People's Democratic Republic of Algeria
الجمهورية الجزائرية الديمقراطية الشعبية
Ministry of Higher Education and Scientific Research
وزارة التعليم العالي و البحث العلمي

Mouloud Mammeri University of Tizi-Ouzou
Faculty of sciences
Department of Mathematics

Thesis submitted to the department of Mathematics

in fulfillment of requirements for a Master's degree in Mathematics

Option : Probability and Statistics

Bayesian prediction of order statistics

By:
Miss. Imache Fatima

Supervisor :
Pr. Fellag Hocine

Thesis Jury:

Dr. Boualam Karima :	UMMTO	Chairman.
Pr. Fellag Hocine :	UMMTO	Supervisor.
Dr. Belkacem Cherifa :	UMMTO	Examiner.

Academic year : 2022/2023


بسم الله الرحمن الرحيم

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis supervisor, Professor Fellag Hocine, who provided dedicated support throughout my academic development and has always been a great inspiration to me with his long and successful scientific career . ultimately i am grateful for allowing me to undertake this research . His blend of words of motivation and critical yet invaluable insights is what propelled my visions into a reality and truly helped me fulfil my potential, for which I am inexplicably grateful. I am excited to carry the lessons you have imparted with me as i embark on the new adventures ; thank you once again for being not just a professor , but mentor and a guide .

This dissertation could not have happened without the guidance and helpful feedback of the members of the jury including *Mrs* Boualem Karima and *Mrs* Belkacem Cherifa who dedicated their valuable time and expertise to assess my work. I am very grateful for their opinions, which helped bolster my own confidence in this strenuous process ; I am deeply honored to have the privilege of presenting my work before you .

I would like to express a special thanks to the probability and statistics team who encouraged my passion for this field also for motivating me to strive for excellence in all that i do ; and to all my professors in the mathematics department of UMMTO during my academic journey. I am honored to have had the privilege of learning under your guidance .

Tizi-Ouzou, December 5, 2023.

Dedications

As our beloved prophet Mohammed (saws) said in his noble hadith " mother, mother ,mother then father"

To my little sweet parents

Through the highs and lows of this academic journey you have been my pillars of strength , my confidants and my constant source of inspiration . For your countless sacrifices ,unwavering support and encouragements that have fueled my academic pursuits ;i dedicate this dissertation to you with deep gratitude ; your belief in me has been my greatest motivation and the deriving force my determination to succeed.

To my beloved siblings

In your unique ways ,each of you has contributed to shaping my character pushing me to exceed my limits , and reminding me of the importance of perseverance , this work stands as a tribute to the bond we share and the love that knows no bounds.

To the most significant void in my life **GB**
whose friendship ,support and inspiration will forever resonate in my heart.

To probability and statistics mates " M2 PS 2023"

For the laughter and understanding have provided a balance to the demanding academic journey .Thank you for the cherished memories we have created together.
Finally to my little friends **ZUZU** and **BLACK**.

Abstract

The study of order statistics was a curiosity of lot of mathematics researchers, because of their several applicability in different fields; especially in non-parametric statistics.
We tend to predict time-related processes. In this research,we focus on predicting future events based on their order.

Two types of prediction will be discussed in this thesis classical frequentist prediction and Bayesian prediction given both point prediction and interval prediction. Finally an analysis of real data and simulation studies will be conducted to illustrate the theoretical concepts.

Keywords : Order statistics, Prediction, Bayesian prediction, Frequentist prediction, Interval prediction, Point prediction.

Résumé

L'étude des statistiques d'ordre était une curiosité pour des chercheurs en mathématiques, parce qu'ils sont applicable presque dans tous les domaines;surtout ceux lies au non-paramétrique.
On a tendance à prédire des processus liés dans le temps. Dans cette recherche on s'intéresse a la prédiction des évènements futures en se basant sur leur ordre. Deux types de prédiction seront discutées dans ce mémoire la prédiction classique fréquentiste et la prédiction Bayesienne ;en utilisant la prédiction ponctuelle et la prédiction par intervalle.En fin une analyse de données réelles et étude de simulation sera mise en place a fin de bien illustrer les concepts théoriques.

Mots clés : Statistique d'ordre, Prédiction ,Prédiction Bayesienne, Prédiction fréquentiste ,Intervalle prédiction , Prédiction ponctuelle.

Contents

List of Figures

List of Tables

list of abbreviations and acronyms

r.v.	*random variable .*
df	*density function .*
cdf	*Cumulative distribution function.*
pdf	*Probability density function.*
pmf	*Probability mass function.*
iid	*Independent and identically distributed.*
inid	*Independent nonidentically distributed.*
iif	*if and only if.*
MSE	*Mean square error .*
BLUE	*Best linear unbaised estimator.*
BLUP	*Best linear unbaised predictor.*
BLIP	*Best linear invariant predictor.*
ABLUP	*Asymptotically best linear unbaised predictor.*
ABLIP	*Asymptotically best linear invariant predictor .*
LHS	*Left Hand Side .*
RHS	*Right Hand Side .*

PLF	*Predictive Likelihood function .*
PMLE	*Predictive Maximum Likelihood Estimator .*
MSPE	*Mean Square Error Estimator .*
MUP	*Median Unbaised Estimator .*
SEL	*squared error loss function*
RRSS	*record ranked set sample.*
PL	*precautionary loss function*
WSEL	*weighted squared loss function.*
LINEX	*linear exponential loss function.*

Introduction

The term "order statistics" was introduced only in 1942, by Wilks. However, the subject is much older, astronomers having long been interested in estimates of location beyond the sample mean. By early in the nineteenth century measures considered included the median, symmetrically trimmed means, the mid-range, and related functions of order statistics. In 1818, Laplace obtained (essentially) the distribution of the rth-order statistic in random samples and also derived a condition on the parent density under which the median is asymptotically more efficient than the mean. Other topics considered are of more recent origin: extreme-value theory and the estimation of location and scale parameters by order statistics.
The history is much older than the term which was introduced by Wilks for the ordered variates in a sample.For example, he asked for the probability that at least N_0 of N measurements on a second random sample will lie between the smallest and the largest value of a first sample of n taken from the same population.This is a nonparametric use of order statistics.This is in contrast to nonparametric tolerance limits and nonparametric confidence intervals, which do require the order statistics.As far back as the second century b.c, the Greek astronomer Hipparchus noticed variation in the length of the year. He estimated this as at most 3/4 day, "apparently by taking half the range of his observations" [Plackett (1958)]. This early date is an extreme outlier in the history of order statistics, but outliers in observational data have long drawn astronomers and others to pay special attention to extreme observations.
Order statistics have been used in a wide range of problems, including robust statistical estimation, detection of outliers, characterization of probability distribution, goodness-of-fit tests, entropy estimation, analysis censored samples; reliability analysis, quality control and strength of materials ...

In the realm of statistical analysis, order statistics play a pivotal role in the prediction of future observations and reconstruction of past unobserved values; that comes up naturally in several real life situations.
This dissertation endeavors to explore the prediction of order statistics. To do this and to embark on this adventure successfully, all of this research is based on the work of **Balakrishnan**, who has many work and researches on order statistics, as well as their prediction, [3, 14] and [24, 25]. The purpose of this statistical prediction is to infer the values of future ordered statistics based on the available observations.
The major aim of this thesis is to provide a comprehensive exploration of the Bayesian framework for predicting order statistics; while duly considering the frequentist viewpoint.
Throughout the subsequent sections we will shed light on the fundamental concepts of prediction of order statistics; while structuring it as follows around four chapters:
In the first chapter we revise some general and important notions on order statistics which will be useful in the forecasting process; for instance : distribution of some order statistics, using several references as: [3, 12, 21] and [24, 25].
The second one entitled the frequentist prediction of order statistics will provide a comprehensive overview of their usage in both parametric and nonparametric estimation of parameters; it will encompass point and interval prediction,[13, 18] and [19, 20]; concluding with informativeness measurement and underlying distribution selection; [16, 2].
Bayesian prediction of order statistics is the third chapter which contains important notions on the Bayesian paradigm, following by two big parts: Bayesian point prediction then Bayesian interval prediction, [1, 5, 6], and [7, 8, 9, 14].
Finally we conclude this academic journey with the fourth chapter which will be a simulations and real data studies of the foregoing obtained theoretical results, [15].
Indeed; the objectives that we aimed to achieve through this thesis are:

- Address the issue of prediction in statistics and step beyond the realm of estimation.

- Highlight the key aspects of both frequentist and Bayesian prediction, thereby establishing a connection between the two and elucidating the

differences between these two approaches.

- The practical part is not only meant to showcase the obtained results, but also to bring order statistics application from an abstract context to concrete reality.

To establish this essay; we have formulated the following research question: How one can predict the future order statistics in the different statistics perspectives, in several cases and under various conditions?

Chapter 1

General Notions on order statistics

Many years ago the study of order statistics was a curiosity of lot of mathematics researchers, because of their several applicability in different fields as meteorology, hydrology, business analysis ...
Order statistics is one of the fundamental tools of nonparametric statistics and statistical inference.
In this chapter we will give the main notions of order statistics, which will be used in the following chapters;using some important references like [21],[25] and [26].

1.1 Basic definitions and notations

Definition 1.1
Order statistic is that knowledge extracted of an observation which utilizes the rank or order of this observations.
If the initial random variables X_1,X_2 $\ldots$,,X_n *are independent and have a common distribution function (d.f.) F i.e; (*X_i*, are assumed to be statistically* idd*) are arranged in order of magnitude* [1] *and then written as:*

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

or, in more explicit notation:

$$X_{(1:n)} \leq X_{(2:n)} \leq \ldots \leq X_{(n:n)}$$

[1] magnitude of the data is also called "size ". For example, if every male in our sample was found to have a higher white cell count "WCC" than any female in the sample, we could say that the magnitude of the relation between the two variables (Gender and WCC) is very high in our sample. In other words, we could predict one based on the other (at least among the members of our sample).

then $X_{(i)}$ is called the ith order statistic $(i = 1, \dots n)$ in sample of size n from some distribution so we denote the order statistics by:

$X_{(1)}$ is the minimum, written: $\min_{1 \le i \le n} X_i$

$X_{(2)}$ is the 2nd smallest of $X_1, X_2 \dots X_n$

$\vdots \qquad = \qquad \vdots$

$X_{(n)}$ is the maximum, written: $\max_{1 \le i \le n} X_i$

$W = X_{(n)} - X_{(1)}$, is called range

Example 1.1 • *population peak .*

- *flow of a river .*
- *sports ranking order statistics can be used to determine the ranking of teams based on their performance, for example we can find the top 5 teams with the highest scores .*

Let $X_1, X_2, \dots X_{15}$ be a random sample of size 15 from the uniform distribution over the interval $(0, 1)$. Here are three different realization of such samples.

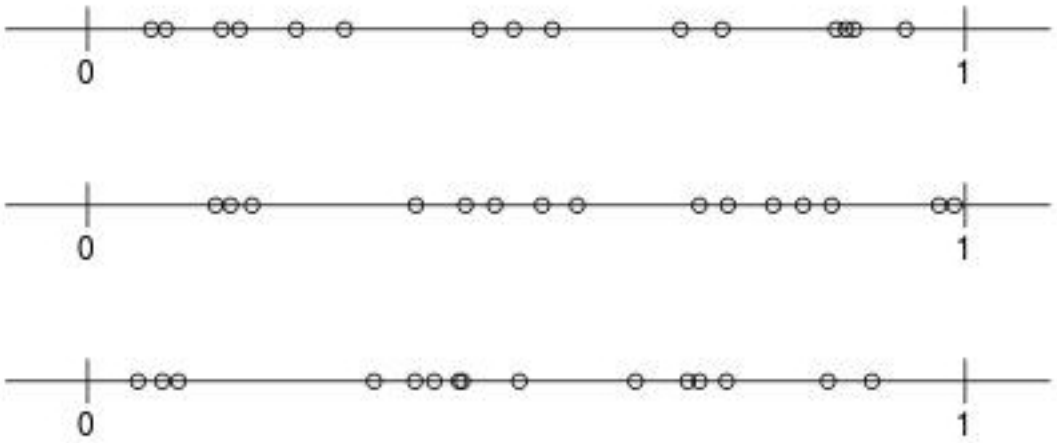


Figure 1.1: example of 3 samples

Consider the single smallest value from each of these three sample, as follows:

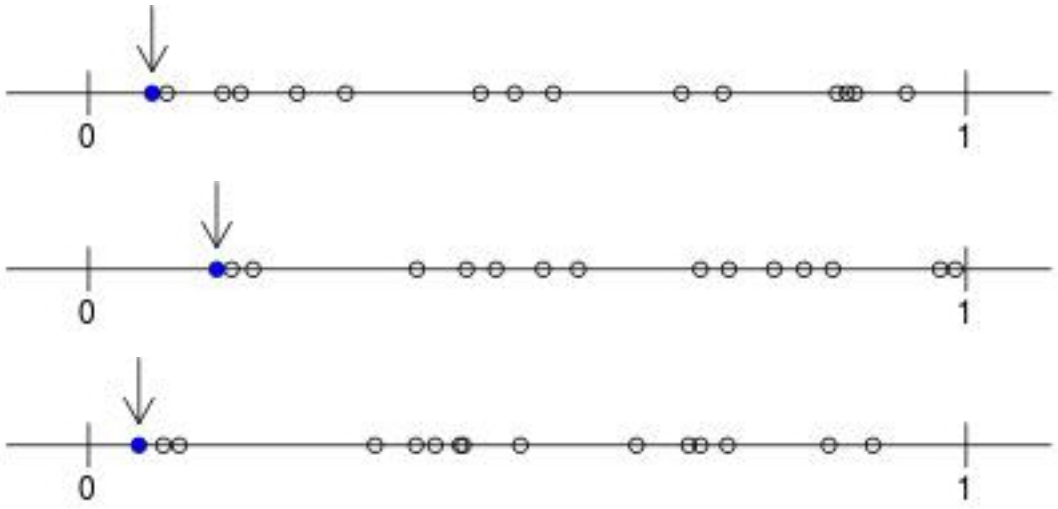


Figure 1.2: ordered sample

Collecting the minimums from the three samples onto a single graph we get

They are down towards zero!

Remark 1.1
The order statistics X_i are necessarily dependent, because of the inequality relations among them.
By order statistics we will mean either ordered variates or ordered observations, thus:

$$\left.\begin{array}{ll} X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} & \textbf{\textit{ordered variations}} \\ x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} & \textbf{\textit{ordered observations}} \end{array}\right\} = order statistics$$

1.2 Continuous Order Statistics

All the results and functions given in this section all about a continuous distributions. Let $X_1, X_2, \ldots, X_n$ be independent, absolutely continuous r.v. with common cdf $\mathbf{F}$ and pdf $\mathbf{f}(x)$ and let's $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be the corresponding order statistics.

1.2.1 Distribution of a single order statistic

Proposition 1.1
The cdf F of a single order statistic is given by the following formula:

$$\begin{aligned}\mathbf{F}_{X_{(k)}}(x) &= P(X_{(k)} \leq x) \\ &= \sum_{i=k}^{n} C\binom{n}{i}\mathbf{F}_X^i(x)(1-\mathbf{F}(x))^{n-i}.\end{aligned} \tag{1.1}$$

Also written as follows:

$$\begin{aligned}&\mathbf{P}(x < X_{(i)} \leq x + \delta x) = \frac{n!}{(i-1)!(n-i)!} \\ &[\mathbf{F}(x)]^{(i-1)}[1-\mathbf{F}(x+\delta x)]^{(n-i)} \\ &[\mathbf{F}(x+\delta x) - \mathbf{F}(x)] + \mathbf{O}((\delta x)^2)\end{aligned} \tag{1.2}$$

By deriving the previous formula we have the density function of the k*th order statistic $X_{(k)}$ and its writing as:*

$$\mathbf{f}_{X_{(k)}} = \frac{n!}{(k-1)(n-k)!}\mathbf{f}(x)\mathbf{F}^{k-1}(x)(1-\mathbf{F}(x))^{n-k} \tag{1.3}$$

Example 1.2
We set this example when $(X_i)_{1\leq i\leq n}$ n r.v. $\sim \mathbf{U}[0,1]$, the density function of the k*th order statistics using the formula mentioned in the 1.1 above provides the following result:*
$\forall 0 \leq x \leq$ *we have*

$$\begin{aligned}\mathbf{f}_{X_{(k)}} &= \frac{n!}{(k-1)!(n-k)!}\mathbf{1}x^{k-1}(1-x)^{n-k} \\ \mathbf{f}_{X_{(k)}} &= \frac{1}{\beta(k,n-k+1)}x^{k-1}(1-x)^{n-k}\end{aligned} \tag{1.4}$$

Then $X_{(k)} \sim \beta(k, n-k+1)$.

1.2.2 Joint Distribution of Two Order Statistics

Let us consider the order statistics X_i and $X_j (1 \leq i < j \leq n)$ their joint density is given:

Proposition 1.2
Let X_i for $i = 1, \ldots, n$, r.v. iid *with the distribution function* $\mathbf{F}$ *.Let $X_{(r)}$*

and $X_{(s)}$ 2 ordered r.v., r< s: their joint distribution is given by:

$$\mathbf{F}_{(X_{(r)},X_{(s)})}(x,y) = \begin{cases} \dfrac{n!\sum_{j=1}^{n}\sum_{i=r}^{j}\mathbf{F}^{i}(x)(\mathbf{F}(y)-\mathbf{F}(x))^{j-i}(1-\mathbf{F}(y))^{n-j}}{(i-1)!(j-i)!(n-j)!} & \text{if} \quad x<y \\ \mathbf{F}_{X_{(s)}}(y) \qquad\qquad \text{otherwise.} \end{cases} \tag{1.5}$$

We can also write from 1.5 the *pdf*

$$\mathbf{f}_{(X_{(r)},X_{(s)})}(x,y) = n!\mathbf{f}_{X_{(r)}}(x)\mathbf{f}_{X_{(s)}}(y)$$

$$\mathbf{f}_{(X_{(r)},X_{(s)})}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\mathbf{F}^{i-1}(x)(\mathbf{F}(y)-\mathbf{F}(x))^{j-i-1}$$

$$(1-\mathbf{F}(y))^{n-j}\mathbf{f}(x)\mathbf{f}(y) \qquad for -\infty < x < y < +\infty \tag{1.6}$$

1.2.3 Joint Distribution of k Order Statistics

All the results and functions given in this section all about a continuous distribution. Let $X_1, X_2, \ldots, X_n$ be independent, absolutely continuous r.v. with common pdf $\mathbf{f}$ and cdf $\mathbf{f}(x)$ and let's $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be the corresponding order statistics.
The joint pdf of $X_{(n_1)} \ldots X_{(n_k)}$ for $(n1 \leq n_2 \leq \ldots \leq n_k)$ is given by:

$$\mathbf{f}_{(X_{(1)},X_{(2)}\ldots X_{(k)})}(x_1,x_2,\ldots,x_k)) = \frac{n!}{(n_1-1)!(n_2-n_1-1)!\ldots(n-n_k)!}$$

$$\mathbf{F}^{n_1-1}(x_1)\mathbf{f}(x_1)[\mathbf{F}(x_2)-\mathbf{F}(x_1)]^{n_2-n_1-1}\mathbf{f}(x_2)\ldots[1-\mathbf{F}(x_k)]^{n-n_k}\mathbf{f}(x_k) \tag{1.7}$$

If we define: $x_0 = -\infty, X_{k+1} = +\infty, n_0 = 0, n_{k+1} = n+1$
In particular, the joint pdf of all n order statistics becomes simply:

Theorem 1.1

The joint density function of order statistics $X_1, X_2, \ldots, , X_n$ has the form

$$f_{(X_{(1)},X_{(2)}\ldots X_{(n)})}(x_1,x_2,\ldots,,x_n)) = \begin{cases} n!\prod_{i=1}^{n}\mathbf{f}(x_i) & \text{if } x_1 \leq x_2 \leq \ldots \leq x_n \\ 0 \qquad\qquad & \text{otherwise.} \end{cases}$$

Example 1.3

When the random sample X_i $(i = 1, 2, \ldots, , n)$ is from a standard exponential distribution with pdf $\mathbf{f}(x) = \exp^{-x}\, 0 \leq x < -\infty$. *we have the joint density of all n order statistics as:*

$$\mathbf{f}_{(X_{(1)},X_{(2)}\ldots X_{(n)})}(x_1, x_2, \ldots, , x_n)) = \begin{cases} n! \exp^{-\sum_{i=1}^{n} x_i} & \text{if } x_1 \leq x_2 \leq \ldots \leq x_n \\ 0 & \text{otherwise.} \end{cases}$$

1.3 Conditional Distribution of order statistics

Conditional distribution refers to the probability distribution of one variable given the value of another value in our case we consider sets of order statistics $(X_{(1)}, X_{(2)} \ldots, X_{(r-1)})$ and $(X_{(r+1)}, X_{(r+2)} \ldots, X_{(n)})$ become conditionally independent if $X_{(r)}$ is fixed; in other words this help us to understand the behaviour of any X_{r+1} when $X_{(r)}$ is fixed
From 1.7 it follows that, for $i \leq r < s \leq n$, the joint conditional pdf of $X_{(r+1)} \ldots X_{(s-1)}$ given $X(i) = x_i$ for $i \leq r$ and $i \geq s$ is:

$$\mathbf{f}_{X_{(r+1)}\ldots X_{(s-1)}|X_i = x_{i\, i\leq r\, i\geq s}(x_{r+1},\ldots X_{s-1}) = (s-r-1)! \prod_{j=r+1}^{s-1} \frac{\mathbf{f}(x_i)}{\mathbf{F}(x_s) - \mathbf{F}(x_r)} \qquad x_{(1)} < \ldots < x_{(n)} \tag{1.8}$$

Theorem 1.2

Let $X_1, X_2, \ldots, X_n$ be i.i.d r.v from a population with cdf F(x) and pdf f(x), and let $X_1 \leq X_2 \leq \ldots \leq X_n$ be the corresponding order statistics. Then the conditional distribution of X_j given X_i is the same as the distribution of the (j-i)th order statistic in a sample size (n-i) from a population with distribution F truncated on the left at x

Theorem 1.3

Let $X_1, X_2, \ldots, X_n$ be i.i.d r.v from a population with cdf F(x) and pdf f(x), and let $X_1 \leq X_2 \leq \ldots \leq X_n$ be the corresponding order statistics. Then the conditional distribution of X_i given X_j is the same as the distribution of the (i)th order statistic in a sample size (j-1) from a population with distribution F truncated on the right at x_j.

1.4 Distribution of some usual order statistics

Some order statistics used more than others, like the minimum, the maximum, the range and the median and in this section we will give their distribution function and some examples to illustrate them.

1.4.1 Distribution of the minimum

The game plan will be to relate the cdf of the minimum to the behavior of the individual sampled values $X_1, X_2, \ldots, X_n$ for which we know the pdf and cdf.
Imagine a random sample falling in such a way that the minimum is below a fixed value x. It might look like in these different way

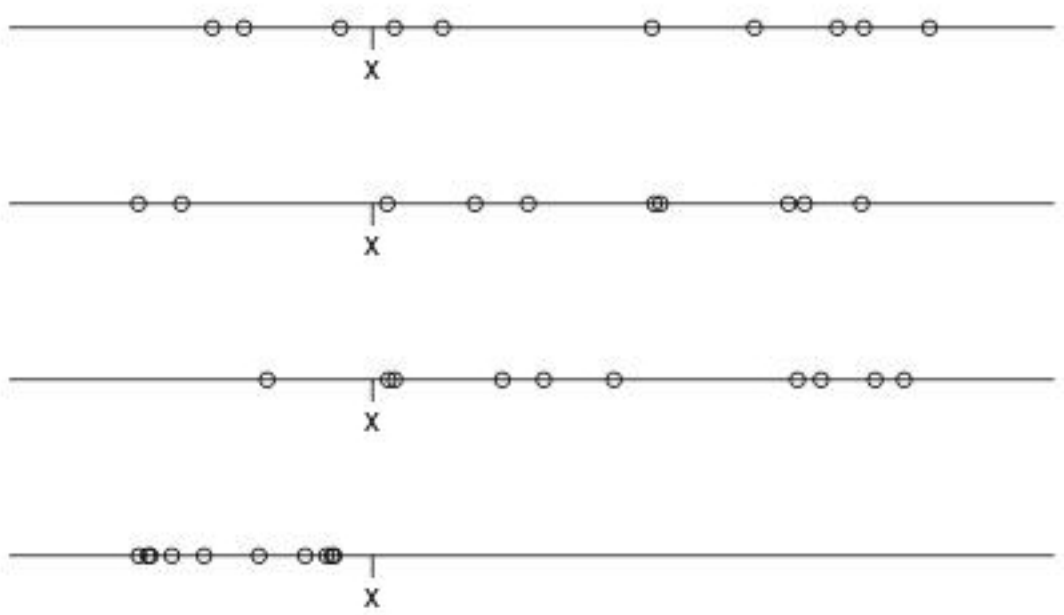


Figure 1.3: minimum distribution

There are many ways for the individual X_i to fall so that the minimum is less than or equal to x. we can not Considering all of the possibilities because there are a lot of them.On the other hand, the minimum is greater than x if and only if all the X_i are greater than x. So, it is easy to relate the probability $\mathbf{P}(X_{(1)}\, x)$ back to the individualX_i.
Then the distribution function of the minimum $\mathbf{F}_{X_{(1)}}(x)$ is given:

$X_{(1)} = \min_{1\leq i\leq n} X_i$.

$$\begin{aligned} \mathbf{F}_{X_{(1)}}(x) &= \mathbf{P}(X_{(1)} \leq x) = \mathbf{P}(\min X_i \leq x) \\ &= 1 - \mathbf{P}(\min X_i > x) \\ \textit{then}: \qquad \mathbf{F}_{X_{(1)}}(x) &= 1 - [1 - \mathbf{F}_X(x)]^n \end{aligned}$$

And the density function is written :

$$\begin{aligned} \mathbf{f}_{X_{(n)}}(x) &= [\mathbf{F}_{X_{(1)}}(x)]' \\ \mathbf{f}_{X_{(n)}}(x) &= n\mathbf{f}(x)[1 - \mathbf{F}(x)]^{n-1} \end{aligned}$$

Example 1.4

According to the uniform distribution $\mathbf{U}[0,1]$ *already mentioned in 1.4; we take* $\mathbf{f}(x) = I_{(0,1)}(x)$ *and:*

$$F(x) = \begin{cases} 0 \ , & x < 0 \\ x \ , & 0 \leq x < 1 \\ 1 \ , & x \geq 1. \end{cases}$$

So the pdf for the minimum in this case is

$$\mathbf{f}_{X_{(1)}}(x) = n[1 - x]^{n-1} I_{(0,\infty)}(x)$$

As we can notice this is the pdf for the Beta distribution with parameters 1 *and* n. *Thus we can write:* $X_{(1)} \sim Beta(1, n)$

1.4.2 Distribution of the maximum

Again consider our random sample $X_1, X_2, \ldots, X_n$ from a continuous distribution with pdf f and cdf F.As with the minimum, we will consider the cdf and try to relate it to the behavior of the individual sampled values $X_1, X_2, \ldots, X_n$. Imagine a random sample falling in such a way that the maximum is below a fixed value x.This will happen if and only if all of the X_i are below x; like a following

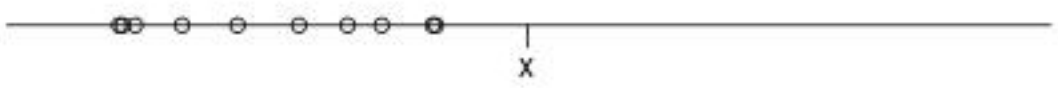


Figure 1.4: maximum sample

Then the distribution of the maximum $\mathbf{F}_{X_{(n)}}(x)$ is given: $X_{(n)} = \max_{1 \leq i \leq n} X_i$.

$$\begin{aligned} & \mathbf{F}_{X_{(1)}}(x) = \mathbf{P}(X_{(n)} \leq x) \\ & = \mathbf{P}(\max X_i \leq x) \\ & = \mathbf{P}(X \leq x) \\ \textit{then}: \quad & \mathbf{F}_{X_{(n)}}(x) = [\mathbf{F}_X(x)]^n \end{aligned}$$

And its density is given by:

$$\begin{aligned} \mathbf{f}_{X_{(n)}}(x) &= [\mathbf{F}_{X_{(1)}}(x)]' \\ \mathbf{f}_{X_{(n)}}(x) &= n\mathbf{f}(x)[\mathbf{F}(x)]^{n-1} \end{aligned}$$

Example 1.5

In the case of the random sample of size 15 *from the uniform distribution on* $(0, 1)$ *the pdf is*

$$\mathbf{f}_{X_{(n)}}(x) = nx^{n-1} I_{(0,1)}(x)$$

Which is the pdf of the Beta $(n, 1)$ *distribution and its graph is as following:*

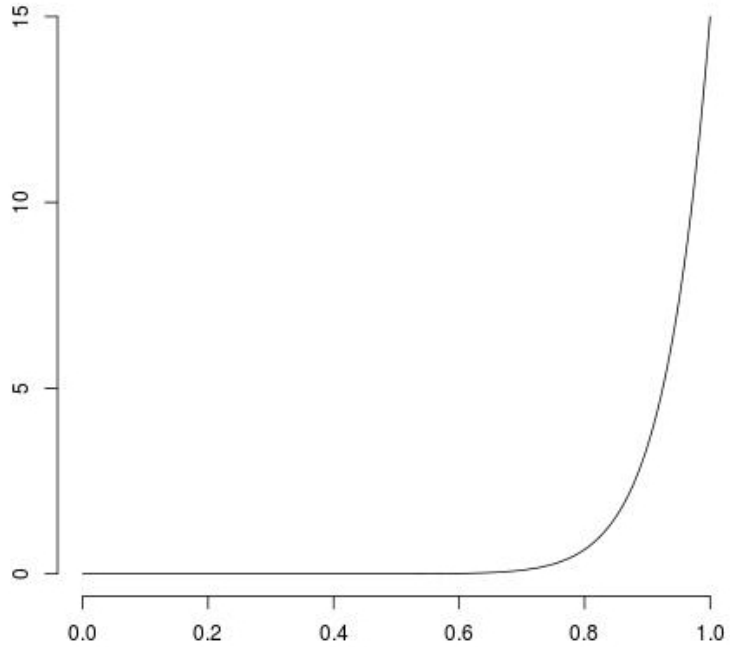


Figure 1.5: Maximum distribution graph

As expected, all most of the probability or "mass" for the maximum is piled up near the right endpoint of 1.

1.4.3 Distribution of the Median

As we have already mentioned there are some order statistics more used than others like the median.For example in population peak using the percentiles and at the time which the median and other percentiles exhibit a turning point transitioning from increasing to decreasing, could be considered an estimate of the population peak.
Consider the sample size n to be odd.Then from the equation 1.3 we have the pdf of the sample median $\tilde{X}_n = X_{(n+1)/2}$ to be:

$$\mathbf{f}_{\tilde{X}_n}(x) = \frac{n!}{\{[(n-1)/2]!\}^2}[\mathbf{F}(x)(1-\mathbf{F}(x))]^{(n-1)/2}\mathbf{f}(x) \qquad -\infty < x < \infty \tag{1.9}$$

We see at once that it is symmetric about 0, if the population distribution is symmetric about 0.

Example 1.6

In the case of standard uniform population, the pdf *of sample median given above become:*

$$\mathbf{f}_{\tilde{U}_n}(u) = \frac{n!}{\{[(n-1)/2]!\}^2}u^{(n-1)/2}(1-u)^{(n-1)/2} \qquad 0 < u < 1 \tag{1.10}$$

1.4.4 Distribution of the range of the order statistics

The range of order statistics is the difference between the largest and smallest order statistics in a sample.It is a measure of the variability of the sample; is also used in various fields and real situations; for example in finance can be used to analyze the variability or dispersion of stock prices over a specific time period, let's have a closer look about this process imagine a financial analyst want to study the performance of particular stock over the past year.He has daily closing prices for that stock recorded over the entire year.To assess the stock's volatility or price variability, he could calculate the range of the ordered values.
Now let's take a view about the calculation of the range of order statistics. We have the joint *pdf*:

$$\mathbf{f}_{(X_{(r)},X_{(s)})}(x,y) = n!\mathbf{f}_{X_{(r)}}(x)\mathbf{f}_{X_{(s)})}(y)$$

$$\mathrm{f}_{(X_{(r)},X_{(s)})}(x,y) = \frac{n!}{i!(j-i)!(n-j)!}\mathbf{F}^{i-1}(x)(\mathbf{F}(y)-\mathbf{F}(x))^{j-i-1}(1-\mathbf{F}(y))^{n-j}\mathbf{f}(x)\mathbf{f}(y)$$

$$\text{for } -\infty < x < y < +\infty$$

Now denoting $\mathbf{W} = X_{(n)} - X_{(1)}$ for the sample range.The joint density function of $X_{(1)} and \mathbf{W}$ from 1.6 to be:

$$\mathbf{g}(x_1, w) = n(n-1)[\mathbf{F}(x_1+w) - \mathbf{F}(x_1)]^{n-2}\mathbf{f}(x_1)\mathbf{f}(x_1+w) \tag{1.11}$$

$$\text{for } -\infty < x_1 < +\infty$$

$$\text{and } 0 < w < +\infty$$

By integrating out the formula in 1.11 we obtain the cdf of the range $\mathbf{W} \forall 0 < w < +\infty$

In exponential case: using the formula above we will get the range distribution as follows:

$$\mathbf{g}_w(w) = n(n-1)\int_{-\infty}^{+\infty}[\mathbf{F}(x_1+w) - \mathbf{F}(x_1)]^{n-2}\mathbf{f}(x_1)\mathbf{f}(x_1+w)\,dx_1 \tag{1.12}$$

Then the cdf of the range is :

$$\mathbf{G}_W(w_0) = \mathbf{P}(W \le w_0)$$
$$\mathbf{G}_W(w_0) = n(n-1)\int_0^{w_0}\int_{-\infty}^{+\infty}[\mathbf{F}(x_1+w) - \mathbf{F}(x_1)]^{n-2}\mathbf{f}(x_1)\mathbf{f}(x_1+w)\,dx_1\,dw$$
$$0 \le w_0 < +\infty$$

$$\mathbf{G}_W(w_0) = \int_0^{w_0}(n-1)(1-\exp^{-w})^{n-1}\exp^{-w}\,dw$$

$$\mathrm{G}_W(w_0) = (1-\exp^{-w_0})^{n-1} \qquad \forall 0 \le w_0 < +\infty$$

To see more about the integration process to obtain this formula check [25].

Remark 1.2

The result on the sample range presented above may be generalized for the sub-range: $\mathbf{W}_{i,j} = X_{(j)} - X_{(i)}$ $1 \leq i < j \leq n$. *using the joint distribution from 1.6.*
we have this result for $0 < w_{i,j} < +\infty$ *:*

$$\mathbf{g}_{w_{i,j}} = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_{-\infty}^{+\infty} [\mathbf{F}(x)]^{i-1}[\mathbf{F}(x+w_{i,j}) - \mathbf{F}(x)]^{j-i-1} * [1- F(x+w_{i,j})]^{n-j}\mathbf{f}(x)\mathbf{f}(x+w_{ij})\,dx \quad (1.13)$$

Example 1.7

For the uniform [0,1]distribution, we obtain the density function of $w_{i,j} from 1.13$. *For* $0 < w_{i,j} < 1$

$$\mathbf{g}_{w_{i,j}} = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_{0}^{1-w_{i,j}} x^{i-1}w^{j-i-1}(1-w_{i,j}-x)^{n-j}\,dx$$

$$g_{w_{i,j}} = \frac{n!}{(j-i-1)!(n-j+1)!} w^{j-i-1}(1-w_{i,j}-x)^{n-j+1}$$

In addition to the finance example discussed above, it's important to note that by analyzing the range of order statistics; we can quickly get an idea of how much the stock's price varied over the year.This information can be valuable for investors and traders who are interested in understanding the potential risks and rewards associated with particular stock.

1.5 Order statistics and record value

The theory of records is tied very closely with the theory of extreme order statistics.
Record values are the extreme values in a data set that represent a new high or low value as you progress through the data; there are two types of record values: upper records (new high values) and lower records(new low values).
Let $X_1, X_2, \ldots$ *be a sequence of independent identically distributed random variables with common distribution function* $\mathbf{F}, and X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, $n = 1, 2 \ldots$ *be the corresponding order statistics.*

We will assume that $\mathbf{F}$ *is continuous so that ties are not possible.There are a few results available on record values corresponding to discrete distributions, the more elegant results are associated with the continuous case.*
Denote:

$$M_n = \max\{X_{(1)}, X_{(2)}, \dots X_{(n)}\}$$

and consider the increasing sequence of these sequential maximal value

$$X(1) \leq X(2) \leq \dots X(i) \leq X(i+1) \leq \dots \leq X(j) \dots$$

An observation X(j) will be called a record (precisely an upper record) if it exceeds in value all preceding observation. Let us fix the times, when signs of the strong appear in this sequence such time corresponding to the situation when $X(j) > X(i)$ *for* $i < j$*. The lower records are analogously defined.Usually X(1) is taken as the first record.*

- *The sequence of record times* $\{T\}_{n=0}^{\infty}$ *is defined as follows:*

$$T_0 = 1 \qquad \textit{with probability 1}$$

And for n>1

$$T_n = \min\{j : j > T_{n-1}, X_j > X_{T_{n-1}}\}$$

- *The corresponding record value sequence* $\{R_n\}_{n=0}^{\infty}$ *is defined by:*

$$R_n = X_{T_{n'}}, n = 0, 1, 2 \dots$$

- *Inter-record times* Δ_n *are defined by:*

$$\Delta_n = T_n - T_{n-1} n = 1, 2 \dots$$

- *The record counting process* $\{N_n\}_{n=1}^{\infty}$ *where:*

$$N_n = \{ \textit{ number of records among } \} \qquad \{X_1, X_2 \dots\}$$

Denote $X(1) < X(2) \dots$ *the record value in the sequence* $X_1, X_2 \dots$ *and let* $1 = T_0 < T_1 < T_2 < \dots$ *be the corresponding record times, introduce also record indicators* ξ_n, $n = 1, 2, \dots$ *which take values 0 and 1, mark the appearance of record values, that is:* $\xi_n = \begin{cases} 1 & \textit{if } X_n > X_{n-1} \\ 0 & \textit{otherwise.} \end{cases}$

Example 1.8

i) Tracking daily temperatures,each time the temperature reaches a new high value, it is considered an upper record; similarly, when it reaches a new low value, it is a lower record.

ii) meteorologists use record values to monitor and assess the impact of extreme weather events,study climate trends and make informed predictions about the future weather patterns. record rainfall can trigger emergency responses, resource allocation and risk assessment for communities.
For instance let's say we have a collected daily rainfall data (in millimeters) for a month in a given region the values are:
$$15, 12, 18, 10, 25, 8, 22, 5, 28, 30, 20, 14, 6, 32, 9, 35, 11, 27, 7, 33$$
Let's identify the record values:

- *Record high rainfall:* 35 *mm (new high)*
- *Record low rainfall:* 5 *mm (new low).*

1.6 Markov property of order statistics

Markov property of Order statistics is also under investigation.It turns out that Order statistics from a Markov chain if the underlying distribution is continuous, while this property fails as a rule for distribution functions having discontinuity points.
The Markov property is not valid for order statistics if the underlying distribution function has three jumps points or more, because in this situation there exists one point "a" at least such that:
$\mathbf{P}[X = a] > 0 \quad ; \mathbf{P}[X < a] > 0 \quad ; \mathbf{P}[X > a] > 0$
Hence it remains to discuss distributions with one or two atoms.Which coincide with end-point. $\alpha = \inf\{x : \mathbf{F}(x) > 0\}$ *or* $\beta = \sup\{x : \mathbf{F}(x) < 1\}$

Example 1.9

Let X have a degenerate distribution , say $\mathbf{P}[X = a] = 1$ *. Then it is evident that order statistics from the Markov property.*

In particular since the conditional distribution in 1.8 is free of x_i for $i < r$ and $i > s$ $X_{(r+1)} \dots X_{(s-1)}$ are independent of $X_{(1)} \dots X_{(r-1)}$ and $X_{(s+1)} \dots X_{(n)}$

when $X_{(r)}$ and $X_{(s)}$are given upon conditioning only on the lower order statistics 1.8 leads us to conclude that:

$$\mathbf{f}_{X_{(r+1)}\ldots X_{(n)}|X_{(1)}=X_1\ldots x_{(r)}} = X_r(x_{(r+1)}, \ldots x_n)$$
$$\mathbf{f}_{X_{(r+1)}\ldots X_{(n)}|X_{(r)}=X_r(x_{(r+1)},\ldots x_n)}$$

Which establish that the order statistics in a sample from continuous parent from Markov chain. Further:

$$\mathbf{f}_{X_{(r+1)}|X_{(r)}=x}(y) = (n-r)\left\{\frac{1-\mathbf{F}(y)}{1-\mathbf{F}(x)}\right\}^{n-r-1}\frac{\mathbf{f}(y)}{1-\mathbf{F}(x)} \qquad y > x$$

Provides the transition density.
In practical terms, the Markov property of order statistics means that once we know the current order statistics, we don't need to consider the entire history of previous order statistics to predict the distribution of the next order statistic.This property is particularly useful in various applications where order statistics encountered such as extreme value theory (EVT).
The Markov property of order statistics is often utilized in EVT to simplify calculations and make predictions about extreme events.For instance:

Example 1.10

Suppose a study of extreme rainfall events in a specific region.We have historical data of maximum annual rainfall (in mm) recorded over a several decades, using the EVT to model and predict the likelihood of even higher maximum annual rainfall in the future.
We have the following dataset of maximum annual rainfall (in mm) for the past 30 *years:*

215, 180, 230, 190, 205, 220, 195, 200, 210, 225, 235, 245, 198, 204, 215,
250, 260, 240230, 210, 190, 280, 210, 235, 260, 270, 245, 225, 215, 230

Using the EVT we can model the distribution of the maximum annual rainfall using extreme value distributions such as GEV [2] *distribution .*
For example, suppose we have calculated that the current maximum annual rainfall is 280 *mm; by applying the Markov property of order statistics we can estimate the probability of observing a maximum annual rain-*

[2] generalized extreme value

fall greater than 280 *mm in the next year without needing to consider the entire history of previous maximums.*

Remark 1.3

It seems that the following hypotheses must be true.

a) Let the d.f. F have one discontinuity point only, which coincides with the left end-point $\alpha = \inf\{x : \mathbf{F}(x) > 0\} > -\infty$ *or* $\beta = \sup\{x : \mathbf{F}(x) < 1\} < +\infty$ *of the distribution support.Then order statistics* $X_{(1)}, X_{(2)} \ldots and X_{(n)}$ *possess the Markov property.*

b) If α *and* β *are the only jump points of c.d.f.* $\mathbf{F}$*, then order statistics form a Markov chain.*

c) In contrast to the continuous case, the X_i*'s do not form a Markov chain.*

1.7 Generalized order statistics

Generalized order statistics are used in various fields, such as reliability analysis, survival modeling, EVT and finance to better understand the distribution and characteristics of data when it follows a specific distribution other than uniform. We will present the so-called generalized order statistics which will be noted throughout this section (gos).For univariate absolutely continuous distributions.

Definition 1.2

Kamps (1995) introduced the generalized order statistics (gos).The order statistics, record values and sequential order statistics are special cases of this generalized order statistics.
Suppose $X(1, n, m, k), \ldots X(n, n, m, k)$ *(*$1 \leq k$*,* m *is a real number) are* n *generalized order statistics.*
The joint pdf $\mathbf{f}_{1\ldots n}(x_1, \ldots, x_n)$ *can be written:*

$$f_{1\ldots n}(x_1,\ldots,x_n)=\begin{cases} k\prod_{j=1}^{n-1}\gamma_j\prod_{i=1}^{n-1}(1-\mathbf{F}(x_i))^m\mathbf{f}(x_i)(1-\mathbf{F}(x_n))^{k-1}\mathbf{f}(x_n) & \text{for} \\ \qquad\qquad \mathbf{F}^{-1}(0)<x_1<\ldots<x_n<\mathbf{F}^{-1}(1) & \\ 0 \qquad\qquad \text{otherwise.} & \end{cases}$$

(1.14)

where : $\gamma_j = k+(n-j)(m+1)$ *if* $m=0$ *and* $k=1$ *then* $X(r,n,m,1)$ *reduces to the ordinary rth order statistics.*
If $k=1$ *and* $m=-1$ *then gos is the joint pdf of the first n upper record values of the iid r.v. with distribution function* $\mathbf{F}(x)$ *and the corresponding probability density function f (x).*

Example 1.11

Let $\mathbf{F}_{r,n,m,k}(x)$ *be the distribution function of* (r,n,m,k) *the following inequalities are equal:*

$$\mathbf{F}_{r,n,m,k}(x)=I_\alpha(r,\frac{\gamma_r}{m+1}) \qquad if \qquad m>-1 \tag{1.15}$$

And

$$\mathbf{F}_{r,n,m,k}(x)=\Gamma_\beta(r) \qquad if \qquad m=-1 \tag{1.16}$$

Where:

$$\alpha(x)=1-(\bar{\mathbf{F}}(x))^{m+1}$$

$$\beta(x)=-k\ln(\bar{\mathbf{F}}(x))$$

$$I_x(p,q)=\frac{1}{B(p,q)}\int_0^x u^{p-1}(1-u)^{q-1}\,du \quad ,B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

$$and \qquad \Gamma_x(r)=\int_0^1\frac{1}{\Gamma(r)}u^{r-1}\exp^{-u}\,du$$

See [1] for the proof.

Example 1.12 ***gos of the exponential distribution***

Let X be a r.v. whose pdf f is given by:

$$\mathbf{f}(x)=\begin{cases}\sigma^{-1}\exp(-\sigma^{-1}x) & for \quad x>0,\sigma>0\\ 0 & otherwise.\end{cases}$$

We will denote $X \in E(\sigma)$ if the pdf of X is of the form given by 1.15. If $X_i \in E(\sigma), i = 1, 2, \ldots$ and X's are independent, then $X(1, n, m, k) \in E(\sigma/\gamma_1)$ and:

$$X(r+1, n, m, k) \xrightarrow{d} X(r, n, m, k) + \sigma \frac{W}{\gamma_{r+1}}$$

where W is independent of X(r,n,m,k) and $W \in E(1)$.

Example 1.13 ***gos in EVT***

Suppose we want to analyze the distribution of daily maximum temperatures (in degrees Celsius) recorded in a given city.Instead of assuming a uniform distribution, the temperatures follow a specific distribution such as the "Weibull" distribution.
we have a collection of daily maximum temperatures data for a month, and we want to find gos based on Weibull distribution.
Let's say the daily maximum temperatures for the month are as follows:

$$32, 28, 30, 34, 31, 33, 29, 35, 36, 37$$
$$31, 33, 30, 38, 33, 39, 28, 36, 32, 32, 35$$

To find the gos we would follow these steps:

a) Fit the Weibull distribution to the data using statistical software or methods.

b) Obtain the cumulative distribution function (CDF) of the fitted Weibull distribution.

c) Calculate the inverse CDF (quantile function) for various probabilities (e.g.; $(\mathbf{P} = 0.1, 0.2, \ldots, 0.9)$).

d) The inverse CDF values for each probability represent the generalized order statistics corresponding to those probabilities.

```
Probability: 0.1  - Generalized Order Statistic: 28.25473
Probability: 0.2  - Generalized Order Statistic: 30.18547
Probability: 0.3  - Generalized Order Statistic: 31.4586
Probability: 0.4  - Generalized Order Statistic: 32.46985
Probability: 0.5  - Generalized Order Statistic: 33.35462
Probability: 0.6  - Generalized Order Statistic: 34.18474.
```

Chapter 2

Frequentist Prediction Of order statistics

Statistical prediction is an important part of decision-making, and many of our decisions are based on the prediction of future unknown events.The order statistics play an important role in the prediction of future observations and reconstruction of past unobserved values, they are major item in several optimal inference procedures. In quite a few instances the order statistics become sufficient statistics and, thus, provide minimum variance unbiased estimators (MVUE) of and most powerful test procedures for the unknown parameters.
Order statistics appear in a natural way in inference procedures when the sample is censored.They also provide some quick and simple estimators which are quite often highly efficient.
In this chapter based on [15, 18] and [19], we begin with some definitions about censoring data, some important estimators based on order statistics and its proprieties.Then we will give important estimators of order statistics, point and interval estimation will be discussed, also some new methods of prediction which have been practically verified.
In our work we confine our selves to location-scale family of continuous pdf's of the form

$$\mathbf{f}(x|\theta) = (1/\sigma)g((x-\mu)/\sigma). \tag{2.1}$$

g is known pdf generates a distribution with finite second moment; but we will give some important remarks about the discrete case.Pay attention that location and scale doesn't mean specially the mean and the standard deviation.

2.1 Censoring data

A very important characteristic in a set of data is the presence of censored observations which mean the sampling is from an unrestricted population but the exact values of some observations in the sample are unknown.
Let us consider a life-testing experiment where n items are kept under observation until failure, these items could be some systems, components, or they could be patients put under certain drug or clinical conditions.Suppose the life lengths of these n items are i.i.d. random variables with a common absolutely continuous cdf $\mathbf{F}(x;\theta)$ *and pdf* $\mathbf{f}(x;\theta)$*, where* θ *is the unknown parameter.Then we have a random sample* $(x_1, \ldots x_n)$ *from the cdf* $\mathbf{F}(x;\theta)$*.Note, that these values are recorded in increasing order of magnitude; that is, the data appear as the vector of order statistics.For some reason or other (because a barrier has been imposed either by the observer or the measuring process), suppose that we have to terminate the experiment before all items have failed.*
Censored data are not to be confused with the term truncated which is applied to observations obtained when the sampling is from a restricted population .The numbers of observations that would have occurred in a sample above and below the truncation point are unknown; for example to study heights of American military personnel, the sample would be from a truncated population of heights i.e; members of the military forces were previously screened to eliminate those falling below minimum or above maximum allowable heights.

2.1.1 Censoring data type I

Suppose it is decided to terminate the experiment at a predetermined time t, so that only the failure times of the items that failed prior to this time are recorded.The data so obtained constitute a Type I censored sample also called time censoring.It corresponds to right censoring in which large observations are missing. Clearly, the number of observed order statistics "r" is a random variable; it could even be 0.

The likelihood of this type of censoring is given by:

$$\left.\begin{aligned}\mathbf{L}(\theta|r,x) &= \frac{n!}{(n-r)!}\mathbf{f}(x_1;\theta)\ldots\mathbf{f}(x_r;\theta)\{1-\mathbf{F}(t,\theta)\}^{n-r}, \\ & \qquad x_1 < x_2 < \ldots < x_r < t, \quad 0 < r \leq n \\ &= \{1-\mathbf{F}(t;\theta)\}^n, \qquad r = 0, \quad t < x_1.\end{aligned}\right\}$$

x is the vector of observed order statistics.

Remark 2.1

Having two distinct forms for $\mathbf{L}(\theta)$ *complicates the study of the finite sample properties of the MLE of* θ *.*

2.1.2 Censoring data type II

If the experiment is terminated at the rth failure, that is, at time $X_{(r)}$*, we obtain Type II censored sample. Here r is fixed, while* $X_{(r)}$*, the duration of the experiment, is random.This type is also called ("failure censoring"). The likelihood function can be written as :*

$$\left.\begin{aligned}\mathbf{L}(\theta|x) &= \frac{n!}{(n-r)!}\mathbf{f}(x_1;\theta)\ldots\mathbf{f}(x_r;\theta)\{1-\mathbf{F}(x_r;\theta)\}^{n-r} \\ & \qquad x_1 < x_2 < \ldots < x_r\end{aligned}\right\}$$

As in the Type I censored case, the likelihood above corresponds to a right-censored sample; that is, large values are censored.One can have left censoring wherein smaller values are censored or one can have censoring of multiple regions.When there is just left and right censoring, the name used often is double censoring.

Remark 2.2

- *There is a difference between the censored data and truncated ones.Truncated observations are obtained when the sampling is from restricted . The number of the observations that would have occurred in sample above and below the truncation point are known.*

- *Other types of censoring data exist like: Random Censoring; Progressive Censoring, interval censoring ...*

2.2 Prediction of order statistics

Prediction problems come up naturally in several real-life situations. They can be classified under two categories:

a) the random variable to be predicted comes from the same experiment so that it may be correlated with the observed data.

b) It comes from an independent future experiment.

With order statistics both of these situations are feasible.

- ***One-sample problem***
  Suppose a machine consists of n components and fails whenever k of these components fail. Observations consist of the first r failure times, and the goal is to predict the failure time of the machine. Assuming the components' life lengths are i.i.d., we have a prediction problem involving a Type II censored sample, and it falls into category i mentioned above.
- ***Two-sample problem*** *A manufacturer of certain equipment is interested in setting up a warranty for the equipment in a lot being sent out to the market. Using the information based on a small sample, possibly censored, the goal is to predict and set a lower prediction limit for the weakest item in the lot. This falls into category ii above.*

2.2.1 Preliminaries

It may be of interest to predict in one sample problem: (a.) The time at which all the components will have failed.(b.) certain sample quantile.(c.) The mean failure time of the unobserved lifetimes.
In the two sample problem we may predict:

(a.) the range. (b.) quartiles. (c.) the smallest lifetime.
Let $\boldsymbol{X} = (X_{(1)}, X_{(2)}, \dots X_{(m)})'$ *and* $\boldsymbol{Y} = (Y_{(1)}, Y_{(2)}, \dots Y_{(n)})'$ *be the order statistics of two independent random sample from the family of continuous pdf* $\mathbf{f}(x|\theta)$, θ *is unknown parameter vector.Our goal is to predict from some observation of the components of* $\boldsymbol{X}$*:* $\boldsymbol{X}_1$ *:* $\boldsymbol{X} = (X_{(1)}, X_{(2)}, \dots X_{(r)})'$ *the remaining components of* $\boldsymbol{X}$*:* $\boldsymbol{X}_2 = (X_{(r+1)}, X_{(r+2)}, \dots X_{(m+r)})'$ *called*

one sample problem or predict $\boldsymbol{Y}$ *the two sample problem; this type of prediction arises generally in life testing.* $\boldsymbol{X}$ *represent the ordered life-times of m components simultaneously put on the test, if the test stopped after the rth failure so* $\boldsymbol{X}_1$*represents the only available data; or vector of observed failure times.*
Let $\mathbf{U}$ *and* $\mathbf{W}$ *be vectors of r.v. whose joint distribution depends on unknown parameters* θ*. Having observed* $\mathbf{U} = u$ *it is desired to predict* $T = T(\mathbf{W})$ *some function of* $\mathbf{W}$*; let* $\hat{T} = \hat{T}(\mathbf{U})$ *a function of* $\mathbf{U}$ *denote a generic predictor of T.*
Good choices for $\hat{T}$ *may be defined relative to the loss* $L(\hat{T}, T)$*, when* $\hat{T}$ *is used to predict T, an optimal choice of for* $\hat{T}$ *is finding that function if possible which minimizes:*

$$\mathbf{E}\{L[\hat{T}(U), T(W)]\}$$

$\mathbf{E}$ *expectation over all joint realizations of* $\mathbf{U}$ *and* $\mathbf{W}$*.*
We associate the vector $\mathbf{U}$ *with elements of* $\boldsymbol{X}_1$*, associate* $\mathbf{W}$ *with:*

$$\begin{cases} \boldsymbol{X}_2 & \text{in the one sample problem} \\ \boldsymbol{Y} & \text{in the two sample problem.} \end{cases}$$

Remark 2.3

We consider the function being predicted $T = T(\mathbf{W})$ *to be linear.*

2.3 Point prediction

In this section we are interested to finding a point predictor of single order statistic $X_{(s)}$ *from a sample* $X_{(1)} \leq \ldots X_{(r)}$ $r < s$*. The most important and efficient methods for constructing a point predictor of order statistic* $X_{(s)}$ *based on the observed* $X_{(1)} \leq \ldots X_{(r)}$ $r < s$ *will be discussed in the following.Using [22] as main reference.*

2.3.1 Best unbiased predictor BUP

Let $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ *be order statistics from a random sample from an absolutely continuous cdf* $\mathbf{F}(x;\theta)$ *having pdf* $\mathbf{f}$*, suppose we observe only* $X_1 = X_{(1)}, \ldots, X_{(r)}$ *the goal is to predict* $X_{(s)}$*.Since* $\mathbf{F}$ *is continuous, the conditional distribution of* $X_{(s)}$ *given* $X_{(r)}$ *is:*

$$\mathbf{f}_{X_{(s)}|X_1}(x) = \mathbf{f}_{X_{(s)}|X_{(r)}}(x) \tag{2.2}$$

This is Markov property 1.6 using 2.2 the BUP of $X_{(s)}$*is the conditional expectation* $\mathbf{E}(X_{(s)}|X_1)$ *which can be written* $\mathbf{E}(X_{(s)}|X_{(r)})$ *according to the Markov property, if* θ *is unknown parameter we use its estimator.*

2.3.2 Linear prediction

using the precedent notation 2.2.1 ,$\boldsymbol{X}_1$ *is observed* $\boldsymbol{X}_2$, $\boldsymbol{Y}$ *have yet to be observed or censoring.Let*

$$W_{(s)} = \begin{cases} \boldsymbol{X}_2 & \text{in the one sample problem} \\ \boldsymbol{Y} & \text{in the two sample problem} \end{cases}$$

We will focus on the prediction of $W_{(s)}$ *single component of* $\mathbf{W}$*.A linear predictor* $\hat{W}_{(s)} = \xi' \boldsymbol{X}_1$ *is called the best linear unbiased predictor (BLUP) of* $W_{(s)}$ *if:*

$$\begin{aligned} &\mathbf{E}(\hat{W_{(s)}}) = \mathbf{E}(W_{(s)}) \quad \textit{and} \\ &\mathbf{E}(\hat{W_{(s)}} - \mathbf{E}(W_{(s)}))^2 \quad \textit{is minimum.} \end{aligned}$$

ξ *is a vector of constants.Denotes the vectors of the standardized order statistics* $\boldsymbol{Z}_{(x_i)}$ $\boldsymbol{Z}_{(\boldsymbol{Y})}$*:*

$\boldsymbol{Z}_{(x_i)} = (\boldsymbol{X}_i - \mu)/\sigma$
$\boldsymbol{Z}_{(\boldsymbol{Y})} = (\boldsymbol{Y}_i - \mu)/\sigma$
α_i *the expected values of the standardizes order statistics*
The covariance matrix is given as : $\sigma^2\mathbf{V} = \sigma^2(v_{ij})$
$Cov(Z_{(i)}, Z_{(s)}) = w_{(s)}, \qquad w' = (w_1, \nu_2, \ldots, w_r)$

The BLUP of $W_{(s)}$ *can be written:*

$$\hat{W}_{(s)} = \begin{cases} \hat{X}_{(s)} = (\hat{\mu} + \hat{\sigma}\alpha_{s,m}) + w'\Omega_{11}(\boldsymbol{X}_1 - \hat{\mu}\mathbf{1}_1 - \hat{\sigma}\alpha_1) \textit{in the one sample problem} \\ \hat{\boldsymbol{Y}}_{(s)} = \hat{\mu} + \hat{\sigma}\alpha_{s,n} \textit{in the two sample problem.} \end{cases}$$

Note that Ω *is the inverse of the covariance matrix*
the above point predictor is accompanied by an estimates of its precision:

$$MSE(\hat{W}_{(s)}) = \begin{cases} \sigma^2\{\nu_{ss} - w'\Omega_{11}w + c_{11}\} \quad \textit{in the one sample problem} \\ \sigma^2\{\alpha_1'\Omega\alpha_1 + \alpha_{s,n}^2 \mathbf{1'}_1\Omega\mathbf{1}_1 - 2\alpha_{s,n}1'_1\Omega\alpha_1\}/\Delta \quad \textit{in the two sample case.} \end{cases}$$

$$c_{11} = var\{(1 -' \Omega_{11}\mathbf{1}_1)\hat{\mu} + (\alpha_{s,m} -')\Omega_{11}\alpha_1)\hat{\sigma}\}/\sigma^2 \tag{2.3}$$

$$\Delta = (\mathbf{1}'_1\Omega_{11}\mathbf{1}_1)(\alpha'_1\Omega_{11}\alpha_1) - (\mathbf{1}'_1\Omega_{11}\mathbf{1}_1)^2 \tag{2.4}$$

Remark 2.4

We point out that in the two sample case, the BLUP of $Y_{(s)}$ *is just the BLUE of* $\mathbf{E}(Y_{(s)})$.

We say that the linear combination of components of $\hat{\boldsymbol{W}}$ *is the BLUP of* $\boldsymbol{W}$ *if and only if:*

- $\mathbf{E}((W_{(s)}) - (\hat{W}_{(s)})) = 0$
- $\mathbf{E}(W_{(s)} - \mathbf{E}(\hat{W}_{(s)}))$ *is a minimum*

It follows from [18] that the BLUP is:

$$\hat{W}_{(s)} = \hat{\mathbf{E}}(W_{(s)}) + w'\mathbf{V}^{-1}(\boldsymbol{X} - \hat{\mu}\mathbf{1} - \hat{\sigma}\alpha) \tag{2.5}$$

Similarly in large samples we say that the linear combination $\tilde{W}_{(s)}$ *of the data* $\{X_{(i)}\}_{i=1}^{k}$ *the ABLUP of* $W_{(s)}$ *if and only if in the asymptotic distribution of the sample quantiles.*

- $\mathbf{E}((W_{(s)}) - (\tilde{W}_{(s)})) = 0$
- $\mathbf{E}(W_{(s)} - \mathbf{E}(\tilde{W}_{(s)}))$ *is a minimum*

Then the ABLUP of $W_{(s)}$ *is:*

$$\tilde{W}_{(s)} = \tilde{\mathbf{E}}(W_{(s)}) + \nu'\Gamma^{-1}(\boldsymbol{X} - \hat{\mu}\mathbf{1} - \tilde{\sigma}v) \tag{2.6}$$

$v' = (v_1, \ldots, v_k)$,$\lambda_i = \mathbf{F}(v_i)$; $\Gamma = (\gamma_{ij}) = \{\lambda_i(1 - \lambda_j)/(\mathbf{f}_i\mathbf{f}_j)\}$
$(1 \leq i \leq j \leq k)$.

Remark 2.5

A linear predictor $\tilde{\boldsymbol{W}} == \xi'\boldsymbol{X}_1$ *is called the best linear invariant predictor (BLIP) of* $W_{(s)}$ *if* $\mathbf{E}(\tilde{\boldsymbol{W}}_{(s)} - \boldsymbol{W}_{(s)})^2$ *is a minimum in the class of the linear predictors of* $\boldsymbol{W}_{(s)}$ *whose MSE is proportional to* σ^2 *; this condition imposes the constraint* $\xi'\mathbf{1}_1 = 1$ *on the coefficient vector* ξ*.we say that* $\xi'\boldsymbol{X}_1$ *is an unbaised predictor of* $\boldsymbol{W}_{(s)}$ *if we have these addition conditions:*

a) $\xi'\alpha_1 = \alpha_{s,k}$ *in the one sample case ; where* $k = m$.

b) for $k = n$ $\xi'\alpha_1 = \alpha_{s,k}$ *in the two sample case.*

The MSE of the BLIP of $\boldsymbol{W}_{(s)}$ *never exceeds that of the BLUP.*

Remark 2.6

The results derived in the preceding sections are easily extended to contain the case of a vector of future observations or linear combination of order statistics.

2.3.3 Maximum likelihood predictor

Let the sample be taken from the location-scale family of pdfs $\mathbf{f}((x-\mu)/\sigma)$*; the predictive likelihood function of* $X_{(s)}$ *and* $\theta = (\mu, \sigma)$ *is given by:*

$$\mathbf{L}(X_{(s)}, \theta, x) = C_n(r, s) \left\{ \prod_{j=1}^{r} \mathbf{f}_i \right\} [\mathbf{F}_s - \mathbf{F}_r]^{s-r-1} \mathbf{f}_s] 1 - \mathbf{F}_s]^{n-s} \tag{2.7}$$

$\mathrm{C}_n(r, s) = n!/\{(s-r-1)!(n-s)!\}$. Suppose $\delta_m = t(x)$ and $\hat{\theta}_m = u(x)$ are statistics for which

$$\mathbf{L}(t(x), u(x); x) = \sup_{(X_{(s)}, \theta)} \mathbf{L}(X_{(s)}, \theta, x)$$

Then we call $\delta_m = t(x)$ the MLP of $X_{(s)}$ and $\hat{\theta}_m = u(x)$ the predictive maximum likelihood estimate (PMLE)of θ.
The PLF in 2.7 can be written as:

$$\mathbf{L}(X_{(s)}, \theta, x) = \mathbf{f}_{X_{(s)}|X_{(r)}}(X_{(s)}, X_{(r)}, \theta) \mathbf{f}(x, \theta)$$

Remark 2.7

When θ *is known,* δ_m *is* $\mod (X_{(s)}|X_{(r)})$.
Note that:
$\mathbf{E}(\delta_m - X_{(s)}) = 0 \Longleftrightarrow \mathbf{E}(\delta_m) - \mathbf{E}\{\mathbf{E}(X_{(s)}|X_{(r)})\} = 0$
$\Longleftrightarrow \mathbf{E}\{\delta_m - \mathbf{E}(X_{(s)}|X_{(r)})\} = 0$. *these leads to the following results:*

- *When the parameters are known; the MLP of* $X_{(s)}$ *is unbiased if the mode and the mean of the conditional distribution of* $X_{(s)}$ *given* $X_{(r)}$ *coincide.*

- *When the parameters are unknown, simple sufficient conditions for the unbiasedness of δ_m do not seem to exist in general.*

A condition of consistency of δ_m by analyzing its MSPE in [12].Considering:

$$\mathbf{E}\{\delta_m - X_{(s)}\}^2 = \mathbf{E}\{\delta_m - \mathbf{E}(X_{(s)}|X_{(r)})\}^2 + \mathbf{E}\{Var(X_{(s)}|X_{(r)})\}.$$

Suppose the following regularity conditions:

$$\begin{array}{ll} C1 & s = \lambda_s]; \qquad 0 < \lambda_s < 1 \\ C2 & \mathbf{f}(\mathbf{F}^{-1}(\lambda_s)) > 0 \end{array}$$

It follows that:

$$\mathbf{E}^{k/2}|X_{(s)} - \mathbf{F}^{-1}(\lambda_s)|^k] \to \mathbf{E}|Y|^k \qquad \text{as} \quad n \to \infty \tag{2.8}$$

iif $\mathbf{E}|X|^\epsilon$ exists for some $\epsilon > 0$, Y is a normal r.v. (means = 0, and Var $= \{\lambda_s(1-\lambda_s)/fn^2(\mathbf{F}^{-1}(\lambda_s))\}$).
Hence if C1 and C2 hold and $\mathbf{E}|x|^\epsilon < \infty$ for some $\epsilon > 0$
$\mathbf{E}\{Var(X_{(s)}|X_{(r)})\} \to 0$ as $n \to \infty$ we have the following result:

- The MLP δ_m of $X_{(s)}$ is consistent if $|E(\delta_m - \mathbf{E}(X_{(s)}|X_{(r)}))^2 \to 0$ as $n \to \infty$

Remark 2.8

When the parameters are unknown, it states an unbiased predictor δ of $X_{(s)}$ is its BUP iif:

$$\mathbf{E}_\theta\{[X_{(s)} - \delta]\gamma(X)\} = 0$$

$\gamma(X)$is an unbiased estimator of zero.

Example 2.1 ***Exponential distribution***

δ_m the MLP of $X_{(s)}$ is unique when exists when θ is known. Consider a censored sample from Exp(σ) pdf; our goal is to examine the properties of the δ_m, we define:

$$\pi_0(r,s) = \log\left(\frac{n-r}{n-s+1}\right); \;\; \pi_1(r,s) = \sum_{j=n-s+1}^{n-r} j^{-1}; \;\; \pi_2 = \sum_{j=n-s+1}^{n-r} j^{-2}.$$

π_0, π_1, π_2 are the mode, mean and the variance respectively; of Z_{s-r} order statistic from a random sample of size $n-r$ from the EXP(1) distribution; since $X_{(r)}$ and $X_{(s)} - (r)$ are independent and $X_{(s)} - X_{(r)} \xrightarrow{d} \sigma Z_{(s-r)}$

in $n-r$ sample size. you can see [22] form the calculation process.The MLP of $X_{(s)}$ is obtained as:

$$\delta_m = \quad \text{mod } (X_{(s)}|X_{(r)}) = X_{(r)} + \sigma\pi_0 \tag{2.9}$$

δ_m is negatively biased, we have:

$$0 < \pi_1 - \pi_0 < \log\left\{\left(\frac{n-r+1}{n-s}\right)\left(\frac{n-s+1}{n-r}\right)\right\}$$

$$= \log\{[1+(n-r)^{-1}][1+(n-s)^{-1}]\} \tag{2.10}$$

The bias approaches 0 if $n-s \to \infty$.
The consistency of δ_m; consider

$$\mathbf{E}(\delta_m - \delta_{(s)})^2 = \mathbf{E}(X_{(s)} - X_{(r)} - \sigma\pi_0)^2 = \sigma^2\{\pi_2 + (\pi_1 - \pi_0)^2\} \tag{2.11}$$

from π_2 and 2.10 it follows that the second terms on the RHS of 2.21 also approaches to 0.SO δ_m is consistent.

Remark 2.9

When σ is unknown (Kaminsky and Rhodin 1985) have obtained the MLP and PMLE of $X_{(s)}$ and σ:

$$\delta_m = X_{(r)} + \pi_0\hat{\sigma}_m \qquad \hat{\sigma} = \frac{\mathbf{T}_r}{(r+1)}$$

where $\mathbf{T}_r = \sum_{i=1}^{r} X_{(j)} + (n-r)X_{(r)}$ is the known total time on test up to the time of r-th failure.

2.3.4 Conditional median predictor

In 1991 (TAKADA) has shown that the MUP of the form:

$$\delta_U(2) = \delta_U(1) + k_1\hat{\sigma} \tag{2.12}$$

Where $\delta_U(1)$ is the BUP of $X_{(s)}$ and $K_1 = \{(X_{(s)} - \delta_U(1))/\hat{\sigma}\}$ is better than $\delta_U(1)$ and the best invariant predictor in terms of the PMC A.3.2,[17].
Exponential distribution *The median of the conditional distribution of $X_{(s)}$ given $X_{(r)}$ as a predictor of $X_{(s)}$ for the exponential population.For*

the EXP(σ) parent, when σ is known, we have $\delta_U(1) = X_{(r)} + \sigma\pi_1$; $k_2 = Med\{(X_{(s)} - \delta_U(1))/\sigma\}$ an MUP is given:

$$\delta_U(2) = \delta_U(1) + k_2\sigma \tag{2.13}$$

Now let $k_2 = med\{(X_{(s)} - X_{(r)})/\sigma\} - \pi_1 = Med(Z_{(s-r)}) - \pi_1\}$; then we obtain

$$\delta_U(2) = X_{(r)} + \sigma Med(Z_{(s-r)}) \tag{2.14}$$

Since $\{X_{(s)} - x|X_{(r)} = x\} \xrightarrow{d} \sigma Z_{(s-r)}$ the CMP of $X_{(s)}$ is given by:

$$\delta_c = X_{(r)} + \sigma Med(W_{(s-r)}) \tag{2.15}$$

From 2.14 and 2.22 $\delta_U(2)$ and CMP are identical.
When σ is unknown, the UMVUE of σ is $\hat{\sigma} = (\mathbf{T}_r/r)$. Let $\delta_U(1) = X_{(r)} + \hat{\sigma}\pi_1$ and $\gamma(X_1)$ be any unbiased estimator of 0 based on X_1 we can conclude that:

$$\mathbf{E}\{(X_{(s)} - \delta_U(1))\gamma(X_1)\} = 0$$

hence $\delta_U(1)$ is the BUP of $X_{(s)}$ and $\delta_U(2)$ can be simplified:

$$\delta_U(2) = X_{(r)} + \hat{\sigma}Med((X_{(s)} - X_{(r)})/\hat{\sigma}) \tag{2.16}$$

Remark 2.10

With the conditional median approach, we can obtain another predictor of $X_{(s)}$ to be called MCMP (modified conditional median predictor) as:

$$\hat{\delta}_c = X_{(r)} + \hat{\sigma}Med(Z_{(s-r)})$$

2.3.5 Performance of some predictors

We want to compare the performance of the MCMP, MUP, BLUP and the MLP of $X_{(S)}$ for the EXP(σ) where σ is unknown, $\hat{\sigma}$ and $X_{(s)} - X_{(r)}$ are independent.$\mathbf{E}(\hat{\sigma}^2) = [(r+1)/r]\sigma^2$ and $\hat{\sigma}$ is unbiased estimator of σ.In the following table we will give the MSPE and the relative efficiency of $\hat{\delta}_c$ defined by:
$e(\hat{\delta}_c, \delta) = \dfrac{MSPE(\delta)}{MSPE(\hat{\delta}_c)}$ *where δ is the other predictor of $X_{(s)}$.*
It's important to note that we have a sample of size 10 from exponential population.

		MSPE in terms of σ^2				relative efficiency of $\hat{\delta}_c$		
r	s	$\hat{\delta}_c$	$\delta_U(2)$	δ_L	δ_m	$e(\hat{\delta}_c, \delta_U(2))$	$e(\hat{\delta}_c, \delta_L)$	$e(\hat{\delta}_c, \delta_m)$
4	5	0.0337	0.0334	0.0347	0.0556	0.991	1.030	1.650
	7	0.2100	0.2197	0.2254	0.2420	1.046	1.073	1.152
	9	0.9481	1.0212	1.0171	1.0107	1.077	1.073	1.066
	10	2.7735	2.9815	2.9921	3.0385	1.075	1.079	1.096
5	6	0.0476	0.0471	0.0471	0.0800	0.989	1.008	1.681
	8	0.3178	0.3257	0.3364	0.3778	1.025	1.059	1.189
	10	2.3566	2.4449	2.5068	2.7110	1.037	1.064	1.150
6	8	0.2224	0.2225	0.2303	0.2971	1.000	1.036	1.336
	10	2.0.469	2.0786	2.1473	2.4601	1.015	1.049	1.202
7	9	0.4494	0.4479	0.4603	0.6081	0.997	1.024	1.353
	10	1.7820	1.7854	1.8414	2.2536	1.002	1.033	1.265

Table 2.1: MSPE of predictors of $X_{(s)}$ in sample size 10 for the exponential population

From the table it is clear that the efficiency of:

- $\hat{\delta}_c$*relative to MUP is very high; its performance is better than the MUP in most cases of considered r and s.*
- *Moreover* $\hat{\delta}_c$ *is easier to compute than the MUP.*
- *The efficiency of* $\hat{\delta}_c$ *when compared to the BLUP* δ_L *and MLP* δ_m *is found to exceed 1 in all cases for considered r and s.*

To see the example in Uniform cases you are referred to [22].

2.3.6 New method for prediction of Order statistics

New method of predicting order statistics based on some characterization properties of order statistics via their cdf;[13].

Definition 2.1*(point D-predictor)*

Let W be a continuous r.v. with strictly increasing cdf, which has an infinite support.Then, a point predictor $\hat{W}$ *of W is said to be a point D-predictor of W if,* $\hat{W} \xrightarrow{d} W$ *where* $X \xrightarrow{d} Y$*; means that the r.v. X and Y have the same cdf.*

This new method for treating two prediction problems of two different type .

a) The classical problem of predicting $X_{(s)}$ $1 \leq r < s \leq n$ *by a D-predictor based on the observed order statistics* $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(r)}$ *which will be denote CP2 .*

b) The prediction problem of $X_{(r:M)}$ [1] *by a D-predictor when the sample size of the test is enlarged from M to n (by adding some extra items* $X_{(M+1)}, \ldots, X_{(n)}$*) . This will be denote CP1.*

Remark 2.11

The problem CP1 was not discussed in the literature till now.

The following two theorems are used to suggest the precedents predictors.

Lemma 2.1

Let $X_{(r)}$ *be the rth order statistic from a sample size n drawn from a continuous cdf* $\mathbf{F}_X$*. Furthermore let* U_j,$j = 0, 1, \ldots, n-m-1$ *be the r.v. which are independent of X and satisfy the relation* $X_{(r:m+j)}$ [2] $\xrightarrow{d} X_{(r)}+U_j$,$1 \leq r < m < n$*.Then ,*$U_j \xrightarrow{d} Y_{n-m-j:n-r}$ [3]$Y_{n-m-j:n-r}$ *order statistic from a sample drawn from cdf* $\mathbf{F}_Y$ *and* $Y \sim Exp(\alpha)$ *(i.e;* $E_\alpha(y) = 1 - \exp^{-\alpha y}, 0 < y < \infty; \alpha > 0$*) iif* $X \sim Logistic(\alpha)$ *i.e;* $(L_\alpha(x) = [1 + \exp^{-\alpha y}]^{-1}, -\infty < x < \infty)$*.*

Lemma 2.2

Let $X_{(r)}$ *be the rth order statistic from a sample of size n drawn from a continuous cdf* $\mathbf{F}_X$ *. Furthermore, let* $U_j, j = 0, 1, \ldots, r-1$ *be r.v. which are independent of X and satisfy the relation:* $X_{(r+n-m-j)} \xrightarrow{d} X_{(n-m)} + U_j; 1 \leq r < m < n$*.Then* $U_j \xrightarrow{d} Y_{(r-j:m)}$[4]*, drawn from the cdf* $E_\alpha(y)$ *iif* $X \sim Exp(\alpha)$*.*

Theorem 2.1*(problem CP1)*

Let $X_{(1)} < X_{(2)}, \ldots X_{(M)}$ *be order statistics from a sample of size M drawn from a continuous strictly increasing cdf* $\mathbf{F}_x$ *with an infinite support.Then the* $D_predictorof_{(r:M)}$[5] *where* $1 \leq r < M < n$ *is given by:*

[1] $X_{(r:M)}$ is the rth order statistic in sample of size M .
[2] $X_{(r:m+j)}$ the rth order statistic in a sample of size m+j.
[3] $Y_{n-m-j:n-r}$ the (n-m-j) th order statistic in (n-r) size sample.
[4] $Y_{(r-j:m)}$ the (r-j)th order statistic from a sample size m
[5] $\hat{X}_r$ is the rth order statistic in a sample of size n

$$\hat{X}_{(r:M)} = \mathbf{F}_X^{-1}(L_\alpha(L_\alpha^{-1}(\mathbf{F}_X(X_{(r:M)})) - Y_{(n-M:n-r)})) \tag{2.17}$$

Where $Y_{(n-M:n-r)}$ is the (n-M)th order statistic from a sample of size $n-r$ drawn from the cdf $E_\alpha(y)$, $L_\alpha^{-1}(y) = \frac{1}{\alpha}[\log(y) - \log(1-y)]$,$0 \leq y \leq 1$. For predicting $X_{(r:n)}$ based on the observed value $X_{(r:m)}$, where $M < n$ we generate an ordered random sample of size $n-r$ from the exponential distribution $E_\alpha(y)$ and determine $Y_{(n-M:n-r)}$.From 2.17 $\hat{X}_{(r:M)}$ can be easily computed.

Theorem 2.2*(problem CP2)*

Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be order statistic from a sample of size n drawn from a continuous strictly increasing cdf $\mathbf{F}_x$ with an infinite support.Then, a point predictor $\hat{X}_{(R+k)}$ of $X_{(R+k)}$ is given by:

$$\hat{X}_{(R+k)} = \mathbf{F}_X^{-1}((E_\alpha(E_\alpha^{-1}(\mathbf{F}(X_{(R)})) + Y_{(k:n-R)} \tag{2.18}$$

where $1 \geq k, R \leq n-2$ and $2 \leq R+k \leq n$;$E_\alpha^{-1}(y) = -\frac{1}{\alpha}\log(1-y), 0 \leq y \leq 1$, and $Y_{(k:n-R)}$ is the kth order statistic from a sample of size (n-R) drawn from the cdf $E_\alpha(y)$.

For predicting $X_{(R+k)}$ based on the observed value $X_{(R)}$; we generate an ordered random sample of size $n-R$ from the exponential distribution $E_\alpha(y)$ and determine $Y_{(k:n-R)}$.

Remark 2.12

*Since both the r.v. $X_{(r)}$ and $X_{(R+k)}$ as well as their cdf do not depend on the parameter α, the proof **??** of the theorems above; reveals an interested and useful property of the D-predictors (CP1) and (CP2), that they are stable with respect to α, the change of the value of α does not alter the value of the corresponding D-predictor.*

The following theorem gives an explicit form of quantile $Q(y : a; \mu_1, \lambda_1; \mu_2, \lambda_2)$

Theorem 2.3

For $y \in (0,1)$*, the quantile function for cdf* $\mathbf{F}(x : a, \mu_1, \lambda_1; \mu_2, \lambda_2)$ *is*

$$Q(y : a, \mu_1, \lambda_1; \mu_2, \lambda_2) = \mu_2 - \frac{1}{\lambda_2}\log(1-\beta_0) \qquad 0 \le a \le 1; \lambda_i > 0$$

Where $\beta_0 \in (0,1)$ *is the minimum root of the non linear equation of* (β)*.*

See [13].

2.4 Interval prediction

When predicting future outcomes from past outcomes, prediction intervals and regions play a more important role than point prediction.Using notation in 2.2.1, we want to predict $T = T(\boldsymbol{W})$ *from* $\hat{T}(\mathbf{U})$*.*

2.4.1 Intervals based on pivotals

$\mathbf{Q} = \psi(T\hat{T})$ *in the parametric case this means that for data generated by a pdf* $\mathbf{f}$ *as given in 2.1* $\mathbf{Q}$ *has a distribution free of* (μ, σ) *and in the nonparametric case the distribution of* $\mathbf{Q}$ *is the same for all continuous densities.In either case a* $1-2\gamma$ *prediction region for* T *is obtained by finding a set* B *contains* $\mathbf{Q}$ *with probability* $1-2\gamma$ *and then inverting:*

$$\begin{aligned} 1-2\gamma &= \mathbf{P}(\mathbf{Q} \in B) \\ &= \mathbf{P}(T \in A(\mathbf{U})) \end{aligned}$$

Yielding $A(\mathbf{U})$ *as* $1-2\gamma$ *prediction region .*
It is important to note that the concept of invariance allows this procedure to be used for location scale families.
We consider the problem of predicting linear combination of unobserved r.v. let $\boldsymbol{k}$ *be a generic vector of constants;* $T = \boldsymbol{K}\boldsymbol{X}_2$ *in the one sample problem and* $T = \boldsymbol{k}\boldsymbol{Y}$ *in the two sample problem.We have the following pivotals for any equivariant estimators*[6] $(\hat{\mu}, \hat{\sigma})$[7]

[6] $(\hat{\mu}(X_1), \sigma(\hat{X}_1))$ computed from X_1 are called equivariant if for all constants c and d $d > 0$: $\hat{\mu}(c\mathbf{1} + d\mathbf{X}_1) = c + d\hat{\mu}(\mathbf{X}_1)$
$\hat{\sigma}(c\mathbf{1} + d\mathbf{X}_1) = d\hat{\sigma}(\mathbf{X}_1)$

[7] It's important to note that BLUE, MLE and BLIE based on location-scale families are equivariant.

$$\Psi_1 = (\hat{\mu} - \mu)/\hat{\sigma} \qquad \Psi_2 = \sigma/\hat{\sigma}$$

Let $\hat{\boldsymbol{X}}_s = a_s'\boldsymbol{X}_1$ *and* $\hat{Y}_s = b_s'\boldsymbol{X}_1$ *be linear unbiased predictor of* X_s *from arrays of predicted values and predicting coefficients denote by:*

$$\begin{aligned}\hat{\boldsymbol{X}}_2 &= (\hat{X}_{(r+1)}, \hat{X}_{(r+2)}, \ldots, \hat{X}_{(m)})'\\ \hat{\boldsymbol{Y}} &= \hat{y}_{(1)}, \hat{y}_{(2)}, \ldots, \hat{y}_{(n)})'\\ A &= (a_{r+1}, a_{r+2}, \ldots a_m)\\ B &= (b_1, b_2, \ldots, b_n).\end{aligned}$$

Using the standardized variable and the 2.4.1 that the following are pivotals:

$$\begin{aligned}Q_{x,y} &= \boldsymbol{k}'(\boldsymbol{X}_2 - \hat{\boldsymbol{X}}_2)/\hat{\sigma} \quad &= \Psi_2\boldsymbol{k}'(\boldsymbol{Z}_{x2} - A'\boldsymbol{Z}_{x1})\\ Q_{x,y} &= \boldsymbol{k}'(\boldsymbol{Y} - \hat{\boldsymbol{Y}})/\hat{\sigma} \quad &= \Psi_2\boldsymbol{k}'(\boldsymbol{Z}_y - B'\boldsymbol{Z}_x)(2.19)\end{aligned}$$

Let $Q^{\delta}_{x,\boldsymbol{k}}$ *and* $Q^{\delta}_{y,\boldsymbol{k}}$ *represent the* $100\delta\%$ *points of the distributions of* $Q^{\delta}_{x,\boldsymbol{k}}$ *and* $Q^{\delta}_{y,\boldsymbol{k}}$ *respectively:*

$$\begin{aligned}\mathbf{P}(Q_{x,\boldsymbol{k}} \leq Q^{\delta}_{x,\boldsymbol{k}}) &= \delta\\ \mathbf{P}(Q_{y,\boldsymbol{k}} \leq Q^{\delta}_{y,\boldsymbol{k}}) &= \delta\end{aligned}$$

$Q^{\delta}_{x,\boldsymbol{k}}$ *and* $Q^{\delta}_{y,\boldsymbol{k}}$ *can easily be found by simulating data from the standardized pdf and computing* $Q_{x,\boldsymbol{k}}$ *and* $Q_{y,\boldsymbol{k}}$.
Other versions of $Q_{x,\boldsymbol{k}}$ *and* $Q_{y,\boldsymbol{k}}$ *of the form:*

$$\begin{aligned}&k_1(\boldsymbol{X}_1)Q_{x,\boldsymbol{k}} + k_2(\boldsymbol{X}_1)\\ &k_3(\boldsymbol{Y}_1)Q_{y,\boldsymbol{k}} + k_4(\boldsymbol{Y})Q\end{aligned}$$

Where $\{k_j\}$ *are r.v. free of unknown parameters; in either case level* $1 - 2\gamma$ *prediction intervals are given:*

$$\begin{aligned}&For \quad \boldsymbol{k'X}_2 \quad [\boldsymbol{k'}\hat{\boldsymbol{X}}_2 + \hat{\sigma}Q^{\gamma}_{x,k}; \boldsymbol{k'}\hat{\boldsymbol{X}}_2 + \hat{\sigma}Q^{\gamma-1}_{x,k}]\\ &For \quad \boldsymbol{k'Y} \quad [\boldsymbol{k'}\hat{\boldsymbol{Y}} + \hat{\sigma}Q^{\gamma}_{y,k}; \boldsymbol{k'}\hat{\boldsymbol{Y}} + \hat{\sigma}Q^{\gamma-1}_{y,k}]\end{aligned}$$

these are two sided interval can be written: $[L_x(\boldsymbol{X}_1); U_x(\boldsymbol{X}_1)]$ *and* $[L_y(\boldsymbol{X}_1); U_y(\boldsymbol{X}_1)]$. *Th one sided lower* $1 - \gamma$ *prediction interval are given as: for one sample problem and two sample respectively problem*

$[L_x(\boldsymbol{X}_1); \infty[$
$[L_y(\boldsymbol{X}_1); \infty[$ (2.20)

2.4.2 Intervals based on best linear predictors

Pivotals based on the form of the BLP of single order statistic $X_{(s)}$ *given by:*

$$Q_{x,s}(\hat{\mu},\hat{\sigma}) = (X_{(s)} - \hat{X}(\hat{\mu},\hat{\sigma}))/\hat{\sigma}$$
$$= \Psi_2(\boldsymbol{Z}_{X,s} - \mathbf{1} + w_s'\mathbf{V}_X^{-1}(\boldsymbol{Z}_X + \alpha_1)) - \Psi_1(\mathbf{1} + w'\mathbf{V}_X^{-1}) - \alpha_{s,m} - w_s'\mathbf{V}^{-1}\alpha_1$$

$$Q_{y,s}(\hat{\mu},\hat{\sigma}) = (y_{(s)} - \hat{Y}(\hat{\mu},\hat{\sigma}))/\hat{\sigma}$$
$$= \Psi_2(\boldsymbol{Z}_{Y,s} + \alpha_{s,n}) - \Psi_1 - \alpha_{s,n}.$$

Here, letting $Q_{x\boldsymbol{k}}(\hat{\mu},\hat{\sigma}) = \sum \boldsymbol{k}_s Q_{x,s}(\hat{\mu},\hat{\sigma})$ *and* $Q_{y\boldsymbol{k}}(\hat{\mu},\hat{\sigma}) = \sum \boldsymbol{k}_s Q_{y,s}(\hat{\mu},\hat{\sigma})$ *prediction intervals constructed from the above pivotals and precedents* $100\delta\%$ *we have:*

$$For \quad \boldsymbol{k'X}_2 \quad [\boldsymbol{k'}\hat{\boldsymbol{X}}_2(\hat{\mu},\hat{\sigma}) + \hat{\sigma}Q_{x,k}^{\gamma}; \boldsymbol{k'}\hat{\boldsymbol{X}}_2(\hat{\mu},\hat{\sigma}) + \hat{\sigma}Q_{x,k}^{\gamma-1}]$$

$$For \quad \boldsymbol{k'Y} \quad [\boldsymbol{k'}\hat{\boldsymbol{Y}}(\hat{\mu},\hat{\sigma}) + \hat{\sigma}Q_{y,k}^{\gamma}; \boldsymbol{k'}\hat{\boldsymbol{Y}}(\hat{\mu},\hat{\sigma}) + \hat{\sigma}Q_{y,k}^{\gamma-1}]$$

Example 2.2

This example represent some analytic results for the interval prediction for the exponential model $\mathbf{f}(x|\mu,\sigma) = (1/\sigma)\exp((x-\mu)/\sigma)$ *where* $x > \mu$ *and* $\mathbf{f}(x|\mu,\sigma) = 0$ *elsewhere. Let* $S_r = \sum_{i=1}^{r} X_{(i)} + (m-r)X_{(r)}$, *r times the mle of* σ *for predicting* $X_{(s)}$ *from* $\boldsymbol{X}_1$ *with* μ *known and equal to 0 the* $1-2\gamma$ *prediction interval for* $X_{(s)}$ *has the form:*

$$[X_{(r)} + R^{\gamma}(\boldsymbol{X}_1, X_{(s)})S_r; X + (r) + R^{1-\gamma}(\boldsymbol{X}_1, X_{(s)})S_r] \tag{2.21}$$

where R is the derived distribution of the pivotal written as:

$$R(\boldsymbol{X}_1, X_s) = (X_{(s)} - X_{(r)})/S_r$$

$$P\ (R(\boldsymbol{X}_1, X_s) \geq t > 0) = (r/B(s-r), m-s+1)) \sum_{i=0}^{s-r-1} \binom{s-r-1}{i}(-1)^i$$
$$*[1+(m\text{-}s+i+1)t]^{-r}/[m-s+i+1] \tag{2.22}$$
$$B(a,b) = (a-1)!(b-1)!/(a+b-1)!$$

Remark 2.13

a) The percentage points of the distribution 2.22 can be approximated by scaled percentage points of appropriate $\mathbf{F}$ *distribution .See [9].*

b) The example above extended to an exponential family with unknown location parameter μ; also (Abu-salih and al 1987) generalize prediction to samples from the mixture of exponential pdf; with the mixing proportion β is assumed known.

c) Consider the class $C(1-\gamma)$ of lower $1-\gamma$ prediction interval for $X_{(s)}$ of the type 2.20 which are invariant; the conditional mean length $\mathbf{E}(X_{(s)} - L_x(X_1))|X_{(s)} \geq L_x(\boldsymbol{X}_1))$ is minimized over $C(1-\gamma)$ for all σ by the lower bound interval [19].

d) $\delta(\boldsymbol{X}_1) = X_{(r)} + R^{\gamma}(\boldsymbol{X}_1, X_{(s)})$ the lower bound interval as given in 2.21 minimizes $\mathbf{P}(X_{(s)} > \delta(\boldsymbol{X}_1) + a)$ for all positive constants a.

e) the minimization probabilities given above are type of optimality; these probabilities minimize the value being predicted is more than any specified distance above the lower endpoint of the prediction interval.

2.4.3 Multiple future samples

Let $\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \ldots \mathbf{Y}_p$ be vectors of order statistics obtained from independent random samples of sizes $m, n_1, n_2, \ldots, n_p$ from a continuous pdf use $y_{(i,j)}$ to denote the jth component of $\mathbf{Y}_i$, $j = 1, 2, \ldots n_i$; $i = 1, 2, \ldots, p$ given indices $\{q_i, q_i \leq n_i\}$; based on $\mathbf{X}$ it is desired to construct lower prediction intervals for $\{y_{(i,n_i+q_i+1)}\}$ of the form $\{I_i = [L_i(\mathbf{X}), \infty[; i = 1, 2, \ldots, p\}$ $1-\gamma$ represent the probability that all p intervals are simultaneously correct.Note that if the ith interval contains $y_{(i,n_i-q_i+1)}$ then at least q_i of the components of $\mathbf{Y}_i$ lie in I_i This type of prediction is complex because the distribution or the joint coverage probability has a complex expression, you refer to [13]to see example about exponential distribution and normal distribution.

2.5 Informativeness of order statistics

In this section our goal is to find the subset of order statistics that contains the most information about a sample of r.v. drawn independently from some known parametric distribution $\mathbf{f}(x|\theta)$ θ can be known or unknown parameter.To quantify the level of the informativeness and char-

acterize the amount of information contained in any subset of the complete collection of order statistics,[2].
consider order statistics $X_1, X_2, \ldots X_n$ drawn from **f**; let $X_{(1)}, \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be the corresponding order statistics; let denote the collection of the random samples as $X^n := (X_1, X_2, \ldots, X_n)$ and we use $[n]$ to denote the collection $\{1, 2, \ldots, n\}$we will use the mutual information **??** as a base measures of informativeness.
Starting by considering the mutual information between the sample X^n and any order statistic $X_{(i)}$ i.e; $I(X_{(i)}; X^n)$ and find the index $i \in [n]$ that results in the largest mutual information this approach works only with discrete r.v. not in the continuous case, then in the discrete case we have:

$$I(X_{(i)}; X^n) = H(X^n) - H(X^n|X_{(i)})$$
$$\sum_{x_{(i)}} \sum_{x^n} p(x_{(i)}, x^n) \log \left(\frac{p(x_{(i)}, x^n)}{p(x_{(i)}) p(x^n)} \right)$$

Definition 2.2

Let $Z^n := (Z_1, Z_2, \ldots, Z_n)$ be a vector of iid standard Gaussian r.v. independent of X^n let $S \subset [n]$ be defined as:

$$S = \{(i_1, i_2, \ldots, i_k) : 1 \leq i_1 < i_2 < \ldots < i_k \leq n\}$$

with $|s| = k$. We define the following three measures of order statistic informativeness:

$$r_1(S, X^n) = I(X^n; X_{(S)}) \tag{2.23}$$
$$r_2(S, X^n) = \lim_{\sigma \to \infty} 2\sigma^2 I(X^n + \sigma Z^n; X_{(S)}) \tag{2.24}$$
$$r_3(S, X^n) = \lim_{\sigma \to \infty} 2\sigma^2 I(X^n; X_{(S)} + \sigma Z^k) \tag{2.25}$$

So the measures r_1 computes the mutual information between a subset $X_{(S)}$ and sample X^n, the measure r_2 computes the slope of mutual information at $\sigma = \infty$ as noise becomes large only the most informative $X_{(S)}$ maintain the largest information; r_3 is alternative to r_2. The limits of r_2 and r_3 always exist but may be infinity.

We can consider similar measures to r_2 and r_3 given as:

$$r_4(S, X^n) = \lim_{\sigma \to 0} \frac{I(X^n + \sigma Z^n; X_{(S)})}{\frac{1}{2}\log(1 + \frac{1}{\sigma^2})} \tag{2.26}$$

$$r_5(S, X^n) = \lim_{\sigma \to 0} \frac{I(X^n; X_{(S)} + \sigma Z^k)}{\frac{1}{2}\log(1 + \frac{1}{\sigma^2})} \tag{2.27}$$

r_4 is that the most informative set $X_{(S)}$ should have the largest increase in the mutual information as the observed sample becomes less noisy; r_5 is alternative to r_4.It is important to note that these two measures $r_4; r_5$ are not useful.The three measures introduced previously has a shortcomings; the elements of the most informative set are not ordered based on the amount of information that each element provides, in this case we are unable to quantify the amount of information that an additional order statistic adds to the given collection, this can be remedied by using the conditional version of the measures introduced in definition 1.

Definition 2.3

Under the assumption in definition 1; let $\nu \subset [n]$ such that $S \cap \nu = \varnothing$; then we define three conditional measures of order statistic informativeness:

$$r_1(S, X^n|\nu) = I(X^n; X_{(S)}|X_{(\nu)}) \tag{2.28}$$

$$r_2(S, X^n|\nu) = \lim_{\sigma \to \infty} 2\sigma^2 I(X^n + \sigma Z^n; X_{(S)}|X_{(\nu)}) \tag{2.29}$$

$$r_3(S, X^n|\nu) = \lim_{\sigma \to \infty} 2\sigma^2 I(X^n; X_{(S)} + \sigma Z^k|X_{(\nu)}) \tag{2.30}$$

Characterization of the measures introduced in definition 1 and definition 2 are given in the following theorem.

Theorem 2.4

Let $S \subseteq [n]$ such that $|S| = k$ and $\nu \subset [n]$ such that $S \cap \nu = \varnothing$ then the metrics in definition 2 evaluate to

$$r_1(S, X^n|\nu) = \begin{cases} r_1(S, X^n|\nu) = \{H(X_{(S)}|X_{(\nu)}) & \text{for discrete r.v.} \\ \infty \quad \text{otherwise} \end{cases}$$

$$r_2(S, X^n|\nu) = \mathbf{E}[||\mathbf{E}[X^n|X_{(\nu)}] - \mathbf{E}^n|X_{(S)}, X_{(\nu)}]||^2] \tag{2.31}$$

$$r_3(S, X^n|\nu) = \mathbf{E}[||X_{(S)} - \mathbf{E}[X_{(S)}|X_{(\nu)}]||^2] \tag{2.32}$$

Taking $\nu = \varnothing$ gives an evaluation of the metrics in definition 1:

$$r_1(S, X^n) = \begin{cases} H(X_{(S)}) & \text{For the discrete r.v.} \\ \infty & \text{otherwise} \end{cases}$$

$$r_2(S, X^n) = \mathbf{E}[||\mathbf{E}[X^n] - \mathbf{E}[X^n|X_{(S)}]||^2] \tag{2.33}$$

$$r_3(S, X^n) = \mathbf{E}[||X_{(S)} - \mathbf{E}[X_{(S)}]||^2] \tag{2.34}$$

Proof 2.1

For simplicity, we focus on the case $\nu = \varnothing$ but for arbitrary ν follows the same lines.

- *Characterization of $r_1(S, X^n)$: we assume that X^n is sequence of discrete r.v. by using the relationship mentioned in A.3.4 we have; $I(X^n; X_{(S)}) = H(X_{(S)}) - H(X_{(S)}|X^n) = H(X_{(S)})$ where $H(X_{(S)}|X^n) = 0$ since $X_{(S)}$ is fully determined given the value of the sequenceX^n. If X^n is a sequence of continuous r.v. then $I(X^n; X_{(S)}) = \infty$ since $h(X_{(S)}|X^n) = -\infty$.*

- *Characterization of $r_2(S, X^n)$ we have $r_2(S, X^n) = 2\lim_{\sigma\to\infty} \sigma^2 I(X^n + \sigma Z^n; X_{(S)})$ defining $snr = \frac{1}{\sigma^2}$ then set the limit to 0 so we have:*
  $2\lim_{snr\to 0} \frac{I(\sqrt{snr}X^n + Z^n; X_{(S)})}{snr}$
  using the fact that : $\lim_{snr\to 0} \frac{\mathbf{f}(snr) - \mathbf{f}(0)}{snr} = \frac{d}{da}\mathbf{f}(a)|_{a=0}$
  $= 2\frac{d}{dsnr} I(\sqrt{snr}X^n + Z^n; X_{(S)})|_{snr=0}$ *using the generalized I-MMSE.*[8] *Since $X_{(S)} \to X^n \to (\sqrt{snr}X^n + Z^n)$ is a Markov chain we have :*
  $\mathbf{E}[||X^n - \mathbf{E}^n|Z^n]||^2 - ||X^n - \mathbf{E}[X^n|Z^n, X_{(S)}]||^2]$
  Since Z^n is independent of X^n then:
  $\mathbf{E}[||X^n - \mathbf{E}[X^n]||^2] - \mathbf{E}[||X^n - \mathbf{E}[X^n|X_{(S)}]||^2]$ *which is equal to*
  $\mathbf{E}[||\mathbf{E}[X^n] - \mathbf{E}[X^n|X_{(S)}||^2]$[3]
  Hence $r_2(S, X^n) = \mathbf{E}[||\mathbf{E}[X^n] - \mathbf{E}[X^n|X_{(S)}||^2]$

- *Characterization of $r_3(S, X^n)$ it follows by the data processing inequality that: $I(X, Z) = I(X, Y)$ for a Markov chain $X \to Y \to Z$ if $I(X; Y|Z) = 0$ notice that: $(X_{(S)} + ^k) \to X_{(S)} \to X^n$ forms a Markov chain with: $I(X_{(S)} + \sigma Z^k; X_{(S)}|X^n) = 0$ thus;*

  $$I(X_{(S)} + \sigma Z^k; X^n) = I(X_{(S)} + \sigma Z^k; X_{(S)})$$
  so $r_3(S, X^n) = \lim_{\sigma\to\infty} 2\sigma^2 I(X^n; X_{(S)} + \sigma Z^k)$
  $\lim_{\sigma\to\infty} 2\sigma^2 I(X_{(S)}; X_{(S)} + \sigma Z^k) = \mathbf{E}[||X_{(S)} - \mathbf{E}[X_{(S)}]||^2]$

Using the precedent theorem 2.4 procedures are constructed to answer these questions:

a) How much information does a single order statistic $X_{(i)}$ contain about the random sample X^n for each i in $[n]$?

b) Let $S \subseteq [n]$ be a set of cardinality $|S| = k$ and let $X_{(S)} = \{X_{(i)}\}_{i \in S}$ which subset of order statistics $X_{(S)}$ of size k is the most informative with respect to sample X^n?

c) Given a set $S \subseteq [n]$ and the collection of order statistics $X_{(S)}$ which additional order statistic $X_{(i)}$ where $i \in [n]$ but $i \notin S$ adds the most information about X^n ?

Then the three approaches are proposed note out that $m \in [3]$

a Marginal approach : generate one set of cardinality k according to

$$\bar{S}_m^M = \{(i_1, \ldots, i_k) : r_m(i_k, X^n), 1 \leq i_1 < \ldots < i_k \leq n\}$$

This approach generates an ordered set $\bar{S}_m^M$ of indices of order statistics, listed from the first most informative to the k-th most informative; and quantifies the amount of information that an individual order statistic contains about the sample.

b Joint approach: generates one set of cardinality k with

$$\bar{S}_m^J \in \arg_{S \subseteq [n],} \max_{|S|=k} r_m(S, X^n)$$

$\bar{S}_m^J$ contains indices of the k order statistics that are the most informative about the sample.

c Sequential approach: generate also one set of cardinality k according to

$$\bar{S}_m^S = \{(i_1, \ldots, i_k) : r_m(i_t, X^n|\nu_{t-1}|) \geq \max_{j \in [n]: j \notin \nu_{t-1}} r_m(j, X^n|\nu_{t-1}),$$

$$\nu_t = (i_1, \ldots i_t), t \in [k], \nu_0 = \varnothing\}$$This approach produces an ordered set, $\bar{S}_m^S$ of indices of order statistics where i_t is the most informative order statistic given that the information of $t-1$ order statistics has already been incorporated.

Remark 2.14

The sets $\bar{S}_m^M$,$\bar{S}_m^J$ and $\bar{S}_m^S$ may not be the same, even in simple cases, thus the application of interest and target analysis should guide the choice of which approach to use.

2.5.1 Evaluation of the informativeness measures

After using a metric to measures the informativeness of a set of order statistics now we will concentrate to evaluate this measures in two different cases discrete and continuous one.

2.5.1.1 Discrete case : Bernoulli

The following lemma studies the Bernoulli case

Lemma 2.3

Let X^n be sampled as iid Bernoulli with success probability p. Let B be a binomial $(n, 1-p)$ r.v. and B' be a binomial $(n-1, 1-p)$ r.v. then:

$$r_1(i, X^n) = h_b(\mathbf{P}(B < i)),$$
$$r_2(i, X^n) = \frac{np^2}{\mathbf{P}(B < i)}[\mathbf{P}(B' < i)]^2 + \frac{np^2}{\mathbf{P}(B \geq i)}[\mathbf{P}(B' \geq i)]^2 - np^2$$
$$r_3(i, X^n) = \mathbf{P}(B < i)\mathbf{P}(B \geq i)$$

Where $h_t(t) := -t\log(t) - (1-t)\log(1-t)$ is the binary entropy function.

When X^n is sampled iid Bernoulli with probability p, the information in the order statistics $0 \leq X_{(1)} \leq X_{(2)} \leq \ldots X_{(n)} \leq 1$ is simply the counts of $0's$ and $1's$ present in the data.In terms of order statistics the information lies in the location of the switch point if it exists i.e; the i where $X_{(i)} = 0$ but $X_{(i+1)} = 1$.Since we expect $\mathbf{E}(X^n) = np$ of the sample to take the value 1, the switch point is expected to occur at round (n(1-p)) and using the fact that when we consider $x(1-x)$ for $x \in [0;1]$;which is symmetric and convex with the maximum occurring at $x = 1/2$, thus $r_3(i, X^n)$ above is maximized by i such that $\mathbf{P}(B \geq i)$ or $1 - \mathbf{P}(B \geq i)$ is as close to 1/2 as possible so the maximizer i is a median of B and written like

$$i_3(X^n) = \arg\max_{i\in[n]} = \arg\min_{i\in\{\lfloor n(1-p)\rfloor, \lceil n(1-p)\rceil\}}$$

and the maximizer of r_1 is the same as the mentioned one above since the binary entropy function $h_b(t)$ is increasing on $0 < t < 1/2$ and decreasing on $1/2 \leq t < 1$ tells us that the most informative order statistic is where we expect the switch point to occur.
The following proposition shows that as the sample size grows; the most informative order statistic significantly dominates the other statistics for measures r_1 and r_3.

Proposition 2.1

Let X^n be iid Bernoulli with success probability $p \in [0,1]$ independent of n, we obtain

$$\lim_{n\to\infty} r_1(\lfloor cn \rfloor, X^n) = \begin{cases} \log(2), & c = (1-p) \\ 0, & otherwise \end{cases}$$

$$\lim_{n\to\infty} r_3(\lfloor cn \rfloor, X^n) = \begin{cases} 1/4, & c = (1-p) \\ 0, & otherwise \end{cases}$$

For ceiling and floor so put $\lceil . \rceil$ the same result holds see [2].

2.5.1.2 Continuous case : Uniform

From theorem 2.4 we have that the metric r_1 is infinity for continuous r.v. so we focus on r_2 and r_3; in particular the following lemma studies a uniform sample

Lemma 2.4

Let X^n be sampled as iid $U[0;a]$ for $a > 0$; for $k \in \{2,3\}$ define $i_k(X^n) = \arg\max_{i\in[n]} r_k(i, X^n)$ then

$$r_2(i, X^n) = \frac{a^2 i(n+1-i)}{4n(n+2)}$$
$$r_3(i, X^n) = \frac{a^2 i(n+1-i)}{(n+1)^2(n+2)}$$
$$and \quad i_2(X^n) = i_3(X^n) \in \left\{ \left\lceil \frac{n+1}{2} \right\rceil, \left\lfloor \frac{n+1}{2} \right\rfloor \right\}$$

For $c \in [0,1]$independent of n, metrics $r_2$and $r_3$have the following behaviors as n goes to infinity :
$\lim_{\to\infty} r_2(\lfloor cn \rfloor, X^n) = a^2 c(1-c)/4$
$\lim_{\to\infty} r_3(\lfloor cn \rfloor, X^n) = a^2 c(1-c)$

2.6 Selection approaches for predicting order statistics

Before making prediction of the future events it is important to manipulate the data, knowing the underlying distribution is not always easy process, for that there is approaches to choose the one which suits the best; to make a good predictions in aim to take the best possible decisions,[16]. In this section we focus on case when we have a type II censored sample if the population's distribution is known we follow this process: first we calculate the asymptotic MLE for the two parameters i.e; location and scale then we can calculate MLP to predict future order statistics but if the the population's distribution is unknown we must select the best underlying distribution so an extra step in the process above should be include which is selection of the best underlying distribution. The following schema show this process:

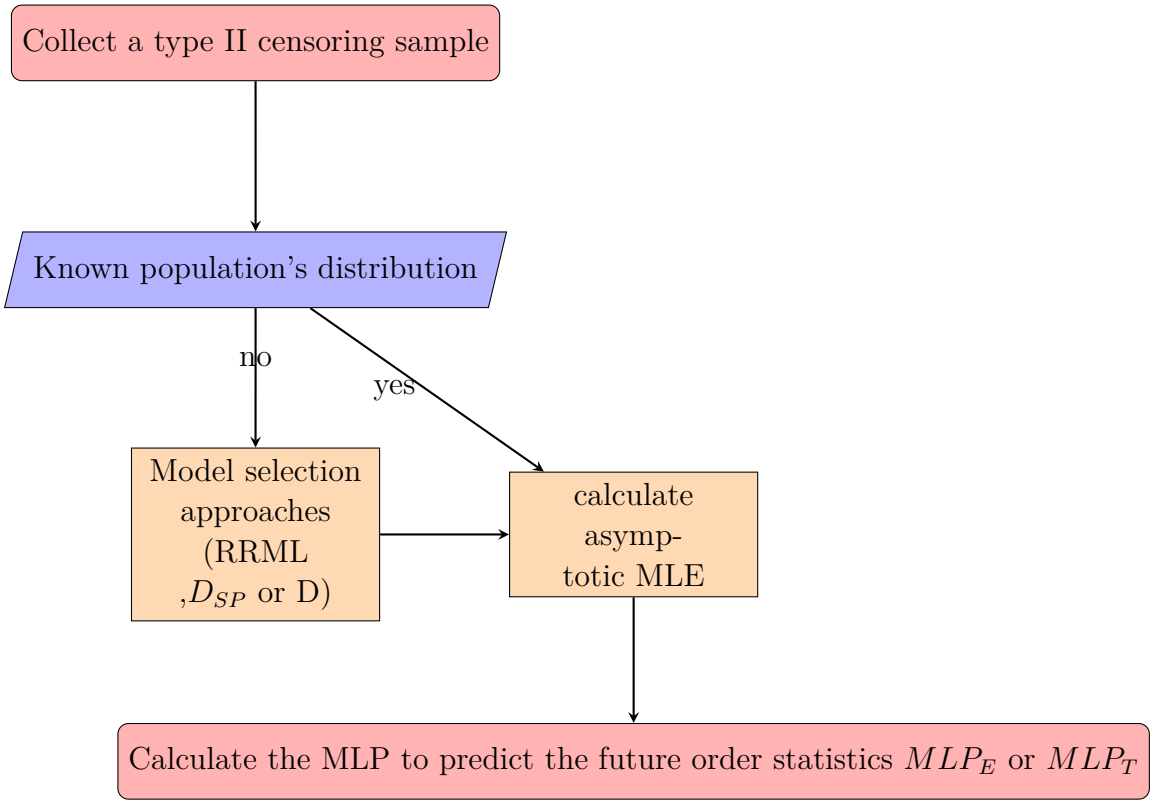


Figure 2.1: selection model

2.6.1 Three model selection approaches

When there is several candidate distributions are competing for the best underlying distribution and it is difficult to identify which one is the best three approaches are suggested to discriminate these candidates.We use the following approaches to obtain the predictor of $\hat{X}_{(s)}$

- The ratio of the maximized likelihood "RRML".
- Modification D_{SP} approach commonly shorted D_{SP}.
- Modification D approach commonly shorted D.

See [22] for more information about this.

Remark 2.15

- *The idea of the D_{SP} approach and D approach is based on goodness-of-fit test methods.*

- *We can't choose the best approach from these three mentioned above, for that the searchers proposed a Monte Carlo simulation to evaluate the performance of these approaches according to others parameters for example which MLP is used MLP_T one or the MLP_E ....*

2.6.2 Methods for approximate predictors

According to the last step mentioned in 2.1 before calculating the MLP we should calculate the MLE of the location-scale parameters, because almost we have a empirical studies thus an estimation is the key of the study, so here we will give an overview of the approximate maximum likelihood estimation of μ and σ then use the estimators to calculate approximate Maximum likelihood predictors using two approximation methods.

2.6.2.1 Approximate maximum likelihood estimation

Let Y_i denote the failure time of the ith item and $X_i = \log(Y_i)$,which follows a location-scale family, having the pdf mentioned in 2.1 and cdf

given: $\mathbf{G}$[9]

$$\mathbf{F}(x;\mu,\sigma) = \mathbf{G}(\frac{x-\mu}{\sigma}), \qquad -\infty < \mu < \infty, \sigma > 0; -\infty < x < \infty$$

$$Denotethesamplesizeby$$

n, $anddenotetypeIIcencoredsamplewithrfailuresbyx_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(r)}$ where $1 \leq r < s \leq n$; we want to predict $x_{(s)}$ for $r < s \leq n$
For simplification of notations we will write:
$\mathbf{f}(x) \equiv \mathbf{f}(x;\mu,\sigma)$ $\mathbf{F}(x) \equiv \mathbf{F}(x;\mu,\sigma)$ and the cpital notation $X_{(s)}$ is unknown and can be predicted based on the sample $X = (x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(r)})$
Kaminsky and Rhodin [12] considered prediction of $x_{(s)}$ having observed the sample X above, then they have the following predictive likelihood functions (PLF) of $X_{(s)}; \mu$ and σ is:

$$L(X_{(s)},\mu,\sigma;x) \equiv \mathbf{f}(x, X_{(s)};\mu,\sigma) = \frac{n!}{(s-r-1)!(n-s)!}\prod_{j=1}^{r}\mathbf{f}(x_{(j)})$$

$$[\mathrm{F}(\mathrm{X}_{(s)}) - \mathbf{F}(x_{(r)})]^{s-r-1}\mathbf{f}(X_{(s)})[1-\mathbf{F}(X_{(s)})]^{n-s}. \qquad (2.35)$$

[10]

The PLF of $X_{(s)}, \mu$ and σ in 2.35 can be represented as a product of two likelihood functions , the PLF of μ and σ which is called L_1 and the PLF of $X_{(S)}$ which is denoted as L_2 and they are represented respectively by:

$$L_1(\mu,\sigma;x) = \frac{n!}{(n-r)!}\prod_{j=1}^{r}\mathbf{f}(x_{(j)})[1-\mathbf{F}(x_{(r)}]^{n-r} \qquad (2.36)$$

and

$$L_2(X_{(s)};\mu,\sigma,x) = \frac{(n-r)!}{(s-r-1)!(n-s)!}\frac{[\mathbf{F}(X_{(s)})-\mathbf{F}(x_{(r)})]^{s-r-1}}{[1-\mathbf{F}(x_{(r)})]^{n-r}} * [1-\mathbf{F}(X_{(s)})]^{n-s}\mathbf{f}(X_{(s)}) \qquad (2.37)$$

In practice, we can obtain the MLE of $\hat{\mu}$ and $\hat{\sigma}$ by maximazing $L_1(\mu,\sigma;x)$ in 2.36,then use $\hat{\mu}$ and $\hat{\sigma}$ to replace μ and σ as the plug-in parameters in 2.37 to predict $X_{(s)}$.

[9] is the cdf of a member in the location-scale family

[10] PLF :The predictive likelihood function is a measure of how well a statistical model predicts new unseen data points. It is calculated by integrating the likelihood of the new data under the model's parameter values , weighted by the posterior distribution of these parameters , simply we can say this function evaluate how well a model's predictions match actual observations when considering uncertainly in the model's parameters.

Let $c_1 = \dfrac{n!}{(n-r)!}$ and $c_2 = \dfrac{(n-r)!}{(s-r-1)!(n-s)!}$
Let $z_{(j)} = (x_{(j)} - \mu)/\sigma$ for $j = 1, \ldots, r$ and $Z_{(s)} = (X_{(s)} - \mu)/\sigma$ for $s = r+1, \ldots, n$ and $z = (x_{(1)}, x_{(2)}, \ldots z_{(r)})$ then 2.36 and 2.37 can be rewrited:

$$L_1 \equiv L_1(\mu, \sigma; z) = c_1 \prod_{j=1}^{r} \sigma^{-1} \mathbf{f}(z_{(j)})[1 - \mathbf{F}(z_{(r)})]^{n-r} \tag{2.38}$$

and

$$L_2 \equiv L_2(Z_{(s)}; \hat{\mu}, \hat{\sigma}; z) = c_2 \sigma^{-1} \frac{[\mathbf{F}(Z_{(s)}) - \mathbf{F}(z_{(r)})]^{s-r-1}}{[1 - \mathbf{F}(z_{(r)})]^{n-r}} [1 - \mathbf{F}(Z_{(s)})]^{n-s} . \mathbf{f}(Z_{(s)}) \tag{2.39}$$

After straightforward computations, the MLE of μ, σ and $Z_{(s)}$ can be obtained respectively as solutions of:

$$\begin{aligned}
&\frac{\partial \log(L_1)}{\partial \mu} = \frac{1}{\sigma}\left[\sum_{j=1}^{r} \Psi(z_{(j)}) + (n-r)h(z_{(r)})\right] = 0 \\
&\frac{\partial \log(L_1)}{\partial \mu} = \frac{1}{\sigma}\left[-r + \sum_{j=1}^{r} \Psi(z_{(j)}) + (n-r)h(z_{(r)})z_{(r)}\right] = 0 \\
&\text{and} \quad \frac{\log(L_2)}{(s)} = (s-r-1)h_1(z_{(r)}, Z_{(s)}) - \Psi(Z_{(s)}) - (n-s)h(Z_{(s)}) = 0
\end{aligned} \tag{2.40}$$

$$\Psi(Z_{(j)}) = -\frac{\mathbf{f}'(Z_{(j)})}{\mathbf{f}(Z_{(j)})} \quad j = 1, \ldots, n \text{ where} Z_{(j)} = z_{(j)} \text{ if } j \geq r,$$

Where: $h(Z_{(j)}) = \dfrac{\mathbf{f}(Z_{(j)})}{1 - \mathbf{F}(Z_{(j)}}, \quad j = 1, \ldots, n \text{ where} Z_{(j)} = z_{(j)} \text{ if } j \geq r$

and $h_1(z_{(r)}, Z_{(s)}) = \dfrac{\mathbf{f}(Z_{(s)})}{\mathbf{F}(Z_{(s)}) - \mathbf{F}(z_{(r)})}$

It is important to note that there is no analytic presentations of $\hat{\mu}$ and $\hat{\sigma}$ thus there is a need to use numerical gradient computation methods, for instance Newton-Raphson method.

2.6.2.2 Approximate maximum likelihood predictors

Once we obtain the MLE of $\hat{\mu}$ and $\hat{\sigma}$ we can predict $X_{(s)}$ by using two approximation methods:

- Expected value prediction method; the resulting predictors is denoted by MLP_E
- Taylor series prediction method; the predictors getting via this methods is denoted by MLP_T

The two approximate methods mainly use two different methods to get approximate of $h_1(Z_{(s)}, z_{(r)})$ and $h(Z_{(s)}$.[20] Mehrotra and Nanda proposed approximate maximum likelihood estimators for the normal and gamma distribution by replacing $h(x)$ and $xh(x)$ by their respective expected value and efficiencies compared to those for the best linear unbaised estimators for these distribution; another method proposed by Balakrishnan and Cohen [24] using the taylor series expansion of $h(x)$and $\mathbf{f}(x)/\mathbf{F}(x)$ at point$\mathbf{F}^{-1}(p_s)$ where $(p_i = i/(n+1), i = 1, 2, \dots n)$.The problem with their approach is that likelihood equations involve complicated terms and it impossible to obtain an explicit form for MLE.
According to the article [12] an explicit form for the predictor of $X_{(s)}$ is found:

- Based on the expected value prediction method we have respectively the expected value of :$\mathbf{f}(Z_{(j)}), h_1(z_{(r)}, Z_{(s)})$ and $h(Z_{(s)})$ given by:

$$
\begin{aligned}
E[\mathbf{f}(Z_{(j)})] &= \frac{1}{n+1} \sum_{k=j+1}^{n+1} E[\Psi(Z_{(k:n+1)})] \;\; j \geq n \text{ and } Z_{(j)} = z_{(j)} \;\; \text{if } j \geq r \\
E[h(Z_{(j)})] &= \frac{1}{n-j} \sum_{k=j+1}^{n} E[\Psi(Z_{(k)})] \;\; j \geq n-1 \text{ and } Z_{(j)} = z_{(j)} \text{ if } j \geq r, \\
E[h_1(z_{(r)}, Z_{(s)})] &= \frac{1}{j-i-1} \sum_{k=j}^{n} E[\Psi(Z_{(k)})] \;\; j - i \leq 2, \text{ and } Z_{(j)} = z_{(j)} \text{ if } j \geq r.
\end{aligned}
\tag{2.41}
$$

- Based on the Taylor series prediction method, replacing (μ, σ) with $\hat{\mu}, \hat{\sigma}$ also replacing $h_1(z_{(r)}, Z_{(s)})$ and $h(Z_{(s)})$ with their Taylor series approximations at point $\mathbf{F}^{-1}(p_s)$ under distribution candidate, as we know there are many common distribution in location-scale family, the widely used the normal distribution the SEV[11] ...

[22] to see the illustrating application of this methods using the normal distribution and the SEV .

[11] SEV: smallest extreme value distribution.

Chapter 3

Bayesian Prediction Of Order statistics

Building upon our exploration of frequentist prediction techniques and in continuation of this research into prediction methodologies.This chapter shifts our focus to Bayesian prediction in the context of the order statistics.
By transitioning from the frequentist viewpoint to the Bayesian framework we embark on a journey to uncover a new dimension of predictive modeling that not only complements but also challenges and extends the traditional methods; broadening our analytical understanding.
Through a meticulous analysis we illuminate the transformative potential of Bayesian prediction shedding light on its ability to enhance our comprehension and decision-making in sphere of order statistics.
The integration of Bayesian methods into the realm of order statistics introduces a powerful approach to enhance predictive accuracy, make informed decisions and uncover latent patterns that conventional methods may overlook. This chapter is based on [7, 8, 9, 10] and [1, 23]

3.1 Fondamental concepts of Bayesian paradigm

There are two main philosophical approaches to statistics.The first is often referred to as the frequentist approach.Sometimes it is called the classical approach, the alternative approach is the Bayesian approach It applies the laws of probability directly to the problem.This offers many fundamental advantages over the more commonly used frequentist approach,[6, 7].

3.1.1 Bayesian statistical model

A parametric Bayesian statistical model is a framework for making statistical inferences and predictions using Bayes theorem.

3.1.1.1 Bayes theorem

Let A and B be events such that $P(B) \neq 0$, the conditional probability of these two events are related by the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The continuous version of the formula above, namely, that given two r.v. X and Y, with conditional distribution $\mathbf{f}(Y|X)$ and marginal distribution $\mathbf{f}(X)$ then the conditional distribution of X given Y is

$$\mathbf{f}(X|Y) = \frac{\mathbf{f}(Y|X)\mathbf{f}(X)}{\int \mathbf{f}(Y|X)\mathbf{f}(X)\,dx}$$

3.1.1.2 Definitions

Bayesian statistics consists of the observations of a r.v. x distributed following to $\mathbf{f}(x|\theta)$ in other words the likelihood distribution and incorporates a prior knowledge about parameter $\theta \in \Theta$ commonly named $\pi(\theta)$, given these two distributions we constructed the following distributions

a) Posterior distribution

$$\pi(\theta|X) = \frac{\mathbf{f}(x|\theta)\pi(\theta)}{\int\limits_{\Theta} \mathbf{f}(x|\theta)\pi(\theta)\, d\theta}$$

b) Marginal distribution of x represent the distribution of the observed data as following:

$$\mathbf{f}(x) = \int_{\Theta} \mathbf{f}(x|\theta)\pi(\theta)\, d\theta \tag{3.1}$$

c) Joint distribution of (θ, x) noted $h(\theta, x)$:

$$h(\theta, x) = \mathbf{f}(x|\theta)\pi(\theta)$$

In the continuation of building a Bayesian model four spaces are involved; let's define them:

- Observations space denoted X represents set results of an experimentation or observations of studied phenomenon.
- Actions space Symbolized "A", it represents set of decisions or actions (denoted a_i) to be taken according to the observed information.
- States of nature space indicated by Θ it represents set of unknown parameters θ.
- Decisions rules space represented by "D" it indicate decision rules set (decision rule denoted δ), which is defined as an application from X to "A" by:

$$\begin{aligned} \delta : \quad & X \longrightarrow A \\ & x \longrightarrow \delta(x_i) = a_i \end{aligned}$$

3.1.2 Prior distribution

The prior distribution is the key of the Bayesian inference, but also represent the most critical and criticized point in the same time because the choice of this prior distribution is quite challenging especially in real-life situations. There are two main types of prior distributions: the conjugate and the non-informative distribution.

- **Conjugate prior distribution:** also known as informative prior distribution, is called conjugate if the likelihood distribution $\mathbf{f}(x|\theta)$ and the prior distribution $\pi(\theta)$ have the same shape.

  Definition 3.1

  A family F of probability distribution on Θ is said to be conjugate for a likelihood function $\mathbf{f}(x|\theta)$, if for every $\pi \in F$, the posterior distribution $\pi(\theta|x)$ also belongs to F.

  The following table give us some examples about conjugate distributions.

  Example 3.1

$\mathbf{f}(x\|\theta)$	$\pi(\theta)$	$\pi(\theta\|x)$
$\mathrm{N}(\theta\|\sigma^2)$	$\mathrm{N}(\mu,\tau^2)$	$\mathrm{N}(x/\sigma^2+\mu/\tau^2,(1/\sigma^2+1/\tau^2)^{-1})$
$\mathrm{Binomial}(n,\theta)$	$\mathrm{Beta}(\alpha,\beta)$	$\mathrm{Beta}(\alpha+x,\beta+n-x)$
$\mathrm{Gamma}(\nu,\theta)$	$\mathrm{Gamma}(\alpha,\beta)$	$\mathrm{Gamma}\ (\alpha+\nu,\beta+x)$
$\mathrm{N}(\mu,1/\theta)$	$\mathrm{Gamma}(\alpha,\beta)$	$\mathrm{Gamma}(\alpha+1/2,\beta+(\mu-x)^2/2)$

Table 3.1: Examples of conjugate distribution

- **non-informative prior distribution**: When no prior information is available instead of turning back to classical alternatives like maximum likelihood estimation it still preferable to use Bayesian techniques if only because they underlie classical optimality criteria, these particular prior distributions must be derived from the sample distribution, since this is the only available information, that's why they are called noninformative priors.Some of the most used are given below:

  a) **Generalized prior distribution** commonly known as improper prior distribution if:

  $$\int_{\Theta} \pi(\theta)\, d\theta = \infty \quad \text{or} \quad \sum_{\theta\in\Theta} \pi(\theta) = \infty$$

  b) **Jeffrey's prior** (1946, 1961).It is based on the Fisher information and is defined by:

  $$\pi(\theta) \propto \sqrt{I(\theta)} \qquad or \;\; \pi(\theta) = C\sqrt{I(\theta)}$$

Where :C is normalization constant and I is Fisher information $I(\theta) = E[(\frac{\partial}{\partial\theta}\log \mathbf{f}(x|\theta))^2]$ or even $I(\theta) = -E[\frac{\partial^2 \log \mathbf{f}(x|\theta)}{\partial^2\theta}]$.

c) **Uniform prior** is the simplest one, assuming that Θ is finite set of size k, then $\pi(\theta) = \frac{1}{k}$. is closely tied with Laplace's prior.See[6]

3.1.3 Predictive posterior distribution

Predictive distribution is the centerpiece of the Bayesian prediction, it can be used to forecast future values and quantify uncertainty around those prediction or to check whether the model is consistent with data. The posterior predictive distribution is the distribution of unobserved observations (prediction) conditional on the observed data (likelihood). Let x be the observed data, θ be the parameter, and x_{pred} be the unobserved data; the posterior predictive distribution is defined to be the following:

$$\begin{aligned} p(x_{pred}|x) &= \int p(x_{pred}, \theta|x)\, d\theta \\ &= \int (x_{pred}, \theta|x)\pi(\theta|x)\, d\theta \end{aligned} \tag{3.2}$$

Given the assumption that the observed data and unobserved data are conditional independent given θ, the posterior predictive distribution given in 3.2 can be simplified as the following:

$$p(x_{pred}|x) = \int p(x_{pred}|\theta)p(\theta|y)\, d\theta$$

In other words we can say that The posterior predictive distribution is an integral of the likelihood function$\mathbf{f}(x|\theta)$ with respect to the posterior distribution $\pi(\theta|x)$.
The following figures show the difference between a posterior distribution and the posterior predictive distribution.

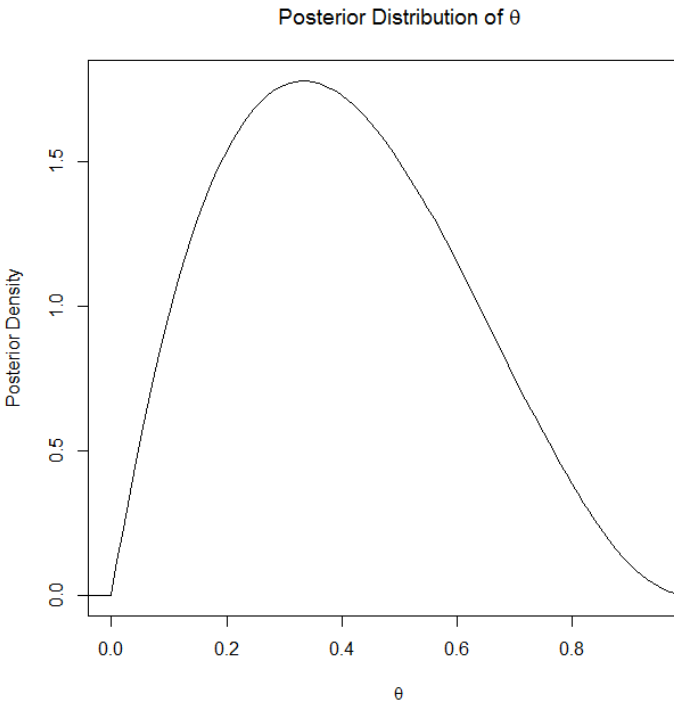


Figure 3.1: Posterior distribution

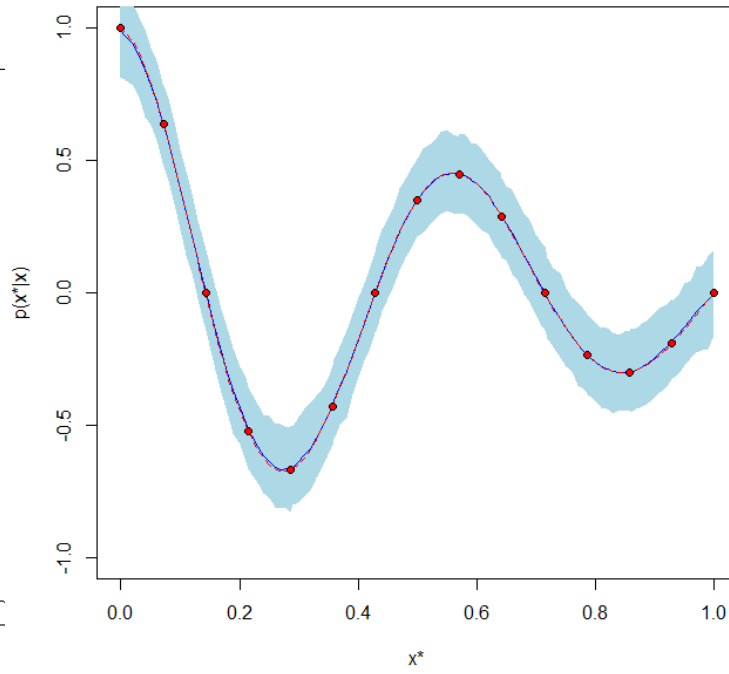


Figure 3.2: Predictive posterior distribution

According to the above figure we can say that the posterior distribution refer to the distribution of the parameter θ and the predictive posterior distribution refers to the distribution of the future observations of data.

Remark 3.1

Note that the posterior predictive distribution is not the same as the prior predictive distribution, the prior predictive distribution is $\pi(x)$ given in 4.2, which is the marginal distribution of the data.

3.1.4 Usual loss functions

Loss function also called a cost function denoted L, is supposed to evaluate error $L(\theta, \delta)$ associated with the decision d when the parameter take the value θ is defined from $(\Theta \times D)$ in $[0, +\infty[$ by;

$$L(\Theta; D) \longrightarrow \mathcal{R}^+$$
$$(\theta, \delta) \longrightarrow L(\Theta, \delta(x))$$

Here we give the most common loss functions:

- **The squared error loss"SEL" function:** or even called the quadratic loss function defined by: $L(\theta, \delta) = (\theta - \delta(x))^2$.

- **Entropy loss function :** defined by $L(\theta,\delta)=\frac{\delta(x)}{\theta}-\log(\frac{\delta(x)}{\theta})-1$
- **Linear loss function** commonly noted the absolute error loss function defined by:

$$L(\theta,\delta)=\begin{cases} K_1(\theta-\delta(x))\text{if} & \theta>\delta \\ K_2(\delta(x)-\theta)\text{if} & \theta\leq\delta. \end{cases}$$

  Under the assumption the loss function is symmetric and the constant $K_1=K_2$ the absolute error function is written as:

$$L(\theta,\delta)=|\theta-\delta|.$$

- **LINEX loss function** is defined by:

$$L(\theta,\delta)=(\frac{\delta(x)}{\theta})^{a'}-a'\log(\frac{\delta(x)}{\theta})-1 \quad for \;\; a'\neq 0$$

- **The** $0-1$ **loss function** is defined:

$$L(\theta,\delta)=\begin{cases} 0 & \text{if} \quad |\theta-\delta|<\epsilon \\ 1 & \text{if} \quad |\theta-\delta|\geq\epsilon \end{cases}$$

3.2 Bayesian prediction of Order statistics

In this section We delve into the Bayesian methodologies to predict order statistics , this problem has received considerable attention recently . Unlike the frequentist approach when the prediction are based only on the likelihood function the Bayesian model consist to incorporate a prior beliefs which represent not only an advantage but provides another methods of calculation using a different notations and concepts does not include in the frequentist viewpoint .
Through this section we will explore the two methods of prediction the point and interval one.

3.3 Bayesian point prediction of order statistics

In this section we will discuss the point prediction of order statistics in the Bayesian approach ; from different distributions and under different loss functions .

3.3.1 Bayesian point predictor based on record-value

At this point we will discuss the prediction of order statistics based on observed upper record value (already discussed in 1.5) from a exponential distribution under SEL function.
The exponential distribution is usually used in the prediction process, due to the memoryless property facilitated, even obtaining simple analytic results easily usable "exploitable". We denote $\exp(\theta)$ an exponential distribution with pdf

$$\begin{aligned} &\mathbf{f}(x;\theta) = \theta \exp^{-\theta x}, \quad x > 0, \ \ \theta > 0. \\ &\text{and its cdf is as follows :} \\ &\mathbf{F}(x,\theta) = 1 - \exp^{-\theta x}, \quad x > 0, \ \ \theta > 0. \end{aligned} \tag{3.3}$$

Let's $X_1, X_2, \ldots$ be an infinite sequence of r.v.; an observation X_j is an upper record. The sequence of upper k-records is then defined by $R_{n_{(k)}} = X_{T_{n,k}-k+1:T_{n,k}}$ for $n \geq 1$, is already mentioned in 1.5.
Suppose $R_{(1k)}, R_{(2k)}, \ldots, R_{(nk)}$ are the first n upper records from $\exp(\theta)$, using these upper records will predict $Y_{(j)}$ from sample size m.
The likelihood function for the parameter θ given $r_{(k)} = r_{(1k)}, r_{(2k)}, \ldots, r_{(nk)}$ is as following:

$$L(\theta|r) = (k\theta)^n \exp^{-k\theta r_{(nk)}}, \quad 0 < r_{(1k)} < r_{(2k)} < \ldots < r_{(nk)}. \tag{3.4}$$

For the prior distribution$\pi_{a,b}(\theta)$; where a and b are positive hyperparameters that could be chosen.we take the conjugate one:

$$\pi_{a,b}(\theta|r_{(k)}) \propto \theta^{a-1} \exp^{-b\theta} \quad \theta > 0, \tag{3.5}$$

Then from equations 3.4 and 3.5 the posterior distribution of θ given $r_{(k)}$ is:

$$\pi_{post(a,b)}(\theta|r_{(k)}) = \theta^{n+a-1} \exp^{-\theta(kr_{(nk)}+b)} \frac{(kr_{(nk)}+b)^{n+a}}{\Gamma(n+a)}, \quad \theta > 0 \tag{3.6}$$

So $\pi_{post(a,b)}(\theta|r_{(k)}) \sim \Gamma(n+a, kr_{(nk)}+b)$ where $\Gamma(.)$ is the complete gamma function.
Under the SEL function we obtain the Bayesian estimator of θ as follows:

$$\hat{\theta_B} = \frac{n+a}{kR_{(nk)}+b} \tag{3.7}$$

Using the pdf of the jth order statistics $Y_{(j)}$ given in 1.3 from a sample of size m in the exponential case we get:

$$\mathbf{f}_{Y_{(j)}}(y|\theta) = \frac{\theta(1-\exp^{-\theta y})^{j-1}\exp^{-\theta(m-j+1)}y}{B(j,m-j+1)}, \quad y>0, \tag{3.8}$$

$B(.)$ is a the complete beta function.
Then we need to calculate the focal point of the Bayesian prediction which is the Bayesian predictive density function of $Y_{(j)}$ given $r_{(k)}$,i.e, is a result of combining the posterior density given in 3.6 and the pdf in 3.8 then integrated out the parameter θ as following:

$$\mathbf{f}*_{Y_{(j)}}(y|r_{(k)}) = \int_{\theta} \mathbf{f}_{Y_{(j)}}(y|\theta)\pi_{(a,b)}(\theta|r_{(k)})\,d\theta. \tag{3.9}$$

then the explicit formulas of 3.9 is :

$$\mathbf{f}^*_{Y_{(j)}}(y|r_{(k)}) = \frac{(kr_{(nk)}+b)^{n+a}}{B(j,m-j+1)\Gamma(n+a)}\int_0^\infty \theta^{n+a}\exp^{-\theta(kr_{(nk)}+b+(m-j+1)y)}$$
$$(1-\exp^{\theta*y})^{j-1}\,d\theta$$

$$= \frac{(n+a)(kr_{(nk)}+b)^{n+a}}{B(j,m-j+1)}\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}\times\{Kr_{(nk)}+b+(m-j+i+1)y\}^{-(n+a+1)},$$
$$y>0. \tag{3.10}$$

The Bayes point predictor under SEL is **the mean of the predictive density** mentioned in 3.10 and is written as:

$$\hat{Y_{(j)}} = E(Y_{(j)|r_{(k)}})$$
$$= \frac{(Kr_{(nk)}+b)^{n+a}}{B(j,m-j+1)\Gamma(n+a)}\int_0^\infty\int_0^\infty y\theta^{n+a}\exp^{-\theta(kr_{(nk)}+b+(m-j+1)y)}(1-\exp^{\theta y})^{j-1}\,dy\,d\theta$$

Using the computation process in [17] we will have:

$$\hat{Y_{(j)}} = \frac{kR_{(nk)}+b}{(n+a-1)B(j,m-j+1)}\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}\frac{1}{(m-j+i+1)^2}$$

Let's $A(j,m)$ be a function depends on j and m defined by:

$$A(j,m) = \frac{1}{(n+a-1)B(j,m-j+1)}\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}\frac{1}{(m-j+i+1)^2}$$
$$= \psi(m+1)-\psi(m-j+1)$$

Remark 3.2

$\psi(x)$ psi function also known by the digamma function is the logarithmic derivative of the gamma function is defined by:

$$\psi(x) = \frac{\partial}{\partial x \ln \Gamma(x)} = \frac{\Gamma'(x)}{\Gamma(x)}$$

Then: $\hat{Y}_{(j)} = A(j,m)(kR_{(nk)} + b)/(n + a - 1)$.
Given the equation mentioned in 3.10 we can simply have the Bayesian predictive density function for the the minimum and the maximum, $j = 1$ and $j = m$

- **For the minimum case:** the equation 3.10 it simplifies to:

$$\mathbf{f} *_{Y_{(1)}} (y|r_{(k)}) = \frac{m(n+a)}{kr_{(nk)} + b}(1 + \frac{my}{kr_{(nk)} + b})^{-(n+a+1)}, \quad y > 0. \tag{3.11}$$

- **For the maximum in future sample case:** the equation 3.10 it simplifies to:

$$\mathbf{f} *_{Y_{(1)}} (y|r_{(k)}) = \frac{m(n+a)}{(kr_{(nk)} + b)^{-(n+a)}} \sum_{i=0}^{m-1} \frac{(-1)^i \binom{m-1}{i}}{(kr_{(nk)} + b + (i+1)y)^{(n+a+1)}}, \quad y > 0. \tag{3.12}$$

The Bayesian point predictor for the minimum and the maximum in future sample is respectively:

$$\hat{Y}_{(1)} = \frac{kR_{(nk)} + b}{(n + a - 1)m}$$

$$\hat{Y}_{(m)} = \frac{m(kR_{(nk)} + b)}{n + a - 1} \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i} \frac{1}{(i+1)^2}$$

Remark 3.3

These results are important in the case if we change the loss function or the prior distribution.

- ***Under the absolute error loss function:*** *we arrive at the median of the predictive density 3.10 and the point predictor of the minimum $j = 1$ is given*

$$\hat{Y}_{(1)} = \frac{kR_{(nk)} + b}{m}(2^{1/(n+a)} - 1).$$

- ***The case of the use another prior distribution:*** *if we take* $\pi(\theta) = \sum_{i=1}^{l} \beta_i \pi_{(a_i,b_i)}(\theta)$ *then the posterior distribution will be* $\pi(\theta|r_{(k)}) = \sum_{i=1}^{l} \beta_i(r_{(k)})\pi_{(a_i,b_i)}(\theta|r_{(k)})$*, we will obtain the Bayes point predictor of* $Y_{(j)}$*:*

$$\hat{Y}_{(j)} = \frac{1}{B(j, m-j+1)} \sum_{s=1}^{l} \sum_{i=0}^{j-1} \beta_s(r_{(k)})(-1)^i \binom{j-1}{i} \frac{kR_{(nk)} + b_s}{(n + a_s - 1)(m - j + i + 1)^2}.$$

Where $\beta_s(r_{(k)}) = \beta_s m_s(r_{(k)}) / \sum_{i=1}^{l} \beta_i m_i(r_{(k)}), s = 1, \ldots, l$ *and* $m_s(r_{(k)}) = \Gamma(n + a_s)/(kr_{(nk)} + b_s)^{n+a_s}$

3.3.1.1 Prediction of the mean of a future sample

As we see in the precedent section we get the Bayesian point predictor of $Y_{(j)}$ using the first n k-records from $EXP(\theta)$, here we focus on predicting the mean of the future sample namely $\bar{Y}_{(j)} = 1/m \sum_{i=1}^{m} Y_i$.
The pdf of the sample mean of a future sample of size m $\bar{Y}_{(j)}$ from an exponential distribution is given as:

$$\mathbf{f}_{\bar{Y}_m}(y) = \frac{(m\theta)^m y^{m-1} \exp^{-m\theta y}}{\Gamma(m)}, \quad y > 0. \tag{3.13}$$

From equations 3.10 and 3.13 the Bayesian predictive density of $\bar{Y}_m$ given the first n k-records $r_{(k)} = (r_{(1k)}, \ldots, r_{(nk)})$ is given by:

$$\mathbf{f}*_{\bar{Y}_m}(y|r_{(k)}) = \int_0^\infty \frac{m^m \theta^{n+a+m-1}}{\Gamma(n+a)\Gamma(m)} y^{m-1} (kr_{(nk)} + b)^{n+a} \exp^{-\theta(kr_{(nk)}+b+my)} \, d\theta$$

$$= \frac{1}{B(m, n+a)y} \left(\frac{kr_{(nk)+b}}{kr_{(nk)} + b + my}\right)^{n+a} \left(1 - \frac{kr_{(nk)+b}}{kr_{(nk)} + b + my}\right)^m \quad for \quad y > 0. \tag{3.14}$$

Under the SEL function, the Bayes predictive estimator called $\hat{\bar{Y}}_m$ is given by:

$$\begin{aligned} \hat{\bar{Y}}_m &= E(\bar{Y}_m|r_{(k)}) \\ &= \frac{kR_{(nk)} + b}{n + a - 1} \end{aligned}$$

The computation process is detailed in [14].
We remark that the predictor of $\bar{Y}_m$ does not depend on m.

Remark 3.4

- ***Under the*** $0-1$ ***loss function*** $\hat{\tilde{Y}}_m$ *is the predictive mode given:*

$$\hat{\tilde{Y}}_m = \frac{(m-1)(kR_{(nk)}+b)}{m(n+a+1)}$$

- ***Using another prior prediction*** *If we take* $\pi(\theta) = \sum_{i=1}^{l} \gamma_i \pi_{(a_i,b_i)}(\theta)$, *then the posterior distribution will be* $\pi(\theta|r_{(k)}) = \sum_{i=1}^{l} \gamma_i(r_{(k)})\pi_{(a_i,b_i)}(\theta|r_{(k)})$; *we obtain the Bayes predictive estimator of* $\hat{Y}_m$*:*

$$\hat{\tilde{Y}}_m = \sum_{i=1}^{l} \gamma_i(r_{(k)}) \frac{kR_{(nk)}+b}{n+a_i-1}$$

Where $\gamma_j(r_{(k)}) = \gamma_j u_j(r_{(k)}) / \sum_{i=1}^{l} \gamma_i u_i(r_{(k)}), j = 1, \ldots, l$ *with* $u_j(r_{(k)}) = \Gamma(n+a_j)/(kr_{(nk)}+b_j)^{n+a_j}$.

3.3.2 Bayesian point prediction for a grouped data

In lifetimes test usually we assume that the data has an exponential distribution 3.3. In this section have a survival test for n items from a sample X where $X = (X_1, X_2, \ldots, X_n)$. and E define the experiment where the n items supposed to be to a survival test. Our aim is to predict using the Bayesian method the future sample of size N $Y_1, Y_2, \ldots, Y_N$; independent of X. Given these notations:

- $Y = Y_{(r)} \quad 1 < r < N$ the ith order statistic.
- $Z = \sum_{i=1}^{k} Y_{(i)} + (N_k)Y_{(k)}$ the global lifetime of N till the kth failure.

Let T be predetermined duration, involving initially the n items submitted to a survival test. The time interval $[0, T]$ is subdivided into k intervals of length T/k. Let $(t_1, t_2, \ldots, t_k)$ such that $t_j = jT/k; j = 1, \ldots, k$ inspection times already fixed.
Assuming that $t_0 = 0$ and $t_{k+1\to\infty}$ Let :

- p_1 be the probability of failure in $[0, \frac{T}{k}] = \mathbf{F}_\theta(t_1)$
- p_j be the probability of failure in $[t_{j-1}, t_j] = \mathbf{F}_\theta(t_j) - \mathbf{F}_\theta(t_{j-1})$ for $j = 1, \ldots, k$

Using the memoryless property of the exponential distribution the above probabilities verify the following recurrence relation:

$$\begin{array}{l} p_j = p_1 exp\{-\theta(j-1)T/k\} \\ \text{and} \\ \sum_{j=1}^{k} p_j = \mathbf{F}_\theta(T) \end{array} \tag{3.15}$$

Now let's see the process of the prediction using grouped data A.1.1; following these steps:

a) **The likelihood function** is a multinomial written like follows:

$$L(x|\theta) = C \prod_{j=1}^{k} p_j^{xj} \{1 - \sum_{i=1}^{k}\}^{n-s}$$

Where $S = \sum_{i=1}^{k} x_i$; $S' = nk - \sum_{i=2}^{k}(k-i+1)x_i$ and

$$C = \frac{n!}{\prod_{i=1}^{k} x_i!(n-s)!}.$$

The recurrence relation given in 3.15 allows us to get the likelihood function only for p_1 and is written like:

$$L(x|\theta) = C \sum_{i=0}^{S} (-1)^i \binom{S}{i} exp\{-\theta \frac{T}{k}(S'+i)\}$$

Using the conjugate prior distribution$\pi(\theta)$ = gamma $\Gamma(g,h)$ we will obtain the posterior distribution .

b) **The posterior distribution** is given:

$$\pi(\theta|x) = \frac{1}{K\Gamma(g)} \sum_{i=0}^{S} (-1)^i \binom{S}{i} \theta^{g-1} exp\{-\theta[\frac{T}{k}(S'+i)+h]\}$$

Where: $K = \sum_{i=0}^{S}(-1)^i \binom{S}{i} \{\frac{T}{k}(S'+i)+h\}^{-g}$

c) **Density function for Y** using the prior distribution $\pi(\theta) = \Gamma(g,h)$ and under SEL function we get the pdf of Y as:

$$\mathbf{f}(y|\theta) = r\binom{N}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \theta \exp^{-(N-r+1-j)}$$

d) **Predictive density** then is:

$$\mathbf{f}*(y|x) = \int_0^\infty \mathbf{f}(y|\theta)\pi(\theta|x)\,d\theta$$
$$= \frac{gr\binom{N}{r}}{K}\sum_{i=0}^{S}\sum_{j=0}^{r-1}(-1)^{i+j}\binom{S}{i}\binom{r-1}{j}\left\{(N-r+1+j)g+H+i\frac{T}{k}\right\}^{-(g+1)}$$
where $H = \frac{T}{k}S' + h$

e) **The predictive cdf:** $\mathbf{F}(z|x)$ is given as follows:

$$\mathbf{F}(z|x) = \int_0^z \mathbf{f}*(y|x)\,dy$$
$$= \frac{r\binom{N}{r}}{K}\sum_{i=0}^{S}\sum_{j=0}^{r-1}(-1)^{i+j}\binom{S}{i}\binom{r-1}{j}(N-r+1+j)^i$$
$$\times\left\{(H+iT/k)^{-g} - [z(N-r+1+j)+(H+iT/k)]^{-g}\right\}$$

f) **The Bayesian point predictor of** $Y_{(r)}$ under the assumption $g > 1$ is given as follows:

$$E(Y|x) = \frac{r\binom{N}{r}}{(g-1)K}\sum_{i=0}^{S}(-1)^i\binom{S}{i}\left\{H+iT/k\right\}^{-(g+1)}$$
$$\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}\left\{N+r+1+j\right\}^{-2}$$

Remark 3.5

In special case when $r = 1$ then the prediction is about the first order statistic called in reliability analysis the occurrence of the first failure, the predictive density function and the predictive cumulative function are given respectively:

$$\mathbf{f}*(y|x) = \frac{gN}{K}\sum_{i=0}^{S}(-1)^i\binom{S}{i}\left\{Ny+H+i\frac{T}{k}\right\}^{-(g+1)}$$
$$\mathbf{F}(z|x) = 1 - \frac{1}{K}\sum_{i=0}^{S}(-1)^i\binom{S}{i}\left\{z(N-r+1+j)+H+iT/k\right\}^{-g}.$$

3.3.3 Bayesian predictor for a type II censored data from Weibull distribution

Weibull distribution is one of the most popular distribution in analyzing skewed data.In this part we consider the Bayes prediction of future or-

der statistics, based on the current type II censored sample; unlike the previous part here it is further assumed that the lifetimes of items being tested have a Weibull distribution with two parameters $\alpha > 0$ and $\lambda > 0$ are respectively the shape and scale parameters will be denoted $WE(\alpha, \lambda)$, its pdf is given by:

$$\mathbf{f}(t, \alpha, \lambda) = \begin{cases} \alpha\lambda t^{\alpha-1} \exp^{-\lambda t^{\alpha}} & \text{if} \quad t > 0 \\ 0 & \text{if} \quad t \leq 0. \end{cases} \tag{3.16}$$

If the shape parameter α is known we choose naturally as prior on the scale parameter λ the conjugate gamma prior, in case the parameter α is unknown does not exist a continuous conjugate although there is a continuous discrete joint prior.([7] and therein references).

Assumptions and notations:

Let $X = \{X_{(1)} < \ldots < X_{(m)}\}$ be the observed data, known as informative sample.

And $X_{(m+1)} < \ldots < X_{(n)}$ the unobserved future order statistics from the same sample. The prediction problem involves the prediction of the future order statistic $X_{(m+k)}$ for $0 < k \leq n - m$

For predicting the future order statistic $X_{(m+k)}$ first we obtain the posterior predictive density of $X_{(m+k)}$ given the first m ordered observations and is given by:

$$\pi_{X_{(m+k)}}(y|X) = \int_0^{\infty} \int_0^{\infty} \mathbf{f}_{X_{(m+k)}}(y|\alpha, \lambda, X)\pi(\alpha, \lambda|X)\, d\alpha\, d\lambda, \quad y > x_{(m)}.$$

Where $\mathbf{f}_{X_{(m+k)}}(.|\alpha, \lambda, X)$ is the conditional density of $X_{(m+k)}$ given $x_{(1)} < X_{(2)} < \ldots < x_{(m)}$.

Given the Markov property of the conditional order statistics we get:

$$\mathbf{f}_{X_{(m+k)}}(y|\alpha, \lambda, X) = \mathbf{f}_{X_{(m+k)}|X_{(m)}=x_{(m)}}(y|\alpha, \lambda, X)$$

$$= \frac{(n-m)!}{(k-1)!(n-k-m)!}\alpha\lambda y^{\alpha-1} \exp^{-\lambda(n-k-m+1)y^{\alpha}}$$
$$(\exp^{-\lambda x_{(m)}^{\alpha}} - \exp^{-\lambda y^{\alpha}})^{k-1} \exp^{\lambda(n-m)x_{(m)}^{\alpha}} \quad y > x_{(m)}.$$

Then the predictive density of $X_{(m+k)}$ at any point $y > x_{(m)}$ is the following:

$$\begin{aligned} \mathbf{f}_{X_{(m+k)}}(y|X) &= E_{posterior}[\mathbf{f}_{X_{(m+k)}|X_{(m)}=t_{(m)}}(y|\alpha, \lambda, X)] \\ &= \int_0^{\infty} \int_0^{\infty} \mathbf{f}_{X_{(m+k)}|X_{(m)}=t_{(m)}}(y|\alpha, \lambda, X)\pi(\alpha, \lambda|X)\, d\alpha\, d\lambda. \end{aligned} \tag{3.17}$$

As expected the Bayesian point predictor in this case also is the mean of the predictive density given above.

Remark 3.6

The results given above about the type-II censored sample are valid for type-I and other censoring schemes.

3.3.4 Bayesian point prediction using RRSS

Predicting future record statistics is one of the important problems in real life situations, for instance after the heatwave of last July one would be interested in predicting the degree of temperature in the future when the present record will be broken.
Here , we will discuss the prediction of records based on upper record ranked set sample it's a two-sample prediction A.4.
In practical experiments obtaining observations for the variable is costly and time consuming such as oil and mining surveys, sports ...; so there are some experiments where have been done sequentially, and only record-breaking data are observed, in order to reduce the cost and increase the efficiency of the sampling schemes, an alternative called record ranked set sample (RRSS)scheme for generating record-breaking data has been proposed [23].

Definition 3.2 ***Formal definition of upper RRSS*** *[8].*
Suppose we have n independent random sequences where the ith sequence sampling is finished when the ith upper record is observed.The only observations available for analysis are the last upper record value in each sequence.Let us denote the last upper record for the ith sequence in this plan by $R_{i,i}$, then $R = (R_{1,1}, R_{2,2} \ldots, R_{n,n})^T$ will be an upper RRSS of size n.The following observational procedure illustrates this plan

$$\begin{array}{lccccc} 1: & \underline{R_{(1)1}} & & & \to R_{1,1} = R_{(1)1} \\ 2: & R_{(1)2} & \underline{R_{(2)2}} & & \to R_{2,2} = R_{(2)2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ n: & R_{(1)n} & R_{(2)n} & \underline{R_{(n)n}} & \to R_{n,n} = R_{(n)n} \end{array}$$

Where $R_{(i)j}$ is the ith ordinary upper record in the jth sequence.It is important to note that $R_{i,i}$ are independent random variables but not necessarily ordered with probability 1.According to Lemma 2.1 *in [23],$R'_{i,i}s$*

have a stochastic orders in probability i.e; for $i < j; P(R_{i,i} < R_{j,j}) > \frac{1}{2}$. *Thus; if* $R_{i,i}$ *are upper records then using the marginal density of ordinary records the joint pdf of the elements of* R *is given as follows:*

$$\mathbf{f}_R(r;\theta) = \prod_{i=1}^{n} \frac{\{-\log \bar{\mathbf{F}}(r_{i,i};\theta)\}^{i-1}}{(i-1)!}\mathbf{f}(r_{i,i};\theta), \quad \theta \in \Theta \tag{3.18}$$

Where $\bar{\mathbf{F}}(.) = 1-\mathbf{F}(.)$ *the survival function;* $r = (r_{1,1}, r_{2,2}, \ldots, r_{n,n})^T$ *is the observed vector of R.Substituting* $\bar{\mathbf{F}}$ *by* $\mathbf{F}$ *gives that of the lower RRSS.In RRSS scheme, a specific statistic of a future sequence is predicted based on an observed sample.These two samples are supposed to be independent and called the future samples and informative samples; respectively.*

- **The likelihood function of** θ is:

  $$L(\theta|r) \propto \theta^N \exp^{-\theta\sum_{i=1}^{n} r_{i,i}}$$

  Let $t = \sum_{i=1}^{n} r_{i,i}$ is the observed value of $T = \sum_{i=1}^{n} R_{i,i}$.T is a complete sufficient statistic for θ and $N = \frac{n(n+1)}{2}$.

- **The prior distribution** Because θ is nonnegative, the natural choice for the prior of θ would be assume that its density has the following form:

  $$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\exp^{-\beta\theta}, \quad \theta > 0.$$

  To note that $\alpha > 0$ and $\beta > 0$ are the hyperparameters chosen to reflect the prior knowledge about θ,$\Gamma(.)$ is the complete gamma function.

- **The posterior distribution of** θ given r is given by:

  $$\pi(\theta|r) = \frac{(\beta+t)^{N+\alpha}}{\Gamma(N+\alpha)}\theta^{N+\alpha-1}\exp^{-\theta(\beta+t)}.$$

  i.e;$(\theta|r) \sim Gamma(N+\alpha, \beta+T)$.
  Assuming that Y_s is the *sth* upper record value from a future sequence;

- **The Bayesian predictive density function** of Y_s given r is written as follows:

  $$\begin{aligned} \mathbf{f}*_{Y_s}(y|r) &= \textstyle\int_\theta \mathbf{f}_{Y_s}(y|\theta)\pi(\theta|r)\,d\theta \\ \mathbf{f}*_{Y_s}(y|r) &= \frac{1}{B(s,N+\alpha)y}p(y)^s(1-p(y))^{N+\alpha}, \quad y > 0. \end{aligned} \tag{3.19}$$

Where $p(y) = \dfrac{y}{\beta + t + y}$. Using the notations given above we will discuss how to predict future record statistics arising from a future sequence based on an observed upper RRSS; Under different loss functions such as the general entropy loss function (GEL) is specially used because of different choices of a parameter value involved in the loss function can produce different symmetric and asymmetric loss functions , also the precautionary loss (PL) function, SEL, weighted SEL(WSEL) and LINEX loss function.

a) **Using GEL function** it is a generalization of the entropy loss function, its minimum occurs at$\hat{\theta} = \theta$ where $\hat{\theta}$ the estimate of θ of the form:

$$L_1(\theta, \hat{\theta}) = q\left[(\frac{\hat{\theta}}{\theta})^p - p\ln(\frac{\hat{\theta}}{\theta}) - 1\right]; \quad q > 0; p \neq 0,$$

Because the value of q does not play any role in the optimization of the loss function; we assume that $q = 1$.
Then Bayes predictor for Y_s under GEL is:

$$\hat{Y}_s = (E_{\mathbf{f}*}(Y_s^{-p}|r))^{-\frac{1}{p}}$$

The above expectation exist and is finite.
The choice of p is a challenging task in analyst term because it reflects the asymmetry of the loss function.

b) **Precautionary loss function PL** is an asymmetric loss function we get it when p of the GEL equal $p = -2$ and has the form:

$$L_2(\theta, \hat{\theta}) = \frac{(\hat{\theta} - \theta)^2}{\hat{\theta}}$$

Then the Bayesian predictor for Y_s is given as the following:

$$\hat{Y}_s = E_{\mathbf{f}*}(Y_s^2|r)^{1/2}$$

$$\hat{Y}_s = \left[\frac{(s-1)(s-2)}{(N+\alpha)(N+\alpha+1)}\right]^{\frac{1}{2}}(\beta + T), \quad s > 2.$$

And its MSE is given by

$$MSE(\hat{Y}_s) = E(\hat{Y}_s - Y_s)^2 =$$

$$\frac{(s-1)(s-2)}{(N+\alpha)(N+\alpha+1)}\left(\frac{N}{\theta^2}+\left(\beta+\frac{N}{\theta}\right)^2\right)-\frac{2s}{\theta}\left(\frac{(s-1)(s-2)}{(N+\alpha)(N+\alpha+1)}\right)^{\frac{1}{2}}$$
$$(\beta+\frac{N}{\theta})+\frac{s+s^2}{\theta^2}.$$

c) **Under the SEL function** The Bayesian point predictor in case the parameter of the GEL $p = -1$ coincide with the Bayes point predictor under a symmetric loss function SEL of the form :

$$L_3(\theta,\hat{\theta}) = (\hat{\theta}-\theta)^2$$

The Bayes point predictor for Y_s under the above loss function is:

$$\hat{Y}_s = \frac{s}{N+\alpha-1}(\beta+T), \quad s>0.$$

Its MSE is:

$$MSE(\hat{Y}_s) = \frac{s^2}{(N+\alpha-1)^2}\left(\frac{N}{\Theta^2}+\left(\beta+\frac{N}{\theta}\right)^2\right)-\frac{2s^2}{\theta(N+\alpha-1)}(\beta+\frac{N}{\theta})+\frac{s+s^2}{\theta^2}.$$

d) **Under WSEL function** The Bayes point predictor in this case it can be obtained with using the GEL function with $p = 1$; in this case we will get an asymmetric loss function WSEL (weighted squared error loss function) having the following form:

$$L_4(\theta,\hat{\theta}) = \frac{(\hat{\theta}-\theta)^2}{\theta}$$

Then the Bayesian point predictor for Y_s under WSEL may be defined using the predictive function $(E_{\mathbf{f}*}(Y_s^{-1}|r))^{-1}$ is:

$$E_{\mathbf{f}*}(Y_s^{-1}|r) = \frac{N+\alpha}{(s-1)(\beta+T)}$$
$$\text{therefore}: \quad \hat{Y}_s = \frac{s-1}{N+\alpha}(\beta+1), \quad s>1.$$

Then the MSE is as following:

$$E(\hat{Y}_s - Y_s)^2 = \frac{(s-1)^2}{(N+\alpha)^2}\left(\frac{N}{\theta^2}+\left(\beta+\frac{N}{\theta}\right)^2\right)-\frac{2s(s-1)}{\theta(N+\alpha)}\left(\beta+\frac{N}{\theta}\right)+\frac{s+s^2}{\theta^2}.$$

e) **Under the modified SEL function** Which has the form:

$$L_5(\theta, \hat{\theta}) = \left(\frac{\hat{\theta} - \theta}{\theta} \right)^2$$

The Bayes point predictor for Y_s is given by:

$$\hat{Y}_s = \frac{E_{\mathbf{f}*}(Y_s^{-1}|r)}{E_{\mathbf{f}*}(Y_s^{-2}|r)} = \frac{s-2}{N+\alpha+1}(\beta + T), \quad s > 2.$$

f) **Under LINEX loss function** It's important to note that the LINEX (Linear exponential) is convexe but asymmetric loss function is defined as following:

$$L_6(\theta, \hat{\theta}) = b[\exp^{a(\hat{\theta}-\theta)} - a(\hat{\theta} - \theta) - 1];$$

$b > 0$ is scale parameter and $a \neq 0$ is the shape parameter,without loss of generality it can be assumed $b = 1$.For a close to zero, this loss function is approximately SEL, therefore almost symmetric. Under the LINEX loss function the Bayes point predictor has not an explicit form, it must be solved by numerical methods, but we can give its implicit form which is:

$$\hat{Y}_s = -\frac{1}{a} \ln E_{\mathbf{f}*}(\exp^{-aY_s}|r) = -\frac{1}{a} \ln \left[\int_0^\infty \exp^{-ay} \mathbf{f}*_{Y_s}(y|r)\, dy \right]$$

It is important to note that expectation exist and is finite.

g) **Under the 0-1 loss function** of the form :

$$L_7(\theta, \hat{\theta}_i) = \begin{cases} 0 & \theta \in \Theta_i \\ 1 & \theta \in \Theta_j \end{cases}$$

For $i \neq j$, the Bayesian point predictor for Y_s is the mode of the Bayesian predictive density function given in 3.19, written like follows:

$$\hat{Y}_s = \frac{s-1}{N+\alpha+1}(\beta + T), \quad s > 1.$$

3.4 Bayesian interval prediction of order statistics

In this section we will discuss another way of prediction, is the interval prediction, which is a quite similar to the frequentist confidence intervals;

however, while their goal is the same but their statistical definition and meaning is very different.
Credible intervals are an important concepts in Bayesian statistics, its core purpose to describe and summarise the uncertainty related to the unknown parameters to estimate or to predict; indeed, the confidence interval is obtained through a complex algorithm full of rarely-tested assumptions and approximations, the Bayesian credible intervals are fairly straightforward to compute.

Definition 3.3

Let $(X, A, P_0, \pi(\theta))$ be the Bayesian model, assuming that θ is unknown parameter, I is credible interval of level α if:

$$P_{\pi(.|x)}(\theta \in I) \geq 1 - \alpha$$

for a prior distribution π , a set C_x is said to be an $\alpha-$credible set if:

$$P^{\pi}(\theta \in C_x|x) \geq 1 - \alpha$$

This region is called an HPD (highest posterior density) $\alpha -$ credible region if we can write it following this form:

$$C_x^{\pi} = \{\theta, \pi(\theta|x) \geq k\}$$

Where k is the largest bound such that :

$$P(\theta \in C_x^{\pi}) \geq 1 - \alpha.$$

Remark 3.7

C_s^{π} the HPD is $\alpha-$credible, but the converse is false.

3.4.1 Bayesian interval prediction for the future order statistics

In the continuation of the prediction process in 3.3.1 we will use the same notation and the same sample distribution 3.3.
In order to construct Bayesian prediction intervals, we first need the predictive survival function of the jth order statistic in future sample of size

m from 3.10, for $z > 0$ we simply obtain

$$\bar{\mathbf{F}} *_{Y_{(j)}} (z|r_{(k)}) = \frac{(n+a)(kr_{(nk)}+b)^{n+a}}{B(j, m-j+1)} \times \sum_{i=0}^{j-1} (-1)^i \binom{j-1}{i}$$
$$= \frac{1}{B(j, m-j+1)} \sum_{i=0}^{j-1} (-1)^i \binom{j-1}{i} \frac{1 + (m-j+i+1)z/(kr_{(nk)}+b)^{-(n+a)}}{m-j+i+1}. \tag{3.20}$$

The Bayesian predictive bounds of a two-sided equitailed $100(1-\alpha)\%$ interval for $Y_{(j)}$ in a future sample of size m, $j \geq 1$, can be obtained by solving this two equations:

$$\bar{\mathbf{F}} *_{Y_{(j)}} (L|r_{(k)}) = 1 - \frac{\alpha}{2} \quad \text{and} \quad \bar{\mathbf{F}} *_{Y_{(j)}} (U|r_{(k)}) = \frac{\alpha}{2} \tag{3.21}$$

Where L and U are the lower and upper bounds respectively.
Let $\zeta_{Y_{(j)},\alpha}(a,b)$ be the upper $\alpha-$quantile of the predictive distribution of $Y_{(j)}$ i.e;

$$\bar{\mathbf{F}} *_{Y_{(j)}} (\zeta_{Y_{(j)},\alpha}(a,b)|r_{(k)}) = \alpha;$$

And $\zeta_{Y_{(j)},1-\alpha}(a,b)$ is the lower $100(1-\alpha)\%$ Bayesian prediction limits or bound for $Y_{(j)}$.

Remark 3.8

It is clear that we have one-sided bound i.e; lower and upper and not two-sided Bayesian prediction limits.
In general we do not have an explicit form expression for the quantiles given above, so we use numerical methods to determinate them .

For HPD method, when $j \geq 2$ we need to solve the following equations:

$$\int_{uY_{(j)};1}^{uY_{(j)};2} \mathbf{f} *_{Y_{(j)}} (y|r_{(k)})dy = 1-\gamma \quad \text{or} \quad \bar{\mathbf{F}} *_{Y_{(j)}} (uY_{(j)};1|r_{(k)}) - \bar{\mathbf{F}} *_{Y_{(j)}} (uY_{(j)};2|r_{(k)})$$
$$\text{and} \quad \mathbf{f} *_{Y_{(j)}} (uY_{(j)};1|r_{(k)}) = \mathbf{f} *_{Y_{(j)}} (uY_{(j)};1|r_{(k)}) \tag{3.22}$$

In a special case when $j = 1$,we get the predictive survival function of the minimum which is written:

$$\bar{\mathbf{F}} *_{Y_{(1)}} (z|r_{(k)}) = \left(1 + \frac{mz}{kr_{(nk)}+b}\right)^{-(n+a)}, \quad z > 0 \tag{3.23}$$

For which the upper α quantile is obtained as:

$$\zeta_{Y_{(1)};\alpha}(a,b) = \frac{kr_{(nk)}+b}{m}\{\alpha^{-1/(n+a)} - 1\}$$

Substituting equation 3.23 into equation 3.21 an explicit Bayesian prediction interval $100(1-\alpha)\%$ for the minimum of a future sample of size m is given by:

$$\left(\frac{kR_{(nk)}+b}{m}\left\{\left(1-\frac{\alpha}{2}\right)^{-1/(n+a)}-1\right\},\frac{kR_{(nk)}+b}{m}\left\{\left(\frac{\alpha}{2}\right)^{-1/(n+a)}-1\right\}\right) \tag{3.24}$$

Since $R_{(nk)} \sim \Gamma(n, k\theta)$, the expected width(EW) of the above interval is:

$$EW(l)=\left\{\frac{b\theta+n}{m\theta}\left(\frac{\alpha}{2}\right)^{-1/(n+a)}-\left(1-\frac{\alpha}{2}\right)^{-1/(n+a)}\right\}$$

When all the parameters are fixed, the width is decreasing with respect to m.
Let $Q_m = 1 + my/(kR_{(nk)} + b)$, the Q_m has the Pareto distribution with shape parameter $n + a$ and scale= 1 consequently, it can be considered as a Bayesian pivotal quantity for constructing prediction interval for the minimum of the future sample; i.e; in this case of the Pareto distribution we can easily get the same Bayesian prediction interval as in 3.24.

Remark 3.9

Since $\mathbf{f}*_{Y_{(1)}(y|r_{(k)})}$ *is strictly decreasing with respect to y, a two sided HPD prediction interval can not be constructed ; however, the HPD prediction interval for the minimum* $Y_{(1)}$ *with Bayesian cover* $1-\alpha$ *is given by:*

$$\left(0;\frac{kR_{(nk)}+b}{m}[\alpha^{-1/(n+a)}-1]\right).$$

3.4.1.1 Interval prediction of the mean of a future sample

We want to predict the mean of the future sample , using the k-record values from exponential distribution , we have already the pdf of the sample mean given in 3.13 also the predictive density of $\bar{Y}_{(m)}$ which is given in 3.14,from the latter one we have the Bayesian predictive survival func-

tion of $\bar{Y}_{(m)}$ is obtained as follows:

$$F *_{Y_{(m)}} (z|r_{(k)}) = \int_0^\infty \frac{\theta^{n+a-1} \exp^{-\theta(kr_{(nk)}+b)}(kr_{(nk)}+b)^{n+a}}{\Gamma(n+a)} \int_z^\infty \frac{(m\theta)^m y^{m-1} \exp^{-m}}{\Gamma(m)} \, dy \, d\theta$$
$$= \sum_{i=0}^{m-1} \frac{\Gamma(n+a+i)}{\Gamma(n+a)i!} \{q(z)\}^{n+a} \{1-q(z)\}^i, \quad z > 0 \tag{3.25}$$

where

$$q(z) = \frac{kr_{(nk)}+b}{kr_{(nk)}+b+mz} \tag{3.26}$$

The Bayesian predictive bounds of two-sided equitailed $100(1-\alpha)\%$ interval for $\bar{Y}_{(m)}$,$m \geq 1$ can be obtained by solving the following equations:

$$\bar{\mathbf{F}} *_{\bar{Y}_{(m)}} (L|r_{(k)}) = P(\bar{Y}_{(m)} > L|r_{(k)}) = 1 - \frac{\alpha}{2}$$
$$\text{and} \quad \bar{\mathbf{F}} *_{\bar{Y}_{(m)}} (U|r_{(k)}) = P(\bar{Y}_{(m)} > U|r_{(k)}) = \frac{\alpha}{2}$$

Let $\zeta_{\bar{Y}_{(m)},\alpha}(a,b)$ and $\zeta_{\bar{Y}_{(m)},1-\alpha}(a,b)$ be respectively the α upper and lower quantile of the Bayesian predictive distribution;

$$\bar{\mathbf{F}} *_{\bar{Y}_{(m)}} (\zeta_{\bar{Y}_{(m)},\alpha}(a,b|r_{(k)})) = \alpha$$
$$\text{and} \quad \bar{\mathbf{F}} *_{\bar{Y}_{(m)}} (\zeta_{\bar{Y}_{(m)},1-\alpha}(a,b|r_{(k)})) = 1-\alpha$$

It is clear that the values $\zeta_{\bar{Y}_{(m)},\alpha}(a,b)and\zeta_{\bar{Y}_{(m)},1-\alpha}(a,b)$ are the one-sided upper and lower $100(1-\alpha)\%$ Bayesian prediction bounds for $\bar{Y}_{(m)}$; these values can be determined numerically.
The two-sided equitailed $100(1-\alpha)\%$ Bayesian prediction interval is given by:

$$\left[\zeta_{\bar{Y}_{(m)};1-\alpha/2}(a,b); \zeta_{\bar{Y}_{(m)};\alpha/2}(a,b)\right]$$

Remark 3.10

a) For positive integer values of a, we have these results follows from the relationship between negative binomial and binomial distribution: $\bar{\mathbf{F}} *_{\bar{Y}_{(m)}} (Z|r_{(k)}) = P(T_1 \leq m-1)$ *where* T_1 *has a negative binomial distribution with parameters* $(n+a)$ *and* $q(z)$*, equivalently* $\bar{\mathbf{F}} *_{\bar{Y}_{(m)}}$ $(Z|r_{(k)}) = P(T_2 \geq n+a-1)$*; where* T_2 *has a binomial distribution with parameters* $(n+a+m-2)$ *and* $q(z)$*.*

b) It is easy to show that $q(\bar{Y}_{(m)})$ *in 3.26 has beta distribution with parameters* $(n+a-1)$ *and* m*;Therefore,we can find prediction interval for*$\bar{Y}_{(m)}$ *based on the pivotal quantity* $q(\bar{Y}_{(m)})$*; let* $\beta_\gamma(n_1,n_2)$ *denote the upper* γ*th of beta distribution with* n_1,n_2 *as shape parameters; that is* $P(W \geq \beta_\gamma(n_1,n_2)) = \gamma$ *i.e;* $W \sim Beta(n_1,n_2)$*.In this case two-sided equitailed* $100(1-\alpha)\%$ *Bayesian prediction interval for*$\bar{Y}_{(m)}$ *is given as follows:*

$$\left[\frac{kR_{(nk)}+b}{m}\left(\frac{1}{\beta_{\alpha/2}(n+a-1,m)}\right); \frac{kR_{(nk)}+b}{m}\left(\frac{1}{\beta_{1-\alpha/2}(n+a-1,m)}-1\right)\right].$$

For the HPD method, we need to solve the following equations:

$$\int_{v\bar{Y}_{(m),1}}^{v\bar{Y}_{(m),2}} \mathbf{f}*_{\bar{Y}_{(m)}}(y|r_{(k)})dy = 1-\gamma$$
$$\text{and}$$
$$\mathbf{f}*_{\bar{Y}_{(m)}}(v\bar{Y}_{(m)};1|r_{(k)}) = \mathbf{f}*_{\bar{Y}_{(m)}}(v\bar{Y}_{(m)};2|r_{(k)})$$

3.4.2 Bayesian prediction interval for a type-II censored data

Recall the notions used in 3.3.3, the focus now is to construct a prediction interval for type II censored data; its pdf given in 3.16, and the predictive density is given in 3.17, now let's show the survival predictive function at any point $y > x_{(m)}$:

$$\bar{\mathbf{F}}*_{X_{(m+k)}}(y|X) = E_{posterior}[\mathbf{F}*_{X_{(m+k)|X_{(m)}=x_{(m)}}}(y|\alpha,\lambda,X)]$$
$$= \int_0^\infty \int_0^\int \mathbf{F}*_{X_{(m+k)|X_{(m)}=x_{(m)}}}(y|\alpha,\lambda,X)\pi(\alpha,\lambda|X)\,d\alpha\,d\lambda.$$

Where $\mathbf{F}*_{X_{(m+k)|X_{(m)}=x_{(m)}}}(y|\alpha,\lambda,X) = \int_y^\infty \mathbf{f}_{X_{(m+k)|X_{(m)}=x_{(m)}}}(u|\alpha,\lambda,X)\,du$. To construct a two-sided predictive interval for $X_{(m+k)}$; a symmetric $100\gamma\%$ interval can be obtained by solving the following nonlinear equations simultaneously for the lower and upper bounds L and U respectively:

$$L: \quad \frac{1+\gamma}{2} = P(\mathbf{F}_{X_{(m+k)}} > L|X) \rightarrow \mathbf{F}*_{X_{(m+k)|X}}(L) = \frac{1+\gamma}{2}$$
$$U: \quad \frac{1-\gamma}{2} = P(\mathbf{F}_{X_{(m+k)}} > U|X) \rightarrow \mathbf{F}*_{X_{(m+k)|X}}(U) = \frac{1-\gamma}{2}$$

The above equations can not be solved analytically, we need to apply a numerical method to solve them.

3.4.3 Bayesian interval prediction using RRSS

3.4.3.1 Survival method

In order to construct Bayesian prediction interval based on the survival method, using the RRSS from 3.3.4 and all the notation given there. From 3.19 we can find $\bar{\mathbf{F}} *_{Y_{(s)}} (y|r)$ the Bayesian predictive survival function for $Y_{(s)}$ and is written:

$$\bar{\mathbf{F}} *_{Y_{(s)}} (y|r) = \int_y^\infty \frac{1}{B(s, N+\alpha)z} P(z)^s (1-P(z))^{N+\alpha}\, dz$$

$L(r)$ and $U(r)$ denote the lower and upper limits; we get a $100(1-\gamma)\%$ Bayesian prediction interval as follows:

$$\begin{aligned} &P(L(r) < Y_{(s)} < U(r)|r) = 1-\gamma \\ &i.e; \quad \bar{\mathbf{F}} *_{Y_{(s)}} (L(r)|r) = 1 - \frac{\gamma}{2}; \qquad \bar{\mathbf{F}} *_{Y_{(s)}} (U(r)|r) = \frac{\gamma}{2} \end{aligned}$$

$L(r)$ and $U(r)$ satisfy:

$$L(r) = \frac{B_{1-\frac{\gamma}{2}}(s, N+\alpha)}{1 - B_{1-\frac{\gamma}{2}}(s, N+\alpha)}(+T), \quad U(r) = \frac{B_{\frac{\gamma}{2}}(s, N+\alpha)}{1 - B_{\frac{\gamma}{2}}(s, N+\alpha)}(+T).$$

Where $B_\gamma(c_1, c_2)$ represent the $100\gamma - th$ percentile of Beta(c_1, c_2) distribution.

3.4.3.2 HPD method

From 3.19,one can conclude that the predictive function $\mathbf{f}*_{Y_{(s)}(y|r)}$ is continuous and unimodal pdf, then a $100(1-\gamma)\%$ HPD interval for $Y_{(s)}$ is given by (ζ_1, ζ_2) where ζ_1 and ζ_2 are the solutions of the following equations:

$$\begin{aligned} &\textstyle\int_{\zeta_1}^{\zeta_2} \mathbf{f} *_{Y_{(s)}} (y|r)\, dy = 1-\gamma, \quad 0 < \gamma < 1, \\ &\mathbf{f} *_{Y_{(s)}} (\zeta_1|r) = \mathbf{f} *_{Y_{(s)}} (\zeta_2|r); \end{aligned}$$

Or equivalently we can solve these equations to get the HPD interval for $Y_{(s)}$:

$$\begin{aligned} &\textstyle\int_{\zeta_1}^{\zeta_2} \frac{1}{B(s, N+\alpha)y} P(y)^s (1-P(y))^{N+\alpha}\, dy = 1-\gamma, \quad 0 < \gamma < 1, \\ &\left(\frac{\zeta_1}{\zeta_2}\right)^{s-1} = \left(\frac{\beta + t + \zeta_1}{\beta + t + \zeta_2}\right)^{N+\alpha+s}. \end{aligned}$$

A Monte Carlo simulation has to be used to solve this.

Remark 3.11

In the case $s = 1$, the Bayesian predictive density function in 3.19 for y is strictly decreasing function i.e;

$$\frac{\partial \mathbf{f} *_{Y_{(s)}}(y|r)}{\partial y} = -\frac{N+\alpha+1}{\beta+t+y} < 0.$$

The $100(1-\gamma)\%$ Bayesian prediction interval based on the HPD method is given by :

$$\left[0, (\beta+T)\left\{\gamma^{-\frac{1}{N+\alpha}} - 1\right\}\right].$$

To conclude this chapter it is important to note that the performance of the Bayesian predictor depends on the prior distribution and also on the loss function; indeed the frequentist interpretation of prediction regions does not apply to probabilities obtained from this distribution and inferences drawn from it can be highly dependent on the prior chosen, the use of noninformative priors provides some level of objectivity.

0.

Chapter 4

Applications And Simulations

As we reach the culmination of this exploration into order statistics prediction, after reviewing the frequentist prediction in chapter two and we delve further into prediction methodologies by introducing the Bayesian framework in chapter three.We focus in this part on evaluating the theoretical results obtained in the previous chapters; in order to monitor the performance of these predictors, provide an illustrative framework; and make this process a bit concrete, away from the abstract notions seen throughout the previous chapters; a simulation study and real data analysis are carried out using R software.[15].

4.1 Simulation study

to model and understand the behaviour of the order statistics especially their use in estimation and prediction processes there is different methods to simulate the foregoing results obtained in chapter 2 *and chapter* 3.*The most used methods in our study are Monte Carlo and MCMC.*

4.1.1 Monte Carlo simulation

Monte Carlo simulation is a type of simulation that relies on repeated random sampling and statistical analysis to compute the results.
In Monte Carlo simulation,we identify a statistical distribution which we can use as the source for each of the input parameters,then we draw random samples from each distribution which represent the values of the input variables ,for each set of input parameters we get a set of output parameters.The value of each output parameters is one particular outcome scenario in the simulation run.We collect such output values from a num-

ber of simulation runs.Finally we perform statistical analysis on the values of the output parameters to make decisions about the course of action.
The following steps are typically performed for the Monte Carlo simulation of a physical process:

a) ***Statistic Model generation:*** *We start with developing a deterministic model which closely resembles the real scenario,in this deterministic model we use the most likely value of the input parameters, then we apply mathematical relationships which use the values of the input variables and transform them into desired output.*

b) ***Input distribution identification:*** *After having a statistic model,we try to identify the underlying distributions.*

c) ***Random variables generation:*** *In this step we generate a set of random numbers (or random samples) from the identified underlying distribution.One set of random numbers consisting of one value for each of the input variables,will be used in the deterministic model to provide one set of output values;this process will be repeated to generate more sets of random numbers;one for each input distribution and collect different sets of possible output values.*

d) ***Analysis and decision making*** *Now we can perform statistical analysis on value collected from a sample of output values.*

4.1.2 Markov chains Monte Carlo methods

Markov chain Monte Carlo simulations or simply MCMC simulations are a fundamental tool in statistics and data science and is a powerful technique used to approximate complex probability distributions especially when direct sampling or computation is challenging. It allows to obtain approximate samples from high dimensional and intricate probability distributions, which is crucial in many real-life applications.
Using the ergodic theory of Markov chains offers an indirect solution to simulate a probability target density π based on the observation that it is much easier to construct an ergodic Markov chain with π as stationary probability measure ,than to simulate directly from π.

The general idea is to construct a Markov chain update to generate state space X_{t+1} given X_t such that π is a stationary distribution for the chain. i.e; if X_t has π as density so it will be the same for X_{t+1}.

4.1.3 Simulation of Order Statistics

4.1.3.1 Uniform Distribution

Some methods of simulating order statistics from a distribution $\mathbf{F}(x)$ is discussed in this section.it's important to mention that a straightforward way of simulating order statistics is to generate a pseudo-random sample from distribution $\mathbf{F}(x)$ and then sort the sample through an efficient algorithm like quick-sort, this general method"expensive and time consuming " may be avoided in many instances by making use of some of the distributional properties [3].
If we wish to generate order statistics based on the Uniform(0,1) distribution, we may use the following theorems and avoid sorting methods.

Theorem 4.1

For the uniform (0,1) distribution, the r.v. $V_1 = U_{(i)}/U_{(j)}$ and $V_2 = U_{(j)}$ $1 \leq i < j \leq n$ are statistically independent with V_1 and V_2 having Beta (i,j-i) and Beta (j,n-j+1) distribution respectively.

Theorem 4.2

For the uniform (0,1) distribution, the r.v.

$$V_1^* = \frac{U_{(1)}}{U_{(2)}}, V_2^* = \frac{U_{(2)}}{U_{(3)}}^2 \dots V_{n-1}^* = \frac{U_{(n-1)}}{U_{(n)}}^{n-1} and V_n^* = U_{(n)}^n$$

are all independent uniform (0,1) r.v.

For example, if we need only the ith order statistic u_i, it may simply be generated as a pseudorandom observation from $Beta(i, n-i+1)$ distribution. If, in particular, the largest observation u_n is required, it may be generated as $\nu_1^{1/n}$, where ν_1 is a pseudorandom observation from the Uniform(0,1) distribution. This approach, as noted by Schucany (1972), can be used efficiently to produce some extreme observations or even a complete sample $(u_1, u_2, \dots u_n)$. For example, after generating three pseudorandom Uniform (0,1) observations ν_1, ν_2, and ν_3, we may use Theorem

4.2 to produce three largest uniform order statistics from a sample of size n by setting

$$u_n = \nu_1^{1/n}, u_{n-1} = \nu_1^{1/n}\nu_2^{1/(n-1)}, u_{n-2} = \nu_1^{1/n}\nu_2^{1/(n-1)}\nu_3^{1/(n-2)} \tag{4.1}$$

Remark 4.1

This method is referred to in the literature as the descending method. It should be mentioned here that the descending method has been found to be slightly faster than the ascending method by Lurie and Mason (1973) through an empirical comparison.

Instead, if we require only the smallest and largest uniform order statistics from a sample of size n, they may be produced from two pseudorandom uniform (0,1) observations ν_1 and ν_2 with the use of the theorem 4.1 and setting

$$u_n = \nu_1^{1/n}, u_{n-1} = \nu_1^{1/n}\nu_2^{1/(n-1)} \tag{4.2}$$

We may note here that this is a combination of the ascending and descending methods.
Suppose some central uniform order statistics, say only $(u_i, u_{i+1}, \ldots u_j)$, are required; we may simulate these by first generating u_j as a pseudorandom observation from Beta(j, n - j + 1) distribution and then producing the remaining order statistics through the descending method. This simulation procedure due to Ramberg and Tadikamalla (1978), has been extended further by Horn and Schlipf (1986).
Another interesting method of generating uniform order statistics is due to Lurie and Hartley (1972). This method makes use of the fact that if $X_1, X_2, \ldots, X_{n+1}$ are independent standard exponential random variables, then

$$\frac{X_1}{\sum_{i=1}^{n+1} X_i}\frac{X_2}{\sum_{i=1}^{n+1} X_i}\cdots\frac{X_n}{\sum_{i=1}^{n+1} X_i}$$

are distributed as $U_{(1)}, U_{(2)} - U_{(1)}, \ldots U_{(n)} - U_{(n-1)}$.Now, after generating n + 1 pseudorandom Uniform (0,1) observations $\nu_1, \nu_2, \ldots \nu_{n+1}$ and setting $x_i = -\log \nu_i$ we may then produce the uniform order statistics u_i as:

$$u_i = \frac{\sum_{r=1}^{i} \log \nu_r}{\sum_{r=1}^{n+1} \log \nu_r}, for \quad i = 1, 2, \ldots n \tag{4.3}$$

Note that this method has fewer steps to perform than the descending method, but needs one extra Uniform(0,1) observation. The just-described

methods of simulating uniform order statistics may also be used easily to generate order statistics from any known distribution $\mathbf{F}(x)$ *for which* $\mathbf{F}^{-1}(.)$ *is relatively easy to compute. we may simply obtain the order statistics* $X_{(1)}, X_{(2)}, \ldots, X(n)$ *from the required distribution* $\mathbf{F}(.)$ *by setting :*

$$X_{(i)} = \mathbf{F}^{-1}(u_{(i)}), i = 1, 2, \ldots, n \tag{4.4}$$

4.1.3.2 Exponential Distribution

If we wish to generate the complete sample $(x_1, x_2, \ldots, x_n)$ *or a type II censored sample* $(x_1, x_2, \ldots, x_r)$ *from a standard exponential distribution , we may use the following theorem and avoid sorting .*

Theorem 4.3

Let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ *be the order statistics from the standard exponential population ; the r.v.* $Z_1, Z_2, \ldots Z_n$ *for* $i = 1 \ldots n$

$$Z_i = (n - i + 1)(X_{(i)} - X_{(i-1)})$$

With $X_{(0)} \equiv 0$ *are statistically iid we have :*

$$X_{(i)} = \sum_{r=1}^{i} Z_R/(n - r - 1) \qquad for i = 1, \ldots n \tag{4.5}$$

Which express the ith order statistic in a sample of size n from the standard exponential distribution as a linear combination of i independent standard exponential r.v.

This may be done simply be generating a pseudorandom sample $y_1, y_2, \ldots y_r$ *from the standard exponential distribution first use the equation 4.5.*

4.1.3.3 Using R software

With R it is easy to simulate order statistics without using this long algorithm , just using a package ***"orderstats"*** *; it's important to note that : All the methods in this package generate a vector of uniform order statistics using a beta distribution and use an inverse cumulative distribution function for some distribution to give a vector of random order statistic variables for some distribution ; the process described above. This is*

much more efficient than using a loop since it is directly sampling from the order statistic distribution.

4.1.3.4 examples

a) ***Uniform (0,1)***

order_ probs(draw_ size, k, n) .

b) ***Cauchy*** *:*
order_ rcauchy(draw_ size = 1, location = 0, scale = 1, k = 1, n = 1)

c) χ^2
order_rchisq(draw_size, df, k, n)

d) ***exponential distribution*** *:*
order_rexp(draw_size, rate, k, n)

e) ***normal distribution*** *:*
order_rnorm(draw_size = 1, mean = 0, sd = 1, k = 1, n = 1)

Where :

- *draw_ size : The size of the output sample*
- *k : The Kth smallest value from a sample .*
- *n : The size the sample to compute the order statistic from.*
- *location : The location parameter in the Cauchy distribution.*
- *scale : The scale parameter in the Cauchy distribution.*
- *df : degree of freedom.*
- *rate : The shape parameter in the exponential distribution.*

```
    order_ probs(15,3,100)
 [1] 0.017734578 0.054960082 0.012222320 0.024839220 0.027622212
0.011707373
 [7]  0.051729510 0.008486045 0.027338375 0.016745944
0.009281785 0.011902904
[13] 0.011455905 0.040999721 0.036476345
```

```
    order_rnorm(10,0,1,100,10000)
 [1] -2.327898 -2.287177 -2.338512 -2.307891 -2.305122 -2.333735
-2.294784
 [8] -2.281453 -2.265546 -2.323333
```

See [11].

4.1.4 Simulation study using Weibull distribution

The data in this illustrative example, already considered by "Balakrishnan, Beutner and Cramer (2010)" in "exact two sample non-parametric confidence prediction and tolerance intervals based on ordinary and progressively type-II right censored data" and analyzed by the authors of the article [7].
The data represents the life-times in hours of appliance cord put under a specific test; the sample consists of the smallest 9 *observations of* 12 *appliance cords as follows:*

57.5	77.8	88.0	98.4	102.1	105.3	139.3	143.9	148.0

Table 4.1: Data for Weibull model

The aim is to predict the unobserved order statistics and provides their credible intervals.
Weibull distribution is used to model this data set because the empirical hazard function A.8 shows an increasing trend.
Dividing all the points by 100*;the MLE of* $\alpha = 4.3490$ *and* $\lambda = 0.5112$ *and their confidence interval based on the fisher information i.e; using the Jeffery's prior are given respectively :*$[1.8476; 6.8504]$ *and* $[0.0354; 0.9871]$.
It is important to note that: when α *is known, it is assumed that* $\lambda > 0$ *has a gamma* (a, b) *prior given:*

$$\pi_1(\lambda|a,b) = \begin{cases} \dfrac{b^a}{\gamma(a)}\lambda^{a-1}\exp^{-b\lambda} & \lambda > 0. \\ 0 & \textit{otherwise.} \end{cases}$$

In the case the both parameters are unknown it is assumed that λ *has the same prior* $\pi_1(\lambda|a,b)$ *and the prior on* α *is* $\pi_2(\alpha)$ *is log-concave on the support* $[0, \infty[$.

The posterior density function of α *is given by:*

$$g(\alpha|X) \propto \pi_2(\alpha)\alpha^m \prod_{i=1}^{m} x_i^{\alpha-1} \times \frac{1}{(b + \sum_{i=1}^{m} x_i^{\alpha}(n-m)x_m^{\alpha})^{a+m}}$$

For simplification of the calculation process the above posterior density function it is approximated with a two-parameter gamma density function with shape and scale parameters 13.7217 *and* 3.3782 *respectively.*
The posterior density function and its approximate are very close as we can see in the figure below:

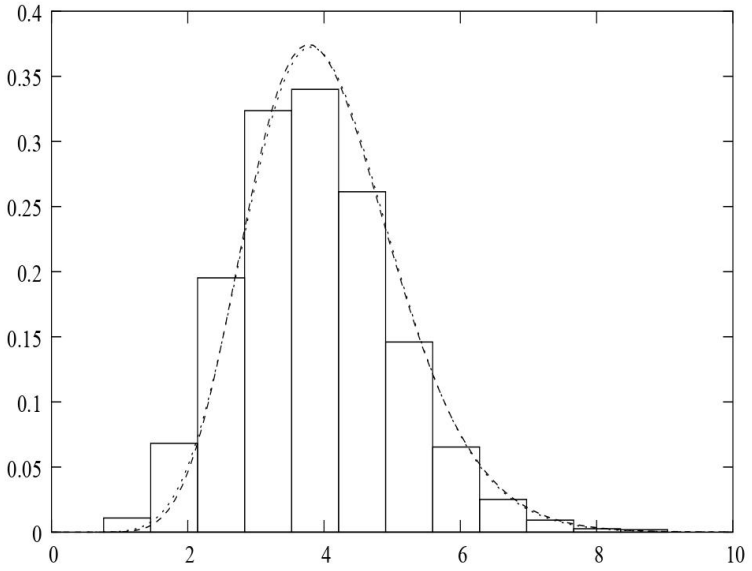


Figure 4.1: Posterior density function and its approximate

Using the MCMC method; 10000 samples have been generated for α, λ*; we remark that the posterior density function of* λ *can be approximated by two-parameters gamma distribution with shape and scale parameters as* 5.1560 *and* 9.3009 *respectively.*
Considering the prediction of 10^{th},11^{th} *and* 12^{th} *order statistics , which are unobserved. The* 95% *predictive interval are given:* 10^{th} $[148.0; 191.9]$, *The*11^{th} $[148.0; 233.7]$ *and the*12^{th} $[148.0; 302.2]$.

4.1.5 Simulation of CP1 predictor

This point predictor has been proposed by authors of [13] and it's important to note that the CP1 predictor was not discussed in the literature

till now, in this part we will see its performance using a simulation study and using the algorithm mentioned in [13].

Presentation of the process:

We will focus in this study in checking the performance of the CP1 predictor using exponential distribution denoted $E_3(x)$ which is important in lifetime studies also we use a mixture model $\mathbf{F}(x : 0.5, 3, 2, 5, 4)$; for sample size $n = 20$,$M = 15$ and different values of $r = 5 \ldots, 14$ the results of prediction and their MSE, SD and coefficient of variation (CV) in the following table.

Exponential $E_3(x)$						
r	$X_{(r:M)}$	$X_{(r:n)}$	$\hat{X}_{(r:n)}$	MSE	SD	CV
5	0.12	0.0931	0.0977	0.0010	0.04	0.32
6	0.16	0.1152	0.1219	0.0013	0.05	0.29
7	0.20	0.1392	0.1466	0.0015	0.05	0.27
8	0.24	0.1655	0.1719	0.0017	0.06	0.24
9	0.28	0.1927	0.2030	0.0024	0.07	0.24
10	0.34	0.2232	0.2336	0.0026	0.08	0.22
11	0.41	0.2570	0.2735	0.0037	0.09	0.22
12	0.49	0.2940	0.3146	0.0050	0.10	0.22
13	0.60	0.3368	0.3652	0.0075	0.12	0.24
14	0.77	0.3828	0.4235	0.0110	0.14	0.25

Mixture Model $\mathbf{F}(x : 0.5, 3, 2, 5, 4)$						
r	$X_{(r:M)}$	$X_{(r:n)}$	$\hat{X}_{(r:n)}$	MSE	SD	CV
5	4.9280	4.9000	4.9016	0.0008	0.0355	0.006
6	4.9525	4.9166	4.9183	0.0010	0.0412	0.007
7	4.9805	4.9347	4.9369	0.0014	0.0480	0.008
8	5.0116	4.9537	4.9565	0.0020	0.0551	0.009
9	5.0480	4.9744	4.9791	0.0029	0.0640	0.011
10	5.0903	4.9980	5.0029	0.0040	0.0741	0.013
11	5.1419	5.0237	5.0306	0.0060	0.0894	0.015
12	5.2051	5.0515	5.0610	0.0093	0.1079	0.019
13	5.2892	5.0835	5.0970	0.0158	0.1365	0.025
14	5.4150	5.1187	5.1422	0.0319	0.1871	0.035

Table 4.2: Simulation of the proposed CP1 predictor

Discussion: *from this simulation study and referring to the other simulations studies given in [13]we can conclude that:*

- *The performance of the predictors are good, using different distribution allows to compare which one is the best in term of performance.*

- *The other distribution provides results little bit good comparing to the exponential distribution in terms of MSE especially.*

- *The exponential distribution provides the best CV which measure the accuracy stability comparing to others.*

4.1.5.1 Simulation Using RRSS

RRSS is Record ranked set sampling.
To monitor the performance of the Bayes predictors we follow some steps of a specific algorithm then showing the results and discuss them. We can summarize these algorithm like following:

a) we carry out a Monte Carlo simulation in order to get a random samples.

b) A prior distribution should be already chosen and specify its hyper-parameters denote α and β, then we can generate θ from this prior.

c) Using the definition of RRSS we generate an upper RRSS of size n for generated θ.

d) For a specific r we compute the Bayes point predictors under a given loss function both symmetric and asymmetric loss.

e) We repeat this process 1000 *or more to compute the mean squared prediction error "MSPE" for the point predictors and the expected length "EL" and coverage probability CP for the Bayesian prediction interval.*

Application *To illustrate this process we consider both symmetric SEL and asymmetric PL functions, and using two priors informative and Jeffery's prior in order to predict $Y_{(s)}$for $s = 2, 4, 6, 7$ upper record values, $n = 2 \dots 5$ and $\theta = 1.17$.The results are the following:*

PL								
n	$s=2$	$MSPE$	$s=4$	MSPE	$s=6$	MSPE	$s=7$	MSPE
2	–	–	1.86	6.55	3.39	11.26	4.15	14.22
3	–	–	1.95	5.71	3.57	8.94	4.37	10.89
4	–	–	2.00	5.38	3.65	7.86	4.47	9.35
5	–	–	2.03	5.10	3.70	7.28	4.54	8.53
SEL								
n	$s=2$	$MSPE$	$s=4$	MSPE	$s=6$	MSPE	$s=7$	MSPE
2	2.62	4.35	5.25	14.51	7.87	30.48	9.19	40.64
3	2.07	2.26	4.13	6.15	6.20	11.67	7.23	15.04
4	1.90	1.85	3.81	4.48	5.71	7.90	6.66	9.91
5	1.83	1.69	3.67	3.85	5.50	6.49	6.42	7.98

Table 4.3: Bayes point predictors for $Y_{(s)}$ for non-informative prior.

Now we will give the case when an informative prior with hyperparameters $\alpha = 1.5$ and $\beta = 1.5$; under both SEL and PL functions and the results are the following

PL								
n	$s=2$	$MSPE$	$s=4$	MSPE	$s=6$	MSPE	$s=7$	MSPE
2	–	–	2.03	5.42	3.71	8.26	4.54	9.96
3	–	–	2.05	5.21	3.73	7.70	4.57	9.17
4	–	–	2.06	5.06	3.75	7.25	4.60	8.52
5	–	–	2.07	4.95	3.77	6.94	4.62	8.0.6
SEL								
n	$s=2$	$MSPE$	$s=4$	MSPE	$s=6$	MSPE	$s=7$	MSPE
2	2.36	2.54	4.71	7.25	7.07	14.14	8.25	18.40
3	2.05	1.97	4.10	4.99	6.15	9.05	7.18	11.47
4	1.92	1.76	3.83	4.12	5.75	7.10	6.71	8.82
5	1.85	1.65	3.70	3.70	5.55	6.16	6.48	7.54

Table 4.4: Simulation of the Bayes point predictors for $Y_{(s)}$ for informative prior.

[1] *From this simulation we note that:*

a) The MSPE are decreasing with respect to the sample sizes of upper record values n when the other parameters are fixed, but we remark that MSPE are increasing with respect to s. You can see more results about using different loss functions in [8].

b) The MSPE decrease when the hyperparameters α and β are increasing.

[1]MSPE:is mean squared prediction error

c) *The Bayes point predictors based on the informative prior perform better than those based on the non-informative prior because the prior information and the choice of the hyperparameters give us much information and perfectly fits the model.*

Since the above study we want to know more about the behaviour of the MSPE; the following figures present the plot of the MSPE, which give us a general idea about its behaviour

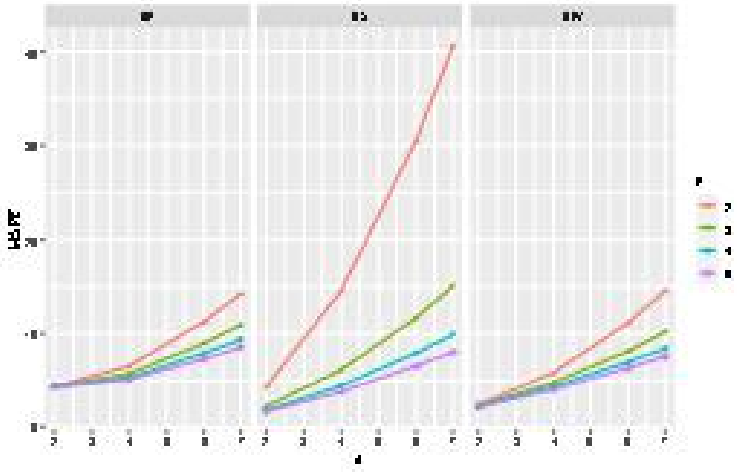


Figure 4.2: Plot the MSPE for non-informative prior

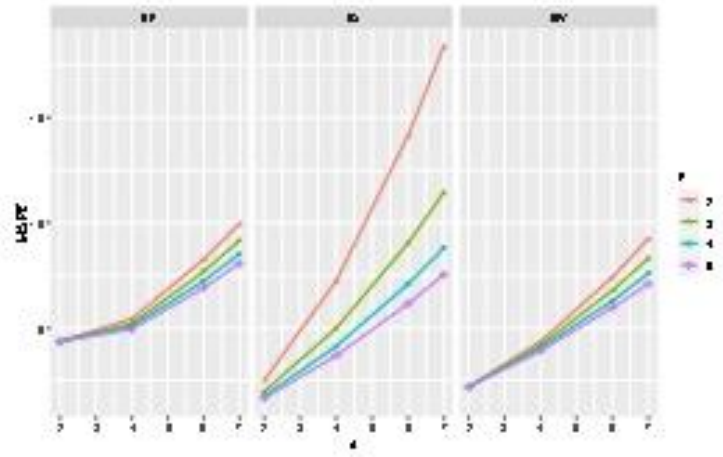


Figure 4.3: Plot the MSPE for informative prior

The plot above are given for $s = 2, 4, 6, 7$ *in different case of* $n = 2$ *the red plot,*$n = 3$ *is the green plot,*$n = 4$*is represented with the blue plot,and the purple plot present the behaviour of the MSPE for* $n = 5$ *which are decreasing with respect to sample size of upper record values. The following plot present the behaviour of the MSPE for Jeffery's prior and an informative prior , under different loss functions.*

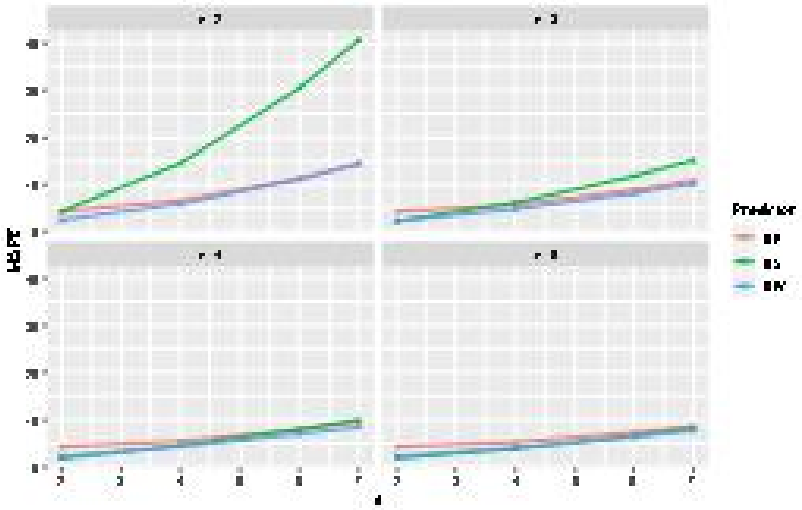


Figure 4.4: Behaviour of the MSPE under non-informative prior

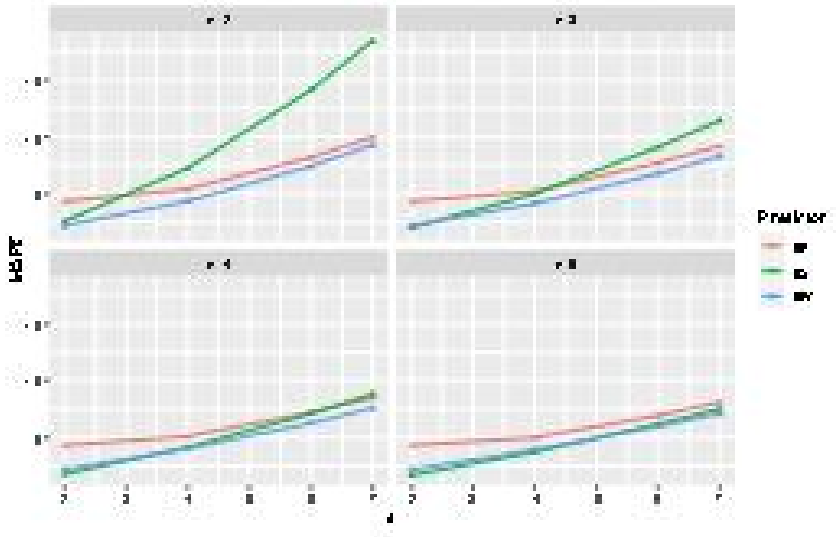


Figure 4.5: Behaviour of the MSPE under informative prior

The behaviour of the MSPE is not constant but its performance is better under a symmetric loss functions than under asymmetric loss functions.[8] for more information.
Here our focus on the interval prediction using the survival method and the HPD method and compare the intervals based on the informative prior to those based on the non-informative in term of their Expected length (EL) and coverage probability(CP) the results are the following:

Survival method				
n	$s=2$	$s=4$	$s=6$	$s=7$
	EL and CP	EL and CP	EL and CP	EL and CP
2	10.44 and 0.937	18.37 and 0.942	26.09 and 0.939	29.92 and 0.938
3	6.80 and 0.942	11.12 and 0.943	15.14 and 0.946	17.11 and 0.947
4	5.80 and 0.942	9.09 and 0.947	12.03 and 0.951	13.44 and 0.950
5	5.37 and 0.946	8.21 and 0.951	10.67 and 0.951	11.84 and 0.952

HPD method				
n	$s=2$	$s=4$	$s=6$	$s=7$
	EL and CP	EL and CP	EL and CP	EL and CP
2	6.88 and 0.959	12.99 and 0.960	18.73 and 0.958	21.44 and 0.959
3	5.51 and 0.958	9.62 and 0.957	13.08 and 0.959	14.81 and 0.958
4	5.02 and 0.960	8.17 and 0.959	10.91 and 0.958	12.23 and 0.952
5	4.68 and 0.961	7.53 and 0.962	9.92 and 0.960	11.01 and 0.961

Table 4.5: Simulation of an interval Prediction for $Y_{(s)}$ for non-informative prior.

Using an informative prior and compare the above results with the result given in the table below.
From this simulation study using the table below and the table above we can conclude that:

- *When n is large EL and CP improve.*
- *When n increases the CP become closer to prediction coefficient* 95%*; and EL decrease with increasing n.*
- *The Bayesian intervals based on informative priors perform better than those based on the non-informative prior.*

Survival method				
n	$s=2$	$s=4$	$s=6$	$s=7$
	EL and CP	EL and CP	EL and CP	EL and CP
2	8.31 and 0.949	13.97 and 0.950	19.36 and 0.949	22.01 and 0.951
3	6.49 and 0.949	10.39 and 0.950	13.96 and 0.951	15.71 and 0.952
4	5.74 and 0.950	8.91 and 0.952	11.72 and 0.953	13.06 and 0.952
5	5.37 and 0.953	8.18 and 0.9554	10.58 and 0.955	11.72 and 0.954
HPD method				
n	s= 2	$s=4$	$s=6$	$s=7$
	EL and CP	EL and CP	EL and CP	EL and CP
2	6.40 and 0.959	11.21 and 0.960	15.86 and 0.963	18.01 and 0.964
3	5.44 and 0.963	9.08*and*0.964	12.32 and 0.965	13.89 and 0.964
4	4.98 and 0.964	8.03 and 0.963	10.70 and 0.971	12.02 and 0.972
5	4.70 and 0.970	7.52 and 0.972	9.84 and 0.975	10.92 and 0.974

Table 4.6: Simulation of an interval Prediction for $Y_{(s)}$ for an informative prior.

4.2 Real data application

In this section we will discuss and some examples of application of order statistics in real situations and analyze the obtaining results.

4.2.1 Application of order statistics to health data

The applications of order statistics have been generally of the following two classes:

- *Instances in which short-cut or time-saving devices are appropriate for problems of estimation and/or tests of significance, the estimates and short-cut tests do not always attain the highest power efficiency, but are preferred because of ease of calculations without great sacrifice in efficiency.*
- *Applications in which no other known techniques provides a suitable answer with as high a power-efficiency; for example in case of censored data.*

Starting with an example of censored data, the concept is crucial throughout our study.

The data in this study is mentioned in [4], but initially are taken from "Sarhan" and "Greenberg" from "estimation of location and scale parameters by order statistics from singly and doubly censored samples".
The data represent a sample of 10 *individual systolic blood pressure readings, for technical reasons in measurement* 20% *of the smallest and* 30% *of largest observations were censored; so we have only the* 5 *central values.*
The aim of this study is to estimate the mean μ *and standard deviation* σ *of the population knowing the magnitude of the five central values and only the relative position of the censored observations. The following table contains the arranged data in size order and the coefficients selected from "Sarhan" and "Greenberg" from "estimation of location and scale parameters by order statistics from singly and doubly censored samples"; these coefficients are predetermined to provide the best linear unbiased estimate of* μ *and* σ*.*

ordered observations	Coefficients for	
	Mean = μ	standard deviation σ
1. Censored	0	0
2. Censored	0	0
3. 108	0.2050	-0.8898
4. 111	0.1038	-0.1101
5. 119	0.1122	-0.0262
6. 121	0.1198	0.0549
7. 125	0.4592	0.9711
8. Censored	0	0
9. Censored	0	0
10. Censored	0	0
Estimates	$\hat{\mu} = 118.9$	$\hat{\sigma} = 16.6$
Variance (in terms of σ^2)	0.1180	0.1713
Efficiency relative to uncensored sample %	84.78	33.62

Table 4.7: Health censored data

The calculations are as follows :

$$\hat{\mu} = 108(0.2050) + 111(0.1038) + 119(0.1122) + 121(0.1198) + 125(0.4592) = 118.9$$

$$\hat{\sigma} = 108(-0.8898) + 111(-0.1101) + 119(-0.0262) + 121(0.0549) + 125(0.9711) = 16.6$$

The estimate of the mean is 84.78% *efficient although* 50% *of the sample values only were known; as we can see the efficiency of estimating the mean were to be maximized with* 50% *of the sample observations.*
In the foregoing example we have obtained a maximized efficiency only with 50% *of the sample observations; here in the actual example we will illustrate short-cut procedure i.e; instead of using all observations in the estimation process specially when the there is large series of observations such as the infant deaths by age of death, U.S.A* .1955 *, some of them will be used and compare their efficiency.*
The data are taken from " Infant Mortality . Vital stat.special rep. Vol.46,N0.14,table 4.Washington,D,C.:National Office of vital statistics, (Aug.22),1957". Clearly the data are arranged in ascending order, as shown in table below.
The data given below follows the one-parameter exponential distribution . Suppose the distribution of infant deaths in is to be used as an independent check of the separation factor employed in life table to calculate life expectancy that is the average age of death under one year of age is desired.

Age at death	Average of death
Under 1 hour	8.008
1 hour or over , under 1 day	32.472
1 day or over ,under 2 days	11.692
2 days or over , under 3 days	7.449
3 days or over , under 4 days	3.783
4 days or over , under 5 days	2.232
5 days or over , under 6 days	1.739
6 days or over , under 1 week	1.279
1 week or over , under 2 weeks	4.374
2 weeks or over , under 3 weeks	2.475
3 weeks or over , under 4 weeks	1.848
Over 4 weeks , under 2 months	6.571
Over 2 months , under 3 months	5.008
Over 3 months , under 4 months	3.858
Over 4 months , under 5 months	3.125
Over 5 months , under 6 months	2.517
Over 6 months , under 7 months	2.094
Over 7 months , under 8 months	1.679
Over 8 months , under 9 months	1.459
Over 9 months , under 10 months	1.177
Over 10 months , under 11 months	1.049
Over 11 months , under 1 year	1.015
Total	106.903

Table 4.8: Infant deaths

Using only five observations to estimate the average age at death, using procedure "Ogawa's short-cut" procedure, the values chosen would be as follows:

Observation order number	Age at death
42.064	1.14
71.673	11.83
90.262	100.36
100.903	216.77
105.679	324.02

Table 4.9: The 5 observations of infant deaths

This short-cut procedure leads to an average age at death equal to 40.2

days. Dividing this by 365 *days of the year we get the separation factor:* $\frac{40.2}{365} = 0.110$ *while the separation factor in the abridged life tables A.7 for* 1955 *is* $f_0 = \frac{L_0 - l_1}{d_0} = \frac{97.670 - 97.354}{2.646} = 0.119.$
There is a slight discrepancy between the two value, this do not due to the fact that the former is based upon only five observations but it due to the method of estimating the life table separation factor for infants under one year age is slight overestimate .
The advantage of this "Ogawa's short-cut " method of estimating the actual average age at death is that with $k = 5$ *the distribution of the first* 39% *of the deaths need not to be known. Furthermore, if the estimate were to be based upon* $k = 3$ *optimally spaced points, less than one half (or the last part) of the distribution of deaths need be known.*

4.2.2 Prediction of order statistics in real situations

In the previous part we have see how using order statistics in estimation of unknown parameters, provide efficient estimators; in this section we focus on prediction of order statistics in real situations.

4.2.3 D_{SP} selection approach for predicting

When several underlying distribution are presenting, we need to choose the best to fit the data. In chapter 2 *we have discussed* 3 *selection approaches, in this study we will focus on the* D_{SP} *method.*
The data of this example provided in (Mann and Fertig in thier article "tables for obtaining Weibull confidence bounds and tolerance bounds based on the best linear invariant estimates of parameters of the extreme value distribution"technometrics A journal of staistics for the physical, chemical and engineering sciences vol.15,pp87-101,1973) and analyzed by the authors of [13].
Presentation of the data
A test airplane component's failure time, in which 13 *components were placed on test, and the test was terminated at the time of the* 10^{th} *failure; the failure times (in hours) of the* 10^{th} *components are the following*

D_1: 0.22 0.50 0.88 1.00 1.32 1.33 1.54 1.76 2.50 3.00.

Let $Y_1 = \ln D_1$.
Our aim is to predict the censored data noted $_{(i)}$ *applying Taylor series predicting method and using* D_SP *approach to discriminate the best underlying distribution between the normal distribution noted (ND) and smallest extreme value distribution (SEV), according to the following figure the both candidate distribution provide a good fitting for the data set. The MLE of the location* μ *and scale* σ *parameters are obtained with Newton-*

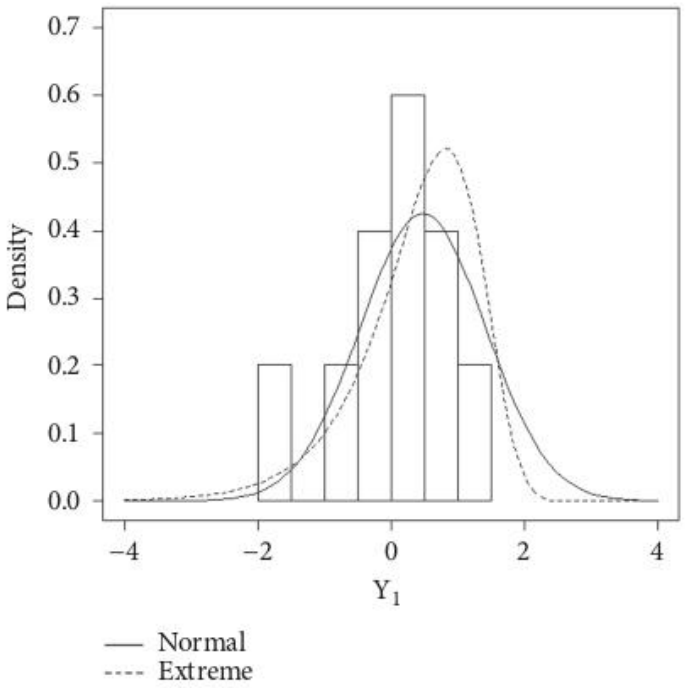


Figure 4.6: The histogram and the estimated density functions of the data D_1

Raphson algorithm and they are:

$$(\hat{\mu}_{ND}, \hat{\sigma}_{ND}) = (0.479, 0.938) \quad ; \quad (\hat{\mu}_{SEV}, \hat{\sigma}_{SEV}) = (0.821, 0.705)$$

The values of D_{SP} *using ND and SEV are respectively* 0.223 *and* 0.212, *it's evident that the best distribution of this data is SEV because the smaller value of* D_{SP} *obtained from this distribution.*
We can apply the Taylor series prediction to predict $Y_{(11)}, Y_{(12)}, Y_{(13)}$ *under the extreme value distribution, we obtain*

$$\hat{Y}_{(11)}, \hat{Y}_{(12)}, \hat{Y}_{(13)} = (1.098, 1.281, 1.567)$$

Consult [13] to see the performance evaluation of the other approaches D and RRML a Monte Carlo simulation carried out in [13].

4.2.4 Analysis of prediction using Health data

In chapter 2 *we have encountered several predictors either point or interval prediction. Egyptian researchers have introduced new method for prediction of order statistics, these predictors have yielded remarkable practical results and their performance when compared to other predictors was satisfactory.*
The data in this analysis study was treated by "Wu and Wu (2005)" and also analyzed by "Lawless (2003)" and "Barakat et al(2008b)", and analyzed again in objective of comparison in article [13] by "Barakat", O.Khaled and H.A .Ghonem (2020) ".
Cancer remission does not mean the cancer is cured, but it is an important milestone. Remission means that the signs and symptoms of cancer are reduced.In any clinical trial, for the cancer patients, the prediction of the duration of remission achieved by any remedy is an extremely important issue in facing disease, it marks a major turn in patient care.
The data set represents the duration of remission of 20 *leukemia patients, who are treated by one drug.The* 20 *ordered times in year are given in the table below:*

1.013	1.034	1.109	1.169	1.266	1.509	1.533	1.563	1.716	1.929
1.965	2.061	2.344	2.546	2.626	2.778	2.951	3.413	4.118	5.136

Table 4.10: The duration of remission of 20 leukemia patients.

The data was fitted by two-parameters exponential distribution and its validity for the complete data with parameters $\hat{\mu} = 1.013$ *and* $\hat{\gamma} = (1/0.876568)$ *is checked by Kolomorov-Smirnov* $(K - S)$ *statistic with the observed value is* 0.10259*and p-value* 0.9701

its distribution is written as follows :

$$E_{\gamma}(x-\mu) = \mathbf{f}(x,\mu,\gamma) = \frac{1}{\gamma}\exp\{-\frac{x-\mu}{\gamma}\}$$

$\mu < x$ *is the threshold parameter and* $\gamma > 0$ *is a scale parameter.*
Assume that we have observed the first 15 *observations and we we want to predict the next unobserved data by using the method CP2; when we*

put $n = 20$,$R = 15$ *and* $k = 1, \ldots, 5$*; the process is repeated* $10,000$ *times because* $Y_{(1)}, \ldots, Y_{(5)}$ *in sample of size* $n - R = 5$ *are random; then we compute the average to obtain suggested point D-prediction.*
The results of this study and the comparison with result of "Barakat et al.(2018b)" concerning the MLP method are presented in the following table.

k	Exact value	MLP	CP2
1	2.7780	2.6260	2.8549
2	2.9510	2.8806	3.1349
3	3.4130	3.2088	3.5125
4	4.1180	3.6713	4.0805
5	5.1360	4.4621	5.2135
MSE		0.1447	0.0114

Table 4.11: Prediction using CP2

Remark 4.2

- *The CP2 method outperform the MLP one in term of MSE, we remark that*

$$\frac{MSE(MLP)}{10} = MSE(CP2)$$

- *CP2 provides an accurate way to predict the duration of remission of cancer.*

4.2.5 Prediction using CP2

The data of this example handled by "Lawless (1982)", analyzed also by "Valiollahi et al.(2017)" in goal to compare the four methods MLP,BUP ,CMP and BP; and used for another comparison study in article [13] to complete this comparison with CP2; was taken from accelerator life test of 59 *conductors.The data set is represented in the table below as* 59 *observed order statistics.*

2.997	4.137	4.288	4.531	4.700	4.706	5.009	5.381	5.434	5.459
5.589	5.640	5.807	5.923	6.033	6.071	6.087	6.129	6.352	6.369
6.479	6.492	6.515	6.522	6.538	6.545	6.573	6.725	6.869	6.923
6.948	6.956	6.958	7.024	7.224	7.365	7.398	7.459	7.489	7.495
7.496	7.543	7.683	7.937	7.945	7.974	8.120	8.336	8.532	8.591
8.687	8.799	9.218	9.254	9.289	9.663	10.092	10.092	10.491	11.038

Table 4.12: The accelerator life test of conductors

The appropriate model for this data is the generalized exponential distribution $E_\gamma^\alpha(x)$ *with parameters* $\hat{\alpha} = 52.411$ *and* $\hat{\gamma} = 0.642$ *and its pdf is given by :*

$$\mathbf{f}(\alpha, \gamma, x) = \alpha\gamma(1 - \exp^{-\gamma x})^{\alpha-1} \exp^{-\gamma x} \quad x > 0.$$

The validity of this model is verified by $(K - S)$ *test with p-value*$= 0.532$. *We assume that we have observed the first* 20 *observations and we want to predict the next* 39 *unobserved data for different values of* k $(1, 3, 5, 8)$. *As the previous example the order statistics in the unobserved sample are random so the process have been repeated* 10.000 *times then compute the average to obtain point D-prediction; to diminish the resulted error from the simulation process.*
The results of this study and the comparison are presented in the following table.

k	Exact value	MLP	BUP	CMP	BP	CP2
1	6.4790	6.3690	6.4640	6.4360	6.5210	6.4318
3	6.5150	6.5570	6.6530	6.6260	6.8520	6.5584
5	6.5380	6.7450	6.8430	6.8190	7.1620	6.6880
8	6.7250	7.0300	7.1560	7.1120	7.8020	6.8896
MSE		0.0374	0.0745	0.0607	0.4162	0.0134

Table 4.13: Predictors for the accelerator life test

We remark that the CP2 performs better than the other predictors, and its MSE is minimized.

4.2.6 Prediction of order statistics using RRSS in real situation

The data in this illustrative study was treated by "E.G.Gervi, P.Nasiri, M.Salehi)[8].

To illustrate the performances of some predictors, they consider a real data set (noted Y) represents daily heat degree (Celsius) of the month of January for 3 *years* $(2009, 2014 and 2018)$ *of Nova Scotia province and BACCARO PT station for government Canada. Also this data are available in the historical section at ("www.climate.weather.gc.ca")*

Notation and explications of the data

- *Weibull distribution was the model fitted on* Y. $Y \sim We(a = 5.4, b = 33.3)$.
- *Let* $X = aY^{1/b}$ *has an exponential distribution.*
- *So we can transform* Y *to* X *using the relation above.*
- *the p-value of the goodness of fit* $(K - S)$ *test of the one parameter exponential is* 0.997 *so the hypothesis to fit this data with exponential data is approved.*

From the transformed data set X *we have* $r = (5.946, 5.935, 5.959)$ *so* $n = 3$ *or* $N = \dfrac{n(n+1)}{2} = 6$,$T = 17.84$. *Now we can apply the theoretical obtained results on* X *and* r.
The performance of the following predictors is evaluated using the bootstrap method A.9.
Our aim is to predict the upper record values $(s = 3, 4, 6, 7)$ *and their MSPE also to give the Bayesian prediction interval. The following results are for a symmetric SEL function and asymmetric PL function using both the noninformative and informative prior respectively.*

$n = 3$	PL	PL MSPE	SEL	SEL MSPE
$s = 3$	3.89	46.56	10.70	48.80
$s = 4$	6.74	59.54	14.27	74.97
$s = 6$	12.31	93.32	21.41	142.16
$s = 7$	15.08	113.72	24.98	183.18

Table 4.14: Bayes point predictors for $Y_{(s)}$ for Jeffery's prior

In this case we will compare the results above using an informative prior with hyperparameters $\alpha = 1.5$ *and* $\beta = 1.5$ *the results are the following:*

$n=3$	PL	PL MSPE	SEL	SEL MSPE
$s=3$	3.42	38.30	8.93	28.46
$s=4$	5.93	49.54	11.90	41.72
$s=6$	10.83	77.08	17.85	73.93
$s=7$	13.27	93.12	20.83	92.87

Table 4.15: Bayes point predictors for $Y_{(s)}$ for informative prior

Now predicting with an interval using both survival and HPD methods for two different cases noninformative and informative prior.

$n=3$	Survival	its expected length	HPD	Expected length
$s=3$	$[1.66; 33.26]$	31.6	$[0.40; 25.11]$	24.71
$s=4$	$[2.83; 41.77]$	38.94	$[1.38; 30.27]$	28.89
$s=6$	$[5.44; 58.47]$	53.03	$[3.49; 43.13]$	39.64
$s=7$	$[6.82; 66.73]$	59.91	$[4.49; 59.04]$	54.10

Table 4.16: Bayesian 95% prediction interval for $Y_{(s)}$ for Jeffery's prior

Using an informative prior, we compute both the interval based on the survival and HPD methods for this real data set and the results are below.

$n=3$	Survival	its expected length	HPD	Expected length
$s=3$	$[1.47; 26.42]$	24.95	$[0.45; 17.51]$	17.06
$s=4$	$[2.52; 32.99]$	30.47	$[1.98; 25.40]$	23.42
$s=6$	$[4.87; 45.85]$	40.98	$[3.15; 29.70]$	26.55
$s=7$	$[6.12; 52.19]$	46.07	$[4.32; 38.76]$	34.44

Table 4.17: Bayesian 95% prediction interval for $Y_{(s)}$ for Jeffery's prior

Concluding remarks

- *The predictors based on informative prior performs better than the case noninformative.*
- *The HPD method has a better performance than the survival method.*
- *All the MSPE are increasing.*

4.2.7 Prediction based on record values

The data used in this example are taken from "I.R.Dunsmore"(the future of occurrence of records ,Ann.Inst.Stat.Math.35(1983), pp.267-277.) to be

analyzed by the authors of the article [17].
A rock crushing machine has to be rest if any operation, the size of the rock being crushed is larger than any other rock that has been crushed before.The following data represent the sizes dealt with up to the third time that the machine has been rest.

9.3 0.6 24.4 18.1 6.6 9.0 14.3 6.6 13.0 2.4 5.6 33.8.

Using the assumption that these data come from an $Exp(\theta)$ distribution. The record values represent the sizes at the operation when resetting was necessary, the k-records extracted from the given data set are:

i	1	2	3	4	5	6
$R_{(i)1}$	9.3	24.4	33.8	–	–	–
$R_{(i)2}$	0.6	9.3	18.1	24.4	–	–
$R_{(i)3}$	0.6	9.3	14.3	18.1	–	–
$R_{(i)4}$	0.6	6.6	9	9.3	13	14.3

Let consider two case of the prior distribution , informative case with the hyperparameters $\alpha = 2$ and $\beta = 2$, and the case of Jeffery's prior i.e; the non-informative prior; considering $k = 2$ is the number of the used record values, m represent the future sample of size 5,10 *and* 20 *the results of prediction: the interval prediction* 95%*, the point prediction, the prediction of the mean in future sample denoted mean and the estimated MSE i.e; we don't use the real MSE because is a function of θ using the estimate of θ we have the estimated MSE; are given below:*

m	j	Survival method	pivotal method	HPD method	point predictor	MSE
5	1	[0.043; 8.63]	[0.043; 8.63]	[0; 6.58]	2.03	2.69
	4	[2.48; 39.10]	–	[2.41; 37.22]	13.04	101.12
	5	[4.74; 69.65]	–	[4.69; 68.25]	23.20	297.15
	Mean	[2.34; 28.56]	[2.73; 37.76]	[2.27; 26.69]	10.16	37.64
10	1	[0.031; 7.39]	[0.031; 7.39]	[0; 5.44]	1.63	3.21
	4	[1.31; 25.98]	–	[1.25; 22.99]	7.79	75.82
	9	[7.33; 97.55]	–	[7.25; 92.98]	31.38	1221.37
	10	[10.68; 151.56]	–	[10.61; 147.01]	47.64	2647.76
	Mean	[4.19; 48.79]	[3.60; 58.23]	[4.12; 45.66]	16.27	244.95
20	1	[0.015; 3.70]	[0.015; 3.70]	[0; 2.72]	0.81	0.81
	4	[0.60; 11.47]	–	[0.54; 11.02]	3.53	15.56
	10	[2.78; 32.73]	–	[2.71; 30.85]	10.88	147.93
	18	[9.47; 100.66]	–	[9.42; 97.67]	34.12	1455.19
	20	[15.08; 178.76]	–	[15.03; 172.64]	58.52	4132.54
	Mean	[4.82; 46.85]	[14.11; 56.47]	[4.75; 44.21]	16.27	244.34

Table 4.18: Prediction using records with non-informative prior

In this following table we will see the results of prediction based on informative prior with hyperparameters $\alpha = 2$ and $\beta = 2$.

m	j	Survival method	pivotal method	HPD method	point predictor	MSE
5	1	[0.062; 14.79]	[0.062; 14.79]	[0; 10.88]	3.25	12.44
	4	[3.38; 70.72]	–	[3.31; 67.79]	20.88	527.76
	5	[6.44; 125.78]	–	[6.36; 120.93]	37.14	1509.33
	Mean	[3.16; 52.40]	[2.77; 61.53]	[3.11; 50.12]	16.27	246.17
10	1	[0.021; 4.31]	[0.021; 4.31]	[0; 3.29]	1.02	0.65
	4	[0.96; 14.29]	–	[0.94; 14.09]	4.78	14.43
	9	[5.50; 24.31]	–	[5.31; 23.40]	9.60	232.52
	10	[8.01; 82.55]	–	[7.82; 80.91]	29.76	508.24
	Mean	[3.16; 26.02]	[3.66; 34.73]	[3.12; 25.75]	10.16	36.79
20	1	[0.011; 2.16]	[0.011; 2.16]	[0; 1.64]	0.51	0.16
	4	[0.44; 6.46]	–	[0.40; 6.33]	2.20	2.96
	10	[2.10; 17.47]	–	[1.98; 16.95]	6.79	28.14
	18	[7.22; 53.39]	–	[7.14; 51.98]	21.31	276.85
	20	[11.44; 96.14]	–	[11.38; 94.25]	36.55	788.00
	Mean	[3.70; 24.61]	[4.25; 33.07]	[3.53; 24.25]	10.16	36.37

Table 4.19: Prediction using records with informative prior

Concluding remarks: *Using the results above and referring to [14] we can conclude that:*

- *The classical results and those of the Bayesian approach are quite*

close when we use non-informative prior.

- *Using the informative priors allows more precise inference.*
- *The Bayesian methods performs better than the classical method in case using the informative prior with suitable hyperparameters.*
- *the width of the Bayesian interval decreases as the sample size m increases.*
- *When the hyperparameters increase, the Bayesian intervals get shorter.*
- *The HPD method is more precise than the others.*
- *The estimated MSE are increasing with respect to j when other parameters are fixed, also we can see that it become smaller as m increase.*
- *The estimated MSE are decreasing when the hyperparameters are increasing.*

Conclusion and perspectives

Much has been learned about the prediction of order statistics in the past 25 years.
In this dissertation we considered prediction of order statistics, which has essentially covered the different method of prediction, ranging from the frequentist perspective to exploring the Bayesian framework.starting our journey from the frequentist side of prediction we have shed light on the point prediction given a several predictors as best unbiased predictor, then the interval prediction; without overlooking the method of selecting the most informative order statistics to improve the prediction process; as delving into the Bayesian framework by considering different parent distribution as exponential and Weibull, we have derived several Bayes point predictors with respect to different loss functions, then we have seen the Bayesian interval prediction using different samples schemes.
We have also compared the point predictors obtained using the MSE and coverage probability in case of interval prediction.
Some of the obtained results are illustrated using real data and simulations.
Finally still a lot of work to be done in terms of predicting order statistics; moving away from location-scale families and venturing into using more general families of distribution, also as future perspectives it is time to research more about nonlinear prediction.

Bibliography

[1] A.chadli ,A.Meradji . Prediction Bayesienne de statistiques d'ordre avec des donnees groupees, cas d'une loi exponentielle; Rev.Sci. Technol., synthese 29 :36-41(2014).

[2] Alex Dysto , Martina Cardone , Cynthia Rush. " The Most Informative Order Statistics and its Application to Image Denoising ", 27 Jan 2021.

[3] Barry C. Arnold, N. Balakrishnan,H. N. Nagaraja. A First Course in Order Statistics, the Society for Industrial and Applied Mathematics 2008.

[4] Bernard G.Greenberg, Ph.D,F.A.P.H.A.,and Ahmed E.sarhan, Ph.D. Application of order statistics to health data, vol.48,NO.10. A.J.P.H.

[5] Christian P.Robert.Le choix Bayesien principes et pratique ;Springer-Verlag France, Paris, 2006.

[6] Christian P.Robert .The Bayesain choice from decision-theoretic foundations to computational implementation; Springer science +Business media 2007.

[7] Debasis Kundu, Mohammad Z. Raqab2. Bayesian inference and prediction of order statistics for a Type-II censored Weibull distribution.

[8] Eshan Golzade Gervi, Parviz Nasiri, Mahdi Salehi. An overview of bayesian prediction of future record statistics using upper recor ranked set sampling scheme, Nonlinear Anal Appl; 12(2021) No.1 493-507.

[9] Essam K. AL-Hussaini , Abd EL-Baset A. Ahmad On Bayesian predictive distributions of generalized order statistics , Metrika Springer-Verlag 2003.

[10] Gelman, al. Bayesian Data Analysis; Journal of the American Statistical Association 45 (2),2004.

[11] Gentle, James E. (2009), Computational Statistics, Springer, p. 63, ISBN 9780387981444.

[12] H.A. David, H.N.Nagaraja. Order statistics, third edition, Wiley USA , Canada.

[13] Haroon M .Barakat, Osama M.KHaled, Hadeer A.Ghonem. New methods for prediction of future order statistics . Quality technology and quantitative management, Taylor and Francis group; 30 Jun 2020.

[14] Jafar Ahmadi, S.M.T.K MirMostafaee; N.Balakrishnan. Bayesian pridiction of order statistics based on k-record value from exponential distribution, Taylor and Francis group; 4, august 2011.

[15] Jean-Michel Marin, Christian P. Robert. Bayesian Essentials with R; Springer Science+Business Media New York 2014,second edition.

[16] Jyun-you Chiang, Shuai Wang, Tzong-Ru Tsai, Ting Li.Model selection approaches for predicting future order statistics from type II censored data, Hindawi Mathematical problems in engineering, 8 October 2018.

[17] Keating, Jerome P and Mason, Robert L,Pitman's measure of closeness.Sankhy: The Indian Journal of Statistics, Series B,22.32,1985.JSTOR.

[18] Kenneth S. Kaminsky and Paul I.Nelson. Best linear unbaised prediction of order statistics in location and scale families.Department of mathematics, Bucknell university, Lewisburg, Penn, 17837, usa, 05 Apr 2012.

[19] Kenneth S. Kaminsky and Paul I.Nelson . Prediction of Order statistics.

[20] K. G. Mehrotra and P. Nanda. "Unbiased estimation of parameters by order statistics in the case of censored samples," Biometrika, vol. 61, pp. 601–606, 1974.

[21] Mohammad Ahsanullah, Valery B.Nevzorov, mohammad Shakil.An introduction to order statistics, atlantis studies in probability and

statistics series editor| C. P. TSOKOS, Volume 3, University of South Florida Tampa, Tampa, USA, 2013.

[22] Mohamed Z.Raqab, Haikady N Nagaraja. On some predictors of future order statistics. Metran January 1995. (Researchgate).

[23] Mahdi Salehi, JAfar Ahmadi.Record ranked set sampling scheme, Metron 72,351-365 (2014).

[24] N. Balakrishnan and A. C. Cohen, Order statistics and inference, Statistical Modeling and Decision Science, Academic Press, Inc., Boston, MA, 1991.

[25] N. Balakrishnan ,A. Clifford Cohen. Order Statistics and Inference Estimation Methods, ACADEMIC PRESS, INC. Harcourt Brace Jovanovich, Publishers.

[26] Turkson, Anthony Joe, Francis Ayiah-Mensah, and Vivian Nimoh. "Handling Censoring and Censored Data in Survival Analysis: A Standalone Systematic Literature Review." International Journal of Mathematics and Mathematical Sciences 2021 (September 24, 2021): e9307475.

Appendices

Appendix A

Definitions

A.1 Most used notions

A.1.1 Grouped data

We assume that only $n_j, 1 < j < m$, are available.

Theorem A.1 *The MLE of θ based on the grouped data $\{n_1 \dots n_m\}$ uniquely exists if and only if $n_m < n$. If $n_m < n$ then the MLE of θ is given as the unique solution of the following equation:*

$$\sum_{j=1}^{m-1} n_j \frac{\alpha_j(t_j - t_{j-1})}{1 - \exp^{\alpha_j(t_j - t_{j-1})/\theta}} = \sum_{j=1}^{m-1} n_j \alpha_j (t_j - t_{j-1}) + \sum_{j=1}^{m} n_j (\sum_{l=1}^{j-1} \alpha_l (t_l - t_{l-1}))$$

$\alpha_i < \infty$ are known constants .

A.1.2 Random Censoring

With the ith item let us now associate a random variable C_i called the censoring time whose cdf $\mathbf{F}_c$ is free of θ. Define: $\mathbf{T}_i = \min(X_i, C_i)$ and $D_i = 1$ if $\mathbf{T}_i = X_i$ and $D_i = 0$ otherwise. Let us assume X_i and C_i are independent, and we observe $(\mathbf{T}_i, D_i)$, $i = 1$ to n. Thus, each lifetime is censored by an independent time, and we also know whether our observation is the life length or the corresponding censoring time. This scheme is known as a random censoring scheme *and is very common in clinical trials: patients enter into the experiment at random time points, while the study itself is terminated at a prespecified time; the likelihood of this*

situation is given by the following formula:

$$\mathbf{L}(\theta|t,d) = \prod_{i=1}^{n} \{\mathbf{f}(x_i;\theta)[1-\mathbf{F}_c(x_i)]\}^{d_i} \{\mathbf{f}_c(c_i)[1-\mathbf{F}(c_i;\theta)]\}^{1-d_i}$$

Where $\mathbf{f}_c$ *is the pdf of* C_i.
If the $C_i's$ *are constants, say t, then the previous equation reduce to:*

$$\mathbf{L}(\theta|t,d) = \prod_{i=1}^{r} \mathbf{f}(x_i;\theta)\{1-\mathbf{F}(t;\theta)\}^{n-r}$$

$x_1, \ldots x_r < t; d_i = 0,1$ *and* $\sum d_i = r$*.So we obtain a type I censored sample.*

A.1.3 Progressive Censoring

Of the n items put on test, suppose we remove n_1 *unfailed items at time* $t_1; n_2$ *unfailed items at time* $t_2, \ldots,$ *where* $t_1 \leq t_2 < \ldots$ *are prespecified times. the experiment will be terminated at time* t_i *if* n_i *exceeds the number of unfailed items remaining at that time: this is known as a type I progressive censoring scheme.*
A type II version of this sample in that scheme at the time of r_i*th failure,* n_i *unfailed items are removed from the study* $(i \geq 1)$*; this is continued until each item is taken care of either due to its failure or due to its removal from the experiment.*
to note that the likelihoods are more complex in these cases.But the Type II case is tractable.See Lawless (1982,pp.33-34) for more information.

A.2 EMVLUE

The variance and covariance of the estimators **??** *and* **??** *are given as:*

$$Var(\hat{\mu}) = \frac{\sigma^2(\alpha'\Sigma^{-1}\alpha)}{\triangle} \tag{A.1}$$

$$Cov(\hat{\mu},\hat{\sigma}) = -\frac{\sigma^2(\alpha'\Sigma^{-1}1)}{\triangle} \tag{A.2}$$

For the symmetric case the best linear unbaised estimates of μ and σ are:

$$\hat{\mu^*} = \frac{1'\Sigma^{-1}X}{1'\Sigma^{-1}1} \tag{A.3}$$

$$\hat{\sigma^*} = \frac{\alpha'\Sigma^{-1}X}{\alpha'\Sigma^{-1}\alpha} \tag{A.4}$$

and the corresponding covariance of the estimators is zero and their variances are given as:

$$Var(\hat{\mu^*}) = \frac{\sigma^2}{1'\Sigma^{-1}1}$$

$$Var(\hat{\sigma^*}) = \frac{\sigma^2}{\alpha'\Sigma^{-1}\alpha}$$

For some distributions the MVLUE *of the location and scale parameters can be expressed in simplified form.The following lemma is useful to find the inverse of the covariance matrix.*

Lemma A.1*Let $\Sigma = (\sigma_{r,s})$ be $n \times n$ matrix with elements, which satisfy the relation:*

$\sigma_{rs} = c_r d_s \quad 1 \leq r, s \leq n$. c_i and d_i are some positive.Then the inverse: $\Sigma^{-1} = (\sigma^{r,s})$ has elements given as follows:

$$\sigma^{1,1} = \frac{c_2}{c_1(c_2 d_1 - c_1 d_2)}$$

$$\sigma^{n,n} = \frac{d_{n-1}}{d_n(c_n d_{n-1} - c_{n-1} d_n)}$$

$$\sigma^{k+1,k} = \sigma^{k,k+1} = -\frac{1}{c_{k+1} d_k (c_k d_{k+1}}$$

$$\sigma^{k,k} = \frac{c_{k+1} d_{k-1} - c_{k-1} d_{k+1}}{(c_k d_{k-1} - c_{k-1} d_k)(c_{k+1} d_k - c_k d_{k+1})} \quad k = 2, \ldots, n-1 \tag{A.5}$$

and $\sigma^{i,j} = 0$, if $|i - j| > 1$.

A.3 Functions

A.3.1 Underlying distribution

Before making modeling decisions, we need to know the underlying data distribution.For any given data-set, there's some process that generated

that data-set; The underlying process may be known or unknown. In some areas of researches it is possible to make assumptions about the specific distribution of the variable under the study; for example the variables whose are determined by an infinite number of independent random events follows the normal distribution as the height and weight of the world population.From predictive perspective, it is often desirable to understand the underlying shape of a population's distribution.In order to determinate this underlying distribution, it is common to fit the observed distribution by a theoretical distribution by comparing the observed frequencies in the data to the frequencies expected by the the theoretical distribution; i.e; a χ^2 *test of goodness of fit .*

A.3.2 Pitman's measures of closeness

[14] Is set of statistical measures used to compare the closeness or similarity between two probability distribution or statistical models as estimators predictors ...; these measures are used in statistical inference and decision theory to compare competing models or estimators and select the one that best represents the underlying population or provides more accurate estimates.There are three commonly used Pitman's measures of closeness:

- *Pitman's measure of difference: it quantifies the discrepancy between two distribution by calculating the maximum difference in probability assigned to any event; finally it provides a single numerical value representing the larges discrepancy between them .*
- *Pitman's measure of efficiency: evaluates the relative efficiency of two statistical models or estimators, it compares the variances of the estimates produced by two models and indicates the more precise estimates.*
- *Pitman's measure of closeness: assess the overall closeness or similarity between two distributions by combining the concepts of difference and efficiency.*

Let $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_k$ *k real valued estimators of a real parameter* θ*. Denote the absolute error loss of ith estimator by* $L_i = |\hat{\theta}_i - \theta|$ *for* $i = 1, 2, \ldots k$*. For*

$k = 2$ Pitman introduced $\mathbf{P}(L_1, L_2)$ as a pairewise measure of closeness (nearness) for the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$.

A.3.3 Mixture model

One of the most important models in reliability is the finite mixture model.Finite mixture of exponential distributions is considered as a possible model for a lifetime distribution. The mixture of two exponential distributions $E_{\lambda i}(x-\mu_i), x \geq \mu_i, i = 1, 2$, is given by:

$$\mathbf{F}(x : a; \mu_1; \lambda_1; \mu_2; \lambda_2) = aE_{\lambda 1}(x - \mu_1) + \bar{a}E_{\lambda 2}(x - \mu_2) \tag{A.6}$$

$$= 1 - a\exp(-\lambda_1(x - \mu_1)) - \bar{a}\exp(-\lambda_2(x - \mu_2)), \tag{A.7}$$

where $0 \leq a \leq 1$ and $\bar{a} = 1 - a$. the parameter a is known as the mixing proportion.The quantile function $Q(y : a; \mu_1, \lambda_1; \mu_2, \lambda_2)$, $0 < y < 1$; i,e; the inverse function $\mathbf{F}(x : a; \mu_1, \lambda_1; \mu_2, \lambda_2)$ is the only obstacle that encounters us to get the point D-pridector for the mixture cdf $\mathbf{F}(x : a; \mu_1, \lambda_1; \mu_2, \lambda_2)$ either for CP1 or for the CP2 problem.

A.3.4 Mutual information

mutual information between two variable X and Y denote as: I(X;Y)=I(Y;X) measures the reduction in uncertainty about one of the variables given the knowledge of the other, letting p(x,y) be the joint density of (X, Y) and p(x), p(y) be the marginals; the mutual information is calculated as:

$$I(X;Y) = \int\int p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx\, dy \tag{A.8}$$

.

Notice that there is a relationship between the mutual information and the differential entropy:

$$I(X;Y) = h(y) - h(Y|X)$$

where the entropy is defined as

$$h(Y) = -\int p(y) \log p(y)\, dy \tag{A.9}$$

and the conditional entropy is given by :

$$h(X|Y) = \int\int p(x,y) \log(p(x)/p(x,y))\, dy\, dx. \tag{A.10}$$

If X and Y are independent knowing one delivers no information about the other, so the mutual information is zero, if X is deterministic function of Y and Y is deterministic function of X then knowing one give the complete information on the other; if X and Y are discrete the mutual information is the same amount of information contained in X or Y alone.If X and Y are continuous the mutual information is infinite since $h(Y|X) = -\infty$.

A.4 One and Two sample problem

In statistics one-sample problem and two-sample problem refer to different types of hypothesis testing or statistical analysis.

A.4.1 One-sample problem

The one-sample problem studies the analysis of a single set of data points or observations taken from a population, the goal is to make inferences about a given parameters, in one-sample problem,you compare the sample data to hypothesized value or a known population parameter and assess whether the data provides enough evidence to support or reject a particular claim.
For instance: using one-sample t-test to determine if the average blood pressure of a sample of patients is significantly different from known average blood pressure value.
One-sample prediction: *refers to make predictions for a new or unobserved data point based on the information provided by a single sample.*

A.4.2 Two-sample problem

The two-sample problem deals with comparing two distinct sets of data points or observations, typically from two different groups or populations; the objective is to examine if there is a statistically significant difference between the characteristics or measurements of these two groups; this type of analysis helps to determine whether the observed differences are likely due to random variation or if they indicate a genuine difference in the populations.

As an example: comparing the average exam scores of students who attended two different types of test preparation courses to see which course leads to higher scores than other.
Two-sample prediction: *might focus on making predictions between two different samples of data.*

A.5 Coverage probability

Is a concept used in statistics to measure the reliability of confidence interval in capturing the true population parameter, also named confidence interval coverage probability.The coverage probability indicates the proportion of times that the confidence interval procedures will result in intervals that contain the true value of the parameter being estimated.A common misconception is that a 95% *confidence interval means that the parameter value is estimated with* 95% *probability. In reality, the confidence level (such as* 95%*) refers to the long-term success rate of the confidence interval procedure in capturing the true parameter value. If the coverage probability matches the specified confidence level, it indicates that the confidence interval method is working as intended.*

A.6 Goodness-of-fit test

Is a procedure used to assess how well observed data fits a theoretical distribution or model, in other words these tests evaluates whether the collected data follows a specific model or distribution that was already expected to follow, it quantifies the level of agreement or disagreement between the observed data and the theoretical distribution testing against.For instance we have:

- χ^2*goodness-of-fit test: is used for testing the fit between observed and expected frequencies.*
- *Kolmogorov-Smirnov goodness-of-fit test usually is used to evaluate whether the observation follows a known normal distribution; by and large it assesses the similarity between the cdf of the observed data and the cdf of the theoretical distribution.*

A.7 Life table

Life table is a table that displays the life expectancy and the probability of dying at each age or age group for a given population according to the age specific death rates prevailing at that time.The life table gives an organized complete picture of a population's mortality.
A complete life table contains data for every single year of age, an abridged life table typically contains data by 5 *or* 10 *years age intervals. For example see the following picture.*

Age interval	Years in interval	Linearity Adjustment	Reported pop/ births	Adjusted deaths	Mortality rate	Probability of dying	Probability of surviving	Pop surviving (expected)	Deaths (expected)	Years lived in interval	Cumulative yrs lived in interval	LE: Life Expectancy
(years)	n_x	a_x	N_x	$d_{(adj)}$	m_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
<1	1	0.1	141158	678.1	0.00480	0.00478	0.99522	100000	478	99570	7828154	78.28
1-4	4	0.5	531025	153.0	0.00029	0.00115	0.99885	99522	115	397858	7728585	77.66
5-9	5	0.5	687357	88.0	0.00013	0.00064	0.99936	99407	64	496876	7330727	73.74
10-14	5	0.5	719258	86.0	0.00012	0.00060	0.99940	99343	59	496569	6833851	68.79
15-19	5	0.5	726266	446.0	0.00061	0.00307	0.99693	99284	304	495659	6337282	63.83
20-24	5	0.5	747927	621.1	0.00083	0.00414	0.99586	98980	410	493873	5841623	59.02
25-29	5	0.5	708376	695.1	0.00098	0.00489	0.99511	98570	482	491642	5347750	54.25
30-34	5	0.5	743386	799.1	0.00107	0.00536	0.99464	98087	526	489121	4856108	49.51
35-39	5	0.5	759543	957.1	0.00126	0.00628	0.99372	97561	613	486275	4366987	44.76
40-44	5	0.5	762579	1344.1	0.00176	0.00877	0.99123	96949	851	482616	3880712	40.03
45-49	5	0.5	741136	1788.2	0.00241	0.01199	0.98801	96098	1152	477609	3398096	35.36
50-54	5	0.5	679033	2250.2	0.00331	0.01643	0.98357	94946	1560	470827	2920487	30.76
55-59	5	0.5	636723	3393.3	0.00533	0.02630	0.97370	93385	2456	460787	2449660	26.23
60-64	5	0.5	496072	4223.4	0.00851	0.04168	0.95832	90930	3790	445173	1988872	21.87
65-69	5	0.5	385226	5691.6	0.01477	0.07124	0.92876	87140	6208	420178	1543699	17.72
70-74	5	0.5	302778	8290.8	0.02738	0.12814	0.87186	80932	10371	378731	1123522	13.88
75-79	5	0.5	252158	11004.1	0.04364	0.19674	0.80326	70561	13882	318100	744791	10.56
80-84	5	0.5	166000	11358.2	0.06842	0.29214	0.70786	56679	16558	242000	426691	7.53
≥85	14	0.5	104337	14453.5	0.13853	1.00000	0.00000	40121	40121	184691	184691	4.60

A.8 Hazard function

The Hazard function is a concept in survival analysis and reliability engineering.It represents the instantaneous rate of failure or hazard at given

time (t) *,i.e; it describes the probability that an event will occur at time* (t) *given that the subject has survived up to that time.It is denoted* $h(t)$ *and its formulas is:*

$$h(t) = \lim_{h\to 0} \frac{P[t \leq T < t + h | T \geq t]}{h} = \frac{\mathbf{f}(t)}{\bar{\mathbf{F}}(t^-)} \quad \textit{Where:}$$

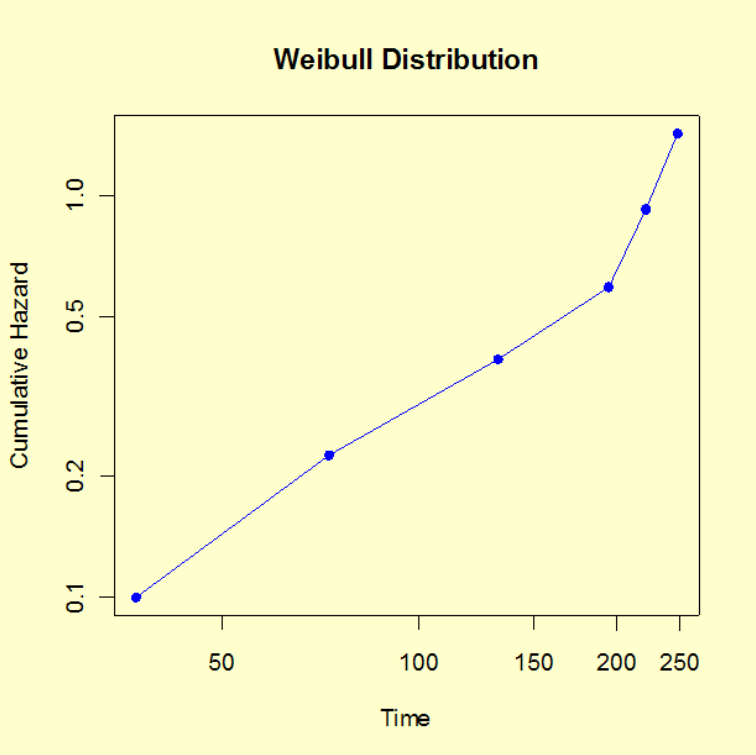
T: is a non-negative r.v. representing the time until some event of interest.

$\mathbf{f}(t)$*: is the pdf of the r.v. T.*

$\bar{\mathbf{F}}(t)$ *is the survivor function.*

It is important to note that in contrast to the survival function, which focuses on note failing, the Hazard function focuses on failing, that is on the event occurring. Thus in some sense the Hazard function Can be considered as giving the opposite side of the information given by the survivor function.

The following figure present an hazard function of Weibull distribution.



A.9 Bootstrap

Bootstrapping is a statistical procedure that resamples a single data set to create many simulated samples. This process allows to calculate standard errors, construct confidence intervals, and perform hypothesis testing for

numerous types of sample statistics. Bootstrap methods are alternative approaches to traditional hypothesis testing and are notable for being easier to understand and valid for more conditions.
Bootstrapping resamples the original dataset with replacement many thousands of times to create simulated datasets. This process involves drawing random samples from the original dataset. Here's how it works:

- *The bootstrap method has an equal probability of randomly drawing each original data point for inclusion in the resampled datasets.*
- *The procedure can select a data point more than once for a resampled dataset. This property is the "with replacement" aspect of the process.*
- *The procedure creates resampled datasets that are the same size as the original dataset.*

The process ends with your simulated datasets having many different combinations of the values that exist in the original dataset. Each simulated dataset has its own set of sample statistics, such as the mean, median, and standard deviation. Bootstrapping procedures use the distribution of the sample statistics across the simulated samples as the sampling distribution.