UNIVERSITI PUTRA MALAYSIA

PARAMETER ESTIMATION OF KUMARASWAMY BURR TYPE X MODELS BASED ON CURE MODELS WITH OR WITHOUT COVARIATES

MADAKI UMAR YUSUF

FS 2017 55

PARAMETER ESTIMATION OF KUMARASWAMY BURR TYPE X MODELS BASED ON CURE MODELS WITH OR WITHOUT COVARIATES

By

MADAKI UMAR YUSUF

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

June 2017

COPYRIGHT

DEDICATIONS

The entire work is dedicated to my father Alh. Yusuf Madaki and my mother Haj. Maryam Umar Kakke Y. Madaki and also the entire Madaki's family for giving me their support and prayers.

I also dedicated this work to my paternal grand father late Alh. Abubakar Dawi Madaki and my paternal grand mother late Haj. Hamsatu Abubakar Madaki and also the my maternal grand father and (namesake) late Alh. Umaru Kakke and my maternal grand mother late Haj. Maimuna Kakke and lastly my nursing grand mother to the enire Madaki's family at large, we are indempted to her as our special mother late Haj. Fatima Kakke (Baabbaa) may their gentle soul rest in Jannatul Firdausi, Ameen ya rabbil Alamin.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

PARAMETER ESTIMATION OF KUMARASWAMY BURR TYPE X MODELS BASED ON CURE MODELS WITH OR WITHOUT COVARIATES

By

MADAKI UMAR YUSUF

June 2017

Chairman : **Mohd Rizam Abu Bakar, PhD**
Faculty : **Science**

In the last few years, many attempts have been made to define new models that extend well known Beta Kumaraswamy-G (BK-G) and Kumaraswamy-G (K-G) families of distribution that can provide a greater flexibility in modeling real-life data. Furthermore, one of the weakness of Beta distribution is that it is not fairly tractable and in a particular case, its cumulative distribution function (CDF) involves the incomplete beta function ratio.

Kumaraswamy distribution has a closed form of probability density function (PDF) and CDF, which makes it tractable. This motivated us to extend the BK-G family which has four shape parameters and K-G family which has two shape parameters. The Burr Type X (BX) distribution was chosen because of its PDF and CDF are of a closed form. As a consequence of this, it can be used suitably for censored data.

Based on the problem stated, we develop a new model using the method of confounding the existing parametric models by adopting the BX model as the baseline distribution. This proposed model is called Beta Kumaraswamy Burr-Type X (BKBX) distribution with six parameters. Due to the intricacy and non-close form solution of the BKBX model, we provide the modified better version of the model by reducing its parameters to four and called this as Kumaraswamy Burr-Type X (KBX) distribution.

In this thesis, we considered two methods via the classical maximum likelihood estimation (MLE) and the Bayes estimation using the Gibbs sampling (G-S) algorithm to estimate the parameters of BKBX, KBX and Beta-Weibull (BWB) models. We obtained the posterior summaries considering the cure models with covariates by the method of Gibbs sampling of the Markov Chain Monte Carlo (MCMC). Series of simulation

studies were conducted to evaluate the performance of the proposed estimation approaches. The two common types of cure fraction models, namely; mixture and the non-mixture models for the survival data based on the BKBX, KBX and BWB distributions were provided.

Hence, to obtain effective results for the cure models with censored data and covariates, the estimation of the parameters was done under a Bayesian approach using G-S method. The comparison was done between the BKBX, KBX and BWB models to validate the usefulness of the modified distributions. The KBX signifies and proves to be less time consuming, have a close form solution of both its survival and hazard function unlike BKBX and BWB models and yet have similar features as the Kumaraswamy Weibull (KWB) distribution. Based on the results of the cure models, with or without covariates for the censored dataset used at all levels of comparison, the KBX model ok to be the best choice.

The application of real datasets which are uncensored and not cure model of right skewed, left skewed and approximate symmetry were considered. Based on the results obtained, the KBX distribution has provided a better fit compared to the BKBX, the baseline BX, and non-nested models based on the model selection criteria using the MLE.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENGANGGARAN BERPARAMETER MODEL KUMARASWAMY BURR JENIS X BERDASARKAN MODEL SEMBUH DENGAN ATAU TANPA KOVARIAT

Oleh

MADAKI UMAR YUSUF

Jun 2017

Pengerusi : **Mohd Rizam Abu Bakar, PhD**
Fakulti : **Sains**

Dalam beberapa tahun kebelakang ini, terdapat banyak percubaan dibuat untuk membentuk model baharu dengan meluaskan taburan famili terkenal Beta Kumaraswamy-G (BK-G) dan Kumaraswamy-G (K-G) yang mampu membuat pemodelan lebih fleksibal terhadap data sebenar. Selain itu, salah satu kelemahan taburan Beta adalah ia tidak mudah dikawal dan dalam kes tertentu fungsi taburan kumulatif (CDF) nya melibatkan fungsi nisbah Beta tak lengkap.

Taburan Kumaraswamy memiliki bentuk fungsi kepadatan kebarangkalian (PDF) dan CDF yang tertutup dan mudah dikawal. Ini mendorang kami untuk meluaskan famili BK-G yang mempunyai empat parameter bentuk dan famili K-G yang mempunyai dua parameter bentuk. Taburan Burr jenis X (BX) dipilih kerana PDF dan CDF nya adalah dalam bentuk tertutup. Susulan itu, ia sesuai untuk digunakan pada data tertapis.

Berdasarkan masalah yang dinyatakan, kami membina model baharu menggunakan kaedah pembauran terhadap model berparameter sedia ada dengan mengambil model BX sebagai taburan dasar. Model yang dicadangkan ini dinamakan dengan taburan Beta Kumaraswamy Burr jenis X (BKBX) dengan enam parameter. Oleh kerana penyelesaian terhadap model BKBX ini rumit dan berbentuk tak tertutup, kami menghasilkan model terubahsuai yang lebih baik dengan mengurangkan parameter kepada empat dan memanggil model ini dengan taburan Kumaraswamy Burr jenis X (KBX).

Dalam tesis ini, kami mempertimbangkan dua kaedah melalui penganggaran kebolehjadian maksimum klasik (MLE) dan penganggaran Bayes yang menggunakan algoritma persampelan Gibbs (G-S) untuk menganggar parameter model BKBX, KBX

dan Beta-Weibull (BWB). Kami memperoleh ringkasan posterior berdasarkan anggapan model sembuh dengan kovariat menggunakan kaedah pensampelan Gibbs rantai Markov Monte Carlo. Siri kajian simulasi dijalankan untuk menilai prestasi pendekatan penganggaran yang disaran. Dua jenis model pecahan sembuh yang lazim iaitu model campuran dan tak campuran bagi data mandirian berdasarkan taburan BKBX, KBX dan BWB disediakan.

Oleh yang demikian, untuk mendapatkan hasil yang berkesan untuk model sembuh dengan data tertapis dan kovariat, penganggaran parameter telah dilakukan melalui pendekatan Bayes menggunakan kaedah G-S. Perbandingan telah dilakukan antara BKBX, KBX dan BWB untuk mengesahkan kebergunaan taburan terubahsuai. KBX menunjukkan dan membuktikan penggunaan masa yang rendah, memiliki bentuk penyelesaian tertutup bagi kedua-dua fungsi mandirian dan bahaya tidak seperti model BKBX dan BWB dan malah memiliki ciri yang hampir sama dengan taburan Kumaraswamy Weibull (KWB). Berdasarkan kepada keputusan model sembuh, sama ada dengan kovariat atau tidak untuk data tertapis pada semua tahap perbandingan, model KBX terbukti satu pilihan terbaik.

Penggunaan data sebenar yang tidak tertapis dan bukan model sembuh yang pencong kanan, pencong kiri dan hampir simetri telah dilakukan. Berdasar kepada keputusan yang dicapai, ternyata taburan KBX lebih sesuai berbanding dengan BKBX, asas BX dan model tak tersarang pada kriteria pemilihan model menggunakan MLE.

ACKNOWLEDGEMENTS

All praise be to Allah the lord of the world, the Beneficent, the Merciful who in his infinite mercy gives me life, good health, strength, hope, guidance and perseverance to pursue this program to the successful completion. May Allah's Mercy and Peace be upon our noble prophet Muhammad Rasulillah Sallallahu Alaihi Wasallam, his family and companions.

I would like to express my profound gratitude to my supervisor an the chairman supervisory committee, in person of Assoc. Prof. Haji. Dr. Mohd Rizam Abu Bakar for his excellent supervision, criticism and fortitude throughout the period of my studies and his kindness would forever not going to be forgotten.

My sincere appreciation also goes to my members of the supervisory committee, Prof Dr. Noor Akma Ibrahim and Assoc. Prof. Dr. Jayanthi Arasan for their supportive advices and guidance. I would also like to express my thanks to the entire staff of Mathematics Department (Jabatan Matematik) and Institute for Mathematical Research (Institut Penyelidikan Matematik) for their help and support.

I remained grateful to my entire family for their support and continuous for their prayers throughout my studies here in Universiti Putra Malaysia (UPM). My brothers: (Ya Aliyu, Ya Abubakar, Ibrahim (Sultan), Abubakar Sadiq and Usman), my sisters: (Ya Hamsatu (Uwani), Ya Hadiza, Ya Fatima, Hadiza (Gona), Aisha (Meme) and Fatima (Yasmin)), my lovely wife: (Amina Bukar Modu Jumbam) for her patience and encouragement and my dearest sons: (Yusuf Farhan Umar Madaki and Umar Farooq Umar Madaki), my best friends: (Buba Siddi (G-Side), Musa I. Dikko (Biggie), Musa A. Dori (Lil-Moore), Sheriff W. Geidam (Sharpson), Hussaini L. Mashio, Abubakar A. Nuhu (Emir), Mohd Y. Yunusa (Doc.), Abubakar K. Bizi (Babsy) and Hassan Saidu), my colleagues: (Babagana Modu, Musa Chiwa Dalah, Dr. Sani Isa, Prof. Vijay Vir Singh, Dr. Mundher A. Khaleel, Chris B. Guure, Dr. Firas Adel, Dr. Mohd D. Goni, Alhassan Abare, Dr. Mohd I. Jalo, Umar M. Mustapha, Ali I. Abbas, Suleiman Dahiru, Zarah A. Musa, Abdurrahman Idris, Abubakar U. Idris, Ali G. Kolo, Ibrahim A. Alamin, Abubakar S. Tanko, Ibrahim Arzuka, Yusuf D. Jikan-Toro, Usman Waziri, Ahmed I. Karage, Dr. L.B. Mohammed, Dr. Ibrahim M. Idris, Babangida I. Babura, Buba A. Tarajo, Dr. Sadiq Shehu, Saidu Abubakar, Idris Ajoge, Auwal Chul, Umar S. Aliyu Tilde, Mohammed Sani and Ishaq A. Baba etc), and also the prayers given to me by all my entire staffs/students and well-wishers around the globe that helped directly or indirectly to the successful completion of my PhD programme.

Lastly, I really appreciate the understanding, support and sponsorship by the Tertiary Education Trust Fund (TETFund) through my place of work Yobe State University (YSU) and Yobe State Government (YSG) in Nigeria.

I certify that a Thesis Examination Committee has met on 6 June 2017 to conduct the final examination of Madaki Umar Yusuf on his thesis entitled "Parameter Estimation of Kumaraswamy Burr Type X Models Based on Cure Models with or Without Covariates" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Zarina Bibi binti Ibrahim, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Habshah binti Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Mohd Bakri bin Adam, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Olimjon Shukurovich Sharipov, PhD
Professor
Academy of Sciences of Uzbekistan
Uzbekistan
(External Examiner)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 28 September 2017

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy.
The members of the Supervisory Committee were as follows:

Mohd Rizam Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Noor Akma Ibrahim, PhD
Professor
Institute of Mathematical Research
Universiti Putra Malaysia
(Member)

Jayanthi Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature:_____ Date:_____

Name and Matric No: Madaki Umar Yusuf, GS41688

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of
Chairman of
Supervisory Committee: Associate Professor Dr. Mohd Rizam Abu Bakar

Signature: _____
Name of
Member of
Supervisory Committee: Prof. Dr. Noor Akma Ibrahim

Signature: _____
Name of
Member of
Supervisory Committee: Associate Professor Dr. Jayanthi Arasan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACTG	AIDS clinical trials group study data
BKBX	Beta Kumaraswamy Burr Type X Distribution
KBX	Kumaraswamy Burr Type X Distribution
BBX	Beta Burr Type X Distribution
B	Beta Distribution
BX	Burr Type X Distribution
BX1	Burr Type X with One Shape Parameter Distribution
BE	Beta Exponential Distribution
BWB	Beta Weibull
BR	Beta Rayleigh Distribution
BKR	Beta Kumaraswamy Rayleigh Distribution
B-G	Beta-G Family of Distribution
K-G	Kumarawamy-G Family of Distribution
BK-G	Beta Kumaraswamy-G Family of Distribution
E	Exponential Distribution
GE	Generalized Exponential Distribution
G	Gompertz Distribution
GG	Generalized Gompertz Distribution
ER	Exponential Rayleigh Distribution
GR	Generalized Rayleigh Distribution
AIC	Akaike Information Criterion
CAIC	Corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
K-SM	Kolmogorov-Smirnov Test
K	Kumaraswamy Distribution
KWB	Kumaraswamy Weibull Distribution
KR	Kumaraswamy Rayleigh Distribution
HW	Heidelberger and Welch
PDF	Probability Density Function
CDF	Cumulative Distribution Function
SF	Survival Function
HPD	Highest Posterior Density
JPD	Joint Probability Distribution
LPML	Log Pseudo Marginal Likelihood
MLE	Maximum Likelihood Estimation
MCMC	Markov Chain Monte Carlo
MSE	Mean Square Error
MGF	Moment Generating Function
PH	Proportional Hazard
PWM	Probability Weighted r^{th} Moments
M-H	Metropolis-Hasting Algorithm
G-S	Gibbs Sampling Algorithm
LR	Likelihood Ratio Test

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

The general conceptual ideology of probability distribution theory of a newly developed models, by confounding and modifying existing distributions which is also apply by the method of Bayesian survival data analysis by estimating the unknown parameters using iterative algorithm which refers to as an approach which simply define to be a process that corresponds to the time to event of a study, meaning a time assigned for a period until the occurrence of an event of interest, the end or terminal points.

It is very important to decide what actually the event and how far or long it will take the follow-up period would be in the case of what we called censoring study event. In a cancer research base on clinical trials, the time origin mostly maintained with the time of patients into a trial while on the other hand the endpoint is the death time of patient which experience different event or simply end of the trial.

A time to event data generally violates the assumptions of normality which makes them rarely follow the normal (Gaussian) distribution and also on the other hand they are skewed in nature mostly follows the assumption of Weibull and Burr Type X distributions. It is very uncommon to include many conjugate prior distributions for an event in the study to obtain the posterior distribution.

Base on the above issues that its generally makes it necessary to the particular methods of survival analysis. The major challenges that comes into consideration regarding the time to event data analysis, normally arises base on the fact that the lifetimes of the target group subset will always be an unknown after follow-up and only some few (susceptible) patients may experience the event. It is define in survival data analysis as an incomplete observation or censoring, which varies in many ways as mention below:

1. When a patient does not experienced the event (death) or is terminated.
2. An insufficient or irregular follow-up of a patient between the period of trial, and
3. A patient or individual experiences of some other different study event other than event of interest.

Frequently, the period of survival time for a given group of cancer patients which can be seen as a time line where the event is exempted, at least in a theoretical aspect to terminate the follow-up time. This is an example of what we called right censoring which is the most common among all the types of censoring in the survival data analysis.

Alternatively, the second type of censoring which is the interval censored data which consist the general case of the left and right censoring which are the special cases of the interval censored data. These type of censoring can be generated when the hazard or failure time for the observed patients data that includes the trial is not the same or exactly known but it lies between an interval of two successive assigned time.

Generally, in the presence of censored data to make it necessary for analyzing the survival data using the statistical techniques for the data exploration cannot be used. In many trials related to cancer studies, the environment and cohort of the individuals which are included in the study are called "heterogeneous cohorts " which are divided into two groups, the first group constitute of individual patients who become those that will not experience the event (insusceptible ones) to the cancer disease and included to be cured.

On the other hand, the second group which contains the individuals that do not respond to treatments while on the follow-up time period and also remain for their lifetime as uncured. One of the major interest of the clinical trials in estimating the cured patients proportion is more importantly by determining the survival cancer trends patients of the cohort. Moreover, the models that normally takes the cure fraction is accounted and widely used in survival modeling.

The most familiar and attentive models are the mixture otherwise known as (standard parametric cure) and the bounded cumulative hazard (BCH) models. In recent years these models received more attention in the literature. This model mostly proposed to be a class or group of a cancer patients that attend a clinical trials and at the initial stage process of a follow-up to treatment, while some cancer patients still be left uncured.

These cancer cells are called clonogens, which grows rapidly and replaces the normal tissue in human body tissue. Moreover, the BCH model base on the number of the cancer cells assumes that it follows a Poisson model. The estimation base on the cure fraction model generally depends on estimating the survival function of the model. Its gives clear view and clue to the probability of the time-to-event of interest.

The major approach used for estimating the hazard and survival functions, also the parametric and non-parametric models. The parametric method of estimation involves models like: beta, exponential, gamma and Weibull distributions etc, are represented as the distribution function of the survival data analysis.

Additionally, in this process sometimes it attracts more incorporation to some covariates in the function to which leads to knowledge gain about a particular prognostic factors that might probably affect the decision of a cure probability base on the assumption of parametric models.

On the other hand, we have the non-parametric approach which involves some methods that does not follow any existing distributional function but rather where a parametric estimation might not be suitable it applies to that by simply violating the assumption of normality and other skewed models like: Weibull and Burr Type X distributions. This approach simply leans on by estimating the survival and hazard functions by the means of some wide range of usage estimators, depends on the censoring type.

In the presence or case with censored data where an information is partially completed, its generally one of the primary role and most basic function of survival data analysis which does not work without censoring. For example, in the case of right censoring the most commonly used method for estimating the functions is called Kaplan-Meier estimator or a product limit estimator is the limit of the life-table estimator when intervals are taken so small that only at most one distinct observation occurs within an interval.

Kaplan-Meier estimator is a maximum likelihood estimate, while in the interval censoring case, the Turn-bull estimator is very common and had been widely used for the time been in recent years. Though, with some important limitations it is very much obvious to develop a method and implement it using Kaplan-Meier estimator with interval and left censored survival data.

1.1.1 Motivations for Choosing Beta Generated Kumarawamy-G and Kumaraswamy-G Families

Furthermore, the basic motivations for the Beta Generated Kum-G and Kum-G families in practice are the following:

1. To make the kurtosis more flexible compared to the baseline model;
2. To produce a skewness for symmetrical distributions;
3. To construct heavy-tailed distributions for modeling real data;
4. To generate distributions with symmetric, left-skewed and right-skewed.
5. To define special models with all types of the hazard rate function;
6. To provide consistently better fits than other generated models under the same underlying distribution.

1.1.1.1 Motivations for Choosing Burr Type X as a Baseline Distribution

The motivations for choosing Burr Type X distribution are:

1. It is quite common to the two parameter gamma, Weibull and generalized exponential distributions.

2. The density function of Burr Type X has close form.
3. The Burr Type X can be used very conveniently even for censored data.
4. Unlike gamma, Weibull and Generalized exponential distributions, it has non-monotone hazard form which can be very useful in many practical applications.
5. The Burr Type X hazard function is monotonically increasing, monotonically decreasing and bathtub shapes unlike Unlike gamma, Weibull and Generalized exponential distributions with only monotonically increasing, monotonically decreasing shapes respectively.
6. The Burr Type X model, as a special case to the due to the relative flexibility of its hazard function and the ease for estimation of its parameters, ever since it has been widely used for analyzing reliability and agricultural lifetime data.

1.1.1.2 Motivations for BKBX and KBX Distributions

The motivations for choosing BKBX and KBX distribution are:

1. Model flexibility due to number of shape parameters provided.
2. Ability to tackle and solve an old problem by proposing a joint models.
3. We were motivated to introduced a new model called BKBX with six parameters ($\theta = \nu, \kappa, \varphi, \psi, \vartheta, \tau$) and also the modification of BKBX called the KBX model with four ($\theta = \varphi, \psi, \vartheta, \tau$) parameters, which generalized many of its baseline distributions, sub-models with their properties respectively
4. The fact that, this new models consists of several varieties of model fitting many areas of applied statistics, engineering, medicine and agriculture.
5. These two models serve as a generalization to other existing models which are very flexible and versatile models with properties of their densities which can be expressed as a mixture of many sub-models of Beta-G and Kumaraswamy-G families. For example, Burr Type X, Beta exponential, generalized exponential and Rayleigh distributions respectively.

Beta Kumaraswamy Burr Type X model is a very large family sub-model with quite a lot great properties of three most efficient models (Beta, Kumaraswamy and Burr Type X), which interest me to choose this flexible parametric high family of distribution with varieties of data fittings based on the shapes of the models parameters (right and left skewed datasets).

Reference to Kumaraswamy model as he fails to accept the fact that beta distribution does not fit hydrological datasets while Burr Type X model fits Reliability, survival and medical datasets.

This is the main reason we adopt this three models to fit one or more large datasets in different areas, as such, this model (BKBX) is actually very intricate model to use for simple cases, this needs a big data to fit it well and also we suggest academicians, researchers, doctors and nurses etc, which are use in modeling general datasets in agriculture, medical and engineering fields.

1.2 General Formulation of Survival Model

The survival analysis normally involves censored data, which makes the regression analysis techniques unsuitable with case of incomplete data. Censoring generally occurs when recorded survival times are only known for a proportion of the duration of the experiment and the remainder are known only to exceed certain times. This partial observations are too precious to be ignored and certain informations an be gained from knowledge of their existence. The most common in survival analysis are the right-censored data. However, the left censored data is theoretically possible but generally uncommon in these application. The basic problems in survival studies are determine by:

- The sampling distribution,
- The likelihood function, and
- The properties of any statistical methods derived.

We shall encounter the right-censored data in our analysis and we rely on large-sample methods.

1.2.1 Survival Distributions

Let T be a positive continuous random variable representing the survival time of an individual from a homogeneous population. Suppose $f(t)$ is the probability density function of the survival time T, of any specific individual which is defined by:

$$f(t) = \lim_{t \to 0} \frac{Pr(t < T \leq t + \Delta t)}{\Delta t}. \tag{1.1}$$

From Equation (1.1) above represents the probability of failure in a small interval $(t, t + \Delta t)$ upon the unit time, which has the following properties:

- $\int_0^\infty f(v)dv = 1$, and
- $f(t) \geq 0$ for all $t \geq 0$.

It is assume that m $f(t) \geq 0$ for all $t \geq 0$, and the cumulative distribution function F(t) is defined by:

$$F(t) = Pr(T \leq t) = \int_0^t f(v)dv.$$

By the fundamental theorem of calculus, $(t) = f(t)$. The cumulative distribution function $F(t)$ is monotonically increasing function of t such that F(0)=0 and $F(\infty) = \lim_{t\to 0} F(t) = 1$.

1.2.2 Survival Function $S(t)$

$S(t)$ is the probability that an individual is still alive at any time t, and is defined as follows:

$$S(t) = Pr(T > t) = \int_0^t f(v)dv.$$

Thus, $S(t) = 1 - F(t)$ and $S(t)$=-F'(t). It s generally assumed that $S(t) = 1$ for all $t \geq 0$, and that the survival function satisfies these three conditions below:

- $S(0) = 1$.
- $\lim_{t\to\infty} S(t) = 0$ and
- $S(t)$ is monotonically non-increasing in t.

1.2.3 Hazard Function $h(t)$

The $h(t)$ is defined as an instantaneous conditional probability of failure immediately after t, given survival until time t and is defined by:

$$\begin{aligned}
h(t) &= \lim_{\Delta t\to 0} \frac{Pr(t \leq T < t+\Delta t \mid T \geq t)}{\Delta t}, \\
&= \lim_{\Delta t\to 0} \frac{\int_t^{t+\Delta t} f(v)dv}{\int_t^{\infty} f(v)dv} \times \frac{1}{\Delta t}, \\
&= \lim_{\Delta t\to 0} \frac{F(t+\Delta t) - F(t)}{S(t)}, \\
&= \lim_{\Delta t\to 0} \frac{F(t+\Delta t) - F(t)}{S(t)}, \\
h(t) &= \frac{\text{Probability Density Function}}{\text{Survival Function}}, \\
h(t) &= \frac{f(t)}{S(t)}. \qquad (1.2)
\end{aligned}$$

Thus, the hazard $h(t)$, is the ratio of the $f(t)$ to the survival function $S(t)$. It is positive and also satisfies $\int_0^\infty h(v)dv = \infty$, for all small value of $\Delta t, h(t)\Delta t$ is the conditional probability of failure in the time interval $(t+\Delta t)$ given that the individual has survived until time t. The hazard or failure time is variously known as the force of mortality, age specific failure rate and instantaneous failure rate. From Equation (1.2) above in terms of the following:

$$h(t) = \frac{f(t)}{1-F(t)},$$

since,

$$h(t) = \frac{f(t)}{S(t)},$$

then,

$$\int_0^\infty h(v)dv = \int_0^\infty \frac{f(v)}{S(v)}dv = -\ln S(t).$$

Therefore, the $S(t)$ can also be written as:

$$S(t) = \exp\Big[-\int_0^t h(v)dv\Big]. \tag{1.3}$$

We can write the distribution function in terms of the hazard function as:

$$F(t) = 1 - \exp\Big[-\int_0^t h(v)dv\Big], \tag{1.4}$$

and by using equation (1.2) above, the density function can also be written as:

$$f(t) = h(t)\exp\Big[-\int_0^t h(v)dv\Big].$$

The cumulative hazard function (known as the integrated hazard function) can be defined by:

$$H(t) = \int_0^t h(v)dv. \tag{1.5}$$

Where $t \geq 0$. It has the following properties:

- $S(0) = 0$.
- $\lim_{t\to\infty} H(t) = \infty$, and
- $S(t)$ is monotonically non-decreasing in t.

Using Equation (1.3) and Equation (1.4), we can write the integrated hazard function as:

$$H(t) = -\ln S(t).$$

Other representations of the survival time distribution function include the mean residual life function which is the expected remaining life $(T-t)$, given the individual has survived to time t. It is defined as:

$$\begin{aligned}\mu(t) &= \Big[T-t|T>t\Big],\\ &= \int_t^{\infty}(v-t)\frac{f(v)}{S(t)}dv,,\\ &= \frac{1}{S(t)}\int_t^{\infty} vf(v)dv - t.\end{aligned}$$

Where $0 \leq t < \infty$ from Equation (1.5), above

$$\mu(t)S(t) = \int_t^{\infty}(v-t)f(v)dv,$$

Let, $u = v-t$, and integrating by parts,

$$\begin{aligned}\mu(t)S(t) &= \int_t^{\infty} S(u)du,\\ \text{where, } S(0) &= 0,\\ \mu(0) &= \int_t^{\infty} S(u)du,\\ &= \int_t^{\infty} uf(u)du = E[T].\end{aligned}$$

Let,

$$Q(t) = \int_t^{\infty} S(u)du.$$

This implies that :

$$\acute{Q}(t) = -S(t), Q(0) = \mu(0),$$

and also,

$$\begin{aligned}\frac{d}{dt}(\ln Q(t)) &= \frac{\acute{Q}(t)}{Q(t)},\\ &= \frac{S(t)}{\int_t^{\infty} S(v)dv} = \frac{-1}{\mu(t)},\end{aligned}$$

Hence,

$$\ln\acute{Q}(t) - \ln Q(0) = -\int_0^t \frac{du}{\mu(u)},$$

This implies that,

$$\frac{Q(t)}{Q(0)} = \exp\Big[-\int_0^t \frac{du}{\mu(u)}\Big],$$

therefore,

$$\mu(t)S(t) = Q(0)\exp\Big[-\int_0^t \frac{du}{\mu(u)}\Big].$$

Finally, the survival function can be expressed as:

$$S(t) = \frac{\mu(0)}{\mu(t)}$$
$$S(t) = \exp\Big[-\int_0^t \frac{du}{\mu(u)}\Big] \tag{1.6}$$

Mathematically, all six distributed representations above, namely:

$$f(t), F(t), S(t), h(t), H(t) \text{ and } \mu(\text{t}),$$

are equivalent. Any one distributional representation implies the other five functions. Discrete survival times distribution are applied less frequently than continuous distributions. It is applicable for example where a grouped data are involved. The previous distribution representations.

1.2.4 Bounded Cumulative Hazard (BCH) and Cure Fraction Models

The BCH model initially was introduced by Yakovlev et al. (1993), assuming that, cancer patients on clinical trial and also after the initial treatment have some active cells in their body called "clonogens", these cancer cells grow simultaneously and produces a detectable cancer mass later on called cancer relapse.

Yakovlev et al. (1993), assume that the number of clonogens denoted by N, follows a Poisson distribution with mean θ. Let $Z_i, i = 1,2,3,...,N$ be the time needed by the i^{th} clonogen to reproduce a detectable cancer mass. Where, the time it takes the cancer to relapse can be defined by the random variable T=min$\{Z_i, i = 1,2,3,...,N\}$ and Z_i are identically independently distributed iid. The survival function of the entire population, $S_o(t)$, conditional on N can be obtained as follows:

$$\begin{aligned} S_o(t) &= P[\text{There is no detectable cancer by time t/N}], \\ &= P[N=0] + P[t_1 > t, t_2 > t, ..., t_N > t/N \geq 1], \\ &= \exp(-\theta) + \sum_{\infty}^{N=1} (S(t))^N \Big[\frac{\exp(-\theta)\theta^N}{N!}\Big]. \\ &= \exp(-\theta F(t)). \end{aligned}$$

Where, *F(t)* is a proper cumulative distribution function that represents the time to development of detectable tumor mass. The parametric BCH model comes up when some of the common parametric choices of *F(t)*, for example like: Weibull and exponential models we employed. Moreover, an alternative for the non parametric models for the *F(t)* can be considered. The cure fraction, which is denoted by p, based on this model

is defined as:

$$p = \lim_{t\to\infty} S_o(t) \equiv P(N = 0) = \exp(-\theta).$$

Consequently, the estimation of the cure fraction with censored data can be achieved by the population methods of complexity Markov Chain Monte Carlo (MCMC) methods because of suitability and flexibility in co-ordinating highly complex models like: Beta Kumaraswamy Burr Type X (BKBX) and Beta-Weibull (BWB) distributions which produces a stable computations and faster than expectation maximization (EM) algorithm in term of convergence of the posterior summaries. An important property of the BCH model is that it can be written in the shape of mixture model as follows:

$$\begin{aligned}
S_u(t) =& \exp(-\theta F(t)),\\
=& \exp(-\theta) + \exp(-\theta F(t)) - \exp(-\theta),\\
=& \exp(-\theta) + [1-\exp(-\theta)]\Big[\frac{\exp(-\theta F(t))\exp(-\theta)}{1-\exp(-\theta)}\Big],\\
=& p + [1-p]\Big[\frac{\exp(-\theta F(t))\exp(-\theta)}{1-\exp(-\theta)}\Big],
\end{aligned}$$

which is equals to the mixture model formula given as:

$$S_u(t) = p + [1-p]S_o(t),$$

where, $S_o(t)$ is the survival function of the uncured patients.

1.2.5 Kaplan-Meier (KM) Estimator

Kaplan-Meier or simply (KM) estimator, otherwise called the product limit estimate of the survival function from life-time datasets. It is the limit of the life-time table when intervals constructed in the life tables are taken so small that at most one only individual dies within each interval. An important property of this estimator, is that the results for the survival function estimates do not depend on the intervals of the survival data.

The Kaplan-Meier estimate of the survival function is an empirical or non-parametric method of estimating $S(t)$ from non- or right-censored data. It is extremely popular as it requires only very weak assumptions and yet utilizes the information content of both fully observed and right-censored data.

It drops only at times when a failure has been observed. If we write $t_{(i)}$ as the i^{th} ordered event time, and $d_{(i)}$, $q_{(i)}$ and $n_{(i-)}$ accordingly, the Kaplan-Meier formula can be written as:

$$\hat{S(t)} = \prod t_{(i)\leq t} \frac{n_{(i-)} - d_{(i)}}{n_{(i-)}}. \tag{1.7}$$

Note: that the Kaplan-Meier estimate in Equation (1.7) above does not change between events, nor at times when only censoring

1.3 Censoring

1.3.1 Right Censored Data

It is often called the suspended censoring, which the data are mostly the common type in the incomplete population, otherwise called censored data. In this case of censored survival data, the right censoring generally occurs when some when some individuals have their future time past the final observed time.

As such, we may be interested to know for how long cancer patients will survive after cancer diagnosis. In such situation we subject the group of cancer patients to a clinical treatment trial at time t, follow-up for sufficient and regular periods of time, and record the patients' times to failure. The commonest form of censoring is right censoring. Here, the subject is followed until some time, at which the event has yet to occur, but then takes no further part in the study. This may be because:

1. The subject dies from another cause, independently of the cause of interest;
2. The study ends while the subject survives; or
3. the subject is lost to the study, by dropping out, moving to a different area, etc.

Some patients may take long time to experience the event of interest and nothing wrong happens to them during the experiment. Let T_i^* and C_i denote the survival and right censoring times for the i^{th} individual, where δ_i is a censoring indicator variable, that is, δ_i =1 for an observed life time and δ_i =0 for censored lifetime given as: observe $T_i = \min\{T_i^*, C_i\}$ censoring indicator

$$\delta_i = \begin{cases} 1 & \text{if failure } T_i = T_i^* \\ 0 & \text{if right censored } T_i = C_i \end{cases}$$

observe (T_i, δ_i) for $i = 1, 2, \ldots, n$.

1.3.2 Left Censored Data

This type of censoring occurs normally when the time of entry into study for some individuals is unknown. Left censoring is much rarer. This occurs when the event of interest has already occurred at the observation time, but it is not known exactly when it occur. Examples of left censoring include:

1. Infection with a sexually-transmitted disease such as HIV/AIDS;

2. Onset of a pre-symptomatic illness such as cancer; and

3. Time at which teenagers begin to drink alcohol, later forgotten.

1.3.3 Interval Censored Data

This type of censoring arises when the failure time cannot be observed, but can only be determined to lie in an interval obtained from a sequence of examination times. However, interval censoring for cancer patients arises when the time to the event of interest T can be known that it lies within a random interval of time when a group of cancer patients involved in a clinical trial have periodic follow ups. In this case, the inly information we have is that the lifetime T lied between two points of time left (L) and right (R), where $L < R$.

Among the three main types of censored data, right censoring received more attention, because of the availability and vast of its datasets in all areas of research makes it famous among all, on the other hand both the left and right censored data are special case of the interval censored data which stated in the literature that the failure time based on interval censored cannot be directly observed, which only lies between a limit interval. Interval censoring is both extremely common and very rare. This oc- curs when the exact time the event occurs is not known precisely, but an interval bounding this time is known. Examples of interval censoring include:

1. Infection with a sexually-transmitted disease such as HIV/AIDS with regular testing (e.g. annually); and

2. Failure of a machine during the Chinese new year.

1.4 Scope of the Study

This thesis focuses mostly on developing new models, estimation of parameters and also applying the cure models (mixture and non-mixture) comparison and application of datasets by employing a powerful simulation algorithms, collectively known as Markov Chain Monte Carlo (MCMC) methods. These algorithms, largely developed during the nineties and have have revolutionized the practice of statistics by allowing us to tackle problems of real complexity that were impossible (or extremely difficult) to handle before.

This means that we can now be much more realistic in our modeling, in which is extremely important when dealing with a real world applications. These methods are particularly well suited to computing the posterior distribution (via simulation) in Bayesian inference and are to a large extent responsible for the great upsurge in popularity of Bayesian statistics during the last decade. Thus, we shall make some remarks about

the use of MCMC algorithms in other contexts, this thesis will be focused on the application of MCMC methods to Bayesian inference and probability theory.

1.5 Problem Statement

In recent years there has been a renewed interest in constructing new statistical lifetime distributions. We proposed a new six and modified four- parameters distribution by using the Burr Type X distribution as the baseline model in the Beta Kumaraswamy-G and Kumaraswamy-G versatile families of distribution respectively. The new lifetime are interesting and potentially good as a theoretical contribution to the literature on distributions.

The model proposed by Kumaraswamy (1980), is an old time probability distribution two random processes with applications to hydrological data. The Kumaraswamy distribution with two parameters φ and ψ received quite few attention for decades which fits hydrological and climatological lifetime datasets.

In reliability and life testing experiments, many times the data are modeled by finite range distributions. We start with the Kumaraswamy model on the interval (0,1), having the probability density function and the cumulative distribution function with two shape parameters $\varphi, \psi > 0$ defined by Kumaraswamy distribution which fails to accept the fact that beta does not strictly fit hydrological data such as daily rainfall etc.

This model is unimodal, increasing, decreasing and constant based on the values of its parameters. It has shown that both the beta and Kumaraswamy distribution have same shape properties. The Kumaraswamy model is not as famous as beta model to statisticians over the years. Also, Kumaraswamy (1980), emphasize some advantages of Kumaraswamy model over the beta model and also stated that:

- The normality assumption constant is easy.
- It has an easy formula for the model.
- A quantile probability functions does not consist any special functions.
- A simplicity formula in generating a random numbers and moments of order statistics and L-moments.

Jones (2009), mentioned that the beta model has some influence over the Kumaraswamy model:

1. An easy procedure for moments, MGF and the normality assumption.
2. It has a single parameter of symmetric models in its sub-family.

3. An easy estimation process of parameter.

4. It has a simple formula for random generator.

5. It has many ways of generating data of the distribution.

Kumaraswamy model approximates the beta distribution with two parameters ν and κ, but is easier to work with in some ways. The Kumaraswamy distribution, apparently it came out of hydrology (water). In general the main limitation of the parametric cure fraction models are sometimes hard to find a distribution flexible enough to fit the data. on these context, the non-parametric techniques are considered to be more attractive under the violation of parametric assumptions.

One of the limitation of BKBX model is that the survival and hazard functions cannot be expressed in a closed form and KBX model can be expressed in a closed form, specifically when more covariates are considered, thus numerical approach that is the integration techniques are required to determine the estimate of parameters in the model.

In medical and health studies, the concept of cured patients refers to one who will never experience the event of interest (fail) within the pre-assigned follow-up time. Estimation of the probability of cure, especially in different cancer clinical trials, is badly needed. The survival models which incorporate patients who are expected to be cured are growing rapidly since these models addressed the proportion of cured patients which is important in understanding prognosis in potentially hard diseases and also serve to expose unknown healthy issues related to the studied population.

Many common and survival models are available for this purpose, such as the mixture cure model, developed by Boag (1949), which is the most widely used model in survival data analysis. It has been extensively employed by a large number of researchers in many different fields. However, this model has some drawbacks as was highlighted and proposed by Chen et al. (1999). These limitations are:

- Chen et al. (1999), suggested that when covariates are included in the analysis, on the other hand mixture model does not require a proportional hazard context.

- Mixture model provides improper posterior distributions for some types of non-informative improper priors when covariates are included through the parameter π via a standard regression model.

- Mixture model does not seems to outline the biological process for obtaining the failure time, in a situation where cancer relapse is involved.

1.6 Research Objectives

This research will be carry out with the following objectives, as follows:

1. To introduce a new model Beta Kumaraswamy Burr-Type X (BKBX) with properties.
2. To develop a modified new model Kumaraswamy Burr-Type X (KBX) with properties.
3. To propose a Bayesian parameter estimation and simulation studies on the new models (BKBX and KBX) using the long-term survival models with censored data and covariates.
4. To applied a real life data on the new proposed models using maximum likelihood estimator.

1.7 Contribution

The main contribution of this work to the existing vast literature regarding the family of probability distribution by constructing a credible models and baselines which enable and also motivate us to fill some gap in the benchmark of some useful numerous models.

1. Demonstrating that are inadequate and insufficient to model some kind of datasets in different area of research like: agriculture, engineering and medicine etc.
2. Proposing a new flexible models (BKBX and KBX) and comparing with existing ones and validating the results with a simulation study.
3. Providing a gap in the literature by proposing a new model and use it to incorporate covariates in cure models by solving censored and uncensored data respectively.
4. The BKBX and KBX are very good in modeling right skewed, left skewed and symmetric datasets.
5. We provide some contributions based the literature that no researcher developed a new model, incorporate it with cure fraction and covariates by checking the importance of the proposed methodology.

1.8 Outline of the Thesis

This thesis comprises of two parts dealing with the probability distribution theory and parametric approaches to model the cure rate estimation with censored data and covariates.

The first part is about developing a new model using the method of joining two or model existing parametric models and employed an empirical simulation studies. On the other part it handles the estimation of the cure fraction using the Bayesian analysis approaches via Markov Chain Monte Carlo (MCMC) methods using the Gibbs sampling algorithm to obtain the posterior summaries.

Chapter 1: This chapter, provides some introduction based of the new models Beta Kumaraswamy Burr-Type X and Kumaraswamy Burr-Type X, with their respective properties and application as well as the Bayesian survival analysis with censoring respectively.

Chapter 2: This chapter, presents a review of the literature related to the main themes of this research is presented. The probability theory, survival data analysis and cure fraction models.

Chapter 3: In this chapter, we present and describes the new model (BKBX) distribution with six parameters its properties where a simulation study which was done to compare and contrast, also check the efficiency and suitability of the new (BKBX) model.

Chapter 4: In this Chapter, we also proposed the second new modified model (KBX) distribution with four parameters and its properties which is also the modified version of the (BKBX) where a simulation study which was done to compare and contrast, also check the efficiency and suitability of the new (KBX) model.

Chapter 5: This chapter involves the parametric estimation based on cure models with or with-out covariates and also a simulation study was done using MCMC algorithm.

Chapter 6: In this chapter, we provide an application of real data sets to the new and existing models by different criterion of model selections as well as the likelihood ratio and goodness-of-fit tests respectively.

Chapter 7: This chapter, contains the summary, conclusion, recommendations and future research work.

BIBLIOGRAPHY

Aarset, M. V. 1987. How to Identify a Bathtub Hazard Rate. *IEEE Transactions on Reliability* 36 (1): 106–108.

Achcar, J. A., Coelho-Barros, E. A. and Mazucheli, J. 2012. Cure Fraction Models Using Mixture and Non-Mixture Models. *Tatra Mountains Mathematical Publications* 51 (1): 1–9.

Ahmad, Z. and Hussain, Z. 2017. Kumaraswamy Inverse Flexible Weibull Extension Distribution. *CEPAL Review Journal* 120 (1): 1–4.

Akaike, H., Petrov, B. and Csaki, F. 1973. Information Theory and an Extension of the Maximum Likelihood Principle 10 (1): 2–8.

Akinsete, A., Famoye, F. and Lee, C. 2008. The Beta-Pareto Distribution. *Statistics* 42 (6): 547–563.

Alizadeh, M., Cordeiro, G. M., De Brito, E. and Demétrio, C. G. B. 2015a. The Beta Marshall-Olkin Family of Distributions. *Journal of Statistical Distributions and Applications* 2 (1): 1–18.

Alizadeh, M., Tahir, M., Cordeiro, G. M., Mansoor, M., Zubair, M. and Hamedani, G. 2015b. The Kumaraswamy Marshal-Olkin Family of Distributions. *Journal of the Egyptian Mathematical Society* 23 (3): 546–557.

Aljawadi, B., Bakar, M. R. A., Ibrahim, N. A. and Midi, H. 2011. Parametric Estimation of the Cure Fraction Based on BCH Model Using Left-Censored Data with Covariates. *Modern Applied Science* 5 (3): 2–5.

Alzaatreh, A., Famoye, F. and Lee, C. 2014. The Gamma-Normal Distribution: Properties and Applications. *Computational Statistics and Data Analysis* 69 (1): 67–80.

Alzaatreh, A., Lee, C. and Famoye, F. 2013. A New Method for Generating Families of Continuous Distributions. *METRON Journal* 71 (1): 63–79.

Amin, N. A. M. and Adam, M. B. 2016. Markov Chain Monte Carlo Convergence Diagnostics For Gumbel Model. *Journal of Applied Environmental and Biological Sciences* 6 (2S): 130–136.

Amin, N. A. M., Adam, M. B. and Ibrahim, N. A. 2015. Bayesian Inference using Multiple-try Metropolis Hastings Scheme for the Efficiency of Estimating Gumbel Distribution Parameters. *MATEMATIKA* 31 (1): 25–36.

Andrews, D. and Herzberg, A. 1985. Stress-Rupture Life of Kevlar 49/epoxy Spherical Pressure Vessels. *Data* 23 (3): 181–186.

Arasan, J. and Lunn, M. 2009. Survival Model of a Parallel System With Dependent Failures and Time Varying Covariates. *Journal of Statistical Planning and Inference* 139 (3): 944–951.

Bakar, M. R. A., Salah, K. A., Ibrahim, N. A. and Haron, K. 2009. Bayesian Approach for Joint Longitudinal and Time-to-Event Data With Survival Fraction. *Bulletin of the Malaysian Mathematical Sciences Society* 32 (1): 75–100.

Bakoban, R. and Abu-Zinadah, H. H. 2017. The Beta Generalized Inverted Exponential Distribution With Real Data Applications. *REVSTAT Statistical Journal* 15 (1): 65–88.

Barreto-Souza, W., Cordeiro, G. M. and Simas, A. B. 2011. Some Results for Beta Fréchet Distribution. *Communications in Statistics Theory and Methods* 40 (5): 798–811.

Barreto-Souza, W., Santos, A. H. and Cordeiro, G. M. 2010. The Beta Generalized Exponential Distribution. *Journal of Statistical Computation and Simulation* 80 (2): 159–172.

Behairy, S., AL-Dayian, G. and El-Helbawy, A. 2016. The Kumaraswamy-Burr Type III Distribution: Properties and Estimation. *British Journal of Mathematics and Computer Science* 14 (2): 1–21.

Benkhelifa, L. 2017. The Beta Generalized Gompertz Distribution. *Applied Mathematical Modelling* 36 (2): 3–8.

Berkson, J. and Gage, R. P. 1952. Survival Curve for Cancer Patients Following Treatment. *Journal of the American Statistical Association* 47 (259): 501–515.

Boag, J. W. 1949. Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy. *Journal of the Royal Statistical Society. Series B (Methodological)* 11 (1): 15–53.

Burr, I. W. 1942. Cumulative frequency functions. *The Annals of Mathematical Statistics* 13 (2): 215–232.

Carrasco, J. M. and Cordeiro, G. M. 2017. An Extension of the Kumaraswamy Distribution. *International Journal of Statistics and Probability* 6 (3): 61.

Chen, M.-H., Harrington, D. P. and Ibrahim, J. G. 2002. Bayesian Cure Rate Models for Malignant Melanoma: A Case-Study of Eastern Cooperative Oncology Group Trial E1690. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51 (2): 135–150.

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 (447): 909–919.

Chen, M.-H. and Shao, Q.-M. 1999. Monte Carlo Estimation of Bayesian Credible and HPD intervals. *Journal of Computational and Graphical Statistics* 8 (1): 69–92.

Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. 2007. Flexible Cure Rate Modeling Under Latent Activation Schemes. *Journal of the American Statistical Association* 102 (478): 560–572.

Cordeiro, G. M., Cristino, C. T., Hashimoto, E. M. and Ortega, E. M. 2013a. The Beta Generalized Rayleigh Distribution With Applications to Lifetime Data. *Statistical Papers* 54 (1): 133–161.

Cordeiro, G. M. and de Castro, M. 2011. A New Family of Generalized Distributions. *Journal of Statistical Computation and Simulation* 81 (7): 883–898.

Cordeiro, G. M., Gomes, A. E., da Silva, C. Q. and Ortega, E. M. 2013b. The Beta Exponentiated Weibull Distribution. *Journal of Statistical Computation and Simulation* 83 (1): 114–138.

Cordeiro, G. M., Nadarajah, S. and Ortega, E. M. 2013c. General Results for the Beta Weibull Distribution. *Journal of Statistical Computation and Simulation* 83 (6): 1082–1114.

Cordeiro, G. M., Ortega, E. and Silva, G. 2012a. The Beta Extended Weibull Family. *Journal of Probability and Statistical Science* 10 (10): 15–40.

Cordeiro, G. M., Ortega, E. M. and Nadarajah, S. 2010. The Kumaraswamy Weibull Distribution With Application to Failure Data. *Journal of the Franklin Institute* 347 (8): 1399–1429.

Cordeiro, G. M., Ortega, E. M. and Silva, G. O. 2014. The Kumaraswamy Modified Weibull Distribution: Theory and Applications. *Journal of Statistical Computation and Simulation* 84 (7): 1387–1411.

Cordeiro, G. M., Pescim, R. R. and Ortega, E. M. 2012b. The Kumaraswamy Generalized Half-Normal Distribution for Skewed Positive Data. *Journal of Data Science* 10 (2): 195–224.

Cordeiro, G. M., Simas, A. B. and Stošic, B. D. 2011. Closed Form Expressions for Moments of the Beta Weibull Distribution. *Anais da Academia Brasileira de Ciências* 83 (2): 357–373.

Correa, M. A., Nogueira, D. A. and Ferreira, E. B. 2012. Kumaraswamy Normal and Azzalini's Skew Normal Modeling Asymmetry. *Sigmae* 1 (1): 65–83.

De Castro, M., Cancho, V. G. and Rodrigues, J. 2010. A Hands on Approach for Fitting Long-Term Survival Models Under the GAMLSS Framework. *Computer Methods and Programs in Biomedicine* 97 (2): 168–177.

Domma, F. and Condino, F. 2016. Use of the Beta-Dagum and Beta-Singh-Maddala Distributions for Modeling Hydrologic Data. *Stochastic Environmental Research and Risk Assessment* 1–15.

El-Gohary, A., Alshamrani, A. and Al-Otaibi, A. N. 2013. The Generalized Gompertz Distribution. *Applied Mathematical Modelling* 37 (1): 13–24.

Elbatal, I. 2013. Kumaraswamy Linear Exponential Distribution. *Pioneer Journal of Theoretical and Applied Statistics* 5 (1): 59–73.

Elbatal, I., El Gebaly, Y. M. and Amin, E. A. 2017. The Beta Generalized Inverse Weibull Geometric Distribution. *Pakistan Journal of Statistics and Operation Research* 13 (1): 75–90.

Eugene, N., Lee, C. and Famoye, F. 2002. Beta-Normal Distribution and Its Applications. *Communications in Statistics Theory and Methods* 31 (4): 497–512.

Famoye, F., Lee, C. and Olumolade, O. 2005. The Beta-Weibull Distribution. *Journal of Statistical Theory and Applications* 4 (2): 121–136.

Fan, Y. and van den Dool, H. 2004. Climate Prediction Center Global Monthly Soil Moisture Data Set at 0.5 Resolution for 1948 to Present. *Journal of Geophysical Research: Atmospheres* 109 (1): 8–13.

Gamerman, D. and Lopes, H. F. 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference 2 (1): 125–137.

Gelfand, A. E. and Dey, D. K. 1994. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society. Series B (Methodological)* 10 (1): 501–514.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. 2014. Bayesian Data Analysis 2 (1): 217–234.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. 1995. Markov Chain Monte Carlo in Practice 2 (1): 25–37.

Gomes, A. E., da Silva, C. Q., Cordeiro, G. M. and Ortega, E. M. 2014. A New Lifetime Model: The Kumaraswamy Generalized Rayleigh Distribution. *Journal of Statistical Computation and Simulation* 84 (2): 290–309.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. 1979. Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. *Water Resources Research* 15 (5): 1049–1054.

Gu, Y., Sinha, D. and Banerjee, S. 2011. Analysis of Cure Rate Survival Data Under Proportional Odds Model. *Lifetime Data Analysis* 17 (1): 123–134.

Gupta, A. K. and Nadarajah, S. 2004. Handbook of Beta Distribution and Its Applications 2 (1): 3–33.

Gupta, R. D. and Kundu, D. 2001. Generalized Exponential Distribution: Different Method of Estimations. *Journal of Statistical Computation and Simulation* 69 (4): 315–337.

Guure, C. B., Ibrahim, N. A., Adam, M. B. and Bosomprah, S. 2014. Bayesian Parameter and Reliability Estimate of Weibull Failure Time Distribution. *The Bulletin of the Malaysian Mathematical Sciences Society, Series* 2 (14): 611–632.

Guure, C. B., Ibrahim, N. A. and Ahmed, A. O. M. 2012. Bayesian Estimation of Two-Parameter Weibull Distribution Using Extension of Jeffreys' Prior Information With Three Loss Functions. *Mathematical Problems in Engineering* 2012 (4): 2–11.

Harris, J. W. and Stöcker, H. 1998. Handbook of Mathematics and Computational Science 6 (2): 3–19.

Hassan, A. S. and Elgarhy, M. 2016. Kumaraswamy Weibull-Generated Family of Distributions With Applications. *Advances and Applications in Statistics* 48 (3): 2–5.

Hosmer, D. W., Lemeshow, S. and May, S. 2008. Model Development. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition* 132–168.

Ibrahim, J. G., Chen, M. H. and Sinha, D. 2001. Bayesian Semiparametric Models for Survival Data With a Cure Fraction. *Biometrics* 57 (2): 383–388.

Ibrahim, N. A., Taweab, F. and Arasan, J. 2014. A Parametric Non-Mixture Cure Survival Model with Censored Data. *Computational Problems in Engineering* 307 (6): 231–238.

Iversen, G. R. 1984. Bayesian Statistical Inference. *Sage Journals* 1 (43): 9–12.

Jiang, W., Sun, H. and Peng, Y. 2017. Prediction Accuracy for the Cure Probabilities in Mixture Cure Models. *Statistical Methods in Medical Research* 5 (45): 2–8.

Johnson, K. and Kotz, S. 1994. Balakrishnan (1994), Continuous Univariate Distributions. *New York: Wiley* 23 (1): 70–81.

Jones, M. 2004. Families of Distributions Arising From Distributions of Order Statistics. *TEST* 13 (1): 1–43.

Jones, M. 2009. Kumaraswamys Distribution: A Beta-Type Distribution With Some Tractability Advantages. *Statistical Methodology* 6 (1): 70–81.

Jones, M. and Faddy, M. 2003. A Skew Extension of the T-Distribution, With Applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (1): 159–174.

Kalbfleisch, J. D. and Prentice, R. L. 2011. The Statistical Analysis Of Failure Time Data. *John Wiley and Sons* 360 (1): 25–57.

Kenney, J. F. and Keeping, E. S. 1954. Mathematics of Statistics-Part One 2–4.

Khaleel, M. A., Ibrahim, N. A., Shitan, M. and Merovci, F. 2016. Some Properties of Gamma Burr Type X Distribution with Application. *AIP Conference Proceedings* 1739 (1): 2–4.

Kim, S., Chen, M.-H. and Dey, D. K. 2011. A New Threshold Regression Model for Survival Data With a Cure Fraction. *Lifetime Data Analysis* 17 (1): 101–122.

Kumaraswamy, P. 1980. A Generalized Probability Density Function for Double-Bounded Random Processes. *Journal of Hydrology* 46 (1): 79–88.

Lai, X., Yau, K. K. and Liu, L. 2017. Competing Risk Model With Bivariate Random Effects for Clustered Survival Data. *Computational Statistics Data Analysis* 112: 215–223.

Lee, C., Famoye, F. and Olumolade, O. 2007. Beta-Weibull distribution: Some Properties and Applications to Censored Data. *Journal of Modern Applied Statistical Methods* 6 (1): 17.

Lemonte, A. J. et al. 2014. The Beta Log-Logistic Distribution. *Brazilian Journal of Probability and Statistics* 28 (3): 313–332.

Li, Y., Wileyto, E. P. and Heitjan, D. F. 2010. Modeling smoking cessation data with alternating states and a cure fraction using frailty models. *Statistics in medicine* 29 (6): 627–638.

Madaki, U. Y. and Bakar, M. R. A. 2016a. A Bayesian Estimation on Right Censored Survival Data With Mixture and Non-Mixture Cured Fraction Model Based on Beta-Weibull Distribution. *AIP Conference Proceedings* 1739 (1): 2–4.

Madaki, U. Y. and Bakar, M. R. A. 2016b. Cure Models Based on Weibull Distribution With and Without Covariates Using Right Censored Data. *Indian Journal of Science and Technology* 9 (28): 4–7.

Madaki, U. Y., Bakar, M. R. A., Ibrahim, N. A., Arasan, J. and Husain, Q. N. 2016. A Bayesian via Laplace Approximation on Log-Gamma Model With Censored Data. *Modern Applied Science* 11 (1): 2–9.

Madi, Mohammad Z Raqab, M. T. 2009. Bayesian Analysis for the Exponentiated Rayleigh Distribution. *METRON* 67 (3): 269–288.

Maller, R. A. and Zhou, X. 1996. Survival Analysis With Long-Term Survivors. *Wiley New York* 31 (2): 3–8.

Marshall, A. W. and Olkin, I. 1967. A Multivariate Exponential Distribution. *Journal of the American Statistical Association* 62 (317): 30–44.

Martinez, E. Z., Achcar, J. A., Jácome, 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. *Computer Methods and Programs in Biomedicine* 112 (3): 343–355.

Mead, M., Afify, A. Z., Hamedani, G. and Ghosh, I. 2017. The Beta Exponential Fréchet Distribution With Applications. *Austrian Journal of Statistics* 46 (1): 41–63.

Merovci, F., Khaleel, M. A., Ibrahim, N. A. and Shitan, M. 2016. The beta Burr Type X distribution properties with application. *SpringerPlus* 1–18.

Moors, J. 1988. A Quantile Alternative for Kurtosis. *Journal of the Royal Statistical Society: Series D, The Statistician* 21 (5): 25–32.

Nadarajah, S., Cordeiro, G. M. and Ortega, E. M. 2012. General Results for the Kumaraswamy-G Distribution. *Journal of Statistical Computation and Simulation* 82 (7): 951–979.

Nadarajah, S., Cordeiro, G. M. and Ortega, E. M. 2015. The Zografos–Balakrishnan-G Family of Distributions: Mathematical Properties and Applications. *Communications in Statistics-Theory and Methods* 44 (1): 186–215.

Nadarajah, S. and Eljabri, S. 2013. The Kumaraswamy GP Distribution. *Journal of Data Science* 11 (4): 739–766.

Nadarajah, S. and Kotz, S. 2004. The Beta Gumbel Distribution. *Mathematical Problems in Engineering* 2004 (4): 323–332.

Nadarajah, S. and Kotz, S. 2006. The Beta Exponential Distribution. *Reliability Engineering and System Safety* 91 (6): 689–697.

Nadarajah, S. and Zografos, K. 2003. Formulas for Rényi Information and Related Measures for Univariate Distributions. *Information Sciences* 155 (1): 119–138.

Ortega, E. M., Cordeiro, G. M. and Kattan, M. W. 2012. The Negative Binomial–Beta Weibull Regression Model to Predict the Cure of Prostate Cancer. *Journal of Applied Statistics* 39 (6): 1191–1210.

Osborn, D. and Madey, R. 1968. The Incomplete Beta Function and Its Ratio to the Complete Beta Function. *Mathematics of Computation* 22 (101): 159–162.

Paranaíba, P. F., Ortega, E. M., Cordeiro, G. M. and Pascoa, M. A. 2013. The Kumaraswamy Burr Xii Distribution: Theory and Practice. *Journal of Statistical Computation and Simulation* 83 (11): 2117–2143.

Paranaíba, P. F., Ortega, E. M., Cordeiro, G. M. and Pescim, R. R. 2011. The Beta Burr Xii Distribution With Application to Lifetime Data. *Computational Statistics and Data Analysis* 55 (2): 1118–1136.

Pararai, M., Oluyede, B. O., Warahena-Liyanage and Gayan. 2016. The Beta Lindley-Poisson Distribution with Applications. *Journal of Statistical and Econometric Methods* 5 (4): 1–37.

Peng, Y. and Taylor, J. M. 2011. Mixture Cure Model with Random Effects for the Analysis of a Multi-Center Tonsil Cancer Study. *Statistics in Medicine* 30 (3): 211–223.

Pescim, R. R., Cordeiro, G. M., Demétrio, C. G., Ortega, E. M. and Nadarajah, S. 2012. The New Class of Kummer Beta Generalized Distributions. *Statistics and Operations Research Transactions (SORT)* 36 (2): 153–180.

Pescim, R. R., Demétrio, C. G., Cordeiro, G. M., Ortega, E. M. and Urbano, M. R. 2010. The Beta Generalized Half-Normal Distribution. *Computational Statistics and Data Analysis* 54 (4): 945–957.

Pollard, J. H. and Valkovics, E. J. 1992. The Gompertz Distribution and Its Applications. *Genus* 2 (1): 15–28.

Raqab, M. 1998. Order Statistics From the Burr Type X Model. *Computers and Mathematics with Applications* 36 (4): 111–120.

Raqab, M. Z. and Kundu, D. 2005. Comparison of Different Estimators of P$[Y < X]$ for a Scaled Burr Type X Distribution. *Communications in Statistics: Simulation and Computation* 34 (2): 465–483.

Raqab, M. Z. and Kundu, D. 2006. Burr Type X Distribution: Revisited. *Journal of Probability and Statistical Sciences* 4 (2): 179–193.

Raqab, M. Z. and Madi, M. T. 2011. Generalized Rayleigh Distribution. *International Encyclopedia of Statistical Science* 5 (2): 599–603.

Renyi, A. 1961. On Measures of Entropy and Information. *Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1 (18): 547–561.

Santos, C. A. and Alberto Achcar, J. 2011. A Bayesian Analysis in the Presence of Covariates for Multivariate Survival Data: An example of Application. *Colombian Journal of Statistics* 34 (1): 111–131.

Seppä, K., Hakulinen, T., Kim, H.-J. and Läärä, E. 2010. Cure Fraction Model With Random Effects For Regional Variation In Cancer Survival. *Statistics in Medicine* 29 (27): 2781–2793.

Shams, T. 2013. The Kumaraswamy-generalized Lomax distribution. *Middle East Journal of Scientific Research* 17 (3): 641–646.

Shiffman, S., Waters, A. J. and Hickcox, M. 2004. The Nicotine Dependence Syndrome Scale: A Multidimensional Measure of Nicotine Dependence. *Nicotine and Tobacco Research* 6 (2): 327–348.

Silva, G. O., Ortega, E. M. and Cordeiro, G. M. 2010. The Beta Modified Weibull Distribution. *Lifetime data analysis* 16 (3): 409–430.

Singla, N., Jain, K. and Sharma, S. K. 2012. The Beta Generalized Weibull Distribution: Properties and Applications. *Reliability Engineering and System Safety* 102: 5–15.

Smith, R. L. and Naylor, J. 1987. A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull Distribution. *Applied Statistics* 22 (1): 358–369.

Surles, J. and Padgett, W. 1998. Inference for $P(y < x)$ in the Burr Type X Model. *Communications in Statistics-Theory and Methods* 7 (4): 225–238.

Surles, J. and Padgett, W. 2001. Inference for Reliability and Stress-Strength for a Scaled Burr Type X Distribution. *Lifetime Data Analysis* 7 (2): 187–200.

Surles, J. and Padgett, W. 2005. Some Properties of a Scaled Burr Type X Distribution. *Journal of Statistical Planning and Inference* 128 (1): 271–280.

Tahir, M., Cordeiro, G. M. and Mansoor, M. 2015. The Kumaraswamy Pareto IV Distribution. *Austrian Journal of Statistics* .

Tsodikov, A., Ibrahim, J. G. and Yakovlev, A. 2011. Estimating Cure Rates From Survival Data. *Journal of the American Statistical Association* 11 (1): 5–7.

Uddin, L. Q., Molnar-Szakacs, I., Zaidel, E. and Iacoboni, M. 2006. Rtms to the Right Inferior Parietal Lobule Disrupts Self–Other Discrimination. *Social cognitive and affective neuroscience* 1 (1): 65–71.

Wahed, A. 2006. A General Method of Constructing Extended Families of Distributions From an Existing Continuous Class. *Journal of Probability and Statistical Science* 4: 165–177.

Wahed, A. S., Luong, T. M. and Jeong, J.-H. 2009. A new generalization of Weibull distribution with application to a breast cancer data set. *Statistics in medicine* 28 (16): 2077–2094.

Weisstein, E. W. 2004. Rayleigh distribution. *Sigma Journal of Engineering and Natural Sciences* 13 (3): 14.

Xiang, L., Ma, X. and Yau, K. K. 2011. Mixture Cure Model With Random Effects for Clustered Interval-Censored Survival Data. *Statistics in Medicine* 30 (9): 995–1006.

Yakovlev, A. Y., Asselain, B., Bardou, V., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A. 1993. A Simple Stochastic Model of Tumor Recurrence and Its Application to Data on Premenopausal Breast Cancer. *Biometrie Et Analyse De Donnees Spatio-Temporelles* 12 (3): 66–82.

Yin, G. and Ibrahim, J. G. 2005. Cure Rate Models: A Unified Approach. *Canadian Journal of Statistics* 33 (4): 559–570.

Yousof, H. M., Rasekhi, M., Afify, A. Z., Ghosh, I., Alizadeh, M. and Hamedani, G. 2017. The Beta Weibull-G Family of Distributions: Theory, Characterizations and Applications. *Pakistan Journal of Statistics* 33 (2): 3–8.

Zea, L. M., Silva, R. B., Bourguignon, M., Santos, A. M. and Cordeiro, G. M. 2012. The Beta Exponentiated Pareto Distribution With Application to Bladder Cancer Susceptibility. *International Journal of Statistics and Probability* 1 (2): 3–8.

Zhang, J. and Peng, Y. 2009. Accelerated Hazards Mixture Cure Model. *Lifetime Data Analysis* 15 (4): 301–335.

Zhao, X. and Zhou, X. 2008. Discrete-Time Survival Models With Long-Term Survivors. *Statistics in Medicine* 27 (8): 1261–1281.

Zografos, K. and Nadarajah, S. 2005. Expressions for Rényi and Shannon Entropies for Multivariate Distributions. *Statistics & Probability Letters* 71 (1): 71–84.