6-2013

Statistical Inference Based on Upper Record Values

Daniel J. Luckett

College of William and Mary

Follow this and additional works at: http://publish.wm.edu/honortheses

Recommended Citation
http://publish.wm.edu/honortheses/577

This Honors Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M Publish. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M Publish. For more information, please contact wmpublish@wm.edu.
Statistical Inference Based on Upper Record Values

A thesis submitted for partial fulfillment of the requirements
for the degree of Bachelor of Science in Mathematics from
The College of William and Mary

by
Daniel J. Luckett

Accepted for Honors

Tanujit Dey
Ross J. Iaci
M. Drew LaMar
Lawrence M. Leemis

Williamsburg, VA
24 April, 2013
## Contents

1 Introduction ........................................... 1
   1.1 Overview and Motivation ............................. 1
   1.2 Applications of Records ........................... 2
   1.3 Definition of Upper and Lower Records .......... 3
   1.4 Distributions of Record Times and Record Values ... 4
      1.4.1 Record Times .................................. 4
      1.4.2 Record Values ................................ 5

2 An Efficient Algorithm for Generating Record Values .... 7
   2.1 Introduction ....................................... 7
   2.2 Algorithm for Upper Records ....................... 8
   2.3 Algorithm for Lower Records ....................... 9
   2.4 Illustration and Analysis ........................... 10
   2.5 Conclusion ......................................... 11

3 Statistical Inference for the Generalized Inverted Exponen-
tial Distribution Based on Upper Record Values .......... 12
   3.1 Introduction ....................................... 12
   3.2 Non-Bayesian Estimation ............................ 14
      3.2.1 Maximum Likelihood Estimation ................. 14
      3.2.2 Approximate Confidence Intervals ............... 15
      3.2.3 Bootstrap Confidence Intervals ................. 16
   3.3 Bayesian Estimation ................................ 17
      3.3.1 Bayes Estimators Under the General Entropy Loss Function 18
      3.3.2 Two-Sided Bayes Probability Intervals .......... 19
      3.3.3 Empirical Bayes Estimation ..................... 20
   3.4 Prediction of Future Record Values ................. 21
      3.4.1 Non-Bayesian Prediction ....................... 21
      3.4.2 Bayes Prediction ................................ 22
      3.4.3 Conditional Median Prediction ................... 24
   3.5 Numerical Results .................................. 25
      3.5.1 Simulation Study ............................... 25
      3.5.2 Data Analysis ................................ 27
   3.6 Conclusion ......................................... 29

4 Statistical Inference for the Generalized Rayleigh Distribu-
tion Based on Upper Record Values ........................ 30
   4.1 Introduction ....................................... 30
   4.2 Non-Bayesian Estimation ............................ 32
      4.2.1 Maximum Likelihood Estimation ................. 33
4.2.2 Approximate Confidence Intervals ........................................ 33
4.2.3 Bootstrap Confidence Intervals ........................................... 34
4.3 Bayes Estimation ................................................................. 36
  4.3.1 Bayes Estimation Under the General Entropy Loss Function ... 37
  4.3.2 Bayes Estimation Under the Squared-Error Loss Function .... 37
  4.3.3 Two-Sided Bayes Probability Intervals ............................ 38
  4.3.4 Empirical Bayes Estimation ............................................. 39
4.4 Prediction of Future Records .................................................. 40
  4.4.1 Non-Bayesian Prediction ............................................... 40
  4.4.2 Bayesian Prediction ...................................................... 41
  4.4.3 Conditional Median Prediction ........................................ 42
4.5 Numerical Results .............................................................. 43
  4.5.1 Simulation Study ......................................................... 43
  4.5.2 Data Analysis ............................................................ 45
4.6 Conclusion ........................................................................... 47
Acknowledgments

I would like to thank my thesis advisor, Dr. Tanujit Dey, as well as the thesis committee members, Dr. Lawrence M. Leemis, Dr. Ross J. Iaci, and Dr. M. Drew LaMar. I would also like to thank Dr. Sanku Dey for his valuable contributions to this research. Finally, I would like to thank the many faculty members at the William & Mary Mathematics Department, too numerous to name.
Abstract

The goal of this thesis is to examine methods of statistical inference based on upper record values. This includes estimation of parameters based on samples of record values and prediction of future record values. We first define and discuss record times and record values and their distributions. Then we propose an efficient algorithm to generate random samples of record values. The algorithm, based on the conditional survivor function, has a time complexity that is linear with respect to the sample size. It is quite efficient and can be useful in simulation. Next, we discuss inference problems related to two distributions, the generalized inverted exponential distribution and the generalized Rayleigh distribution. A number of techniques are considered, including both frequentist and Bayesian techniques. The following are considered for each of these two distributions. We first consider maximum likelihood and Bayesian estimation of the unknown parameters based on upper record values. We also consider empirical Bayes estimation. Then, we derive approximate confidence intervals based on a normal approximation, bootstrap confidence intervals, and two-sided Bayes probability intervals based on record values. We derive both maximum likelihood and Bayesian methods for predicting future record values. Numerical results include both simulation and data analysis. We conduct a simulation study for each distribution. Simulation is used to compare the performance of maximum likelihood estimators, Bayesian estimators, and empirical Bayes estimators. We consider the average bias and mean squared error for these estimators. We also consider the average length and coverage probability for approximate confidence intervals. As many of the results in this paper cannot be found in closed form, we use numerical methods to find the maximum likelihood estimates and we implement the Metropolis–Hastings algorithm to compute the Bayes estimates. We use each distribution to model one real data set by estimating the parameters, computing approximate confidence intervals and two-sided Bayes probability intervals, and evaluating the fit. Finally, we illustrate Bayesian prediction methods on a simulated sample for each distribution.
Chapter 1

Introduction

1.1 Overview and Motivation

Record values and record times have been of interest to humans throughout history. Meteorologists frequently deal with upper and lower record temperatures and precipitation levels. A seismologist may be interested in earthquakes of record magnitude. Record values appear often in sporting events. For example, an analyst may be concerned with record performances in the Olympic one hundred meter dash. There is something about record breaking performances that makes them fascinating to humans. Thus, it is natural to try to quantify the study of records in a statistical sense.

The study of statistical records first emerged in a landmark paper by Chandler (1952). Numerous papers on records were published throughout the 1950s. A number of statisticians have worked on interesting problems pertaining to the study of records. Wilks (1959) posed the question of how many additional observations are needed to observe a value exceeding the maximum of $n$ observations. Rényi (1962) showed that record times are independent. The notion of $k$th record times was introduced by Dziubdziela and Kopocinsky (1976). Weak records were introduced by Vervaat (1973). Arnold et al. (1998) give a likelihood function for estimating unknown parameters based on record samples. Basak and Balakrishnan (2003) give a predictive likelihood function for future record values.


The goal of this paper is to study statistical inference based on upper record values for the generalized inverted exponential distribution (GIED) and the generalized
Rayleigh (GR) distribution. Frequentist and Bayesian methods are used to estimate unknown parameters and predict future record values. After introducing theoretical results pertaining to inference based on records, we assess the performance of different estimators and predictors using simulation. Multiple real data sets are used for illustrative purposes. As simulation is such an important part of this research, we address the problem of generating samples of record values by proposing an efficient algorithm to generate record samples.

There are many situations where a data set may consist of only record values. As an example, consider a sporting event where only record-breaking performances are recorded. The analyst needs to fit a model to the data when the record values are known but the whole sample is not. Such situations require an ability to accurately estimate parameters of a distribution based only on a record sample. There are many situations where the prediction of future records is essential. A meteorologist may want to know how much flooding will occur the next time the current rainfall record is broken. The statistician must estimate the next record value of rainfall from a data set consisting of past record values. The goal of this paper is to evaluate methods for inference and prediction problems like these.

This thesis is organized as follows. The rest of Chapter 1 defines the mathematical properties of record times and record values and their distributions. Chapter 2 proposes an algorithm to quickly and efficiently generate samples of record values. Chapters 3 and 4 discuss statistical inference based on upper records for the generalized inverted exponential distribution and the generalized Rayleigh distribution, respectively. The results presented are compared using simulation and data analysis.

1.2 Applications of Records

The mathematical theory of records has extensive applications to real world situations. We will consider two of them here as motivating examples. First, consider an example from reliability theory. Suppose we have a system of \( n \) identical components such that the system functions when \( k \) or more out of the \( n \) components are functioning. Thus, the system will fail when the \( (n-k+1) \)th component fails. Note that this includes the cases of series and parallel systems when \( k = n \) and \( k = 1 \), respectively. There are many examples of systems that fit this description. Suppose we are interested in predicting the lifetime of the system when \( r \) components have failed, for some \( 1 \leq r < k \leq n \). The lifetime of the system corresponds to the \( (n-k-r+1) \)th future record value.
As a second example, consider a Type-II censoring scheme. Suppose there are $n$ items on test and the test will conclude after $k$ failures, for some $k < n$. It may be of interest to predict how long the test will last. To do this, one must predict the $k$th future record value. Such a situation would occur, for example, in an accelerated life test to test the lifetime of mechanical components, when the failure time for each component is recorded. Alternatively, the same situation would occur in a clinical trial when the time at which a patient is cured is recorded.

### 1.3 Definition of Upper and Lower Records

Suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed (iid) random variables from a continuous distribution. We say that $X_i$ is an upper record value and $i$ is an upper record time if $X_i > X_j$ for all $j < i$. In other words, an upper record value is a value that is greater than all observed values so far. The record times are the indices at which record values occur. We note that, by this definition, an observation must be strictly greater than and not equal to the previous upper record in order to be considered a record value.

To formalize this notion, let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a continuous distribution. Define the first record time, $U(1) = 1$ and the $m$th record time, for $m > 1$ by the recursive formula

$$U(m) = \min\{j > U(m-1) : X_j > X_{U(m-1)}\}.$$  

The $m$th record value, then, is $X_{U(m)}$.

Next we will define the notion of $k$th records. Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a continuous cumulative distribution function. Define the first $k$th record time as $U(1, k) = k$ and the $m$th $k$th record time, for $m > 1$ by the recursive formula

$$U(m, k) = \min\{j > U(m-1, k) : X_j > X_{U(j-k,k-1)}\}.$$  

For details on $k$th records, see Nevzorov (2001).

Finally, we say that $X_i$ is a weak record if $X_i \geq X_j$ for all $j < i$. Weak records generalize the notion of records by including as an upper record any observation that is equal to the most recent upper record. This corresponds to the situation where a record is “tied” but not “broken.” Analogous definitions for all of these exist for lower records as well.

An important question that arises is that of how many records one should expect to observe in a sample of size $n$. Let $M_n$ be the number of record observations in a
random sample of size $n$. It is shown in Ahsanullah (1995), that the expected value of $M_n$ is

$$E[M_n] = \sum_{i=1}^{n} \frac{1}{i}.$$  

Thus, in a sample of size $n$, one can expect to see only approximately $\log n$ record values (Doostparast (2009)).

Another interesting question is that of how long one would expect to wait before observing a new upper record. Wilks (1959) addresses this question. For a random sample of size $n$, let $N(n)$ be the minimum number of additional observations needed in order to observe a value greater than $M_n = \max\{X_1, X_2, \ldots, X_n\}$. Let $N(1)$ be the minimum number of additional observations needed in order to observe a value less than $m_n = \min\{X_1, X_2, \ldots, X_n\}$. Define $N(1, n)$ as the minimum number of additional observations needed to observe a value that falls outside of the interval $[m_n, M_n]$. It is shown in Nevzorov (2001) that while $E[N(1, n)] = n$, $E[N(1)] = E[N(n)] = \infty$. Thus, the expected time to observe a new record value is infinite. These two results show how difficult it can be to generate simulated samples of record values.

### 1.4 Distributions of Record Times and Record Values

We will now derive the distributions of record times and record values. We will concern ourselves only with upper records. Results for lower records can be derived analogously.

#### 1.4.1 Record Times

Let $U(1) = 1$ be the first record time. Nevzorov (2001) shows that the survivor function of the second record time is

$$S_2(j) = P(U(2) > j) = \frac{1}{j}, \quad j = 1, 2, \ldots.$$  

Thus, the probability mass function of $U(2)$ is

$$f_2(j) = P(U(2) > j - 1) - P(U(2) > j) = \frac{1}{j(j - 1)}, \quad j = 2, 3, \ldots.$$  

We know from Rényi (1962) that record times are independent. Thus, the joint probability mass function of the first $n$ record times is

$$f(j_2, j_3, \ldots, j_n) = P(U(1) = 1, U(2) = j_2, \ldots, U(n) = j_n) = \frac{1}{(j_2 - 1)(j_3 - 1) \cdots (j_n - 1)(j_n)}.$$
for \(1 = j_1 < j_2 < \ldots < j_n\). The marginal probability mass function of the \(m\)th record time is

\[
f_m(j) = \frac{1}{(k_2 - 1)(k_3 - 1) \cdots (j - 1)(j)}.\]

It can be shown that record times \(U(1), U(2), \ldots\) form a Markov chain.

Along with the record times \(U(i)\), we can define the inter-record times \(\Delta(1) = 1\) and \(\Delta(i) = U(i) - U(i - 1)\) for \(i > 1\). Inter-record times correspond roughly to the number of non-record observations between record values. Nevzorov (2001) shows that the inter-record times are conditionally independent given the record values, and the \(i\)th inter-record time has probability mass function

\[
f_i(m) = P(\Delta(i) = m \mid X_{U(1)}, X_{U(2)}, \ldots) = (1 - F(X_{U(i-1)}))(F(X_{U(i-1)}))^{m-1},\]

for \(m = 1, 2, \ldots\), and \(i = 2, 3, \ldots\). Thus, the \(i\)th inter-record time follows a geometric distribution.

### 1.4.2 Record Values

Let \(X_1, X_2, \ldots\) be a sequence of iid random variables with continuous cumulative distribution function \(F\) and probability density function \(f\). The survivor function of the \(m\)th record value, \(X_{U(m)}\), given the past record values \(X_{U(1)}, X_{U(2)}, \ldots, X_{U(m-1)}\), is

\[
F_m(x) = P(X_{U(m)} > x \mid X_{U(1)} = x_2, X_{U(2)} = x_2, \ldots, X_{U(m-1)} = x_{m-1}) = \frac{1 - F(x)}{1 - F(x_{m-1})}.
\]

The conditional density function is

\[
f_m(x \mid x_1, x_2, \ldots, x_{m-1}) = \frac{f(x)}{1 - F(x_{m-1})}.
\]

It follows that record values form a Markov chain as well. The joint probability density function of the record values \(X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}\) is

\[
f_n(x_1, x_2, \ldots, x_n) = \frac{f(x_1)f(x_2)\cdots f(x_n)}{(1 - F(x_1))(1 - F(x_2))\cdots(1 - F(x_{n-1}))}.
\]

This joint density function is important for maximum likelihood estimation based on record values.

Nevzorov (2001) notes one more important result dealing with extending the probability integral transformation to record values. Let \(X_1, X_2, \ldots, X_n\) be the first \(n\) upper record values from a continuous cumulative distribution function, \(F\). Let \(U_1, U_2, \ldots, U_n\) be the first \(n\) upper record values from a uniform distribution on the interval \([0, 1]\). Then, the random vector \((F(X_1), F(X_2), \ldots, F(X_n))\) has the same distribution as \((U_1, U_2, \ldots, U_n)\).
The mathematical theory of records is a growing area of study with great potential for research. The results we have presented comprise only a small number of the developments that have been made.
Chapter 2

An Efficient Algorithm for Generating Record Values

2.1 Introduction

The study of record values is popular in many areas because of its wide applicability in the modeling and analysis of lifetime data. Motivated by the study of extreme weather conditions, Chandler (1952) introduced the study of record values and documented many of the basic properties of records. There is a volume of statistical literature on the subject of record statistics, and the underlying theory for the likelihood of a record-breaking event taking place in a stable system is remarkably simple (see Benestad (2003)). Properties of record values have been extensively studied in the literature by Ahsanullah (1988, 1995), Arnold and Balakrishnan (1989), Arnold et al. (1992, 1998), Nevzorov (2001), Kamps (1995) and Jaheen (2004).

Simulation studies are an integral part of statistical research. Simulation is used advantageously in a number of situations. This includes providing the empirical estimation of sampling distributions, studying the misspecification of assumptions in statistical procedures, determining the power in hypothesis tests, and many others. However, in order to design simulation studies relating to record values, one must be able to efficiently generate record samples from a wide variety of distributions. Doostparast (2009) claims that from a sequence of \( n \) independent and identically distributed random variables, only about \( \log n \) variates will be record values. Algorithms that generate records by generating \( n \) random variates and accepting only those which are record values are very inefficient. To the best of our knowledge, there is no algorithm at hand that can efficiently generate record values from several distributions. Given this condition, generating moderately large samples of record values or a large number of samples of record values will take a significant amount of time. A vast amount of research has been conducted where simulation studies have
been carried out only based on small samples. Simulation studies come with their own set of limitations. In the framework of multilevel studies an important problem is generating an adequate sample size that provides unbiased and accurate estimates.

In this chapter we present an efficient algorithm for generating records that is based on the conditional survivor function and will work for any distribution with a closed form inverse survivor function. We proceed as follows. Section 2.2 discusses the algorithm for upper record values and section 2.3 discusses the analogous case for lower record values. Section 2.4 contains an example for implementing our algorithm. We conclude with final remarks in section 2.5.

### 2.2 Algorithm for Upper Records

Let $X$ be a random variable with population probability density function $f(x)$ on positive support and cumulative distribution function $F(x)$. We wish to generate a random sample of $n$ upper records from the distribution of $X$. That is, we need a sample $x_{U(1)}, x_{U(2)}, \ldots, x_{U(n)}$ where $x_{U(i)}$ is drawn from the distribution of $X$ and $x_{U(i)} < x_{U(j)}$ for all $i < j$. This is challenging because it may take many repeated sample points before $n$ upper records are generated, even for relatively small $n$.

We set the condition that each record value $x_{U(i)}$ is greater than $x_{U(i-1)}$. Thus, the first sample point generated will be the first upper record, $x_{U(1)}$. The second upper record is obtained by sampling from the conditional distribution $F_{X|X>x_{U(1)}}(x)$. Iterating this process will yield a random sample of upper records, where each record is drawn from the desired distribution and each record is greater than all those before it. Assuming that random variates can be generated from the distribution of $X$ and that the conditional distribution of $X$ is easily obtained, each iteration can be performed in constant time. Thus, the entire algorithm has a time complexity that is linear with respect to the desired sample size $n$. Next, we show how this algorithm can be easily implemented for a variety of distributions.

Assuming that the distribution of $X$ has survivor function $S(x)$ and inverse survivor function $S^{-1}(x)$ that can be obtained in closed form, it is well known that a simple variate generation algorithm for $X$ is

\[
\begin{align*}
generate & \quad U \sim U(0,1) \\
return & \quad X = S^{-1}(U)
\end{align*}
\]

where $U(0,1)$ denotes the uniform distribution on the interval (0,1). It is also well
known that the conditional survivor function of $X$ given that $X > t$ is
\[ S_{X|X>t}(x) = \frac{S(x)}{S(t)} \quad x > t. \]

The inverse of the conditional survivor function is found as
\[ S_{X|X>t}^{-1}(x) = S(t)S^{-1}(x) \quad x > t. \]

So, an efficient algorithm for generating a random sample of upper records can be obtained as in Algorithm 1.

**Algorithm 1** To generate upper records

1: generate $U \sim U(0, 1)$
2: calculate $X_1 = S^{-1}(U)$
3: for $i = 2$ to $n$ do
4: generate $U \sim U(0, 1)$
5: calculate $X_i = S(X_{i-1})S^{-1}(U)$
6: end for
7: return $X_1, X_2, \ldots, X_n$
8: the sample $X_1, X_2, \ldots, X_n$ will represent $n$ upper record values $(x_{U(1)}, x_{U(2)}, \ldots, x_{U(n)})$

### 2.3 Algorithm for Lower Records

Here we propose an analogous algorithm for generating samples of lower records. Let $X$ be a random variable with probability density function $f(x)$ on positive support and cumulative distribution function $F(x)$. We wish to generate a random sample of lower records from the distribution of $X$. That is, we need a sample $x_{L(1)}, x_{L(2)}, \ldots, x_{L(n)}$ where $x_{L(i)}$ is drawn from the distribution of $X$ and $x_{L(i)} > x_{L(j)}$ for all $i < j$.

An analogous algorithm can be defined for generating samples of lower records. Assume that the distribution of $X$ has survivor function $S(x)$, inverse survivor function $S^{-1}(x)$, and cumulative distribution function $F(x)$ that can be obtained in closed form. The conditional survivor function given that $X$ is less than $t$ is given by
\[ S_{X|X<t}(x) = \frac{S(x)}{F(t)} \quad x < t. \]

The inverse of the conditional survivor function is found as
\[ S_{X|X<t}^{-1}(x) = F(t)S^{-1}(x) \quad x < t. \]

So, an algorithm for generating a random sample of lower records can be obtained as in Algorithm 2.
Algorithm 2 To generate lower records
1: generate $U \sim U(0, 1)$
2: calculate $X_1 = S^{-1}(U)$
3: for $i = 2$ to $n$ do
4: generate $U \sim U(0, 1)$
5: calculate $X_i = F(X_{i-1})S^{-1}(U)$
6: end for
7: return $X_1, X_2, \ldots, X_n$
8: the sample $X_1, X_2, \ldots, X_n$ will represent $n$ lower record values ($x_{L(1)}, x_{L(2)}, \ldots, x_{L(n)}$)

2.4 Illustration and Analysis

As an example, we can easily implement the above algorithms for the one parameter exponential distribution as follows. Let $X$ be an exponentially distributed random variable with parameter $\lambda$ with survivor function

$$S(x) = e^{-\lambda x} \quad x \geq 0.$$  

Then,

$$S_{X|X > t}(x) = \frac{e^{-\lambda x}}{e^{-\lambda t}} = e^{-\lambda(x-t)} \quad x \geq t.$$  

So the corresponding inverse survivor function is

$$S^{-1}(u) = -\log(u)/\lambda,$$

where $u$ is a sample from the $U(0, 1)$ distribution. So, the inverse conditional survivor function is

$$S_{X|X > t}^{-1}(u) = S(t)S^{-1}(u) = -\frac{e^{-\lambda t}\log(u)}{\lambda}.$$  

We can now use Algorithm 1 to generate a sample of upper records from the exponential distribution. The algorithm is reasonably efficient. We have implemented and executed our algorithm in R software to generate upper records from the exponential(1) distribution. The running times (in seconds) for generating a sample of upper records of size $n$ from an exponential(1) is shown in Table 2.1. A plot of execution times against the sample size appears in Figure 2.1.

<table>
<thead>
<tr>
<th>n</th>
<th>1000</th>
<th>2000</th>
<th>3000</th>
<th>4000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.0115</td>
<td>0.0266</td>
<td>0.0435</td>
<td>0.0641</td>
<td>0.0857</td>
</tr>
</tbody>
</table>

Table 2.1: Execution times (in seconds) for generating exponential(1) upper record samples

Table 2.1 and Figure 2.1 both provide visual evidence that the proposed algorithm achieves a linear time complexity. Running times featured in Table 2.1 and Figure 2.1 are average running times over 100 executions of the algorithm for each sample size.
The execution time grows no faster than linearly with respect to the sample size. Any algorithm that produces random data must go through a certain degree of exertion to generate each value. In other words, the number of steps required to produce $n$ values must be a linear function of $n$. Thus, any algorithm that produces random data must feature an execution time that grows linearly with respect to the sample size, and one would not expect any algorithm to achieve a better time complexity than the algorithm presented here.

### 2.5 Conclusion

Statistical research involving record values is vital as it has countless applications in several areas. Simulation studies are an imperative part of this research. However, an efficient variate generation algorithm is needed to perform any kind of simulation study. In the past, record samples have been generated by constructing many sample points and accepting only those random variates which are records. This technique is vastly wasteful and inept. In this chapter we propose a faster algorithm for generating records. It is simple and easy to implement under any circumstances. The algorithm presented here generates record samples for large sample sizes in milliseconds. Even for sample sizes as large as $n = 5000$, the algorithm executes in under 0.1 seconds. This algorithm should greatly improve computational efficiency for simulation studies related to record data.
Chapter 3

Statistical Inference for the Generalized Inverted Exponential Distribution Based on Upper Record Values

3.1 Introduction

The study of record values and associated statistics are of great significance in many real life situations such as meteorology, seismology, athletic events, economics, and life testing. The frequency of weather conditions inspired Chandler (1952) to study the distributions of lower records, record times, and inter-record times for independent and identically distributed (iid) sequences of random variables. Since then, numerous papers on record values and their distributional properties have appeared in the statistical literature. Among them are Ahsanullah (1988, 1995), Arnold and Balakrishnan (1989), Arnold et al. (1992, 1998), and Kamps (1995).

Let $X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}$ be the first $n$ upper record values from the two-parameter generalized inverted exponential distribution (GIED) with probability density function

$$f(x; \alpha, \lambda) = \frac{\alpha \lambda}{x^2} e^{-\lambda/x} (1 - e^{-\lambda/x})^{\alpha - 1} \quad x \geq 0, \alpha, \lambda > 0$$

and cumulative distribution function

$$F(x; \alpha, \lambda) = 1 - (1 - e^{-\lambda/x})^\alpha \quad x \geq 0.$$  

(3.2)

The hazard function, then, is

$$h(x; \alpha, \lambda) = \frac{\alpha \lambda}{x^2(e^{\lambda/x} - 1)} \quad x \geq 0.$$ 

(3.3)

Here $\alpha, \lambda > 0$ are the shape and scale parameters, respectively. The generalized inverted exponential distribution with shape parameter $\alpha$ and scale parameter $\lambda$ will be
denoted by GIED(\(\alpha, \lambda\)). The generalized inverted exponential distribution (GIED) was introduced in the literature by Abouammoh and Alshingiti (2009) as a generalization of the inverted exponential distribution. Abouammoh and Alshingiti (2009) carried out extensive studies on the properties of the GIED. Due to its practicality, the GIED can be used for many applications, including accelerated life testing, horse racing, supermarket queues, sea currents, wind speeds, and others (see Nadarajah and Kotz (2000)). The hazard function of the GIED can never be constant. The GIED has a unimodal and right skewed density function. Abouammoh and Alshingiti (2009) observed that in many situations, the GIED provides a better fit than the gamma, Weibull, generalized exponential, and inverted exponential distributions. Plotted below on the left are probability density functions for the GIED with \(\alpha = 1, \lambda = 1\) plotted as the solid line, \(\alpha = 2, \lambda = 1\) plotted as the dashed line, and \(\alpha = .25, \lambda = 1\) plotted as the dotted line. Plotted on the right are various hazard functions for the GIED using the same parameter settings.

![GIED density functions](image1)

![GIED hazard functions](image2)

Figure 3.1: GIED density functions    Figure 3.2: GIED hazard functions

Recently, Krishna and Kumar (2012) studied the reliability estimation of this distribution under progressive type-II censoring.

The rest of this chapter is organized as follows. In Section 3.2, we derive the maximum likelihood estimators for the two unknown parameters of the GIED. In Section 3.3 we derive the Bayes estimators based on the general entropy loss function. Section 3.4 presents the predictive distribution of the future record values based on a given set of the first \(n\) upper records. Numerical comparison results and data analysis are presented in Section 3.5. We conclude in Section 3.6.
3.2 Non-Bayesian Estimation

In this section we discuss the process of obtaining the maximum likelihood estimators of the parameters $\alpha$ and $\lambda$ based on upper record values. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with cumulative distribution function $F(x)$ and probability density function $f(x)$ on positive support. Let $Y_n = \max \{X_1, X_2, \ldots, X_n\}$ for $n \geq 1$. The observation $X_j$ is an upper record value of $\{X_i\}_{i \geq 1}$ if it is greater than all preceding observations, that is, if $Y_j > Y_{j-1}$ for $j > 1$.

Suppose we observe $n$ upper record values $X = \{X_U(1), X_U(2), \ldots, X_U(n)\}$ from a sequence of iid random variables following a GIED($\alpha, \lambda$) with probability density function in (3.1). Arnold et al. (1998) give the likelihood function as

$$L(\alpha, \lambda | x) = f(x_U(n); \alpha, \lambda) \prod_{i=1}^{n-1} \frac{f(x_U(i); \alpha, \lambda)}{1 - F(x_U(i); \alpha, \lambda)} 0 \leq x_U(1) < x_U(2) < \ldots < x_U(n) < \infty. \quad (3.4)$$

Substituting (3.1) and (3.2) in (3.4), we get the likelihood function

$$L(\alpha, \lambda | x) = \alpha^n \lambda^n e^{\alpha \ln \left(1 - e^{-\lambda/x_U(n)}\right)} \prod_{i=1}^{n} \frac{1}{x_U^2(i)} \cdot \frac{e^{-\lambda/x_U(i)}}{1 - e^{-\lambda/x_U(i)}}. \quad (3.5)$$

The maximum likelihood estimates are the values of $\alpha$ and $\lambda$ that maximize this likelihood function.

3.2.1 Maximum Likelihood Estimation

The log likelihood function $\ell(\alpha, \lambda | x) = \log L(\alpha, \lambda | x)$, dropping terms that do not involve $\alpha$ and $\lambda$, is

$$\ell(\alpha, \lambda | x) = n (\ln \alpha + \ln \lambda) - \sum_{i=1}^{n} \frac{\lambda}{x_U(i)} + \alpha \ln \left(1 - e^{-\lambda/x_U(n)}\right) - \sum_{i=1}^{n} \ln \left(1 - e^{-\lambda/x_U(i)}\right). \quad (3.6)$$

We assume that the parameters $\alpha$ and $\lambda$ are unknown. To obtain the normal equations for the unknown parameters, we differentiate (3.6) partially with respect to $\alpha$ and $\lambda$ and equate to zero. The resulting equations are

$$0 = \frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \ln \left(1 - e^{-\lambda/x_U(n)}\right) \quad (3.7)$$

and

$$0 = \frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{1}{x_U(i)} + \alpha \frac{e^{-\lambda/x_U(n)}}{x_U(n) \left(1 - e^{-\lambda/x_U(n)}\right)} - \sum_{i=1}^{n} \frac{e^{-\lambda/x_U(i)}}{x_U(i) \left(1 - e^{-\lambda/x_U(i)}\right)}. \quad (3.8)$$
The solutions of the above equations are the maximum likelihood estimators of the GIED parameters $\alpha$ and $\lambda$, denoted $\hat{\alpha}_{MLE}$ and $\hat{\lambda}_{MLE}$, respectively. As the equations expressed in (3.7) and (3.8) cannot be solved analytically, one must use a numerical procedure to solve them.

### 3.2.2 Approximate Confidence Intervals

Since the maximum likelihood estimators of the unknown parameters $\alpha$ and $\lambda$ cannot be derived in closed form, it is not easy to derive the exact distributions of the MLEs. Hence, we cannot obtain exact confidence intervals for the parameters. We must use the large sample approximation. It is known that the asymptotic distribution of the MLE $\hat{\theta}$ is $(\hat{\theta} - \theta) \rightarrow N_2(0, I^{-1}(\theta))$ (see Lawless (1982)), where $I^{-1}(\theta)$, the inverse of the observed information matrix of the unknown parameters $\theta = (\alpha, \lambda)$, is

\[
I^{-1}(\theta) = \left( \begin{array}{cc} -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha^2} & -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha \partial \lambda} \\ -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha \partial \lambda} & -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \lambda^2} \end{array} \right) \bigg|_{(\alpha, \lambda) = (\hat{\alpha}, \hat{\lambda})}^{-1} = \left( \begin{array}{cc} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\lambda}) \\ \text{cov}(\hat{\lambda}, \hat{\alpha}) & \text{var}(\hat{\lambda}) \end{array} \right).
\]

The derivatives in $I(\theta)$ are given by

\[
\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha^2} = -\frac{n}{\alpha^2} \quad \text{(3.9)}
\]

\[
\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha \partial \lambda} = \frac{e^{-\lambda/x_U(n)}}{x_U(n) \left( 1 - e^{-\lambda/x_U(n)} \right)} = \frac{\partial^2 \log L}{\partial \alpha \partial \lambda} \quad \text{(3.10)}
\]

\[
\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \lambda^2} = -\frac{n}{\lambda^2} - \frac{\alpha}{x_U(n) \left( 1 - e^{-\lambda/x_U(n)} \right)^2} + \sum_{i=1}^{n} \frac{e^{-\lambda/x_U(i)}}{x_U(i) \left( 1 - e^{-\lambda/x_U(i)} \right)^2} \quad \text{(3.11)}
\]

The above approach is used to derive approximate $100(1 - \tau)%$ confidence intervals of the parameters $\alpha$ and $\lambda$ of the forms

\[
\hat{\alpha} \pm z_{\tau/2} \sqrt{\text{var}(\hat{\alpha})}
\]

and

\[
\hat{\lambda} \pm z_{\tau/2} \sqrt{\text{var}(\hat{\lambda})},
\]

where $z_{\tau/2}$ is the upper $(\tau/2)$th percentile of the standard normal distribution.
3.2.3 Bootstrap Confidence Intervals

Here we present bootstrap confidence intervals for the GIED. First, we will use a bootstrap method based on percentiles as presented in Efron (1982). The algorithm uses the following steps.

1. From the sample \( x_{U(1)}, x_{U(2)}, \ldots, x_{U(n)} \), compute the maximum likelihood estimates \( \hat{\alpha} \) and \( \hat{\lambda} \).

2. Generate a bootstrap sample \( x^*_{U(1)}, x^*_{U(2)}, \ldots, x^*_{U(n)} \), using \( \hat{\alpha} \) and \( \hat{\lambda} \) as parameters.

3. Compute the maximum likelihood estimates \( \hat{\alpha}^*_1 \) and \( \hat{\lambda}^*_1 \) based on the bootstrap sample.

4. Repeat steps 2 and 3 \( N \) times, and store \( \hat{\alpha}^*_i \) and \( \hat{\lambda}^*_i \) from step 3 for \( i = 1, 2, \ldots, N \).

5. Arrange the \( \hat{\alpha}^*_i \) in ascending order and obtain \( \alpha_U \) and \( \alpha_L \), the upper and lower limits of a 100(1 - \( \gamma \))% bootstrap confidence interval for \( \alpha \), as the \( \gamma/2 \) and \( (1 - \gamma/2) \) percentiles of the \( \hat{\alpha}^*_i \).

6. Arrange the \( \hat{\lambda}^*_i \) in ascending order and obtain \( \lambda_U \) and \( \lambda_L \), the upper and lower limits of a 100(1 - \( \gamma \))% bootstrap confidence interval for \( \lambda \), as the \( \gamma/2 \) and \( (1 - \gamma/2) \) percentiles of the \( \hat{\lambda}^*_i \).

A second bootstrap method is based on Hall (1988). The algorithm proceeds as follows.

1. From the sample \( x_{U(1)}, x_{U(2)}, \ldots, x_{U(n)} \), compute the maximum likelihood estimates \( \hat{\alpha} \) and \( \hat{\lambda} \).

2. Generate a bootstrap sample \( x^*_{U(1)}, x^*_{U(2)}, \ldots, x^*_{U(n)} \), using \( \hat{\alpha} \) and \( \hat{\lambda} \) as parameters.

3. Based on the bootstrap sample, compute the maximum likelihood estimates, \( \hat{\alpha}^* \) and \( \hat{\lambda}^* \), as well as the statistics.

\[
T_1^* = \frac{\sqrt{n}(\hat{\alpha}^* - \hat{\alpha})}{\sqrt{\text{Var}(\hat{\alpha}^*)}} \quad \text{and} \quad T_2^* = \frac{\sqrt{n}(\hat{\lambda}^* - \hat{\lambda})}{\sqrt{\text{Var}(\hat{\lambda}^*)}}
\]

where \( \text{Var}(\hat{\lambda}^*) \) and \( \text{Var}(\hat{\alpha}^*) \) are found from the observed Fisher information matrix.

4. Repeat steps 2 and 3 \( N \) times.
5. For the values of \( T^*_1 \) and \( T^*_2 \) found in step 3, let \( F(x) = P(T^*_i < x) \), for \( i = 1, 2 \) be the cumulative distribution function of \( T^*_1 \) and \( T^*_2 \). Define

\[
\hat{\alpha}_t(x) = \hat{\alpha} + \frac{\sqrt{Var(\hat{\alpha})}}{\sqrt{n}} \quad \text{and} \quad \hat{\lambda}_t(x) = \hat{\lambda} + \frac{\sqrt{Var(\hat{\lambda})}}{\sqrt{n}}
\]

where \( Var(\hat{\lambda}) \) and \( Var(\hat{\alpha}) \) are also found from the observed Fisher information matrix.

6. Approximate 100\((1 - \gamma)\)% confidence intervals of \( \alpha \) and \( \lambda \) are given by

\[
\left( \hat{\alpha}_t(\gamma/2), \hat{\alpha}_t(1 - \gamma/2) \right) \quad \text{and} \quad \left( \hat{\lambda}_t(\gamma/2), \hat{\lambda}_t(1 - \gamma/2) \right).
\]

These bootstrap confidence intervals provide an alternate to the approximate confidence intervals derived in Section 3.2.2.

### 3.3 Bayesian Estimation

In this section we consider Bayesian inference of the unknown parameters of the GIED. It is assumed that \( \alpha \) and \( \lambda \) have the independent gamma prior distributions with probability density functions

\[
g(\alpha) \propto \alpha^{a-1}e^{-b\alpha} \quad \alpha > 0 \quad (3.12)
\]

and

\[
g(\lambda) \propto \lambda^{c-1}e^{-d\lambda} \quad \lambda > 0. \quad (3.13)
\]

The hyper-parameters \( a, b, c, \) and \( d \) are known and non-negative. If both the parameters \( \alpha \) and \( \lambda \) are unknown, joint conjugate priors do not exist. It is not unreasonable to assume independent gamma priors on the shape and scale parameters for a two-parameter GIED, because gamma distributions are very flexible, and the Jeffreys (non-informative) prior, introduced by Jeffreys (1946) is a special case of this.

The joint prior distribution in this case is

\[
g(\alpha, \lambda) \propto \alpha^{a-1}e^{-b\alpha} \lambda^{c-1}e^{-d\lambda} \quad \alpha, \lambda > 0. \quad (3.14)
\]

Combining (3.14) with (3.5) and using Bayes theorem, the joint posterior distribution is derived as

\[
\pi(\alpha, \lambda|x) = \frac{\alpha^{n+a-1}\lambda^{n+c-1}e^{-d\lambda-b\alpha}(1 - e^{-\lambda/x_U(n)})^\alpha}{\int_0^\infty \prod_{i=1}^n \frac{e^{-\lambda/x_U(i)}}{(1 - e^{-\lambda/x_U(i)})}}, \quad (3.15)
\]

17
where,

$$J_o = \int_0^\infty \int_0^\infty \alpha^{n+a-1} \lambda^{n+c-1} e^{-d\lambda - ba} (1 - e^{-\lambda/x_{U(n)}})^\alpha \prod_{i=1}^n \frac{e^{-\lambda/x_{U(i)}}}{(1 - e^{-\lambda/x_{U(i)}})} d\alpha d\lambda. \quad (3.16)$$

The marginal posterior distribution of a parameter is obtained by integrating the joint posterior distribution with respect to the other parameter. Hence, the marginal posterior probability density functions of $\alpha$ and $\lambda$ are given, respectively, by

$$\pi_1(\alpha|x) = \frac{\alpha^{n+a-1} e^{-ba}}{J_0} \int_0^\infty e^{-d\lambda \lambda^{n+c-1}} (1 - e^{-\lambda/x_{U(n)}})^\alpha \prod_{i=1}^n \frac{e^{-\lambda/x_{U(i)}}}{(1 - e^{-\lambda/x_{U(i)}})} d\lambda \quad (3.17)$$

and

$$\pi_2(\lambda|x) = \frac{e^{-d\lambda \lambda^{n+c-1}} \prod_{i=1}^n e^{-\lambda/x_{U(i)}}}{J_0} \int_0^\infty \alpha^{n+a-1} e^{-ba} (1 - e^{-\lambda/x_{U(n)}})^\alpha d\alpha. \quad (3.18)$$

Next, we must consider the question of what loss function will be used to derive the estimators from the marginal posterior distributions.

### 3.3.1 Bayes Estimators Under the General Entropy Loss Function

Following Calabria and Pulcini (1996), the Bayes estimators for the parameters $\alpha$ and $\lambda$ for the probability density function (3.1) under the general entropy loss function may be defined as

$$\hat{\alpha}_{BGE} = \left[ E(\alpha^{-p}) \right]^{-1/p} \quad (3.19)$$

and

$$\hat{\lambda}_{BGE} = \left[ E(\lambda^{-p}) \right]^{-1/p} \quad (3.20)$$

respectively, provided that $E(\alpha^{-p})$ and $E(\lambda^{-p})$ exist and are finite. These estimators can be expressed as

$$\hat{\alpha}_{BGE} = \left[ \frac{J_\alpha}{J_0} \right]^{-1/p}$$

and

$$\hat{\lambda}_{BGE} = \left[ \frac{J_\lambda}{J_0} \right]^{-1/p}.$$

18
where

\[
J_\alpha = \int_0^\infty \int_0^\infty \alpha^{n+a-p-1} \lambda^{n+c-1} e^{-d\lambda} e^{-\alpha \left( b - \ln \left( 1 - e^{-\lambda/x U(n)} \right) \right)} \prod_{i=1}^n \frac{e^{-\lambda/x U(i)}}{1 - e^{-\lambda/x U(i)}} \, da d\lambda
\]

and

\[
J_\lambda = \int_0^\infty \int_0^\infty \alpha^{n+a-1} \lambda^{n+c-p-1} e^{-d\alpha} e^{-\alpha \left( b - \ln \left( 1 - e^{-\lambda/x U(n)} \right) \right)} \prod_{i=1}^n \frac{e^{-\lambda/x U(i)}}{J_o 1 - e^{-\lambda/x U(i)}} \, da d\lambda.
\]

All the double integrals above have no closed form. Therefore, we will implement the Metropolis–Hastings (M-H) algorithm to compute the estimators. The M-H algorithm is a powerful Markov Chain Monte Carlo (MCMC) algorithm. The M-H algorithm was introduced by Metropolis et al. (1953). For a discussion of the algorithm, the reader is referred to any Bayesian statistics textbook. In this chapter, we consider three special cases of the general entropy loss function, corresponding to \( p = -1, p = 1, \) and \( p = -2 \). It should be mentioned that for \( p = -1 \) the general entropy loss function simplifies to the squared-error loss (SEL) function. The weighted squared-error loss (WSEL) function results from \( p = 1 \). For \( p = -2 \), the general entropy loss function is referred to as the precautionary loss function.

### 3.3.2 Two-Sided Bayes Probability Intervals

The Bayesian method to interval estimation is much more direct than the frequentist method based on confidence intervals. Once the marginal posterior distribution of \( \alpha \) has been obtained, a symmetric 100(1 – \( \gamma \))% two-sided Bayes probability interval estimate of \( \alpha \), denoted by \([\alpha_L, \alpha_U]\), can be obtained by solving the two equations (see Martz and Waller (1982), pages 208–209)

\[
\int_0^{\alpha_L} \pi_1(w|x)dw = \frac{\gamma}{2} \quad (3.21)
\]

and

\[
\int_{\alpha_U}^\infty \pi_1(w|x)dw = \frac{\gamma}{2} \quad (3.22)
\]

for the limits \( \alpha_L \) and \( \alpha_U \).

Similarly, a symmetric 100(1 – \( \gamma \))% two-sided Bayes probability interval estimate of \( \lambda \), denoted by \([\lambda_L, \lambda_U]\), can be obtained by solving

\[
\int_0^{\lambda_L} \pi_2(w|x)dw = \frac{\gamma}{2} \quad (3.23)
\]

19
\[ \int_{\lambda_u}^{\infty} \pi_2(w|x)dw = \frac{\gamma}{2} \quad (3.24) \]

for the limits \( \lambda_L \) and \( \lambda_U \). Again, these equations cannot be solved in closed form.

### 3.3.3 Empirical Bayes Estimation

In the preceding sections, we have assumed the hyper-parameters \( a, b, c, \) and \( d \) are known. Empirical Bayes estimation addresses the question of estimating the hyper-parameters from existing data. When the current sample is observed, assume that \( m \) past samples \( X_{j,U(1)}, X_{j,U(2)}, \ldots, X_{j,U(n)} \), for \( j = 1, 2, \ldots, m \), are available. Each sample is assumed to be an upper record sample of size \( n \) from a \( \text{GIED}(\alpha, \lambda) \) distribution. The likelihood function for each sample \( j \) is given by

\[
L(\alpha, \lambda|x_j) = \alpha^n \lambda^n e^{\alpha \ln(1 - e^{-\lambda/x_{j,U(n)}})} \prod_{i=1}^{n} \frac{1}{x_{j,U(i)}} \cdot \frac{e^{-\lambda/x_{j,U(i)}}}{1 - e^{-\lambda/x_{j,U(i)}}}. \quad (3.25)
\]

For each sample \( j \), let \( \hat{\alpha}_j \) and \( \hat{\lambda}_j \) be the maximum likelihood estimates for \( \alpha \) and \( \lambda \), respectively, based on sample \( j \), which are obtained from (3.25). We then calculate the mean and variance of the maximum likelihood estimators for each of the \( j \) samples, equate these to the mean and variance of the gamma prior distribution, and solve for the hyper-parameters. We can find \( \hat{a} \) and \( \hat{b} \), estimators for \( a \) and \( b \), by solving

\[
\frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j = \frac{b}{a}
\]

and

\[
\frac{1}{m - 1} \sum_{j=1}^{m} \left( \hat{\alpha}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 = \frac{b}{a^2}.
\]

We can find \( \hat{c} \) and \( \hat{d} \), estimators for \( c \) and \( d \), by solving

\[
\frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j = \frac{d}{c}
\]

and

\[
\frac{1}{m - 1} \sum_{j=1}^{m} \left( \hat{\lambda}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 = \frac{d}{c^2}.
\]

Solving the above equations yields the estimators for the hyper-parameters

\[
\hat{a} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right) \left/ \left( \frac{1}{m - 1} \sum_{j=1}^{m} \left( \hat{\alpha}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 \right) \right.
\]
and
\[ \hat{b} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 \bigg/ \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\alpha}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 \right) \]
for the prior distribution for \( \alpha \). Similarly, estimators for the hyper-parameters for the prior distribution for \( \lambda \) can be found as
\[ \hat{c} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 \bigg/ \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\lambda}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 \right) \]
and
\[ \hat{d} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 \bigg/ \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\lambda}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 \right) . \]

The empirical Bayes estimators of \( \alpha \) and \( \lambda \) are found by substituting \( \hat{a} \), \( \hat{b} \), \( \hat{c} \), and \( \hat{d} \) into (3.17) and (3.18) and proceeding as before.

### 3.4 Prediction of Future Record Values

Next, we consider the problem of predicting future record values given a sample of observed record values.

#### 3.4.1 Non-Bayesian Prediction

Suppose that we observe the first \( n \) upper record values from a population with probability density function \( f(x; \theta) \). Our aim is to predict \( z = X_{U(m)} \), \( m > n \), having observed records \( X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)} \). The joint predictive likelihood function of \( z = X_{U(m)} \), and the possibly vector-valued parameter \( \theta \) is given by Basak and Balakrishnan (2003) as
\[
L(z, \theta, X) = \prod_{i=1}^{n} h(X_{u(i)}, \theta) \frac{[H(z; \theta) - H(X_{U(n)}; \theta)]^{m-n-1}}{\Gamma(m-n)} f(z; \theta)
\]
where,
\[
f(z; \alpha, \lambda) = \frac{\alpha \lambda}{z^2} e^{-\lambda/z} (1 - e^{-\lambda/z})^{\alpha-1},
\]
\[
F(z; \alpha, \lambda) = 1 - (1 - e^{-\lambda/z})^{\alpha},
\]
\[
H(z; \theta) = -\ln(1 - F(z; \theta)),
\]
and
The predictive likelihood function for the GIED is

\[ h(X_U; \theta) = \frac{f(X_U; \alpha, \lambda)}{S(X_U; \alpha, \lambda)} = e^{\lambda z_u(i) (1 - e^{-\lambda/z_u(i)})}. \]

The predictive likelihood function of \( y/u(i) \) records based on a Bayesian approach using the general entropy loss function. Prediction of future records has

\[ L(y; \alpha, \lambda) = \prod_{i=1}^{n} \frac{\alpha^{\lambda} \lambda^{\lambda/y_u(i)} e^{-\lambda/y_u(i)} \left[ -\ln \left( 1 - e^{-\lambda/y} \right)^{\alpha} + \ln \left( 1 - e^{-\lambda/y_u(i)} \right) \right]^{m-n-1}}{\Gamma(m-n)} \]

(3.26)

The log predictive likelihood is given by

\[ \log L = m \ln \alpha + (n + 1) \ln \lambda + \sum_{i=1}^{n} \ln e^{-\lambda/y_u(i)} - \sum_{i=1}^{n} \ln y_u(i)^2 - \sum_{i=1}^{n} \ln \left( 1 - e^{-\lambda/y_u(i)} \right) \]
\[ + (m - n - 1) \ln \left[ \ln \left( 1 - e^{-\lambda/y_u(n)} \right) - \ln \left( 1 - e^{-\lambda/y} \right) \right] - \ln \Gamma(m-n) \]
\[ + \ln e^{-\lambda/y} - \ln y^2 - (1 - \alpha) \ln \left( 1 - e^{-\lambda/y} \right). \]

So, the normal equations are as follows.

\[ 0 = \frac{\partial \log L}{\partial \alpha} = \frac{m}{\alpha} + \ln(1 - e^{-\lambda/y}) \]
\[ 0 = \frac{\partial \log L}{\partial \lambda} = \frac{n + 1}{\lambda} - \sum_{i=1}^{n} \frac{1}{y_u(i)} - \sum_{i=1}^{n} \frac{e^{-\lambda/y_u(i)}}{y_u(i) \left( 1 - e^{-\lambda/y_u(i)} \right)} - \frac{1}{y} - (1 - \alpha) \frac{e^{-\lambda/y}}{y \left( 1 - e^{-\lambda/y} \right)} \]
\[ + (m - n - 1) \left[ \frac{\ln \left( 1 - e^{-\lambda/y_u(n)} \right) - \ln \left( 1 - e^{-\lambda/y} \right)}{\ln \left( 1 - e^{-\lambda/y_u(n)} \right) - \ln \left( 1 - e^{-\lambda/y} \right)} \right] \]
\[ 0 = \frac{\partial \log L}{\partial y} = (m - n - 1) \frac{\lambda e^{-\lambda/y}}{y^2 (1 - e^{-\lambda/y})} \]
\[ + \frac{\lambda}{y^2} - \frac{2}{y} - (1 - \alpha) \frac{e^{-\lambda/y}}{y \left( 1 - e^{-\lambda/y} \right)}. \]

The above equations cannot be solved in closed form. Thus, we must use a numerical procedure to find the maximum likelihood predictor.

### 3.4.2 Bayes Prediction

In this section, we consider the prediction of future records based on a Bayesian approach using the general entropy loss function. Prediction of future records has
been studied by a number of statisticians (see Berred (1998), Dunsmore (1983), Ahsanullah (1980), and Arnold et al. (1998)). Suppose that we have $n$ upper records $X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}$ from a GIED. Based on such a record sample, we are interested in obtaining a Bayesian prediction interval for the future upper record $X_{U(m)}$, for some $m > n$, with a certain confidence. The conditional probability density function of $z = X_{U(m)}$ for a given $y = X_{U(n)}$ is given (see Ahsanullah (1995)) by

$$f(z|y; \alpha, \lambda) = \frac{\log S(y; \alpha, \lambda) - \log S(z; \alpha, \lambda)}{\Gamma(m-n)} \frac{f(z; \alpha, \lambda)}{S(y; \alpha, \lambda)} \quad z > y \quad (3.27)$$

where

$$S(y; \alpha, \lambda) = (1 - e^{-\lambda/y})^\alpha$$

and

$$S(z; \alpha, \lambda) = (1 - e^{-\lambda/z})^\alpha.$$

For the GIED, with probability density function given in (3.1), the function $f(z|y; \alpha, \lambda)$ can be shown to be

$$f(z|y; \alpha, \lambda) = \frac{\alpha^{m-n} \lambda e^{-\lambda/z} (1 - e^{-\lambda/z})^{\alpha-1}}{z^2 (1 - e^{-\lambda/y})^\alpha \Gamma(m-n)} \left[ \log \left( \frac{1 - e^{-\lambda/y}}{1 - e^{-\lambda/z}} \right) \right]^{m-n-1} \quad 0 < y_n < \infty. \quad (3.28)$$

As we know that future record values satisfy the Markovian (memoryless) property, the future upper record $z = X_{U(m)}$ given the set of the first $n$ upper records $X = \{X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}\}$ depends only on the current upper record $y = X_{U(n)}$. Therefore, the conditional probability density function of $z$ given $x$ is the same as the conditional probability density function of $z$ given $y$. The predictive probability density function of $z$ given $x$ is

$$f^*(z|x) = \int_0^\infty \int_0^\infty f(z|y; \alpha, \lambda) \pi(\alpha, \lambda|x) \, d\alpha \, d\lambda \quad (3.29)$$

where $f(z|y; \alpha, \lambda)$ and $\pi(\alpha, \lambda|x)$ are given respectively by (3.28) and (3.15).

The predictive limits of the $100(1 - \gamma)$% two sided interval of the future upper record $z$ can be obtained by solving

$$\int_y^{z_L} f^*(z|x) \, dz = \frac{\gamma}{2} \quad (3.30)$$

and
\[ \int_{z_L}^{\infty} f^*(z|x) dz = \frac{\gamma}{2} \]  
(3.31)

with respect to the lower and upper limits \( z_L \) and \( z_U \).

The Bayesian prediction bounds for \( Y = X_{U(m)} \) are obtained by evaluating \( P(Y \geq \eta|x) \), for some given positive value of \( \eta \). From (3.29), we have

\[ P(Y \geq \eta|x) = \int_{\eta}^{\infty} f^*(z|x) dz. \]

The 100\((1-\gamma)\)% predictive interval for \( Y = X_{U(m)} \) is obtained by evaluating both the lower \( L(x) \) and the upper \( U(x) \) limits which satisfy \( P(Y \geq L(x)|x) = 1 - \gamma/2 \) and \( P(Y \geq U(x)|x) = \gamma/2 \).

### 3.4.3 Conditional Median Prediction

We now consider the conditional median prediction of future record values. Suppose that we have \( n \) upper records \( X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)} \) from a GIED(\( \alpha, \lambda \)) distribution and we are interested in predicting \( z = X_{U(m)} \), the \( m \)th upper record, for some \( m > n \).

It is well known that the distribution of \( z = X_{U(m)} \) depends only on the current upper record, \( y = X_{U(n)} \). The conditional median predictor is found as the median of the conditional distribution of \( z \) given \( y \).

First, we consider the case where \( \alpha \) and \( \lambda \) are unknown. The conditional distribution of \( z \) given \( y \) is found in (3.29). To find the median, let the cumulative distribution function of \( z \) given \( y \) be

\[ F(z|y) = \int_{0}^{z} f^*(\tau|y) d\tau. \]

Let \( F^{-1}(u) \) be the inverse distribution function. Equating \( F^{-1}(u) = 1/2 \) and solving for \( u \) yields \( \hat{z}_{CMP} \), the conditional median predictor of \( z \) when \( \alpha \) and \( \lambda \) are unknown. The solution does not exist in closed form.

In the case where \( \alpha \) and \( \lambda \) are known, the conditional median predictor can be found as the median of (3.28). Let the cumulative distribution function of \( z \) given \( y \) be

\[ F(z|y; \alpha, \lambda) = \int_{0}^{z} f(\tau|y; \alpha, \lambda) d\tau. \]

Again, let \( F^{-1}(u) \) be the inverse distribution function. Equating \( F^{-1}(u) = 1/2 \) and solving for \( u \) yields \( \hat{z}_{CMP} \), the conditional median predictor of \( z \) when \( \alpha \) and \( \lambda \) are known. Again, the solution does not exist in closed form.
3.5 Numerical Results

In this section we first present some results from a simulation study. Then we present some data analysis results.

3.5.1 Simulation Study

The goal of this simulation study is to examine and compare the behaviors of the maximum likelihood and Bayes estimators for the two parameters of the GIED based on record values for various sample sizes. For each simulation, we consider $\alpha = 1$ and $\lambda = 2$. We consider three different sample sizes, $n = 5, 7, 10$. First we compute the maximum likelihood estimators and approximate confidence intervals using the methods described in Section 3.2. We report the average bias and mean squared error (MSE) over 1000 replications. We also report the average interval length and coverage probability for 95% approximate confidence intervals. As the maximum likelihood estimators cannot be found in closed form, we calculate estimates using the BFGS method as implemented in the R package maxLik.

The Bayes estimators cannot be found in closed form. Therefore, we use the Metropolis–Hastings algorithm to compute Bayes estimates. We use informative priors for both $\alpha$ and $\lambda$. The chosen hyper-parameters are $a = c = 4$ and $b = d = 1$. Bayes estimators are computed using the general entropy loss function with $p = -2, -1, 1$. It should be noted that $p = -2$ corresponds to the precautionary loss function, $p = -1$ corresponds to the squared-error loss function, and $p = 1$ corresponds to the weighted squared-error loss function. This allows us to consider the Bayes estimators under both symmetric and asymmetric loss functions. The proposal distribution used for the M-H algorithm is a chi-square distribution. We generate 5000 samples after 5000 burn-in samples.

To evaluate the convergence of the M-H algorithm, we use Gelman–Rubin diagnostics, using the R package coda. For both $\alpha$ and $\lambda$, three different Markov chains are simulated. The potential scale reduction factor (psrf) for $\alpha$ is 1.01, and the psrf for lambda is 1. Since both are close to 1, we can conclude that the Markov chains converged quite well to the stationary distribution.

We also perform graphical diagnostics. Trace plots and density plots of the MCMC are plotted here. One can see from the trace plots that the chains have converged quite well. The density plots reflect the fact that the M-H algorithm uses the proposal distribution to create a mixture distribution.

Table 3.1 contains the bias and mean squared error (MSE) for the maximum
Table 3.1: Average bias and MSE of estimators for $\alpha$ and $\lambda$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\hat{\alpha}_{MLE}$</th>
<th>$\hat{\alpha}_{SEL}$</th>
<th>$\hat{\alpha}_{WSEL}$</th>
<th>$\hat{\alpha}_{PL}$</th>
<th>$\hat{\lambda}_{MLE}$</th>
<th>$\hat{\lambda}_{SEL}$</th>
<th>$\hat{\lambda}_{WSEL}$</th>
<th>$\hat{\lambda}_{PL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.80</td>
<td>0.021</td>
<td>0.016</td>
<td>0.024</td>
<td>0.33</td>
<td>0.17</td>
<td>0.22</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(2.85)</td>
<td>(0.0017)</td>
<td>(0.00093)</td>
<td>(0.0023)</td>
<td>(1.63)</td>
<td>(0.75)</td>
<td>(0.65)</td>
<td>(0.79)</td>
</tr>
<tr>
<td>7</td>
<td>0.47</td>
<td>0.018</td>
<td>0.014</td>
<td>0.021</td>
<td>0.32</td>
<td>0.15</td>
<td>0.21</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(0.88)</td>
<td>(0.0010)</td>
<td>(0.00062)</td>
<td>(0.0013)</td>
<td>(1.52)</td>
<td>(0.74)</td>
<td>(0.64)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>10</td>
<td>0.45</td>
<td>0.017</td>
<td>0.012</td>
<td>0.019</td>
<td>0.30</td>
<td>0.14</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.00099)</td>
<td>(0.00051)</td>
<td>(0.0027)</td>
<td>(1.49)</td>
<td>(0.039)</td>
<td>(0.067)</td>
<td>(0.030)</td>
</tr>
</tbody>
</table>

Table 3.2 includes the average interval length and coverage probability for 95% approximate confidence intervals for $\alpha$ and $\lambda$. Average interval length is listed first and coverage probability follows in parentheses.

Finally, we present simulation pertaining to empirical Bayes estimation. Table 3.3 includes the average bias and MSE of the empirical Bayes estimators for $\alpha$ and $\lambda$. 
<table>
<thead>
<tr>
<th>( n )</th>
<th>( CI_\alpha )</th>
<th>( CI_\lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.66</td>
<td>8.59</td>
</tr>
<tr>
<td></td>
<td>(0.984)</td>
<td>(0.943)</td>
</tr>
<tr>
<td>7</td>
<td>2.36</td>
<td>8.49</td>
</tr>
<tr>
<td></td>
<td>(0.997)</td>
<td>(0.952)</td>
</tr>
<tr>
<td>10</td>
<td>1.91</td>
<td>7.79</td>
</tr>
<tr>
<td></td>
<td>(0.998)</td>
<td>(0.959)</td>
</tr>
</tbody>
</table>

Table 3.2: Average interval length and coverage probability for confidence intervals

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \hat{\alpha}_{SEL} )</th>
<th>( \hat{\alpha}_{WSEL} )</th>
<th>( \hat{\alpha}_{PL} )</th>
<th>( \hat{\lambda}_{SEL} )</th>
<th>( \hat{\lambda}_{WSEL} )</th>
<th>( \hat{\lambda}_{PL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.59</td>
<td>0.46</td>
<td>0.68</td>
<td>1.25</td>
<td>0.19</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.22)</td>
<td>(0.64)</td>
<td>(2.75)</td>
<td>(0.22)</td>
<td>(6.80)</td>
</tr>
<tr>
<td>7</td>
<td>0.49</td>
<td>0.39</td>
<td>0.54</td>
<td>0.73</td>
<td>0.0031</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.19)</td>
<td>(0.36)</td>
<td>(1.34)</td>
<td>(0.010)</td>
<td>(2.67)</td>
</tr>
<tr>
<td>10</td>
<td>0.41</td>
<td>0.34</td>
<td>0.45</td>
<td>0.39</td>
<td>0.15</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.12)</td>
<td>(0.22)</td>
<td>(1.04)</td>
<td>(0.0099)</td>
<td>(2.23)</td>
</tr>
</tbody>
</table>

Table 3.3: Average bias and MSE of empirical Bayes estimators for \( \alpha \) and \( \lambda \)

Empirical Bayes simulation results are averages over 1000 replications. The hyper-parameters are estimated from 100 previous samples.

### 3.5.2 Data Analysis

To illustrate the results of this chapter, we analyze one real data set. The data is presented on page 3 of Lawless (1982). It consists of \( n = 11 \) times to breakdown of electrical insulating fluid subjected to 30 kilovolts. The data, under a logarithm transformation, is 2.836, 3.120, 3.045, 5.169, 4.934, 4.970, 3.018, 3.770, 5.272, 3.856, 2.046. From this data set, we extract the \( n = 4 \) upper record values 2.836, 3.120, 5.169, 5.272. Using the methods described in this chapter, we compute the maximum likelihood estimates, as well as Bayes estimates under the squared-error, weighted squared-error, and precautionary loss functions. Bayes estimates were computed with the assumption of independent gamma priors with hyper-parameters \( a = c = 0.5 \) and \( b = d = 18 \). The estimates are in Table 3.4. In the table, the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( MLE )</th>
<th>( SEL )</th>
<th>( WSEL )</th>
<th>( PL )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>45.523</td>
<td>43.768</td>
<td>41.259</td>
<td>44.961</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>13.055</td>
<td>17.560</td>
<td>17.092</td>
<td>17.792</td>
</tr>
</tbody>
</table>

Table 3.4: Estimates for \( \alpha \) and \( \lambda \) based on the data

abbreviation MLE stands for maximum likelihood estimates, and the abbreviations SEL, WSEL, and PL refer to Bayes estimates under the squared error loss function, the weighted squared error loss function, and the precautionary loss function, respectively.
After fitting the GIED, we must assess the fit. Figures 3.7 and 3.8 contain graphs of empirical and fitted survivor functions. Plotted first is the fitted survivor function using maximum likelihood estimates. Plotted second is the fitted survivor function using Bayes estimates under the weighted squared-error loss function. The weighted squared-error loss function appears to perform best of the three loss functions we consider in this situation. One can see that, while not an ideal fit, the GIED is acceptable for this data set. It appears that the Bayes estimators under the weighted squared-error loss function provide a better fit than the maximum likelihood estimators do.

![Figure 3.7: Empirical and fitted survivor function using maximum likelihood estimators](image)

![Figure 3.8: Empirical and fitted survivor function using Bayes estimators under the WSEL function](image)

Finally, a two-sided Bayes probability interval for $\alpha$ is $(28.174, 61.066)$. A two-sided Bayes probability interval for $\lambda$ is $(13.450, 22.397)$.

To illustrate the prediction methods we have presented, we consider a synthetic sample of $n = 10$ simulated record values. We will then predict the 11th upper record value. A randomly generated sample of upper records from a GIED with $\alpha = 1$ and $\lambda = 2$ turns out to be $1.089, 3.638, 5.255, 6.093, 9.695, 31.931, 43.306, 92.525, 102.818, 4752.639$. We use the M-H algorithm to generate a sample from (3.29). Again, we assume independent gamma priors. The hyper-parameters chosen are $a = c = 90$ and $b = d = 0.8$.

The conditional median predictor, found as the median of the sample generated by the M-H algorithm, is 6491. A 95% Bayesian predictive interval for the next future record is $(5777, 7875)$. 

28
3.6 Conclusion

The generalized inverted exponential distribution provides an excellent model for lifetime data for a variety of situations. Thus, it is important for the analyst to have reliable statistical methods to use for this distribution. We have provided both frequentist and Bayesian methods of estimating the parameters based on samples of upper record values and methods of predicting future record values. We have examined and compared the different methods, including Bayesian methods under different loss functions. Simulation and data analysis help to illustrate these results.
4.1 Introduction

A rich class of probability distributions was introduced by Burr (1942), which includes twelve different forms of cumulative distribution functions for modeling lifetime data. Among them, Burr Type X and Burr Type XII are the most popular ones. Several authors consider different aspects of the Burr Type X and Burr Type XII distributions. See, for example, Ahmad, Fakhry and Jaheen (1997), Jaheen (1995, 1996), Raqab (1998), Rodriguez (1977), Sartawi and Abu–Salih (1991), Surles and Padgett (1998), and Wongo (1993). For an excellent review of these two distributions the readers are refereed to Johnson et al. (1995). Recently, Surles and Padgett (2001) (see also Surles and Padgett (2004)) introduced a two-parameter Burr Type X distribution and correctly named it the generalized Rayleigh distribution. Note that the two-parameter generalized Rayleigh distribution is a particular case of the generalized Weibull distribution, originally proposed by Mudholkar and Srivastava (1993). Also see Mudholkar, Srivastava and Freimer (1995). In this chapter, we call the two-parameter Burr Type X distribution the generalized Rayleigh (GR) distribution. For \( \alpha > 0 \) and \( \theta > 0 \), the two-parameter GR distribution has the distribution function

\[
F(x; \alpha, \theta) = \left(1 - e^{-(\theta x)^2}\right)^\alpha \quad x \geq 0, \alpha, \theta > 0.
\]

Therefore, the GR distribution has the density function

\[
f(x; \alpha, \theta) = 2\alpha \theta^2 xe^{-\theta x^2} \left(1 - e^{-(\theta x)^2}\right)^{\alpha-1} \quad x \geq 0, \alpha, \theta > 0.
\]

Letting \( \lambda = \theta^2 \), the probability density function of the two-parameter generalized Rayleigh distribution takes the form

\[
f(x; \alpha, \lambda) = 2\alpha \lambda xe^{-\lambda x^2} \left(1 - e^{-\lambda x^2}\right)^{\alpha-1} \quad x \geq 0, \alpha, \lambda > 0. \tag{4.1}
\]
The corresponding cumulative distribution function is
\[ F(x; \alpha, \lambda) = \left(1 - e^{-\lambda x^2}\right)^\alpha \quad x \geq 0. \tag{4.2} \]

The survivor function takes the form
\[ S(x; \alpha, \lambda) = 1 - \left(1 - e^{-\lambda x^2}\right)^\alpha \quad x \geq 0, \tag{4.3} \]

and the corresponding hazard function is
\[ h(x; \alpha, \lambda) = \frac{2x\alpha \lambda e^{-\lambda x^2} \left(1 - e^{-\lambda x^2}\right)^{\alpha-1}}{1 - (1 - e^{-\lambda x^2})^\alpha} \quad x \geq 0. \tag{4.4} \]

Here \( \alpha, \lambda > 0 \) are the shape and scale parameters respectively. From now on, the two-parameter generalized Rayleigh distribution with parameters \( \alpha \) and \( \lambda \) will be denoted by GR(\( \alpha, \lambda \)). It is observed in Raqab and Kundu (2003) that for \( \alpha \leq 1/2 \), the probability density function of a GR distribution is a decreasing function and it is a right skewed unimodal function for \( \alpha > 1/2 \). Different forms of the density functions can be found in Raqab and Kundu (2003). It is also observed that the hazard function of a GR distribution can be either bathtub shaped or an increasing function, depending on the shape parameter, \( \alpha \). For \( \alpha < 1/2 \), the hazard function of a \( \text{GR} (\alpha, \lambda) \) is bathtub type and for \( \alpha \geq 1/2 \), it is increasing. Surles and Padgett (2001) showed that the two-parameter GR distribution can be used quite effectively in modeling strength data and general lifetime data. Plotted below in Figure 4.1 are GR probability density functions for certain parameter settings. The solid line uses \( \alpha = 0.25 \), the dashed line uses \( \alpha = 2 \), and the dotted line uses \( \alpha = 1 \). All three density function plots use scale parameter \( \lambda = 1 \). Plotted in Figure 4.2 are hazard functions for certain parameter settings. The solid line uses \( \alpha = 0.25 \), the dashed line uses \( \alpha = 1 \), and the dotted line uses \( \alpha = 4 \). All hazard function plots use scale parameter \( \lambda = 0.25 \).
Record values are significant in many real life situations such as industry, weather, seismology, athletic events and economics. Chandler (1952) was the first to introduce record values, record times and inter-record times for independent and identically distributed (iid) sequences of random variables. Feller (1966) cited some examples of record values with respect to gambling problems. The theory of record values and their distributional properties has been extensively studied in the literature. See Ahsanullah (1988, 1995), Arnold and Balakrishnan (1989), Arnold et al. (1992, 1998), and Kamps (1995) for reviews on various developments in the area of records.

The rest of this chapter is organized as follows. In Section 4.2, we derive the maximum likelihood estimators for the two parameters of a GR(α, λ) based on upper record values. In Section 4.3, we derive the Bayes estimators for the unknown parameters under the general entropy loss function. Two-sided Bayes probability intervals are also presented. Section 4.4 presents the Bayesian predictive distribution of the future record values based on a given set of the first n upper records. Numerical comparisons of the maximum likelihood and Bayes estimators and analysis of one data set are presented in Section 4.5. Finally, we conclude the chapter in Section 4.6.

4.2 Non-Bayesian Estimation

In this section we discuss the process of obtaining the maximum likelihood estimators of the parameters α and λ based on upper record values. Suppose we observe n upper record values \( X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)} \) from a sequence of independent and identically distributed random variables from a GR(α, λ) with probability density function in (4.1). Arnold et al. (1998) give the likelihood function as

\[
L(\alpha, \lambda|x) = f(x_{U(n)}; \alpha, \lambda) \prod_{i=1}^{n-1} \frac{f(x_{U(i)}; \alpha, \lambda)}{1 - F(x_{U(i)}; \alpha, \lambda)} \quad 0 \leq x_{U(1)} < \ldots < x_{U(n)} < \infty.
\]

(4.5)

Substituting (4.1) and (4.2) in (4.5), we get the likelihood function

\[
L(\alpha, \lambda|x) = 2^n \alpha^n \lambda^n \prod_{i=1}^{n} x_{U(i)} e^{-\lambda \sum_{i=1}^{n} x_{U(i)}^2} \frac{\prod_{i=1}^{n} \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha-1}}{\prod_{i=1}^{n-1} \left[1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha}\right]}.
\]

(4.6)

The maximum likelihood estimators are the values of α and λ that maximize the likelihood function.
4.2.1 Maximum Likelihood Estimation

The log likelihood function \( \ell(\alpha, \lambda|x) = \log L(\alpha, \lambda|x) \), dropping terms that do not involve \( \alpha \) and \( \lambda \), is

\[
\ell(\alpha, \lambda|x) = n(\log \alpha + \log \lambda) - \lambda \sum_{i=1}^{n} x_{U(i)}^2 + (\alpha - 1) \sum_{i=1}^{n} \log(1 - e^{-\lambda x_{U(i)}^2}) \] (4.7)

\[
- \sum_{i=1}^{n-1} \log \left[ 1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha} \right].
\]

We assume that the parameters \( \alpha \) and \( \lambda \) are unknown. To obtain the normal equations for the unknown parameters, we differentiate (4.7) partially with respect to \( \alpha \) and \( \lambda \) and equate to zero. The resulting equations are given below in (4.8) and (4.9), respectively.

\[
0 = \frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} x_{U(i)}^2 \left(1 - e^{-\lambda x_{U(i)}^2}\right) + \sum_{i=1}^{n-1} \frac{(1 - e^{-\lambda x_{U(i)}^2})^\alpha \log \left(1 - e^{-\lambda x_{U(i)}^2}\right)}{\left[1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha}\right]} \] (4.8)

\[
0 = \frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_{U(i)}^2 + (\alpha - 1) \sum_{i=1}^{n} \frac{x_{U(i)}^2 e^{-\lambda x_{U(i)}^2}}{\left(1 - e^{-\lambda x_{U(i)}^2}\right)} + \alpha \sum_{i=1}^{n-1} \frac{x_{U(i)}^2 e^{-\lambda x_{U(i)}^2} \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha-1}}{\left[1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha}\right]} \] (4.9)

The solutions of the above equations are the MLEs of the parameters \( \alpha \) and \( \lambda \) denoted by \( \hat{\alpha}_{MLE} \) and \( \hat{\lambda}_{MLE} \), respectively. The equations expressed in (4.8) and (4.9) cannot be solved analytically. Therefore, we must use iterative methods to find numerical solutions.

4.2.2 Approximate Confidence Intervals

Because the maximum likelihood estimators of the unknown parameters cannot be solved for in closed form, it is not easy to derive confidence intervals for \( \alpha \) and \( \lambda \). We must use the large sample approximation. It is known that the asymptotic distribution of the MLE \( \hat{\theta} \) is \( \hat{\theta} \rightarrow \mathcal{N}(\theta, I^{-1}(\theta)) \), where \( I^{-1}(\theta) \), the inverse of the observed information matrix of the unknown parameters \( \theta = (\alpha, \lambda) \), is

\[
I^{-1}(\theta) = \left( \begin{array}{cc} -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha^2} & -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \alpha \partial \lambda} \\ -\frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \lambda \partial \alpha} & \frac{\partial^2 \ell(\alpha, \lambda|x)}{\partial \lambda^2} \end{array} \right)^{-1}_{(\alpha, \lambda)=(\hat{\alpha}, \hat{\lambda})} = \left( \begin{array}{cc} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\lambda}) \\ \text{cov}(\hat{\lambda}, \hat{\alpha}) & \text{var}(\hat{\lambda}) \end{array} \right).
\]

The derivatives above are

33
\[
\frac{\partial^2 \ell(\alpha, \lambda | x)}{\partial \alpha^2} = -\frac{n}{\alpha^2} + \sum_{i=1}^{n-1} \left( 1 - e^{-\lambda x_{U(i)}} \right)^\alpha \left[ \log \left( 1 - e^{-\lambda x_{U(i)}} \right) \right]^2,
\]
(4.10)

\[
\frac{\partial^2 \ell(\alpha, \lambda | x)}{\partial \alpha \partial \lambda} = \sum_{i=1}^{n} \frac{x_{U(i)}^2 e^{\lambda x_{U(i)}^2}}{e^{\lambda x_{U(i)}^2} - 1} - \sum_{i=1}^{n-1} \frac{n-1}{x_{U(i)}^2 \left( 1 - e^{-\lambda x_{U(i)}^2} \right)^\alpha \left[ (1 - e^{-\lambda x_{U(i)}^2}) \right] - \alpha \log \left( 1 - e^{-\lambda x_{U(i)}^2} \right) - 1}{(e^{\lambda x_{U(i)}^2} - 1) \left[ (1 - e^{-\lambda x_{U(i)}^2}) \right] - 1}^2,
\]
(4.11)

\[
\frac{\partial^2 \ell(\alpha, \lambda | x)}{\partial \lambda^2} = -\frac{n}{\lambda^2} - (\alpha - 1) \sum_{i=1}^{n} \frac{x_{U(i)}^4 e^{\lambda x_{U(i)}^2}}{e^{\lambda x_{U(i)}^2} - 1}^2 + \alpha \sum_{i=1}^{n-1} \frac{n-1}{x_{U(i)}^2 \left( 1 - e^{-\lambda x_{U(i)}^2} \right)^\alpha \left( 1 - e^{-\lambda x_{U(i)}^2} \right)^\alpha - e^{\lambda x_{U(i)}^2}} \left[ (1 - e^{-\lambda x_{U(i)}^2}) \right] - 1}^2 \left. \right|_{\alpha}^2.
\]
(4.12)

Using this approach, we can derive $100(1 - \tau)\%$ confidence intervals of the form

\[
\hat{\alpha} \pm z_{\tau/2} \sqrt{\text{var}(\hat{\alpha})}
\]

and

\[
\hat{\lambda} \pm z_{\tau/2} \sqrt{\text{var}(\hat{\lambda})}
\]

where, $z_{\tau/2}$ is the upper $(\tau/2)\text{th}$ percentile of the standard normal distribution.

### 4.2.3 Bootstrap Confidence Intervals

Here we present bootstrap confidence intervals for the GR distribution. First, we will use a bootstrap method based on percentiles as presented in Efron (1982). The algorithm uses the following steps.

1. From the sample $x_{U(1)}, x_{U(2)}, \ldots, x_{U(n)}$, compute the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}$.

2. Generate a bootstrap sample $x_{U(1)}^*, x_{U(2)}^*, \ldots, x_{U(n)}^*$, using $\hat{\alpha}$ and $\hat{\lambda}$ as parameters.

3. Compute the maximum likelihood estimates $\hat{\alpha}_i^*$ and $\hat{\lambda}_i^*$ based on the bootstrap sample.

4. Repeat steps 2 and 3 $N$ times, and compute $\hat{\alpha}_i^*$ and $\hat{\lambda}_i^*$ for $i = 1, 2, \ldots, N$. 

34
5. Arrange the $\hat{\alpha}^*_i$ in ascending order and obtain $\alpha_U$ and $\alpha_L$, the upper and lower limits of a $100(1 - \gamma)$% bootstrap confidence interval for $\alpha$, as the $\gamma/2$ and $(1 - \gamma/2)$ percentiles of the $\hat{\alpha}^*_i$.

6. Arrange the $\hat{\lambda}^*_i$ in ascending order and obtain $\lambda_U$ and $\lambda_L$, the upper and lower limits of a $100(1 - \gamma)$% bootstrap confidence interval for $\lambda$, as the $\gamma/2$ and $(1 - \gamma/2)$ percentiles of the $\hat{\lambda}^*_i$.

A second bootstrap method is based on Hall (1988). The algorithm proceeds as follows.

1. From the sample $x_{U(1)}, x_{U(2)}, \ldots, x_{U(n)}$, compute the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}$.

2. Generate a bootstrap sample $x_{U(1)}^*, x_{U(2)}^*, \ldots, x_{U(n)}^*$, using $\hat{\alpha}$ and $\hat{\lambda}$ as parameters.

3. Based on the bootstrap sample, compute the maximum likelihood estimates, $\hat{\alpha}^*$ and $\hat{\lambda}^*$, as well as the statistics

$$T_1^* = \frac{\sqrt{n}(\hat{\alpha}^* - \hat{\alpha})}{\sqrt{Var(\hat{\alpha}^*)}} \quad \text{and} \quad T_2^* = \frac{\sqrt{n}(\hat{\lambda}^* - \hat{\lambda})}{\sqrt{Var(\hat{\lambda}^*)}}$$

where $Var(\hat{\lambda}^*)$ and $Var(\hat{\alpha}^*)$ are found from the observed Fisher information matrix.

4. Repeat steps 2 and 3 $N$ times.

5. For the values of $T_1^*$ and $T_2^*$ found in step 3, let $F(x) = P(T_i^* < x)$, for $i = 1, 2$ be the cumulative distribution function of $T_1^*$ and $T_2^*$. Define

$$\hat{\alpha}_t(x) = \hat{\alpha} + \frac{\sqrt{Var(\hat{\alpha})}}{\sqrt{n}} \quad \text{and} \quad \hat{\lambda}_t(x) = \hat{\lambda} + \frac{\sqrt{Var(\hat{\lambda})}}{\sqrt{n}}$$

where $Var(\hat{\lambda})$ and $Var(\hat{\alpha})$ are also found from the observed Fisher information matrix.

6. Approximate $100(1 - \gamma)$% confidence intervals of $\alpha$ and $\lambda$ are given by

$$\left(\hat{\alpha}_t(\gamma/2), \hat{\alpha}_t(1 - \gamma/2)\right) \quad \text{and} \quad \left(\hat{\lambda}_t(\gamma/2), \hat{\lambda}_t(1 - \gamma/2)\right).$$

These bootstrap confidence intervals provide an alternate to the approximate confidence intervals derived in section 4.2.2.
4.3 Bayes Estimation

Next, we will consider Bayesian estimation of the unknown parameters $\alpha$ and $\lambda$. It is assumed that $\alpha$ and $\lambda$ have the independent gamma prior distributions with probability density functions

$$
g(\alpha) \propto \alpha^{d-1} e^{-ca} \quad \alpha > 0 \quad (4.13)
$$

and

$$
g(\lambda) \propto \lambda^{b-1} e^{-a\lambda} \quad \lambda > 0. \quad (4.14)
$$

The hyper parameters $a$, $b$, $c$, and $d$ are known and non-negative. If both the parameters are unknown, joint conjugate priors do not exist. Note that it is not unreasonable to assume independent gamma priors on the shape and scale parameters for a two-parameter lifetime distribution, because gamma distributions are very flexible and the Jeffreys (non-informative) prior, introduced by Jeffreys (1946), is a special case of this. Independent gamma priors have been used in the Bayesian analysis of the Weibull distribution. See, for example, Kundu (2008).

Combining (4.13) and (4.14) with (4.6) and using Bayes theorem, the joint posterior distribution is derived as

$$
\pi(\alpha, \lambda|x) = \alpha^{n+d-1} \lambda^{n+b-1} e^{-ca} e^{-\lambda \sum_{i=1}^{n} x_{u(i)}^2 + a} \frac{\prod_{i=1}^{n} \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a-1}}{J_0 \prod_{i=1}^{n-1} \left[ 1 - \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a} \right]} \int_{J_0} \prod_{i=1}^{n-1} \left[ 1 - \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a} \right] d\lambda d\alpha. \quad (4.15)
$$

where,

$$
J_0 = \int_{0}^{\infty} \int_{0}^{\infty} \alpha^{n+d-1} \lambda^{n+b-1} e^{-ca} e^{-\lambda \sum_{i=1}^{n} x_{u(i)}^2 + a} \frac{\prod_{i=1}^{n} \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a-1}}{\prod_{i=1}^{n-1} \left[ 1 - \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a} \right]} d\lambda d\alpha.
$$

The marginal posterior of a parameter is obtained by integrating the joint posterior with respect to the other parameter. Hence, the marginal posterior probability density functions of $\alpha$ and $\lambda$ are given, respectively, by

$$
\pi_1(\alpha|x) = \frac{\alpha^{n+d-1} e^{-ca}}{J_0} \int_{0}^{\infty} \lambda^{n+b-1} e^{-\lambda \sum_{i=1}^{n} x_{u(i)}^2 + a} \frac{\prod_{i=1}^{n} \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a-1}}{\prod_{i=1}^{n-1} \left[ 1 - \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a} \right]} d\lambda \quad (4.16)
$$

and

$$
\pi_2(\lambda|x) = \frac{\lambda^{n+b-1} e^{-\lambda \sum_{i=1}^{n} x_{u(i)}^2 + a}}{J_0} \int_{0}^{\infty} \alpha^{n+d-1} e^{-ca} \frac{\prod_{i=1}^{n} \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a-1}}{\prod_{i=1}^{n-1} \left[ 1 - \left( 1 - e^{-\lambda x_{u(i)}^2} \right)^{a} \right]} d\alpha. \quad (4.17)
$$

Next we will consider what loss functions to use with the marginal posteriors.
4.3.1 Bayes Estimation Under the General Entropy Loss Function

Following Calabria and Pulcini (1996), the Bayes estimators for the parameters $\alpha$ and $\lambda$ for the probability density function (4.1) under the general entropy loss function may be defined as
\[
\hat{\alpha}_{BGE} = \left[ E(\alpha^{-p}) \right]^{-1/p} \tag{4.18}
\]
and
\[
\hat{\lambda}_{BGE} = \left[ E(\lambda^{-p}) \right]^{-1/p}, \tag{4.19}
\]
respectively, provided that $E(\alpha^{-p})$ and $E(\lambda^{-p})$ exist and are finite. These estimators can be expressed as
\[
\hat{\alpha}_{BGE} = \left( \frac{J_\alpha'}{J_0} \right)^{-1/p}
\]
and
\[
\hat{\lambda}_{BGE} = \left( \frac{J_\lambda'}{J_0} \right)^{-1/p},
\]
where
\[
J_\alpha' = \int_0^\infty \int_0^\infty \alpha^{n+d-p-1} \lambda^{n+b-1} e^{-\alpha} \frac{\lambda^{2} \left( \sum_{i=1}^{n} x_{i(i)}^2 + a \right) \prod_{i=1}^{n} \left( 1 - e^{-\lambda x_{i(i)}^2} \right)^{\alpha-1}}{\prod_{i=1}^{n} \left[ 1 - \left( 1 - e^{-\lambda x_{i(i)}^2} \right)^\alpha \right]} \, d\lambda \, d\alpha
\]
and
\[
J_\lambda' = \int_0^\infty \int_0^\infty \alpha^{n+d-1} \lambda^{n+b-p-1} e^{-\alpha} \frac{\lambda^{2} \left( \sum_{i=1}^{n} x_{i(i)}^2 + a \right) \prod_{i=1}^{n} \left( 1 - e^{-\lambda x_{i(i)}^2} \right)^{\alpha-1}}{\prod_{i=1}^{n} \left[ 1 - \left( 1 - e^{-\lambda x_{i(i)}^2} \right)^\alpha \right]} \, d\alpha \, d\lambda. \tag{4.21}
\]
All the above double integrals have no closed form. Therefore, we will implement a Markov Chain Monte Carlo (MCMC) algorithm, the Metropolis–Hastings algorithm, to sample from the posterior distributions of $\alpha$ and $\lambda$ in order to compute the Bayes estimates. For details on the Metropolis–Hastings algorithm, see Metropolis et al. (1953).

4.3.2 Bayes Estimation Under the Squared-Error Loss Function

The Bayes estimators for parameters $\alpha$ and $\lambda$ of the GR distribution under the squared error loss function may be defined as
\[
\hat{\alpha}_{SEL} = E(\alpha|x) = \int_0^\infty \alpha \, \pi_1(\alpha|x) \, d\alpha
\]
and

\[ \hat{\lambda}_{SEL} = E(\lambda|x) = \int_0^\infty \lambda \pi_2(\lambda|x) d\lambda, \]

respectively. We note that this corresponds to the general entropy loss function with \( p = -1 \). These estimators can be expressed as

\[ \hat{\alpha}_{SEL} = \frac{J_\alpha}{J_0} \quad \text{and} \quad \hat{\lambda}_{SEL} = \frac{J_\lambda}{J_0}, \tag{4.22} \]

where

\[ J_\alpha = \int_0^\infty \int_0^\infty \alpha^{n+d} \lambda^{n+k-1} e^{-\alpha} \frac{e^{-\lambda(\sum_{i=1}^n x_{u(i)}^2 + a)}}{\prod_{i=1}^{n-1}\left[1 - \left(1 - e^{-\lambda x_{u(i)}^2}\right)\right]} d\lambda d\alpha, \tag{4.23} \]

and

\[ J_\lambda = \int_0^\infty \int_0^\infty \alpha^{n+d-1} \lambda^{n+d} e^{-\alpha} \frac{e^{-\lambda(\sum_{i=1}^n x_{u(i)}^2 + a)}}{\prod_{i=1}^{n-1}\left[1 - \left(1 - e^{-\lambda x_{u(i)}^2}\right)\right]} d\alpha d\lambda. \tag{4.24} \]

It may be noted that equations (4.23) and (4.24) cannot be solved in closed form. We must use the Metropolis–Hastings algorithm to compute the estimates.

### 4.3.3 Two-Sided Bayes Probability Intervals

The Bayesian method to interval estimation is much more direct than the frequentist method based on confidence intervals. Once the marginal posterior distribution of \( \alpha \) has been obtained, a symmetric 100(1 - \( \gamma \))% two-sided Bayes probability interval estimate of \( \alpha \), denoted by \([\alpha_L, \alpha_U]\), can be obtained by solving the two equations (see Martz and Waller (1982), pages 208–209)

\[ \int_0^{\alpha_L} \pi_1(w|x)dw = \frac{\gamma}{2} \tag{4.25} \]

\[ \int_{\alpha_U}^c \pi_1(w|x)dw = \frac{\gamma}{2} \tag{4.26} \]

for the limits \( \alpha_L \) and \( \alpha_U \).

Similarly, a symmetric 100(1 - \( \gamma \))% two-sided Bayes probability interval estimate of \( \lambda \), denoted by \([\lambda_L, \lambda_U]\), can be obtained by solving the two equations

\[ \int_0^{\lambda_L} \pi_2(w|x)dw = \frac{\gamma}{2} \tag{4.27} \]

\[ \int_{\lambda_U}^\infty \pi_2(w|x)dw = \frac{\gamma}{2} \tag{4.28} \]

for the limits \( \lambda_L \) and \( \lambda_U \). The integrals cannot be found in closed form.
4.3.4 Empirical Bayes Estimation

In the preceding sections, we have assumed that the hyper-parameters $a, b, c, \text{ and } d$ are known. Empirical Bayes estimation addresses the question of estimating the hyper-parameters from existing data. When the current sample is observed, assume that $m$ past samples $X_{j,U(1)}, X_{j,U(2)}, \ldots, X_{j,U(m)}$, for $j = 1, 2, \ldots, m$, are available. Each sample is assumed to be an upper record sample of size $n$ from a GR($\alpha, \lambda$) distribution. For each sample $j$, let $\hat{\alpha}_j$ and $\hat{\lambda}_j$ be the maximum likelihood estimates for $\alpha$ and $\lambda$, respectively, based on sample $j$, which are obtained from (4.7). We then calculate the mean and variance of the maximum likelihood estimates for each of the $j$ samples, equate these to the mean and variance of the gamma prior distribution, and solve for the hyper-parameters. We can find $\hat{a}$ and $\hat{b}$, estimates for $a$ and $b$, by solving

\[
\frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j = \frac{b}{a}
\]

and

\[
\frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\alpha}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 = \frac{b}{a^2}.
\]

We can find $\hat{c}$ and $\hat{d}$, estimates for $c$ and $d$, by solving

\[
\frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j = \frac{d}{c}
\]

and

\[
\frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\lambda}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 = \frac{d}{c^2}.
\]

Solving the above equations yields the estimates for the hyper-parameters

\[
\hat{a} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right) / \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\alpha}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 \right)
\]

and

\[
\hat{b} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 / \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\alpha}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j \right)^2 \right)
\]

for the prior distribution for $\alpha$. Similarly, estimates for the hyper-parameters for the prior distribution for $\lambda$ can be found as

\[
\hat{c} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right) / \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\lambda}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 \right),
\]

and

\[
\hat{d} = \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 / \left( \frac{1}{m-1} \sum_{j=1}^{m} \left( \hat{\lambda}_j - \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j \right)^2 \right).
\]
The empirical Bayes estimators of $\alpha$ and $\lambda$ are found by substituting $\hat{a}$, $\hat{b}$, $\hat{c}$, and $\hat{d}$ into (4.16) and (4.17) and proceeding as before.

4.4 Prediction of Future Records

In this section, we consider the problem of predicting future record values from a sample of $n$ upper records.

4.4.1 Non-Bayesian Prediction

Suppose that we observe the first $n$ upper record values from a population with probability density function $f(x; \theta)$. Our aim is to predict $z = X_{U(m)}$, for some $m > n$, having observed records $X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}$. The joint predictive likelihood function of $z = X_{U(m)}$, and the possibly vector valued parameter $\theta$ is given by Basak and Balakrishnan (2003) as

$$L(z, \theta, X) = \prod_{i=1}^{n} h(X_{U(i)}, \theta) \frac{H(z; \theta) - H(X_{U(n)}; \theta)}{\Gamma(m-n)} f(z; \theta).$$

For the GR distribution, $h(z)$ and $f(z)$ are given in (4.4) and (4.1), respectively, and the cumulative hazard function $H(z)$ can be found as

$$H(z; \theta) = -\ln(1 - F(z; \theta)).$$

Thus, the predictive likelihood function for the GR distribution is

$$L(z, \alpha, \lambda) = \prod_{i=1}^{n} \left[ \frac{2x_{U(i)}\alpha e^{-\lambda x_{U(i)}^2} \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha-1}}{1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha}} \right]^{m-n-1} \times \frac{-\log\left[1 - \left(1 - e^{-\lambda z^2}\right)^{\alpha} + \log\left[1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha}\right]\right]}{\Gamma(m-n)} \times 2\alpha\lambda z e^{-\lambda z^2} \left(1 - e^{-\lambda z^2}\right)^{\alpha-1}. \quad (4.29)$$

The log predictive likelihood function is

$$\log L(z, \alpha, \lambda) = (n+1) \log \alpha + (n+1) \log \lambda + \log z - \lambda z^2 + (\alpha - 1) \log \left(1 - e^{-\lambda z^2}\right) \quad (4.30)$$

$$+ (m-n-1) \log \left[ -\log\left[1 - \left(1 - e^{-\lambda z^2}\right)^{\alpha}\right] + \log\left[1 - \left(1 - e^{-\lambda x_{U(n)}^2}\right)^{\alpha}\right]\right]$$

$$- \log \Gamma(m-n) + \sum_{i=1}^{n} \log x_{U(i)} - \lambda \sum_{i=1}^{n} x_{U(i)}^2 + (\alpha - 1) \sum_{i=1}^{n} \log \left(1 - e^{-\lambda x_{U(i)}^2}\right)$$

$$- \sum_{i=1}^{n} \log \left[1 - \left(1 - e^{-\lambda x_{U(i)}^2}\right)^{\alpha}\right].$$
The normal equations are as follows.

\[
0 = \frac{\partial \log L}{\partial \lambda} = \frac{n + 1}{\lambda} - z^2 + (\alpha - 1)\frac{z^2}{e^{\lambda z^2} - 1} - \sum_{i=1}^{n} x_{U(i)}^2
\]

\[
+ (\alpha - 1)\sum_{i=1}^{n} \frac{x_{U(i)}^2}{e^{\lambda x_{U(i)}^2} - 1} - \sum_{i=1}^{n} \frac{\alpha x_{U(i)}^2 (1 - e^{-\lambda x_{U(i)}^2})^\alpha}{(e^{\lambda x_{U(i)}^2} - 1)((1 - e^{-\lambda x_{U(i)}^2})^\alpha - 1)}
\]

\[
+ (m - n - 1) \left[ \frac{\alpha x_{U(i)}^2 e^{-\lambda z^2} (1 - e^{-\lambda z^2})^{\alpha - 1}}{1 - (1 - e^{-\lambda z^2})^\alpha} - \frac{\alpha x_{U(i)}^2 e^{-\lambda z^2} (1 - e^{-\lambda z^2})^{\alpha - 1}}{1 - (1 - e^{-\lambda z^2})^\alpha} \right]
\]

\[
/ \left[ -\log \left( 1 - \left( 1 - e^{-\lambda z^2} \right)^\alpha \right) + \log \left( 1 - \left( 1 - e^{-\lambda z^2} \right)^\alpha \right) \right]
\]

The normal equations cannot be solved in closed form. Again, we must use an iterative algorithm to solve them numerically.

### 4.4.2 Bayesian Prediction

Predicting future record values is a problem of notable interest. For example, while studying record temperatures, rainfalls or snowfalls, we would be interested in predicting the degree of temperature or amount of rainfall or snowfall in the future when the present record will be broken (see Arnold et al. (1998)). Prediction of future records has been studied by a number of statisticians (see Berred (1998), Dunsmore (1983), Ahsanullah (1980), and Arnold et al. (1998)).

Suppose that we have \( n \) upper records, \( X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)} \), from a GR distribution. Based on such a record sample, we are interested in obtaining Bayesian prediction intervals for the next upper record \( X_{U(m)} \), for some \( m > n \) with a certain
The conditional probability density function of \( z = X_{U(m)} \) for a given \( y = X_{U(n)} \) is given (see Ahsanullah (1995)) by

\[
f(z|y; \alpha, \lambda) = \frac{\log S(y; \alpha, \lambda) - \log S(z; \alpha, \lambda) \cdot f(z; \alpha, \lambda)}{S(y; \alpha, \lambda)} f(z; \alpha, \lambda) \quad z > y.
\]

For the GR distribution, this becomes

\[
f(z|y; \alpha, \lambda) = \left[ \log \left( 1 - \left( 1 - e^{-\lambda y^2} \right)^\alpha \right) \right] m-n-1
\]

\[
\times \frac{2z\alpha e^{-\lambda z^2} \left( 1 - e^{-\lambda z^2} \right)^{\alpha-1}}{1 - (1 - e^{-\lambda y^2})^\alpha}
\]

\[
= \left[ \log \left( 1 - \left( 1 - e^{-\lambda y^2} \right)^\alpha \right) \right] m-n-1
\]

\[
× \frac{2z\alpha e^{-\lambda z^2} \left( 1 - e^{-\lambda z^2} \right)^{\alpha-1}}{1 - (1 - e^{-\lambda y^2})^\alpha} \Gamma(m-n) \ln \left( \frac{1 - \left( 1 - e^{-\lambda y^2} \right)^\alpha}{1 - (1 - e^{-\lambda z^2})^\alpha} \right) \]

\[
\forall z > y.
\]

As we know that future record values satisfy the Markovian (memoryless) property, the future upper record \( z = X_{U(m)} \) given the set of the first \( n \) upper records \( X = \{X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}\} \) depends only on the current upper record \( y = X_{U(n)} \). Therefore, the conditional probability density function of \( z \) given \( x \) is the same as the conditional probability density function of \( z \) given \( y \). The predictive probability density function of \( z \) given \( x \) is

\[
f^*(z|x) = \int_0^\infty \int_0^\infty f(z|y; \alpha, \lambda) \pi(\alpha, \lambda|x) \, d\alpha \, d\lambda
\]

where \( f(z|y; \alpha, \lambda) \) and \( \pi(\alpha, \lambda|x) \) are given, respectively, by (4.31) and (4.15). The predictive limits of the 100(1 - \( \gamma \))% two-sided interval of the future upper record \( z \) can be obtained by solving the equations

\[
\int_y^{z_L} f^*(z|x) \, dz = \frac{\gamma}{2}
\]

and

\[
\int_{z_U}^\infty f^*(z|x) \, dz = \frac{\gamma}{2}
\]

with respect to the lower and upper limits \( z_L \) and \( z_U \).

### 4.4.3 Conditional Median Prediction

Next we consider the conditional median prediction of future record values. Suppose that we have \( n \) upper records \( X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)} \) from a GR(\( \alpha, \lambda \)) distribution and we are interested in predicting \( z = X_{U(m)} \), the \( m \)th upper record, for some
m > n. The conditional median predictor is found as the median of the conditional
distribution of \( z \) given \( y \). In the case where \( \alpha \) and \( \lambda \) are known, this is simply the
median of (4.31). When \( \alpha \) and \( \lambda \) are unknown, this is found as the median of (4.32).

The conditional median predictor cannot be found in closed form. However, if
a sample can be generated from equation (4.32) using the Metropolis–Hastings al-
gorithm, then the conditional median predictor can be found as the median of the
sample.

4.5 Numerical Results

Here we present the results of a simulation study, along with results pertaining to
a real data set.

4.5.1 Simulation Study

This simulation study was designed to compare the performance of the maximum
likelihood estimators and Bayes estimators for the GR distribution based on record
samples. For each simulation, we generate record samples from a GR distribution with
parameters \( \alpha = 1 \) and \( \lambda = 2 \). We consider sample sizes \( n = 5, 7, 10 \). We report average
bias and mean squared error (MSE) over 1000 replications for maximum likelihood
estimators. We also report the average interval length and coverage probability for
approximate confidence intervals. We use the methods presented in Section 4.2.
Because the maximum likelihood estimators cannot be found in closed form, we use
the BFGS method implemented in the \texttt{R} package \texttt{maxLik} to compute the estimates.

The Bayes estimators cannot be found in closed form. Thus, we must resort to us-
ing a Markov Chain Monte Carlo (MCMC) method to compute Bayes estimators. We
implement the Metropolis–Hastings (M-H) algorithm, a powerful MCMC algorithm,
to sample from the posterior distributions of \( \alpha \) and \( \lambda \). We use an informative prior
distribution for both \( \alpha \) and \( \lambda \), with hyperparameters \( a = c = 4 \) and \( b = d = 1 \). We
compute the Bayes estimates under the general entropy loss function for \( p = -2, -1, 1 \).
We note that \( p = -2 \) corresponds to the precautionary loss function, \( p = -1 \) cor-
responds to the squared-error loss function, and \( p = 1 \) corresponds to the weighted
squared-error loss function. These include both symmetric and asymmetric loss func-
tions. The MH algorithm is implemented using the chi-square distribution as the
proposal distribution. We generate 5000 samples after 5000 burn-in samples.

To evaluate the convergence of the algorithm, we use the Gelman–Rubin diag-
nostics. Computations are made using the \texttt{R} package \texttt{coda}. We run three different
Markov chains for each parameter. The potential scale reduction factor (psrf) for both \( \alpha \) and \( \lambda \) is 1. Therefore, we can conclude that the algorithm converges quite well.

Further evidence for the convergence of the MH algorithm can be obtained graphically. Trace plots and density plots for each parameter are plotted here. The trace plots indicate that the algorithm converged quite well. The density plots, as one would expect, show a nice mixture distribution.

Table 4.1 includes the average bias and MSE for the maximum likelihood and Bayes estimators for \( \alpha \) and \( \lambda \). Bias is listed first, and MSE follows in parentheses.

Here, the abbreviations MLE, SEL, WSEL, and PL refer to maximum likelihood estimators, Bayes estimators under the squared-error loss function, Bayes estimators under the weighted squared-error loss function, and Bayes estimators under the precautionary loss function, respectively.
\[ \hat{\alpha}_{MLE}, \hat{\alpha}_{SEL}, \hat{\alpha}_{WSEL}, \hat{\alpha}_{PL}, \hat{\lambda}_{MLE}, \hat{\lambda}_{SEL}, \hat{\lambda}_{WSEL}, \hat{\lambda}_{PL} \]

\begin{table}[h]
\centering
\begin{tabular}{cccccccc}
\hline
\( n \) & \( \hat{\alpha}_{MLE} \) & \( \hat{\alpha}_{SEL} \) & \( \hat{\alpha}_{WSEL} \) & \( \hat{\alpha}_{PL} \) & \( \hat{\lambda}_{MLE} \) & \( \hat{\lambda}_{SEL} \) & \( \hat{\lambda}_{WSEL} \) & \( \hat{\lambda}_{PL} \) \\
\hline
5 & 0.65 & 0.69 & 0.84 & 0.62 & 0.83 & 1.27 & 1.52 & 1.16 \\
    & (1.87) & (0.48) & (0.71) & (0.39) & (2.53) & (1.62) & (2.30) & (1.37) \\
7 & 0.64 & 0.66 & 0.84 & 0.58 & 0.78 & 0.96 & 1.20 & 0.86 \\
    & (1.78) & (0.44) & (0.70) & (0.34) & (2.01) & (0.95) & (1.45) & (0.76) \\
10 & 0.51 & 0.64 & 0.83 & 0.55 & 0.75 & 0.61 & 0.83 & 0.51 \\
    & (1.44) & (0.43) & (0.68) & (0.31) & (1.03) & (0.43) & (0.73) & (0.32) \\
\hline
\end{tabular}
\caption{Average bias and MSE of estimators for \( \alpha \) and \( \lambda \)}
\end{table}

Table 4.2 contains the average interval length and coverage probability of simulated 95\% approximate confidence intervals.

\begin{table}[h]
\centering
\begin{tabular}{c|cc}
\hline
\( n \) & \( CI_{\alpha} \) & \( CI_{\lambda} \) \\
\hline
5 & 6.25 & 4.13 \\
    & (0.959) & (0.910) \\
7 & 6.19 & 3.53 \\
    & (0.947) & (0.939) \\
10 & 5.64 & 3.28 \\
    & (0.948) & (0.935) \\
\hline
\end{tabular}
\caption{Average interval length and coverage probability for confidence intervals}
\end{table}

Finally, we consider simulation results pertaining to empirical Bayes estimation. We generate 100 prior samples from a GR distribution with \( \alpha = 1 \) and \( \lambda = 2 \) to calculate estimates for the hyperparameters. Then, we generate 1000 informative samples and compute the empirical Bayes estimators using the methods described in section 3.3.5. Table 4.3 contains the average bias and MSE of the empirical Bayes estimators under three different loss functions.

\begin{table}[h]
\centering
\begin{tabular}{cccccccc}
\hline
\( n \) & \( \hat{\alpha}_{SEL} \) & \( \hat{\alpha}_{WSEL} \) & \( \hat{\alpha}_{PL} \) & \( \hat{\lambda}_{SEL} \) & \( \hat{\lambda}_{WSEL} \) & \( \hat{\lambda}_{PL} \) \\
\hline
5 & 0.42 & 0.92 & 0.11 & 0.34 & 0.79 & 0.24 \\
    & (0.25) & (0.84) & (0.21) & (0.28) & (0.71) & (0.31) \\
7 & 0.40 & 0.91 & 0.10 & 0.22 & 0.60 & 0.17 \\
    & (0.24) & (0.83) & (0.20) & (0.25) & (0.47) & (0.27) \\
10 & 0.37 & 0.86 & 0.047 & 0.064 & 0.25 & 0.019 \\
    & (0.19) & (0.75) & (0.12) & (0.14) & (0.15) & (0.17) \\
\hline
\end{tabular}
\caption{Average bias and MSE of empirical Bayes estimators for \( \alpha \) and \( \lambda \)}
\end{table}

The abbreviations SEL, WSEL, and PL are as before.

### 4.5.2 Data Analysis

We apply the results of this paper to one real data set. The data is presented in Lawless (1982), page 319. It consists of the survival times in days of 16 lung cancer patients:

\begin{verbatim}
6.96 9.30 6.96 7.24 9.30 4.90 8.42 6.05 10.18 6.82 8.58 7.77 11.94
\end{verbatim}
From this data set, we extract a sample of upper records: 6.96, 9.30, 10.18, 11.94, 12.94.

First, we compute the maximum likelihood and Bayes estimators of the parameters $\alpha$ and $\lambda$ based on the data. The estimates are found in Table 4.4. The two-sided Bayes probability interval (TBPI) for $\alpha$ is $(6.54, 10.08)$. The TBPI for $\lambda$ is $(0.016, 0.060)$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>SEL</th>
<th>WSEL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>10.82</td>
<td>8.92</td>
<td>8.73</td>
<td>9.00</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.045</td>
<td>0.034</td>
<td>0.028</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Table 4.4: Estimates for $\alpha$ and $\lambda$ based on the data

After fitting the GR distribution to the data, we must evaluate how good of a fit it is. Plots of the empirical and fitted survivor functions are plotted here. The first uses maximum likelihood estimates and the second uses Bayes estimates under the precautionary loss function. The precautionary loss function appears to yield a better fit than the other loss functions we consider.

![Empirical and fitted survivor function using maximum likelihood estimators](image1.png)

![Empirical and fitted survivor functions using Bayes estimators under the PL function](image2.png)

Figure 4.7: Empirical and fitted survivor function using maximum likelihood estimators

Figure 4.8: Empirical and fitted survivor functions using Bayes estimators under the PL function

One can see that, while not an ideal fit, the GR distribution is an acceptable model for this data.

To demonstrate prediction methods, we generate a synthetic sample from a GR distribution with $\alpha = 1$ and $\lambda = 2$. We simulate 10 record values and then use the methods described here to predict the 11th record value. The synthetic sample of records is 0.2346 0.6754 0.9442 1.033 1.102 1.263 1.621 1.880 1.948 1.981. To compute the predictors, we use the MH algorithm to sample from (4.32). The
conditional median predictor of the next upper record is found as the median of the
simulated sample. For our data set, this turns out to be $z_{CM} = 2.095$. A 95% predictive interval for the next record value is $(0.8490, 4.0666)$. For the Bayesian computation, we again use independent gamma priors with the hyperparameters $a = c = 15$ and $b = d = 0.25$.

4.6 Conclusion

The generalized Rayleigh distribution was introduced as a model for lifetime data. We have provided a number of statistical methods for this distribution, including both frequentist and Bayesian methods for estimating the parameters based on upper record values, and methods for predicting future record values. Such methods are essential for analysts to adequately use this distribution as a model. Simulation and data analysis help to compare the different methods we have presented, including Bayesian methods under multiple loss functions.
Bibliography


