People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Mohamed Khider University, Biskra

Faculty of Exact Sciences and Natural and Life Sciences

DEPARTEMENT OF MATHEMATICS

Master dissertation presented with a view to obtaining the Diploma :

MASTER of Mathematics

Option: **Statistic**

By

Gasmi Rayane

Title :

Application of records statistics

Examination Committee Members:

Pr.	**YAHIA Djabrane**	UMKB	President
Dr.	**ZOUAOUI Nour El Houda**	UMKB	Supervisor
Dr.	**KHEIREDDINE Souraya**	UMKB	Examinator

11/06/2024

Dedication

I express my deepest dedication in this humble work to:

My beloved **Mother**, whose prayers, love, guidance, and unwavering wisdom have been the foundation of my journey. Her strength, resilience, and compassion inspire me every day. Thank you, **Mom**, for your endless support and for being the light that guides me in every endeavor.

To my dear **Father**, whose unwavering support, steadfast encouragement, sacrifices, dedication, and tireless efforts have made me feel comfort and security.

To my dear brothers: **Sif Eddine, Imad Eddine, Ala Eddine, Zin Eddine**, who encouraged and inspired me in conducting this study and gave me strength and hope. They provided me with a great sense of enthusiasm and perseverance to continue this.

To my whole family, especially my uncle **Ammar**.

To all my friends, especially **Ikram** and **Ibtissem**, who continually provide their moral, spiritual, and emotional support.

Gasmi Rayane

.

Acknowledgment

First of all, I would like to thank the Almighty God who gave me strength, wisdom, mind, power of thinking, competence, and skills, and gave me good health while doing this.

Dear supervisor **DR.Zouaoui Nour El Houda**, Your leadership has been crucial in overcoming challenges and achieving goals, and your guidance has enhanced my skills and inspired me to strive for excellence in everything I do and to enrich the quality of this thesis. I truly appreciate the time and effort you've dedicated to my professional growth. Your patience, wisdom, and encouragement have made a significant difference in my journey. Thank you for your steadfast support.

With great honor, I would like to express my thanks and gratitude to the members of the jury **Pr.DJABRANE Yahia** for his dedication to teaching, and I am truly grateful for the time and effort you've put in to help me succeed." and **DR.KHEIREDDINE Souraya** to have examined and evaluated my work.

To all those who have contributed to the realization of this memory :

Thank's

Abstract

Record statistics is a broad term; it can refer to various types of statistical information related to records. In a general sense, record statistics could involve analyzing and presenting data related to records. The subject of the study is to introduce the concept of record as applied to probability distribution function and present some their characterization results based on the records values. In addition, we discuss the problem of estimation of the location and scale parameters in Best Linear Unbiased Estimator (BLUE) and how applied it in both different types of records with differents distributions functions.

Notations and symbols

iid	Independent and identically distributed
cdf	Cumulative distribution function note $F(\cdot)$
pdf	Probability density function also density function note $f(\cdot)$
$h(\cdot)$	Hazard or failure rate function
$H(\cdot)$	Cumulative hazard function
$U(n)$	Upper record times
U_n^X	Upper record values
$U_{n,k}^X$	nth value of upper k-records
$f_{U_n^X}(x)$	Probability density function of nth upper record values
$f_{U_m^X,U_n^X}(x,y)$	The joint probability density function of mth and nth upper records
$f_{U_n^X\|U_m^X}(y\|x)$	The conditional probability density function of nth upper given mth upper record
$f_{U_{n,k}^X}(x)$	Probability density function of nth value of upper k-records
$f_{U_{m,k}^X,U_{n,k}^X}(x,y)$	The joint probability density function of mth and nth upper k-records
$f_{U_{n,k}^X\|U_{m,k}^X}(y\|x)$	The conditional probability density function of nth upper given mth upper k-records
$L(n)$	Lower record times

L_n^X	Lower record values
$L_{n,k}^X$	nth value of lower k-records
$f_{L_n^X}(x)$	Probability density function of nth lower record values
$f_{L_m^X,L_n^X}(x,y)$	The joint probability density function of mth and nth lower records
$f_{L_n^X \mid L_m^X}(y \mid x)$	The conditional probability density function of nth lower given mth lower record
$f_{L_{n,k}^X}(x)$	Probability density function of nth value of lower k-records
$f_{L_{m,k}^X,L_{n,k}^X}(x,y)$	The joint probability density function of mth and nth lower k-records
$f_{L_{n,k}^X \mid L_{m,k}^X}(y \mid x)$	The conditional probability density function of nth lower given mth lower k-record
X_i	Random variables with $i = 1, 2, \ldots$
et al	and others
$E(\cdot)$	Mathematical esperance
$Var(\cdot)$	Mathematical variance
BLUE	Best Linear Unbiased Estimation
MLM	Maximum Likelihood Method

Contents

List of Figures

List of Tables

Introduction

The word "record" is derived from the Latin verb "recordari" - to remember, and everyone has a basic understanding of what a recorded event entails. People have learned that a new record is something important that will be remembered. Therefore, many people feel impressed by achieving and then breaking new and exciting records, making the "Guinness World Records" book itself a record holder as the best-selling copyrighted book in history.

The study of record values has been undertaken by a relatively small but highly talented group of individuals. The statistical study of record values began with Chandler[[8]], where he formulated the theory of record values as a model for successive extremes in a sequence of independently and identically distributed random variables. Stuart was a pioneer in discovering the quantitative statistical applications of record keeping in inference. Barton, Mallows, and F.N.David were fascinated by the combinatorial aspects of record sequences. While Dwass established the independence of record indicators, Rényi developed some of the initial limit theorems. Resnick (1973) and Shorrock (1973) documented the asymptotic theory of records. The theory of record values and their distributional properties have been extensively studied in the literature; for example, see Ahsanullah (1995, 1997), Nagaraga[[10]], Arnold and Balakrishnan [[6]], Balakrishnan and Ahsanullah (1994), Raqab[[2]], Bieniek and Szynal (2002), Saran and Singh (2008). Many books on order statistics and related processes have chapters on record values. Many of them emphasize the close parallel between record value theory and the theory of sample maxima.

In life, during experimental testing, an experimenter may find it necessary to terminate the experiment prematurely if a certain number of units fail, rather than waiting for all units to fail. This approach is efficient in terms of time and cost, and such observations are referred to as type II controlled sampling. Therefore, we use estimation, which is considered the cornerstone of the decision-making, planning, and resource allocation process. It involves the process of approximating or predicting values, quantities, or outcomes based on available information, experience, and logical reasoning.

In Chapter 2 we see the Best linear unbiased estimation (BLUE) wich is a widely utilized method for estimating population parameters such as scale or location, particularly when the available sample is either complete or Type-II censored. Key references for this method include works by Rao (1973), David [[10]], Balakrishnan and Cohen (1991), and Arnold, Balakrishnan, and Nagaraja (1992). It's important to note that the concept of "best" is well-defined when estimating a single parameter, typically meaning the minimum variance. However, when estimating multiple parameters, there is no single definition of "best." Therefore, two criteria, namely trace and determinant, are commonly considered. These criteria are based on the variance-covariance matrix of parameter estimators, with "best" being defined as the minimum trace or determinant of this matrix. Consequently, estimators are categorized as either "trace-efficient" or "determinant-efficient." It's worth noting that only a limited number of distributions allow for the explicit derivation of BLUE for any distribution. In certain cases, such as the exponential distribution, BLUE coincides with maximum likelihood estimators. Moreover, BLUE estimators are not only explicit linear estimators but also highly efficient ones. In fact, they are asymptotically fully efficient compared to maximum likelihood estimators. An interesting finding is that, for the two-parameter exponential distribution, trace-efficient and determinant-efficient estimators align with BLUE estimators.

Chapter 1

Record statistics

In this chapter we start with a definition of upper-record and lower-record in the section 1.1, then we introduce record time and value with their characteristics for continuous random variables in the section 1.2 of K-record statistics, in section 1.3 we present its special case when$(k = 1)$, which called ordinary or classical record statistics. At the end, we extract the density of the lower and upper records from specific continuous distribution in the section 1.4.

1.1 Introduction

Record values are the local maxima or minima of a sequence of random variables, records occur naturally in various fields of study such as medicine, sports, and engineering among others.

We will examine scenarios where the record values, such as the consecutive highest insurance claims in non-life insurance, peak water levels, or extreme temperatures, are considered as "outliers." Consequently, the second or third largest values become particularly noteworthy. Examples could include insurance claims in certain non-life insurance contexts. Dziubdziela and Kopocinski (1976) [[11]]introduced a model for

K-th record values, with K representing a positive integer. Record values can be used to characterize various distributions.

Let X_1,X_2,... be a sequence of independent observations on a random variable X having the cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$. An observation X_j is called a record value (or simply a record) if its value is smaller than or greater than all the preceding observations. If it is smaller than all the preceding observations, it is called lower record. Hence X_j is a record, if X_j<X_i for every $i < j$. An analogous definition can be given for the upper record also. It is interesting to note here that the first observation is always a lower record as well as upper record.

An observation is called is called a record if it is greater than (or less than) all the preceding observations. Frequently, our attention is drawn towards witnessing fresh achievements and documenting them, such as Olympic or global records. These record values serve a purpose in reliability theory. Additionally, these statistical metrics are intricately linked to the timing of certain non-uniform Poisson processes utilized in models concerning shocks.

1.2 K-Record Statistics

Chandler [[8]]introduced the record values and the record times. Feller (1966) provided some examples of record values concerning issues related to gambling.

1.2.1 Definition of k-Record Values and k-Record Times

Suppose that $X_1, X_2, ..., X_n$ is a sequence of iid random variables with *cdf* F and *pdf* f . Let$X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$, be the order statistics of X_1, ..., X_n.

1/k-Record Values

Definition 1.2.1 *Record values refer to the maximum or minimum observations within a dataset, representing the extreme points or peaks of a particular variable, let's define the sequences of upper k-record values, for a fixed integer $k \geq 1$ and $n \geq 2$ as follows:*

$$U_{n,k}^{X} = X_{U_{n,K}-k+1:Un,k}$$

and the sequences of lower k-record values are as follows:

$$L_{n,k}^{X} = X_{L_{n,K}-k+1:L_{n,k}}$$

2/k-Record Times

Definition 1.2.2 *Record times denote the moments or instances when these extreme values occur, indicating the times at which the highest or lowest observations are recorded within a given dataset or timeframe. let's clarify the sequences of upper k-record times, for $(n \geq 1)$ as follows:*

$$U_{n,k} = min\left\{i : i > U_{n-1,k}, Xi > X_{U_{n-1,K}-k+1:U_{n-1,k}}\right\}, U_{1,k} = k$$

and the sequences of lower k-record times are as follows:

$$L_{n,k} = min\left\{i : i > L_{n-1,k}, Xi < X_{L_{n-1,K}-k+1:L_{n-1,k}}\right\}, L_{1,k} = k$$

1.2.2 K-Record Probability Density Function

The model of k-record values is proposed at first by (Dziubdziela and Kopocinski[[11]])

Lemma 1.2.1 *The hazard (function also known as the failure rate, hazard rate, or force of function) $h(x)$ is the ratio of the probability density function f to the survival* *function* $1 - F(x)$, *given by:*

$$h(x) = \frac{f(x)}{1 - F(x)}.$$

The cumulative hazard function

$$H(x) = -log[1 - F(x)]. \tag{1.1}$$

Theorem 1.2.1 *The pdf of* $U_{n,k}^X$ *is obtained to be:*

$$f_{U_{n,k}^X}(x) = k^n \frac{[H\,(x)]^{n-1}}{(n-1)!}[1 - F(x)]^{k-1} f(x), -\infty < x < \infty. \tag{1.2}$$

Proof. By induction ,the densities$f_{U_n^X}(x_1, ...x_k), n = 1, 2, ...,$ satisfy the equations

$$f_{U_n^X}(x_1, ...x_k) = \begin{cases} k! g_{U_n^X}(x_1) f(x_1) f(x_2) ... f(x_k), x_1 < x_2 < ... < x_k \\ 0, otherwise \end{cases}$$

where

$$g_{U_1^X}(x) = 1$$

and

$$g_{U_{n+1}^X}(x) = k \int_{-\infty}^{x} g_{U_n^X}(y) \frac{f(y)}{1 - F(y)} dy, n = 1, 2, ...$$

Then we find

$$g_{U_n^X(}x) = \frac{1}{(n-1)!}[-klog(1 - F(x))]^{n-1}, n = 1, 2, ...$$

In the end, we have

$$f_{U_{n,k}^X}(x) = \int_x^\infty \int_{x_2}^\infty \ldots \int_{x_{k-1}}^\infty \frac{k!}{(n-1)!}[-klog(1-F(x))]^{n-1} f(x)f(x_2)\ldots f(x_k)dx_k\ldots dx_2$$
$$= \frac{k}{(n-1)!}[-klog(1-F(x))]^{n-1}[1-F(x)]^{k-1} f(x).$$

■

Corollairy 1.2.1 *The cumulative distribution function of the k-record value is of the form*

$$F_{U_{n,k}^X}(x) = \int_0^{-klog(1-F(x))} \frac{u^{n-1}}{(n-1)!} e^{-u} du, n = 1, 2, \ldots$$

Definition 1.2.3 *Let X be a continuous random variable. We define*

1. *The joint pdf of $U_{m,k}^X$ and $U_{n,k}^X$,where $1 \leq m \leq n$, is given by Grudzien(1982) :*

$$f_{U_{m,k}^X, U_{n,k}^X}(x, y) = \frac{k^n}{(m-1)!(n-m-1)!}[H(x)]^{m-1}[H(y) - H(x)]^{n-m-1}[1 - F(y)]^{k-1}$$
$$h(x)f(y), -\infty < x < y < \infty. \tag{1.3}$$

2. *The conditional pdf of $U_{n,k}^X$ given $U_{m,k}^X (1 \leq m \leq n)$ can be written as:*

$$f_{U_{m,k}^X | U_{n,k}^X}(y|x) = \frac{k^{n-m}}{(n-m-1)!}[H(y) - H(x)]^{n-m-1}[1 - F(y)]^{k-1} \frac{f(y)}{[1 - F(x)]^k},$$
$$x < y. \tag{1.4}$$

3. *An analogous pdf's can be given respectively for lower k record The pdf of $L_{n,k}^X$ is given by:*

$$f_{L_{n,k}^X}(x) = k^n \frac{[-logF(x)]^{n-1}}{(n-1)!}[F(x)]^{k-1} f(x), -\infty < x < \infty. \tag{1.5}$$

4. *The joint pdf of* $L_{m,k}^X$ *and* $L_{n,k}^X$ *,where* $1 \leq m \leq n$ *, can be written as (see, Pawlas and Szynal)[[16]],[[17]]*

$$f_{L_{m,k}^X, L_{n,k}^X}(x,y) = \frac{[-logF(x)]^{m-1}}{(m-1)!} \frac{[logF(x) - logF(y)]^{n-m-1}}{(n-m-1)!}$$
$$[F(y)]^{k-1} \frac{f(x)}{F(x)} f(y), -\infty < x < y < \infty. \quad (1.6)$$

5. *The conditional pdf of* $L_{n,k}^X$ *given* $L_{m,k}^X$*,where* $1 \leq m \leq n$*, can be written as:*

$$f_{L_{n,k}^X | L_{m,k}^X}(y|x) = \frac{k^{n-m}}{(n-m-1)!} [logF(x) - logF(y)]^{n-m-1} [F(y)]^{k-1}$$
$$\frac{f(y)}{[F(x)]^k}, x < y. \quad (1.7)$$

Remark 1.2.1 *we obtain ordinary upper and lower record values when k=1, which we see in the next item.*

Example 1.2.1 *The data set available below is the pulse rates in beats per minute of a random sample of adult females to test the claim that the mean is less than 73 bpm, by using a 0.05 significance level:*

89, 71, 66, 47, 42, 86, 54, 48, 105, 51, 92, 59, 58, 36, 47, 105, 77, 74, 42, 43.

we create lower k-records from the data where $k = 2, 3, 4, 5, 6$ *when n=1,...,20 ,by using the definition we compute:*

- ***for k=2***

  1. *for n=1:*

$$L_{1,2} = 2, L_{1,2}^X = X_{L_{1,2}-2+1:L_{1,2}} = X_{1,2} = 89$$

2. *for n=2:*

$$L_{2,2} = min\left\{i : i > L_{2-1,2}, X_i < X_{L_{2-1,2}-2+1:L_{2-1,2}}\right\}$$
$$= min\left\{i : i > 2, X_i < X_{1,2}\right\}. \tag{1.8}$$

Now, we search the value of i that satisfies eq(1.8)*. Thus, we find for* ***i = 3****:*

$$L_{2,2} = min\left\{i : 3 > 2, X_3 < 89\right\},$$

$$L_{2,2} = 3, L^X_{2,2} = X_{L_{2,2}-2+1:L_{2,2}} = X_{2,3} = 71.$$

(because 89 > 71 > 66).

3. *for n=3:*

$$L_{3,2} = min\left\{i : i > L_{3-1,2}, X_i < X_{L_{3-1,2}-2+1:L_{3-1,2}}\right\}$$
$$= min\left\{i : i > 3, X_i < X_{2,3}\right\}. \tag{1.9}$$

for ***i=4****:*

$$L_{3,2} = min\left\{i : 4 > 3, X_4 < 71\right\},$$

$$L_{3,2} = 4, L^X_{23,2} = X_{L_{3,2}-2+1:L_{3,2}} = X_{3,4} = 66.$$

(because 89 > 71 > 66 > 47).

- *We continue in the same way (we stop when the value is the same with the one before it) to get the sequence of the lower 2-record which is given as follows:*

$$89, 71, 66, 47$$

As a consequence, the following table presents the lower k-record for

k = 2, 3, 4, 5, 6.

2-records	89	71	66	47
3-records	89	71	66	
4-records	89	71		
5-records	89	86	71	66
6-records	89	86	71	

Table 1.1: The Lower K-record for $k = 2, 3, 4, 5, 6$.

1.3 Ordinary Record Statistics

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables and $X_{1,n} \leq ... \leq X_{n,n}$ be the corresponding order statistics where:
$X_{1,n} = min\{X_1, X_2, ..., X_n\}$and $X_{n,n} = max\{X_1, X_2, ..., X_n\}$.

Definition 1.3.1 *the classical upper record times $U(n)$ and upper record values U_n^X as follows:*
$U(1) = 1, U(n+1) = min\{j : j \geq U(n), Xj \geq U_n^X\}, U_n^X = X_{U(n)}, n = 1, 2, ...$

Definition 1.3.2 *the sequences of lower record times $L(n)(n \geq 1)$ and lower record values L_n^X as follows :*
$L(1) = 1, L(n+1) = min\{j : j \geq L(n), Xj \leq L_n^X\}, L_n^X = X_{L(n)}, n = 1, 2, ...$

Remark 1.3.1 *Moving from lower records to upper records by replacing the original sequence of random variables $\{X_j\}$ by$\{-X_j, j > 1\}$ or if $P(X_j > 0) = 1$ by $\{1/X_j, j > 1\}, j = 1, 2, \ldots$*

1.3.1 Record Probability Density Function

In this subsection, we begin by presenting an expression of the density function of U_n^X and L_n^X. Furthermore, we introduce the joint pdf of upper and lower records, concluding with the definition of the conditional pdf of the upper and the lower records.

Theorem 1.3.1 *The pdf of* U_n^X *is obtained to be (Arnold et al.[[6]])*

$$f_{U_n^X}(x) = \frac{[H(x)]^{n-1}}{(n-1)!} f(x), -\infty < x < \infty, n = 1, 2, ... \tag{1.10}$$

Proof. To prove the theorem, we are going to consider exponential observations since this distribution has a lack of memory property then the differences between successive records will be *iid* standard exponential random variable. Let $\{X_j^*, j > 1\}$ be a sequence of $iidExp(1)$ random variables, and consequently, it follows that the nth upper record U_n^* has a gamma distribution with shape $n+1$ and rate 1.

$$U_n^* \sim Gamma(n+1, 1), n = 1, 2, ... \tag{1.11}$$

These results will be useful to obtain the distribution of the nth record corresponding to an iid sequence of random variables $\{Xj\}$ with common continuous *cdf* F. If X has a continuous *cdf* F, then the cumulative hazard function has a standard exponential distribution. we have

$$H(X) \equiv -log[1 - F(X)],$$

then

$$X \stackrel{d}{=} F^{-1}(1 - e^{-X^*})$$

$\{X_j^*\}$follows standard exponential, therefore, we have:

$$U_n \stackrel{d}{=} F^{-1}(1 - e^{-U_n^*}), n = 1, 2... \tag{1.12}$$

Repeated integration by parts can be used to justify the following expression for the survival function of U_n^*(a Gamma$(n+1,1)$ random variable):

$$P(U_n^* > x) = e^{x^*} \sum_{k=0}^{n} \frac{(x^*)^k}{(k-1)!}, x^* > 0.$$

We may then use the relation (1.12) to immediately derive the survival function of the nth record corresponding to an *iid* F sequence.

$$P(U_n > x) = 1 - F_{U_n^X}(x) = [1 - F(x)] \sum_{k=0}^{n} \frac{[-log(1 - F(x))]^k}{(k-1)!}.$$

which is equivalent to:

$$P(U_n < x) = \int_0^{-log(1-F(x))} y^n e^{-y}/(n-1)! dy$$

If F is absolutely continuous with the corresponding probability density function, we may differentiate either of the above expressions to derive the *pdf* for U_n^X. We obtain:

$$\begin{aligned} f_{U_n^X}(x) &= \frac{[H(x)]^{n-1}}{(n-1)!} f(x) \\ &= \frac{1}{(n-1)!} [-log(1 - F(x))]^{n-1} f(x). \end{aligned}$$

For more details see Arnold et al. [[6]] ,page (31) and Karlin[[13]]. ■

Theorem 1.3.2 *The joint pdf of U_m^X and U_n^X ,where $1 \leq m \leq n$, is given by (Arnold et al(1998))*[[6]]

$$f_{U_m^X, U_n^X}(x, y) = \frac{[H(x)]^{m-1}}{(m-1)!} \frac{[H(y) - H(x)]^{n-m-1}}{(n-m-1)!} h(x) f(y), -\infty < x < y < \infty. \quad (1.13)$$

Proof. To start the proof, we begin by noting that the joint pdf of a set of $Exp(1)$ records $(U_0^*, U_1^*, ..., U_n^*)$ can be readily formulated. Additionally, given that the record spacings $U_n^* - U_{n-1}^*$are *iid* $Exp(1)$. Thus, we find:

$$f_{U_0^*,U_1^*,...,U_n^*}(x_0^*, x_1^*, ..., x_n^*) = e^{x^*}, 0 < x_0^* < x_1^* < ... < x_n^*$$

Now,we apply the transformation eq(1.12) coordinatewise to obtain the joint *pdf* of the set of records $U_0, U_1, ..., U_n$ corresponding to an *iid* F sequence expressed as:

$$\begin{aligned} f_{U_0,U_1,...,U_n}(x_0, x_1, ..., x_n) &= \frac{\displaystyle\prod_{i=0}^{n} f\,(x_i)}{\displaystyle\prod_{i=1}^{n-1}[1 - F(xi)]}, \\ &= f(x_n)\prod_{i=1}^{n-1} h(xi), -\infty < x_0 < ... < x_n. \end{aligned} \tag{1.14}$$

The joint *pdf* of any pair of records U_m, U_n naturally be derived from eq(1.14) by integration.it might be more straightforward to first find the joint distribution for two $Exp(1)$ records (U_m^*, U_n^*) and then use the transformation eq(1.12) where $m < n$, since $(U_m^*, U_n^*) = (Y_1, Y_1 + Y_2)$ where $Y_1 \backsim Gamma(m+1, 1)$ and $Y_2 \backsim Gamma(n-m, 1)$ are independent, we may use a simple Jacobian argument starting from the joint *pdf* of (Y_1, Y_2) to obtain:

$$f_{U_m^*,U_n^*}(x_m^*, x_n^*) = \frac{1}{m!(n-m-1)!} x_m^{*m}(x_n^* - x_m^*)^{n-m-1} e^{-x_n^*}, 0 < x_m^* < x_n^* < \infty.$$

With eq(1.12)applied U_m^* to U_n^*, we obtain

$$f_{U_m,U_n}(x_m, x_n) = \frac{[-log(1 - F(x_m))]^m}{m!} \frac{\left[-log\left(\frac{1-F(x_n)}{1-F(x_m)}\right)\right]}{(n-m-1)!} \frac{f(x_m)h(x_n)}{1 - F(x_m)}, -\infty < x_m < x_n < \infty.$$

■

Corollairy 1.3.1 *The conditional pdf* U_n^X *given* U_m^X $(1 < n < m)$ *using eq*(1.10) *and eq*(1.13)*is as follows:*

$$f_{U_n^X|U_m^X}(y|x) = \frac{f_{U_m^x,U_n^x}(x,y)}{f_{U_m^X}(x)} = \frac{[H(y) - H(x)]^{n-m-1}}{(n-m-1)!}\frac{f(y)}{1-F(x)}, x < y. \tag{1.15}$$

Remark 1.3.2 *The survival function of the nth upper record can be represented as:*

$$1 - F_{U_n^X}(x) = [1 - F(x)]\sum_{j=0}^{n-1}\frac{[-log(1-F(x))]^j}{j!} \tag{1.16}$$

where

$$F_{U_n^X}(x) = \int_0^{-log(1-F(x))}\frac{u^{n-1}}{(n-1)!}e^{-u}du, n = 1, 2, \ldots$$

The following corollary presents the *pdf* of lower records and the joint and conditional pdf of L_m^X and L_n^X .

Corollairy 1.3.2 *We have:*

- *The pdf of* L_n^X *is given by :*

$$f_{L_n^X}(x) = \frac{[-logF(x)]^{n-1}}{(n-1)!}f(x), -\infty < x < \infty, n = 1, 2, \ldots \tag{1.17}$$

- *The joint pdf of* L_m^X *and* L_n^X *,where* $1 \le m \le n$*, can be written as:*

$$f_{L_m^X,L_n^X}(x,y) = \frac{[-logF(x)]^{m-1}}{(m-1)!}\frac{[logF(x) - logF(y)]^{n-m-1}}{(n-m-1)!}\frac{f(x)}{F(y)}f(y), \tag{1.18}$$
$$-\infty < x < y < \infty.$$

- *The conditional pdf of L_n^X given L_m^X where $1 < m < n$, can be written as:*

$$f_{L_n^X|L_m^X}(y|x) = \frac{[logF(x) - logF(y)]^{n-m-1}}{(n-m-1)!}\frac{f(y)}{F(x)}, x < y. \qquad (1.19)$$

Example 1.3.1 *Consider the following 20 observations:*

10, 10, 10, 8, 12, 5, 9, 36, 3, 2, 1, 30, 14, 20, 22, 22, 61, 4, 7, 90.

By the definition of upper record and lower record, we obtain:

upper record values:

$$10, 12, 36, 61, 90$$

lower record values:

$$10, 8, 5, 3, 2, 1$$

1.4 Record From Specific Continuous Distribution

In statistical analysis, it's often necessary to understand and visualize the behavior of random variables and their distributions. In this example, we will focus on extracting and analyzing specific record values from a continuous distribution, using R for the computation and plotting.

1.4.1 Weibull Distribution

Record values are significant in understanding extreme observations in a dataset. In continuous distributions, the upper k-record value is the k-th largest observation seen so far. We will use the Weibull distribution, a widely used continuous probability distribution, to illustrate this concept. Our objective is to:

1. Extract upper k-record values from a sample generated from the Weibull distribution.

2. Define the probability density function (pdf) for these upper k-record values.

3. Implement the theoretical concepts in R and visualize the results using a plot.

Theoretical concepts

$$
\begin{aligned}
f(x) &= \tfrac{r}{\lambda^r} (x)^{r-1} e^{-\left(\frac{x}{\lambda}\right)^r} \\
F(x) &= 1 - e^{-\left(\frac{x}{\lambda}\right)^r} \\
f_{U_{n,k}^X}(x) &= \tfrac{rk^n}{\lambda^r (n-1)!} \left(\tfrac{x}{\lambda}\right)^{r(n-1)} (x)^{r-1} \left(e^{-\left(\frac{x}{\lambda}\right)^r}\right)^k \\
f_{U_n^X}(x) &= \tfrac{r}{\lambda^r (n-1)!} \left(\tfrac{x}{\lambda}\right)^{r(n-1)} (x)^{r-1} \exp\left(-\left(\tfrac{x}{\lambda}\right)^r\right) \\
f_{L_{n,k}^X}(x) &= \tfrac{rk^n}{\lambda^r (n-1)!} \left[-log\left(1 - e^{-\left(\frac{x}{\lambda}\right)^r}\right)\right]^{n-1} \left[1 - e^{-\left(\frac{x}{\lambda}\right)^r}\right]^{k-1} (x)^{r-1} e^{-\left(\frac{x}{\lambda}\right)^r} \\
f_{L_n^X}(x) &= \tfrac{r}{\lambda^r (n-1)!} \left[-log\left(1 - e^{-\left(\frac{x}{\lambda}\right)^r}\right)\right]^{n-1} (x)^{r-1} e^{-\left(\frac{x}{\lambda}\right)^r}
\end{aligned}
$$

R implementation and plotting:

Here's the step-by-step implementation in R:

- Generate a sample from the Weibull distribution:

  We use the Weibull distribution, parameterized by shape ($r = 2$) and scale ($\lambda = 5$) as provided in the code below (See the figure 1.1).

```r
set.seed(10)
z <- rweibull(100, 5, 2)
```

```
 [1] 1.850609 2.067896 1.936578 1.636301 2.395194 2.165961 2.105373 2.108018
 [9] 1.730374 1.933633 1.687980 1.784876 2.336451 1.753217 2.010718 1.934551
[17] 2.484509 2.117748 1.966632 1.417700 1.359915 1.730926 1.521434 2.013433
[25] 1.959067 1.618646 1.413601 2.147960 1.528079 2.013073 1.820199 2.377573
[33] 2.242732 1.275613 1.941131 1.562127 1.442565 1.082296 1.646101 1.858099
[41] 2.104243 2.161505 2.669720 1.588558 2.135163 2.255759 2.648453 1.873519
[49] 2.357154 1.479080 2.014792 1.160149 2.140127 1.887349 2.211367 1.767575
[57] 1.901908 1.893431 1.965515 1.852965 2.561421 2.334642 1.891831 1.968568
[65] 1.422440 1.542590 1.778622 1.914643 2.398257 2.174139 2.417934 1.821472
[73] 1.700356 1.830903 2.529640 1.808995 1.994697 1.047792 2.126025 2.188945
[81] 1.367079 1.896671 2.169284 1.721387 2.194744 2.629511 1.484975 2.105624
[89] 2.247515 2.242213 1.870947 1.612335 1.757744 1.840183 1.985104 1.420261
[97] 2.136239 1.814744 1.567718 2.570375
```

Figure 1.1: The sample of Weibull distribution

- Extract Upper k-Record Values: Derive the upper record of the previous sample we extract the K-upper records when $k = 1, 2, 3, 4, 5, 6$ as shown in the following table 1.2.

1-record	1.850609 2.067896 2.395194 2.484509 2.669720
2-record	1.850609 1.936578 2.067896 2.165961 2.336451 2.395194 2.484509 2.648453
3-record	1.850609 1.936578 2.067896 2.105373 2.108018 2.165961 2.336451 2.377573 2.395194 2.484509 2.561421 2.629511
4-record	1.636301 1.850609 1.936578 2.067896 2.105373 2.108018 2.165961 2.336451 2.377573 2.395194 2.484509 2.529640 2.561421 2.570375
5-record	1.636301 1.850609 1.936578 2.067896 2.105373 2.108018 2.117748 2.147960 2.165961 2.242732 2.336451 2.377573 2.395194 2.398257 2.417934 2.484509 2.529640 2.561421
6-record	1.636301 1.850609 1.936578 2.067896 2.105373 2.108018 2.117748 2.147960 2.165961 2.242732 2.255759 2.336451 2.357154 2.377573 2.395194 2.398257 2.417934 2.484509 2.529640

Table 1.2: Upper K-record of $k = 1, 2, 3, 4, 5, 6$.

Plotting the PDF in R

The Probability Density Function (PDF) of Upper k-Record Values in R is defined in the appendix A section 2.2.2 and by utilizing the code below, we can visualize the density of the upper record value of a Weibull distribution (with parameters $r = 2$, $\lambda = 5$) on the graph 1.2.

```
f <- function(n, k, r, l, w ) {
  g <- ((r * k^n) / (l^r * factorial(n - 1))) * (w / l)^(r * (n
      - 1)) * (w)^(r - 1) * (exp(-(w / l)^r))^(k)
  return(g)
           }
f(100, 3, 2, 5, w )
DATAA <- data.frame(URV = z, DOURV = f(100, 3, 2, 5, z))
library(ggplot2)
ggplot(data = DATAA, aes(x = URV, y = DOURV)) +
    geom_line() +
    geom_point()
```

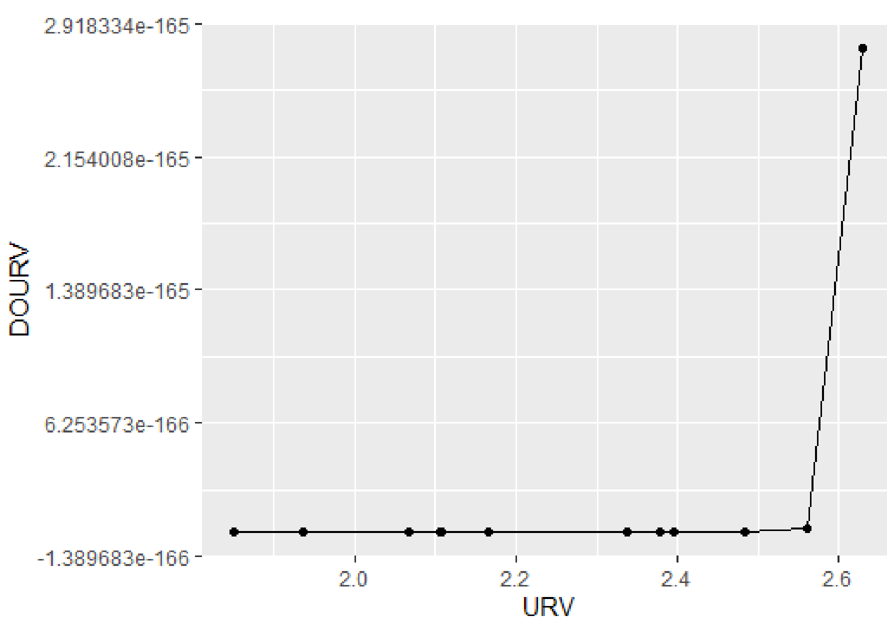

Figure 1.2: Weibull distribution's density in terms of upper k-record values

Results and Discussion

This example demonstrates the extraction and analysis of upper k-record values from a Weibull distribution using R. The theoretical function combined with R implementation and visualization provides a comprehensive approach to studying upper k-records in continuous distributions. The plot generated by the R code shows the pdf of the upper k-record values from the Weibull distribution. This visualization helps in understanding the distribution and behavior of the extreme values in the sample.

Remark 1.4.1 *The following results are just theoretical findings.*

1.4.2 Gumbel Distribution

$$
\begin{aligned}
f(x) &= \tfrac{1}{\beta} e^{\left(-\frac{x-\mu}{\beta}\right)} e^{\left(-e^{\left(-\frac{x-\mu}{\beta}\right)}\right)} \\
F(x) &= e^{\left(-e^{\left(-\frac{x-\mu}{\beta}\right)}\right)} \\
f_{U_{n,k}^{X}}(x) &= \tfrac{k^n}{\beta(n-1)!} \left[-log\left(1-e^{-e^{-\frac{x-\mu}{\beta}}}\right)\right]^{n-1} \left[1-e^{-e^{-\frac{x-\mu}{\beta}}}\right]^{k-1} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}} \\
f_{U_{n}^{X}}(x) &= \tfrac{1}{\beta(n-1)!} \left[-log\left(1-e^{-e^{-\frac{x-\mu}{\beta}}}\right)\right]^{n-1} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}} \\
f_{L_{n,k}^{X}}(x) &= \tfrac{k^n}{\beta(n-1)!} \left[-log\left(e^{-e^{-\frac{x-\mu}{\beta}}}\right)\right]^{n-1} \left[e^{-e^{-\frac{x-\mu}{\beta}}}\right]^{k-1} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}} \\
f_{L_{n}^{X}}(x) &= \tfrac{1}{\beta(n-1)!} \left[-log\left(e^{-e^{-\frac{x-\mu}{\beta}}}\right)\right]^{n-1} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}}
\end{aligned}
$$

1.4.3 Frechet Distribution

$$
\begin{aligned}
f(x) &= \tfrac{\alpha}{\beta^{-\alpha}}(x-\mu)^{-\alpha-1}\,e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}} \\
F(x) &= e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}, x>\mu \\
f_{U_{n,k}^X}(x) &= \tfrac{\alpha k^n (x-\mu)^{-\alpha-1}}{\beta^{-\alpha}(n-1)!}\left[-log\left(1-e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}\right)\right]^{n-1}\left[1-e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}\right]^{k-1} e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}} \\
f_{U_n^X}(x) &= \tfrac{\alpha}{\beta^{-\alpha}(n-1)!}\left[-log\left(1-e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}\right)\right]^{n-1}(x-\mu)^{-\alpha-1}\,e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}} \\
f_{L_{n,k}^X}(x) &= \tfrac{\alpha k^n}{\beta^{-\alpha}(n-1)!}\left[-log\left(e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}\right)\right]^{n-1}\left[e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}\right]^{k-1}(x-\mu)^{-\alpha-1}\,e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}} \\
f_{L_n^X}(x) &= \tfrac{\alpha}{\beta^{-\alpha}(n-1)!}\left[-log\left(e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}\right)\right]^{n-1}(x-\mu)^{-\alpha-1}\,e^{-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}}
\end{aligned}
$$

Chapter 2

Power and Exponential distribution: Records application

In this chapter we study two cases which are the power function distribution in the section 2.1 and the Exponential distribution in the section 2.2, each one containing some characteristics and parameters estimator with the BLUE methode and MLM.

2.1 Power function distribution

2.1.1 Characterization

The Power function distribution has a probability distribution used to model phenomena where a small number of instances carry a significant proportion of the total value. This distribution is notable for its heavy right tail, indicating that rare events or extreme values occur more frequently than in a normal distribution. Understanding the power function distribution helps in analyzing and modeling complex systems where rare events play a significant role. Let $X_{L(1)}, X_{L(2)}, ..., X_{L(r)}$ be the first r lower record values from the Power **pdf**

$$f(x)=\begin{cases}\delta x^{\delta-1}, 0\leq x\leq 1 \\ 0, \qquad\qquad\qquad \text{otherwise,}\end{cases}$$

and **cdf**

$$F\left(x\right)=x^{\delta}, 0\leq x\leq 1$$

where $\delta>0$, a simple description of the record value sequence is possible.

We introduce some basic rules that play a central role in the present study. The probability density function of $X_{L(r)}$ is given by

$$\begin{aligned}f_r\left(x\right)&=\frac{1}{\Gamma\left(r\right)}\left[-\ln\left(F\left(x\right)\right)\right]^{r-1}f\left(x\right), -\infty<x<\infty,\\&=\frac{1}{\Gamma\left(r\right)}\left[-\ln\left(x^{\delta}\right)\right]^{r-1}\delta x^{\delta-1}.\end{aligned}$$

and the joint probability density function of two lower record values $X_{L(r)}$ and $X_{L(s)}$is given by

$$f\left(x_r,x_s\right)=\frac{1}{\Gamma\left(r\right)\Gamma\left(s-r\right)}\left[-\ln\left(F\left(x_r\right)\right)\right]^{r-1}\left[\ln\left(F\left(x_r\right)\right)-\ln\left(F\left(x_s\right)\right)\right]^{s-r-1}\frac{f\left(x_r\right)}{F\left(x_r\right)}f\left(x_s\right),$$

$$-\infty<x_s<x_r<\infty.$$

$$=\frac{\delta^2}{\Gamma\left(r\right)\Gamma\left(s-r\right)}\left[-\ln\left(x_r^{\delta}\right)\right]^{r-1}\left[\ln\left(x_r^{\delta}\right)-\ln\left(x_s^{\delta}\right)\right]^{s-r-1}\frac{1}{x_r}x_s^{\delta-1}$$

Corollairy 2.1.1 $\bullet$*The first moment of $X_{L(r)}$ from the Power function probability-distribution is given by*

$$E(X_{L(r)})=\left(\frac{\delta}{\delta+1}\right)^r.$$

• The second moment of $X_{L(r)}$ is (similar to the first)

$$E(X_{L(r)})^2 = \left(\frac{\delta}{\delta+2}\right)^r.$$

• We compute the variance of $X_{L(r)}$ to be

$$\begin{aligned} Var(X_{L(r)}) &= E(X_{L(r)})^2 - \left(E(X_{L(r)})\right)^2 \\ &= \left(\frac{\delta}{\delta+2}\right)^r - \left(\frac{\delta}{\delta+1}\right)^{2r} \\ &= \left(\frac{\delta}{\delta+1}\right)^r \left[\left(\frac{\delta+1}{\delta+2}\right)^r - \left(\frac{\delta}{\delta+1}\right)^r\right] \\ &= a_r b_r. \end{aligned} \tag{2.1}$$

Proof. let's define

$$\begin{aligned} E(X_{L(r)}) &= \int_0^1 x f_r(x)\, dx \\ &= \int_0^1 x \frac{1}{\Gamma(r)} \left[-\ln\left(x^\delta\right)\right]^{r-1} \delta x^{\delta-1} dx \end{aligned}$$

substituting $y = -\ln\left(x^\delta\right)$ and simplifying we get

$$E(X_{L(r)}) = \int_0^{+\infty} \frac{1}{\Gamma(r)} y^{r-1} \exp\left(-y\left(1+\frac{1}{\delta}\right)\right) dy$$

by using the Gama distribution($\int_0^{+\infty} \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} = 1$) and by matching we find

$$E(X_{L(r)}) = \left(\frac{\delta}{\delta+1}\right)^r = b_r. \tag{2.2}$$

By using the same method we can prove the other statements. ■

2.1.2 Parameters estimator

The following concepts, Mbah and Tsokos, [[15]], identifies the analytical estimator of the parameters for μ and σ when δ is known.

Lemma 2.1.1 *Let $x_1, x_2, \ldots, x_r$ be r lower record values from the power function distribution is given by equation ($f(x) = \delta\left(\left(x-\mu\right)/\sigma\right)^{\delta-1}$), where*

$$h^t = (x_1, x_2, \ldots, x_r),$$

then

$$E\left(h^t\right) = \mu\mathbf{1} + \sigma\alpha,$$

and

$$Var\left(h^t\right) = \sigma^2\mathbf{V},$$

Proof. We have from Lloyd (1952) that:

$$h_1 \leqslant h_2 \leqslant \ldots \leqslant h_r,$$

$$z_1 \leqslant z_2 \leqslant \ldots \leqslant z_r;$$

where $(h_1, h_2, \ldots, h_r)$ and$(z_1, z_2, \ldots, z_r)$ in ascending order of magnitude of $(x_1, x_2, \ldots, x_r)$ and y_r in the order when

$$y_r = \left(x_r - \mu\right)/\sigma,$$

$$z_r = \left(h_r - \mu\right)/\sigma.$$

Let

$$E\left(z_r\right) = \alpha_r \quad , \quad Var\left(z_r\right) = V;$$

these quantities have known values depending on the form of the parent distribution but not on the parameters μ and σ. we clearly have in the matrix form:

$$E\left(h^t\right) = \mu\mathbf{1} + \sigma\alpha \quad , \quad Var\left(h^t\right) = \sigma^2\mathbf{V}.$$

where, from equations (2.2) and (2.1)

$$\mathbf{1}^t = (1, 1, \ldots, 1),$$

$$\alpha^t = (b_1, b_2, \ldots, b_r),$$

■

Lemma 2.1.2 *The inverse of covariace martix*

$$V^{-1} = \left(V^{ij}\right), 1 \leq i < j \leq r,$$

where

$$V^{ij} = \begin{cases} \frac{a_{i+1}b_{i-1} - a_{i-1}b_{i+1}}{(a_i b_{i-1} - a_{i-1}b_i)(a_{i+1}b_i - a_i b_{i+1})}, & \textit{if } i = j = 1 \textit{ to } r-1 \\ \frac{-1}{a_{i+1}b_i - a_i b_{i+1}}, & \textit{if } j = i+1 \textit{ and } i = 1 \textit{ to } r-1 \\ \frac{b_{r-1}}{b_r(a_r b_{r-1} - a_{r-1}b_r)}, & \textit{if } i = j = r \\ 0, & \textit{if } j > i+1 \end{cases}$$

Proof. From the lemma (2.1.1) we can write the Esperance and the variance as

$$E\left(h^t\right) = M\theta \quad , \quad Var\left(h^t\right) = \sigma^2\mathbf{V};$$

where M is the $(r \times 2)$ matrix $(1, \alpha)$,and $\theta^t = (\mu, \sigma)$ .With V is the $(r \times r)$ symmetric positive definite matrix

$$V = v_{ij} = \begin{pmatrix} Var\left(x_1, x_1\right) & Cov\left(x_1, x_2\right) & \ldots & Cov\left(x_1, x_r\right) \\ Cov\left(x_1, x_2\right) & \ddots & & Cov\left(x_2, x_r\right) \\ \vdots & & \ddots & \vdots \\ Cov\left(x_1, x_r\right) & Cov\left(x_2, x_r\right) & \cdots & Var\left(x_r, x_r\right) \end{pmatrix}$$

For $A = V^{-1}$,the required estimator of the vector θ of parameters are given by

$$\widehat{\theta} = \left(M^t A M\right)^{-1} M^t A h^t.$$

The variance matrix of the estimates is $\left(M^t A M\right)^{-1} \sigma^2$,where

$$M^t A M = \begin{pmatrix} 1^t A 1 & 1^t A \alpha \\ 1^t A \alpha & \alpha^t A \alpha \end{pmatrix},$$

which his inverse

$$\left(M^t A M\right)^{-1} = \frac{1}{\Delta} \begin{pmatrix} \alpha^t A \alpha & -1^t A \alpha \\ -1^t A \alpha & 1^t A 1 \end{pmatrix},$$

where Δ is the determinant of the matrix $M^t A M$. ■

Theorem 2.1.1 *The **BLUE**,$\widehat{\mu}$ and $\widehat{\sigma}$ for μ and σ based on r lower record values from the power function probability distribution is given by*

$$\widehat{\mu} = \frac{\alpha^t \mathbf{V}^{-1} \left(\alpha \mathbf{1}^t - \mathbf{1} \alpha^t\right) \mathbf{V}^{-1} \mathbf{h}}{\Delta},$$

and

$$\widehat{\sigma} = \frac{\mathbf{1}^t \mathbf{V}^{-1} \left(\mathbf{1} \alpha^t - \alpha \mathbf{1}^t\right) \mathbf{V}^{-1} \mathbf{h}}{\Delta}.$$

Proof. By using lemma (2.1.1) and lemma (2.1.2) (using the results of the inverse matrix in the equation of $\widehat{\theta}$ We find the following estimates)

$$\widehat{\mu} = -\alpha^t W h^t \quad , \quad \widehat{\sigma} = 1 W h^t,$$

where W is the skew-symmetric matrix defined by

$$W = \frac{A\left(1\alpha^t - \alpha 1^t\right) A}{\Delta}.$$

■

Corollairy 2.1.2 *The variance of these estimates are*

$$Var\left(\widehat{\mu}\right) = \frac{\alpha^t A \alpha \sigma^2}{\Delta} \quad , \quad Var\left(\widehat{\sigma}\right) = \frac{1^t A 1 \sigma^2}{\Delta}.$$

Lemma 2.1.3 *The best linear unbiased estimates (BLUE),$\widehat{\mu}$ and $\widehat{\sigma}$ for μ and σ for known δ are respectively*

$$\widehat{\mu} = -\frac{1}{D_0}\left[\frac{(\delta+2)^2 x_1}{\delta} + \left(\frac{\delta+2}{\delta}\right)^2 x_2 + \ldots + \left(\frac{\delta+2}{\delta}\right)^{r-1} x_{r-1} - (\delta+1)\left(\frac{\delta+2}{\delta}\right)^r x_r\right],$$

and

$$\widehat{\sigma} = \frac{\delta+1}{\delta D_0}\left[\left(D_0 + \frac{(\delta+2)^2}{\delta}\right) x_1 + \left(\frac{\delta+2}{\delta}\right)^2 x_2 + \ldots + \left(\frac{\delta+2}{\delta}\right)^{r-1} x_{r-1} - (\delta+1)\left(\frac{\delta+2}{\delta}\right)^r x_r\right],$$

where

$$D_0 = \sum_{i=2}^{r}\left(\frac{\delta+2}{\delta}\right)^i.$$

Proof. Using the method introduced by Lloyd, [[14]]; the entries of $\mathbf{V}^{-1}$ are given by

$$V^{ii} = \frac{A}{BC} = \left(2\delta^2 + 4\delta + 1\right)\left(\frac{\delta+2}{\delta}\right)^i, i = 1, \ldots, r-1;$$

where:

$$A = \left(\frac{\delta}{\delta+1}\right)^{i-1}\left[\left(\frac{\delta+1}{\delta+2}\right)^{i+1} - \left(\frac{\delta}{\delta+1}\right)^{i+1}\right] - \left(\frac{\delta}{\delta+1}\right)^{i+1}\left[\left(\frac{\delta+1}{\delta+2}\right)^{i-1} - \left(\frac{\delta}{\delta+1}\right)^{i-1}\right]$$

$$B = \left[\left(\frac{\delta}{\delta+1}\right)^{i-1}\left(\left(\frac{\delta+1}{\delta+2}\right)^{i} - \left(\frac{\delta}{\delta+1}\right)^{i}\right) - \left(\frac{\delta}{\delta+1}\right)^{i}\left(\left(\frac{\delta+1}{\delta+2}\right)^{i-1} - \left(\frac{\delta}{\delta+1}\right)^{i-1}\right)\right]$$

$$C = \left[\left(\frac{\delta}{\delta+1}\right)^{i}\left(\left(\frac{\delta+1}{\delta+2}\right)^{i+1} - \left(\frac{\delta}{\delta+1}\right)^{i+1}\right) - \left(\frac{\delta}{\delta+1}\right)^{i+1}\left(\left(\frac{\delta+1}{\delta+2}\right)^{i} - \left(\frac{\delta}{\delta+1}\right)^{i}\right)\right]$$

$$\begin{aligned} V^{ij} &= V^{ji} \\ &= \frac{-1}{\left(\frac{\delta}{\delta+1}\right)^{i}\left[\left(\frac{\delta+1}{\delta+2}\right)^{i+1} - \left(\frac{\delta}{\delta+1}\right)^{i+1}\right] - \left(\frac{\delta}{\delta+1}\right)^{i+1}\left[\left(\frac{\delta+1}{\delta+2}\right)^{i} - \left(\frac{\delta}{\delta+1}\right)^{i}\right]} \\ &= -\left(\delta+1\right)\frac{(\delta+2)^{i+1}}{\delta^i}, j = i+1, i = 1, \ldots, r-1, \end{aligned}$$

$$V^{ij} = 0 \qquad \text{for} \quad |i-j| > 1,$$

and

$$\begin{aligned} V^{rr} &= \frac{\left(\frac{\delta}{\delta+1}\right)^{r-1}}{\left(\frac{\delta}{\delta+1}\right)^{r}\left[\left(\frac{\delta}{\delta+1}\right)^{r-1}\left(\left(\frac{\delta+1}{\delta+2}\right)^{r} - \left(\frac{\delta}{\delta+1}\right)^{r}\right) - \left(\frac{\delta}{\delta+1}\right)^{r}\left(\left(\frac{\delta+1}{\delta+2}\right)^{r-1} - \left(\frac{\delta}{\delta+1}\right)^{r-1}\right)\right]} \\ &= (\delta+1)^2\left(\frac{\delta+2}{\delta}\right)^r \end{aligned}$$

we have that by Mbah[[15]]

$$\alpha^t \mathbf{V}^{-1} = \left[(\delta+1)(\delta+2), 0, \ldots, 0 \right],$$

and

$$\mathbf{1}^t \mathbf{V}^{-1} = \left[-(2\delta+3)\frac{\delta+2}{\delta}, -\left(\frac{\delta+2}{\delta}\right)^2, \ldots, -\left(\frac{\delta+2}{\delta}\right)^{r-1}, (\delta+1)\left(\frac{\delta+2}{\delta}\right)^r \delta^{r-2} \right].$$

Therefore,

$$\alpha^t \mathbf{V}^{-1} \alpha = \delta(\delta+2),$$

$$\alpha^t \mathbf{V}^{-1} \mathbf{1} = (\delta+1)(\delta+2),$$

and

$$\mathbf{1}^t \mathbf{V}^{-1} \mathbf{1} = \frac{(\delta+1)^2(\delta+2)}{\delta} + D_0.$$

Let

$$\begin{aligned} \Delta &= \left(\alpha^t \mathbf{V}^{-1} \alpha\right)\left(\mathbf{1}^t \mathbf{V}^{-1} \mathbf{1}\right) - \left(\alpha^t \mathbf{V}^{-1} \mathbf{1}\right)^2 \\ &= \delta(\delta+2) D_0, \end{aligned}$$

then

$$\begin{aligned} \alpha^t \mathbf{V}^{-1}\left(\alpha \mathbf{1}^t - \mathbf{1}\alpha^t\right)\mathbf{V}^{-1} = \delta(\delta+2)^3 \Bigg[&-(\delta+2), -\frac{\delta+2}{\delta}, \ldots, -\left(\frac{\delta+2}{\delta}\right)^{r-2}, \\ &(\delta+1)\left(\frac{\delta+2}{\delta}\right)^{r-1} \Bigg], \end{aligned}$$

and

$$\mathbf{1}^t\mathbf{V}^{-1}\left(\mathbf{1}\alpha^t-\alpha\mathbf{1}^t\right)\mathbf{V}^{-1}=(\delta+1)(\delta+2)\left[D_0+\frac{(\delta+2)^2}{\delta},\frac{(\delta+2)^2}{\delta^2},\ldots,\frac{(\delta+2)^{r-1}}{\delta^{r-1}},\right.$$
$$\left.-\frac{(\delta+1)(\delta+2)^r}{\delta^r}\right].$$

Thus, the best linear unbiased estimates (BLUE),$\widehat{\mu}$ and $\widehat{\sigma}$ for μ and σ for known δ are respectively

$$\widehat{\mu}=-\frac{1}{D_0}\left[\frac{(\delta+2)^2x_1}{\delta}+\left(\frac{\delta+2}{\delta}\right)^2x_2+\ldots+\left(\frac{\delta+2}{\delta}\right)^{r-1}x_{r-1}-(\delta+1)\left(\frac{\delta+2}{\delta}\right)^rx_r\right],$$

and

$$\widehat{\sigma}=\frac{\delta+1}{\delta D_0}\left[\left(D_0+\frac{(\delta+2)^2}{\delta}\right)x_1+\left(\frac{\delta+2}{\delta}\right)^2x_2+\ldots+\left(\frac{\delta+2}{\delta}\right)^{r-1}x_{r-1}\right.$$
$$\left.-(\delta+1)\left(\frac{\delta+2}{\delta}\right)^rx_r\right],$$

where

$$D_0=\sum_{i=2}^{r}\left(\frac{\delta+2}{\delta}\right)^i.$$

■

Corollairy 2.1.3 *The variance of $\widehat{\mu},\widehat{\sigma}$ and covariance of $\widehat{\mu},\widehat{\sigma}$ are given by*

$$Var\left(\widehat{\mu}\right)=\frac{\alpha^t\mathbf{V}^{-1}\alpha}{\Delta}\sigma^2=\frac{\sigma^2}{D_0},$$

$$\begin{aligned}Var\left(\widehat{\sigma}\right)&=\frac{\mathbf{1}^t\mathbf{V}^{-1}\mathbf{1}}{\Delta}\sigma^2\\&=\left(\frac{(\delta+1)^2}{\delta^2D_0}+\frac{1}{\delta(\delta+2)}\right)\sigma^2.\end{aligned}$$

Remark 2.1.1 *If $\delta = 1$, then $D_0 = 9(3^{r-1} - 1)/2$, and BLUE $\widehat{\mu}, \widehat{\sigma}$ of μ and σ from the uniform distribution based on lower record values $X_{L(1)}, X_{L(2)}, \ldots, X_{L(r)}$ are*

$$\widehat{\mu} = \frac{1}{D_0}\left[-3^2 X_{L(1)} + \sum_{i=2}^{r-1} 3^i X_{L(i)} + 2 \times 3^r X_{L(r)}\right],$$

and

$$\widehat{\sigma} = 2\left(X_{L(1)} - \widehat{\mu}\right),$$

with

$$Var\left(\widehat{\mu}\right) = \frac{\sigma^2}{D_0}$$

$$Var\left(\widehat{\sigma}\right) = \frac{3^r + 5}{9\left(3^{r-1} - 1\right)},$$

Method of Maximum likelihood:

Lemma 2.1.4 *The joint pdf of r lower record values $X_{L_1}, X_{L_2}, \ldots, X_{L_r}$ from a continuous cumulative probability distribution function $F(x)$ is given by:*

$$f_{1,2,\ldots,r}(x_1, x_2, \ldots, x_r) = f(x_r) \prod_{i=1}^{r-1} \frac{f(x_i)}{F(x_i)}, \quad -\infty < x_1 < \ldots < x_r < \infty.$$

Observe from the previous lemma that the likelihood of a standard power function distribution is

$$L\left(\delta \mid x_1, x_2, \ldots, x_n\right) = x_r^{\delta} \prod_{i=1}^{r} \frac{\delta}{x_i}.$$

The loglikelihood function is given by

$$\begin{aligned}\log\left(L\left(\delta \mid x_1, x_2, \ldots, x_n\right)\right) &= \log x_r^{\delta} + \log\left(\prod_{i=1}^{r}\frac{\delta}{x_i}\right)\\ &= \delta\log\left(x_r\right) + \sum_{i=1}^{r}\left(\log\delta - \log x_i\right)\\ &= -\delta\log\left(x_r\right) - \sum_{i=1}^{r}\log x_i + r\log\left(\delta\right). \qquad (2.3)\end{aligned}$$

Differentiating (2.3) with respect to δ and equating to zero gives the MLE estimate of δ,$\widehat{\delta}_{MLE}$ to be

$$\widehat{\delta}_{MLE} = -\frac{r}{\ln\left(x_r\right)}.$$

2.2 Exponential Distribution

2.2.1 Characterization

The exponential distribution is a probability distribution that describes the time between events in a Poisson process, where events occur continuously and independently at a constant average rate. It is commonly used in various fields such as reliability engineering, queueing theory, and survival analysis, it is often used to model phenomena such as the time until the next radioactive decay, the lifespan of certain devices, or the waiting time between arrivals of customers at a service point. Suppose that $X_{U(1)}, X_{U(2)}, \ldots, X_{U(r)}$ are r upper record values with probability density function (pdf) given by:

$$f(x) = \begin{cases}\lambda\exp(-\lambda x) & 0 < x < \infty,\\ 0 & \text{otherwise.}\end{cases}$$

and cumulative distribution function (cdf)

$$F(x) = 1 - \exp(-\lambda x).$$

We introduce some basic rules that play a central role in the present study. The probability density function of $X_{U(r)}$ is given by

$$\begin{aligned} f_r(x) &= \frac{1}{\Gamma(r)}[-\ln(1-F(x))]^{r-1}f(x), \quad -\infty < x < \infty, \\ &= \frac{1}{\Gamma(r)}\lambda^r x^{r-1}\exp(-\lambda x). \end{aligned}$$

The first moment of $X_{U(r)}$ from the Exponential distribution is given by

$$E(X_{U(r)}) = \frac{\Gamma(r+1)}{\lambda\Gamma(r)}.$$

Proof. Let's define

$$\begin{aligned} E(X_{U(r)}) &= \int_0^\infty x f_r(x)dx \\ &= \int_0^1 x\frac{1}{\Gamma(r)}[-\ln(\exp(-\lambda x))]^{r-1}\lambda\exp(-\lambda x)dx. \end{aligned}$$

By using the Gamma distribution $\left(\int_0^{+\infty} \frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}e^{-(y/\beta)}dy = 1\right)$ and by matching we find

$$E(X_{U(r)}) = \frac{\Gamma(r+1)}{\lambda\Gamma(r)} = b_r.$$

The second moment of $X_{U(r)}$ is (similar to the first)

$$E(X_{U(r)})^2 = \frac{\Gamma(r+2)}{\lambda^2\Gamma(r)}.$$

We compute the variance of $X_{U(r)}$ to be

$$\begin{aligned}
Var(X_{Ur)}) &= E(X_{U(r)})^2 - (E(X_{U(r)}))^2 \\
&= \frac{\Gamma(r+2)}{\lambda^2\Gamma(r)} - \left(\frac{\Gamma(r+1)}{\lambda\Gamma(r)}\right)^2 \\
&= \frac{1}{\lambda}\frac{\Gamma(r+1)}{\Gamma(r)}\left(\frac{\Gamma(r+2)}{\lambda\Gamma(r+1)} - \frac{\Gamma(r+1)}{\lambda\Gamma(r)}\right) \\
&= a_r b_r.
\end{aligned}$$

■

2.2.2 Parameters estimator

Best Linear Unbiased Estimator

Suppose that $x_1, x_2, \ldots, x_r$ are the r (upper) record values from $E(\mu, \sigma)$ (by Ahsanullah[[3]]), with pdf

$$f(x) = \begin{cases} \frac{1}{\sigma}\exp\left(\sigma^{-1}(x-\mu)\right), & -\infty < \mu < x < \infty, \quad \sigma > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Without any loss of generality, we will consider in this part the standard exponential distribution, $E(0,1)$, with pdf $f(x) = \exp(-x)$, $0 \leq x < \infty$, in which case we have $f(x) = 1 - F(x)$,

$$1^t = (1, 1, \ldots, 1), \quad \alpha^t = (1, 2, \ldots, r),$$

$$V = (v_{ij}), \quad v_{ij} = a_i b_i, \quad i, j = 1, 2, \ldots, r.$$

The inverse of V can be written as

$$V^{-1} = \begin{bmatrix} 2 & i = j = 1 \text{ to } r-1 \\ -1 & |i-j| = 1 \text{ and } i, j = 1 \text{ to } r \\ 1 & i = j = r \\ 0 & \text{otherwise.} \end{bmatrix}$$

Using the method introduced by Lloyd, [14], we have that (by Ahsanullah [3])

$$\alpha^t V^{-1} = (0, 0, \ldots, 0, 1),$$

and

$$1^t V^{-1} = (1, 0, \ldots, 0).$$

Therefore,

$$\alpha^t V^{-1} \alpha = r,$$

$$\alpha^t V^{-1} 1 = 1,$$

and

$$1^t V^{-1} 1 = 1.$$

Let

$$\Delta = r - 1,$$

Thus, the best linear unbiased estimates (BLUE), $\widehat{\mu}$ and $\widehat{\sigma}$ for μ and σ are respectively:

$$\widehat{\mu} = \frac{r x_1 - x_r}{r - 1},$$

and

$$\widehat{\sigma} = \frac{x_r - x_1}{r - 1},$$

The variance of $\widehat{\mu}$, $\widehat{\sigma}$ and covariance of μ, σ are given by

$$\mathrm{Var}(\widehat{\mu}) = \frac{r\sigma^2}{r-1},$$

$$\mathrm{Var}(\widehat{\sigma}) = \frac{\sigma^2}{r-1}.$$

Method of Maximum Likelihood

Observe from equation (1.14) that the likelihood of the Exponential function distribution is

$$L(\sigma|x_1, x_2, \ldots, x_r) = \frac{1}{\sigma} \exp\left(-\frac{x_r}{\sigma}\right) \prod_{i=1}^{r-1} \frac{1}{\sigma}.$$

The log-likelihood function is given by

$$\begin{aligned}
\log(L(\sigma|x_1, x_2, \ldots, x_r)) &= \log\left(\frac{1}{\sigma}\right) - \frac{x_r}{\sigma} + \log\left(\prod_{i=1}^{r-1} \frac{1}{\sigma}\right) \\
&= -\log\sigma - \frac{x_r}{\sigma} - \sum_{i=1}^{r-1} \log\sigma \\
&= -\log\sigma - \frac{x_r}{\sigma} - (r-1)\log\sigma.
\end{aligned}$$

On differentiating the above equation concerning the parameter σ and equating to zero, we get

$$\widehat{\sigma}_{\mathrm{MLE}} = -\frac{x_r}{r}.$$

To determine the Best Linear Unbiased Estimator (BLUE) and the Maximum Likelihood Estimator (MLE), one typically relies on statistical models and assumptions about the data. BLUE is derived from the Gauss-Markov theorem, which assumes linearity, unbiasedness, and minimal variance within linear estimators. Conversely, MLE involves maximizing the likelihood function based on the observed data to estimate the parameters that make the observed data most probable. Both methods are essential in statistical inference, offering different approaches depending on the context and assumptions.

Conclusion

In the present study, we have introduced the concept of records and explained record time, record value, and lower and upper records including some important properties of K-record and ordinary record which are the special case when (k=1), then we provide them by giving two examples.

We present records from specific continuous distributions, in addition, we discuss it by an application for the Weibull distribution.

On the other hand, we investigate the problem of estimation of the location and scale parameters in the best linear unbiased estimator (BLUE) of the Power function and the Exponential distributions by using different types of record (Record in the first and the upper-record in the second), where we started with some basic characteristics by extracting the single and the second moment with the variance of that distribution, then we summarized the method of the Best Linear Unbiased estimator and give the estimation of parameters with their first moment and variance, at the end we use the Likelihood Estimation.

In conclusion, record statistics is a broad term; it can refer to various types of statistical information related to records. In a general sense, record statistics could involve analyzing and presenting data related to records.

The Best Linear Unbiased Estimator is the most accurate method to estimate compared with Likelihood or moment estimation methods but it has negative points of which: Limited Applicability; BLUE is specifically designed for linear regression models. It may not be suitable for non-linear relationships or models with complex functional forms.

The study of other distribution with respect to records is not very much exploited. This could be because of the complex form of the probability density function of record observation since close form solutions will not be obtained for the maximum likelihood estimates.Another method is to investigating the concept of records with respect to using the kernel density approach to characterize the behavior of records is an interesting extension of the theory of records.

Bibliography

[1] Ahsanullah , M. (2004) .Record statistics : Theory and applications .University Press of America .

[2] Ahsanullah, M., & Raqab, M. Z. (2006). Bounds and characterizations of record statistics. Nova Publishers.

[3] Ahsanullah, M., & Nevzorov, V. B. (2015). Records via probability theory. Atlantis Press.

[4] Ahsanullah, M. (2020). Record Values. International Journal of Statistical Sciences, 20(2), 257-266.

[5] Amin, E. A. (2011). K-th upper record values and their moments. International Mathematical Forum, 6(61), 3013-3021. https://doi.org/10.12988/imf.2011.130100

[6] Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1998). Records. John Wiley & Sons.

[7] Balakrishnan, N., & Rao, C. R. (n.d.). A note on the best linear unbiased estimation based on order statistics.

[8] Chandler, K. N. (1952). The distribution and frequency of record values. Journal of the Royal Statistical Society, Series B, 14, 220–228.

[9] Chegg. (n.d.). Use pulse rates in beats per minute (BPM) of a random sample of adult females listed in the data set available. Retrieved from https://www.chegg.com/homework-help/questions-and-answers/use-pulserates-beats-per-minute-bpm-random-sample-adult-females-listed-data-setavailable-q44686006

[10] David, H. A., & Nagaraja, H. N. (2003). Order statistics. Wiley.

[11] Dziubdziela, W., & Kopocinsky, B. (1976). Limiting properties of the kth record values. Applications Mathematica, 15, 187–190.

[12] Grudzien, Z. (1982). Characterization of distribution of time limits in record statistics as well as distribution and moments of linear record statistics from the sample of random numbers. Praca Doktorska, UMCS, Lublin.

[13] Karlin, S. (1966). A First Course in Stochastic Processes. Academic Press.

[14] Lloyd, E. H. (1952). Least Squares estimation of location and scale parameters using order statistics. Biometrika, 39, 88-95.

[15] Mbah, A. K. (2007). On the theory of records and applications (Doctoral dissertation, University of South Florida). Retrieved from https://digitalcommons.usf.edu/etd/2282

[16] Pawlas, P., & Szynal, D. (1998). Relations for single and product moments of kth record values from exponential and Gumbel distributions. Journal of Applied Statistical Science, 7, 53-61.

[17] Pawlas, P., & Szynal, D. (2000). Characterization of the inverse Weibull distribution and generalized extreme value distribution by moments of kth-record values. Applications Mathematicae, 27, 197-202.

[18] Sajeevkumar, N. K., Priya, R. S., & Sumi, A. R. (2022). Use of concomitants of record values in the estimation of parameters μ_2 and σ_2 involved in Morgenstern type bivariate exponential distribution. International Journal of Statistics and Applied Mathematics, 7(4), 110-118.

[19] Tripathi, A., Singh, U., & Singh, S. K. (2020). Does the type of records affect the estimates of the parameters? Journal of Modern Applied Statistical Methods, 19(1), eP2863. https://doi.org/10.22237/jmasm/1608552300

[20] Wergen, G. (2013). Records statistics beyond the standard model - Theory and applications (Doctoral dissertation). University of Cologne.

Appendix A: Software R

2.3 What is the R language?

• The R language is a programming language and mathematical environment used for data processing. It allows you to perform both simple and complex statistical analyzes such as linear or non-linear models, hypothesis testing, time series modeling, classification, etc. It also has many very useful, professional-quality graphics functions.

• R was created by Ross Ihaka and Robert Gentleman in 1993 at the University of Auckland, New Zealand, and is now developed by the R Development Core Team.

The origin of the name of the language comes, on the one hand, from the initials of the first names of the two authors (Ross Ihaka and Robert Gentleman) and, on the other hand, from a play on words on the name of the language S to which it is related.

2.4 R code used at chapter 1

```
upper.record.values <- function(sqnc, k) {
    if (sum(is.na(sqnc)) > 0) {
        stop("There are missing values in data. The function is not
            designated for such data!")
    } else {
```

```
        if (length(k) == 1 && k >= 1 && k <= length(sqnc) && is.
            numeric(k) && floor(k) == k) {
            sqnc <- unique(sqnc)
            dl <- length(sqnc)
            wndw <- sort(sqnc[1:k])
            records <- wndw[1]
            if (k < dl) {
                for (i in (k+1):dl) {
                    if (sqnc[i] > wndw[1]) {
                        if (i <= dl-k+2) {
                            w <- which(sqnc[i] > wndw)
                        } else {
                            w <- which(sqnc[i] > wndw[1:(dl-i+2)])
                        }
                        where <- w[length(w)]
                        if (where > 1) {
                            wndw[1:(where-1)] <- wndw[2:where]
                        }
                        wndw[where] <- sqnc[i]
                        records <- c(records, wndw[1])
                    }
                }
            }
            return(records)
        } else {
            stop("k must be an integer between 1 and length(sqnc)")
        }
    }
}
```

Abstract

Record statistics is a broad term; it can refer to various types of statistical information related to records. In a general sense, record statistics could involve analyzing and presenting data related to records. The subject of the study is to introduce the concept of record as applied to probability distribution function and present some their characterization results based on the records values. In addition, we discuss the problem of estimation of the location and scale parameters in Best Linear Unbiased Estimator (BLUE) and how applied it in both different types of records with differents distributions functions.

Résumé

Les statistiques des records sont des termes larges; elles peuvent se référer à divers types d'informations statistiques liées aux records. Dans un sens général, les statistiques des records peuvent impliquer l'analyse et la présentation des données relatives aux records. Le sujet de l'étude consiste à introduire le concept de record tel qu'appliqué à la fonction de distribution de probabilité et à présenter certains résultats de leur caractérisation basés sur les valeurs des records. De plus, nous discutons du problème de l'estimation des paramètres de localisation et d'échelle dans le Meilleur Estimateur Linéaire Non Biaisé (BLUE) et de son application à différents types de records avec différentes fonctions de distribution.

الملخص

إحصاءات القيم القياسية هو مصطلح واسع؛ يمكن أن يشير إلى أنواع مختلفة من المعلومات الإحصائية المتعلقة بالقيم القياسية. بشكل عام، قد تتضمن إحصاءات القيم القياسية تحليل وتقديم البيانات المتعلقة بها.

موضوع الدراسة هو تقديم مفهوم القيم القياسية كما يُطبق على دالة توزيع الاحتمالات وتقديم بعض نتائج التوصيف استنادًا إليها. بالإضافة إلى ذلك، نناقش مشكلة تقدير معلمات الموقع والمقياس في أفضل مقدر خطي غير منحاز وكيفية (BLUE)تطبيقه في أنواع مختلفة من القيم القياسية مع دوال توزيع مختلفة