.0321*
	$\hat{\beta}$ = 394.6432	6.3112	62.5308	2.2000×10^{-16}***
	$\hat{\gamma}$ = 2.2630	0.4777	4.7375	2.1630×10^{-6}***
	$\hat{\theta}$ = 0.2076	0.0390	5.3196	1.0400×10^{-7}***
KIE	$\hat{a}$ = 57.1140	0.0075	7603.2528	2.2000×10^{-16}***
	$\hat{b}$ =1.1679	0.2433	4.8011	1.5780×10^{-6}***
	$\hat{\theta}$ =1.4856	0.2888	5.1443	2.686×10^{-7}***
GOIEL	$\hat{\alpha}$ = 38.2754	23.4346	1.6333	0.1024
	$\hat{\beta}$ = 398.0173	2.3885	166.6381	2.2000×10^{-16}***
	$\hat{\gamma}$ = 0.6344	0.6026	1.0528	0.2924
	$\hat{\theta}$ = 1.4199	0.2931	4.8450	1.2660×10^{-6}***
EOFW	$\hat{\alpha}$ = 3.5351	43.6890	0.9049	0.3655
	$\hat{\beta}$ = -0.2809	0.1651	-1.7012	0.0889
	$\hat{\lambda}$ =13.7590	8.0738	1.7042	0.0884
	$\hat{\theta}$ = -0.5717	0.3270	-1.7484	0.0804
GOIEE	$\hat{\alpha}$ = 8.7890×10^{-1}	9.4143×10^{-2}	9.3357	2.2000×10^{-16}***
	$\hat{\beta}$ = 2.5806×10^{2}	3.2936×10^{-5}	7.8352×10^{6}	2.2000×10^{-16}***
	$\hat{\gamma}$ = 9.6392×10^{-3}	7.4380×10^{-4}	1.2959×10^{1}	2.2000×10^{-16}***
EOIEW	$\hat{\alpha}$ = 4.3453×10^{2}	3.7615×10^{-4}	1.1552×10^{6}	2.2000×10^{-16}***
	$\hat{\gamma}$ = 4.0868	1.7454×10^{-1}	2.3416×10^{1}	2.2000×10^{-16}***
	$\hat{\theta}$ =1.0725×10^{-1}	1.1500×10^{-2}	9.3266	2.2000×10^{-16}***

*means significant at 5% level of significance

From Table 5.18, the GOIEW recorded the least values of A^*, W^*, K-S, AIC, AICc and BIC as well as highest value of the log-likelihood when compared to GOIEL, KIE, GOIEE, EOIEW and EOFW. This indicates that the GOIEW performs better than the other models.

Table 5.18: Goodness-of-fit statistics and information criteria for head and neck cancer dataset

Model	$-\ell$	AIC	AICc	*BIC*	A^*	W^*	*K-S*	*P-Value*
GOIEW	**-277.7800**	**563.5564**	**564.5820**	**570.6932**	**0.2061**	**0.0325**	**0.0769**	**0.9394**
KIE	-279.3100	564.6116	565.2116	569.9642	0.4512	0.0730	0.1056	0.6612
GOIEL	-279.6800	567.3698	568.3954	574.5065	0.2652	0.0326	0.1331	0.3829
EOFW	-444.000	564.0247	565.0503	571.1614	0.2613	0.0446	0.0814	0.9095
GOIEE	-306.5200	619.0498	619.6498	624.4023	2.8418	0.5328	0.4995	1.127×10^{-10}
EOIEW	-281.0000	568.0057	568.6057	573.3582	0.6828	0.1135	0.1191	0.5220

*bolded means best based on selection criteria

The GOIEW distribution was compared with its sub-models using LRT in Table 5.19. The LRT statistics and their *P*-values suggest that GOIEW distribution gives a better fit than its sub-models.

Table 5.19: Likelihood ratio test statistic for Head and Neck cancer data

Model	Hypothesis	LRT	P-Value
GOIEE	$H_0:\theta=1\ Vs\ H_1:\ H_0$ is false	57.4930	1.2660×10^{-6}***
EOIEW	$H_0:\alpha=\beta\ Vs\ H_1:\ H_0$ is false	6.4493	0.0111*

Figure 5.10 represents the plots of empirical density with the fitted distributions and the CDF of the empirical distribution with the CDFs of the fitted distribution. The fitted distributions from Figure 5.10 tries to mimic the shape of the CDF and density of the empirical distribution.

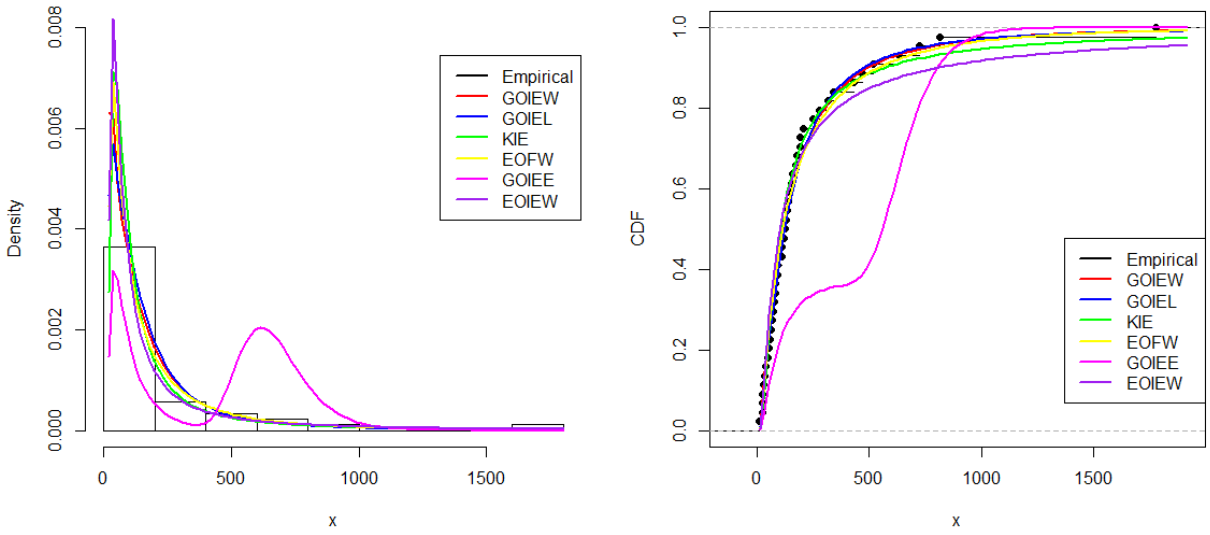

Figure 5.10: Plots of fitted densities and CDFs for head and neck cancer data

The P-P plots in Figure 5.11 further affirms the fact that GOIEW provides a better fit to the head and neck cancer data than the other distributions.

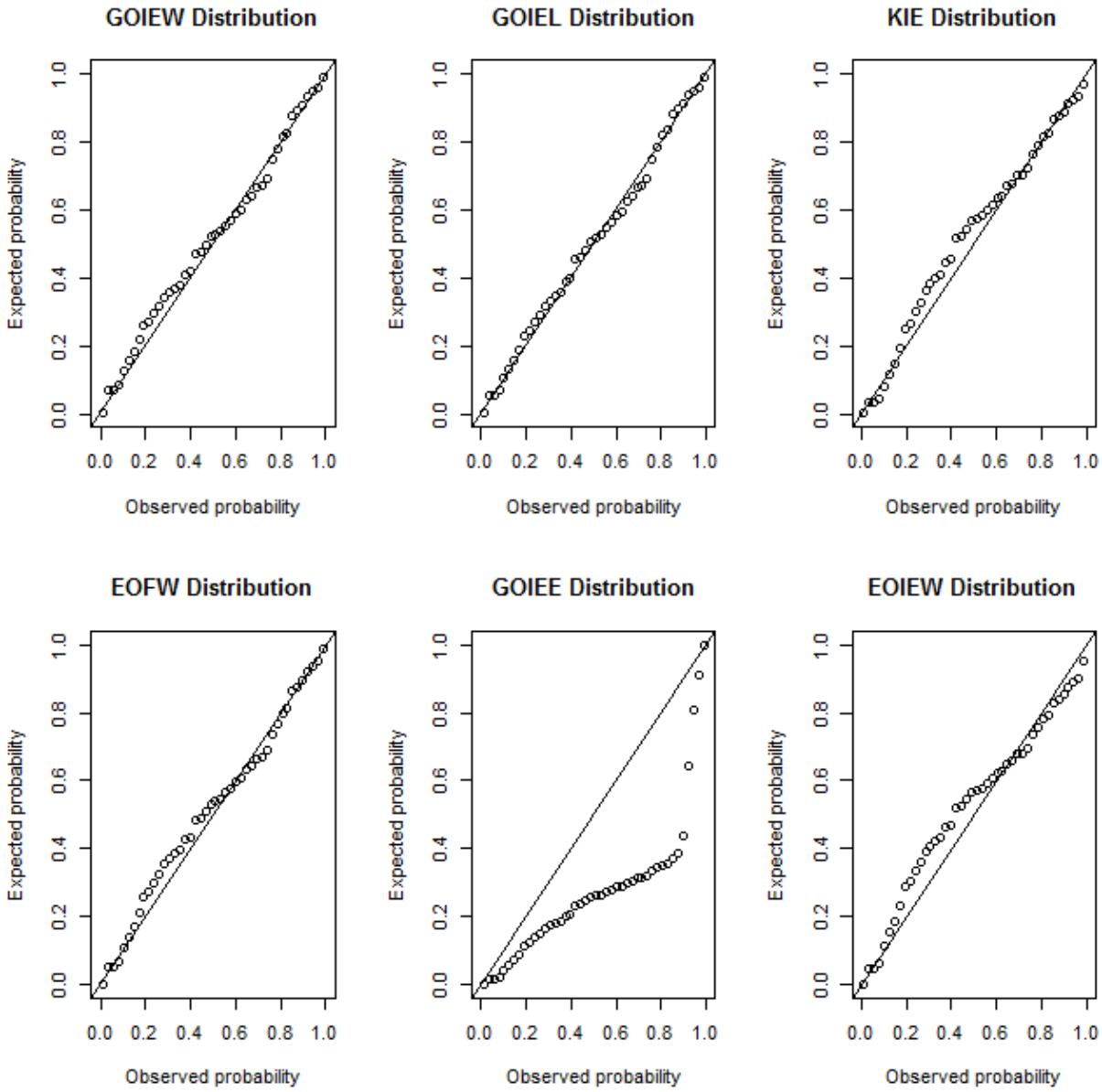

Figure 5.11: P-P plots of the fitted distributions

The variance-covariance matrix for the parameter estimates of the GOIEW distribution for the head and neck cancer dataset is given by

$$J^{-1} = \begin{bmatrix} 1244.2733 & -210.1850 & 15.4294 & -1.2169 \\ -210.1850 & 35.5048 & -2.6064 & 0.2056 \\ 15.4294 & -2.6064 & 0.2223 & -0.0180 \\ -2.2169 & 0.2056 & -0.0180 & 0.0015 \end{bmatrix}.$$

Further, the 95% confidence intervals of the estimated parameters α, β, γ and θ of GOIEW distribution were derived and these are $[6.6112, 148.1734]$, $[382.2733, 407.0132]$, $[1.3267, 3.1993]$ and $[0.1312, 0.2840]$ respectively.

5.2.2 Censored Datasets

This subsection examines the application of the GOIEW and GOIEL distributions to four censored cancer datasets.

5.2.2.1 Cancer of the Tongue

The descriptive statistics for the cancer of the tongue data is displayed in Table 5.20. The results from Table 5.20 suggest that the data was right skewed and leptokurtic in nature. The mean, minimum and maximum values of the cancer of the tongue data were 80.73, 1.0 and 400.0 respectively.

Table 5.20: Descriptive statistics for cancer of the tongue dataset

Minimum	Maximum	Mean	Skewness	Excess kurtosis
1.0000	400.0000	80.7300	2.1702	10.0714

The hazard rate function of the cancer of the tongue dataset showed a modified bathtub shape as shown in Figure 5.12. The TTT transform plot showed a convex shape followed by concave and then convex.

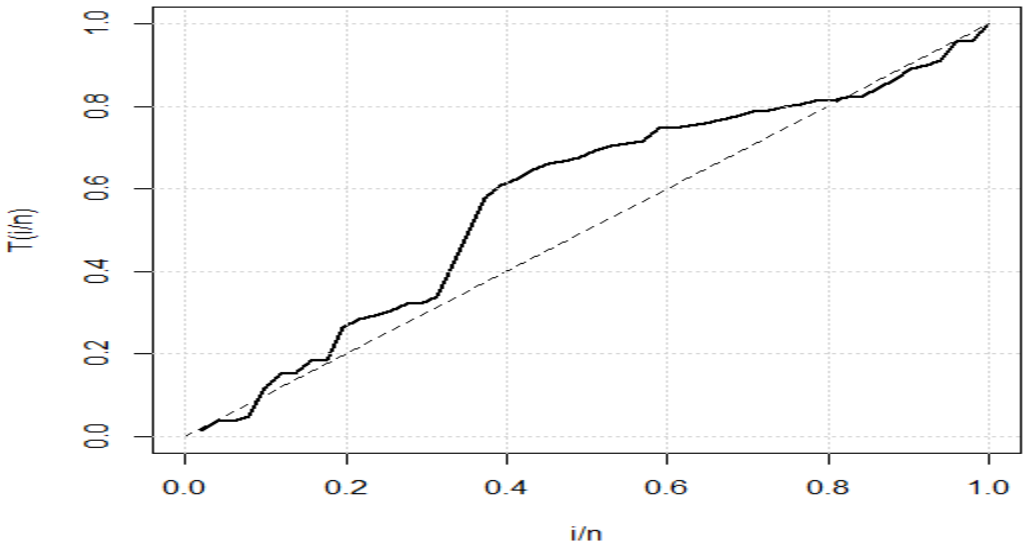

Figure 5.12: TTT transform plot for the cancer of the tongue dataset

The maximum likelihood estimates and its standard errors for GOIEW, GOIEL, Kumaraswamy inverse exponential (KIE) and generalized inverse exponential (GIE) distributions used to model cancer of the tongue data are presented in Table 5.21. All the parameters from Table 5.21 were found to be significant at the 5% level of significance except α in GOIEW as well as a and θ in KIE.

Table 5.21: Maximum likelihood estimates for cancer of tongue dataset

Distribution	Parameter	Standard Error	Z-Value	*P-Value*
GOIEW	$\hat{\alpha}$ = 30.9374	20.3004	1.5240	0.1275
	$\hat{\beta}$ = 197.4260	3.2157	61.3950	2.2000×10^{-16}***
	$\hat{\gamma}$ = 2.8825	0.6222	4.6330	3.6040×10^{-6}***
	$\hat{\theta}$ = 0.1428	0.0469	3.0480	2.3040×10^{-3}***
KIE	$\hat{a}$ = 0.2865	3.6961	0.775	0.9382
	$\hat{b}$ =0.2640	0.0583	4.5293	5.9190×10^{-6}***
	$\hat{\theta}$=22.1098	285.2039	0.0775	0.9382
GOIEL	$\hat{\alpha}$ = 9.7720×10^{-1}	2.0117×10^{-1}	4.8575	1.1890×10^{-6}***
	$\hat{\beta}$ = 2.0488×10^{2}	5.0650×10^{-4}	4.0451×10^{5}	2.2000×10^{-16}***
	$\hat{\gamma}$ = 5.0025×10^{-2}	2.9483×10^{-2}	1.6967	0.0897
	$\hat{\theta}$ = 2.9365	8.0619×10^{-1}	3.6424	1.7010×10^{-4}***
GIE	$\hat{\alpha}$ = 0.2640	0.0583	4.5293	5.9170×10^{-6}***
	$\hat{\theta}$ = 6.3369	1.9909	3.1841	1.4520×10^{-3}*

*means significant at 5% level of significance

The log-likelihood and information criteria for the cancer of the tongue dataset is displayed in Table 5.22. It was clear from Table 5.22 that GOIEW distribution performs better in terms of modeling cancer of the tongue data than the other distributions as it recorded the highest log-likelihood value and minimum values of AIC, AICc and BIC.

Table 5.22: Log-likelihood and information criteria for cancer of the tongue dataset

Model	-ℓ	AIC	AICc	*BIC*
GOIEW	**-177.7800**	**363.5694**	**364.4390**	**371.2967**
KIE	-182.0300	370.0575	370.5682	375.8530
GOIEL	-186.7600	381.5256	382.3951	390.0309
GIE	-182.0300	368.0575	368.3075	371.9212

*bolded means best based on selection criteria

The variance-covariance matrix of the parameter estimates of the GOIEW distribution for the cancer of the tongue dataset is given by

$$J^{-1} = \begin{bmatrix} 412.1063 & -65.2797 & 12.2537 & -0.9087 \\ -65.2697 & 10.3406 & -1.9412 & 0.1440 \\ 12.2537 & -1.9412 & 0.3871 & -0.0285 \\ -0.9087 & 0.1440 & -0.0285 & 0.0022 \end{bmatrix}.$$

In addition, the 95% confidence intervals of the estimated parameters α, β, γ and θ of GOIEW distribution were also obtained and these are $[0,\ 70.7262]$, $[191.1232,\ 203.7288]$,$[1.6630,\ 4.1020]$ and $[0.0509,\ 0.2347]$ respectively.

5.2.2.2 Head and Neck Cancer Data with Censored Observations

Table 5.23 presents the descriptive statistics for the head and neck cancer dataset. The minimum, average and the maximum observations for the head and neck cancer dataset are respectively 7.0, 357.8 and 1417. Results from Table 5.23 also suggest that the data was right skewed and leptokurtic in nature.

Table 5.23: Descriptive statistics for head and neck cancer dataset

Minimum	Maximum	Mean	Skewness	Excess kurtosis
7.0000	1417.0000	357.8000	1.7400	1.8700

As indicated in Figure 5.13, the TTT transform plot for the head and neck cancer data showed a convex shape followed by concave and then convex. This implies that the hazard rate function of the dataset exhibit a modified upside down bathtub shape.

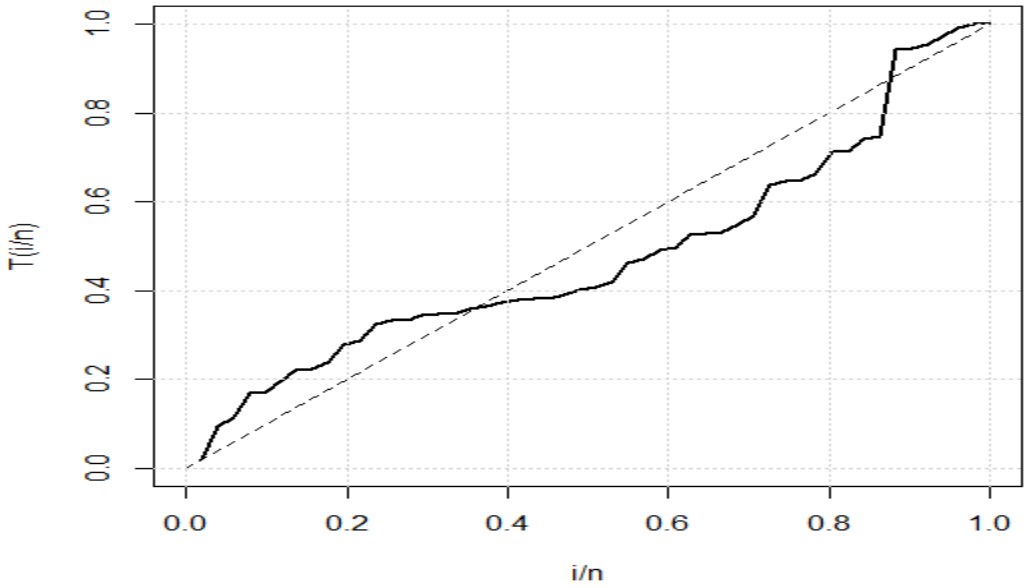

Figure 5.13: TTT transform plot for the head and neck cancer data

The parameters estimates of the models are presented in Table 5.24. It is seen that, apart from the parameters in EOFW as well as β and θ in GOIEW, the rest of the parameters were significant at the 5% level of significance.

Table 5.24: Maximum likelihood estimates for head and neck cancer dataset

Distribution	**Parameter**	**Standard Error**	Z-Value	***P-Value***
GOIEL	$\hat{\alpha}$ = 0.5689	0.1461	3.8944	9.843×10^{-5}***
	$\hat{\beta}$ = 1.0125	1.0452	0.9687	0.3327
	$\hat{\gamma}$ = 0.0025	0.0030	0.8405	0.4006
	$\hat{\theta}$ = 1.5846	0.7543	2.1009	0.0357*
EOFW	$\hat{\alpha}$ = 0.6024	0.5298	1.1371	0.2555
	$\hat{\beta}$ = 0.5785	0.3048	1.8979	0.0577
	$\hat{\lambda}$ = 0.0204	0.0501	0.4072	0.6839
	$\hat{\theta}$ = 1.4066	0.8195	1.7165	0.0861
GOIEW	$\hat{\alpha}$ = 83.3813	63.0927	1.9084	0.0563
	$\hat{\beta}$ = 391.6708	8.4311	46.4554	2.2000×10^{-16}***
	$\hat{\gamma}$ = 2.8113	0.5208	5.3980	6.7370×10^{-8}***
	$\hat{\theta}$ = 0.1466	0.0335	4.3819	1.177×10^{-5}***
OGEW	$\hat{\alpha}$ = 25.5230	4.4521×10^{-5}	5.7328×10^{5}	2.2000×10^{-16}***
	$\hat{\lambda}$ =293.9000	7.9025×10^{-8}	3.7191×10^{9}	2.2000×10^{-16}***
	$\hat{\beta}$ = 0.2066	2.5741×10^{-2}	8.0242	1.0220×10^{-15}***
	$\hat{\theta}$ = 0.0039	5.5144×10^{-4}	7.0205	2.2110×10^{-12}***

*means significant at 5% level of significance

A comparison of the models using the AIC, AICc and BIC model selection criteria as shown in Table 5.25 showed that the GOIEL model had the least AIC, AICc and BIC

values and the highest log likelihood value than the other models; hence the GOIEL model was selected as the best model for fitting this head and neck cancer dataset.

Table 5.25: Log-likelihood and information criteria for head and neck cancer dataset

Model	$-\ell$	AIC	AICc	*BIC*
GOIEL	**-294.0800**	**596.1593**	**597.0289**	**603.8866**
EOFW	-294.23	596.4046	597.3340	604.1937
GOIEW	-298.61	605.2190	606.0886	612.9463
OGEW	-295.13	598.2548	599.1244	605.9821

*bolded means best based on selection criteria

The variance-covariance matrix for the parameter estimates of the GOIEL distribution for head and neck cancer data is given by

$$J^{-1} = \begin{bmatrix} 2.1340\times10^{-2} & 1.1633\times10^{-1} & 3.5584\times10^{-4} & -3.7932\times10^{-2} \\ 1.1633\times10^{-1} & 1.0924\times10^{0} & 2.5492\times10^{-3} & -9.3976\times10^{-2} \\ 3.5584\times10^{-4} & 2.5492\times10^{-3} & 8.7152\times10^{-6} & -1.3981\times10^{-3} \\ -3.7932\times10^{-2} & -9.3976\times10^{-2} & -1.3981\times10^{-3} & 5.6890\times10^{-1} \end{bmatrix}.$$

Also, the 95% confidence intervals of the estimated parameters α, β, γ and θ of GOIEL distribution were also obtained and these are $[0.2825,\ 0.8553]$, $[0,\ 3.0611]$, $[0,\ 0.0084]$ and $[0.1062,\ 3.0630]$ respectively.

5.2.2.3 Leukemia Dataset with Censored Observations

Table 5.26 summarizes the descriptive statistics of the leukemia data. As shown in Table 5.26, the least and the highest observation in the leukemia dataset was recorded as 1.0 and 91.0 respectively. It was also observed from Table 5.26 that the dataset is right skewed with positive excess kurtosis.

Table 5.26: Descriptive statistics for leukemia dataset

Minimum	Maximum	Mean	Skewness	Excess kurtosis
7.0000	1417.0000	357.8000	1.7400	1.8700

The hazard rate function of the leukemia dataset showed a bathtub shape as exhibited in Figure 5.14. The TTT transform plot showed a convex shape followed by concave.

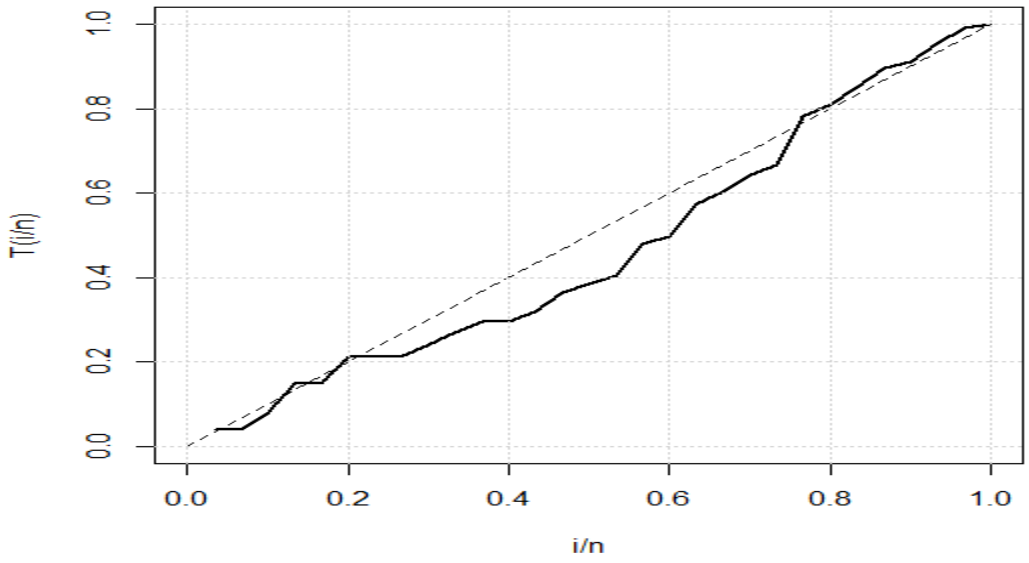

Figure 5.14: TTT transform plot for leukemia dataset

Table 5.27 displays the maximum likelihood estimates of the distributions used in modeling the leukemia dataset. It was realized that at least two of the parameters of GOIEL, GOIEW and exponentiated Lomax (E-Lx) distributions were found to be significant at the 5% level of significance. However, none of the parameters in Kum-BIII distribution were significant at the 5% level of significance.

Table 5.27: Maximum likelihood estimates for leukemia dataset

Distribution	**Parameter**	**Standard Error**	Z-Value	***P-Value***
GOIEW	$\hat{\alpha}$ = 8.8707	5.1746	1.7143	0.0865
	$\hat{\beta}$ = 30.3543	1.2499	24.2849	2.2000×10^{-16}***
	$\hat{\gamma}$ = 1.9311	0.4831	3.9975	6.4010×10^{-5}***
	$\hat{\theta}$ = 0.2329	0.0701	3.3231	8.9010×10^{-4}***
GOIEL	$\hat{\alpha}$ = 7.5036×10^{-1}	1.1451×10^{-1}	6.5527	5.6510×10^{-11}***
	$\hat{\beta}$ = 2.4979×10^{2}	2.7915×10^{-2}	8.9481×10^{3}	2.2000×10^{-16}***
	$\hat{\gamma}$ = 9.3671×10^{-3}	7.0296×10^{-3}	1.3325	0.1827
	$\hat{\theta}$ = 1.2141×10^{1}	7.3691	1.6476	0.0994
E-Lx	$\hat{\alpha}$ = 195.8812	115.3214	1.6986	0.0894
	$\hat{\lambda}$ = 940.2903	5.9546	157.9104	2.2000×10^{-16}***
	$\hat{\theta}$ = 0.5871	0.0695	8.4466	2.2000×10^{-16}***
Kum-BIII	$\hat{a}$ = 1.5861	100.4313	0.01598	0.9874
	$\hat{b}$ =3.4476	6.5281	0.5281	0.5974
	$\hat{c}$ = 0.4878	0.3274	1.4900	0.1363
	$\hat{k}$ = 4.6013	291.3627	0.0158	0.9874

*means significant at 5% level of significance

Among the competing distributions fitted to the leukemia dataset, the GOIEW distribution appears to be the best distribution for modeling the leukemia data as shown in Table 5.28.

This is because the GOIEW distribution had the minimum values of AIC, AICc and BIC compared to other distributions.

Table 5.28: Log-likelihood and information criteria for leukemia dataset

Model	$-\ell$	AIC	AICc	*BIC*
GOIEW	**-108.2700**	**224.5469**	**224.5469**	**230.1517**
GOIEL	-117.1600	242.3216	243.9216	247.9264
E-Lx	-109.7800	225.5637	226.4868	229.7673
Kum-BIII	-108.3200	224.6342	226.2342	230.2989

*bolded means best based on selection criteria

The variance-covariance matrix of the parameter estimates of the GOIEW distribution for leukemia dataset is given by

$$J^{-1} = \begin{bmatrix} 26.7761 & -6.4678 & 2.3586 & -0.3340 \\ -6.4678 & 1.5623 & -0.5704 & 0.0800 \\ 2.3586 & -0.5704 & 0.2334 & -0.0324 \\ -0.3340 & 0.0800 & -0.0324 & 0.0049 \end{bmatrix}.$$

Further, the 95% confidence intervals of the estimated parameters α, β, γ and θ of GOIEW distribution were derived and these are $[0, 19.0129]$, $[27.9045, 32.8041]$, $[0.9842, 2.8780]$ and $[0.0955, 0.3703]$ respectively.

5.2.2.4 Gastric Cancer Data with Censored Observations

The fourth dataset refers to 201 patients with gastric adenocarcinoma. This data can be found in Arslan *et al.* (2018), Ortega *et al.* (2017) and Martinez *et al.* (2013). x_i is the response variable which represents the time in months after surgery until death. The type of therapy represents the only covariate in the analysis that is: v_{i1} (1=surgery alone, representing 76 observations; 0=adjuvant chemo radiotherapy, representing 125 observations). The fits of the GOIEL and GOIEW regression models were compared with other competing regression models in terms of modeling the gastric adenocarcinoma dataset. These other competing regression models are models proposed by Ortega *et al.* (2017) and Arslan *et al.* (2018). The models developed by Ortega *et al.* (2017) includes; Poisson-gamma log-logistic (PGLL), Poisson Birnbaum-Saunders (PGBS), Poisson-gamma generalized half-normal (PGGHN) and Poisson-gamma Weibull (PGW). The Poisson Weibull Burr (PWB), Poisson Weibull (PW), Poisson Burr (PB), Poisson Weibull-Log-Logistic (PWLL), Poisson Rayleigh Burr (PRB), Poisson exponential Burr (PEB) and Poisson Weibull-Lomax (PWLx) were the models proposed by Arslan *et al.* (2018). The information criteria for the fitted regression models with cure fraction to the gastric cancer dataset is displayed in Table 4.29. It was evident from Table 4.29 that GOIEL regression model provided a better fit to the gastric cancer data than the other models as it recorded the least values of AIC, BIC and AICc.

Table 5.29: Information criteria for the fitted regression models with cure fraction to the gastric cancer dataset

Model	AIC	BIC	AICc
PWLx	884.4	900.9	884.7
PWB	886.4	906.2	886.6
PWLL	1011.3	1027.8	1011.5
PW	898.2	911.4	898.4
PRB	887.6	904.1	887.8
PB	944.3	957.5	944.5
PEB	899.2	915.7	899.4
GOIEW	825.0	844.8	825.5
GOIEL	**728.7**	**748.6**	**729.2**
PGGHN	892.9	909.4	893.2
PGBS	893.9	910.4	894.2
PGLL	900.1	916.7	900.4
PGW	900.3	916.8	900.6

*bolded means best based on selection criteria

The maximum likelihood estimates and its standard errors for the full GOIEL regression model with cure rate fraction to the gastric cancer data are shown in Table 4.30. It was observed from Table 5.30 that the regression coefficient for the therapy type was significant.

Table 5.30: The maximum likelihood estimates for the full GOIEL regression model with cure rate fraction to the gastric cancer dataset

Parameter	Estimates	Standard Error	95% C.L.	*P*-value
α	1.4285	0.2720	(0.8954, 1.9616)	1.5140×10^{-7}
β	764.4100	0.0005	(764.4090, 764.4110)	$<2.2000\times10^{-16}$
γ	0.7314	0.3713	(0.0037, 1.4591)	0.0488
θ	2.8630	0.5817	(1.7229, 4.0031)	8.5850×10^{-7}
β_0	1.0273	0.1882	(0.6584, 1.3962)	4.8270×10^{-8}
β_1	-0.3540	0.2079	(-0.7615, 0.0535)	0.0886

Therefore, the cure rate, π_0 of the therapy type was estimated as;

$$\pi_0 = e^{-\hat{\lambda}}.$$

But

$$\lambda_i = \left(\beta_0 + \beta_1 v_{i1}\right), \text{ for } i = 1,...,201.$$

This implies

$$\hat{\lambda} = \frac{1}{201}\sum_{i=1}^{201}\hat{\lambda}_i = 2.2756,$$

where

$$\hat{\lambda}_i = \exp(1.0273 - 0.3540 v_{i1}).$$

Hence,

$$\pi_0 = e^{-\hat{\lambda}} = 0.1027.$$

The cure rate of patients for the individual therapy types were also estimated as follows: When $v_1 = 0$, $\hat{\lambda}_0 = \exp(1.0273)$, hence the cure fraction of chemoradiotherapy alone is $\hat{\pi}_{00} = e^{-\hat{\lambda}_0} = 0.3580$. Also, when $v_1 = 1$, $\hat{\lambda}_0 = \exp(1.0273 - 0.3540)$, implies the cure fraction of surgery alone is $\hat{\pi}_{01} = e^{-\hat{\lambda}_0} = 0.5100$. The results has indicated that the proportion of cured is far greater for patients receiving surgery alone than those receiving chemoradiotherapy.

CHAPTER SIX

SUMMARY, CONCLUSION AND RECOMMENDATIONS

6.0 Introduction

The summary, conclusion and recommendations of the study are presented in this chapter.

6.1 Summary

The development of generalized class of distributions from existing distributions have received an increasing attention in the statistical literature, due to their wider application in different fields of studies. These modified distributions have the tendency of improving the flexibility as well as the goodness-of-fit when modeling lifetime dataset. Thus, it is imperative to derive or generalize the existing distribution for modeling datasets.

In this study, a new family of distributions by name GOIE family of distributions was developed and studied using the concept of relative odds. The statistical properties of the proposed family of distribution, such as quantile function, moment generating function, characteristic function, incomplete moment, inequality measures, mean residual life and order statistics were derived. Some special distributions of the GOIE family of distribution were developed. The shapes of their densities and failure rates for some chosen parameter values were examined. The plots of the failure rates indicates that GOIEW and GOIEL can assume different kind of non-monotonic failure rates. This means that datasets that shows this kind of failure rates can best be modeled by GOIEL and GOIEW distributions.

Further, the MLE, OLS and CVM were the techniques employed to estimate parameters of the distributions. Monte Carlo simulations were also carried out to assess the performance

of these estimation techniques. The results indicates that all the techniques are consistent as the sample size increases. However, the maximum likelihood estimates recorded the least values of the average biases and mean errors making it the best estimator.

Finally, the application of the special distributions were examined using eight cancer datasets. The results from the applications showed that each of the special distributions performed better than the other candidate distributions. For the cancer of the tongue (censored), leukemia, breast, head and neck (censored), leukemia (censored) cancer datasets, GOIEW was the best model. The GOIEL model performs better than the other models in bladder, head and neck and gastric cancer datasets.

6.2 Conclusion

The GOIE family of distribution was developed in this study. The statistical properties of the GOIE family of distributions were also derived and this include; moments, moment generation function, characteristics function, incomplete moment, inequality measures, mean residual life, mean and median deviations, order statistics, quantile function and moment of the p^{th} order statistics. Two special distributions, GOIEW and GOIEL were derived from the GOIE family of distributions. The results of the study shows that these special distributions exhibited different kind of failure rates and this include; increasing, decreasing, bathtub, bimodal, left-skewed, right-skewed, upside-down bathtub and reversed-J shapes. This gives the GOIEW and GOIEL the upper hand for modeling lifetime datasets that exhibits these kinds of failure rates.

The sub-models of the GOIEW and GOIEL such as OIEW, EOIEW, GOIEE, GOIER for GOIEW and OIEL, EOIEL, GOIEE and GOIER for GOIEL were also established. The usefulness of the sub-models were demonstrated using three cancer datasets.

The study also employed three estimation techniques to estimate the parameters of the distributions. Monte Carlo simulation experiments were also carried out to assess the performance of these estimators. The results indicates that all the techniques are consistent as the sample size increases with maximum likelihood technique as the best estimator.

Finally, regression models with cure fractions were established for GOIEW and GOIEL distributions and was applied to a gastric cancer dataset. The results from the application showed that each of the regression model from GOIE family performed better than the other candidate models.

6.3 Recommendations

The application of the special distributions were illustrated using only cancer datasets. More special distributions can be derived from the GOIE family of distributions and use to model other biological datasets as well as data from other fields.

The study also proposed other sub-family generators that could be used to modify distributions. Going forward, further research should concentrate on these generators when modifying other existing distributions and examine their performances in terms of modeling lifetime datasets.

Based on the findings of this study, the percentage of cured is far greater for patients receiving surgery alone than those receiving chemoradiotherapy. The cure rate could be maximized for surgery alone if there were early detection of the cancer, hence awareness should be created amongst the general public by world health organization and Ghana health service on the need for frequent cancer screening in our health facilities.

REFERENCES

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

Abd El-Raheem, A. M. (2019). Optimal design of multiple accelerated life testing for generalized half-normal distribution under type-I censoring. *Journal of Computational and Applied Mathematics.* 112539, https://doi.org/10.1016/j.cam.2019.112539.

Abdus, S. W., The, M. L. and Jong-Hyeon, J. (2009). A new generalization of Weibull distribution with application to a breast cancer dataset. *Stat Med.*, ***28****(16)*, 2077 - 2094.

Abouammoh, A. M. and Alshingiti, A. M. (2009). Reliability of generalized inverted exponential distribution. *Journal of Statistical Computation and Simulation*, ***79,*** 1301 - 1315.

Afify, A. Z., Cordeiro, G. M., Maed, M. E., Alizadeh, M., Al-Mofleh, H. and Nofal, Z. M. (2019). The generalized odd lindley-G family: Properties and applications. *Anais da Academia Brasileira de Ciencias,* ***91***(3), e20180040.

Ahmad, Z., Elgarhy, M., Hamedani, G. G. and Butt, N. S. (2020). Odd generalized N-H generated family of distributions with application to exponential model. *Pakistan Journal of Statistics and Operation Research*, ***16***(1), 53-71.

Akaike, H. (1973). Information theory and an extension of maximum likelihood principle in B. N. Petrov and F. (Saki (Eds.), *Second International Symposium on Information Theory*, pg 267 - 281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, ***19***(6), 716 - 723.

AL Sobhi, M. M. (2020). The Inverse-Power Logistic-Exponential Distribution: properties, estimation methods, and application to insurance data. *Mathematics, **8**(2060)*, 2-15.

Alizadeh, M., Altun, E., Ozel, G., Afshari, M. and Eftekharian, A. (2019). A new odd log-logistic Lindley distribution with properties and applications. *Sankhya A, **81***(2), 323-346.

Alizadeh, M., Benkhelifa, L., Rasekhi, M. and Hosseini, B. (2020). The odd log-logistic generalized Gompertz distribution: properties, applications and different methods of estimation. *Communications in Mathematics and Statistics, **8***(3), 295 -317.

Alizadeh, M., Cordeiro, G. M., Nascimento, A. D. C., Lima, M. D. C. S. and Ortega, E. M. M. (2017). Odd-Burr generalized family of distributions with some applications. *Journal of Statistical Computation and Simulation, **87***(2), 367-389.

Alizadeh, M., Yousof, H. M., Afifiy, A. Z. and Mansoor, M. (2018). The complementary generalized transmuted Poisson –G family of distributions. *Austrian Journal of Statistics, **47***(4), 51 - 71.

Alkarni, S., H. (2016). Generalised extended Weibull power series family of distribution. *Journal of Data Science, **14***(3), 415 – 440.

Anwar, M. and Bibi, A. (2018). The half-Logistic generalised Weibull Distribution. *Journal of Probability and Statistics*, 2018, htt://doi.org/10.1155/2018/8767826.

Aryal, G. R. and Yousof, H. M. (2017). The exponentiated generalized – G Poisson family of distribution. *Stochastic and Quality Control*, ***32***(1), 7 – 23.

Barlow, R. E. and Doksum, K. A. (1972). Isotonic tests for convex orderings. *In proceedings of the 6th Berkeley symposium, **1***, 293 - 323.

Barriga, D., Vicente, G. C., Daniel., V. G., Gauss, M. C. and Edwin, M. O. (2018). A new survival model with surviving fraction: An application to colorectal cancer data. Statistical Methods in Medical Research, 1-16.

Berkson, J. and Gage, R. P. (1952). Survival curves for cancer patients following treatment. *Journal of the American Statistical Association, 47,* 501–515.

Boag, J. M. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society, Ser. B, **11***, 15–44.

Borges, P. (2020). Estimating the turning point of the log-logistic hazard function in the presence of long-term survivors with an application for uterine cervical cancer data. *Journal of Applied Statistics*, 2-11.

Bourguignon, M., Silva, R. B. and Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *J. Data Sci., **12***, 53 - 68.

Boussari, O., Laurent, B., Gaëlle, R., Marc, C., Nadine, B., Laurent, R. and Valérie, J. (2020). Modeling excess hazard with time-to-cure as a parameter. *Biometric Methodology*, 1-14.

Broyden, C. G. (1970). A new method of solving nonlinear simultaneous equations. *Computational journal, **12***, 94 - 99.

Calsavara, F. V., Eder, A. M., Eduardo, B. and Vera, T. (2019). Long-term frailty modeling using a non-proportional hazards model: Application with a melanoma dataset. *Statistical Methods in Medical Research,* 1-19.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A New Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, ***94,*** 909–919.

Cordeiro, G. M., Afify, A. Z., Ortega, E. M. M., Suzuki, A. K. and Mead, M. E. (2018). The odd Lomax generator of distributions: Properties, estimation and applications, *Journal of Computational and Applied Mathematics*, https://doi.org/10.1016/j.cam.2018.08.008.

Cordeiro, G. M., Alizadeh, M., Ozel, G., Hosseini, B., Ortega, E. M. M. and Altun, E. (2017). The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of Statistical Computation and Simulation,* ***87***(5), 908-932.

Cordeiro, G. M., Alizadeh, M., Pescim, R. R. and Ortega, E. M. M. (2017). The odd log-logistic generalized half-normal lifetime distribution: Properties and applications. *Communications in Statistics - Theory and Methods,* ***46***(9), 4195-4214.

Cordeiro, G. M., Alizadeh, M., Ramires, T. G. and Ortega, E. M. M. (2017). The generalized odd half-Cauchy family of distributions: Properties and applications. *Communications in Statistics - Theory and Methods,* ***46(***11), 5685-5705.

Cordeiro, G. M., Alizadehy, M., Tahirz, M. H., Mansoor, M., Bourguignonk, M. and Hamedani, G. G. (2016). The beta odd log-logistic generalized family of distributions. *Hacettepe Journal of Mathematics and Statistics*, ***45***(4), 1175-1202.

Cordeiro, G. M., Ortega, E. M. M., Popovic, B. V. and Pescim, R. R. (2014). The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, ***247***, 465 - 486.

Elbatal, I., Altun, E., Afify, A. Z. and Ozel, G. (2018). The generalised Burr XII power series distributions with properties and applications. *Annals of Data Science*. Doi: 10.1007/s40745-018-0171-2.

Elgarhy, M., Ahsanul Haq, M. and Ain, Q. (2018). Exponential generalized Kumaswamy distribution with applications. *Annals of Data Science*, ***5***(2), 273 – 292.

Elsayed, H. A. H. and Yousof, H.M. (2020). The generalized odd generalized exponential Fréchet model: Univariate, bivariate and multivariate extensions with properties and applications to the univariate version. *Pakistan Journal of Statistics and Operation Research*, ***16***(3), 527-544.

Fletcher, R. (1970). *A class of methods for nonlinear programming with termination and convergence Properties*. In Integer and Nonlinear programming. North-Holland, Amsterdam, pp 157 - 174.

Goldfarb, D. (1970). A family of variable metric methods derived by variational means. *Mathematics of computation*, ***24***, 23 - 26.

Haghbin, H., Ozel G., Alizadeh, M. and Hamedani, G.G. (2017). A new generalized odd log-logistic family of distributions. *Communications in Statistics - Theory and Methods*, ***46***(20), 9897-9920.

Hassan, A., Elshrpieny, E. and Mohamed, R. (2019). Odd generalized exponential power function distribution: Properties and applications. *Gazi University Journal of Science*, ***32***(1), 351-370.

Hosseini, B., Afshari, M. and Alizadeh, M. (2018). The generalized odd gamma-G family of distributions: Properties and applications. *Austrian Journal of Statistics*, ***47***(2), 69-89.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, **76,*** 297 - 307.

Jennings, S. M. (2014). *Preventing chronic disease: defining the problem.* Ireland: Health Service Executive. ISBN: 978-1-908972-05-7.

Kang, S. B. and Han, J. T. (2015). The graphical method for goodness of fit test in the inverse Weibull distribution based on multiply type II censored samples. *SpringerPlus*, 4(768), 1-14.

Khalil, M. G., Hamedani, G. G. and Yousof, H. M. (2019). The Burr X exponentiated weibull model: Characterizations, mathematical properties and applications to failure and survival time's data. *Pakistan Journal of Statistics and Operation Research*, ***15***(1), 141-160.

Korkmaz, M. Ç., Alizadeh, M., Yousof, H. M. and Butt, N. S. (2018). The generalized odd Weibull generated family of distributions: Statistical properties and applications. *Pakistan Journal of Statistics and Operation Research*, ***14***(3), 541-556.

Kumar, M., Sanjay, K. S. and Umesh, S. (2018). Bayesian inference for Poisson-inverse exponential distribution under progressive type-II censoring with binomial removal. *The Society for Reliability Engineering, Quality and Operations Management,* 1-15.

Leão, J., Víctor, L., Helton, S. and Vera, T. (2018). Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Wiley Statistics in Medicine,* 1-20.

Maiti, S. S. and Pramanik, S. (2015). Odds Generalized Exponential – Exponential Distribution. *Journal of Data Science, **13**,* 733-754.

Mallet, A. (1986). A maximum likelihood method for random coefficient regression models. *Biometrika*, ***73***(3), 645 - 656.

Martinez, E. Z., Achcar, J. A., Jácome, A. A. A. and Santos, J. S. (2013). Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput Methods Prog Biomed,* ***112,*** 343–355.

Martinez, Z. E. and Achcar, A. J. (2014). Bayesian bivariate generalized Lindley model for survival data with a cure fraction. *Computer Methods Programs Biomedicine,* 2-13.

Martinez, Z. E. and Achcar, A. J. (2018). A new straightforward defective distribution for survival analysis in the presence of a cure fraction. *Journal of Statistical Theory and Practice,* 1-26.

Megan, C. (2013). Chronic Disease Prevention and Management. *National Conference of State Legislatures* (pp. 4-15). Denver: National Conference of State Legislatures ISBN 978-1-58024-699-6.

Muhammad, M. (2017). Generalized half-logistic Poisson distributions. *Communications for Statistical Applications and Methods,* ***24***(4), 353 – 356.

Muhammad, M. (2018). Poisson-odd generalized exponential family of distributions: Theory and applications. *Hacettepe Journal of Mathematics and Statistics,* ***47***(6), 1652-1670.

Nasir, A., Jammal, F. and Sha, A. A. (2018). *A compounded four-parameter lifetime model: properties, cure rate model and applications.* Hal-01902847v1.

Nasir, A., Yousof, H. M., Jamal, F. and Korkmaz, M. C. (2019). The exponentiated Burr XII power series distribution: properties and applications. *Stats,* ***2***(1), 15 – 31.

Nasiru, S., Mwita, P. N. and Ngesa, O. (2017). Exponentiated generalized Transformed-Transformer family of distributions. *Journal of Statistical and Econometric methods*, **4**, 1-17.

Nasiru, S., Mwita, P. N. and Ngesa, O. (2017). Exponentiated generalized exponential Dagum distribution. *Journal of King Saud University – Science,* ***31***, 362 - 371.

Nwezza, E. E. and Ugwuowo, F. I. (2016). The Marshall-Olkin Gumbel-Lomax distribution: properties and applications. *Heliyon,* ***6,*** e03569.

Oguntunde, P. E. and Adejumo, A. O. (2015). The generalised inverted generalised exponential distribution with an application to a censored data. *Journal of Statistics Application and Probability,* ***4*****(**2), 223-230.

Oguntunde, P. E., Adebowale, O. A. and Enahoro, A. O. (2017). Exponential Inverse Exponential (EIE) Distribution with Applications to Lifetime Data. *Asian Journal of Scientific Research,* ***10***(3), 169-177.

Oguntunde, P. E., Adejumo, A. O. and Balogun, O. S. (2014). Statistical properties of the exponentiated generalized inverted exponential distribution. *Applied Mathematics,* ***24***(2), 47-55.

Oguntunde, P. E., Babatunde, O. S. and Ogunmola, A. O. (2014). Theoretical analysis of the Kumaraswamy-inverse exponential distribution. *International Journal of Statistics and Applications,* ***4***(2), 113-116.

Okashaa, H. M. and Kayid, M. (2015). A new family of Marshall–Olkin extended generalized linear exponential distribution. *Journal of Computational and Applied Mathematics*, ***296***, 576–592.

Oluyede, B. O., Mashabeb, B., Fagbamigbec, A., Makubateb, B. and Wandukua, D. (2020). The exponentiated generalized power series. Family of distributions: theory, properties and applications. *Heliyon*, ***6***, e04653.

Ortega, E. M. M., Cordeiro, G. M., Hashimoto, E. M. and Suzuki, A. K. (2017). Regression models generated by gamma random variables with long-term survivors. *Cummum stat Applic Methods.*, ***24***, 43-65.

Ortega, M. E., Artur, J. L., Giovana, O. S. and Gauss, M. C. (2015). New flexible models generated by gamma random variables for lifetime modeling. *Journal of Applied Statistics*, 2-21.

Prataviera, F., Ortega, E. M. M., Cordeiro, G. M., Pescim, R. R. and Verssani, B. A. W. (2018). A new generalized odd log-logistic ßexible Weibull regression model with applications in repairable systems. *Reliability Engineering and System Safety*. doi: 10.1016/j.ress.2018.03.034.

Rahmouni, M. and Orabi, A. (2018). The exponential - generalized truncated geometric distribution. A new lifetime distribution. *International Journal of Statistics and Probability*, ***7***(1), 1 – 20.

Ramos, M. A., Cordeiro, G.M., Marinho, P. D., Dias, C. B. and Hamadani, G. G. (2013). The zografos-balakrishman log-logistic distribution: properties and applications. *Journal of Statistical Theory and Applications*, ***12***(3), 225-244.

Rodrigues, J., de Castro, M., Cancho, V. G. and Balakrishnan, N. (2009). COM–Poisson cure rate survival models and an application to a cutaneous melanoma data. *J Stat Plan Inference*, ***139***, 3605–3611.

Schumitzky, A. (1991). Nonparametric EM algorithms for estimating prior distributions. *Applied Mathematics and computation*, ***45***(2), 143 - 157.

Schwarz, G. E. (1978). Estimating the Dimensions of a Model. *Annals of Statistics*, ***6***(2), 461 - 464.

Segura, J. L. and Arias, M. A. V. (2020). Location of the zeros of certain parametric families of functions of generalized Fresnel integral type. *Applied Mathematics and Computation*, ***381***, 125253.

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation*, ***24***, 647 - 656.

Sugiura, N. (1978). Further analysis of the Data by Akaike Information Criterion and the finite corrections. *Communication in statistics – theory and Methods*, ***A7,*** 13 - 26.

Sun, S. and He, S. (2019). Generalizing expectation propagation with mixtures of exponential family distributions and an application to Bayesian logistic regression. *Neurocomputing*, ***337***, 180 - 190.

Tahir, M. H., Cordeiro, G. M., Alizadeh, M., Mansoor, M., Zubair, M. and Hamedani, G. G. (2014). The odd generalized exponential family of distributions with applications. *Journal of Statistical Distributions and Applications*, pp 1-28.

Tahir, M. H., Cordeiro, G. M., Alizadeh, M., Mansoor, M., Zubair, M. and Hamedani, G. G. (2015). The odd generalized exponential family of distributions with applications. *Journal of Statistical Distributions and Applications*, ***2***(1), 1.

Tahmasebi, S. and Jafari, A. A. (2015). Generalized Gompertz-power series distributions. https://arxiv.org/abs/1508.07634v1.

Tsodikov, A. D., Ibrahim, J. G. and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *J Am Stat Assoc*, ***98***, 1063–1078.

Vallinayagam, V., Prathap, S. and Venkatesan, P. (2014). Parametric regression models in the analysis of breast cancer survival data. *International Journal of Science and Technology*, 163-167.

Vatto, V. T., Nascimento, A. D. C., Miranda, F. W. R., Lima, M. C. S., Pinho, L. G. B. and Cordeiro, G. M. (2016). Exponentiated generalized Nadarajah-Haghighi distribution. *Chilean Journal of Statistics.* Available at: https://arxiv.org/abs/1610.08876. Accessed on 11th December, 2020.

Wang, L., Tripathi, Y. M. and Lodhi, C. (2019). Inference for Weibull competing risks model with partially observed failure causes under generalized progressive hybrid censoring, *Journal of Computational and Applied Mathematics.* doi: https://doi.org/10.1016/j.cam.2019.112537.

World Health Organization (2005). *Preventing chronic diseases: a vital investment.* Canada: Public Health Agency of Canada.

Wullianallur, R. and Viju, R. (2018). An empirical study of chronic diseases inthe united states: a visual analytics approach to public health. *International Journal of Environmental Research and Public Health*, 1-24.

Yakovlev, A., Asseleani, B., Bardou, V., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A. A. (1993). Stochastic models of tumor latency and their biostatistical applications. *Biometrie et Analyse de Donnes Spatio-Temporelles,* ***12,*** 66–82.

Yousof, H. M., Korkmaz, M. C. and Hamedani, G. G. (2017). The odd Lindley Nadarajah-Haghighi distribution. *Journal of Mathematical Computed Science*, ***7***(5), 864 - 882.

Zhou, Z., Zhang, A., Ding, C. and Xiong, M. (2013). The weight enumerator of three families of cyclic codes. *IEEE Transactions on Information Theory,* ***59***(9), 6002-6009.

Zilber, D. and Katzfuss, M. (2020). Vecchia–Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data. *Computational Statistics and Data Analysis,* ***153***, 107081.

Zubair, M., Pogány, T. K., Cordeiro, G. M. and Tahir, M. H. (2019). The log-odd normal generalized family of distributions with application. *Anais da Academia Brasileira de Ciencias*, ***91***(2), e20180207.

APPENDIX A

PDFs of other competing distributions

1. Poisson Weibull-Burr distribution

$$f(x)=\frac{\tau e^{-ax^b}}{\left(1+x^c\right)^k}\left[abx^{b-1}+\frac{ckx^{c-1}}{\left(1+x^c\right)}\right]\frac{\exp\left\{-\tau\left[1-e^{-ax^b}\left(1+x^c\right)^{-k}\right]\right\}}{1-e^{-\tau}}.$$

2. Poisson Weibull distribution

$$f(x)=\tau e^{-ax^b}\left[abx^{b-1}+\frac{1}{\left(1+x^c\right)}\right]\frac{\exp\left\{-\tau\left[1-e^{-ax^b}\right]\right\}}{1-e^{-\tau}}.$$

3. Poisson Burr distribution

$$f(x)=\frac{\tau}{\left(1+x^c\right)^k}\left[\frac{ckx^{c-1}}{\left(1+x^c\right)}\right]\frac{\exp\left\{-\tau\left[1-\left(1+x^c\right)^{-k}\right]\right\}}{1-e^{-\tau}}.$$

4. Poisson Weibull log-logistic distribution

$$f(x)=\frac{\tau e^{-ax^b}}{\left(1+x^c\right)}\left[abx^{b-1}+\frac{cx^{c-1}}{\left(1+x^c\right)}\right]\frac{\exp\left\{-\tau\left[1-e^{-ax^b}\left(1+x^c\right)^{-1}\right]\right\}}{1-e^{-\tau}}.$$

5. Poisson Rayleigh Burr distribution

$$f(x)=\frac{\tau e^{-ax^2}}{\left(1+x^c\right)^k}\left[2ax+\frac{ckx^{c-1}}{\left(1+x^c\right)}\right]\frac{\exp\left\{-\tau\left[1-e^{-ax^2}\left(1+x^c\right)^{-k}\right]\right\}}{1-e^{-\tau}}.$$

6. Poisson exponential-Burr distribution

$$f(x)=\frac{\tau e^{-ax}}{\left(1+x^{c}\right)^{k}}\left[a+\frac{ckx^{c-1}}{\left(1+x^{c}\right)}\right]\frac{\exp\left\{-\tau\left[1-e^{-ax}\left(1+x^{c}\right)^{-k}\right]\right\}}{1-e^{-\tau}}.$$

7. Poisson Weibull-Lomax distribution

$$f(x)=\frac{\tau e^{-ax^{b}}}{\left(1+x\right)^{k}}\left[abx^{b-1}+\frac{k}{\left(1+x\right)}\right]\frac{\exp\left\{-\tau\left[1-e^{-ax^{b}}\left(1+x\right)^{-k}\right]\right\}}{1-e^{-\tau}}.$$

8. Extended odd Fréchet Weibull distribution

$$f(x)=\frac{\alpha\beta\lambda\theta x^{\beta-1}e^{-\lambda x^{\beta}}\left[1-\left(1-e^{-\lambda x^{\beta}}\right)^{\alpha}\right]^{\theta-1}}{\left(1-e^{-\lambda x^{\beta}}\right)^{\alpha\theta+1}}e^{-\left[\left(1-e^{-\lambda x^{\beta}}\right)^{-\alpha}-1\right]^{\theta}}.$$

9. Odd generalized exponential Weibull distribution

$$f(x)=\alpha\lambda\theta\beta x^{\beta-1}e^{\theta x^{\beta}}e^{-\lambda\left(e^{\theta x^{\beta}}-1\right)}\left(1-e^{-\lambda\left(e^{\theta x^{\beta}}-1\right)}\right)^{\alpha-1}$$

10. Generalized inverse exponential distribution

$$f(x)=\frac{\alpha\theta}{x^{2}}e^{-\frac{\theta}{x}}\left(1-e^{-\frac{\theta}{x}}\right)^{\alpha-1},\ x>0,\alpha>0,\theta>0.$$

11. Kumaraswamy inverse exponential distribution

$$f(x)=ab\frac{a}{x^2}\left[\exp\left(-\frac{\theta}{x}\right)\right]^a\left\{1-\left[\exp\left(-\frac{\theta}{x}\right)\right]^a\right\}^{b-1},\ x>0,a>0,b>0,\theta>0.$$

12. Generalized inverse exponential distribution

$$f(x)=\gamma\beta\alpha^{\beta}x^{-(\beta+1)}\exp\left[-\gamma\left(\frac{\alpha}{x}\right)^{\beta}\right],\ x>0.$$

13. Exponentiated Lomax distribution

$$f(x)=\alpha\theta\lambda\left[1-(1+\lambda x)^{-\theta}\right]^{\alpha-1}(1+\lambda x)^{-(\theta+1)};\ x>0,\alpha,\theta,\lambda>0.$$

14. Inverse Weibull distribution

$$f(x)=\beta\alpha^{\beta}x^{-(\beta+1)}\exp\left[-\left(\frac{\alpha}{x}\right)^{\beta}\right].$$

15. Kumaraswamy Burr type III distribution

$$f(x)=abckx^{-(c+1)}(1+x^{-c})^{-(ak+1)}\left(1-(1+x^{-c})^{-ak}\right)^{b-1}\ x>0;\ a,b,c,k>0.$$

16. Poisson-gamma Weibull distribution

$$f(x)=\frac{\tau\lambda x^{\lambda-1}\exp\left\{-\left(x/\alpha\right)^{\lambda}\right\}}{\alpha^{\lambda}\Gamma(a)\left[1-\exp(-\tau)\right]}\left(\frac{x}{\alpha}\right)^{\lambda(\alpha-1)}\exp\left\{\frac{-\tau\gamma\left[a,\left(x/\alpha\right)^{\lambda}\right]}{\Gamma(a)}\right\}.$$

17. Poisson-gamma log-logistic distribution

$$f(x)=\frac{\tau\lambda x^{\lambda-1}\left[1+\left(x/\alpha\right)^{\lambda}\right]^{-2}}{\alpha^{\lambda}\Gamma(a)\left[1-\exp(-\tau)\right]}\left\{\log\left[1+\left(\frac{x}{\alpha}\right)^{\lambda}\right]\right\}^{\alpha-1}\exp\left\{\frac{-\tau\gamma\left[a,\log\left[1+\left(x/\alpha\right)^{\lambda}\right]\right]}{\Gamma(a)}\right\}.$$

18. Poisson-gamma Birnbaum-Saunders distribution

$$f(x)=\frac{\tau(x+\lambda)x^{-3/2}}{2\sqrt{2\lambda\pi}\alpha\Gamma(a)\left[1-\exp(-\tau)\right]}\exp\left\{\frac{-1}{2\alpha^{2}}\left[\left(\frac{x}{\lambda}\right)+(\lambda x)-2\right]\right\}\times\left\{-\log\left\{1-\Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{x}{\lambda}}-\sqrt{\frac{\lambda}{x}}\right)\right]\right\}\right\}^{\alpha-1}$$
$$\exp\left\{\frac{-\tau\gamma\left(a,-\log\left[1-\Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{x}{\lambda}}-\sqrt{\frac{\lambda}{x}}\right)\right]\right]\right)}{\Gamma(a)}\right\}.$$

19. Poisson-gamma generalized half-normal distribution

$$f(x)=\frac{\tau\sqrt{2}\left(\alpha/x\right)\left(x/\lambda\right)^{\alpha}}{\sqrt{\pi}\Gamma(a)\left[1-\exp(-\tau)\right]}\exp\left[\frac{1}{2}\left(\frac{x}{\lambda}\right)^{2\alpha}\right]\left\{-\log\left\{1-2\Phi\left[\left(\frac{x}{\lambda}\right)^{\alpha}\right]\right\}\right\}^{\alpha-1}$$
$$\exp\left\{\frac{\tau\gamma\left(a,-\log\left\{1-2\Phi\left[\left(\frac{x}{\lambda}\right)^{\alpha}\right]\right\}\right)}{\Gamma(a)}\right\}.$$

APPENDIX B

```
############################ GOIEW Distribution PDF ########################
PDF_GOIEW<-function(x,alpha,beta,gamma,theta){
  A<-theta*gamma*(x^(theta-1))*exp(-gamma*(x^theta))
  B<-1-exp(-gamma*(x^theta))
  C<-((beta-alpha)*(B^(beta-alpha-1)))+(alpha*(B^(-alpha-1)))
  PDF<-A*C*exp(-((1-B^beta)/B^alpha))
  return(PDF)
}
############################ GOIEW Distribution CDF ########################
CDF_GOIEW<-function(x,alpha,beta,gamma,theta){
  A<-1-exp(-gamma*(x^theta))
  CDF<-exp(-((1-A^beta)/A^alpha))
  return(CDF)
}
############################ GOIEW Distribution Hazard function ################
Hazard_GOIEW<-function(x,alpha,beta,gamma,theta){
  A<-theta*gamma*(x^(theta-1))*exp(-gamma*(x^theta))
  B<-1-exp(-gamma*(x^theta))
  C<-((beta-alpha)*(B^(beta-alpha-1))+alpha*(B^(-alpha-1)))
  HAZARD<-(A*C*exp(-(1-B^beta)/B^alpha))/(1-exp(-(1-B^beta)/(B^alpha)))
```

```
return(HAZARD)
}
################## GOIEW Distribution Survival function ##################
SURVIVAL_GOIEW<-function(x,alpha,beta,gamma,theta){
 A<-1-exp(-gamma*(x^theta))
 B<-exp(-((1-A^beta)/A^alpha))
 SURVIVAL<-1-B
 return(SURVIVAL)
}
############################## GOIEW Distribution Quantile function ###############
Qf<-function(alpha,beta,gamma,theta){
 output<-0
 u<-seq(0.1,0.9,0.1)
 for(i in 1:length(u)){
  f<-function(x){(((1-exp(-gamma*(x^theta)))^alpha)*log(u[i]))-((1-exp(-
gamma*(x^theta)))^beta)+1}
  rc<-uniroot(f,lower=0,upper=10000,tol=1e-9)
  output[i]=rc$root
 }
output
}
```

```
u<-seq(0.1,0.9,0.1)

cbind(u,Qf(0.1,0.3,0.7,0.2),Qf(10,20,40,0.5),Qf(1.0,7,0.8,0.5),Qf(15,24,33,12))
```

############################ GOIEW Distribution moment ########################

```
Moment<-function(alpha,beta,gamma,theta){
  results<-0
  r<-seq(1,6,1)
  for(i in 1:length(r)){
    f<-function(x,alpha,beta,gamma,theta,r){(x^r[i])*(PDF_ODIEW(x,alpha,beta,gamma,theta))}
    results[i]<-
integrate(f,lower=0,upper=Inf,subdivisions=1000,alpha=alpha,beta=beta,gamma=gamma,thet
a=theta,r=r)$value
  }
  return(results)
}

r<-seq(1,6,1)

print(cbind(r,Moment(10.5,10.8,5,1.5),Moment(0.5,0.8,5,1.5),Moment(1.5,4.5,2.5,3.5),Moment(1
1.5,8.5,1.5,0.5)))
```

GOIEW Distribution Negative Log-likelihood for Optimization (Complete data)

```
LL_GOIEW<-function(alpha,beta,gamma,theta){
  A<-theta*gamma*(x^(theta-1))*exp(-gamma*(x^theta))
```

```
 B<-1-exp(-gamma*(x^theta))

 C<-((beta-alpha)*(B^(beta-alpha-1)))+(alpha*(B^(-alpha-1)))

 PDF<-A*C*exp(-((1-B^beta)/B^alpha))

 LL<--sum(log(PDF))

 return(LL)

}
```

########################## GOIEW Distribution Optimization (Complete case) ##########

Require (bbmle) ###### calling R package bbmle ########

Fit<-mle2 (GOIEW_LL, start=list (alpha=alpha, beta=beta, gamma=gamma, theta=theta), method "BFGS", data=list(x)

Summary (fit) ########## summary of results ####################

GOIEW Distribution Negative Log-likelihood for Optimization (Censored case)

```
LLL_ODIEW<-function(alpha,beta,gamma,theta){

 A<-theta*gamma*(x^(theta-1))*exp(-gamma*(x^theta))

 B<-1-exp(-gamma*(x^theta))

 C<-((beta-alpha)*(B^(beta-alpha-1)))+(alpha*(B^(-alpha-1)))

 D<-1-exp(-gamma*(x^theta))

 CDF<-exp(-((1-D^beta)/D^alpha))

 PDF<-A*C*exp(-((1-B^beta)/B^alpha))
```

```
  LL<--sum(r*log(PDF))-sum((1-r)*log(1-CDF))
  return(LL)
}
```

```
######################### GOIEW Distribution Optimization (Censored case) ##########
require (bbmle) ###### calling R package bbmle ########
fit<-mle2 (GOIEW_LL, start=list (alpha=alpha, beta=beta, gamma=gamma, theta=theta), method "BFGS", data=list(x,r))
Summary (fit) ########## summary of results ###################
```

```
############################## GOIEL Distribution PDF ##########################
PDF_GOIEL<-function(x,alpha,beta,gamma,theta){
  A<-theta*gamma*((1+gamma*x)^(-theta-1))
  B<-(beta-alpha)*((1-((1+gamma*x)^(-theta)))^(beta-alpha-1))
  C<-(alpha)*((1-((1+gamma*x)^(-theta)))^(-alpha-1))
  D<-1-((1-((1+gamma*x)^(-theta))))^beta
  E<-((1-((1+gamma*x)^(-theta))))^alpha
  PDF<-A*(B+C)*exp(-(D/E))
  return(PDF)
}
```

```
########################## GOIEL Distribution CDF ########################

CDF_GOIEL<-function(x,alpha,beta,gamma,theta){

 A<-(1-(1+gamma*x)^(-theta))

 CDF<-exp(-((1-A^beta)/A^alpha))

 return(CDF)

}


############################ GOIEL Distribution Hazard function ################

HAZARD_GOIEL<-function(x,alpha,beta,gamma,theta){

 A<-theta*gamma*((1+gamma*x)^(-theta-1))

 B<-(beta-alpha)*((1-((1+gamma*x)^(-theta)))^(beta-alpha-1))

 C<-(alpha)*((1-((1+gamma*x)^(-theta)))^(-alpha-1))

 D<-1-((1-((1+gamma*x)^(-theta))))^beta

 E<-((1-((1+gamma*x)^(-theta))))^alpha

 F<-A*(B+C)*exp(-(D/E))

 HAZARD<-F/(1-exp(-(D/E)))

 return(HAZARD)

}

################### GOIEL Distribution Survival function ###################

SURVIVAL_GOIEL<-function(x,alpha,beta,gamma,theta){

 A<-(1-(1+gamma*x)^(-theta))
```

```
  B<-exp(-((1-A^beta)/A^alpha))

  SURVIVAL<-1-B

  return(SURVIVAL)

}

############################## GOIEL Distribution Quantile function ###############

Qf<-function(alpha,beta,gamma,theta){

  output<-0

  u<-seq(0.1,0.9,0.1)

  for(i in 1:length(u)){

    f<-function(x){(((1-(1+gamma*x)^(-theta))^alpha)*log(u[i]))-((1-(1+gamma*x)^(-
theta))))^beta)+1}

    rc<-uniroot(f,lower=0,upper=10000,tol=1e-9)

    output[i]=rc$root

  }

  output

}

u<-seq(0.1,0.9,0.1)

cbind(u,Qf(0.8,0.1,0.3,0.9),Qf(1.2,1.0,3.0,0.1),Qf(3.0,8,0.4,0.6),Qf(14,27,32,16))
```

```
########################## GOIEL distribution moment ########################

Moment<-function(alpha,beta,gamma,theta){

 results<-0

 r<-seq(1,6,1)

 for(i in 1:length(r)){

  f<-function(x,alpha,beta,gamma,theta,r){(x^r[i])*(PDF_GOIEL(x,alpha,beta,gamma,theta))}

  results[i]<-
integrate(f,lower=0,upper=Inf,subdivisions=100000,alpha=alpha,beta=beta,gamma=gamma,th
eta=theta,r=r)$value

 }

 return(results)

}

r<-seq(1,6,1)

print(cbind(r,Moment(12.5,11.8,10.5,11.5),Moment(10.5,13.8,17.5,12.5),Moment(16.5,11.8,14.1,
13.5),Moment(14.7,12.1,18.2,10.5)))
```

```
#### GOIEL Distribution Negative Log-likelihood for Optimization (Complete data) #####

LL_GOIEL<-function(alpha, beta, gamma ,theta){

 A<-theta*gamma*((1+gamma*x)^(-theta-1))

 B<-(beta-alpha)*((1-((1+gamma*x)^(-theta)))^(beta-alpha-1))

 C<-(alpha)*((1-((1+gamma*x)^(-theta)))^(-alpha-1))

 D<-1-((1-((1+gamma*x)^(-theta))))^beta
```

```
  E<-((1-((1+gamma*x)^(-theta))))^alpha

  PDF<-A*(B+C)*exp(-(D/E))

  LL<--sum(log(PDF))

  return(LL)

}
```

############## GOIEL Distribution Optimization (Complete case) ######

Require (bbmle) ###### calling R package bbmle ########

Fit<-mle2 (GOIEL_LL, start=list (alpha=alpha, beta=beta, gamma=gamma, theta=theta), method "BFGS", data=list(x)

Summary (fit) ########## summary of results ####################

GOIEL Distribution Negative Log-likelihood for Optimization (Censored case)

```
LLL_GOIEL<-function(alpha, beta, gamma, theta){

  A<-theta*gamma*((1+gamma*x)^(-theta-1))

  B<-(beta-alpha)*((1-((1+gamma*x)^(-theta)))^(beta-alpha-1))

  C<-(alpha)*((1-((1+gamma*x)^(-theta)))^(-alpha-1))

  D<-1-((1-((1+gamma*x)^(-theta))))^beta

  E<-((1-((1+gamma*x)^(-theta))))^alpha

  H<-(1-(1+gamma*x)^(-theta))

  CDF<-exp(-((1-H^beta)/H^alpha))
```

PDF<-A(B+C)*exp(-(D/E))*

*LL<--sum(r*log(PDF))-sum((1-r)*log(1-CDF))*

return(LL)

############## GOIEL Distribution Optimization (Censored case) ######

Require (bbmle) ###### calling R package bbmle ########

Fit<-mle2 (GOIEL_LL, start=list (alpha=alpha, beta=beta, gamma=gamma, theta=theta), method "BFGS", data=list(x,r))

Summary (fit) ########## summary of results ####################

Algorithm for Monte Carlo Simulation study of MLE for GOIEW Distribution

quantile<-function(alpha,beta,gamma,theta,u){

*f<-function(x){(((1-(1+gamma*x)^(-theta))^alpha)*log(u))-((1-(1+gamma*x)^(-theta))^beta)+1}*

rc<-uniroot(f,lower=0,upper=10000,tol=1e-09)

result=rc$root

return(result)

}

################

P_LL<-function(par){

alpha=par[1]

```
 beta=par[2]
 gamma=par[3]
 theta=par[4]
 A<-theta*gamma*((1+gamma*x)^(-theta-1))
 B<-(beta-alpha)*((1-((1+gamma*x)^(-theta)))^(beta-alpha-1))
 C<-(alpha)*((1-((1+gamma*x)^(-theta)))^(-alpha-1))
 D<-1-((1-((1+gamma*x)^(-theta))))^beta
 E<-((1-((1+gamma*x)^(-theta))))^alpha
 PDF<-A*(B+C)*exp(-(D/E))
 LL<--sum(log(PDF))
 return(LL)
}
###Algorithm for Monte Carlo Simulation Study
library(numDeriv)
library(Matrix)
alpha=0.3
beta=2.2
gamma=0.4
theta=0.8
n1=c(30,50, 80,120,200,250)
for(j in 1:length(n1)){
```

```
n=n1[j]

N=50

mle_alpha<-c(rep(0,N))

mle_beta<-c(rep(0,N))

mle_gamma<-c(rep(0,N))

mle_theta<-c(rep(0,N))

LC_alpha<-c(rep(0,N))

UC_alpha<-c(rep(0,N))

LC_beta<-c(rep(0,N))

UC_beta<-c(rep(0,N))

LC_gamma<-c(rep(0,N))

UC_gamma<-c(rep(0,N))

LC_theta<-c(rep(0,N))

UC_theta<-c(rep(0,N))

count_alpha=0

count_beta=0

count_gamma=0

count_theta=0

temp=1

HH1<-matrix(c(rep(2,16)),nrow=4,ncol=4)

HH2<-matrix(c(rep(2,16)),nrow=4,ncol=4)
```

```
for(i in 1:N)
{
 print(i)
 flush.console()
 repeat{
  x<-c(rep(0,n))
  # Generate a random variable from uniform distribution
  u<-0
  u<-runif(n,min=0,max=1)
  for(k in 1:n){
   x[k]<-quantile(alpha,beta,gamma,theta,u[k])
  }
  #Maximum likelihood estimation
  mle.result<-nlminb(c(alpha,beta,gamma,theta),P_LL,lower=c(0,0,0,0),upper=c(1,1,1,1))
  temp=mle.result$convergence
  if(temp==0){
   temp_alpha<-mle.result$par[1]
   temp_beta<-mle.result$par[2]
   temp_gamma<-mle.result$par[3]
   temp_theta<-mle.result$par[4]
   HH1<-hessian(P_LL, c(temp_alpha,temp_beta,temp_gamma,temp_theta))
```

```
if(sum(is.nan(HH1))==0&(diag(HH1)[1]>0)&(diag(HH1)[2]>0)&(diag(HH1)[3]>0)&(diag(H
H1)[4]>0)){
        HH2<-solve(HH1)
        #print(det(HH1))
      }
      else{
        temp=1}
    }

if((temp==0)&(diag(HH2)[1]>0)&(diag(HH2)[2]>0)&(diag(HH2)[3]>0)&(diag(HH2)[4]>0)
&(sum(is.nan(HH2))==0)){
      break
    }
    else{
      temp=1}
  }
  temp=1
  mle_alpha[i]<-mle.result$par[1]
  mle_beta[i]<-mle.result$par[2]
  mle_gamma[i]<-mle.result$par[3]
  mle_theta[i]<-mle.result$par[4]
```

```
HH<-hessian(P_LL,c(mle_alpha[i],mle_beta[i],mle_gamma[i],mle_theta[i]))
H<-solve(HH)
LC_alpha[i]<-mle_alpha[i]-qnorm(0.975)*sqrt(diag(H)[1])
UC_alpha[i]<-mle_alpha[i]+qnorm(0.975)*sqrt(diag(H)[1])
if((LC_alpha[i]<=alpha)&(alpha<=UC_alpha[i])){
 count_alpha=count_alpha+1
}
LC_beta[i]<-mle_beta[i]-qnorm(0.975)*sqrt(diag(H)[2])
UC_beta[i]<-mle_beta[i]+qnorm(0.975)*sqrt(diag(H)[2])
if((LC_beta[i]<=beta)&(beta<=UC_beta[i])){
 count_beta=count_beta+1
}
LC_gamma[i]<-mle_gamma[i]-qnorm(0.975)*sqrt(diag(H)[3])
UC_gamma[i]<-mle_gamma[i]+qnorm(0.975)*sqrt(diag(H)[3])
if((LC_gamma[i]<=gamma)&(gamma<=UC_gamma[i])){
 count_gamma=count_gamma+1
}
LC_theta[i]<-mle_theta[i]-qnorm(0.975)*sqrt(diag(H)[4])
UC_theta[i]<-mle_theta[i]+qnorm(0.975)*sqrt(diag(H)[4])
if((LC_theta[i]<=theta)&(theta<=UC_theta[i])){
 count_theta=count_theta+1
```


```
  }


}
#Calculate Average Bias
ABias_alpha<-sum(abs(mle_alpha-alpha))/N
ABias_beta<-sum(abs(mle_beta-beta))/N
ABias_gamma<-sum(abs(mle_gamma-gamma))/N
ABias_theta<-sum(abs(mle_theta-theta))/N
print(cbind(ABias_alpha,ABias_beta,ABias_gamma,ABias_theta))
#Calculate MSE
MSE_alpha<-sum((alpha-mle_alpha)^2)/N
MSE_beta<-sum((beta-mle_beta)^2)/N
MSE_gamma<-sum((gamma-mle_gamma)^2)/N
MSE_theta<-sum((theta-mle_theta)^2)/N
print(cbind(MSE_alpha,MSE_beta,MSE_gamma,MSE_theta))
#Average Estimate
AValpha<-sum(mle_alpha)/N
AVbeta<-sum(mle_beta)/N
AVgamma<-sum(mle_gamma)/N
AVtheta<-sum(mle_theta)/N
print(cbind(AValpha,AVbeta,AVgamma,AVtheta))
```


```
}
```

Algorithm for Monte Carlo Simulation study of OLS for GOIEW Distribution

```
quantile<-function(alpha,beta,gamma,theta,u){
  f<-function(x){(((1-exp(-gamma*(x^theta)))^alpha)*log(u))-((1-exp(-
gamma*(x^theta)))^beta)+1}
  rc<-uniroot(f,lower=0,upper=10000,tol=1e-9)
  result=rc$root
  return(result)
}
################
P_LL<-function(par){
  alpha=par[1]
  beta=par[2]
  gamma=par[3]
  theta=par[4]
  y<-sort(x)
  n<-length(y)
  p<-(1:n)/(n+1)
  A<-1-exp(-gamma*(x^theta))
  CDF<-exp(-((1-A^beta)/A^alpha))
```

```
 LSE<-sum((CDF-p)^2)
 return(LSE)
}
###Algorithm for Monte Carlo Simulation Study
library(numDeriv)
library(Matrix)
alpha=0.4
beta=0.7
gamma=2.6
theta=0.4
n1=c(30,50, 80,120,200,250)
for(j in 1:length(n1)){
 n=n1[j]
 N=100
 mle_alpha<-c(rep(0,N))
 mle_beta<-c(rep(0,N))
 mle_gamma<-c(rep(0,N))
 mle_theta<-c(rep(0,N))
 LC_alpha<-c(rep(0,N))
 UC_alpha<-c(rep(0,N))
 LC_beta<-c(rep(0,N))
```

```
UC_beta<-c(rep(0,N))

LC_gamma<-c(rep(0,N))

UC_gamma<-c(rep(0,N))

LC_theta<-c(rep(0,N))

UC_theta<-c(rep(0,N))

count_alpha=0

count_beta=0

count_gamma=0

count_theta=0

temp=1

HH1<-matrix(c(rep(2,16)),nrow=4,ncol=4)

HH2<-matrix(c(rep(2,16)),nrow=4,ncol=4)

for(i in 1:N)

{

 print(i)

 flush.console()

 repeat{

  x<-c(rep(0,n))

  # Generate a random variable from uniform distribution

  u<-0

  u<-runif(n,min=0,max=1)
```


```
for(k in 1:n){
  x[k]<-quantile(alpha,beta,gamma,theta,u[k])
}
#Maximum likelihood estimation
mle.result<-nlminb(c(alpha,beta,gamma,theta),P_LL,lower=c(0,0,0,0),upper=c(1,1,1,1))
temp=mle.result$convergence
if(temp==0){
  temp_alpha<-mle.result$par[1]
  temp_beta<-mle.result$par[2]
  temp_gamma<-mle.result$par[3]
  temp_theta<-mle.result$par[4]
  HH1<-hessian(P_LL, c(temp_alpha,temp_beta,temp_gamma,temp_theta))

if(sum(is.nan(HH1))==0&(diag(HH1)[1]>0)&(diag(HH1)[2]>0)&(diag(HH1)[3]>0)&(diag(H
H1)[4]>0)){
    HH2<-solve(HH1)
    #print(det(HH1))
  }
  else{
    temp=1}
}
```


```
if((temp==0)&(diag(HH2)[1]>0)&(diag(HH2)[2]>0)&(diag(HH2)[3]>0)&(diag(HH2)[4]>0)
&(sum(is.nan(HH2))==0)){
    break
  }
  else{
    temp=1}
 }
 temp=1
 mle_alpha[i]<-mle.result$par[1]
 mle_beta[i]<-mle.result$par[2]
 mle_gamma[i]<-mle.result$par[3]
 mle_theta[i]<-mle.result$par[4]
 HH<-hessian(P_LL,c(mle_alpha[i],mle_beta[i],mle_gamma[i],mle_theta[i]))
 H<-solve(HH)
 LC_alpha[i]<-mle_alpha[i]-qnorm(0.975)*sqrt(diag(H)[1])
 UC_alpha[i]<-mle_alpha[i]+qnorm(0.975)*sqrt(diag(H)[1])
 if((LC_alpha[i]<=alpha)&(alpha<=UC_alpha[i])){
  count_alpha=count_alpha+1
 }
 LC_beta[i]<-mle_beta[i]-qnorm(0.975)*sqrt(diag(H)[2])
 UC_beta[i]<-mle_beta[i]+qnorm(0.975)*sqrt(diag(H)[2])
```

```
 if((LC_beta[i]<=beta)&(beta<=UC_beta[i])){
  count_beta=count_beta+1
 }
 LC_gamma[i]<-mle_gamma[i]-qnorm(0.975)*sqrt(diag(H)[3])
 UC_gamma[i]<-mle_gamma[i]+qnorm(0.975)*sqrt(diag(H)[3])
 if((LC_gamma[i]<=gamma)&(gamma<=UC_gamma[i])){
  count_gamma=count_gamma+1
 }
 LC_theta[i]<-mle_theta[i]-qnorm(0.975)*sqrt(diag(H)[4])
 UC_theta[i]<-mle_theta[i]+qnorm(0.975)*sqrt(diag(H)[4])
 if((LC_theta[i]<=theta)&(theta<=UC_theta[i])){
  count_theta=count_theta+1
 }

}
#Calculate Average Bias
ABias_alpha<-sum(abs(mle_alpha-alpha))/N
ABias_beta<-sum(abs(mle_beta-beta))/N
ABias_gamma<-sum(abs(mle_gamma-gamma))/N
ABias_theta<-sum(abs(mle_theta-theta))/N
print(cbind(ABias_alpha,ABias_beta,ABias_gamma,ABias_theta))
```

```
 #Calculate MSE
 MSE_alpha<-sum((alpha-mle_alpha)^2)/N
 MSE_beta<-sum((beta-mle_beta)^2)/N
 MSE_gamma<-sum((gamma-mle_gamma)^2)/N
 MSE_theta<-sum((theta-mle_theta)^2)/N
 print(cbind(MSE_alpha,MSE_beta,MSE_gamma,MSE_theta))
 #Average Estimate
 AValpha<-sum(mle_alpha)/N
 AVbeta<-sum(mle_beta)/N
 AVgamma<-sum(mle_gamma)/N
 AVtheta<-sum(mle_theta)/N
 print(cbind(AValpha,AVbeta,AVgamma,AVtheta))
}
```


Algorithm for Monte Carlo Simulation study of CVM for GOIEW Distribution

```
quantile<-function(alpha,beta,gamma,theta,u){
 f<-function(x){(((1-exp(-gamma*(x^theta)))^alpha)*log(u))-((1-exp(-
gamma*(x^theta)))^beta)+1}
 rc<-uniroot(f,lower=0,upper=10000,tol=1e-9)
 result=rc$root
 return(result)
```

```
}

###############

P_LL<-function(par){

 alpha=par[1]

 beta=par[2]

 gamma=par[3]

 theta=par[4]

 y<-sort(x)

 n<-length(y)

 i<-(1:n)

 p<-((2*i)-1)/(2*n)

 A<-1-exp(-gamma*(x^theta))

 CDF<-exp(-((1-A^beta)/A^alpha))

 CVM<-(1/(12*n))+sum((CDF-p)^2)

 return(CVM)

}

###Algorithm for Monte Carlo Simulation Study

library(numDeriv)

library(Matrix)

alpha=0.4

beta=0.7
```


```
gamma=2.6

theta=0.4

n1=c(30,50, 80,120,200,250)

for(j in 1:length(n1)){

 n=n1[j]

 N=100

 mle_alpha<-c(rep(0,N))

 mle_beta<-c(rep(0,N))

 mle_gamma<-c(rep(0,N))

 mle_theta<-c(rep(0,N))

 LC_alpha<-c(rep(0,N))

 UC_alpha<-c(rep(0,N))

 LC_beta<-c(rep(0,N))

 UC_beta<-c(rep(0,N))

 LC_gamma<-c(rep(0,N))

 UC_gamma<-c(rep(0,N))

 LC_theta<-c(rep(0,N))

 UC_theta<-c(rep(0,N))

 count_alpha=0

 count_beta=0

 count_gamma=0
```


```
count_theta=0
temp=1
HH1<-matrix(c(rep(2,16)),nrow=4,ncol=4)
HH2<-matrix(c(rep(2,16)),nrow=4,ncol=4)
for(i in 1:N)
{
 print(i)
 flush.console()
 repeat{
  x<-c(rep(0,n))
  # Generate a random variable from uniform distribution
  u<-0
  u<-runif(n,min=0,max=1)
  for(k in 1:n){
   x[k]<-quantile(alpha,beta,gamma,theta,u[k])
  }
  #Maximum likelihood estimation
  mle.result<-nlminb(c(alpha,beta,gamma,theta),P_LL,lower=c(0,0,0,0),upper=c(1,1,1,1))
  temp=mle.result$convergence
  if(temp==0){
   temp_alpha<-mle.result$par[1]
```

```
    temp_beta<-mle.result$par[2]

    temp_gamma<-mle.result$par[3]

    temp_theta<-mle.result$par[4]

    HH1<-hessian(P_LL, c(temp_alpha,temp_beta,temp_gamma,temp_theta))


if(sum(is.nan(HH1))==0&(diag(HH1)[1]>0)&(diag(HH1)[2]>0)&(diag(HH1)[3]>0)&(diag(H
H1)[4]>0)){

     HH2<-solve(HH1)

     #print(det(HH1))

    }

    else{

     temp=1}

  }


if((temp==0)&(diag(HH2)[1]>0)&(diag(HH2)[2]>0)&(diag(HH2)[3]>0)&(diag(HH2)[4]>0)
&(sum(is.nan(HH2))==0)){

    break

  }

 else{

    temp=1}

}

temp=1
```

```
mle_alpha[i]<-mle.result$par[1]

mle_beta[i]<-mle.result$par[2]

mle_gamma[i]<-mle.result$par[3]

mle_theta[i]<-mle.result$par[4]

HH<-hessian(P_LL,c(mle_alpha[i],mle_beta[i],mle_gamma[i],mle_theta[i]))

H<-solve(HH)

LC_alpha[i]<-mle_alpha[i]-qnorm(0.975)*sqrt(diag(H)[1])

UC_alpha[i]<-mle_alpha[i]+qnorm(0.975)*sqrt(diag(H)[1])

if((LC_alpha[i]<=alpha)&(alpha<=UC_alpha[i])){

 count_alpha=count_alpha+1

}

LC_beta[i]<-mle_beta[i]-qnorm(0.975)*sqrt(diag(H)[2])

UC_beta[i]<-mle_beta[i]+qnorm(0.975)*sqrt(diag(H)[2])

if((LC_beta[i]<=beta)&(beta<=UC_beta[i])){

 count_beta=count_beta+1

}

LC_gamma[i]<-mle_gamma[i]-qnorm(0.975)*sqrt(diag(H)[3])

UC_gamma[i]<-mle_gamma[i]+qnorm(0.975)*sqrt(diag(H)[3])

if((LC_gamma[i]<=gamma)&(gamma<=UC_gamma[i])){

 count_gamma=count_gamma+1

}
```

```
  LC_theta[i]<-mle_theta[i]-qnorm(0.975)*sqrt(diag(H)[4])
  UC_theta[i]<-mle_theta[i]+qnorm(0.975)*sqrt(diag(H)[4])
  if((LC_theta[i]<=theta)&(theta<=UC_theta[i])){
    count_theta=count_theta+1
  }

}
#Calculate Average Bias
ABias_alpha<-sum(abs(mle_alpha-alpha))/N
ABias_beta<-sum(abs(mle_beta-beta))/N
ABias_gamma<-sum(abs(mle_gamma-gamma))/N
ABias_theta<-sum(abs(mle_theta-theta))/N
print(cbind(ABias_alpha,ABias_beta,ABias_gamma,ABias_theta))
#Calculate MSE
MSE_alpha<-sum((alpha-mle_alpha)^2)/N
MSE_beta<-sum((beta-mle_beta)^2)/N
MSE_gamma<-sum((gamma-mle_gamma)^2)/N
MSE_theta<-sum((theta-mle_theta)^2)/N
print(cbind(MSE_alpha,MSE_beta,MSE_gamma,MSE_theta))
#Average Estimate
AValpha<-sum(mle_alpha)/N
```


```
 AVbeta<-sum(mle_beta)/N

 AVgamma<-sum(mle_gamma)/N

 AVtheta<-sum(mle_theta)/N

 print(cbind(AValpha,AVbeta,AVgamma,AVtheta))

}
```