جامعــــــــــة باجــــي مختـــــــــار- عنابــــــة
BADJI MOKHTAR - ANNABA UNIVERSITY
UNIVERSITÉ BADJI MOKHTAR - ANNABA
كليــــــــــة العلـــــــــــوم
FACULTY OF SCIENCES
FACULTE DES SCIENCES
قسم الرياضيات
DEPARTMENT OF MATHEMATICS
DEPARTEMENT DE MATHEMATIQUES

THESE

Présentée en vue de l'obtention du diplôme de

Doctorat

Filière : Mathématiques Appliquées

Spécialité : Actuariat

Sur Les Distributions à Deux Paramètres : Censure et Application

Par:

CHOUIA Sarra

Sous la direction de

ZEGHDOUDI Halim	Prof.	U.B.M.Annaba

Co-encadreur

TOUATI Ali-Bey	MCB	U.B.M.Annaba

Devant le jury

Président :	Remita Med Riad	Prof.	U.B.M.Annaba
Examinateurs :	Benchettah Azzedine	Prof.	U.B.M.Annaba
	Ezzebsa Abdelali	MCA	U. Guelma

Année : 2020

DEDICATION

This thesis is lovingly dedicated to my parents whom always where there for encourage me in my hardest moments when there was no one, sisters, my daughter and my Soul mate my husband who supported me during those years.

Thank you

May Allah Bless Them

I didicate this thesis to…

My Mother RACHIDA

My Father Habib

My belover husband NASRO

My Sisters: BALKICE, SELSABIL and ZANOUBIA

My angel: JANA

CHOUIA SARRA

ACKNOWLEDGEMENTS

First and foremost, praise and thanks to **ALLAH** who helped in the times of weakness, reassurance in fear, and hope in the times of despair, guides me throughout my life and gave me the power and determination to accomplish this research work.

I sincerely whould like to thank and express my deeepeast appreciation to my guide and my supervisor **Prof. Zeghdoudi Halim** for all what i learnt from him throughout all my academics years. I really grateful for his continuous encouragement throughout my period of my project work. He was always motivating me to work harder and help me in defferent possible ways. Whithout his active and guidnance and great support, it whould not have been possible to complete this work. Lot of thanks for his patience, supervision and advices. May god reward him immensely. Also i'm deeply grateful to my co-supervisor **Dr. Touati Ali Bey** for his continuous support, his valuable time and encouragement wich added a great value to this study.

I whould like to thank the supervisory committee, for their time and valuable comments on my theis.

- **Ms. Remita Med Riad**. Professor at univeristy badji mokhtar-Annaba
- **Benchettah Azzedine**. Professor at univeristy badji mokhtar-Annaba
- **Ms. Ezzebsa Abdlali**. MCA at univeristy 08 May 1945-Guelma

I thank the referees for their valuable comments, which enrich the quality of my dissertation.

A spacial thanks to **my parents** for their endless love, support and encouragement to finish my PHD. Word cannot express how grateful i am to my lovely mother for your unconditional love, and patient. Your prayer was what sustained me this far. Thank you my father, for always being there for me whenever i needed you.

Thank you my sisters espacialy **Balkice** for helping and encouraging me in difficult times. At last, but not least, thank you Allah for giving me the sweetest angel, my daughter **Jana** that makes my life shiny when i just looking to her. Sincere thanks to my beloved husband **Nasro** for his love, care, continous support and many sacrificies wich made the completion of this work. I am really thankful to my friends and everyone who took part in exhibiting this work to light

Contents

List of Figures

List of Tables

Abstract

The aim of this thesis is to create a new families which are a special cases of *Two Parameters Distribution* under censored case; the first one is a general class of continuous lifetime distribution called "*Generalized Size biased distribution* " and the second is " *The XLindley distribution* "

Further, it can be generated a new lifetime distribution from *GSB* distribution such as **Size biased Zeghdoudi distribution** that we have presented it in the third chapter. The parameter of the new probability distribution function are estimated by the maximum likelihood method under type II censored data. In other hand, several mathematical properties have been studied such as: probability density function, moments, quantile and lambert W function...ect. The maximum likelihood function and Method of moment procedure has been provided to obtain the estimate of the model parameter of the two distribution in two cases: complete and censored.

Finally, Real data sets have been analyzed to illustrate the flexibility of the both new families**.**

Keywords: Size biased, moments, simulation, application, type II censored, Size biased Zeghdoudi distribution, XLindley distribution, LamberW function, quantile function

Résumé en Français

Dans cette thèse, nous introduisons des nouvelles distributions nommées: Size Baised et XLindley. Diverses propriétés statistiques ont été données à savoir: la méthode du moment, l'estimation du maximum de vraisemblance. Une simulation de biais et l'erreur quadratique moyenne des estimateurs sont obtenues par la méthode du maximum de vraisemblance. On a étudié l'estimateur de paramètre de la distribution size biased dans le cas censuré type II. Enfin, une étude comparative entre les distributions déjà connues à été discutée.

Mot-clés: Size biased, les moments, simulation, application, censuré de type II, Size biased Zeghdoudi distribtuion, distribution de XLindley, la fonction de LamberW, fonction de quantile.

الملخص

تلعب التوزيعات الإحصائية دورا هاما في تفسير البيانات الحقيقية أو الظواهر الحياتية لأي مجتمع و دراسة خصائصه . لذلك تهدف هذه الدراسة إلى تقديم عائلات جديدة استنادا على توازيع معروفة أي بالأحرى حالات خاصة من التوازيع ذات المعلمين و دراسة الخصائص الإحصائية و التطبيقية لهاته التوازيع و تقديم معالمهم .

تم تقديم عائلة جديدة سميت (Size biased on General distribution)و إيجاد صيغ مبسطة لها لكل من التوزيع الإحتمالي و دالة المخاطرة ، كذلك قدمنا توزيعا خاصا انطلاقا منها يسمى (Size biased Zeghdoudi ، تم دراسة الخصائص الإحصائية و تطبيق التوزيع الجديد على مجموعة بيانات حقيقة كاملة و مراقبة . ونظرا لأن معاينة خطط الحياة تشمل اختبار الحياة والذي هو عبارة عن وقت مستهلك فى معظم الاحيان ، لذلك فان خطة المراقبة توظف عادة لتقليل وقت الاختبار و التكاليف وذلك من خلال البيانات أو العينات المراقبة(censored data)، كذلك تطرقنا إلى توزيع جديد مستنبط من توزيع زغدودي يسمى XLindley distribution الذي يعتمد على معلم (وسيط) و هو حالة خاصة من توزيع Two parameter lindleyذات المعلمين ، حيث تم تحديد خوائصه الإحصائية المختلفة

و أخيرا في نهاية كل جزء أجرينا دراسة محاكاة و تحليل للبيانات الحقيقة و مقارنة مع توازيع أخرى لتوضيح كفاءة ومرونة كل توزيع بالتطبيق العملي على البيانات الحقيقة .

Introduction en Français

La distribution de Lindley a été introduite par Lindley (1958) en tant que nouvelle distribution utile pour analyser les données sur la durée de vie mais, malgré le peu d'attention dans la littérature statistique, elle est importante pour étudier la modélisation de la fiabilité de la résistance aux contraintes. Ghitany et al (2008) ont discuté diverses propriétés de cette distribution et ont montré qu'à bien des égards, elle fournit un meilleur modèle pour certaines applications que la distribution exponentielle. Aussi, ils ont montré dans un exemple numérique que la distribution de Lindley donne une meilleure modélisation que la distribution exponentielle pour des données sur les temps d'attente et les temps de survie.

Par ailleurs, certains chercheurs ont proposé de nouvelles classes de distributions basées sur des modifications de la distribution Lindley.

Dans ce travail, nous nous proposons deux nouvelles généralisations de la distribution de Lindley. Nous nous référons à cette nouvelle généralisation à savoir la Size biased zeghdoudi distribution (SBZD) et la distribution XLindley distribution (XLD). Elles offrent plus de flexibilité pour analyser des ensembles de données réels complexes. De plus, nous étudions certaines propriétés statistiques de la nouvelle distribution. Les objectifs de ce travail sont:

- Construire une nouvelle classe de distributions (SBZD) de type Lindley;

- Établir les propriétés mathématiques de cette nouvelle classe (SBZD);

- L'étude de la fiabilité résistance-résistance de cette nouvelle classe (SBZD);

- Introduire et étudier les propriétés mathématiques de la distribution XLindley à un paramètre;

- Donner quelques exemples d'ensemble de données sur la durée de vie provenant de différents domaines de connaissances ont été pris en considération;

- Étudier la qualité de l'ajustement de ces distributions pour voir la supériorité sur les autres distributions.

Sankaran (1970) a introduit la distribution discrète de Poisson-Lindley en combinant les distributions de Poisson et Lindley. Ghitany et al.(2008) ont effectué une recherche sur les propriétés statistiques de la distribution de Lindley, et ils ont montré que la distribution peut fournir un meilleur ajustement que la distribution exponentielle. Mahmoudi et Zakerzadeh

(2010) ont proposé une version étendue de la distribution de Poisson composée obtenue en combinant la distribution de Poisson avec la distribution de Lindley généralisée et analysé par Zakerzadeh et Dolati (2010). Récemment, une nouvelle extension de la distribution Lindley, appelée distribution de Zeghdoudi (ZD), qui offre un modèle flexible pour les données de durée de vie, est introduite par Zeghdoudi et Messaadia(2018). Nedjar et Zeghdoudi (2016, 2016a) ont introduit une distribution de durée de vie à deux paramètres avec un taux d'échec décroissant en combinant des distributions gamma et Lindley, appelées la distribution de gamma Lindley (GaL). De la même manière, Zeghdoudi et Nedjar (2016, 2017) et Lazri et al.(2016, 2018), Grine et Zeghdoudi (2017) ont introduit les distributions de pseudo Lindley (PsL), Poisson pseudo Lindley (PPsL) et Lindley Pareto(PL), Poisson Quasi-Lindley (PQL) respectivement. Bouchahed et Zeghdoudi (2018) ont présenté une méthode pour introduire une famille à un paramètre avec application aux familles exponentielles. Récemment, Benatmane et al.(2020) ont introduit une classe de distributions, nommée distributions Rayleigh-Pareto composée (CRP), où la procédure de composition suit la même démarche que précédemment réalisé par Lazri et al.(2017).

Des extensions de la distribution de Lindley ont été obtenues par Nedjar et Zeghdoudi (2016, 2020), Bouchahed et Zeghdoudi (2018), Seghier et al.(2020,2020a). Cette nouvelle classe de distribution a été a reçu une attention considérable par les chercheurs dans ce domaine.

Le but de notre travail est de présenter des nouvelles distributions à un paramètre qui apportera un plus à la littérature existante sur la modélisation des données de survie, dans les sciences biologiques et les sciences actuarielles. Nous avons structuré cette thèse autour de cinq chapitres.

Le chapitre I est une introduction contient le contexte historique, des motivations et les objectives principales de cette thèse. Le chapitre II se focalise sur quelques définitions et concepts de base qui sont nécessaires pour les chapitres suivants.

Le chapitre III consiste en une synthèse sur les nouvelles distributions à deux paramètres à savoir :

La distribution de Lindley à deux paramètres;

La distribution Sujatha à deux paramètres;

La distribution de gamma Lindley.

Au chapitre IV, nous avons introduit la première distribution à un paramètre nommée la distribution Zeghdoudi biaisée et étudié ses propriétés mathématiques.

Enfin, le chapitre V étudier la deuxième nouvelle distribution de type Lindley nommée distribution XLindley. Ses propriétés mathématiques et statistiques sont discutées, y compris la fonction quantile, les moments, l'ordre stochastique et l'estimation des paramètres. De plus, des études de simulation et des exemples illustratifs sur les deux nouvels modèles sont effectués en utilisant des données réelles.

Chapter 1

Introduction

(1.1)Preliminaries

Recently, Many authors has been interested in introducing and defining new families of probability distribution or generalized class of continuous distribution by : mixture of two known distribution also by size biasing known distributions or by transformation, at the some time they have to provide more flexibility for modeling lifetime data. The flexibility of such new families of distributions come in terms of one or more failure rate which may be decreasing or increasing or bathtub shaped.

The researchers used many distributions in survival analysis for analyzing data involving the duration between two events in various fields, the aims of survival analysis is to estimate "survivorship, density, and hazard" functions. The key characteristic that distinguishes survival analysis from other areas in statistics is that survival data are usually censored or incomplete in some way.

The censored data is a key issue for the analysis of survival data, censoring occurs when incomplete information is available about the survival time of some individuals. In this thesis, we present a number of concrete examples extracted from the literature in various fields of public health. The purpose is to use real-life situations to illustrate types of censoring and to motivate the discussion presented in the first chapter.

According to **Miller(1998)** and **Hougaard(2000)** data are said to be censored if the observation time censored survival is only partial, not until the failure event. One major reason for this is that the person studied is alive when the data are evaluated, and thus the complete lifetime is not known at that time. There are many other reasons for censoring. For examples, the patients can be lost to follow-up, patients still alive at the end of the study or patients drop out of the study. There are also several types of censoring, including right censoring, left censoring, interval censoring, random censoring, Type I censoring and Type II censoring (**Collet 1994; Hosmer and Lemeshow 1999; Kalbfleisch and Prentice 2002; Kleinbaum and Klein 2005**).

(1.2) Motivation and Objectives

In the past few years, researchers use the censored data in modeling data in practice to help them finish the study early. So the work objectives are:

- Building a new generating families of distributions on the basis of it obtaining the results of this thesis.

- Some mathematical properties of the new family and estimation of the unknown parameters are derived also the behavior of the hazard rate function of some special cases from this family are explored.

- Study several properties of the new distributions .

- Estimate the unknown parameter of each distributions under complete data & censored data for SBZD.

- In the last contribution, a new distribution has been introduced which is called XLindley distribution.

- Application to real data set (examples of lifetimes data-sets) is given to show the flexibility and potentiality of the new families.

- Study the goodness of fit for each distribution to see the superiority of one over the other known distributions

(1.3) Organization of the Thesis

The thesis is organized as follows. Chapter I is an introduction to the current thesis. Also, historical background of the preceding researchers is reviewed. Chapter II contains some basic definitions and concepts that are needed in the next chapters

In Chapter III, we have introduce already known distributions of two parameters, and we have noticed that we can get our new distributions as special cases from some of them .

In Chapter IV, the Size biased zeghdoudi distribution is introduced as a new member from

Size biased general-one parameter distribution. Closed-form expressions for the density, cumulative distribution, survival and failure rate functions are obtained. The rth moment of the new distribution and lorenz curve are obtained. Estimation of the unknown parameters using the method of moment and maximum likelihood method under censored data is obtained. In addition an application to a real data set demonstrates the usefulness of the new model.

Finally, in Chapter V, we introduce a new Lindley type distribution named XLindley distribution. The properties of it are discussed, including quantile function, moments and stochastic ordering. The estimation of the model parameter is performed by using two methods including the maximum likelihood method. In particular, some mathematical properties of the new distributions are discussed. Simulation studies of the two models are performed. The potentially of them is illustrated based on a real data set.

Chapter 2

Basic concepts and models

2.1 Analyse of Survival data

In a survival analysis, we usually refer to the time variable as **survival time**, because it gives the time that an individual has "survived" over some follow-up period. We also typically refer to the event as a **failure**, because the event of interest usually is "death, disease incidence", or some other **negative individual experience**. However, survival time may be "time to return to work after an elective surgical procedure," in which case **failure is a positive event**.

2.1.1 Introduction

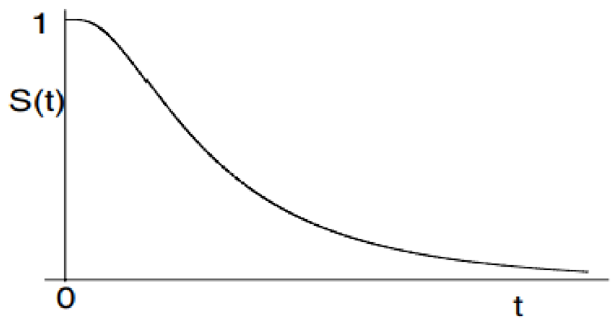
Lifetime distribution methodology is widely used in biomedical and engineering sciences, however, and most of examples in the thesis come from those areas.

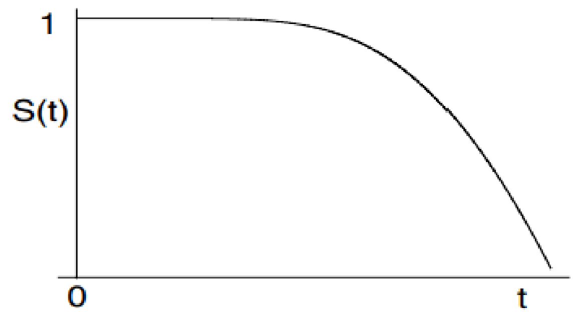
Sometimes the events are actual deaths of individuals and "lifetime" is the length of life measured from some particular starting point. In other instances "lifetime" and the words "death" or "failure", which denote the event of interest, are used in a figurative sense. In discussing applications, other terms such as "survival time" and "failure time" are also frequently used.

In the survival analysis data, we want to use all available data sets, which sometimes are incomplete or include uncertainty as to when a failure occurred. Life data can therefore be separated into two types: complete data (all information is available) or information is missing so we called (censored or trucation data). Each type is explained next.

2.1.2 The Goals of survival analysis

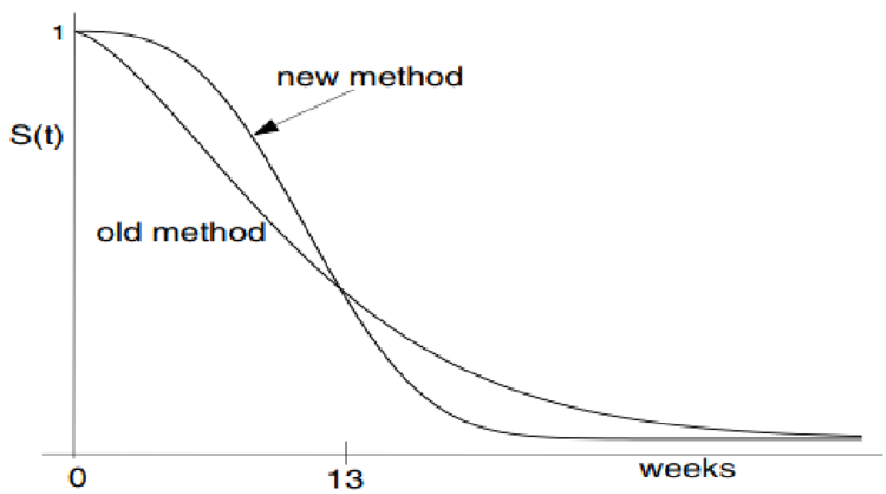
Goal 1. To estimate and interpret survivor function, and/or hazard function from survival data



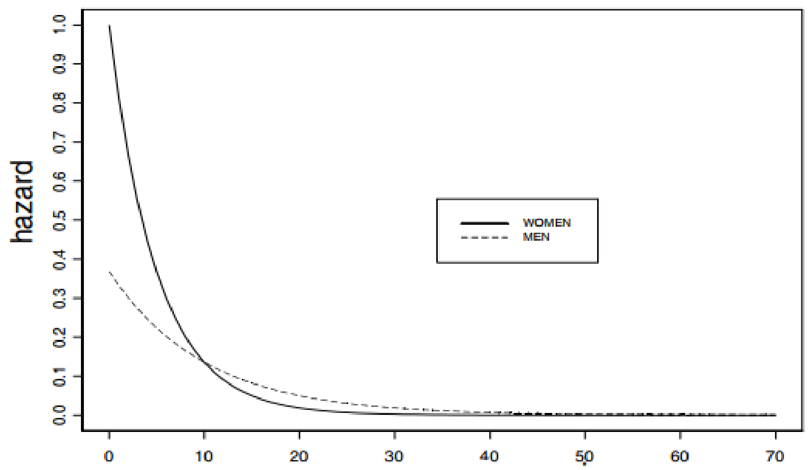


- The graph is smooth curve, decreasing from $S(t) = 1$ at time $t = 0$ to $S(t) = 0$ at $t = \infty$

Goal 2.To compare survivor and/or hazard functions.



Goal 3. To assess the relationship of explanatory variables to survival time



Age at diagnonis(years)

2.2 Examples of Survival Data

The following examples illustrate some ways in which lifetime data arise.

2.2.1 Time to First Use of Marijuana

Turnbull and Weiss (1978) report part of a study conducted at the Stanford-Palo Alto Peer Counseling Program. In this study, 191 California high school boys were asked:**"When did you first use marijuana?"**. The answers were the exact ages (uncensored observations); "I never used it," which are right-censored observations at the boys' current ages; or "I have used it but can not recall just when the first time was," which is a left-censored observation. Notice that a left-censored observation tells us only that the event has occurred prior to the boy's current age. The data is in Table1.1

Age	Number of exact observations	Number of who have not yet smoking it	Number of who have started smoking at earlier age
10	4	0	0
11	12	0	0
12	19	2	0
13	24	15	1
14	20	24	2
15	13	18	3
16	3	14	2
17	1	6	3
18	0	0	1
$\succ 18$	4	0	0

Table1.1.Marijuana use it in high school boys

Another illustrate table to the answers of the boys :

The Answers	Recorded Value
a:I used it but I cannot recall when it was my first time	**a:** T^{-}: age at interview as exact age was earlier but unknown (L.C)
b:I used it when I was....	**b:**T: exact age since it is known(uncensored)
c:I never used it	**c:**T^{+}:age at interview since exact age occurs sometime in the future (R.C)

2.2.2 Times to Death for a Breast-Cancer Trial

In the study of (Sedmak et al., 1989) designed to determine if female breast cancer patients, originally classified as lymph node negative by standard light microscopy (SLM), could be more accurately classified by immunohistochemical (IH) examination of their lymph nodes with an anticytokeratin monoclonal antibody cocktail, identical sections of lymph nodes were sequentially examined by SLM and IH. The significance of this study is that 16% of patients with negative axillary lymph nodes, by standard pathological examination, develop recurrent disease within 10 years. From the Ohio State University Hospital Cancer Hospital registry, have been selected 45 females breast-cancer patients with negative axillary lymph nodes and a minimum 10 year follow-up. Of the 45 patients, 9 were immunoperoxidase positive, and the remaining 36 remained negative. Survival times (in months) for both groups of patients are given in Table1.2 (denotes a censored observation).

immunoperoxidase positive:	$22; 23; 38; 42; 73$
$77; 89; 115; 144^{+}$	
immunoperoxidase negative:	$19; 25; 30; 34; 37$
$46; 47; 51; 56; 57; 61; 66; 67; 74; 78; 86; 122^{+};$	$123^{+}; 130^{+}; 130^{+}; 133^{+}$
$134^{+}; 136^{+}; 141^{+}; 143^{+}; 148^{+}; 151^{+}; 152^{+}; 153^{+};$	$154^{+}; 156^{+}; 162^{+}; 164^{+};$
$165^{+}; 182^{+}; 189^{+}$	

$$^{+}\text{:Censored observation}\begin{cases} \longrightarrow \text{In remission at the end of study} \\ \longrightarrow \text{Lost to follow-up} \\ \longrightarrow \text{Withdraws} \end{cases}$$

Table 1.2. Times to death (in months) for breast cancer patients with different immunohistochemical responses

2.2.3 Death Times of Psychiatric Patients

Woolson (1981) has reported survival data on 26 psychiatric inpatients admitted to the University of Iowa hospitals during the years 1935–1948. This sample is part of a larger study of psychiatric inpatients discussed by Tsuang and Woolson (1977). Data for each patient consists of age at first admission to the hospital, sex, number of years of follow-up (years) from admission to death or censoring, and patient status at the follow-up time. Here, a comparison of the survival experience of these 26 patients is made to the standard mortality of residents of Iowa to determine if psychiatric patients tend to have shorter lifetimes(see Table1.3)

Gender	Age at admission	Time of following-up
Female	51	1
Female	58	1
Female	55	2
Female	28	22
Male	21	30^+
Male	19	28
Female	25	32
Female	48	11
Female	47	14
Female	25	36^+
Female	31	31^+
Male	24	33^+
Male	25	33^+
Female	30	37^+
Female	33	35^+
Male	36	25
Male	30	31^+
Male	41	22
Female	43	26
Female	45	24
Female	35	35^+
Male	29	34^+
Male	35	30^+
Male	32	35
Female	36	40
Male	32	39^+

$^+$Censored observation

Table1.3.Survival data for psychiatric inpatients

2.3 Censoring and Truncation

Time-to-event data present them selves in different ways which create special problems in analyzing such data, there is two features may be present in some survival studies

The $first\ one$ often presented in survival data is known as: **Censoring.** There are various categories of censoring will be discussed in this section. Many several types of censoring schemes within both left and right lead to a different likelihood function as we shall see in this section.

The *second one* often presented in survival data is known as: **Truncation**

2.3.1 Truncated data

Were first encountered quite early in the development of modem statistics by Sir Francis Galton (1897) in connection with an analysis of registered speeds of American trotting horses.Truncated samples are classified according to whether terminals points of truncation are known or unknown. When these points are unknown, they become additional parameters to be estimated from sample data. Truncation of survival data occurs when only those individuals whose event time lies within a certain observational window (Y_L, Y_R) are observed

Right truncated data

When $Y_L = 0$, the right truncation occurs. That is, we observe the survival time X only when $X \leq Y_R$.

Some time right truncation arises:

- In estimating the distribution of stars from the earth in that stars too far away are not visible and are right truncated.

- When it's a mortality study based on death records.

Left truncated data

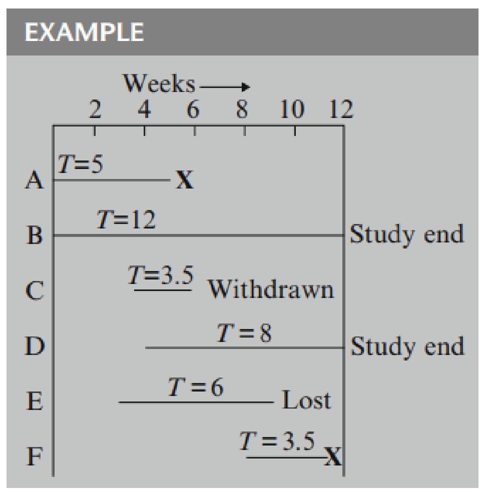
Here we only observe those individuals whose event time X exceeds the truncation time Y_L only if Y_R is infinite. In this type of truncation any subjects who experience the event of interest prior to the truncation time are not observed. The truncation time is often called a delayed entry time since we only observe subjects from this time until they die or are censored.

2.3.2 Censored data

There are three reasons why censoring may occur:

1. A person doesn't experience the event before "the study end"
2. A person is"lost to follow-up" during the study period
3. A person withdrawsd from the study

The graph below illustrate the experience of several persons followed over time.



X: Events occurs

For :

- Person A: He was followed from the beging of the study until the the 5th week when the event occur, so his survival time is 5 weeks and isn't consored

- P*erson B*: Also was followed from the start of the study until the end for 12th week without getting the event, we can say that the survival time here is censored only at least for 12 weeks

- Person C: This one enters the study between the second and third week and was followed until he withdraws at the 6th week, he's survival timle is censored after 3.5 weeks

- Person D: Enters the study at the 4th week and was followed for the remainder time of the study with out getting the event, so this person's censored time is 8 weeks

- Person E: Enters the study at week 3 and is followed until 9thy week, when he is lost to follow-up without reached to the failure time; his censored time is 6 weeks.

- Person F: Enters at the 8th week and was followed until getting the event at week 11.5. As with person A, there is no censoring here; the survival time is 3.5 weeks.

The table(1.4) of the survival time data for the six persons in the graph is now presented for each person. We have indicated whether each time was censored or not with event occurs: (1 denoting failed and 0 denoting censored). This table is a simplified illustration of the type of data to be analyzed in a survival analysis.

Person	Survival time	Event occurs
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1

Table1.4. The type of data analyzed .

Types of censored data:

There are many types of censored data :

Right-censored: Notice that for the 4 persons censored (B, C, D, E), we know that the person's true survival time becomes incomplete at the right side of the follow-up period, occurring when the study ends or when the person is lost to follow-up or is withdrawn. We denote this kind of data: **right-censored data**; this is the common survival data .

For these data, the complete survival time interval, which we don't really know, has been cut off (censored) at the right side of the observed survival time interval.

Left-censored : Time to First Use of Marijuana. See $Table1.1$

If we are following persons until they become **HIV positive**, we may record a failure when a subject first tests positive for the virus. However, we may not know the exact time of first exposure to this virus, and therefore do not know exactly when the failure occurred(event occurs). Thus, the survival time is censored on the left side since the true survival time, which ends at exposure, is shorter than the follow-up time, which ends when the subject's test is positive.

In other words, if a person is **left-censored** at time t, we know that they had an event between time 0 and t, but we don't know the exact time of event. The graph below explain this type.

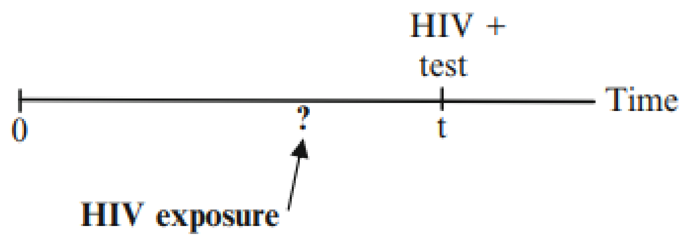


Fig1. The time of exposure the virus of HIV

Interval-censored: Survival analysis data can also be **interval censored**, which can occur

if a subject's true (but unobserved) survival time is within a certain known specified time interval. As an example, again considering HIV surveillance, a subject may have had two HIV tests, where he/she was HIV negative at the time (say, t_1) of the first test and HIV positive at the time (t_2) of the second test. In such a case, the subject's true survival time occurred after time t_1 and before time t_2, i.e, the subject is interval censored in the time interval $[t_1,\, t_2]$.The Figure below illustrate it :

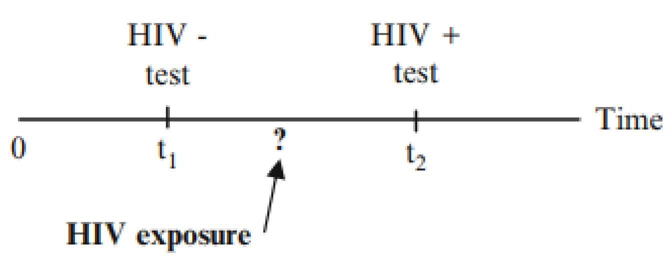


Fig.2.Time of exposure to HIV

This type of censored data incorporates both right-censoring and left-censoring as special cases. Left-censored data occur whenever the value of t_1 is 0 and t_2 is a known upper bound on the true survival time. In contrast, right censored data occurs whenever the value of t_2 is infinity, and t_1 is a known lower bound on the true survival time.

$$\begin{aligned}\text{Left cesoring} \quad &\Rightarrow \quad t_1 = 0, t_2 = \text{upper bound} \\ \text{Right cesoring} \quad &\Rightarrow \quad t_1 = \text{Lower bound}, t_2 = 0\end{aligned}$$

Singly censored data

Type I : we will consider Type I censoring where the event is observed only if it occurs prior to some pre-specified time. These censoring times may vary from individual to individual. In typical experiment on animals or clinical trial starts with a fixed number of animals or patients to which a treatment (or treatments) is (are) applied.

The investigator will terminate this study or report the results before all subjects realize their events because of **time or cost considerations**. In this instance, if there are no accidental losses or subject withdrawals, all censored observations have times equal to the length of the study period.

- If $(X \leq Cr)$ then we can know the exact lifetime X of an individual
- If $(X \succ Cr)$, the individual is a survivor, and his or her event time is censored at Cr

The data from this experiment can be conveniently represented by pairs of random variables(T, δ), where indicates whether the lifetime X corresponds to an event ($\delta = 1$) or is censored ($\delta = 0$), and T is equal to X, if the lifetime is observed, and to Cr if it is censored, i.e:

$$T = min(X, Cr)$$

Type II Experiments involving Type II censoring are often used in testing of equipment life. Here, all items are put on test at the same time, and the test is terminated when $r($ is some predetermined integer $(r \prec n))$. of the n items have failed. Such an experiment may save time and money because it could take a very long time for all items to fail. It is also true that the statistical treatment of Type II censored data is simpler because the data consists of the r smallest lifetimes in a random sample of n

Random censoring This is the 3^{rd}type of censored data. It occur when both the number of censored observations and the censoring levels are random outcomes, with another meaning the period of study is fixed and patients enter the study at different times during that period. Hence the censored times also are different. This type of censoring commonly arises in medical time-to-event studies. A subject who moves away from the study area before the event of interest occurs has a randomly censored value.

2.4 The Survival Function

For specified time t, the survival function $S(t)$ is :

$$S(t) = 1 - F_X(t) = P(X > t) \qquad , t \geq 0$$

2.5 The Hazard Function

The Hazard function is basic quantity, fundamental in survival analysis. This function is also known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic processes. The hazard rate is defined by:

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbf{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t}$$

If X is a continuous random variable, then

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \ln[S(x)]}{dx}$$

A related quantity is the cumulative hazard function $H(x)$, defined by:

$$H(x) = \int_0^x h(u)du = -\ln[S(x)]$$

Thus, for continuous lifetimes:

$$S(x) = \exp[-H(x)] = \left[-\int_0^x h(u)du\right]$$

2.6 Quantities associated with survival distribution

Let be X a real random variable defined on I, with F_X cumulative function .

2.6.1 Moments about the origin (raw moments)

The rth moment about the origin of a random variable X, denoted by $\mu_r^{\shortmid}$ is the expected value of X^r:

$$\mu_r^{\shortmid} = E\left(X^r\right) = \left\{ \begin{array}{cc} \sum\limits_x x^r f(x) & when\ X\ is\ discrete \\ \int_{-\infty}^{+\infty} x^r f\left(x\right)\ dx & when\ X\ is\ continuous \end{array} \right\}$$

For: $r = 0, 1,\ 2, ...$

2.6.2 Central moments

The rth moment about the mean of a random variable X, denoted by μ_r, is the expected value of $(X - \mu_X)^r$symbolically:

$$\mu_r = E\left(\left[X - \mu_X\right]^r\right) = \left\{ \begin{array}{cc} \sum\limits_x \left[x - \mu_X\right]^r f(x) & when\ X\ is\ discrete \\ \int\limits_{-\infty}^{+\infty} \left[x - \mu_X\right]^r f\left(x\right)\ dx & when\ X\ is\ continuous \end{array} \right\}$$

For : $r = 0, 1, 2, ...$

the **first moment** is the expected value $E[X]$. The **second central moment** is the $Var(X)$. Similar to mean and variance, other moments give useful information about random variables.

2.6.3 Moment Generating Functions

The moment generating function (MGF) of a random variable X is a function $M_X(t)$ defined as:

$$M_X(t) = E\left(e^{Xt}\right), t \in R.$$

We say that MGF of X exists, if there exists a positive constant a such that $M_X(t)$ is finite for all $t \in [-a, a]$.

This relation makes it possible to calculate very easily the moments of a distribution if the MGF known. For example:

Mean and variance of X:

$$\begin{aligned} E(X) &= M_X^{'}(0) \\ Var(X) &= E(X^2) - E^2(X) = M_X^{''}(0) - \left[M_X^{'}(0)\right]^2 \end{aligned}$$

Coefficient of variation (γ), Skwness$\left(\sqrt{\beta_1}\right)$ and kurtosis (β_2) are:

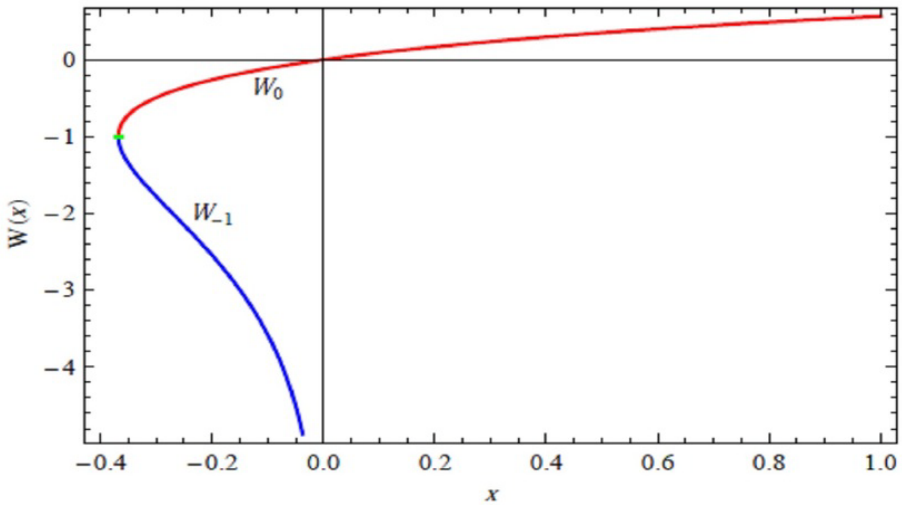
$$\begin{aligned} \gamma &= \frac{\sqrt{Var(X)}}{\mathbb{E}(X)} \\ \sqrt{\beta_1} &= \frac{\mathbb{E}(X^3)}{(Var(X))^{\frac{3}{2}}} \\ \beta_2 &= \frac{\mathbb{E}(X^4)}{(Var(X))^2} \end{aligned}$$

2.7 Lambert W function

The Lambert W function is a multivalued complex function defined as the solution of the equation:

$$W(z) \exp(W(z)) = z. \tag{1.1}$$

Where : z is a complex number



W_{-1} denoted the negative branch. The real branch taking on values in$(-\infty, -1]$

W_0 denoted the principale branch.The real branch taking on values in[-1, ∞)

Both real branches of W are illustrated in Figure above.

Lemma 1 *Let a, b and c be a fixed complex numbers. The solution of the equation $z + abz = c$ with respect to $z \in C$ is :*

$$z = c - \frac{1}{\log(b)} W\left(ab^c \log(b)\right).$$

Where W denotes the Lambert W function.

By multipliying the both sides of this equation on $b^c \log(b)$, we get the following resulting eqation

$$(c - z)\log(b)\exp((c - z)\log(b)) = ab^c \log(b). \tag{1.2}$$

According to equation (1.1) and (1.2).We observed that $(c - z)\log(b)$ is the Lambert W function of the complex argument $ab^c \log(b)$. Therefore, we have:

$$W\left(ab^c \log(b)\right) = (c - z)\log(b).$$

which implies the desired result. This completes the proof of $Lemma 1$

2.8 Quantile Function

The quantile function of random variable X is the inverse of it's cumulative probability function (CDF). We denote the quantile function of X by Q_X, with $u \in [0,1]$:

$$Q_X(u) = F_X^{-1}(u)$$

$$Q_X(u) = \inf\{x \text{ tel que } F_X(x) \geq u\} \quad 0 < u < 1 \tag{1.3}$$

2.9 Estimation

Estimator of unkown parameter θ of probabilty model depends on the random sample. By assuming x is a random variable with probability density function $f(x;\theta)$, and if $x_1, x_2, ..., x_n$ is a random sample of size n from x, then the statistic is used for estimating the unknown parameter θ is called a point estimator :

$$\hat{\theta} = f(x_1, x_2, ..., x_n)$$

There are several methods are used to obtain a point estimates (estimator of θ) for the unknown parameters of a given probability distribution like :

- The Method of Moment
- Maximum Likelihood Estimation

2.9.1 Quantified properties

Bias

The bias of $\hat{\theta}$ is defined as:

$$Biais\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$$

It is the distance between the average of the collection of estimates, and the one-parameter being estimated.

- The estimator $\hat{\theta}$ is an **unbiased** estimator of θ if and only if $Biais\left(\hat{\theta}\right) = 0$
- The estimator is positively biased if $Biais\left(\hat{\theta}\right) > 0$

Convergence

The estimator $\hat{\theta}_n$ is convergent if it's converges in probability$\rightarrow \theta$, soit:

$$\begin{aligned} \hat{\theta}_n \xrightarrow{P} \theta &\iff \lim_{n\rightarrow\infty} P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) \rightarrow 1, \forall\varepsilon > 0 \\ &\iff \lim_{n\rightarrow\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \rightarrow 0 \end{aligned}$$

Theorem 2 *Any unbiased estimator is convergent if it's variance tends towards* 0 *:*

$$\left(E\left(\hat{\theta}_n\right) = \theta \ et \ Var\left(\hat{\theta}_n\right) \rightarrow 0\right) \Rightarrow \hat{\theta}_n \xrightarrow{P} \theta$$

Mean Squared Error

The mean squared error (*Erreur quadratique moyenne in french*)of $\hat{\theta}$ is defined as the expected value (probability-weighted average, over all samples) of the squared errors; that is,

$$MSE\left(\hat{\theta}\right) = E\left(\left(\hat{\theta}_n - \theta\right)^2\right)$$

2.9.2 Asymptotic normality

Let $X_1, X_2, \ldots, X_n$ be a random sample of n size that is, a sequence of i.i.d random variables drawn from a distribution of expected value given by μ(mean) and finite variance given by σ^2.

Theorem central limite

Let $X_1, X_2, \ldots, X_n$ be a set of iid random variables with the same distribution as X, $E(X_i) = m$ and $Var(X_i) = \sigma^2$with $i = 1, ...n$

$$Z_n = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N\left(0, 1\right).$$

The random variable Z_n is converges in law (L) to a Normal reduced centered with *zero* mean and variance $Var(X)$.

$$\sqrt{n}\left(\bar{X}_n - m\right) \xrightarrow{p} N\left(0, \sigma^2\right).$$

The Delta Method

The delta methods used to obtain the asymptotic distribution of a non-linear function of a random variable (usually, an estimator). It uses a first-order Taylor series expansion and Slutsky's theorem.

Let $X_1, X_2, \ldots, X_n$ be random variables with mean θ and standard deviation σ. If $\sqrt{n}\left(X_n - \theta\right) \xrightarrow{p} N\left(0, \sigma^2\right)$, and g differentiable function such that $g^{'}\left(\theta\right) \neq 0$. In this case the method of delta is given by:

$$\sqrt{n}\left(g\left(X_n\right) - g\left(\theta\right)\right) \xrightarrow{p} N\left(0, \sigma^2\left[g^{'}\left(\theta\right)\right]\right).$$

Let $\hat{\theta}$ estimator of θ of the law P_θ of an observed random variable X. We suppose that there are two sequences of real positive functions strictly positive, $a = a_n\left(\theta\right)$ and $b = b_n(\theta)$ as:

$$\frac{\hat{\theta} - a}{b} \xrightarrow{p} N\left(0, 1\right).$$

Then, we can say that $\hat{\theta}$ is an asymptotically normal estimator.

2.9.3 Construction of estimators

Method of Moments

Therefore, if $\theta = E(X)$, then the estimator of θ with method of moment is :

$$\hat{\theta}_n = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

More generally, for $\theta \in \Theta$, if $E(X) = \varphi(\theta)$, with φ is an invertible function, then the estimator of θ with method of moment is:

$$\hat{\theta}_n = \varphi^{-1}(\theta).$$

Similarly, we estimate the variance of the X_i by the empirical variance of the sample $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

Method of Maximum Likelihood

Let X_i be independent random variable and of the same rule. The likelihood function given by:

$$L(\theta; x_1, ..., x_n) = \begin{cases} \prod_{i=1}^{n} P\left(X = x_i; \theta\right) & if\ X_i\ are\ discrete. \\ \prod_{i=1}^{n} f_X\left(x_i; \theta\right) & if\ X_i\ is\ continuous. \end{cases}$$

The estimator of maximum likelihood of θ is $\hat{\theta}_n$, which maximizes the likelihood function $L(\theta; x_1, ..., x_n)$. The estimator of maximum likelihood (ML)of θ is the corresponding random variable. So generally $\hat{\theta}_n$ will be calculated by maximizing the log-likelihood:

$$\hat{\theta}_n = argmax\ \log L(\theta; x_1, ..., x_n).$$

When $\theta = (\theta_1, ..., \theta_d) \in \Theta$ and that all the partial derivatives below exist, $\hat{\theta}_n$ is the solution of the system of equations that called likelihood equations:

$$\forall j \in \{1, ..., d\},\ \frac{\partial}{\partial \theta_j} \log L(\theta; x_1, ..., x_n) = 0.$$

$$where \quad \frac{\partial^2}{\partial \theta_j^2} \log L(\theta; x_1, ..., x_n) < 0.$$

In this case, we solve it by numerical methods, like the Fisher Scoring method.

2.10 Some Measures of Statistical Models

When a number of models are fit to the same data set, we need a method to select the best fitted model for the data. In the next subsections Akaike information criterion (AIC) and Bayesian information criterion (BIC) are discussed respectively.

2.10.1 Akaike information criterion

AIC is the most widely selection model from a set of models. Akaike (1973) introduced an information criterion, or Akaike's information criterion which discerns how chose a fitted model is to the generating or the true model. The idea of is to select the model that minimizes the negative likelihood penalized by the number of parameters as specified in the following equation

$$AIC = 2p - 2LL(x; \hat{\theta}) \tag{1.4}$$

where, p is the number of the estimated parameters in the model and $LL(x; \hat{\theta})$ is the maximized value of log likelihood function in the model. The model with the minimum is chosen as the best to fit the data.

2.10.2 Consistent Akaike information criterion

Bozdogan (1987) proposed a corrected version of AIC in an attempt to overcome the tendency of the AIC to estimate the complexity of the underlying model. Bozdogan (1987) observed that the AIC does not directly depend on sample size n and as a result lacks certain properties of asymptotic consistency. In formulating, the consistent Akaike information criterion ($AICc$) is a correction factor based on the sample size and it is defined as:

$$AICc = -2\log(L) + p[(\log n) + 1]$$

Models that minimize the consistent Akaike information criteria are selected.

2.10.3 Bayesian information criterion

Akaike (1978) and Schwarz (1978) suggested another measure to compare between fitted models called Bayesian information criterion. AIC depends on estimated negative likelihood function and number of the estimated parameter. While BIC penalizes negative likelihood by adding the number of estimated parameters multiplied by the log of the sample size (see Jones (2011). BIC can be obtained by:

$$BIC = p\log(n) - 2LL(x;\hat{\theta}) \tag{1.5}$$

where, p is the number of the estimated parameters in the model, $LL(x;\hat{\theta})$ is the maximized value of log likelihood function in the model and n is the sample size. Models that minimize the Bayesian information criterion are selected. The relation between BIC and AIC criteria can be obtained from (1.4) and (1.5) as follows:

$$BIC = p\log(n) + 2p - AIC$$

.

Chapter 3

A Two Parameters Distributions

In this chapter, we introduce some distributions of two unknown parameters, wich our new distributions are extracted from them.

3.1 Two parameter Lindley distribution$(TPLD)$

This distribution introduced by [33]. A two-parameter Lindley distribution with parameters α and θ is defined by its probability density function (p.d.f)

$$f(x;\alpha,\theta)=\frac{\theta^2(\alpha+x)e^{-\theta x}}{\alpha\theta+1};\ x>0,\ \theta>0,\ \alpha\theta>-1 \tag{1.6}$$

It can easily be seen that at $\alpha=\theta$, the distribution (1.6)reduces to the Lindley distribution and at $\alpha=0$, it reduces to the gamma distribution with parameters $(2,\theta)$, The p.d.f. (1.6) can be shown as a mixture of exponential$(\theta)$and *gamma* $(2,\theta)$distributions as follows:

$$f(x;\alpha,\theta)=pf_1(x)+(1-p)\,f_2(x)$$

Where $p=\frac{\alpha\theta}{\alpha\theta+1}$, $f_1(x)=\theta e^{-\theta x}$ et $f_2(x)=\theta^2 x e^{-\theta x}$

The cumulative distribution function of the distribution is given by:

$$F(x)=1-\frac{1+\alpha\theta+\theta x}{\alpha\theta+1}e^{-\theta x};x>0,\theta>0,\alpha\theta>-1. \tag{1.7}$$

The rth moment about origin of the two-parameter Lindley distribution has been obtained as:

$$\mu_r^{\prime} = E\left(X^r\right) = \frac{\Gamma\left(r+1\right)\left(\alpha+r+1\right)}{\theta^r\left(\alpha+1\right)},\ r = 1, 2, ...$$

By taking $r = 1, 2, 3$, we obtain :

$$\mu_1^{\prime} = \frac{\left(\alpha\theta+2\right)}{\theta\left(\alpha\theta+1\right)}, \mu_2^{\prime} = \frac{2\left(\alpha\theta+3\right)}{\theta^2\left(\alpha\theta+1\right)}, \mu_3^{\prime} = \frac{6\left(\alpha\theta+4\right)}{\theta^3\left(\alpha\theta+1\right)}, \mu_4^{\prime} = \frac{24\left(\alpha\theta+5\right)}{\theta^4\left(\alpha\theta+1\right)}.$$

The log likelihood function of two-parameter Lindley distribution is:

$$logL(x;\alpha,\theta) = n\log\theta^2 - n\log\left(1+\alpha\theta\right) + \sum_{i=0}^{n}\log\left(\alpha\theta+x_i\right) - n\theta\bar{X}.$$

3.1.1 Estimates from moments

Using the first two moments about origin, we have

$$\frac{\mu_2^{\prime}}{\mu_1^{\prime}} = k = \frac{2\left(\alpha\theta+3\right)\left(\alpha\theta+1\right)}{\left(\alpha\theta+2\right)^2}.$$

Taking $b = \alpha\theta$, we get

$$\frac{\mu_2^{\prime}}{\mu_1^{\prime}} = \frac{2\left(b+3\right)\left(b+1\right)}{\left(b+2\right)^2}.$$

This gives a quadratic equation in b. Replacing the first and the second moments $\mu_1^{\prime}$ and $\mu_2^{\prime}$by the respective sample moments $\bar{X}$ and $m_2\prime$ an estimate of k can be obtained. Substituting this estimate of b in the expression for the mean of the two-parameter LD, an estimate of θ can be obtained as :

$$\hat{\theta} = \left(\frac{b+2}{b+1}\right)\frac{1}{\bar{\bar{X}}}, \qquad \bar{X} > 0 \tag{1.8}$$

$$\hat{\alpha} = \frac{b}{\hat{\theta}} \tag{1.9}$$

Later, we observed if we remplace $\alpha = 2+\theta$, we can obtain $XLindley\ distribution$(see last section)

3.2 Two parameter Sujatha distribution

The two parameter Sujatha distribution ($TPSD$) having parameters α et θ defined by its pdf :

$$f(x;\alpha,\theta)=\frac{\theta^3}{\alpha\theta^2+\theta+2}(\alpha+x+x^2)e^{-\theta x};\; x,\theta>0,\alpha\geq 0 \tag{2}$$

It can be easily verified that (2) reduces to *Zeghdoudi distribution* for $\alpha=0$.

The p.d.f (2) can be shown as a mixture of exponential(θ), *gamma* $(2,\theta)$and *gamma* $(3,\theta)$distributions as follows:

$$f(x;\alpha,\theta)=p_1f_1(x)+p_2f_2(x)+(1-p_1-p_2)f_3(x).$$

Where $p_1=\frac{\alpha\theta^2}{\alpha\theta^2+\theta+2}$, $p_2=\frac{\theta}{\alpha\theta^2+\theta+2}$, $f_1(x)=\theta e^{-\theta x}$, $f_2(x)=\frac{\theta^2}{\Gamma(2)}x^{2-1}\theta e^{-\theta x}$ and $f_3(x)=\frac{\theta^3}{\Gamma(3)}x^2e^{-\theta x}$(for more details [34]).
The cumulative distribution function of the distribution can be obtiend as :

$$F_{TPSD}(x)=1-\left[1+\frac{\theta x(\theta+\theta x+2)}{\alpha\theta^2+\theta+2}\right]e^{-\theta x};x>0,\theta>0,\alpha\geq 0. \tag{2.1}$$

The moment generating function of $TPSD$ (2) is :

$$M\left(t\right)=\frac{\theta^3}{\alpha\theta^2+\theta+2}\int\limits_0^\infty\left(\alpha+x+x^2\right)dx. \tag{2.2}$$

Thus, the rth moment about origin of the two-parameter Sujatha distribution has been obtained as the coefficient of $\frac{t^r}{r!}$ in $M_X\left(t\right)$, and by taking $r=1,2,3$, we obtain :

$$\begin{aligned}\mu_1^{\shortmid} &= \frac{\alpha\theta^2+2\theta+6}{\theta\left(\alpha\theta^2+\theta+2\right)},\quad \mu_2^{\shortmid}=\frac{2\left(\alpha\theta^2+3\theta+12\right)}{\theta^2\left(\alpha\theta^2+\theta+2\right)}\\ \mu_3^{\shortmid} &= \frac{6\left(\alpha\theta^2+4\theta+20\right)}{\theta^3\left(\alpha\theta^2+\theta+2\right)},\mu_4^{\shortmid}=\frac{24\left(\alpha\theta^2+5\theta+30\right)}{\theta^4\left(\alpha\theta^2+\theta+2\right)}\end{aligned}$$

3.2.1 Estimation

Method of moment estimator (MoM)

Given a random sample $X_1, ..., X_n$, the two-parameters sujatha distribution (2), by equating the first two moments about origin with the sample mean and moment respectively, we have:

$$\mu_1^{'} = \bar{X} = \frac{\alpha\theta^2 + 2\theta + 6}{\theta\left(\alpha\theta^2 + \theta + 2\right)};\ \mu_2^{'} = m_2^{'} = \frac{2\left(\alpha\theta^2 + 3\theta + 12\right)}{\theta^2\left(\alpha\theta^2 + \theta + 2\right)} \tag{2.3}$$

$$\alpha\theta^2 + \theta + 2 = \frac{\theta + 4}{\theta\bar{X} - 1} = \frac{4\left(\theta + 5\right)}{\theta^2 m_2^{'} - 2} \tag{2.4}$$

We get the following cubic equation in θ:

$$m_2^{'}\theta^3 + 4\left(m_2^{'} - \bar{X}\right)\theta^2 - 2\left(10\bar{X} - 1\right)\theta + 12 = 0 \tag{2.5}$$

Solving the equation (2.5) using iterative method such as Newton-Raphson method. It can obtained the $\hat{\theta}_{MoM}$ and substituting the value of $\hat{\theta}$ in equation(2.4). The $\hat{\alpha}_{MoM}$ obtained as:

$$\hat{\alpha}_{MoM} = \frac{-\bar{x}\hat{\theta}^2 - 2(\bar{x} - 1)\hat{\theta} + 6}{\hat{\theta}^2(\hat{\theta}\bar{x} - 1)} \tag{2.6}$$

Maximum likelihood estimates (MLE)

Let $X_i \sim TPSD(\theta, \alpha), i = \overline{1, n}$, n random variables.The log-likelihood function is:

$$\log L(x; \alpha, \theta) = n\left[3\log\theta - \log\left(\alpha\theta^2 + \theta + 2\right)\right] + \sum_{i=0}^{n}\log\left(\alpha + x_i + x_i^2\right) - n\theta\bar{X}$$

The MLE's $(\hat{\alpha}, \hat{\theta})$ of (α, θ) are the solutions of the following non-linear equations:

$$\frac{\partial\log l(x_i; \theta, \beta)}{\partial\theta} = \frac{3n}{\theta} - \frac{n(2\alpha\theta + 1)}{\alpha\theta^2 + \theta + 2} - n\bar{X} \tag{2.7}$$

$$\frac{\partial\log l(x_i; \beta, \theta)}{\partial\alpha} = \frac{n\theta^2}{\alpha\theta^2 + \theta + 2} + \sum_{i=1}^{n}\left(\frac{1}{\alpha + x_i + x_i^2}\right) \tag{2.8}$$

Chapter 4

New distributions under censored data

In this chapter, we introduce some news distributions by size biasing known distributions.

The size-biased distributions arise when the observations generated from a random process do not have equal probability of being recorded and are recorded according to some weight function. When the sampling mechanism is such that the sample units are selected with probability proportional to some measure of the unit size, the resulting distribution is called "size-biased distribution". Fisher (1934) first introduced such distributions to model ascertainment bias. Let X be a random variable with probability density function (PDF) $f(x;\theta)$ with unknown parameter, then the corresponding weighted distribution function is given by:

$$f^{\alpha}(x;\theta) = \frac{w(x).f(x;\theta)}{\mathbb{E}[w(x)]} \tag{2.8}$$

where $w(x)$ is a non-negative weight function such that $\mathbb{E}[w(x)]$ exists. A special case of the weighted distributions, size-biased distributions is proposed by Rao (1965) when the weighted function has the form $w(x) = x^{\alpha}$ which is called as size-biased distributions of order α, when $\alpha = 1$ or $\alpha = 2$, which are called length-biased and area-biased, respectively. Therefore, the PDF of the length biased distribution is defined by

$$f_L(x;\theta) = \frac{x^{\alpha}.f_0(x;\theta)}{\mu_{\alpha}^{'}} \qquad -\infty < x < +\infty \tag{2.9}$$

where $\mu_{\alpha}^{'} = \mathbb{E}(X^{\alpha}) = \int_{0}^{\infty} x^{\alpha}.f_0(x;\theta)dx$; for $\alpha = 1$ and $\alpha = 2$ we get the size-biased and area-biased distributions and is applicable for area-biased sampling. In many statistical sampling situations care must be taken so that one does not inadvertently sample from size-biased distribution in place of the one intended.

Patil and Ord (1975) studied the size-biased sampling and the related form-invariant weighted distribution whereas Van Deusen (1986) arrived at size - biased distribution theory independently and applied it to fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling (HPS). Later, Lappi and Bailey (1987) analyzed HPS diameter increment data using size- biased distribution. Patil and Rao (1977, 1978) examined some general models leading to size - biased distributions. The results were applied to the analysis of data relating to human populations and wild life management. Gove (2003) reviewed some of the recent results on size- biased distributions pertaining to parameter estimation in forestry with special emphasis on Weibull distribution. Simoj and Maya (2006) introduced some fundamental relationships between weighted and unique variables in the context of maintainability function and inverted repair rate. Mir and Ahmad (2009), Das and Roy (2011) and Ducey and Gove (2015) have also studied the various aspects of size–biased distributions.

Zeghdoudi and Bouchahed (2018) introduced the Zeghdoudi Generated distribution (for more details see [20]) with probability mass function:

$$f_{GZD}^{*}(x;\theta) = \frac{\sum_{k=0}^{n} a_k x^k \exp(-\theta x)}{\sum_{k=0}^{n} a_k \frac{k!}{\theta^{k+1}}}, \qquad \theta, x > 0 \tag{3}$$

In the next sections we propose new size-biased general distribution which is obtained by compounding the size-biased distribution with general distribution without considering its size-biased form and special case of it.

4.1 Size baised General one-parameter distribution and some properties

The probability density function of Size biased Generalized Zeghdoudi Distribution X is:

$$f_{SbGZD}(x;\theta) = \frac{\sum_{k=0}^{n} a_k x^{k+1} \exp(-\theta x)}{\sum_{k=0}^{n} a_k \frac{(k+1)!}{\theta^{k+2}}}, \qquad \theta, x > 0 \tag{3.1}$$

The first and second derivatives of $f_{SbGD,\theta}(x)$

$$\frac{d}{dx} f_{SbGD}(x;\theta) = \frac{\left[a_0 + (2a_1 - \theta a_0) + ... + ((n+1)a_n - \theta a_{n-1})\, x^n - \theta a_n x^{n+1}\right] \exp(-\theta x)}{\sum_{k=0}^{n} a_k \frac{(k+1)!}{\theta^{k+2}}} = 0$$

gives $x_1, x_{2,...,} x_n$ solutions.

We can find easily the cumulative distribution function (c.d.f) of the size biased general one-parameter distribution:

$$F_{SbGD}(x) = 1 - \frac{\sum_{k=0}^{n} \frac{a_k \Gamma(k+2, x\theta)}{\theta^{k+2}}}{\sum_{k=0}^{n} a_k \frac{(k+1)!}{\theta^{k+2}}}; x, \theta > 0 \tag{3.2}$$

4.1.1 Survival and hazard rate function

Let:

$$S_{SbGD}(x) = 1 - F_{SbGD}(x) = \frac{\sum_{k=0}^{n} \frac{a_k \Gamma(k+2, x\theta)}{\theta^{k+2}}}{\sum_{k=0}^{n} a_k \frac{(k+1)!}{\theta^{k+2}}}; x, \theta > 0 \tag{3.3}$$

and:

$$h_{SbGD}(x) = \frac{f_{SbGD}(x)}{1 - F_{SbGD}(x)} = \frac{\sum_{k=0}^{n} a_k x^{k+1} \exp(-\theta x)}{\sum_{k=0}^{n} \frac{a_k \Gamma(k+2, x\theta)}{\theta^{k+2}}} \tag{3.4}$$

be the survival and hazard rate function, respectively.

Proposition 3 *Let $h_\theta(x)$ be the hazard rate function of X. Then $h_\theta(x)$ is increasing for*

$$\sum_{0}^{m}(k+1)(m-2k)a_{m-k}a_{k+1} \geq 0, \;\; m = 0, ..., 2n-1$$

Proof. According to Glaser (1980) and from the density function (3.1) we have

$$\rho(x) = -\frac{f'_{SbGD}(x;\theta)}{f_{SbGD}(x;\theta)} = -\frac{\sum_{k=0}^{n}(k+1)a_k x^k}{\sum_{k=0}^{n} a_k x^{k+1}} + \theta$$

After simple computations we obtain

$$\rho'(x) = \frac{\sum_{m=0}^{2n}\sum_{k=0}^{m}(k+1)(m-2k)a_{m-k}a_{k+1}x^m}{\left(\sum_{k=0}^{n} a_k x^{k+1}\right)^2}.$$

Which implies that $h_\theta(x)$ is increasing *for* $\sum_{k=0}^{m}(k+1)(m-2k)a_{m-k}a_{k+1} \geq 0, \; m = 0, ..., 2n-1$

■

4.1.2 Moments and related measures

The *kth* moment about the origin of the SbGD is:

the Lindley's distribution with applications

$$\mathbb{E}(X^i) = \frac{\sum_{k=0}^{n} \frac{a_k}{\theta^{k+i+2}}\,(k+i+1)!}{\sum_{k=0}^{n} a_k \dfrac{(k+1)!}{\theta^{k+2}}}, i = 1, 2, ...$$

Remark 4 *The kth moment about the origin of the Lindley distribution is:*

$$\mathbb{E}(X^i) = \frac{i!\,(\theta+i+1)}{\theta^i\,(\theta+1)}$$

Corollary 5 *Let $X \sim SbGD(\theta)$, the mean of X is:*

$$\mathbb{E}(X) = \frac{\sum_{k=0}^{n} \frac{a_k}{\theta^{k+3}}\,(k+2)!}{\sum_{k=0}^{n} a_k \dfrac{(k+1)!}{\theta^{k+2}}}. \tag{3.5}$$

Theorem 6 *Let $X \sim SbGD(\theta)$, me =median(X) and $\mu = E(X)$. Then me $< \mu$*

Proof. According to the increasingness of $F(x)$ for all x and θ,

$$F_{SbGD}(me) = \frac{1}{2}$$

and

$$F_{SbGD}(\mu) = 1 - h(\theta) \sum_{k=0}^{n} \frac{a_k \Gamma\left(k+2, \theta h(\theta) \sum_{k=0}^{n} \frac{a_k}{\theta^{k+3}} (k+2)!\right)}{\theta^{k+2}}$$

Note that $\frac{1}{2} < F(\mu) < 1$. It is easy to check that $F(me) < F(\mu)$. To this end, we have $me < \mu$. ■

The coefficients of variation γ, skewness and kurtosis of the SbGD have been obtained as

$$\begin{aligned} \gamma &= \frac{\sqrt{Var(X)}}{\mathbb{E}(X)} \\ skewness &= \frac{\mathbb{E}(X^3)}{(Var(X))^{\frac{3}{2}}} \\ kurtosis &= \frac{\mathbb{E}(X^4)}{(Var(X))^2} \end{aligned}$$

4.1.3 Estimation of parameter

Let $X_1, ..., X_n$ be a random sample of *SbGD*. The ln-likelihood function, $\ln l(x_i;\theta)$ is given by:

$$\ln l(x_i;\theta) = n \ln h(\theta) + \sum_{i=1}^{n} \ln\left(\sum_{k=0}^{m} a_k x_i^{k+1}\right) - \theta \sum_{i=1}^{n} x_i. \quad (3.6)$$

The derivative of $\ln l(x_i;\theta)$ with respect to θ is:

$$\frac{d \ln l(x_i;\theta)}{d\theta} = \frac{n\dot{h}(\theta)}{h(\theta)} - \sum_{i=1}^{n} x_i$$

From the application of Size biased Zeghdoudi distribution in the next section, the method of moments (MoM) and the maximum likelihood (ML) estimators of the parameter θ are the same and it can be obtained by solving the following non-linear equation

$$\frac{\dot{h}\left(\theta\right)}{h\left(\theta\right)}-\bar{x}=0,\ \text{where}\ \dot{h}\left(\theta\right)=\frac{dh\left(\theta\right)}{d\theta} \tag{3.7}$$

The equation to be solved is:

$$\sum_{k=0}^{m}\frac{a_k(k+1)!}{\theta^k}\left((k+2)-\bar{x}\theta\right)=0 \tag{3.8}$$

Note that we can solve the equation (3.8) exactly for $m \leq 4$ and for $m \geq 5$, this equation it can be solved numerically.

Special cases

For $\mathbf{m=0}$, we have $\hat{\theta}_{MV}=\frac{1}{2\bar{x}}$

For $\mathbf{m=1}$, we have $\hat{\theta}_{MV}=\frac{1}{xa_0}\left(a_0-\bar{x}a_1+\sqrt{\bar{x}^2a_1^2+a_0^2+4\bar{x}a_0a_1}\right)$

For $\mathbf{m=2}$, $\hat{\theta}_{MV}$ is one of the two solutions:

$$\begin{cases} \frac{1}{2\bar{x}}a_1\left(a_1\sqrt{9\bar{x}^2a_1^2a_2^3+3\bar{x}a_1^3a_2^2+27\bar{x}a_1^2a_2^2+9a_1^3a_2+a_0a_2}-3\bar{x}a_2+3a_1a_2\right), \\ -\frac{1}{2\bar{x}}a_1\left(a_1\sqrt{9\bar{x}^2a_1^2a_2^3+3\bar{x}a_1^3a_2^2+27\bar{x}a_1^2a_2^2+9a_1^3a_2+a_0a_2}-3\bar{x}a_2+3a_1a_2\right) \end{cases}$$

For $\mathbf{m=3}$ and $\mathbf{m=4}$, we can solve exactly equation(3.8) using methods such as Cardan and Ferrari method

For $\mathbf{m \geq 5}$, according to Galois theorem, there is no general method to solve exactly equation (3.8)

4.1.4 Stochastic orders

Definition 7 *Consider two random variables X and Y. Then X is said to be smaller than Y in the:*

a) Stochastic order ($X <_s Y$), if $F_X(t) \geq F_Y(t)$, $\forall t$.

b) Convex order ($X \leq_{cx} Y$), if for all convex functions ϕ and provided expectation exist, $E[\phi(X)] \leq E[\phi(Y)]$.

c) Hazard rate order ($X <_{hr} Y$), if $h_X(t) \geq h_Y(t)$, $\forall t$.

d) Likelihood ratio order ($X <_{lr} Y$), if $\frac{f_X(t)}{f_Y(t)}$ is decreasing in t.

Remark 8 *Likelihood ratio order⇒Hazard rate order⇒Stochastic order.*
If $E[X] = E[Y]$, *then Convex order⇔Stochastic order.*

Theorem 9 *Let* $X_i \sim SbGD(\theta_i)$, $i = 1, 2$ *be two random variables. If* $\theta_1 \geq \theta_2$, *then* $X_1 <_{lr} X_2$, $X_1 <_{hr} X_2$, $X_1 <_s X_2$ *and* $X_1 \leq_{cx} X_2$

Proof. We have

$$\frac{f_{X_1}(t)}{f_{X_2}(t)} = \frac{\sum_{k=0}^{n} a_k \dfrac{(k+1)!}{\theta_2^{k+2}}}{\sum_{k=0}^{n} a_k \dfrac{(k+1)!}{\theta_1^{k+2}}} e^{-(\theta_1 - \theta_2)t}.$$

For simplification, we use $\ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right)$. Now, we can find

$$\frac{d}{dt} \ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) = -\left(\theta_1 - \theta_2\right)$$

To this end, if $\theta_1 \geq \theta_2$, we have $\frac{d}{dt} \ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) \leq 0$. This means that $X_1 \prec_{lr} X_2$. Also, according to *remark above* the theorem is proved. ■

4.1.5 Mean Deviations

These are two mean deviation: about the mean and about the median, defined as $MD_1 = \int_0^\infty |x - \mu| f(x)\, dx$ and $MD_2 = \int_0^\infty |x - me| f(x)\, dx$ respectively, where $\mu = E(X)$ and me =Median(X). The measures MD_1 and MD_2 can be computed using the following simplified formulas

$$MD_1 = 2\mu F(\mu) - 2\int_0^\mu x f(x)\, dx$$

$$MD_2 = \mu - 2\int_0^{me} x f(x)\, dx$$

4.1.6 Extreme domain of attraction

As to the extreme value stability, the cdf F_{SbGZD} is in the Gumbel extreme value domain of attraction, that is, there exist two sequences $(a_n)_{n \geq 0}$ and $(b_n)_{n \geq 0}$ of real numbers such that for any $x \in \mathbb{R}$, we have

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to+\infty} F_{SbGZD}\left(a_n x + b_n\right)^n = \exp(-\exp(-x)).$$

This follows from Formula 1.2.4 in theorem 1.2.1 (De Haan and Ferreira (2006)) since we have

$$\begin{aligned}
\lim_{t\to\infty} \frac{1 - F_{SbGD}(t + xf(t))}{1 - F_{SbGD}(t)} &= \lim_{t\to\infty} \frac{f_{SbGZD}(t + xf(t))}{f_{SbGZD}(t)} \\
&= \lim_{t\to\infty} \frac{\sum_{k=0}^{n} a_k \left(xf(t) + t\right)^{k+1} e^{(-\theta(xf(t)+t))}}{\sum_{k=0}^{n} a_k t^{k+1} e^{(-\theta t)}} = \exp(-x),
\end{aligned}$$

(such formula is called Γ-variation). Then, F_{SbGD} lies in the Gumbel extreme domain of attraction. In his case, $f(t) = \frac{1}{\theta}$.

So, for (as in the invoked theorem) $a_n = f(F_{SBGZD}^{-1}(1 - 1/n)) = \frac{1}{\theta}$ and $b_n = F_{SBGZD}^{-1}(1 - 1/n)$, we have

$$\lim_{n\to+\infty} F_{SbGZD}\left(a_n x + b_n\right)^n = \exp(-\exp(-x))$$

4.1.7 Estimation of the Stress-Strength Parameter

The stress-strength parameter (R) plays an important role in the reliability analysis as it measures the system performance. Moreover, R provides the probability of a system failure, the system fails whenever the applied stress is greater than its strength, i.e. $R = P(X > Y)$. Here $X \sim SbGD(\theta_1)$ denotes the strength of a system subject to stress Y, and $Y \sim SbGD(\theta_2)$, X and Y are independent of each other. In our case, the stress-strength parameter R is given by

$$\begin{aligned}
R &= P(X > Y) = \int_0^{\infty} S_X\left(y\right) f_Y\left(y\right) dy \qquad (3.9) \\
&= \frac{\int_0^{\infty} \sum_{k=0}^{n} \frac{a_k \Gamma\left(k+2, y\theta_1\right)}{\theta_1^{k+2}} \sum_{k=0}^{n} a_k y^{k+1} \exp\left(-\theta_2 y\right) dy}{\left(\sum_{k=0}^{n} a_k \frac{(k+1)!}{\theta_1^{k+2}}\right)\left(\sum_{k=0}^{n} a_k \frac{(k+1)!}{\theta_2^{k+2}}\right)}
\end{aligned}$$

4.1.8 Lorenz curve

The Lorenz curve is often used to characterize income and wealth distributions. The Lorenz curve for a positive random variable X is defined as the graph of the ratio

$$L(F(x)) = \frac{E(X|X \leq x)F(x)}{E(X)}$$

against $F(x)$ with the properties $L(p) \leq p$, $L(0) = 0$ and $L(1) = 1$. If X represents annual income, $L(p)$ is the proportion of total income that accrues to individuals having the 100 $p\%$ lowest incomes. If all individuals earn the same income then $L(p) = p$ for all p. The area between the line $L(p) = p$ and the Lorenz curve may be regarded as a measure of inequality of income, or more generally, of the variability of X. For the exponential distribution, it is well known that the Lorenz curve is given by:

$$L(p) = p\{p + (1-p)log(1-p)\}$$

For the SbGZ distribution in(3.1);

$$E(X|X \leq x)F_{SbGD}(x) = \frac{\sum_{k=0}^{n} \frac{a_k}{\theta^{k+3}} (k+2)!}{\sum_{k=0}^{n} a_k \dfrac{(k+1)!}{\theta^{k+2}}} \left(1 - \frac{\sum_{k=0}^{n} \dfrac{a_k \Gamma(k+2, x\theta)}{\theta^{k+2}}}{\sum_{k=0}^{n} a_k \dfrac{(k+1)!}{\theta^{k+2}}} \right) \tag{3.10}$$

4.2 Size Biased Zeghdoudi distribution (SBZGD) and some propreties :

Recently, see [16] introduced a new distribution, named Zeghdoudi distribution (ZD) based on mixture of gamma$(2, \theta)$ and gamma$(3, \theta)$, where the density function of random variable X is given by :

$$f_{ZGD}(x;\theta) = \begin{cases} \frac{\theta^3}{\theta+2} x(1+x)e^{-\theta x} & x, \theta \succ 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In this section, we give the size biased Zeghdoudi distribution and study its propreties:

Let X be a random variable with PDF and CDF :

$$f_{SBZGD} = \frac{x.f_{ZGD}(x;\theta)}{E_{ZGD}(x)}$$

Where: $\mathbb{E}(x)=\frac{2(\theta+3)}{\theta(\theta+2)}$

We have:

$$f_{SBZGD}(x;\theta) = \begin{cases} \frac{\theta^4 x^2(1+x)e^{-\theta x}}{2(\theta+3)} & x,\theta > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

And the cumulative distribution function of SBZD is:

$$F_{SBZGD}(x) = \int_0^x f_{SBZD}(x;\theta)dx = \int_0^x \frac{\theta^4 x^2(1+x)e^{-\theta x}}{2(\theta+3)}\, dx$$

$$F_{SBZGD}(x) = \frac{\theta^4}{2(\theta+3)} \int_0^x x^2(1+x)e^{-\theta x}\, dx$$

This gives:

$$F_{SBZGD}(x) = 1 - \frac{x^3\theta^3 + x^2(\theta^3+3\theta^2) + 2x(\theta^2+3\theta) + 2\theta + 6}{2(\theta+3)} e^{-\theta x} \quad ; \quad x > 0 \quad , \quad \theta > 0 \tag{4.2}$$

The first derivative of $f_{SBZGD}(x)$ is:

$$\frac{df_{SBZGD}(x)}{dx} = -x\theta^4 \frac{e^{-x\theta}}{2\theta+6}\left(x^2\theta + x(\theta-3) - 2\right) = 0$$

gives:

$$x = -\frac{1}{2\theta}\left(\theta + \sqrt{2\theta+\theta^2+9} - 3\right) < 0, \frac{1}{2\theta}\left(-\theta + \sqrt{2\theta+\theta^2+9} + 3\right)$$

and the second derivative is:

$$\frac{d^2 f_{SBZGD}(x)}{dx^2} = \theta^4 \frac{e^{-x\theta}}{2\theta+6}\left(x^3\theta^2 + x^2\theta^2 - 6x^2\theta - 4x\theta + 6x + 2\right)$$

$$\text{and } \frac{d^2 f_{SBZGD}(x)}{dx^2} < 0$$

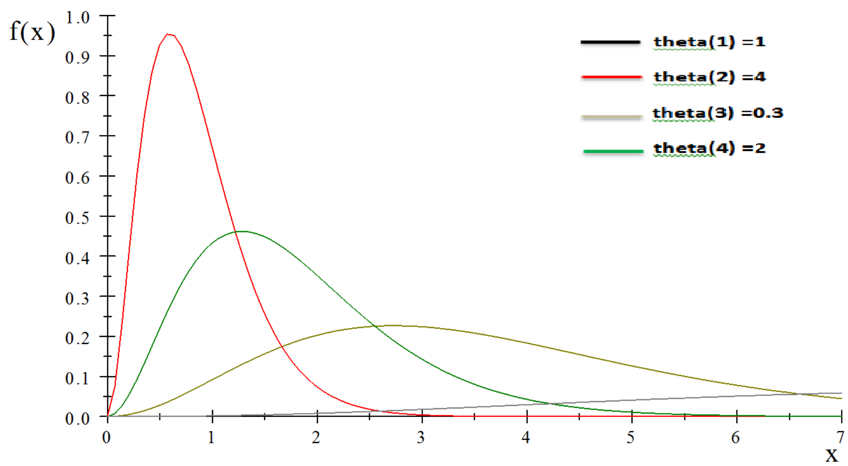


Fig.3: Plots of the density function for some parameter values of θ

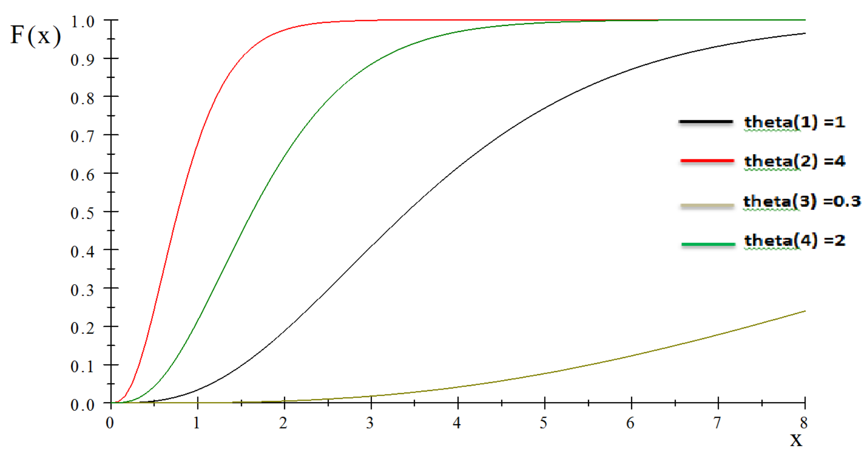


Fig.4: Plots of the cumulative function for some parameter values of θ

Therefore, the mode of SBZGD is given by:

$$Mode(x) = \begin{cases} \frac{1}{2\theta}\left(-\theta + \sqrt{2\theta + \theta^2 + 9} + 3\right) & \text{for } \theta \succ 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

4.2.1 Survival and hazard rate function

The survival function and failure rate(hazard rate)functions for a continuous distribution are defined as :

Let:

$$\begin{aligned} S_{SBZGD}(x) &= 1 - F_{SBZGD}(x) \\ S_{SBZGD}(x) &= \frac{x^3\theta^3 + x^2(\theta^3 + 3\theta^2) + 2x(\theta^2 + 3\theta) + 2\theta + 6}{2(\theta + 3)} e^{-\theta x} \quad ; \quad x, \theta > 0 \end{aligned} \tag{4.4}$$

and:

$$\begin{aligned} H_{SBZGD}(x) &= \frac{f_{SBZGD}(x)}{1 - F_{SBZGD}(x)} \\ H_{SBZGD}(x) &= \frac{\theta^4 x^2(1 + x)}{x^3\theta^3 + x^2(\theta^3 + 3\theta^2) + 2x(\theta^2 + 3\theta) + 2\theta + 6} \end{aligned} \tag{4.5}$$

be the survival and hazard rate function, respectively.

Proposition 10 *Let $H_{SBZGD}(x)$ be the hazard rate function of X. Then $H_{SBZGD}(x)$ is increasing.*

Proof. According to Glaser(1980) and from the density function (4.1):

$$\rho(x) = -\frac{f'_{SBZGD}(x)}{f_{SBZGD}(x)} = \frac{(x\theta - 3x + x^2\theta - 2)}{(1 + x)}$$

$$\rho(x) = \frac{1}{1 + x}(x\theta - 3x + x^2\theta - 2)$$

It follow that:

$$\rho^{'}(x) = \frac{1}{(x+1)^2}\left(\theta x^2 + 2\theta x + \theta - 1\right) \geq 0, \forall x, \theta$$

Imply that h_{SBZGD}(x) is increasing ■

4.2.2 Moments and related measures

The rth moment about the origin of the Size Biased Zeghdoudi Distribution can be obtained as:

$$\begin{aligned} \mu_r^{'} &= \mathbb{E}(X^r) = \int_0^{\infty} x^{(r)} f_{SBZGD}(x)dx \\ &= \int_0^{\infty} x^{(r)} \frac{\theta^4 x^2(1+x)e^{-\theta x}}{2(\theta+3)} dx \\ &= \frac{\theta^4}{2(\theta+3)} \int_0^{\infty} x^{r+2}(1+x)e^{-\theta x} dx \end{aligned}$$

Using gamma integral and little algebraic simplification, we get finally a general expression for the rth factorial moment of SBZGD as:

$$\mu_r^{'} = \frac{\theta^3}{2\theta^{r+3}(\theta+3)}\left[(r+2)!(\theta+r+3\right]; \quad r = 1, 2, 3.... \tag{4.6}$$

Substituting $r = 1, 2, 3$ and 4 in (4.6), the first four moments can be obtained and then using the relationship between moments about origin and moment about mean, the first four moment about origin of SBZGD were obtained as:

$$\mu_1^{'} = \frac{3(\theta+4)}{\theta(\theta+3)}$$

$$\mu_2^{'} = \frac{12(\theta+5)}{\theta^2(\theta+3)}$$

$$\mu_3^{'} = \frac{60(\theta+6)}{\theta^3(\theta+3)}$$

$$\mu_4^{'} = \frac{360(\theta+7)}{\theta^4(\theta+3)}$$

Proposition 11 *Let $X \sim SBZGD(x)$, the mean, variance, coefficients of variation, skewness and kurtosis for X are :*

$$\mu_1 = \mathbb{E}(X) = \frac{3(\theta+4)}{\theta(\theta+3)} \tag{4.7}$$

$$\mathbb{E}(X^2) = \frac{12(\theta+5)}{\theta^2(\theta+3)}$$

$$\mathbb{E}(X^3) = \frac{60(\theta+6)}{\theta^3(\theta+3)}$$

$$\mathbb{E}(X^4) = \frac{360(\theta+7)}{\theta^4(\theta+3)}$$

$$\mu_2 = Var(X) = \frac{3(\theta^2+8\theta+12)}{\theta^2(\theta+3)^2} \tag{4.8}$$

Skewness, Kurtosis and Coefficient of variation of Size Biased Zeghdoudi distribution:

$$Skewness = \sqrt{\beta_1} = \frac{\mathbb{E}(X^3)}{(Var(X))^{\frac{3}{2}}} = \frac{60(\theta+6)(\theta+3)^2}{\left[3(\theta^2+8\theta+12)\right]^{\frac{3}{2}}}$$

$$Kurtosis = \beta_2 = \frac{\mathbb{E}(X^4)}{(Var(X))^2} = \frac{360(\theta+3)^3(\theta+7)}{(3\theta^2+24\theta+36)^2}$$

$$C.V = \gamma = \frac{\sqrt{Var(X)}}{\mathbb{E}(X)} = \frac{\sqrt{3(\theta^2+8\theta+12)}}{3(\theta+4)}$$

Proposition 12 *Let $X_1, X_{2.......}, X_n$ be independant random variables from $SBZGD(\theta)$ distribution, Then the moment generating function (mgf)is given by :*

$$M_x(e^{tX}) = \left(\frac{\theta}{\theta - t}\right)^4 \left(1 - \frac{2t}{2\theta + 6}\right)$$

and

$$M_s(t) = \left(\frac{\theta}{\theta - t}\right)^{4n} \left(1 - \frac{2t}{2\theta + 6}\right)^n$$

Proof. Let $X_1, X_2, \ldots, X_n$ be n independent random variables, we have:

$$M_s(t) = E(e^{tS}) = M_x(t)^n = \left(\frac{\theta}{\theta - t}\right)^{4n} \left(1 - \frac{2t}{2\theta + 6}\right)^n$$

■

Remark 13 *The moment generating function for X and S exists ($E(e^{tX}) < \infty$) only if $t < \theta$*

Theorem 14 *Let $X \sim SBGZD(\theta)$, $M = \operatorname{mod} e(X)$, $me =$median(X) and $\mu = E(X)$. Then $me < \mu$. Then $M < me < \mu$*

Proof. $F_{SBZGD}(M) = 1 - P * e^{-\theta(\frac{1}{2\theta}\left(-\theta+\sqrt{2\theta+\theta^2+9}+3\right))}$; $\theta > 0$

With: ■

$$P = (\frac{\theta^3((\frac{1}{2\theta}\left(-\theta+\sqrt{2\theta+\theta^2+9}+3\right)^3)+\left(\frac{1}{2\theta}\left(-\theta+\sqrt{2\theta+\theta^2+9}+3\right)\right)^4)+\theta^2\ (3(\frac{1}{2\theta}\left(-\theta+\sqrt{2\theta+\theta^2+9}+3\right)^{2)}}{2(\theta+3)})$$
$$+\ \frac{+2(\frac{1}{2\theta}\left(-\theta+\sqrt{2\theta+\theta^2+9}+3\right)))+2\theta(3(\frac{1}{2\theta}\left(-\theta+\sqrt{2\theta+\theta^2+9}+3\right))+1)+6}{2(\theta+3)}$$

$$F_{SBZGD}(\mu) = 1 - \frac{(\theta^3((\frac{3(\theta+4)}{\theta(\theta+3)})^3+(\frac{3(\theta+4)}{\theta(\theta+3)})^4)+\theta^2(3(\frac{3(\theta+4)}{\theta(\theta+3)})^2+2(\frac{3(\theta+4)}{\theta(\theta+3)}))+2\theta(3(\frac{3(\theta+4)}{\theta(\theta+3)})+1)+6}{2(\theta+3)}e^{-\theta(\frac{3(\theta+4)}{\theta(\theta+3)})}$$

4.2.3 Lorenz curve

The Lorenz curve is often used to characterize income and wealth distributions. The Lorenz curve for a positive random variable X is defined as the graph of the ratio:

$$L(F(x)) = \frac{E(X|X \leq x)F(x)}{E(X)}$$

For the SBZGD distribution in(4.1),

$$E(X|X \leq x)F(x) = 1 - \frac{e^{-\theta x}}{2\theta+6}[x^3\theta^3 + x^2\theta(\theta^2+3) + 2x(\theta^2+3\theta) + 2\theta(\theta+3)]$$

$$E(X|X \leq x)F(x) = 1 - \frac{e^{-\theta x}}{2\theta+6}[(1+x)(x^2\theta^3 + 2\theta^2 + 6\theta) + 3x^2\theta]$$

Thus, we obtain the Lorenz curve for the Size Biased Zeghdoudi distribution as:

$$L(p) = \frac{\theta(\theta+3) - \left[x^3\theta^3 + x^2\theta(\theta^2+3) + 2x(\theta^2+3\theta) + 2\theta(\theta+3)\right] e^{-\theta x}}{3(\theta+4)} \tag{4.9}$$

where $x = F^{-1}(p)$ with $F(.)$ given by (4.2).

4.2.4 Stochastic orders

Definition 15 *Consider two random variables X and Y. Then X is said smaller than Y in the:*

a) Stochastic order ($X <_s Y$), if $F_X(t) \geq F_Y(t)$, $\forall t$.
b) Convex order ($X \leq_{cx} Y$), if for all convex functions ϕ and provided expectation exist, $E[\phi(X)] \leq E[\phi(Y)]$.
c) Hazard rate order ($X <_{hr} Y$), if $h_X(t) \geq h_Y(t)$, $\forall t$.
d) Likelihood ratio order ($X <_{lr} Y$), if $\frac{f_X(t)}{f_Y(t)}$ is decreasing in t.

Remark 16 *Likelihood ratio order⇒Hazard rate order⇒Stochastic order. If $E[X] = E[Y]$, then Convex order⇔Stochastic order.*

Theorem 17 *Let $X_i \sim SBZD(\theta_i)$, $i = 1, 2$ be two random variables. If $\theta_1 \geq \theta_2$, then $X_1 <_{lr} X_2$, $X_1 <_{hr} X_2$, $X_1 <_s X_2$ and $X_1 \leq_{cx} X_2$*

Proof. We have:

$$\frac{f_X(t)}{f_Y(t)} = \frac{\theta_1(\theta_2+3)}{\theta_2(\theta_1+3)} e^{-(\theta_1-\theta_2)t} \tag{5}$$

For simplification, we use $ln\frac{f_X(t)}{f_Y(t)}$ Now, we can find:

$$\frac{d}{dt}\ln\left(\frac{f_X(t)}{f_Y(t)}\right) = -(\theta_1 - \theta_2) \tag{5.1}$$

To this end, if $\theta_1 \geq \theta_2$, we have $\frac{d}{dt}\ln\left(\frac{f_X(t)}{f_Y(t)}\right) \leq 0$. This means that $X_1 <_{lr} X_2$. Also, according to Remark 22 the theorem is proved. ∎

4.2.5 Estimation of parameter

Method of Moments Estimation(MME)

Let $\overline{X}$ be the sample mean, equating sample mean and population mean $\mathbb{E}(x)$:

$$\sum_{i=1}^{n}\frac{x_i}{n} = \mathbb{E}(x)$$

Putting the expression of $\mathbb{E}(x)$ from equation (4.7) in the equation and solving the equation for θ, We get :

$$\overline{X} = \frac{3(\theta+4)}{\theta(\theta+3)}$$

$$\hat{\theta}_{MoM} = \frac{3(1-\overline{X}) + \sqrt{3(3\overline{X}^2 + 10\overline{X} + 3)}}{2\overline{X}} \tag{5.2}$$

Maximum Likelihood Estimation (MLE)

Let $X_i\tilde{\ }SBZD(\theta)$, $i = 1 \ldots n$ be n random variables, The ln-likelihood function, $ln\ l(x_i;\theta)$ is:

$$L(\theta) = \left\{\frac{\theta^4}{2(\theta+3)}\right\}^n \prod_{i=1}^{n}(x_i^2 + x_i^3)e^{-\theta\sum_{i=1}^{n} x_i}$$

Logarithm of likelihood function is:

$$\begin{aligned} ln\ l\ (x_i;\theta) &= n\ln(\theta^4) - n\ln(2\theta+6) + \sum_{i=1}^{n}\ln(x_i^2+x_i^3) - \theta\sum_{i=1}^{n}x_i \\ \ln\ l\ (x_i;\theta) &= 4n\ln(\theta) - n\ln(2\theta+6) + \sum_{i=1}^{n}\ln(x_i^2+x_i^3) - \theta\sum_{i=1}^{n}x_i \end{aligned}$$

The derivatives of $ln\ l\ (x_i;\theta)$ with respect to θ is:

$$\begin{aligned} \frac{\partial\ln l(x_i;\theta)}{\partial\theta} &= 0 \\ \frac{\partial\ln l(x_i;\theta)}{\partial\theta} &= \frac{4n}{\theta} - \frac{2n}{2\theta+6} - \sum_{i=1}^{n}x_i \\ \frac{\partial\ln l(x_i;\theta)}{\partial\theta} &= \frac{4}{\theta} - \frac{2}{2\theta+6} - \overline{X} = 0 \end{aligned}$$

From the SBZD (4.1), the method of moments (MoM) and the ML estimators of the parameter θ are the same and it can be obtained by solving the following non-linear equation:

$$\frac{4}{\theta} - \frac{2}{2\theta+6} - \overline{x} = 0$$

$$\hat{\theta}_{MoM} = \hat{\theta}_{ML} = \frac{3(1-\overline{X}) + \sqrt{3(3\overline{X}^2 + 10\overline{X} + 3)}}{2\overline{X}} \tag{5.3}$$

The following theorem shows that the estimator of θ is positively biased.

Theorem 18 *the estimator $\hat{\theta}$ of θ is positively biased, i.e:* $\mathbb{E}(\hat{\theta}) - \theta > 0$.

Proof. Let $g(\overline{x}) = \hat{\theta}$, $g(t) = \frac{1}{2t}[3(1-t) + \sqrt{3(3t^2+10t+3)}]$

$$\frac{d^2g(t)}{dt^2} = \frac{3}{t^3\left(3t^2+10t+3\right)^{\frac{3}{2}}}\left[15\sqrt{3}t + \left(3t^2+10t+3\right)^{\frac{3}{2}} + 17\sqrt{3}t^2 + 5\sqrt{3}t^3 + 3\sqrt{3}\right] \succ 0$$

$g(t)$ is strictly convex. Thus, by Jensen's inequality, we have $\mathbb{E}(g(\overline{x})) > g(\mathbb{E}(\overline{x}))$.

Finally, since $\mathbb{E}(g(\overline{x})) = g(\mu) = g\left(\frac{3(\theta+4)}{\theta(\theta+3)}\right) = \theta$, we obtain $\mathbb{E}\left(\hat{\theta}_{MoM}\right) > \theta$. ■

Theorem 19 *The estimator $\hat{\theta}$ of θ is consistent and asymptotically normal:*

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{P} N\left(0, \frac{1}{\sigma^2}\right)$$

The large-sample $100(1-\alpha)\%$confidence interval for θ is given by:

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n\hat{\sigma}^2}}$$

The proof is omitted because it is very similar to the proof of Theorem 4 (Ghitany et all,2008b)

4.2.6 Simulation I :

We can see that the equation $F(x) = u$, where u is an observation from the uniform distribution on $(0;1)$, cannot be solved explicitly in x (cannot use **lambert W** function in the case $k \geq 2$), the inversion method for generating random data from the SBZD distribution fails. However, we can use the fact that the SBZD distribution is a mixture of gamma $(3;\theta)$ and gamma $(4;\theta)$ distributions:

$$f_{SBZD}(x;\theta) = p(\theta)gamma(3;\theta) + (1-p(\theta))gamma(4;\theta) \qquad 0 < p(\theta) < 1$$

$$f_{SBZD}(x;\theta) = \left(\frac{\theta}{\theta+3}\right)\left(\frac{x^2\theta^3e^{-\theta x}}{2}\right) + \left(\frac{3}{\theta+3}\right)\left(\frac{x^3\theta^4e^{-\theta x}}{6}\right)$$

In this subsection, we investigate the behavior of the ML estimators for a finite sample size (n). A simulation study consisting of the following steps is being carried out $N = 10000$ times for selected values of $(\theta;n)$, where $\theta = 0.1;\ 0.5; 1;\ 3;\ 6$ and $n = 20; 40; 100$

- Generate U_i Uniform$(0;1)$, $i = 1...n$.

- Generate Yi Gamma$(3;\theta)$, $i = 1...n$

- Generate Z_i Gamma$(4;\theta)$, $i = 1...n$

- If $U_i \leq p(\theta)$, then set: $X_i = Y_i$, otherwise, set $X_i = Z_i$, $i = 1...n$

$$average\ bais = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)$$

and the average square error:

$$MSE(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2$$

Table1.5. Average bias of the simulated estimates

	$\theta = 0.1$	$\theta = 0.5$	$\theta = 1$	$\theta = 3$	$\theta = 6$
	$biais(\theta)$	$biais(\theta)$	$biais(\theta)$	$biais(\theta)$	$biais(\theta)$
$n = 20$	0.0024030	0.030394	0.084275	0.2664075	0.4028392
$n = 40$	0.0020183	0.026631	0.076684	0.2509952	0.3547815
$n = 100$	0.001996	0.026566	0.07183	0.2322845	0.3631023

Table1.6. Average MSE of the simulated estimates

	$\theta = 0.1$	$\theta = 0.5$	$\theta = 1$	$\theta = 3$	$\theta = 6$
	$MSE(\theta)$	$MSE(\theta)$	$MSE(\theta)$	$MSE(\theta)$	$MSE(\theta)$
$n = 20$	0.00014896	0.004964	0.02364	0.228012	0.73059
$n = 40$	7.2187×10^{-5}	0.002560	0.0148009	0.139458	0.43883
$n = 100$	3.2452×10^{-5}	0.0015007	0.0085434	0.0838815	0.25438

Remark 20 *Table1.5 shows that the bias is positive(as shown in the Theorem 18), it also shows that bias and MSE (in table1.6)decreases as* n *increases and increases when* θ *increases.*

4.2.7 Application and goodness of fit :

Example1:

Table (2.1) and (2.2) represents the data of survival times (in months) of (94, 91)sierra leone and Liberia individus infected with Ebola virus in 2016, which we compare Lindley distribution(LD), Zeghdoudi distribution(ZD) and SBZD

Table2.1: Comparison between LD, ZD and Size biased zeghdoudi distribution

Survival time $m = 3.17 \quad s = 2.095$	Obs freq	$\text{LD}_{\theta=0.522}$	$\text{ZD}_{\theta=0.852}$	$\text{SBZD}_{\theta=1.173}$
$[0, 2]$	25	38.262	30.339	25.191
$[2, 4]$	43	28.164	37.27	43.75
$[4, 6]$	18	15.075	17.743	18.904
$[6, 8]$	6	7.1187	6.1658	4.9417
$[8, 10]$	2	3.1423	1.828	1.0043
$Total$	94	94	94	94
χ^2	–	13.571	1.8449	**1.2714**

Note: 94 sierra leone individus infected with Ebola virus in 2016.

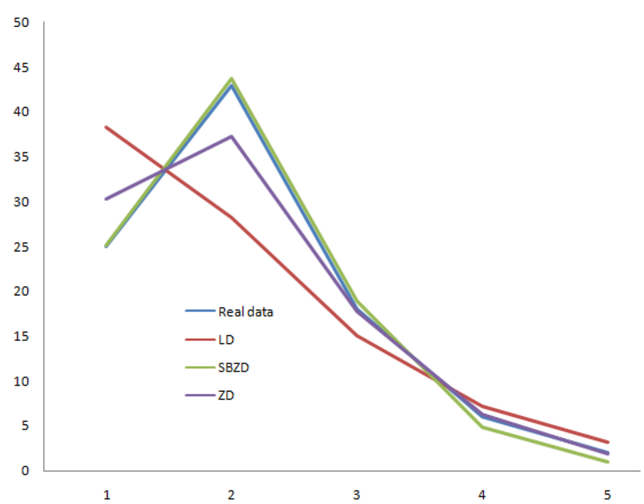


Figure.6. Comparison of SBZD and other models with real data

Table2.2: Comparison between LD, ZD and Size biased zeghdoudi distribution

Survival time $m = 3$ $s = 2.095$	Obs freq	$\text{LD}_{\theta=0.548}$	$\text{ZD}_{\theta=0.896}$	$\text{SBZD}_{\theta=1.2361}$
$[0, 2]$	27	39.100	31.851	27.161
$[2, 4]$	43	27.390	36.361	42.810
$[4, 6]$	16	13.92	15.902	16.428
$[6, 8]$	4	6.2475	5.0682	3.7978
$[8, 10]$	1	2.6189	1.3771	0.68159
$Total$	91	91	91	91
χ^2	–	14.761	2.28	**0.17246**

Note: 91 Liberia individus infected with Ebola virus

Example2:

In this section, a real data represents the lifetime data relating to relief times (in minutes) of 20 patients receiving an analgesic and reported by Gross and Clark $(1975, P.105)$. The data are as follows: 1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3.0, 1.7, 2.3, 1.6, 2.0

According to Table 2.3, we can observe that size-biased Zeghdoudi distribution provide smallest AIC, AICc and BIC values as compared to Exponeniel, Zeghdoudi, Shanker, Xgamma distributions, and hence best fits the data among all the models considered.

Table 2.3. The log-likelihood, AIC, AICc, BIC for 20 patients

Distribution	θ	log-likelihood	AIC	AICc	BIC
Exponential	0.5263	-32.83708	66.72676	67.89638	68.66989
Zeghdoudi	1.365411	-24.85939	52.4496	51.941	52.71451
SBZD	**1.091109**	$\mathbf{-22.10433}$	**46.39088**	**46.43088**	**47.20439**
Shanker	0.838668	-29.89166	61.46066	62.00554	62.77905
Xgamma	1.107468	-31.50824	65.23142	65.2387	66.01221

4.3 Maximun likelihood estimation based on censored data type II :

Consider a n-sample $(X_1, X_2..........X_n)$Generated from (4.1)

We have $(X_{(1)}, X_{(2)}......X_{(m)})$, the first m order statistics. The maximum likelihood estimation for type II censored data is:

$$L(\theta; x) = \frac{n!}{m!(n-m)!} \prod_{i=1}^{m} f(x_i) \left[1 - F(x_m)\right]^{n-m} \tag{5.4}$$

Where: $x_1 \leq x_2...... \leq x_m$

The MLE function can be written:

$$L(\theta; x) \propto \prod_{i=1}^{m} \left[\frac{\theta^4 x_i^2 (1+x_i) e^{-\theta x_i}}{2(\theta+3)}\right] \left[\frac{x_m^3\theta^3 + x_m^2(\theta^3+3\theta^2) + 2x_m(\theta^2+3\theta) + 2\theta+6}{2(\theta+3)} e^{-\theta x_m}\right]^{n-m}$$

$$L(\theta; x) \propto \frac{\theta^{4m}}{(2\theta+6)^m} e^{-\theta \sum_{i=1}^{m} x_i} \left[\prod_{i=1}^{m} (x_i^2 + x_i^3)\right] \left[\frac{x_m^3\theta^3 + x_m^2(\theta^3+3\theta^2) + 2x_m(\theta^2+3\theta) + 2\theta+6}{2\theta+6} e^{-\theta x_m}\right]^{n-m}$$

$$L(\theta; x) \propto \frac{\theta^{4m}}{(2\theta+6)^m} e^{-\theta T_m} \left[\prod_{i=1}^{m} (x_i^2 + x_i^3)\right] \left[\frac{x_m^3\theta^3 + x_m^2(\theta^3+3\theta^2) + 2x_m(\theta^2+3\theta) + 2\theta+6}{2\theta+6}\right]^{n-m}$$

We pose :

$$T_m = \sum_{i=1}^{m} x_i + (n-m)x_m;\ LnL(\theta; x) = l(\theta \backslash \underline{x})$$

And the logarithm of M.likelihood is:

$$l(\theta \backslash \underline{x}) = 4m \ln \theta - m \ln(2\theta+6) + \sum_{i=1}^{m} \ln\left[x_i^2 + x_i^3\right] - \theta T_m + (n-m) \ln\left[\frac{x_m^3\theta^3 + x_m^2(\theta^3+3\theta^2) + 2x_m(\theta^2+3\theta) + 2\theta+6}{2\theta+6}\right]$$

$$l(\theta \backslash \underline{x}) = 4m \ln \theta - n \ln(2\theta+6) + \sum_{i=1}^{m} \ln\left[x_i^2(1+x_i)\right] - \theta T_m + (n-m) \ln\left[x_m^3\theta^3 + x_m^2(\theta^3+3\theta^2) + 2x_m(\theta^2+3\theta)\right.$$

ML estimator of parameter θ is $\hat{\theta}_{ML}$, and when equality to zero, we obtain :

$$\frac{dl(\theta \backslash \underline{x})}{d\theta} = \frac{4m}{\theta} - T_m - \frac{2n}{2\theta+6} + (n-m) \frac{3x_m^3\theta^2 + 3x_m^2(\theta^2+2\theta) + 2x_m(2\theta+3) + 2}{x_m^3\theta^3 + x_m^2(\theta^3+3\theta^2) + 2x_m(\theta^2+3\theta) + 2\theta+6} = 0$$

It is clear that the normal equations do not have explicit solutions. We need some numerical techniques to solve the equations. As part of this work, we will use the R software; who has high abilities to solve a system of nonlinear equations with his **BB solve**(Varadhan and Gilbert, 2009)

4.3.1 Simulation II :

In this section, a simulation is performed to find the parameter estimator of SBZD (θ). The distribution will be generated for $\theta = 1$ and $N = 1000$

Table2.4:Average estimates of parameter and MSE for $\theta = 1$ for varying m

n	m	$\hat{\theta}_{MLE}$	$MSE(\theta)$
20	20	1.157917	0.04738771
	15	1.040943	0.03955424
	10	0.9337355	0.07708183
40	20	0.9275997	0.06249707
	15	0.8753844	0.09010665
	10	0.8313163	0.1609532
100	20	0.7925651	0.1635382
	15	0.7458257	0.2374716
	10	0.7349401	0.343093

Remark 21 *Table 2.4 shows that* $\hat{\theta}_{MLE}$ *decreases as* n *increases and* m *decreases, also* $MSE(\theta)$ *increases as* n *increases and* m *decreases.*

4.4 The XLindley Distribution (XLD): Properties and Application

In this section, we introduce the XLidley distribution and study it's properties. A mixture of two known distributions used to give this new distribution that called XLindley distribution (XLD).

Let X be a random variable following mixture distribution, it's density function (pdf) $f(x)$ given in this form:

$$f(x) = \sum_{i=1}^{k} p_i.f_i(x)$$

With:

$f_i(x)$ probability density function for each i.

p_i, $i = 1....k$ denote mixing proportions that are no-negative, and $\sum_{i=1}^{k} p_i = 1$

We consider: $f_1(x) \sim Exp(\theta)$ and $f_2(x) \sim Lindley(\theta)$ two independent random variables with $p_1 = \frac{\theta}{\theta+1}$ and $p_2 = 1 - \frac{\theta}{\theta+1}$ respectively. Now the density function of XL is given by:

$$f_{XL}(x;\theta) = \begin{cases} \frac{\theta^2(2+\theta+x)}{(1+\theta)^2} e^{-\theta x} & x, \theta \succ 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

The first derivative of $f_{XL}(x)$ is:

$$\frac{df_{XL}(x)}{dx} = \frac{-\theta^2 \left[\theta^2 + \theta(2+x) - 1\right]}{(1+\theta)^2} e^{-x\theta} = 0$$

gives:

$$x = -\frac{\left(\theta^2 + 2\theta - 1\right)}{\theta}$$

For:

1) $0 \prec \theta \prec \sqrt{2} - 1$: $\hat{x} = -\frac{\left(\theta^2+2\theta-1\right)}{\theta}$ is critical point which $f_{XL}(\hat{x};\theta)$ is maximun.

2) $\theta \geq \sqrt{2} - 1$: $\frac{d}{dx} f_{XL}(x;\theta) \leq 0$, then the density function $f_{XL}(x;\theta)$ is decreasing in x.

And second derivative is:

$$\frac{d^2 f(x)}{dx^2} = \frac{\theta^3 \left[\theta^2 + \theta(2+x) - 2\right]}{(1+\theta)^2} e^{-x\theta}$$

$$\text{and } \frac{d^2 f_{SBZGD}(x)}{dx^2} < 0$$

Therefore, the mode of XL is given by:

$$mode(X) = \begin{cases} -\frac{\left(\theta^2+2\theta-1\right)}{\theta} & \text{for } 0 < \theta < \sqrt{2}-1 \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

We can find easily the cumulative distribution function(cdf) of the XL distribution :

$$F_{XL}(x;\theta) = 1 - \left(1 + \frac{\theta x}{(1+\theta)^2}\right) e^{-\theta x} \qquad x > 0,\ \theta > 0 \tag{5.7}$$

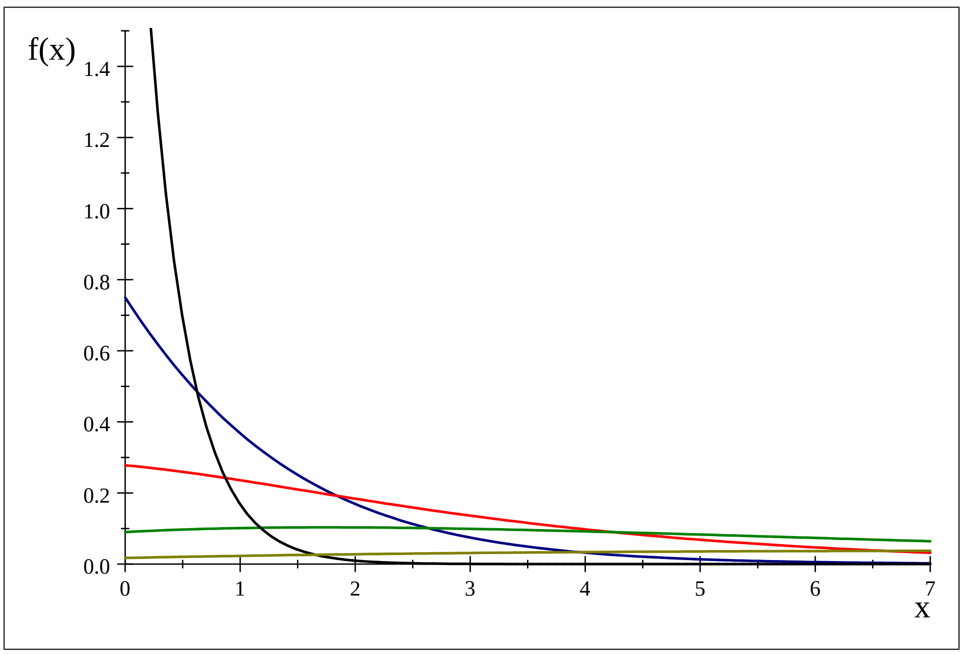


Fig7. Plots of f(x) for some parameter values:bleu(1)red(0.5)green(0.25)black(3)

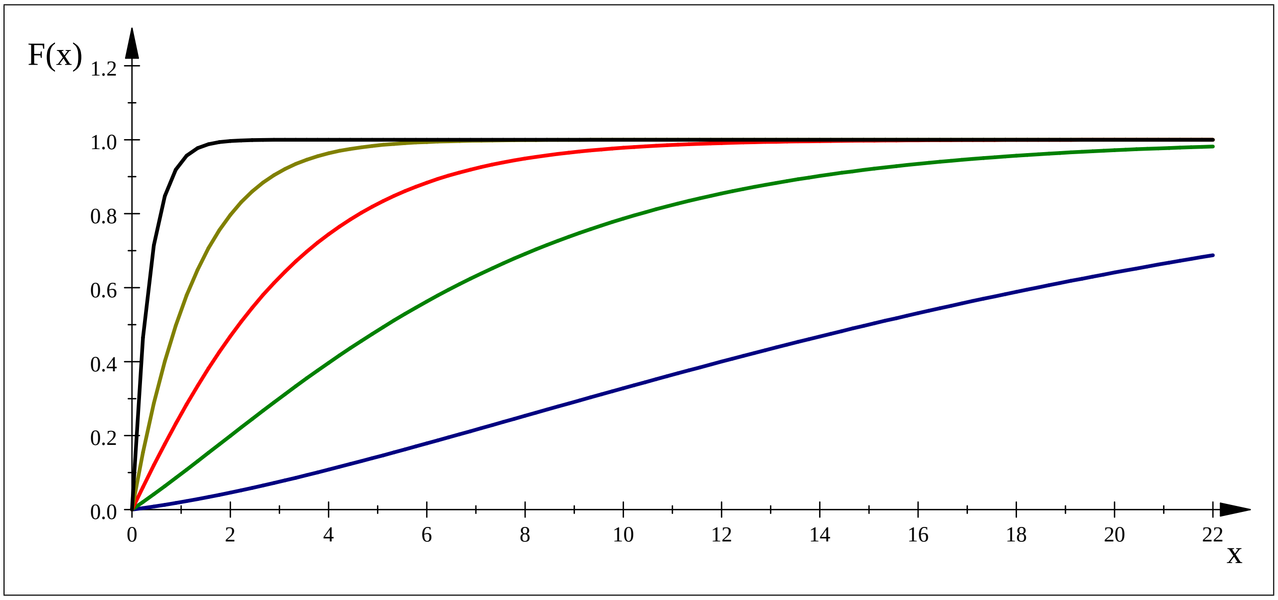


Fig8. Plots of CDF for some parameter values:brown(0.1); green(0.25);red(0.5); blue(1);black(3)

4.4.1 Survival and hazard rate function

The survival function and failure rate(hazard rate) function for a continuous distribution are defined as :

Let:

$$\begin{aligned}
S_{XL}(x) &= 1 - F_{XL}(x) \\
S_{XL}(x) &= 1 - \left[1 - \left(1 + \frac{\theta x}{(1+\theta)^2}\right) e^{-\theta x}\right] \\
S_{XL}(x) &= \left(1 + \frac{\theta x}{(1+\theta)^2}\right) e^{-\theta x} \qquad x \succ 0 \ ; \ \theta \succ 0 \qquad (5.8)
\end{aligned}$$

and:

$$
\begin{aligned}
H_{XL}(x) &= \frac{f_{XL}(x)}{1-F_{XL}(x)} \\
H_{XL}(x) &= \frac{\theta^2\,(x+\theta+2)}{(1+\theta)^2\left(x\frac{\theta}{(1+\theta)^2}+1\right)} \\
H_{XL}(x) &= \frac{\theta^2\,(x+\theta+2)}{(1+\theta)^2+x\theta} \qquad (5.9)
\end{aligned}
$$

be the survival and hazard rate function, respectively.

Proposition 22 *Let $H_{XL}(x)$ be the hazard rate function of X. Then $H_{XL}(x)$ is increasing*

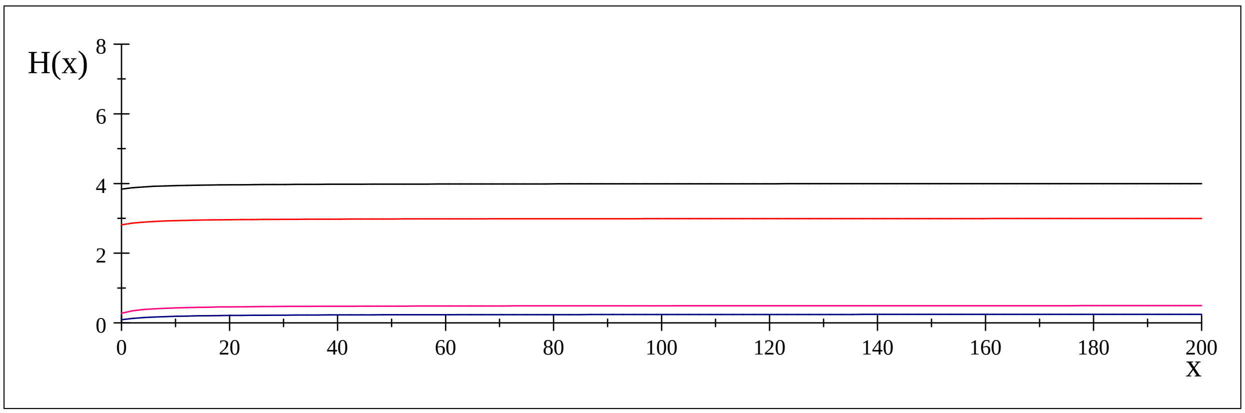


Fig9. Plots of hazard function for some parameter values: blue(0.25);pink(0.5);red(3);black(4)

Proof. According to Glaser(1980) and from the density function(5.5):

$$
\begin{aligned}
\rho(x) &= -\frac{f'_{XL}(x)}{f_{XL}(x)} = \frac{x\theta+\theta^2+2\theta-1}{x+\theta+2} \\
\rho(x) &= \frac{1}{x+\theta+2}\left(x\theta+\theta^2+2\theta-1\right)
\end{aligned}
$$

It follow that:

$$
\rho'(x) = \frac{1}{(x+\theta+2)^2} \qquad (6)
$$

Imply that h_{XL}(x) is increasing

4.4.2 Moments and related measures

The rth moment about the origin of the XLindey distribution can be obtained as:

$$\begin{aligned}
\mu_r^{'} &= \mathbb{E}(X^r) = \int_0^\infty x^{(r)} f_{XL}(x)dx \\
&= \int_0^\infty x^{(r)} \frac{\theta^2(2+\theta+x)}{(1+\theta)^2} e^{-\theta x}dx \\
&= \frac{\theta^2}{(1+\theta)^2} \int_0^\infty x^r(2+\theta+x)e^{-\theta x}dx
\end{aligned}$$

Using gamma integral and little algebraic simplification, we get finally a general expression for the rth factoriel moment of XL distribution as:

$$\mu_r^{'} = \frac{\left(\theta^2+2\theta+r+1\right)r!}{(1+\theta)^2\theta^r} \tag{6.1}$$

Substituting $r = 1, 2, 3$ and 4 in (6.1), the first four moments can be obtained and then using the relationship between moments about origin and moment about mean, the first four moment about origin of XL distrbution were obtained as:

$$\mu_1^{'} = \frac{\theta^2+2\theta+2}{(1+\theta)^2\theta} = \frac{(1+\theta)^2+1}{(1+\theta)^2\theta} = \frac{1}{\theta} + \frac{1}{(1+\theta)^2\theta}$$

$$\mu_2^{'} = \frac{2(\theta^2+2\theta+3)}{(1+\theta)^2\theta^2}$$

$$\mu_3^{'} = \frac{6(\theta^2+2\theta+4)}{(1+\theta)^2\theta^3}$$

$$\mu_4^{'} = \frac{24(\theta^2+2\theta+5)}{(1+\theta)^2\theta^4}$$

Proposition 23 *Let $X \sim XL(x)$, The mean, variance, coefficients of variation, skewness and kurtosis for X are:*

$$\mu_1 = \mathbb{E}(X) = \frac{(1+\theta)^2 + 1}{(1+\theta)^2 \theta} \tag{6.2}$$

$$\mathbb{E}(X^2) = \frac{2(\theta^2 + 2\theta + 3)}{(1+\theta)^2 \theta^2}$$

$$\mu_2 = Var(X) = \frac{\theta^4 + 4\theta^3 + 10\theta^2 + 10\theta + 2}{(1+\theta)^4 \theta^2} = \frac{(1+\theta)^4 + 4\theta^2 + 6\theta + 1}{(1+\theta)^4 \theta^2}$$

Skewness, Kurtosis and Coefficient of variation of XL distribution :

$$Skewness = \sqrt{\beta_1} = \frac{\mathbb{E}(X^3)}{(Var(X))^{\frac{3}{2}}} = \frac{6(\theta^2 + 2\theta + 4)(1+\theta)^4}{\left[(1+\theta)^4 + 4\theta^2 + 6\theta + 1\right]^{\frac{3}{2}}}$$

$$Kurtosis = \beta_2 = \frac{\mathbb{E}(X^4)}{(Var(X))^2} = \frac{24(\theta^2 + 2\theta + 5)(1+\theta)^6}{\left[(1+\theta)^4 + 4\theta^2 + 6\theta + 1\right]^2}$$

$$C.V = \gamma = \frac{\sqrt{Var(X)}}{\mathbb{E}(X)} = \frac{\sqrt{(1+\theta)^4 + 4\theta^2 + 6\theta + 1}}{(1+\theta)^2 + 1}$$

The coefficients are increasing functions in θ (see next figure for the graphe of C.V (γ) and Skewness($\sqrt{\beta_1}$) for varying θ)

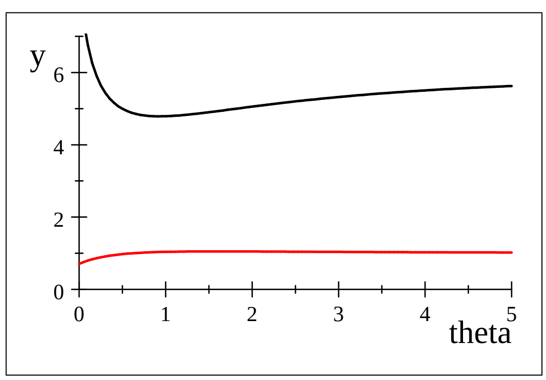


Fig10. Coefficients for variation (red)

and skewness(black)

4.4.3 Stochastics Ordering

Definition 24 *Consider two random variables X and Y. Then X is said smaller than Y in the :*

a) Stochastic order ($X <_s Y$), if $F_X(t) \geq F_Y(t)$, $\forall t$.

b)Convex order ($X \leq_{cx} Y$), if for all convex functions ϕ and provided expectation exist, $E[\phi(X)] \leq E[\phi(Y)]$

c)Hazard rate order ($X <_{hr} Y$), if $h_X(t) \geq h_Y(t)$, $\forall t$

d) Likelihood ratio order ($X <_{lr} Y$), if $\frac{f_X(t)}{f_Y(t)}$ is decreasing in t

Remark 25 *Likelihood ratio order⇒Hazard rate order⇒Stochastic order. If $E[X] = E[Y]$, then Convex order⇔Stochastic order.*

Theorem 26 *Let $X_i \sim XL(\theta_i)$, $i = 1, 2$ be two random variables. If $\theta_1 \geq \theta_2$, then $X_1 <_{lr} X_2$, $X_1 <_{hr} X_2$, $X_1 <_s X_2$ and $X_1 \leq_{cx} X_2$*

Proof. We have

$$\frac{f_X(t)}{f_Y(t)} = \frac{\theta_1^2(2+\theta_1+t)(1+\theta_2)^2}{\theta_2^2(2+\theta_2+t)(1+\theta_1)^2} e^{-(\theta_1-\theta_2)t} \tag{6.3}$$

For simplification, we use $ln\frac{f_X(t)}{f_Y(t)}$ Now, we can find

$$\frac{d}{dt} \ln\left(\frac{f_X(t)}{f_Y(t)}\right) = -\frac{\theta_1-\theta_2}{(t+\theta_1+2)\,(t+\theta_2+2)} - (\theta_1-\theta_2) \tag{6.4}$$

To this end, if $\theta_1 \geq \theta_2$, we have $\frac{d}{dt}\ln\left(\frac{f_X(t)}{f_Y(t)}\right) \leq 0$. This means that $X_1 \prec_{lr} X_2$: Also, according to Remark 25 the theorem is proved.

■

4.4.4 Estimation of parameter

Method of Moments Estimation(MME)

Let $\overline{X}$ be the sample mean, equating sample mean and population mean $\mathbb{E}(x)$:

$$\sum_{i=1}^{n} \frac{x_i}{n} = \mathbb{E}(x)$$

Putting the expression of $\mathbb{E}(x)$ from equation (6.2) in the equation and solving the equation for θ, We get:

$$\overline{X} = \frac{(1+\theta)^2+1}{(1+\theta)^2\,\theta} = \frac{\theta^2+2\theta+2}{\theta^3+2\theta^2+\theta}$$

We obtain equation of 3^{rd} degree: $\overline{X}\theta^3+\theta^2(2\overline{X}-1)+\theta(\overline{X}-2)-2=0$, We take the real part for the solution:

$$\hat{\theta}_{MoM} = -\frac{1}{3\overline{X}}\left(2\overline{X}-1\right) + \frac{\frac{2}{9\overline{X}}+\frac{1}{9\overline{X}^2}+\frac{1}{9}}{\sqrt[3]{\sqrt{\frac{1}{27\overline{X}}+\frac{13}{36\overline{X}^2}+\frac{1}{9\overline{X}^3}+\frac{1}{27\overline{X}^4}}+\frac{11}{18\overline{X}}+\frac{1}{9\overline{X}^2}+\frac{1}{27\overline{X}^3}+\frac{1}{27}}} + \sqrt[3]{\sqrt{\frac{1}{27\overline{X}}+\frac{13}{36\overline{X}^2}+\frac{1}{9\overline{X}^3}+\frac{1}{27\overline{X}^4}}}$$
$$+\sqrt[3]{+\frac{11}{18\overline{X}}+\frac{1}{9\overline{X}^2}+\frac{1}{27\overline{X}^3}+\frac{1}{\overline{X}7}}$$

Maximum Likelihood Estimation (MLE)

Let $X_i\tilde{\ }XLD(\theta)$, $i=1\ldots n$ be n random variables, The ln-likelihood function, $ln\ l(x_i;\theta)$ is:

$$L(\theta) = \left(\frac{\theta^2}{(1+\theta)^2}\right)^n \prod_{i=1}^{n}(2+\theta+x_i)e^{-\theta\sum_{i=1}^{n}x_i} \tag{6.5}$$

Logarithm of likelihood function is:

$$\begin{aligned} ln\ l\ (x_i;\theta) &= 2n\log\theta - 2n\log(1+\theta) + \sum_{i=1}^{n}\log(2+\theta+x_i) - \theta\sum_{i=1}^{n}x_i \\ \ln\ l\ (x_i;\theta) &= 2n\left[\log\theta - \log(1+\theta)\right] + \sum_{i=1}^{n}\log(2+\theta+x_i) - \theta\sum_{i=1}^{n}x_i \end{aligned}$$

The derivatives of $ln\ l\ (x_i;\theta)$ with respect to θ is:

$$\begin{aligned}
\frac{\partial \ln l(x_i;\theta)}{\partial\theta} &= 0 \\
\frac{\partial \ln l(x_i;\theta)}{\partial\theta} &= \frac{2n}{\theta} - \frac{2n}{1+\theta} + \sum_{i=1}^{n}\left(\frac{1}{2+\theta+x_i}\right) - \sum_{i=1}^{n} x_i \\
\frac{\partial \ln l(x_i;\theta)}{\partial\theta} &= \frac{2}{\theta} - \frac{2}{1+\theta} + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2+\theta+x_i}\right) - \bar{X} \\
\frac{\partial \ln l(x_i;\theta)}{\partial\theta} &= \frac{2}{\theta(1+\theta)} + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2+\theta+x_i}\right) - \bar{X} \qquad (6.6)
\end{aligned}$$

To obtain the MLE of θ: $\hat{\theta}_{MLE}$, we can maximize equation(6.6) directly with respect to θ, or we can solve the non-linear equation $\frac{\partial \ln l(x_i;\theta)}{\partial\theta} = 0$. Note that $\hat{\theta}_{MLE}$ cannot solved analytically; so we have to use numerical itération techniques, such as the Newton-Raphson algorithm, are thus adopted to solve the logarithm of likelihood equation for which(6.6) is maximized.

The following theorem shows that the estimator of θ is positively biased.

Theorem 27 *the estimator $\hat{\theta}$ of θ is positively biased, i.e: $\mathbb{E}(\hat{\theta}) - \theta > 0$.*

Proof. Let $g(\overline{x}) = \hat{\theta}$,

$$g(t) = -\frac{1}{3t}(2t-1) + \frac{\frac{2}{9t}+\frac{1}{9t^2}+\frac{1}{9}}{\sqrt[3]{\sqrt{\frac{1}{27t}+\frac{13}{36t^2}+\frac{1}{9t^3}+\frac{1}{27t^4}}+\frac{11}{18t}+\frac{1}{9t^2}+\frac{1}{27t^3}+\frac{1}{27}}} + \sqrt[3]{\sqrt{\frac{1}{27t}+\frac{13}{36t^2}+\frac{1}{9t^3}+\frac{1}{27t^4}} + \frac{11}{18t}+\frac{1}{9t^2}+\frac{1}{27t^3}+\frac{1}{7t}} \quad \blacksquare$$

Proof. $\frac{d^2 g(t)}{dt^2} \geq 0, g(t)$ is strictly convex. Thus, by Jensen's inequality, we have $\mathbb{E}(g(\overline{x})) \succ g(\mathbb{E}(\overline{x}))$.

Finally, since $\mathbb{E}(g(\overline{x})) = g(\mu) = g\left(\frac{(1+\theta)^2+1}{(1+\theta)^2\theta}\right) = \theta$, we obtain $\mathbb{E}\left(\hat{\theta}_{MoM}\right) > \theta$ $\blacksquare$

Theorem 28 *The estimator $\hat{\theta}$ of θ is consistent and asymptotically normal:*

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{P} N\left(0, \frac{1}{\sigma^2}\right)$$

The large-sample $100(1-\alpha)\%$confidence interval for θ is given by:

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n\hat{\sigma}^2}}$$

Proof. The proof is omitted because it is very similar to the proof of Theorem 4 (Ghitany et al,2008b). ■

4.4.5 The quantile function of XL distribution

It may be noted that $F_X(x)$ in equ(5.7) is continuos and strictly increasing, so for the quantile function of X is defined :

$$Q_X(u) = x_u = F_X^{-1}(u) \qquad u\epsilon\,[0.1] \tag{6.7}$$

For $u = F_{XL}(x)$, we give an explicit expression for $Q_X(u)$ in terms of the Lambert W fuction in the following thorem and results.

Theorem 29 *For any $\theta > 0$, the $Q_X(u)$ of the XLindley distribution X is*

$$Q_X(u) = x_u = -\frac{(1+\theta)^2}{\theta} - \frac{1}{\theta}W_{-1}\left[\frac{(1+\theta)^2}{\exp{(1+\theta)^2}}(u-1)\right] \qquad , u\epsilon\,[0.1] \tag{6.8}$$

Where W_{-1} is the negative branch .

Proof. For any $\theta > 0$ let $0 < u < 1$.From equ(6.7) we will solve the equation $u = F_{XL}(x)$ with respect to x, by following the steps bellow:

$$\left(1 + \frac{\theta x}{(1+\theta)^2}\right)e^{-\theta x} = (1-u)$$

$$\left[(1+\theta)^2 + \theta x\right]e^{-\theta x} = (1-u)(1+\theta)^2 \tag{6.9}$$

We multipling the both sides by $\left(-\exp{(-1-\theta)^2}\right)$ of the equ(6.9), we get:

$$-\left((1+\theta)^2 + \theta x\right)e^{-\left[(1+\theta)^2+\theta x\right]} = (1+\theta)^2(u-1)e^{-(1+\theta)^2} \tag{7}$$

By using the definition of Lambert W function ($W(z)\exp(W(z)) = z$) [32], we observe that $-\left[(1+\theta)^2 + \theta x\right]$ is the Lambert W function of the real argument $(1+\theta)^2(u-1)e^{-(1+\theta)^2}$.

So, we have:

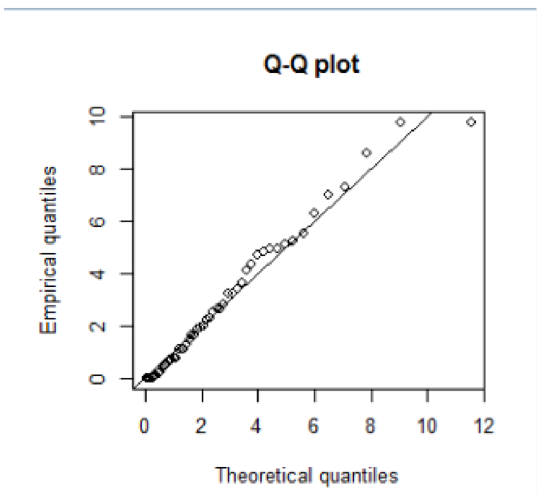
$$W\left((1+\theta)^2(u-1)e^{-(1+\theta)^2}\right) = -\left[(1+\theta)^2+\theta x\right]$$

$$W\left(\frac{(1+\theta)^2}{e^{(1+\theta)^2}}(u-1)\right) = -\left[(1+\theta)^2+\theta x\right] \tag{7.1}$$

In addition to that , for any $\theta > 0$ and $x > 0$ it's obviously that $(1+\theta)^2+\theta x > 0$ and it also checked that $(1+\theta)^2(u-1)e^{-(1+\theta)^2}\epsilon\left(\frac{1}{e},0\right)$ since $0 < u < 1$. Thus, by taking into account the properties of the negative branch W_{-1} of the Lamber W function, so the equation(7.1) become:

$$W_{-1}\left(\frac{(1+\theta)^2}{e^{(1+\theta)^2}}(u-1)\right) = -\left[(1+\theta)^2+\theta x\right] \tag{7.2}$$

Which in turn means the result that given before in Theorem 29 is complete. ■



$Fig11$: The empirical quantiles & theoretical quantiles of XLD

It is clear in $Fig11$ of QQ-Plot that the fitted theoretical quantiles of XL distribution is closer to the empirical quantiles, it's appears as roughly a straight line (although the ends of the Q-Q plot often start to deviate from the straight line).

4.4.6 Simulation

We can see that the equation $F(x) = u$, where u is an observation from the uniform distribution on $(0;1)$, can be solved explicitly in x (we're going to use **lambert W** function, because in this case $k = 1$)

In this subsection, we investigate the behaviour of the ML estimators for a finite sample size (n). A simulation study consisting of the following steps is being carried out $N = 10000$ times for selected values of $(\theta; n)$, where $\theta = 0.1;\ 0.5;\ 1;\ 3;\ 5$ and $n = 20; 40; 100$

- Generate U_i Uniform$(0;1)$, $i = 1...n$.

- Generate Yi exponentiel(θ), $i = 1...n$

- Generate Z_i lindley(θ), $i = 1...n$

- If $U_i \leq p(\theta)$, then set $X_i = Y_i$, otherwise, set $X_i = Z_i$, $i = 1...n$

$$average\ bais = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)$$

and the average square error:

$$MSE(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2$$

Bias	$\theta = 0.1$	$\theta = 0.5$	$\theta = 1.5$	$\theta = 3$	$\theta = 6$
$n = 20$	0.0032	0.01067	0.0485	0.276	0.785
$n = 40$	0.00183	0.0150	0.0135	0.126	0.1770
$n = 100$	0.000321	0.00404	0.0147	0.0452	0.0598

Table2.5. Average bias of the estimator $\hat{\theta}$

MSE	$\theta = 0.1$	$\theta = 0.5$	$\theta = 1$	$\theta = 3$	$\theta = 5$
$n = 20$	$1,0854.10^{-5}$	0.000113	0.00236	0.0765	0.6177
$n = 40$	$3,357.10^{-6}$	0.000225	0.000183	0.01599	0.03135
$n = 100$	$1,0334.10^{-7}$	$1.640.10^{-5}$	0.000217	0.00204	0.00358

Table2.6. The average square error of the estimator $\hat{\theta}$

The result of the simulation are presented in (Tab.2.5) and (Tab.2.6). The following observations are made from the simulation study:

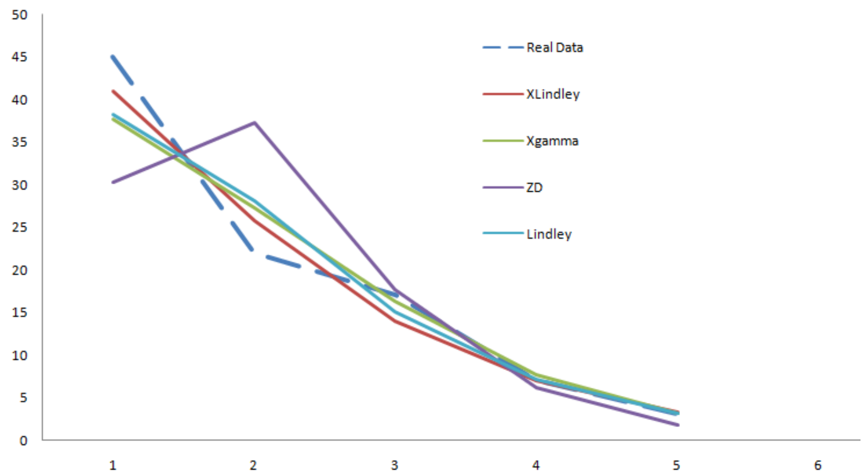
- For some given value of θ, the average of: bias of $\hat{\theta}$ and mean square error of $\hat{\theta}$ are decreases as sample size n increases
- The mean square error(MSE) gets higher and following a similair ways for larger value of θ as we mentioned before.

4.4.7 Application and godness of fit

Now we have used data of survival times (in months) of 94 sierra leone individus infected with Ebola virus showing in table.3, which we compare Lindley distribution(LD), ZD, Exponentiel, Xgamma and XL distributions.

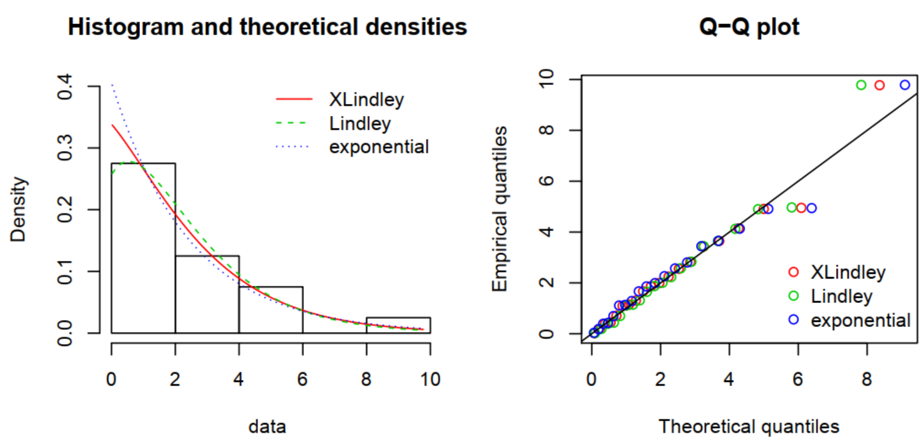
Survival time $m = 3.17$ $s = 2.095$	Obs freq	LD $\hat{\theta}=0.522$	Xgamma $\hat{\theta}=0.689$	ZD $\hat{\theta}=0.852$	Exp $\hat{\theta}=0.315$	XL $\hat{\theta}=0.467$
$[0,2]$	45	38.262	37.652	30.339	43.937	41.028
$[2,4]$	22	28.164	27.197	37.27	23.4	25.855
$[4,6]$	17	15.075	16.342	17.743	12.463	13.984
$[6,8]$	7	7.1187	7.7769	6.1658	6.6375	6.9986
$[8,10]$	3	3.1423	3.2015	1.828	3.5351	3.3409
$Total$	94	94	94	94	94	94
χ^2	–	2.7899	3.2040	14.236	1.8619	1.6446

Table3: Comparison between LD, XG, ZD, Exp and XL distributions



$Fig12$: GoF of Real data with defferent distributions

The $Figure.12$ shows that the XLindley distribution fit the real data better then the other distributions



$Figure13$. The pdf and QQ plots of the models for the data set

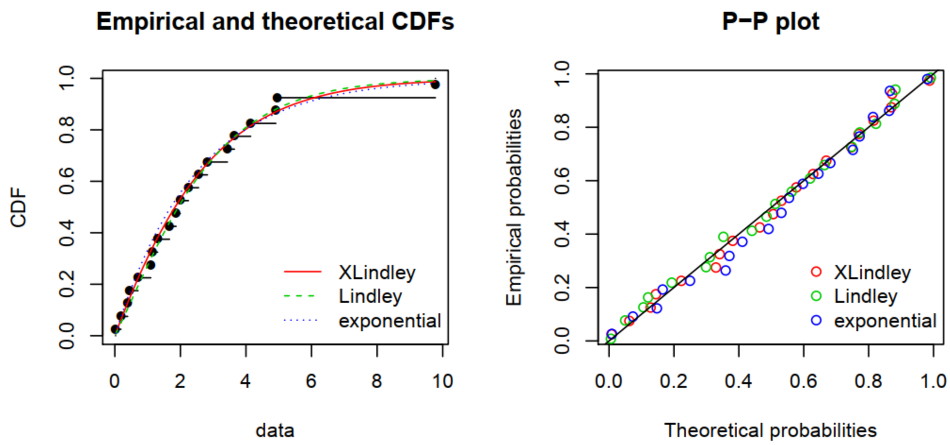


*Figure*14. The CDF and P-P plots of the models for the data set

A density plot compare the fitted densities of the models with the empirical histogram of the data set showing in *figure*.13, it's clear that the fitted probability density and the quantile of XL model are closer to the empirical histogram & quantiles than the Exponential and Lindley. Also in *figure*.14 the fitted cdf for XL model is closer to the empirical cdf of the data set than the other models.

Chapter 5

General conclusion & Perspectives

This study introduces several mathematical properties of two new distributions that are special cases of two parameters distributions: Lorenz curve, moments, quantile and Lambert W functions, methods of point estimate...

Also, the MLE procedure of SBZD was employed to estimate the parameter under two cases: complete & censored data.

In other hand, a simulation studies are carried out to examine the bias and MSE of ML estimators of the unkown parameters of the new models, the efficiency and importance of the SBZ and XL distributions are obtained through a real data sets that showed as the hight flexibility and potentiality for the both N-distributions.

So, according to the pratical part the new distributions are good for modeling in several areas: epidemiology, medical and biochemistry...

Perspectives:

- The Inverse XLindley: Statistical properties with application to COVID-19

- The N-Generalized XLindley families (with coefficient a_k depend on θ)

Bibliography

[1] A. Ahmed, J.A. Reshi and K.A. Mir. Structural properties of size biased Gamma distribution. IOSR Journal of Mathematics (IOSR-JM). 2278-5728. Volume 5, Issue 2 (2013).

[2] H. Akaike. Information theory and extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki.Second International Symposium on Information Theory pp. 267-281 (1973).

[3] A. Arooj. Size Biased Lindley Distribution and Its Properties a Special Case of Weighted Distribution. Applied Mathematics, 8, 808-819 (2017).

[4] A. Beghriche, H. Zeghdoudi. A Size Biased Gamma Lindley Distribution. Thailand Statistician. 17(2): 179-189 (2019).

[5] C. Benatmane, H. Zeghdoudi, R. Shanker and N. Lazri. Composite Rayleigh-Pareto distribution: Application to real fire insurance losses data set. (2020).

[6] S. Chouia, H. Zeghdoudi. The XLindley distribution: Properties and application. Journal of Statistical Theory and Applications, Vol 19(2020)

[7] S. Chouia, H. Zeghdoudi, R.Vinoth and A. Benghriche. A new Size- Biased Distribution with application. Journal of Applied Probability Statistics. Vol 16 issue 1 (2021).

[8] G.M. Cordeiro, M. Castro: A new family of generalized distributions. Journal of Statistical Computation and Simulation. 81(7), 883–898 (2011).

[9] D.R. Cox. A remark on censoring andsurrogate response variable. J. R. Statist.Soc. B45:391–93(1983)

[10] M. Dragulin, V. Preda. A new family of Lindley-type distributions with applications. Conference on Applied and Industrial Mathematics(CAIM), Bacau, 18-21 (2014).

[11] I. Elbatal, A. Gokarna. A new generalization of the exponential Pareto distribution. Journal of Information and Optimization Sciences, 38:5, 675-697 (2017).

[12] L. Fisher, P. Kanarek. Presenting cen-sored survival data when censoring and survival times may not be independent. InRe-liability and Biometry. pp. 303–26. Philadelphia: SIAM(1974).

[13] M.E. Ghitany, F. Alqallaf, D.K. Al-Mutairi and H.A. Husain. A two-parameter weighted Lindley distribution and its applications to survival data. Mathematics and Computers in Simulation 81 1190–1201 (2011).

[14] R. Grine, H. Zeghdoudi. On Poisson Quasi-Lindley Distribution and its Applications. Journal of Modern Applied Statistical Methods, 16(2), 403-417 (2017).

[15] J. Hyde. Testing survival under right censoring and left censoring. Biometrika; 64(2): 225–230 (1977).

[16] P. Jodrá. Computer generation of random variables with Lindley or Poisson Lindley distribution via the Lambert W function. Mathematics and Computers in simulation 81 851-859 (2010).

[17] D. Kalbfleisch, R. L. Prentice. The Statistical Analysis of Failure Time Data. Wiley Sons, New Jersey, 2nd edition(2002).

[18] J.P. Klein, L. Melvin. SURVIVAL ANALYSIS Techniques for Censored and Truncated Data. Second editon. Springer-Verlag (2003).

[19] D.G. Kleinbaum, M. Klein. Survival Analysis: A Self-Learning text, third edition. Statistics for Biology and Health (2012).

[20] H. Messaadia, H. Zeghdoudi. Zeghdoudi Distribution and Its Applications. International Journal of Computing Science and Mathematics. Vol.9 No.1, pp.58 - 65 (2018).

[21] R. Shanker, A. Mishra. A Two parameter lindley distribution. Statistics in Transition-new series. Spring. vol.14, No.1, pp.45-56 (2013).

[22] R. Shanker. A Two parameter Sujatha distribution. Biometrics Biostatistics International Journal. Vol 7, issue3 (2018).

[23] R. Shanker, A. Mishra. A quasi Lindley distribution. African Journal of Mathematics and Computer Science Research, 6(4), 64-71 (2013).

[24] H. Wang, H. Dai. Analysis for Time-to-Event Data under Censoring and runcation. Typeset by SPi Books and Journals (2017).

[25] H. Zeghdoudi, A. Ghouar and M.C. Bouras. New Zero -Truncated Distribution: Properties and Applications. Journal of modern applied statistical methods JMASM.Vol 19 isssue 1 (2020).

[26] H. Zeghdoudi, L. Bouchahed. A new and unified approach in generalizing the Lindley's distribution with applications. Statistics in Transition new series. Vol. 19, No. 1, pp. 61–74 (2018).

[27] H. Zeghdoudi, N. Lazri and D. Yahia. Lindley Pareto Distribution. Statistics in Transition new series. Vol. 19 No. 4, pp. 671–692 (2018).

[28] H. Zeghdoudi, N. Lazri. On Composite Gamma(2,theta)-Pareto Distribution. International Journal of Applied Mathematics and Statistics, Vol 56, Issue N: 5, pp 101 107(2017).

[29] H. Zeghdoudi, N. Lazri. On Lindley-Pareto Distribution: Properties and Application. GSTF Journal of Mathematics, Statistics and Operations Research (JMSOR) Vol.3 No.2(2016).

[30] H. Zeghdoudi, S. Nedjar. Gamma Lindley distribution and its application. Journal of Applied Probability and Statistics Vol. 11, Nř 1, 129-138(2016a)

[31] H. Zeghdoudi, S. Nedjar. On Gamma Lindley distribution: Proprieties and simulation. Journal of Computational and Applied Mathematics 298, pp167-174 (2016c).

[32] H. Zeghdoudi, S. Nedjar. On Poisson Pseudo Lindley Distribution: Properties and Applications. Journal of Probability and Statistical Sciences (JPSS) Vol 15 N 1 (2017).

[33] H. Zeghdoudi, S. Nedjar. On Pseudo Lindley Distribution: Properties and Applications. New Trends in Mathematical Sciences. NTMSCI 5, No. 1, 59-65 (2017).