REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

INFERENCE OF STRESS-STRENGTH RELIABILITY FOR TWO-PARAMETER OF EXPONENTIATED GUMBEL DISTRIBUTION BASED ON LOWER RECORD VALUES

Ehsan FAYYAZISHISHAVAN

DOCTOR OF PHILOSOPHY THESIS
Department of Statistics
Statistics Program

Advisor
Assoc. Prof. Dr. Serpil KILIÇ DEPREN

July, 2021

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

INFERENCE OF STRESS-STRENGTH RELIABILITY FOR TWO-PARAMETER OF EXPONENTIATED GUMBEL DISTRIBUTION BASED ON LOWER RECORD VALUES

A thesis submitted by Ehsan FAYYAZISHISHAVAN in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 12.07.2021 in Department of Statistics, Statistics Program.

Assoc. Prof. Dr. Serpil KILIÇ DEPREN
Yildiz Technical University
Advisor

Approved By the Examining Committee

Assoc. Prof. Dr. Serpil KILIÇ DEPREN, Advisor
Yildiz Technical University ______________________

Prof. Dr. Filiz KARAMAN, Member
Yildiz Technical University ______________________

Assoc. Prof. Dr. Seda BAĞDATLI KALKAN, Member
Istanbul Commerce University ______________________

Assoc. Prof. Dr. Atıf Ahmet EVREN, Member
Yildiz Technical University ______________________

Prof. Dr. Özlem DENİZ BAŞAR, Member
Istanbul Commerce University ______________________

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of Inference of Stress-Strength Reliability for Two-Parameter of Exponentiated Gumbel Distribution Based on Lower Record Values supervised by my supervisor, Assoc. Prof. Dr. Serpil KILIÇ DEPREN. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Ehsan FAYYAZISHISHAVAN

Signature

Dedicated to my wife
and my family

ACKNOWLEDGEMENTS

I would like to thank all my family members and professors who supported me during my Ph.D. journey. Firstly, I would like to thank my advisor, Assoc. Prof. Dr. Serpil KILIÇ DEPREN for her support and patience during my thesis. Special thanks go to Dr. Gholamhossein GHOLAMI for his continuous assists since my Bachelor's, both as a teacher and a good friend. Also, I would like to thank all the examining committee members for their time to read, leave comments, and make guidance to raise the quality of the thesis. Furthermore, I would like to specially thank Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK, for giving me motivation, morality, and great ideas. My special thanks also go to Prof. Dr. Ali Hakan BÜYÜKLÜ, who was the first link between the Yildiz Technical University and me, and for giving me the opportunity to pursue my Ph.D. at the Department of Statistics at Yildiz Technical University.

I want to thank my wife, Sanaz, who has always sacrificed and supported me during my life. Also, I owe my parents Mohammadreza and Kobra, for devoting their life to achieve my goals. I'm indebted to them and appreciate their love and support. I want to thank my elder brother Dr. Ebrahim for always being there, especially when I needed consultation and advice. Also, I would like to give my thanks to my younger brother, Iman, for his help whenever I needed it, especially for English affairs.

Last but not least, I would like to thank all my friends and colleagues who will be happy with my progress, especially Omid Afshani Azar, Arash Mirtorabi, and Hadi Ghani Shayesteh. In addition, I want to thank all the individuals who impacted positively on my academic life directly or indirectly.

Ehsan FAYYAZISHISHAVAN

TABLE OF CONTENTS

LIST OF SYMBOLS

χ^2	Chi-Square Distribution
det	Determinant
exp	Exponential
EG	Exponentiated Gumbel Distribution
F	Fisher Distribution
I	Fisher Information Matrix
$Gamma$	Gamma Distribution
Γ	Gamma Function
G	Gumbel Distribution
E	Mathematical Expectation
ln	Natural Logarithm
N	Normal Distribution
R	Reliability
λ,σ	Scale Parameter
α,β	Shape Parameter
$\mathcal{U}$	Uniform Distribution
$\mathbb{V}$	Variance

LIST OF ABBREVIATIONS

CDF	Cumulative Distribution Function
HRF	Hazard Rate Function
HPD	Highest Posterior Density
i.i.d.	Independent Identically Distributed
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MGF	Moment Generating Function
PDF	Probability Density Function

ABSTRACT

Inference of Stress-Strength Reliability for Two-Parameter of Exponentiated Gumbel Distribution Based on Lower Record Values

Ehsan FAYYAZISHISHAVAN

Department of Statistics
Doctor of Philosophy Thesis

Advisor: Assoc. Prof. Dr. Serpil KILIÇ DEPREN

The two-parameter of exponentiated Gumbel distribution is an important lifetime distribution in survival analysis. This thesis investigates the estimation of the parameters of this distribution by using the lower record values. The maximum likelihood estimator (MLE) procedure of the parameters is considered, and the Fisher information matrix of the unknown parameters is used to construct asymptotic confidence intervals. Bayes estimator of the parameters and the corresponding credible intervals are obtained by using the Gibbs sampling technique. Two real datasets are provided to illustrate the proposed methods.

Keywords: Exponentiated Gumbel Distribution; Maximum likelihood estimator; Bayes estimation; Lower records; Monte Carlo methods

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

ÖZET

Alt Rekor Değerlerine Dayalı İki Parametreli Üstel Gumbel Dağılımı için Stres-Dayanıklılık Güvenilirliğinin Çıkarımı

Ehsan FAYYAZISHISHAVAN

İstatistik Anabilim Dalı
Doktora Tezi

Danışman: Doç. Dr. Serpil KILIÇ DEPREN

İki parametreli üstel Gumbel dağılımı, sağkalım analizinde kullanılan önemli bir yaşam boyu dağılımıdır. Çalışmada daha düşük rekor değerleri kullanılarak bu dağılımın parametrelerinin tahmini incelenmektedir. Parametrelerin En Çok Olabilirlik tahmincisi dikkate alınmış ve bilinmeyen parametrelerin Fisher bilgi matrisi asimptotik güven aralıklarını oluşturmak için kullanılmıştır. Parametrelerin Bayes tahmincisi ve bunlara karşılık gelen güvenilir aralıkları, Gibbs örnekleme tekniği kullanılarak elde edilmiştir. Önerilen yöntemler iki gerçek veri setinde uygulanmıştır.

Anahtar Kelimeler: Üstel Gumbel Dağılımı; En Çok Olabilirlik tahmincisi; Bayes tahmini; Daha düşük rekor değerleri; Monte Carlo yöntemleri

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

1
INTRODUCTION

1.1 Literature Review

1.1.1 The History of Reliability

Reliability may be a common concept that has been recognized for a long time as a positive trait for an individual or an item. It began in 1816 much quicker than individuals thought. The word “reliability” was first coined by a poet named Samuel Taylor Callridge.

Statistically, reliability is the consistency of a set of dimensions or measurement tools which are often used to explain an experiment. Reliability is reversely related to the random errors.

In psychology, reliability is mainly concerned with the consistency of an estimation. A test is set to be reliable when frequent equal results are obtained. For example, in case an experiment is designed to measure a characteristics such as introvertedness, equal results should be obtained every time the experiment is performed concerning a topic.

Therefore, prior to the World War II, reliability was defined as “reliable” or “replicability”. This term was redefined by the U.S. Army in 1940s, which is currently used. The term primary means that a product performs until the expected time. This concept involves additional characteristics, such as the diversity of the products, service-based applications, software packages, or human-related activities. Nowadays, these characteristics have been distributed in all dimensions in a world enriched by the current technologies. Here is the review of the term “reliability” from the past up to present time.

In recent decades, the concept of reliability has been an interest to many researchers. It has been widely developed during the World Wide II since the usage and consequently the occurred errors in military systems have been raised significantly. One of the primary studies on reliability analysis has been done by Abraham Wald, one of the statistical research groups of Columbia University. He discussed the estimation of vulnerability of aircraft used during the war. He analyzed the data gathered from the damages of those aircraft that return from their mission to the shields. His work led to increase the damageable and sensitives body elements of the aircraft and consequently led to raised return rate of those aircraft.

As the time passed and the need for mass production increased to meet the market demands, the production strategy shifted to the use of complex machinery and equipment. With the development of mass production systems, having good and high-quality products and designing high-performance systems were considered, and the reliability became more significant. Today, this theory is used not only for industrial engineering, mechanics, and electricity purposes but also for communications, computer, and information technology engineering dimensions.

Reliability is usually used to express the degree of confidence in the proper functioning of a component or, in general, a set of factors over a specified period of time. All the reliability relationships are based on the probability theory and are calculated under conditions that are somehow vague in relation to the part under study. One can calculate the reliability of a system using the following method. Initially, the system is departed into several components, and then the reliability of components would represent the reliability of the whole system. One of the approaches among the others is that the reliability of each component is calculated based on the available statistical data and by selecting a model for the failure rate, and then its parameters are estimated based on the available data. Numerous random and statistical features of complex systems have been studied to improve their reliability.

Perhaps one of the primary uses of the reliability pertains to the telegraph device. It is a system which is charged by battery and has simple receivers and senders to be connected via cable. In past, the most important failure of this system could be related to the cable disintegration or insufficient voltage so that the electricity lamp, telephone, production and distribution of Ac electricity, and new electronic-based applications were not present in terms of reliability.

In 1916, radio was emerged gradually in society. Cars were used in public in 1920 and represented the mechanical reliability applicability. In 1920s, an improvement was developed in products by Dr. Walter Shewart in laboratory using the statistical quality control. In line i^{th} the reliability of products, a new development was revealed in 20th century. Statistics were taken into consideration as necessary instruments for evaluation along with the development of reliability concepts.

At this point, designers were responsible for the reliability and failure fixing, and there was no economic justification for performing the impeding plans. In 1920s and 1930s, Mr. Taylor worked on solutions through which the products could be stronger and increase the efficiency of production line. He was the first to separate the engineering from management and control.

Charles Lindsberg needed nine cylinders for engine cooling purpose of his ocean plane in 1927 so as to enable 40-hour capacity of continuous work without the need for maintenance. Many individual advances were achieved in 1930s in limited and determined industries. Measuring the process and quality were in their initial stages. In Sweden, Waldy Waibul embarked on testing the materials burnout. During this time, he invented a distribution which is now called the Waibul. In 1930s, Rozen and Ramler found a similar distribution for explaining the coal dust grade.

Until 1940s, there was no reliability and reliability management. The World War II requirements produced many electronic products for the military purposes which included the electronic switches, transportable radios, radar, and electronic detonators. Electronic computers began to work in the early war, but they were not completed by the late war. It was understood in the beginning of war that 50% of the air force electronic equipment in warehouse could not provide the needs of the army and marine forces. More importantly, in this ear most of the works dealing with the reliability should have been done through testing on materials and their burnout. M.A. Miner published a paper in 1945 entitled “the collective failures in burnout” in ASME Journal. Also, Esptein developed paper concerning the statistical view toward the failure problems in 1948 in Applied Physics Journal.

The most important military applications of the reliability were the vacuum tubes, which were used in radar or other electronic products. These systems bore huge expenses and problems using the war. After the war, it was estimated hat half of the ship electronic equipment had failures. The vacuum tube sockets were a natural cause for frequent system problems.

Breaking the system or removal of pipes as well as reinstalling them were the two main solutions for repairing the electronic system.

This process imposed significant expenses on the army in war. They were not able to have their basic equipment without use. In case of lack of modification, the performance and supportive expenses turned out to be greater. In 1948, IEEE established the Association of Reliability with under the supervision of Richard Relman.

Also, in 1948 Z. W. Birunbaum established a laboratory to run statistical research in Washington University, which improved and developed the statistical applications with the cooperation of Naval research Department.

1.1.2 Fundamental Definitions and Concepts

Survival analysis is a branch of statistical science that deals with mortality in biological organizations and breakdown in mechanical systems. This is called validity theory in engineering, sociology, and economics. In other words, survival analysis is a variety of statistical techniques for analyzing random variables that have negative values. One of the most prominent features of survival analysis is the time of failure of a physical, mechanical, or electrical phenomenon, the time of death of a biological unit (human, animal, or a living cell, etc.). In the survival analysis literature, death or accident are considered. For example, in the case of a convicted felon, an incident can be considered as the length of time a person commits a crime again after being released. However, the random variable may not be related to time, for example, the values of the variable are related to the amount paid by an insurance company to the insured in a particular situation.

Data analysis of survival and failure time is considered in many branches of applied statistics. The nature of experiments with this data is often accompanied by the omission of units from the experiment, which is called censorship. The existence of censored data distinguishes the survival analysis from the other statistical analyses. In fact, censorship occurs when the exact time of failure is not known.

1.1.2.1 Survival Functions

Let T be a non-negative and strictly continuous random variable on support of (L, U), which demonstrates the lifetime of a system, unit, component, etc., and follows the PDF of f and the CDF of F. Thus, F demonstrates the probability that the system failure occurs before time t. One of the most important quantities in lifetime analysis, is the survival function (or reliability function).

Survival times are data that measure the time to reach a particular event, such as the loss, death, response, and the onset of a specific disease. Survival times have random changes and, like any other random variable, create a distribution that is usually described by the following three functions:

- Survival function
- Probability density function

- Risk function

The main problem in the survival analysis is estimating such functions in a sample and interpreting them with the pattern of survival in society. The definitions of the survival function and probability density function are as follows.

Definition 1.1. *Survival Function* For a random variable with a CDF of F, survival function is demonstrated by $\bar{F}$ or $S(t)$, and defined by:

$$\bar{F}(t) = S(t) = Pr(T > t) = 1 - F(t), \;\; t > 0 \tag{1.1}$$

Obviously, $\bar{F}$ is a decreasing function, and $\lim_{t\to 0} \bar{F}(t) = 1$, and $\lim_{t\to\infty} \bar{F}(t) = 0$. In fact, survival function represents the probability of failure occurring after the given time.

In clinical studies, if there are no censored data, the survival function is estimated as the portion of patients which survived more than time t:

$$\hat{S}(t) = \text{patients survived more than} t/\text{all patients.} \tag{1.2}$$

In case there are censored data, $\hat{S}(t)$ is not computable, and nonparametric methods should be implemented.

1.1.2.2 Probability Density Function

Similar to every continuous random variable, the survival time has a density function. It is defined as the probability of the fact that a patient is passed away during the time t and $t + \Delta t$:

$$f(t) = \lim_{\Delta t \to 0} (p(t \leq T \leq t + \Delta t)/\Delta t) \tag{1.3}$$

This density function is also called the disability rate. Similar to the survival function in clinical studies, if there is no censored data, the density function $f(t)$ can be estimated using the following formula:

$$f(\hat{t}) = \text{all individuals pass away within this period}/(\text{all individuals})\times(\text{time duration}). \tag{1.4}$$

Based on the definition of survival function and probability density function, we have

$$F(t) = P(T \leq t) = 1 - S(t) \tag{1.5}$$

On the other hand, we have

$$f(t) = F'(t) = \frac{d}{dt}F(t); \tag{1.6}$$

So, the relationship between survival function and probability density function is as below:

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t). \tag{1.7}$$

1.1.2.3 Censoring

A unique feature of survival studies, compared to other studies, is that the information collected is usually censored in these studies. Censorship occurs when we have incomplete information about the survival time of some of the people studied. We know that the primary variable in survival analysis is survival time, which means reaching a specific event. Therefore, to define the time of survival, we need to specify two specific times: one is the time of onset, which can be a time such as the date of birth or the date of diagnosis, and the other is the time of disappearance, which is usually the time of death or a specific condition.

Censorship is divided by two types, namely point censoring and interval censoring, in terms of when censorship occurs and how our information about censored individuals is obtained.

1.1.2.4 Bayesian Inference

Today, Bayesian method is widely used for the statistical inferences in various sciences. The main difference between the classical (frequentist) method and the Bayesian method is that in the classical method, parameter θ is a fixed and unknown quantity in society, which we obtain information about based on a random sample from that society, whereas in Bayesian method θ is not a fixed quantity but a random variable that changes in the domain of its values and its changes are expressed by a prior probability distribution. This distribution, which is based on the experimenter's belief, must be determined before the data can be observed.

In Bayesian inference, the former distribution is combined with sample information to form a new distribution called the posterior distribution. Suppose the random variable Y represents the sample observations that has a density function $f(Y|\theta)$. If we consider the prior density of parameter θ as $\pi(\theta)$, then the posterior density of θ is:

$$\pi(\theta|y) = \pi(\theta)f(y|\theta)f_Y(y) \tag{1.8}$$

where $f(Y(y)$ is the marginal density of Y. In Bayesian method, all inferences about the unknown parameter θ based on the posterior distribution. The basic feature of Bayesian inference is the modification of knowledge and non-sample information in the form of probability distributions and its combination with sample information. In the absence of non-sample information, the use of previously non-informative distributions with a combination of sample information can yield results equivalent to classical inference.

1.1.2.5 Prior Distribution

The prior distribution is the probabilistic distribution that expresses the parameter information before obtaining the sample information. This distribution can depend on parameters called hyper parameters. The prior distribution is generally selected according to the presence or absence of previous information about the model parameters. The most common prior distributions are as follows. A priori distribution that considers the same information about different values of θ and in its parameter space is called an non-informative distribution. Uniform distribution and improper distributions are of this kind.

When the combination of the prior distribution with the likelihood function of the posterior distribution belongs to the same family of the previous distribution, such a is called the conjugate prior distribution.

1.1.2.6 Posterior Distribution

From the combination of the prior distribution and the likelihood function $L(\theta, y)$, a distribution is obtained which is called the posterior distribution:

$$\pi(\theta|Y) \propto L(\theta, y)\pi(\theta). \tag{1.9}$$

Suppose a random sample of a statistical model is observed whose probabilistic distribution depends on unknown parameters $\theta = (\theta_1, \theta_2, ..., \theta_k) \in \Theta$. If there are θ previous beliefs about the distribution of unknown parameters such as $\pi(\theta)$, then the inference about the unknown parameters of the model using the distribution of observations and prior distributions is called Bayesian inference. From Bayesian inference point of view, there is no fundamental difference between the observations and the parameters of a statistical model and both of them are considered as random variables. Suppose $D \in R$ represents the observed data and $P(D|\theta)$ is the probable distribution of observations equal to the probability function $L(\theta|D)$. According to Bayesian inference, we need to calculate the combined distribution of observations and parameters as follows.

$$\pi(D, \theta) = P(D|\theta)\pi(\theta). \tag{1.10}$$

1.1.3 Previous Studies

There are a variety of studies dealing with the stress-strength reliability analysis, different methods used to run the inference, offering approaches toward the estimation of parameters, and etc., which more or less are related to the research topic posed in the current study. Due to the importance of such studies which enlighten the ambiguous dimensions, it becomes more significant to add them to the current review so as to demonstrate how different researchers have concentrated on various aspects of the afore-mentioned areas, and then mention the gap existing in the literature. To this end, the following section is mainly concerned with the experimental studies performed, and the literature gap is discussed at the end of this chapter.

In 1981, [1] drew their attention toward the analysis of some inference results on $\Pr(X < Y)$ in the bivariate exponential mode. To this end, the researchers attempted to obtain three different estimators for $\Pr(X < Y)$ in the case of X and Y following a bivariate exponential distribution as well as to determine the asymptotic variances concerning the aforementioned estimators. In addition, the equality of the X and y means along with the confidence limits pertaining to the difference of two means were illustrated.

An investigation was made by [2] in 1990 with respect to the multivariate normal distribution through the lens of dependent random vectors. The ultimate purpose of the study was to shed light on the way two cases of maximum likelihood as well as minimum variance unbiased could be estimated in line with $R = Pr(A'x > B'y)$. Also, a comparison between the MVUE and MLE of R was made. Findings of the study emphasized the effectiveness of the drug pertaining to the biochemical compounds existing in the brain.

Another inference was led by [3] dealing with the Weibul distribution and investigated the case of $P(Y < X)$. The argumentation presented in this research was the sameness of X and Y while their shape parameter being unknown. It was also recommended in this study to employ the Weibul random variable and it gets lower values when compared to another independent Weibull random variable X. This was further supported through the incorporation of Monte Carlo sampling tables, emphasizing that this probability can be accompanied by 90% confidence limits when estimation of maximum likelihood related to 21 combinations concerning the size of the sample derived from each of the two populations is taken into consideration. In addition, another remark was to address the number of ordered observation at which the samples were type two censored. Last but not least, the extent to which a normal distribution was useful and accurate was another consideration, in a comparison to the exact values as a function of sample size in special condition.

Concerning the Weibull distribution, [4] drew the attention toward the case of three-parameter through running analysis on $P(Y<X)$. Using the stress-strength parameter in which there were independent but the-same shaped Weibul-type X and Y distributions, the study suggested no trace of the maximum likelihood estimators. Two cases, namely the modified maximum likelihood estimator and an approximate modified maximum likelihood estimator of R were developed in the research. In line with the unknown parameters, outcomes of the study suggested an asymptotic distribution concerning the modified maximum likelihood estimators.

The ultimate purpose of the study undertaken by [5] was to address the estimation of R = P (Y<X) in the case of three-parameter Weibull distribution. The study used the stress-strength parameter with X and Y being independent, both being of three-parameter Weibul distribution type, shapes being the same, but with a different scale. Findings of the study proved no evidence of the maximum likelihood estimators; instead, a modified maximum likelihood estimator as well as an approximate modified maximum likelihood estimator of R were suggested. The overall results yielded an asymptotic distribution of the modified maximum likelihood estimators pertaining to the unknown parameters.

In a distinguished study, [6] concentrated on a different situation, in which the trucks deviate to wrong directions and cause accidents in this regard. Two accident types, namely single-vehicle (SV) and multi-vehicle (MV) were analyzed through the use of mixed logit models. The extent to which trucks brought about injuries in SV and MV accidents was modeled individually and further risk factors were taken into consideration. It was found that the level of injury observed in MV vs. the SV accidents differed significantly from different variables perspective. The study suggested the observation when the purpose was to determine the injury severity in this context. Outcomes of the study revealed that two factors, namely snow road surface and light traffic indicators could be better modeled as random parameters in SV and MV models. The study paves the way for further researches highlighting that better and effective strategies are to be considered in order to reduce the injury cases.

Similar to the study led by [6], a mixed logit model was put forth in the study undertaken by [7] to address how the Single-vehicle and multivehicle accidents differ in terms of the probability. According to this research, previous studies have not addressed the difference between two types of accident and failed to indicate the heterogeneity at the same time. Relying on the afore-mentioned model, the study made use of the disaggregated data and no accidents, SV accidents, and MV accidents fell into the category of response variable. In support of [6], the researchers found that there were characteristics pertaining to traffic, road, and surface which made a considerable effect on SV and MV types of accidents feasibility. Also, it was reported that significant random parameters were produced by subjects as hourly traffic volume, inside shoulder width, and wet road surface.

In the similar vein of examining the extent to which the road factors could possibly affect the occurrence of accidents in highways performed by researchers ([6]; [7]), the hourly crash likelihood was analyzed by [8]. It was claimed in this study that some elements such as road surface conditions and traffic states are significantly at work when the crash probability is the center of attention. In order to make an analysis of how these elements are based on the time, the researchers suggested crash prediction models with refined temporal data.

Since the road segment has attracted the attention of many researchers, as indicated in the current review, [9] made use of a Bayesian spatial random parameters Tobit model, compared to the logit model, in order to make analysis on the crash rates on roadways. This was an attempt to make a comparison between the model proposed in this study with the ones suggested by similar researchers who employed the fixed parameters and Tobit model in Bayesian context. The research was performed with a focus on three-year period data collection related to the road segments in Florida. It was found that the proposed model could be of better fit when the spatial correlation was included in the Tobit regression. Furthermore, it was concluded that compared to the spatial Tobit regression, the spatial random parameters Tobit regression had a lower DIC value. Finally, it was reported that the influence of speed limit on crash rates could be better understood when the random parameters Tobit model was taken into consideration, and it could be an alternative when it comes to analyze the crash rate.

To enrich previous researchers' findings concerning the road-related investigations, [10] strived to shed light on the way the roadway and weather conditions impose impacts on crash occurrence. The study was performed in China highways with a focus on data obtained from the crash, traffic, and weather. To establish an examination of the link between the number of accidents occurred and risk factors, the researchers made use of the Bayesian spatio-temporal model. Estimation of parameters suggested that several interactions give rise to the freeway crash risk. The study had also implications dealing with the use of real-time traffic management and control measures.

In 2017, the zonal daytime and nighttime crash frequencies were analyzed by [9] through the use of a Bayesian bivariate conditional autoregressive model. The study gathered the data from among 131 traffic analysis zones in Hong Kong. The current study is also a support of the previous studies in literature which mostly had to do with the factors causing traffic and accidents, the number of accidents occurred in a particular period of time, and road characteristics. The current study proposed a model to bridge the gap between the crash frequencies and traffic attributes. Similar to previous studies, this study embarked on the parameter estimates which manifested that there is a bond between the daytime and nighttime crash with the network which enjoy greater global integration. Also, it was found that there was a positive relationship between the heterogeneous and spatial effects. Ultimately, the proposed model was reported to have an acceptable level of fit.

Similar to the research led by [3] who directed his attention toward the analysis of Weibull distribution, the same distribution was investigated by [11]. It was endeavored to test the reliability of the aforementioned distribution through the lens of the progressively censored samples. Such an affair was facilitated through addressing the X and Y as the independent Weibull distributions whose scale parameters are different but overlap in their shape parameter. Accordingly, two constructs, namely the maximum likelihood estimator and the approximate maximum likelihood estimator of R were yielded. The study further proposed the Bayes estimator of R and the pertinent credible intervals through the incorporation of the Gibbs sampling technique. It is worth mentioning in this study that when the shape parameter gets revealed, the estimation of R is put into practice. In order for more clarification purposes on the details, the real data set analyses were performed and the simulation of Monte Carlo was presented afterwards. It was found that through the integration of an iterative procedure, it is feasible to obtain the MLE of *R*. It was also found that there is a close association among the MSE, and biases of the MLE, and AMLE. Another finding liable to be addressed is that it is possible to make a comparison between the MLS and AMLE in line with the Bayes estimator from the biases and MSEs perspective.

In 2012, [12] developed a study to trigger the stress-strength parameter, where X and Y were taken into consideration as the independent exponential random variables, for which both uniformly minimum variance unbiased estimator and the maximum-likelihood estimator (MLE) were obtained. Relying on three independent inverse gamma priors, Bayes estimates of *R*, and the pertinent credible interval were taken into consideration accordingly. The performances of the different estimators was examined which confirmed the same behavior of three estimators. Also, MLE was suggested to be used due to its easy way of computation.

The review of literature also suggests the estimation of stress-strength parameters or their reliability ([13]; [12]; [5]; [11]), however, the study of [13] was an attempt to address this issue through relying on the data available in the form of record values. In order to take into consideration the two-parameters exponential distributions, the researcher assumed two cases; the one being the location parameter, the second being the scale parameter to be common. Also, in order to obtain the equal tailed confidence intervals, the related pivotal quantities were assumed. Finally, the researchers recommended the study of Bayesian intervals in this regard.

In 2016, [14] took into consideration the Gumbel distribution analysis in confirmation to the study performed by [15]. The researchers made a modification to the so-called Gumbel distribution and proposed the one with Gumbel type-2 distribution. As suggested in the study, the EG type-2 distribution is divided by three subcategories, namely Gumbeltype-2 distribution, the Exponentiated Fréchet (EF) distribution, and the Fréchet distribution. The overall results manifested the effectiveness of EG type-2 distribution compared to the other distributions in terms of the data fit.

In his research, [16] provided comprehensive explanations concerning the survival models, likelihood-based tests, extreme- based tests. The researcher attended to give deep insights into how the previous models are limited and how possibilities can come into a variety of ranges. The chapters entail the distributions of sample extremes, introduction of tests, and applications of tests.

The review of literature has suggested a number of studies performed in the field of research topic. Despite the fact that a plethora of such studies has been carried out to address various aspects concerning the inference on generalized distributions, scant attention has been directed toward the investigation of reliability analysis. In addition, to the best of researcher's knowledge, literature lacks the element of lower record values, which itself is counted as the gap of literature. Therefore, the current study endeavored initially to make analysis on so as to fill the literature gap, and further adds to the literature new insights into the inference on the stress-strength reliability.

1.2 Objective of the Thesis

This thesis aimed at investigating the estimation of the two parameters of exponentiated Gumbel distribution by using lower records values. To this end, the maximum likelihood estimator of the parameters is considered. Also, to construct asymptotic confidence intervals, the Fisher information matrix of the unknown parameters is used. Bayes estimator of the parameters and the corresponding credible intervals are obtained by using the Gibbs sampling technique.

1.3 Hypothesis

There will possibly be better fitting capabilities for those generalized distributions compared to those standard distributions based on literature. In this thesis, exponentiated Gumbel distribution was taken into consideration. Generalized distributions are usually more flexible to cover more data, and they result in a better fit. The goodness of fit (GOF) tests can be used to compare the difference between the standard distributions and the generalized ones. GOF tests are also common to compare the fitness between the MLE and Bayesian estimators. On the other hand, there are frequentistic and Bayesian approaches to estimate the unknown parameters of the distributions. Therefore, in this thesis, we have the following hypotheses:

- The Kolmogorov-Smirnov (K-S) goodness of fit test results confirm the well-fitting of the Exponentiated Gumbel distribution to model the real strength data.
- The methods of estimating (ML and Bayes) have slightly different estimations compered to each other.

1.4 Structure of the Thesis

The rest of this thesis is structured as follows:

- In Chapter 2, generalizing and compounding statistical distributions is reviewed. Also, corresponding developments of these methods are discussed.
- In Chapter 3, the Exponentiated Gumbel distribution and its applications are provided. In order to a better understanding of the distribution behaviour, some characteristics of the EG distribution are mentioned.
- In Chapter 4, the inference of stress strength reliability for two parameters of Exponentiated Gumbel distribution based on lower record values is studied.
- In Chapter 5, two real datasets are implemented to the estimated parameters of Exponentiated Gumbel distribution based on lower record values. The methods of estimating is compered via these two real data set.
- In Chapter 6, the thesis is concluded by summarizing what has been done and the significant results are compared to those in similar studies in literature.

2
GENERALIZATION AND COMPOUNDING STATISTICAL DISTRIBUTIONS

2.1 Introduction

The generalized gamma distribution was first proposed by [17] to estimate the extent to which the income rate distribution fit, followed by which was the initiation of studies on generalized continuous probability distributions. Several studies were undertaken afterwards concerning different types of the generalized distributions, which include the ones led by [18] and [19] dealing with the generalization of the inverse Gaussian distribution. Other instances include the generalization of Pareto distribution [20], [21], and [22]. In the same vein, generalization of beta of the first and second kind was developed by [23] so as to analyze the income distribution. With respect to the stock price returns describing, [24] suggested the generalized beta of the first and second kind.

Greater flexibility in modeling the observed data is a feature which makes the generalization of a "standard distribution" more important and reasonable ([25]). The other case to mention is to determine a suitable model derived from a population, in which inference making is of interest. To this end, a general model characterized by a simpler model as a limiting case is liable to be used, as confirmed by [26]. Subsequently, having fit the general model as regards the data, it becomes feasible to determine the suitability of the special case in mind.

It is evident that the best choice to incorporate the flexibility in a multivariate model is obtaining the dependence function, tagged also as the copula function, which yields maximum flexibility when it comes to developing marginal distributions. In this regard, highly specific and mostly used copula function can be obtained referring to the Marshall–Olkin distribution. Of the other clear characteristics related to the discussion is that it gives a dependence function which features both as an absolutely continuous and a singular part. As regards the plain, in case the survival times or another types of variable are modelled, it is observed that a positive probability is obtained on the part of Marshall–Olkin copula, based on which an occurrence of event can be yielded almost simultaneously, or it goes that the same values are given to the two variables being modelled.

2.2 Generalized Exponential Class

Lifetime analysis is one of the most important topics in statistics and probability. To this end, lifetime distribution is one of the main issues in this section. Lifetime distributions include distribution, Weibull, and gamma, which many researchers have worked on. Of course, these two distributions have other uses as well. But in some cases, these two distributions do not answer many related issues. Because the main problem of the gamma distribution is that if the shape parameter of that number is not an integer, its distribution function or survival function cannot be expressed as an explicit function. Therefore, the distribution function, survival function, or failure rate function should be calculated using the numerical methods.

This has led to the unpopularity of the gamma distribution compared to the Weibull distribution. For Weibull distribution, distribution function, survival function, and failure rate function can be easily calculated. Weibull distribution is often preferred to gamma distribution because of its ease of use, especially in the presence of censored data. It has also been observed in many positive data that the Weibull distribution fits very well. But Weibull distribution also has drawbacks. If the shape parameter is unknown, then the Weibull distribution does not belong to the exponential family. In this case, to test the one-sided hypothesis with respect to the shape parameter, even if the scale parameter is known, the most powerful test is not uniformly Uniformly Most Powerful Test (UMPT). In addition, if the shape parameter is greater than one, the risk rate functions of the two gamma and Weibull distributions are ascending. However, the hazard rate function in the gamma distribution varies from zero to the scale parameter and is finite, but in the Weibull distribution, it changes from zero to infinity and is not bounded.

Based on what has been mentioned, the statisticians sought to introduce new distributions that have the necessary flexibility and to cover these disadvantages so that it can be used as an alternative distribution. One of the generalizations of statistical distributions is to add a shape parameter to the desired distribution. Suppose X is a continuous random variable with a distribution function $F(x)$. In this case $(F(x))^{\beta}$, $\beta > 0)$ can be introduced as a cumulative distribution function of the new generalized distribution.

The Random Variable X has the generalized exponential distribution with parameters

β, σ, and μ if it follows the CDF of

$$F(x;\mu,\sigma,\beta) = (1 - e^{-(x-\mu)/\sigma})^{\beta} \tag{2.1}$$

where $x > \mu, \sigma > 0$, $\beta > 0$. Then, $X \sim GE(\mu,\sigma,\beta)$, which μ, σ, and β denote the location, scale and shape parameters, respectively.

2.3 Marshall-Olkin Class

What is followed in this section is to present the other method concerning the parameter introduction into a family of distributions. More precisely, considering the $\overline{F}$ survival function, the one-parameter family of survival functions

$$\overline{G}(x;\alpha) = \frac{\alpha\overline{F}(x)}{1-\alpha\overline{F}(x)} = \frac{\alpha\overline{F}(x)}{F(x)+\alpha\overline{F}(x)} \quad , \qquad (-\infty < x < \infty,\ 0 < \alpha < \infty), \tag{2.2}$$

where $\overline{\alpha} = 1-\alpha$ is suggested. Note that, when $\alpha = 1$, $\overline{G} = \overline{F}$ [27].

On the condition that F is characterized by a density, then the survival function $\overline{G}$ given by (2.3) would enjoy the easily-computed densities. Especially, if F has a density f and hazard rate r_F, the G has the density g given by ([27]):

$$g(x;\alpha) = \frac{\alpha \mathrm{f}(x)}{\left\{1-\overline{\alpha}\overline{F}(x)\right\}^2}, \tag{2.3}$$

And hazard rate

$$r(x;\alpha) = \frac{r_F(x)}{\left\{1-\overline{\alpha}\overline{F}(x)\right\}^2} \quad , \qquad (-\infty < x < \infty). \tag{2.4}$$

2.4 Beta Class

In this class, one can represent the random variable X as the logit of the beta random variable which is given by [25]:

$$G(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^{F(x)} t^{\alpha-1}(1-t)^{\beta-1}dt \qquad 0 < \alpha,\ \beta < \infty \tag{2.5}$$

The probability density of the generalized class of distribution $G(x)$ is $g(x) = G'(x)$, which is

$$g(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} F(x)^{\alpha-1}(1-F(x)^{\beta-1}F'(x) \tag{2.6}$$

In case one assumes α and β as the integers, the corresponding probability distribution is illustrated by the "order statistics of the random variable X." [28] suggested the use of a distribution generated in this way when the attention was directed toward the survival data modelling through the incorporation of generalized log-logistic model ([25]).

2.5 Zografos-Balakrishnan Class

A number of researchers, namely [29] and [29] attempted to develop a family of univariate distributions obtained by making use of the gamma random variables. Further, they addressed the gamma- G distribution characterized by probability density function (PDF) $f(x)$ and CDF $F(x)$ in line with the baseline cumulative distribution function (CDF) $G(x)$, and $x \in R$, which is shown as

$$f(x) = \frac{1}{\Gamma(\alpha)} \{-\log[1-G(x)]\}^{a-1} g(x) \tag{2.7}$$

and

$$F(x) = \frac{\gamma(a, -\log[1-G(x)])}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_0^{-\log[1-G(x)]} t^{a-1} \exp(-t)\, dt, \tag{2.8}$$

respectively, for $a > 0$, where $g(x) = dG(x)/dx$. We know that $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ denotes the gamma function, and also $\gamma(a,z) = \int_0^z t^{a-1}e^{-t}dt$ denotes the incomplete gamma function. Therefore, the corresponding hazard rate function (HRF) is

$$h(x) = \frac{\{-\log[1-G(x)]\}^{a-1} g(x)}{\Gamma(a, -\log[1-G(x)])}, \tag{2.9}$$

where $\Gamma(a,z) = \int_z^\infty t^{a-1}e^{-t}dt$ denotes the complementary incomplete gamma function.

The only difference between the gamma-G distribution and the G distribution is another parameter which denotes the shape parameter. One can show this parameter with $a > 0$. In case X is known to be a PDF-characterized random variable (2.7), one can write $X \sim$ gamma- $G(a)$. A specified G distribution can yield new gamma- G distribution.

Despite the all findings, not much is definite concerning the general mathematical properties of (2.7) and (2.8). A paricutlar formula was developed by [30] in line with the moments of special gamma-G models which is true for the natural a, a general expression for Shannon entropy and also provided a maximum entropy characterization.

2.6 Exponentiating Class

Another type of exponentiated distribution was introduced by [31]. Their study manifested that the cumulative density function CDF $F(x)$ of the exponentiated family of distributions can be written as follows:

$$F(x;\omega;\alpha) = 1-(1-G(x;\omega))^{\alpha}, \tag{2.10}$$

where $x \in \mathbb{R}$, $\alpha > 0$, and $\omega \in \Omega$. In addition, ω and Ω specify the vector of parameters and parameter space of the baseline distribution ($G(x;\omega)$), respectively. The corresponding probability density function $f(x)$ by differentiating (2.10) with respect to x is given by

$$f(x;\omega;\alpha) = \alpha g(x;\omega)(1-G(x;\omega))^{(\alpha-1)} \tag{2.11}$$

where $x \in \mathbb{R}$, $\alpha > 0$, and $\omega \in \Omega$.

3
EXPONENTIATED GUMBEL DISTRIBUTION AND APPLICATIONS

3.1 Introduction

There are a variety of cases where the extreme value theory can be used, including the materials engineering, traffic control, and economics. When it comes to perform modelling in environmental statistics, extreme values assigned to natural phenomena such as the wind speed, river or sea water level, wave height, or moisture become immensely significant since gaining insight into these values play a key role in preventing the disasters [32].

Accordingly, the probabilistic extreme value theory is thought to be a sophisticated mix of natural phenomena, involving the rainfall, floods, wind gusts, air pollution, and corrosion, as well as sensitive state-of-the-art mathematical conclusions concerning the point processes and regularly variable functions. The theory could be sounded in the activities of E. J. Gumbel, who had continuously devoted his life and activities in the era of pre-World War II events [33].

3.2 Gumbel Distribution

The afore-mentioned theory, i.e. the probabilistic extreme value theory is tied above all to the stochastic behavior of the maximum and the minimum of i.i.d. random variables. It is through the upper and lower tails of the underlying distribution that it becomes feasible to determine the distributional properties of extremes (maximum and minimum), extreme and intermediate order statistics, and exceedances over (below) high (low) thresholds.

Referring to the extreme and intermediate order statistics or exceedances over high thresholds statistical procedures, an estimation can be made dealing with the tail of the underlying distribution function or functional parameters. One can benefit from directing attention toward the tails since in this way it facilitates the suggestion of certain parametric statistical models, specifically tailored for that part of the distribution [33].

The concept of extreme value theory can be interpreted based on the normal origins. If we have $N = 1000$ random variables which are following the normal distribution, their mean is equal to zero, and their standard deviation is unit, choose the maximum value from these $N = 1000$ values. Consequently with repeating this procedure until getting 1000 maximum values as well. Therefore, these maximum values converge to a distribution, namely the Type I extreme value distribution – Gumbel distribution (Figure 3.1).

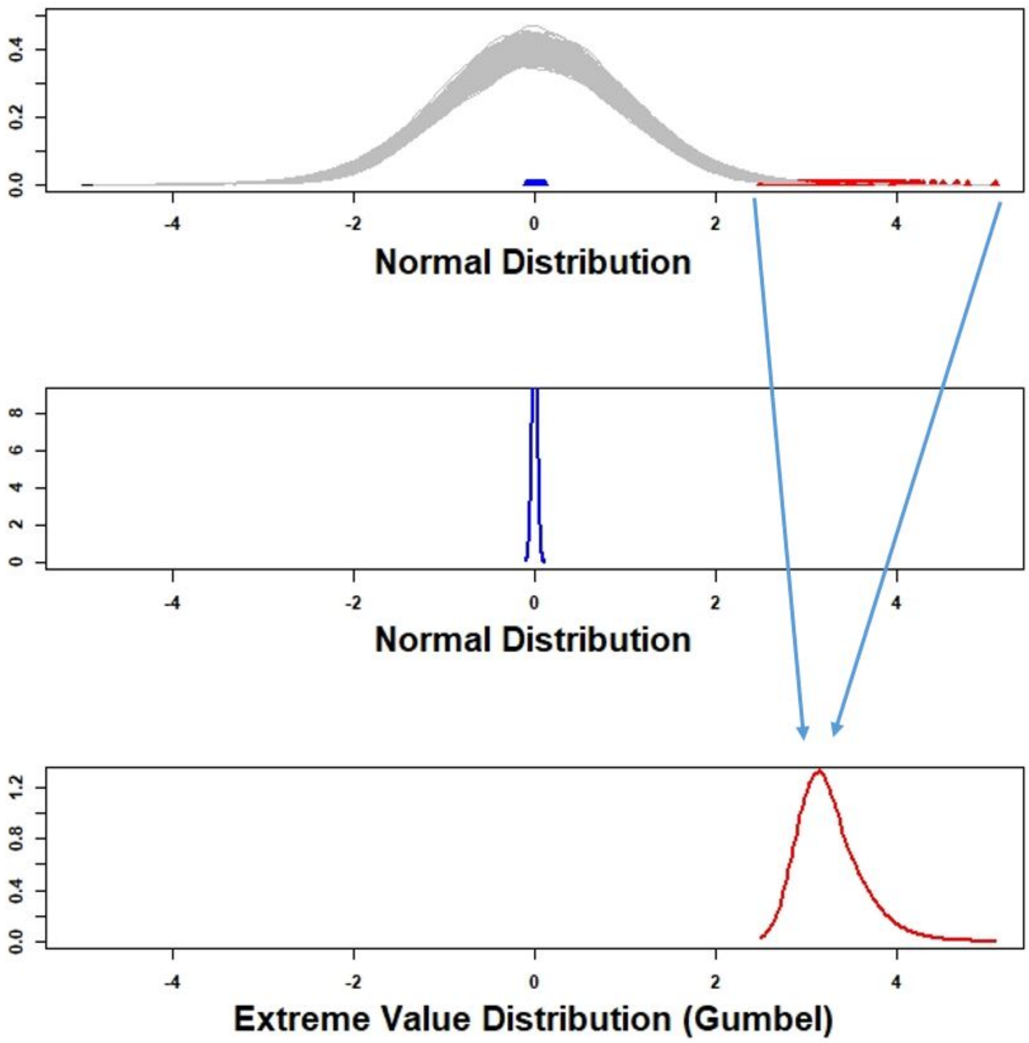

Figure 3.1 The concept of extreme value distribution- Gumbel distribution

There is a wide range of use concerning the Gumbel distribution through the modelling extreme values of natural events and logarithm of lifetime data in the case of survival analysis. The cumulative distribution function (CDF) of the Gumbel distribution can be derived as:

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \tag{3.1}$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is location parameter, and $\sigma > 0$ is scale parameter. The corresponding PDF of the Gumbel distribution is also given by

$$f(x) = \frac{1}{\sigma}\exp\left\{-\left[\frac{x-\mu}{\sigma} + \exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}, \tag{3.2}$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $\sigma > 0$. Figures 3.2 and 3.2 show the PDF and CDF of the Gumbel distribution for certain parameter values, respectively.

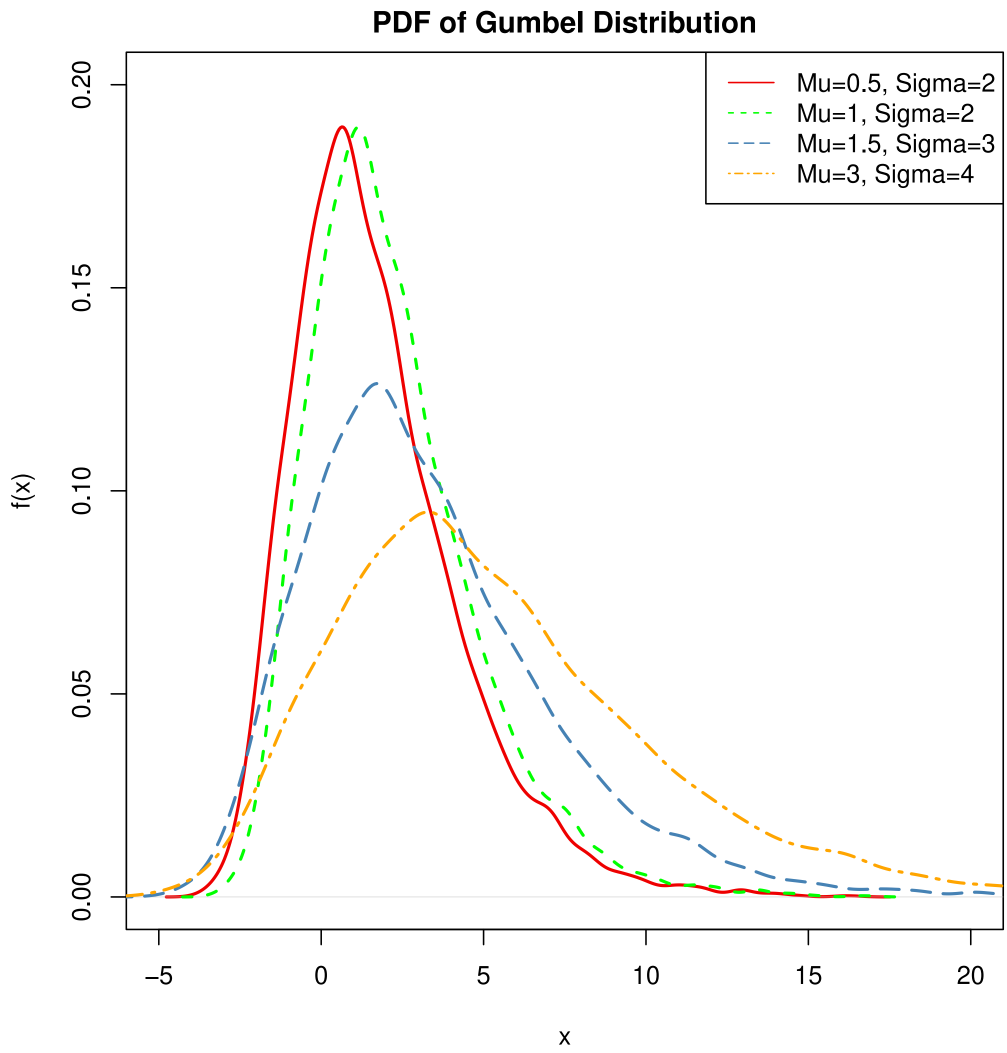

Figure 3.2 Possible shapes of the PDF of Gumbel distribution for different values of μ and σ

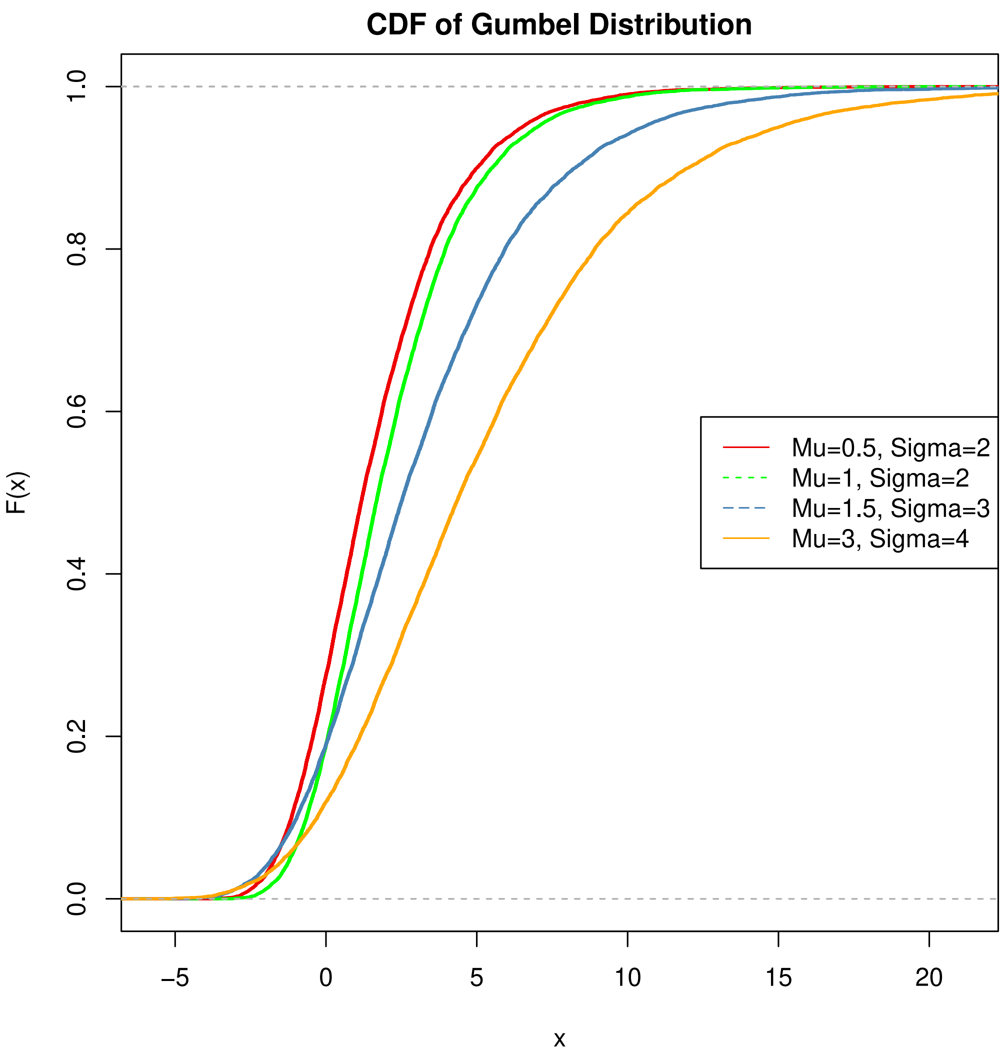

Figure 3.3 Possible shapes of the CDF of Gumbel distribution for different values of μ and σ

3.3 Exponentiated Gumbel Distribution

There are different methods used to embark on generalizing distribution when the purpose is to determine the better data fit, as argued in chapter 2. Adhering to the exponentiating, [34] strived to generalize this distribution, represented in the form of $F(x;\alpha) = 1 - [1 - G(x;\alpha)]$.

Let $G(x)$ depicts the Gumbel distribution as equation (3.1). The new distribution is defined by the CDF of

$$F(x) = 1 - \left[1 - \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}\right]^{\alpha}, \tag{3.3}$$

where $\alpha > 0$. We refer to [33] as the exponentiated Gumbel (EG) distribution. The corresponding PDF is

$$f(x)=\frac{\alpha}{\sigma}\left[1-\exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}\right]^{\alpha-1}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}. \tag{3.4}$$

One advantage considering the standard probability distributions exponentiating is that it performs as a problem solver in cases that there is no fit when using these distributions for modeling complex data [14]. In order to prove the effectiveness of the generalized distribution in modeling the climatology data, the researchers integrated it with their analysis on rainfall data from Orlando, Florida.

In this thesis, we use a slightly different way to define the exponentiated distributions, i.e., $F(x;\alpha) = [G(x;\alpha)]$, which are called the proportional reversed hazard rate models [35]. The random variable X follows the two-parameter Exponentiated Gumbel distribution if it has the following CDF

$$F(x;\alpha,\lambda)=\exp\left\{-\alpha\exp(-\lambda x)\right\}, \tag{3.5}$$

where $x\in\mathbb{R}$, $\alpha>0$, and $\lambda>0$. The PDF corresponding to the CDF (3.5) is

$$f(x;\alpha,\lambda)=\alpha\lambda\exp\{-\lambda x\}\exp\left\{-\alpha\exp(-\lambda x)\right\}. \tag{3.6}$$

Here, α and λ are the shape and scale parameters, respectively. We will denote this distribution by $EG(\alpha,\lambda)$.

4 INFERENCE OF STRESS-STRENGTH RELIABILITY FOR TWO PARAMETERS OF EXPONENTIATED GUMBEL DISTRIBUTION BASED ON LOWER RECORD VALUES

4.1 Introduction

In this chapter, we obtain the maximum likelihood estimation of reliability $R = P(X > Y)$ using lower records, where X and Y are from exponentiated Gumbel distribution. Then, the Fisher information matrix of the unknown parameters of EG distribution will be obtained. Afterward, we attempt to find the Bayes estimator of the parameters. At the end of this chapter, we discuss the inference of R when λ is known. To this end, we obtain the MLE and Bayesian estimation of R.

4.2 Lower Record Values

In chapter 1, several works were mentioned in the literature review section. The only difference in the mentioned works was the different distributions which the authors have chosen for the random quantities. In some situations, one can not obtain the complete data and has to consider certain sampling schemes in order to get incomplete data for X and Y. [11], [12] and [36] have studied the problem of making inference on R based on progressively Type II censored data. [13], has studied ML and Bayesian estimation of R based on the record data by considering one parameter generalized exponential distribution.

Another type of incomplete data is record values which appear usually in many real life applications. Record values arise in climatology, sports, business, medicine, industry, and life testing surveys, among others. These records are commemorating over the period of the time that have been studied. The history of records may show the advancement in science and technology. By considering the record values in various areas of humankind activities, we can evaluate the performance of the societies.

In 1952, Chandler introduced the record values into the statistical world and now after six decades on his original work, hundreds of papers were devoted to various aspects of the record's theory. He provided a foundation for a mathematical theory of records. He defined the record values as consecutive extremes appearing in a sequence of independent identically distributed (*i.i.d.*) random variables. These smallest or largest occurred values are called "lower" or "upper" record values, respectively. Let $X_1, \ldots, X_n$ be a sequence of *i.i.d.* continuous random variables with a cumulative distribution function (cdf) $F(x)$ and its corresponding probability density function (pdf) $f(x)$. For every positive integer $k \geq 1$, the sequence of kth lower record times, $\{L(k), k \geq 1\}$, is defined as follows

$$L(1) = 1, \qquad L(k+1) = \min\left\{j \mid j > L(k), X_j < X_{L(k)}\right\}. \tag{4.1}$$

Then, the kth lower record value will be denoted by $X_{L(k)}$ and the sequence $\{X_{L(k)}, L(k) \geq 1\}$ is called the lower record values. For the sake of simplicity, from now on, we shall refer to $X_{L(k)}$ as X_k. As mentioned before, the record values have many applications in industry and engineering. Consider an electronic system which is subject to some shocks like low or high voltage in which both are dangerous for predefined performance of it. These shocks could be considered as realizations of *i.i.d.* variable and then one can use the record values models to study them. We refer the readers to [38] and [39] for more details on the record values and their applications.

4.3 Maximum Likelihood Estimation

In this section, we consider the maximum likelihood estimation of $R = P(X > Y)$ when $X \sim EG(\alpha, \lambda)$ and $Y \sim EG(\beta, \lambda)$, and X and Y are independently distributed. Formal integration shows that

$$R = P(X > Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} \alpha\beta\lambda^2 e^{-\lambda x} e^{-\alpha e^{-\lambda x}} e^{-\lambda y} e^{-\beta e^{-\lambda y}} \, dy \, dx = \frac{\alpha}{\alpha + \beta}. \tag{4.2}$$

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be two independent set of the lower records from $EG(\alpha, \lambda)$ and $EG(\beta, \sigma)$ respectively. Therefore, the likelihood function of parameters becomes (see [39])

$$L(\alpha, \beta, \lambda) = \left[f(x_n; \alpha, \lambda) \prod_{k=1}^{n-1} \frac{f(x_k; \alpha, \lambda)}{F(x_k; \alpha, \lambda)} \right] \left[f(y_m; \beta, \lambda) \prod_{j=1}^{m-1} \frac{f(y_j; \beta, \lambda)}{F(y_j; \beta, \lambda)} \right]$$

From (3.5) and (3.6), the likelihood function is obtained as

$$L(\alpha,\beta,\lambda)=\alpha^n\beta^m\lambda^{(m+n)}\exp\left\{-\lambda\left(\sum_{k=1}^{n}x_k+\sum_{j=1}^{m}y_j\right)-\alpha e^{-\lambda x_n}-\beta e^{-\lambda y_m}\right\}.$$

The log likelihood function is given by

$$\ell(\alpha,\beta,\lambda)=n\ln\alpha+m\ln\beta+(m+n)\ln\lambda-\lambda\left(\sum_{k=1}^{n}x_k+\sum_{j=1}^{m}y_j\right)-\alpha e^{-\lambda x_n}-\beta e^{-\lambda y_m}.$$

Then, the likelihood equations will be

$$\frac{\partial\ell}{\partial\alpha}=\frac{n}{\alpha}-e^{-\lambda x_n}=0 \tag{4.3}$$

$$\frac{\partial\ell}{\partial\beta}=\frac{n}{\beta}-e^{-\lambda y_m}=0 \tag{4.4}$$

$$\frac{\partial\ell}{\partial\lambda}=\frac{n+m}{\lambda}-\sum_{k=1}^{n}x_k-\sum_{j=1}^{m}y_j+\alpha x_n e^{-\lambda x_n}+\beta y_m e^{-\lambda y_m}=0 \tag{4.5}$$

From above equations, we get

$$\hat{\lambda}=\frac{n+m}{n(\bar{x}-x_n)+m(\bar{y}-y_m)},\qquad \hat{\alpha}=ne^{\hat{\lambda}x_n},\qquad \hat{\beta}=me^{\hat{\lambda}y_m}. \tag{4.6}$$

Note that $\hat{\lambda}$ is the harmonic mean of $\hat{\lambda}_1=1/(\bar{x}-x_n)$ and $\hat{\lambda}_2=1/(\bar{y}-y_m)$, which are the MLEs of independent samples of sizes n and m, respectively.

Therefore, by applying the invariant property of ML estimators, the ML estimation of R will be

$$\hat{R}=\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}. \tag{4.7}$$

In this section, for the EG distribution unknown parameters, we calculate the Fisher information matrix. It will be used to construct asymptotic confidence intervals.

[39] showed that the pdf of the sth lower record, X_s, is given by

$$f_{X_s}(x)=\frac{f(x)[-\ln F(x)]^{s-1}}{\Gamma(s)}=\frac{1}{\Gamma(s)}\alpha^s\lambda\, e^{-s\lambda x-\alpha e^{-\lambda x}},\ x\in\mathbb{R}. \tag{4.8}$$

Therefore, the Fisher information matrix of $\boldsymbol{\theta} = (\alpha, \beta, \lambda)$ will be

$$I(\boldsymbol{\theta}) = -\begin{bmatrix} E(\frac{\partial^2 \ell}{\partial \alpha^2}) & E(\frac{\partial^2 \ell}{\partial \alpha \partial \beta}) & E(\frac{\partial^2 \ell}{\partial \alpha \partial \lambda}) \\ E(\frac{\partial^2 \ell}{\partial \beta \partial \alpha}) & E(\frac{\partial^2 \ell}{\partial \beta^2}) & E(\frac{\partial^2 \ell}{\partial \beta \partial \lambda}) \\ E(\frac{\partial^2 \ell}{\partial \lambda \partial \alpha}) & E(\frac{\partial^2 \ell}{\partial \lambda \partial \beta}) & E(\frac{\partial^2 \ell}{\partial \lambda^2}) \end{bmatrix} \tag{4.9}$$

where

$$I_{11} = \frac{n}{\alpha^2},\ I_{12} = I_{21} = 0,\ I_{13} = I_{31} = \frac{n}{\alpha\lambda}\left(\psi(n+1) - \ln\alpha\right),\ I_{22} = \frac{m}{\beta^2}, \tag{4.10}$$

$$I_{23} = I_{32} = \frac{m}{\beta\lambda}\left(\psi(m+1) - \ln\beta\right), \tag{4.11}$$

$$I_{33} = \frac{m+n}{\lambda^2} + \frac{n}{\lambda^2}\left[\psi'(n+1) + (\psi(n+1) - \ln\alpha)^2\right] + \frac{m}{\lambda^2}\left[\psi'(m+1) + (\psi(m+1) - \ln\beta)^2\right]. \tag{4.12}$$

The asymptotic covariance matrix of the ML estimators could be achieved via inverting the Fisher information matrix as following

$$I^{-1}(\boldsymbol{\theta}) = U(\boldsymbol{\theta}) = \frac{1}{\det(I(\boldsymbol{\theta}))}\begin{bmatrix} I_{22}I_{33} - I_{23}^2 & I_{13}I_{32} & -I_{13}I_{22} \\ I_{23}I_{31} & I_{11}I_{33} - I_{13}^2 & -I_{11}I_{23} \\ -I_{22}I_{31} & -I_{11}I_{32} & I_{11}I_{22} \end{bmatrix}, \tag{4.13}$$

where

$$\det(I(\boldsymbol{\theta})) = I_{11}I_{22}I_{33} - I_{11}I_{23}^2 - I_{22}I_{13}^2. \tag{4.14}$$

Now,by using the delta method, the asymptotic variance of $\hat{R}$ could be obtained as follows.

$$\mathbb{V}(\hat{R}) = C^T U C, \tag{4.15}$$

where $C^T = \left(\frac{\partial R}{\partial \alpha}, \frac{\partial R}{\partial \beta}, \frac{\partial R}{\partial \lambda}\right) = \frac{1}{(\alpha+\beta)^2}(\beta, -\alpha, 0)$.

Note that the $\mathbb{V}(\hat{R})$ is a function of unknown parameters and it needs to be estimated. It can be done by plunging the ML estimators of the parameters. Therefore, the $(1-\gamma)\%$ asymptotic confidence intervals of R will be in the form of $\hat{R} \pm z_{1-\gamma/2}\frac{\sqrt{\mathbb{V}(\hat{R})}}{\sqrt{n}}$.

4.4 Bayesian estimation

In this section, we attempt to find the Bayes estimator of the parameters. To do so, we consider that the parameters are apriori independent and they follow the gamma distributions, i.e. $\alpha \sim Gamma(a_1, b_1)$, $\beta \sim Gamma(a_2, b_2)$ and $\lambda \sim Gamma(a_3, b_3)$. Therefore, the full posterior distribution of the parameters will be

$$\pi(\alpha,\beta,\lambda|data) \propto \alpha^n \beta^m \lambda^{(m+n)} \exp\left\{-\lambda\left(\sum_{k=1}^{n} x_k + \sum_{j=1}^{m} y_j\right) - \alpha e^{-\lambda x_n} - \beta e^{-\lambda y_m}\right\} \quad (4.16)$$

$$\times \quad \alpha^{a_1-1} e^{-b_1\alpha}\, \beta^{a_2-1} e^{-b_2\beta}\, \lambda^{a_3-1} e^{-b_3\lambda} \quad (4.17)$$

The above posterior does not admit a closed form and cannot be used directly in the estimation procedure. Then, to simulate a random sample from such distributions and perform an approximated inference, the Gibbs sampler could be used. The full conditional distributions of the parameters are as follows:

$$\alpha\,|\beta,\lambda,data \sim Gamma(n + a_1, b_1 + e^{-\lambda x_n}), \quad (4.18)$$

$$\beta\,|\alpha,\lambda,data \sim Gamma(m + a_2, b_2 + e^{-\lambda y_m}), \quad (4.19)$$

$$\pi(\lambda\,|\alpha,\beta,data) \propto \lambda^{m+n+a_3-1} \exp\left\{-\lambda\,(b_3 + n\bar{x} + m\bar{y}) - \alpha e^{-\lambda x_n} - \beta e^{-\lambda y_m}\right\}. \quad (4.20)$$

As $\pi(\lambda|\alpha,\beta,data)$ does not have a closed and standard form, one could not produce a sample from this density by using direct methods. The Metropolis-Hastings algorithm is a method which can be used to produce a sample from such distributions. As shown in Figure 4.1, the normal distribution could be a good candidate for the proposal distribution of Metropolis-Hastings algorithm. Therefore, the algorithm of Gibbs sampling is as follows.

Step 1: Start with an initial value $\lambda^{(0)}$.

Step 2: Set $t = 1$.

Step 3: Generate $\alpha^{(t)}$ from $Gamma(n + a_1, b_1 + e^{-\lambda^{(t-1)} x_n})$.

Step 4: Generate $\beta^{(t)}$ from $Gamma(m + a_2, b_2 + e^{-\lambda^{(t-1)} y_m})$,.

Step 5: Use Metropolis-Hastings algorithm to generate $\lambda^{(t)}$ from $\pi(\lambda\,\middle|\alpha^{(t-1)},\beta^{(t-1)},data)$ by using the $N(\lambda^{(t-1)}, \sigma_0^2)$ as a proposal distribution.

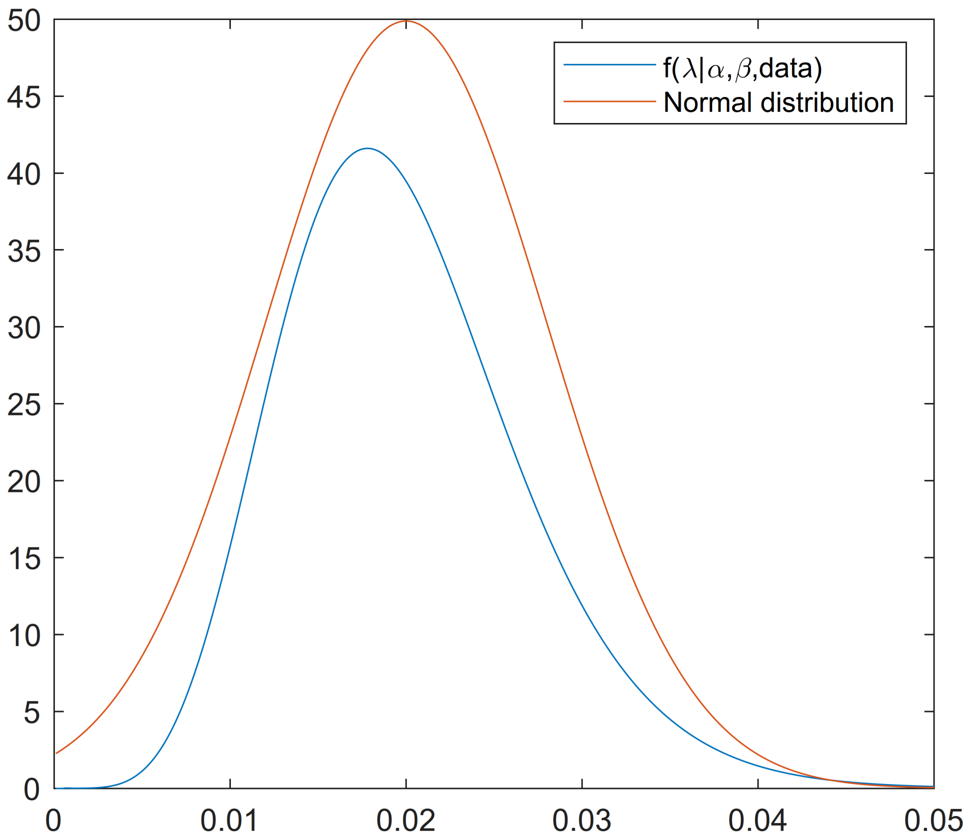

Figure 4.1 Proposal and posterior density functions of scale parameter

Step 5.1: Generate a candidate points λ^* from $N(\lambda^{(t-1)}, \sigma_0^2)$ and u from $\mathscr{U}(0,1)$.

Step 5.2: Set $\lambda^{(t)} = \lambda^*$ if $u \leq \rho(\lambda^{(t-1)}, \lambda^*)$ and $\lambda^{(t)} = \lambda^{(t-1)}$ otherwise, when the acceptance probability is given by $\rho(\lambda^{(t-1)}, \lambda^*) = \min\{1, A\}$, and the acceptance rate is given by

$$\begin{aligned} A &= \frac{\pi(\lambda^*|\alpha^{(t)}, \beta^{(t)}, data)}{\pi(\lambda^{(t-1)}|\alpha^{(t)}, \beta^{(t)}, data)} \cdot \frac{N(\lambda^{(t-1)}|\lambda^*, \sigma_0^2)}{N(\lambda^*|\lambda^{(t-1)}, \sigma_0^2)} \\ &= \frac{\pi(\lambda^*|\alpha^{(t)}, \beta^{(t)}, data)}{\pi(\lambda^{(t-1)}|\alpha^{(t)}, \beta^{(t)}, data)}. \end{aligned}$$

Step 6: Set $t = t + 1$.

Step 7: Repeat steps 3-6, T times.

Once we get a sample from the posteriors, the approximate posterior mean of R and its variance could be computed as following:

$$\hat{E}(R|data) = \frac{1}{T-K}\sum_{t=K+1}^{T} R^{(t)} = \frac{1}{T-K}\sum_{t=K+1}^{T}\frac{\alpha^{(t)}}{\alpha^{(t)}+\beta^{(t)}} \tag{4.21}$$

and

$$\hat{V}ar(R|data) = \frac{1}{T-K}\sum_{t=K+1}^{T}\left(R^{(t)} - \hat{E}(R|data)\right)^2 \tag{4.22}$$

where K is the burn-in period of the chain which helps to vanish the effect of the starting values of the generated Markov chain.

The approximate the highest posterior density (HPD) credible interval of R could be constructed by using the method which was proposed in [40].

Let $R_{(K+1)} < R_{(K+2)} < \cdots < R_{(T)}$ be the ordered output of the chain, $R^{(t)}$. To construct a $100(1-\gamma)\%$ approximate HPD credible interval for R, we consider the following intervals,

$$\left\{\left(R_{(K+1)}, R_{[((1-\gamma)T)]}\right), \left(R_{(K+2)}, R_{[((1-\gamma)T+1)]}\right), \ldots, \left(R_{[(K+\gamma T)]}, R_{(T)}\right)\right\},$$

by choosing the interval with the shortest length, we obtain the HPD credible intervals.

4.5 Inference on R when λ is known

As we showed in section 4.3 the ML estimation of λ does not depend on the value of other parameters, therefore by plunging the MLE of λ, one can assume that the model contains only two parameters. This assumption makes the procedure of inference more easier and straightforward. In other words, we can assume that λ is known and without loss of generality, we set $\lambda = 1$.

4.5.1 MLE of R

As mentioned in section 4.3, the ML estimator of R is

$$\hat{R} = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = \frac{n\,e^{X_n}}{n\,e^{X_n} + m\,e^{Y_m}}. \tag{4.23}$$

By straight computation one can see that

$$2\alpha\,e^{-X_n} \sim \chi^2_{2n} \qquad \text{and} \qquad 2\beta\,e^{-Y_m} \sim \chi^2_{2m} \tag{4.24}$$

By considering (4.24) and the fact that X_n and Y_m are independent, one can show that

$$\frac{R}{1-R}.\frac{1-\hat{R}}{\hat{R}} \sim F_{2n,2m}. \tag{4.25}$$

Therefore, the $100(1-\gamma)\%$ confidence interval for R could be obtained as

$$\left(\frac{1}{1+\frac{m\,e^{Y_m}}{n\,e^{X_n}}F_{1-\frac{\gamma}{2},2m,2n}}, \frac{1}{1+\frac{m\,e^{Y_m}}{n\,e^{X_n}}F_{\frac{\gamma}{2},2m,2n}}, \right) \tag{4.26}$$

where $F_{\gamma,d_1,d_2}$is the $100\gamma^{th}$ percentile of the Fisher distribution with d_1 and d_2 degrees of freedom.

4.5.2 Bayesian Estimation

Since we assumed that the parameters are apriori independent with gamma density, the posterior density of α and β are independent $Gamma(a_1 + n, b_1 + e^{-x_n})$ and $Gamma(a_2 + m, b_2 + e^{-y_m})$, respectively. Therefore, the posterior distribution of R will be

$$\pi(R|data) = c.\, r^{a_1+n-1}(1-r)^{a_2+m-1}(1-zr)^{-(a_1+a_2+n+m)}, \qquad 0 < r < 1, \tag{4.27}$$

where

$$c = \frac{\Gamma(a_1+a_2+n+m)}{\Gamma(a_1+n)\Gamma(a_2+m)}.\left(\frac{b_1+e^{-x_n}}{b_2+e^{-y_m}}\right)^{a_1+n} \quad \text{and} \quad z = 1-\frac{b_1+e^{-x_n}}{b_2+e^{-y_m}}. \tag{4.28}$$

One can use the obtained posterior distribution for Bayesian estimation. We know that in order to estimate the parameters based on Bayesian method considering the assumed loss function, we can calculate different posterior distribution aspects, i.e., mean, median, etc. See [41] or [42] for more details.

By assuming the quadratic loss function, the Bayesian estimation will be the posterior mean which could be computed by considering the following well known equation

$$B(b,c-b)\,{}_2F_1(a,b;c;z) = \int_0^1 x^{b-1}(1-x)^{c-b-1}(1-zx)^{-a}\,dx \qquad c > b > 0, \tag{4.29}$$

in which $B(b,c-b)$ and ${}_2F_1(a,b;c;z)$ are beta and hypergeometric functions, receptively.

Therefore, the Bayesian estimation of R is

$$\begin{aligned}
\hat{R}_{Bayes} &= E(R|data) \qquad (4.30)\\
&= \frac{\Gamma(a_1+a_2+n+m)}{\Gamma(a_1+n)\Gamma(a_2+m)}.\left(\frac{b_1+e^{-x_n}}{b_2+e^{-y_m}}\right)^{a_1+n} B(a_1+n+1,a_2+m)\\
&\times \ {}_2F_1\left(a_1+a_2+n+m, a_1+n+1; a_1+a_2+n+m+1; 1-\frac{b_1+e^{-x_n}}{b_2+e^{-y_m}}\right).
\end{aligned}$$

The variance of the Bayesian estimator could be achieved by using

$$\begin{aligned}
E(R^2|data) &= \frac{\Gamma(a_1+a_2+n+m)}{\Gamma(a_1+n)\Gamma(a_2+m)}.\left(\frac{b_1+e^{-x_n}}{b_2+e^{-y_m}}\right)^{a_1+n} B(a_1+n+2,a_2+m)\\
&\times \ {}_2F_1\left(a_1+a_2+n+m, a_1+n+2; a_1+a_2+n+m+2; 1-\frac{b_1+e^{-x_n}}{b_2+e^{-y_m}}\right).
\end{aligned}$$

To construct the HPD intervals, as the posterior is not tractable, we can generate a sample from the posterior by using indirect sampling algorithm, such as the accept-reject method.

5
REAL DATA ANALYSIS

In this chapter, we analyze a set of real strength data, which were taken from [43, p. 574]. These data are originally from [16] which represent the lifetimes of steel specimens tested at 14 different stress magnitudes. Here, we pickup the dataset corresponding to 35.0 and 35.5 stress levels as Dataset 1 (x) and Dataset 2 (y) in Table 5.1, respectively.

We fitted the EG distribution models for two datasets separately and estimated the scale and shape parameters. The Kolmogorov-Smirnov (K-S) goodness of fit test was applied on the datasets. The reported results in Table 5.2 confirm the well fitting of the EG distribution to model these data. Moreover, Figures 5.1 and 5.2 confirm the appropriate fit of the EG distribution by comparing the empirical and fitted distributions for both datasets.

Now, we can use the lower records based on the datasets to drive the ML and Bayesian estimation of parameters. These records for Dataset 1 are $230, 169, 129, 115$ and based on Dataset 2 are $156, 125, 112$. The corresponding results for ML methods are reported in Table 5.2. According to this table, it is clear that the scale parameters of two data sets are almost the same. By assuming equality of the scale parameter, the MLE and the 95% confidence interval of R based on lower records values become 0.5927 and $(0.4117, 0.7737)$, respectively. Also by using Gibbs sampling, the Bayes estimate and credible interval of R are 0.5893 and $(0.4236, 0.7641)$, respectively.

Dataset 1 (x)					Dataset 2 (y)				
230	169	178	271	129	156	173	125	852	559
568	115	280	305	326	442	168	286	261	227
1101	285	734	177	493	285	253	166	133	309
218	342	431	143	381	247	112	202	365	702

Table 5.1 The real datasets which were reported in [43]. Dataset 1 (x) and Dataset 2 (y) correspond to 35.0 and 35.5 stress level, respectively.

Dataset	Scale	Shape	K-S	p-value
1	0.0071	5.9717	0.1137	0.9326
2	0.0085	6.5564	0.1408	0.7726

Table 5.2 The Kolmogorov-Smirnov test output for real datasets.

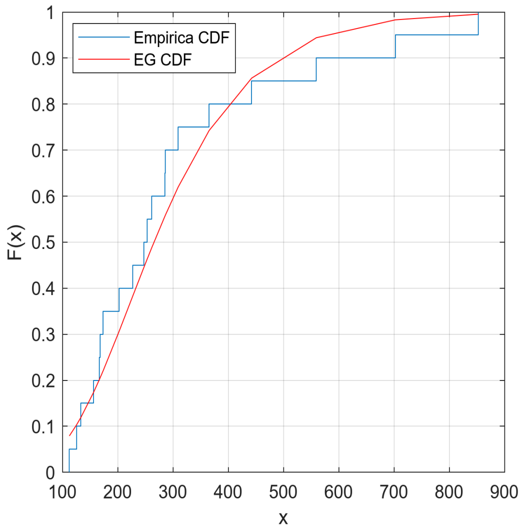

Figure 5.1 Empirical and fitted CDFs for Dataset 1

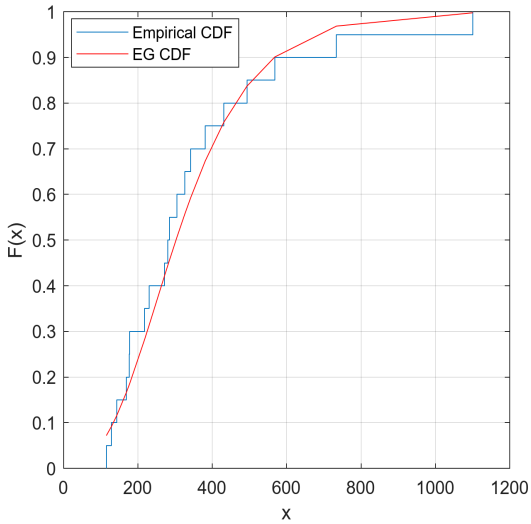

Figure 5.2 Empirical and fitted CDFs for Dataset 2

6
RESULTS AND DISCUSSION

The estimation of the reliability parameter has an extensive literature. It has been studied under different assumptions over distribution of X and Y. [1] studied the ML estimation of R under the assumptions that the stress and strength variables follow a bivariate exponential distribution. By considering a multivariate normal distribution, the MLE has been studied by [2]. The estimation of R, when the distribution is Weibull, were considered by [3]. See [44] and references therein for more details, works on the estimation of R and its applications. In some recent works, [4], estimated R under the assumption that the stress and strength variables are independent and follow the generalized exponential distribution. [5] considered the estimation of R, when X and Y are independent and both follow a three parameter Weibull distributions.

The only difference in the above mentioned works was the different distributions which the authors have been chosen for the random quantities. In some situations, one can not obtain the complete data and has to consider certain sampling schemes in order to get incomplete data for X and Y. [11], [12] and [36] have studied the problem of making inference on R based on progressively Type II censored data. [13], based on the record data, by considering one parameter generalized exponential distribution, has studied ML and Bayesian estimation of R.

The overall objective of the present study was to pinpoint how the parameters relevant to the two-parameter could be estimated in the case of exponentiated Gumbel distribution when the lower record values were applied. In order to attain the models parameter estimation and asymptotic confidence intervals generation purposes, the maximum likelihood and Fisher information matrix of the unknown parameters were attended, respectively. To fulfill the obtaining need related to the Bayes estimator of the parameters and the resulting credible intervals, the Gibbs sampling technique was aimed. Last but not least, outcomes of the study concerning the comparison between two estimation methods, namely ML and Bayes represented the slight difference between the Bayesian estimation from its counterpart, i.e. the ML.

REFERENCES

[1] A. Awad, M. Azzam, M. Hamdan, “Some inference results on $P(X < Y)$ in the bivariate exponential model,” *Communications in Statistics–Theory and Methods*, vol. 10, no. 24, pp. 2515–2525, 1981.

[2] R. Gupta, R. C. Gupta, “Estimation of $P(a'x > b'y)$ in the multivariate normal case,” *Statistics*, vol. 21, no. 1, pp. 91–97, 1990.

[3] J. I. McCool, “Inference on $P\{Y < X\}$ in the weibull case,” *Communications in Statistics-Simulation and Computation*, vol. 20, no. 1, pp. 129–148, 1991.

[4] D. Kundu, R. D. Gupta, “Estimation of $P[Y < X]$ for generalized exponential distribution,” *Metrika*, vol. 61, no. 3, pp. 291–308, 2005.

[5] D. Kundu, M. Z. Raqab, “Estimation of $R = P(Y < X)$ for three-parameter weibull distribution,” *Statistics & Probability Letters*, vol. 79, no. 17, pp. 1839–1846, 2009.

[6] B. Saraçoğlu, I. Kinaci, D. Kundu, “Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways,” *Accident Analysis and Prevention*, vol. 43, pp. 1677–1688, 2011.

[7] B. Dong, X. Ma, F. Chen, S. Chen, “Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model,” *Journal of Advanced Transportation*, no. 3, pp. 1–18, 2018.

[8] F. Chen, S. Chen, X. Ma, “Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data,” *Journal of Safety Research*, vol. 65, pp. 153–159, 2018.

[9] Q. Zeng, H. Wen, H. Huang, M. Abdel-Aty, “A bayesian spatial random parameters tobit model for analyzing crash rates on roadway segments,” *Accident Analysis and Prevention*, vol. 100, pp. 37–43, 2017.

[10] H. Wen, X. Zhang, Q. Zeng, N. Sze, “Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence,” *Accident Analysis and Prevention*, vol. 132, 2019.

[11] A. Asgharzadeh, R. Valiollahi, M. Z. Raqab, “Stress-strength reliability of weibull distribution based on progressively censored samples,” *SORT*, vol. 35, no. 2, pp. 103–124, 2011.

[12] B. Saraçoğlu, I. Kinaci, D. Kundu, “On estimation of $R = P(Y < X)$ for exponential distribution under progressive type-ii censoring,” *Journal of Statistical Computation and Simulation*, vol. 82, no. 5, pp. 729–744, 2012.

[13] A. Baklizi, “Estimation of $P(X < Y)$ using record values in the one and two parameter exponential distributions,” *Communications in Statistics—Theory and Methods*, vol. 37, no. 5, pp. 692–698, 2008.

[14] I. Okorie, A. Akpanta, J. Ohakwe, "The exponentiated gumbel type-2 distribution: Properties and application," *International Journal of Mathematics and Mathematical Sciences*, 2016.

[15] S. Nadarajah, S. Kotz, "The exponentiated type distributions," *Acta Applicandae Mathematica*, vol. 92, no. 2, pp. 97–111, 2006.

[16] M. Crowder, "Tests for a family of survival models based on extremes," in *Recent Advances in Reliability Theory*, Springer, 2000, pp. 307–321.

[17] L. Amoroso, "Ricerche intorno alla curva dei redditi," *Annali de Mathematical engineering*, vol. 4, pp. 123–159, 1925.

[18] I. Good, "The population frequencies of the species and the estimation of population parameters," *Biometrika*, vol. 40, pp. 237–260, 1953.

[19] M. Wise, G. Patil, S. Kotz, J. Ord, E. R. Reidel, "Skew distributions in biomedicine, including some negative powers of time," *Statistical Distributions in Scientific Work*, vol. 2, pp. 241–262, 1975.

[20] M. Ljubo, "Curves and concentration indices for certain generalized pareto distributions," *Statistical Review*, vol. 15, pp. 257–260, 1965.

[21] J. Pickands, "Statistical inference using extreme order statistics," *Annals of Statistic*, vol. 3, pp. 119–131, 1975.

[22] J. Hoskings, J. Wallis, "Parameter and quantile estimation for the generalized pareto distribution," *Technometrics*, vol. 29, pp. 339–349, 1987.

[23] J. McDonald, "Some generalized functions for the size distribution of income," *Econometrica*, vol. 52, pp. 647–663, 1984.

[24] R. Bookstaber, J. McDonald, "A general distribution for describing security price returns," *Journal of Business*, vol. 60, pp. 401–424, 1987.

[25] N. Eugene, C. Lee, F. Famoye, "Beta-normal distribution and its applications," *Communications in Statistics – Theory and Methods*, vol. 31, pp. 497–512, 1997.

[26] W. Nelson, *Applied Life Data Analysis*. John Wileyand Sons Inc.: New York., 1982.

[27] A. N. Marshall, I. Olkin, "A new method for adding a parameter to a family of distributions with applications to the exponential and weibull families," *Biometrika*, vol. 84, pp. 641–652, 1997.

[28] K. Singh, C. Lee, E. George, "On generalized log-logistic model for censored survival data," *Biometrical Journal*, vol. 30, pp. 843–850, 1988.

[29] M. Ristic, N. Balakrishnan, "The gamma exponentiated exponential distribution," *Journal of Statistical Computation and Simulation*, vol. 82(8), pp. 1191–1206, 2012.

[30] S. Nadarajah, G. M. Cordeiro, E. M. M. Ortega, "The zografos–balakrishnan-g family of distributions: Mathematical properties and applications," *Communications in Statistics - Theory and Methods*, vol. 44(1), pp. 186–215, 2015.

[31] S. Nadarajah, S. Kotz, "The exponentiated type distributions," *Acta Applicandae Mathematica*, vol. 92, no. 2, pp. 97–111, 2006.

[32] M. Alizadeh, T. G. Ramires, S. M. T. K. M. Mostafaee, M. Samizadeh, E. M. M. Ortega, “A new useful four-parameter extension of the gumbel distribution: Properties, regression model and applications using the gamlss framework,” *Communications In Statistics—Simulation and Computation*, vol. 48, no. 6, pp. 1746–1767, 2019.

[33] S. Kotz, S. Nadarajah, *Extreme value distributions: theory and applications*. World Scientific, 2000.

[34] S. Nadarajah, “The exponentiated gumbel distribution with climate application,” *Environmetrics: The official journal of the International Environmetrics Society*, vol. 17, no. 1, pp. 13–23, 2006.

[35] R. D. Gupta, A. K. Nanda, “Some results on reversed hazard rate ordering,” *Communications in Statistics-theory and Methods*, vol. 30, no. 11, pp. 2447–2457, 2001.

[36] Y. Lio, T.-R. Tsai, “Estimation of $\delta = p(x < y)$ for burr xii distribution based on the progressively first failure-censored samples,” *Journal of Applied Statistics*, vol. 39, no. 2, pp. 309–322, 2012.

[37] K. Chandler, “The distribution and frequency of record values,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 14, no. 2, pp. 220–228, 1952.

[38] V. B. Nevzorov, *Records: mathematical theory*. American Mathematical Soc., 2001.

[39] B. C. Arnold, N. Balakrishnan, H. N. Nagaraja, *Records*. John Wiley & Sons, 2011, vol. 768.

[40] M.-H. Chen, Q.-M. Shao, “Monte carlo estimation of bayesian credible and hpd intervals,” *Journal of Computational and Graphical Statistics*, vol. 8, no. 1, pp. 69–92, 1999.

[41] C. Robert, *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

[42] P. Congdon, *Bayesian statistical modelling*. John Wiley & Sons, 2007, vol. 704.

[43] J. F. Lawless, *Statistical models and methods for lifetime data*. John Wiley & Sons, 2011, vol. 362.

[44] S. Kotz, M. Pensky, *The stress-strength model and its generalizations: theory and applications*. World Scientific, 2003.

A
CODES

R Code for Figure 3.2

```
u=runif(10000,0,1)
mu=0.5
sigma=2
x1=mu-sigma*(log(-log(u)))
mu=1
sigma=2
x2=mu-sigma*(log(-log(u)))
mu=1.5
sigma=3
x3=mu-sigma*(log(-log(u)))
mu=3
sigma=4
x4=mu-sigma*(log(-log(u)))

par(mar=c(4,4,2,2))
plot(density(x1),type="l",col="red2",lwd=2,xlim=c(-5,20),ylim=
c(0,0.2),lty=1,main="PDF of Gumbel Distribution", xlab="x",
ylab="f(x)")
lines(density(x2),col="green",lwd=2,lty=2)
lines(density(x3),ylim=c(0,1),col="steelblue",lwd=2, lty=5)
lines(density(x4),ylim=c(0,1),col="orange",lwd=2,lty=6)

legend("topright",legend=c("Mu=0.5, Sigma=2","Mu=1, Sigma=2",
"Mu=1.5, Sigma=3","Mu=3, Sigma=4"),col=c("red2","green",
"steelblue","orange"),lty=c(1,2,5,6))
```

R Code for Figure 3.3

```
u=runif(10000,0,1)
mu=0.5
beta=2
x1=mu-beta*(log(-log(u)))
```

```
mu=1
beta=2
x2=mu-beta*(log(-log(u)))
mu=1.5
beta=3
x3=mu-beta*(log(-log(u)))
mu=3
beta=4
x4=mu-beta*(log(-log(u)))

par(mar=c(4,4,2,2))
plot(ecdf(x1),col="red2",lwd=2,lty=1,main="CDF of Gumbel
Distribution", xlab="x", ylab="F(x)")
lines(ecdf(x2),col="green",lwd=2,lty=2)
lines(ecdf(x3),col="steelblue",lwd=2,lty=5)
lines(ecdf(x4),col="orange",lwd=2,lty=7)

legend("right",legend=c("Mu=0.5, Sigma=2","Mu=1, Sigma=2",
"Mu=1.5, Sigma=3","Mu=3, Sigma=4"),col=c("red2","green",
"steelblue","orange"),lty=c(1,2,5,7))
```

MATLAB Code for Figure 4.1

```
    clc
clear all
%xx=[156 173 125 852 559 442 168 286 261 227 285 253 166 133 309
247 112 202 365 702];
xx=[230 169 178 271 129 568 115 280 305 326 1101 285 734 177 493
218 342 431 143 381];
x=sort(xx);
n=length(x);
lam0=1;
m=1;
while m>10^-6
    lam1=n/(sum(x)-(n/sum(exp(-lam0*x))*sum(x.*exp(-lam0*x))));
    m=abs(lam0-lam1);
    lam0=lam1;
end
lam=lam1
alpha=n/sum(exp(-lam*x))
CDF=[x;exp(-alpha*exp(-lam*x))]';
[h,p,k,c] = kstest(x,CDF,0.05,0)
F = cdfplot(x);
hold on
G = plot(x,exp(-alpha*exp(-lam*x)),'r-');
```

MATLAB Code for Figure 5.1

```
    clc
clear all
y=[156 125 112];
x=[230 169 129 115];
xn=115;
ym=112;
n=length(x);
m=length(y);
xb=mean(x);
yb=mean(y);
lambda=(n+m)/(n*(xb-xn)+m*(yb-ym))
alpha=n*exp(lambda*xn)
beta=m*exp(lambda*ym)
I11=n/(alpha^2);
I12=0;
I21=0;
I22=m/(beta^2);
I31=(n/(alpha*lambda))*(psi(0,n+1)-log(alpha));
I13=I31;
I32=(m/(beta*lambda))*(psi(0,m+1)-log(beta));
I23=I32;
I33=(m+n)/(lambda^2)+(n/(lambda^2))*(psi(1,n+1)+(psi(0,n+1)-
    log(alpha))^2)+(m/(lambda^2))*(psi(1,m+1)+(psi(0,m+1)-
    log(beta))^2);
U=[I11 I12 I13;I21 I22 I23;I31 I32 I33];
C=(1/(alpha+beta)^2)*[beta -alpha 0];
var=C*inv(U)*C'
R=alpha/(alpha+beta)
CI=[R-1.96*sqrt(var/n) R+1.96*sqrt(var/n)]
```

MATLAB Code for Figure 5.2

```
    clc
clear all
y=[156 125 112];
x=[230 169 129 115];
xn=115;
ym=112;
n=length(x);
m=length(y);
xb=mean(x);
yb=mean(y);
a1=0.0001;
a2=0.0001;
```

```
a3=0.0001;
b1=0.0001;
b2=0.0001;
b3=0.0001;
alpha=5;
beta=1;
sigma=0.008;
mu=0.020;
lambda=0.0001:0.0001:0.05;
k=length(lambda);
f=(10^15.4)*(lambda.^(m+n+a3-1)).*exp(-lambda.*(b3*+n*xb+m*yb)).
  *exp(-alpha*exp(-lambda.*xn).*exp(-beta*exp(-lambda.*ym)));
q=(1/sqrt(2*3.14*sigma^2)).*exp((-1/(2*sigma^2)).*(lambda-mu).^
  2);
plot(lambda,f)
hold on
plot (lambda,q)
```

B
OBTAINED DATA

In this thesis, two datasets have been used as real strength data. As stated in chapter 5 Real Data Analysis, these datasets can be reached on page 574 of the book "Statistical Models and Methods for Lifetime Data" by Jerald F. Lawless in 2011. So, to reproduce our findings, one can refer to the above-mentioned book. Therefore, the findings of our manuscript can be replicated using the information provided in our manuscript. To ensure that everyone can access those data, it has been provided as below (these data are included in the paper as well):

Figure 4.1 contains the proposal and posterior density functions of the scale parameter. The values are as below:

Posterior	**Normal Density**
3.84E-10	2.260985
2.36E-08	2.33221
2.59E-07	2.405302
1.40E-06	2.480298
5.15E-06	2.557233
1.48E-05	2.636141
3.60E-05	2.717061
7.72E-05	2.800026
0.000151	2.885074
0.000274	2.972241
0.000468	3.061563
0.00076	3.153077
0.001185	3.246819
0.001783	3.342825
0.002603	3.441132
0.0037	3.541778
0.005139	3.644797
0.006991	3.750227
0.009337	3.858103
0.012267	3.968463
0.015879	4.081341
0.020279	4.196774
0.025584	4.314798
0.031917	4.435449

0.039411	4.55876
0.048206	4.684767
0.058451	4.813506
0.070302	4.945009
0.083923	5.079311
0.099485	5.216446
0.117163	5.356446
0.137141	5.499344
0.159608	5.645172
0.184757	5.793961
0.212787	5.945744
0.2439	6.100549
0.278303	6.258407
0.316205	6.419346
0.357817	6.583395
0.403353	6.750582
0.453027	6.920933
0.507056	7.094475
0.565655	7.271231
0.629039	7.451227
0.697424	7.634486
0.771021	7.82103
0.850043	8.01088
0.934696	8.204057
1.025186	8.400579
1.121714	8.600465
1.224477	8.803731
1.333667	9.010394
1.44947	9.220466
1.572068	9.433963
1.701634	9.650895
1.838337	9.871272
1.982337	10.0951
2.133786	10.3224
2.292829	10.55316
2.459603	10.7874
2.634235	11.02511
2.816842	11.2663
3.007534	11.51097
3.206408	11.75912
3.413553	12.01073
3.629046	12.26582
3.852956	12.52436
4.085338	12.78636
4.326237	13.05179

4.575686	13.32066
4.833709	13.59294
5.100316	13.86862
5.375505	14.14768
5.659263	14.4301
5.951566	14.71586
6.252377	15.00493
6.561646	15.29729
6.879313	15.59291
7.205305	15.89176
7.539538	16.1938
7.881916	16.49901
8.232329	16.80735
8.590658	17.11877
8.956772	17.43324
9.330529	17.75071
9.711775	18.07114
10.10034	18.39448
10.49606	18.72068
10.89875	19.04969
11.3082	19.38145
11.72421	19.7159
12.14656	20.053
12.57504	20.39267
13.0094	20.73486
13.44941	21.07949
13.89481	21.42651
14.34535	21.77583
14.80075	22.12739
15.26075	22.48111
15.72505	22.83693
16.19339	23.19474
16.66545	23.55448
17.14095	23.91607
17.61958	24.27941
18.10102	24.64442
18.58498	25.01101
19.07112	25.37909
19.55913	25.74856
20.04868	26.11933
20.53946	26.4913
21.03112	26.86436
21.52335	27.23843
22.0158	27.61339
22.50815	27.98913

23.00006	28.36556
23.49121	28.74255
23.98125	29.12001
24.46986	29.49782
24.95671	29.87585
25.44147	30.25401
25.92381	30.63217
26.40342	31.0102
26.87996	31.388
27.35313	31.76543
27.8226	32.14239
28.28808	32.51873
28.74925	32.89434
29.20582	33.26909
29.65748	33.64285
30.10395	34.0155
30.54495	34.3869
30.98019	34.75692
31.40941	35.12544
31.83233	35.49232
32.2487	35.85742
32.65828	36.22063
33.0608	36.58179
33.45604	36.94079
33.84376	37.29747
34.22375	37.65172
34.59578	38.0034
34.95966	38.35237
35.31519	38.69849
35.66217	39.04164
36.00043	39.38168
36.32979	39.71847
36.65009	40.05188
36.96118	40.38178
37.26291	40.70804
37.55514	41.03052
37.83774	41.3491
38.1106	41.66364
38.37361	41.97401
38.62666	42.28009
38.86967	42.58174
39.10254	42.87885
39.3252	43.17129
39.53759	43.45893
39.73966	43.74164

39.93134	44.01933
40.11261	44.29185
40.28343	44.55909
40.44378	44.82095
40.59364	45.0773
40.73301	45.32803
40.86188	45.57304
40.98028	45.81221
41.0882	46.04544
41.18569	46.27263
41.27277	46.49367
41.34948	46.70847
41.41587	46.91694
41.472	47.11897
41.51791	47.31447
41.55368	47.50337
41.57939	47.68557
41.59511	47.86098
41.60093	48.02954
41.59694	48.19116
41.58324	48.34577
41.55993	48.4933
41.52713	48.63368
41.48493	48.76684
41.43346	48.89273
41.37285	49.01129
41.30322	49.12245
41.2247	49.22618
41.13743	49.32242
41.04155	49.41113
40.9372	49.49226
40.82452	49.56578
40.70367	49.63165
40.5748	49.68985
40.43807	49.74034
40.29362	49.7831
40.14164	49.81812
39.98226	49.84537
39.81567	49.86485
39.64203	49.87653
39.4615	49.88043
39.27426	49.87653
39.08047	49.86485
38.88032	49.84537
38.67396	49.81812

38.46159	49.7831
38.24336	49.74034
38.01946	49.68985
37.79006	49.63165
37.55534	49.56578
37.31548	49.49226
37.07064	49.41113
36.821	49.32242
36.56674	49.22618
36.30803	49.12245
36.04505	49.01129
35.77796	48.89273
35.50694	48.76684
35.23217	48.63368
34.9538	48.4933
34.67201	48.34577
34.38696	48.19116
34.09882	48.02954
33.80775	47.86098
33.51391	47.68557
33.21747	47.50337
32.91859	47.31447
32.61741	47.11897
32.31409	46.91694
32.00879	46.70847
31.70166	46.49367
31.39283	46.27263
31.08247	46.04544
30.77071	45.81221
30.4577	45.57304
30.14356	45.32803
29.82845	45.0773
29.51249	44.82095
29.19581	44.55909
28.87854	44.29185
28.56081	44.01933
28.24275	43.74164
27.92446	43.45893
27.60608	43.17129
27.28771	42.87885
26.96948	42.58174
26.65148	42.28009
26.33383	41.97401
26.01663	41.66364
25.69999	41.3491

25.38399	41.03052
25.06875	40.70804
24.75436	40.38178
24.4409	40.05188
24.12846	39.71847
23.81714	39.38168
23.507	39.04164
23.19815	38.69849
22.89064	38.35237
22.58456	38.0034
22.27999	37.65172
21.97699	37.29747
21.67562	36.94079
21.37596	36.58179
21.07807	36.22063
20.78201	35.85742
20.48783	35.49232
20.1956	35.12544
19.90536	34.75692
19.61716	34.3869
19.33106	34.0155
19.04711	33.64285
18.76534	33.26909
18.48579	32.89434
18.20852	32.51873
17.93354	32.14239
17.66091	31.76543
17.39066	31.388
17.12281	31.0102
16.8574	30.63217
16.59445	30.25401
16.33399	29.87585
16.07604	29.49782
15.82063	29.12001
15.56777	28.74255
15.3175	28.36556
15.06981	27.98913
14.82473	27.61339
14.58227	27.23843
14.34244	26.86436
14.10526	26.4913
13.87073	26.11933
13.63887	25.74856
13.40967	25.37909
13.18314	25.01101

12.95929	24.64442
12.73811	24.27941
12.51962	23.91607
12.30381	23.55448
12.09067	23.19474
11.88021	22.83693
11.67242	22.48111
11.4673	22.12739
11.26484	21.77583
11.06504	21.42651
10.86788	21.07949
10.67337	20.73486
10.48148	20.39267
10.29221	20.053
10.10555	19.7159
9.921485	19.38145
9.740005	19.04969
9.561094	18.72068
9.38474	18.39448
9.210926	18.07114
9.039638	17.75071
8.870859	17.43324
8.704574	17.11877
8.540763	16.80735
8.379411	16.49901
8.220498	16.1938
8.064006	15.89176
7.909916	15.59291
7.758208	15.29729
7.608861	15.00493
7.461856	14.71586
7.317172	14.4301
7.174787	14.14768
7.034681	13.86862
6.896831	13.59294
6.761215	13.32066
6.627812	13.05179
6.496598	12.78636
6.367552	12.52436
6.240649	12.26582
6.115868	12.01073
5.993185	11.75912
5.872576	11.51097
5.754018	11.2663
5.637488	11.02511

5.522962	10.7874
5.410416	10.55316
5.299826	10.3224
5.191169	10.0951
5.08442	9.871272
4.979556	9.650895
4.876553	9.433963
4.775387	9.220466
4.676033	9.010394
4.578469	8.803731
4.48267	8.600465
4.388613	8.400579
4.296274	8.204057
4.205629	8.01088
4.116655	7.82103
4.029327	7.634486
3.943624	7.451227
3.859522	7.271231
3.776997	7.094475
3.696026	6.920933
3.616587	6.750582
3.538657	6.583395
3.462213	6.419346
3.387232	6.258407
3.313693	6.100549
3.241574	5.945744
3.170852	5.793961
3.101505	5.645172
3.033512	5.499344
2.966852	5.356446
2.901502	5.216446
2.837443	5.079311
2.774653	4.945009
2.713111	4.813506
2.652797	4.684767
2.593691	4.55876
2.535772	4.435449
2.479021	4.314798
2.423418	4.196774
2.368944	4.081341
2.315579	3.968463
2.263304	3.858103
2.2121	3.750227
2.161949	3.644797
2.112833	3.541778

2.064733	3.441132
2.017631	3.342825
1.971509	3.246819
1.926351	3.153077
1.882138	3.061563
1.838854	2.972241
1.796481	2.885074
1.755004	2.800026
1.714406	2.717061
1.67467	2.636141
1.635781	2.557233
1.597723	2.480298
1.56048	2.405302
1.524037	2.33221
1.488379	2.260985
1.453492	2.191593
1.41936	2.123999
1.385969	2.058168
1.353305	1.994066
1.321353	1.931658
1.290101	1.870911
1.259535	1.811792
1.22964	1.754266
1.200405	1.698302
1.171815	1.643866
1.143859	1.590926
1.116524	1.539451
1.089797	1.489408
1.063666	1.440767
1.038119	1.393497
1.013146	1.347567
0.988733	1.302947
0.96487	1.259608
0.941545	1.21752
0.918748	1.176655
0.896467	1.136984
0.874693	1.098478
0.853414	1.061111
0.832621	1.024854
0.812303	0.989682
0.792451	0.955568
0.773054	0.922485
0.754104	0.890409
0.73559	0.859314
0.717504	0.829175

0.699837	0.799968
0.68258	0.771669
0.665723	0.744255
0.649259	0.717703
0.63318	0.69199
0.617476	0.667094
0.60214	0.642993
0.587163	0.619666
0.572539	0.597092
0.55826	0.575251
0.544317	0.554122
0.530704	0.533685
0.517413	0.513922
0.504437	0.494813
0.491771	0.476341
0.479405	0.458487
0.467335	0.441232
0.455553	0.424561
0.444053	0.408456
0.432829	0.3929
0.421874	0.377878
0.411183	0.363373
0.40075	0.349371
0.390568	0.335855
0.380633	0.322812
0.370938	0.310228
0.361479	0.298087
0.352249	0.286376
0.343244	0.275083
0.334459	0.264193
0.325888	0.253695
0.317527	0.243576
0.30937	0.233825
0.301414	0.224428
0.293653	0.215376
0.286083	0.206656
0.278699	0.198258
0.271498	0.190172
0.264475	0.182388
0.257625	0.174894
0.250946	0.167682
0.244432	0.160743
0.23808	0.154066
0.231887	0.147644
0.225848	0.141468

0.219959	0.135528
0.214218	0.129818
0.208621	0.124329
0.203164	0.119053
0.197844	0.113984
0.192657	0.109113
0.187602	0.104434
0.182673	0.09994
0.17787	0.095625
0.173187	0.091481
0.168623	0.087504
0.164175	0.083686
0.15984	0.080022
0.155615	0.076507
0.151497	0.073135
0.147485	0.0699
0.143574	0.066799
0.139764	0.063824
0.136051	0.060973
0.132433	0.05824
0.128908	0.055621
0.125473	0.053111
0.122127	0.050707
0.118866	0.048403
0.11569	0.046198
0.112596	0.044086

Figure 5.1 contains the empirical CDF points. The values are as below:

Dataset 1 (x)	**Empirical CDF**
115	0.0719212279684049
129	0.0923341012962292
143	0.115762543390081
169	0.166683622351636
177	0.184075607500274
178	0.186300270395955
218	0.282591210958021
230	0.313419347009354
271	0.420482095983728
280	0.443723918424522
285	0.456520943674426
305	0.506617387685895
326	0.556817405913310
342	0.593068719072731
381	0.673196824095878
431	0.757948464670944
493	0.836781793160388
568	0.900830620381315
734	0.968495854677683
1101	0.997659038479823

Figure 5.2 contains the empirical CDF points. The values are as below:

Dataset 2 (y)	**Empirical CDF**
112	0.0789969342415915
125	0.102939982805443
133	0.119479770907944
156	0.174047431730863
166	0.200610198121388
168	0.206098688035014
173	0.220052491651908
202	0.306024196738187
227	0.383634788334955
247	0.445424477559749
253	0.463641921846360
261	0.487595187256323
285	0.556489633395641
286	0.559248229225808
309	0.619859790763133
365	0.742611428530314
442	0.856432055632269
559	0.944110243253334
702	0.983022936371675

852	0.995207195497212

PUBLICATIONS FROM THE THESIS

Papers

1. Fayyazishishavan E, Kılıç Depren S. (2021). Inference of stress-strength reliability for two parameter of exponentiated Gumbel distribution based on lower record values. *PLoS ONE*, 16(4): e0249028. https://doi.org/10.1371/journal.pone.0249028.

Conference Papers

1. Fayyazi, E., Depren, K. S. (2019). Generalization and Compounding Statistical Distributions, Goals and Benefits: A Review. *Proceedings of the* 11^{th} *International Statistics Congress*, Bodrum, Turkey.