People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

UNIVERSITY MOHAMED KHIDER, BISKRA

FACULTY OF EXACT SCIENCES

DEPARTEMENT OF MATHEMATICS

Thesis Submitted in Partial Execution of the Requirements of the Degree of

Master in "**Mathematics**"

Option: Probability and Statistics

By Melle

BENZAOUI Souhila

Title :

Extropy Estimation

Examination Committee Members:

Dr.	ABDELLI Jihane	UMKB	President
Dr.	ZOUAOUI Nour El Houda	UMKB	Supervisor
Dr.	BERKAN Hassiba	UMKB	Examiner

02/06/2025

Dedication

Praise be to Allah, for no effort is ever complete, and no achievement is ever sealed except by His grace. We walk our path with His guidance and achieve our goals through His mercy. With love, gratitude, and appreciation, we praise Allah at the beginning and the end.

With all the love I hold in my heart, I dedicate my graduation project to:

To the one whose name I carry with pride, the one who has supported me since childhood and illuminated my path toward achieving my dreams. To the one who raised me, sacrificed for me, and whose words will forever remain the guiding stars in my life.

"My dear father"

To the one who surrounded me with love and kindness, who made me feel secure, whose prayers were the secret behind my success, and who guided and supported me whenever I stumbled

" My beloved mother"

To the stars that always light my way, the source of my strength, the most beautiful part of my life, and my partners in every smile, tear, and sigh my dear siblings

"Ziad, Charaf Eddin,AbdElhak, Amani"

To my dear friends" Noura, Khadija, Dalel and Rihab ",who have cheered me on, offered words of encouragement, and stood by my side through it all.

Acknowledgment

I extend my sincerest appreciation and gratitude to Dr "Zouaoui Nour El Houda", my supervisor, whose unwavering dedication and mentorship have been instrumental in shaping my professional growth and success. Their invaluable guidance, encouragement, and expertise have not only enriched my understanding but have also inspired me to strive for excellence in every endeavor.

Thank you, Dr "Zouaoui Nour El Houda", for your unwavering support and for being a beacon of inspiration in my professional journey.

I would like to thank all my teachers who have taught me throughout my academic journey. Additionally, I would also like to thank the jury members, Professor "ABDELLI Jihane" and Dr "BERKAN Hassiba", for accepting to examine and evaluate this work. I also wish to thank all our teachers in the Department of Mathematics at Mohamed Khider University, who contributed to our education during the Bachelor's and Master's years.

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

Abstract

Extropy has recently emerged as a complementary concept to entropy, introduced by Lad et al. (2015). It measures the degree of uncertainty or dispersion in probability distributions.

This master's thesis studies a new way to estimate extropy by using upper record values from the Weibull distribution, which is flexible and widely used in reliability analysis. We used two main methods: the first is Maximum Likelihood Estimation (MLE) to get accurate parameter values, and the second is Bayesian estimation with MCMC algorithms.

We also used numerical simulation to test how well these methods work.

Key words: *Extropy, Upper record values, Weibull distribution, Maximum Likelihood Estimation, Bayesian estimation.*

Notations and symbols

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

L_n^X	Lower record values
$L_{n,k}^X$	nth value of lower k-records
$f_{L_n^X}(x)$	Probability density function of nth lower record values
$f_{L_m^X,L_n^X}(x,y)$	The joint probability density function of mth and nth lower records
$f_{L_n^X \mid L_m^X}(y\mid x)$	The conditional probability density function of nth lower given mth lower record
$f_{L_{n,k}^X}(x)$	Probability density function of nth value of lower k-records
$f_{L_{m,k}^X,L_{n,k}^X}(x,y)$	The joint probability density function of mth and nth lower k-records
$f_{L_{n,k}^X \mid L_{m,k}^X}(y\mid x)$	The conditional probability density function of nth lower given mth lower k-record
X_i	Random variables with $i = 1, 2, \ldots$
et al	and others
$L(\cdot)$	The likelihood function
$l(\cdot)$	The log likelihood function
$S(\cdot)$	The score function
$I(\cdot)$	The value of the ficher information
$MLEs$	Maximum likelihood estimators
$J(X)$	Extropy
FIM	Ficher information matrix

Contents

List of Figures

List of Tables

Introduction

The study of uncertainty concepts is considered a central topic in the field of statistics. Shannon (1948)[21] defined the concept of entropy as a measure of disorder and irregularity within probability distributions. Since then, its applications have expanded to various fields, including psychology, economics, and image recognition.

After seventy years, Lad et al (2015)[11] found a new measure called extropy, which looks at how values are spread out instead of just focusing on where they cluster together, giving us a different way to understand probability distributions. This provides a useful complement for understanding the properties of random data.

The concept of extropy has witnessed significant theoretical and practical development. Qiu (2017)[17] worked on developing extropy properties based on rank statistics and record values,while Qiu et al (2018)[19] proposed specific definitions for extropy to compare uncertainty between two random variables. Additionally, Becerra et al (2018)[4] employed this concept in the field of automatic speech recognition.

When studying the inferential aspects of extropy, the need to develop effective estimators becomes evident. In this context, Qiu and Jia (2018)[18] proposed extropy estimators for use in goodness-of-fit testing applications. Irshad and Maya (2022)[9] developed a basic nonparametric logarithmic estimator for the extropy function. More recently, Zouaoui (2023)[22] explored the assessment of increasing uncertainty in record values and presented related characterization results.

This thesis aims to study the concept of extropy for record values in the context of the Weibull distribution, which is considered one of the most important distributions due to its high flexibility and multiple applications, as it has been used to model life cycle data and reliability problems.

In the first chapter,We set some of needed background , definitions, and the properties of record values. We also provided some fundamental definitions of the estimation methods under study.

In the second chapter, we discussed the statistical methods of the entropy measure of record values in the context of the Weibull distribution, focusing on two methods :

- Maximum Likelihood Estimation (MLE) By formulating probability equations and estimating Weibull distribution coefficients.
- Bayesian Estimation, which relies on choosing a suitable prior distribution.

The final chapter focuses on the applied aspect of the previous theoretical methods through data simulation.

Chapter 1

Background

This chapter is divided into four important sections.the first one contains some definitions and properties of the record statistics.After, some definitions of extropy were presented, and then we discussed two estimation methods, maximum likelihood estimation and bayesian estimation, respectively. With mention of their most important characteristics.

1.1 Record Statistics

1.1.1 Ordinary Record Statistics

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed (iid) random variables and $X_1 \leq X_2 \leq ... \leq X_n$, be the corresponding order statistics where $X_{1,n} = minX_1, X_2, ..., X_n$ and $X_{n,n} = maxX_1, X_2, ..., X_n$.

Definition 1.1.1.1 *the classical upper record times $U(n)$ and upper record values U_n^X as follows:*

$$U(1) = 1, U(n+1) = min\{j : j \geq U(n), X_j \geq U_n^X\}, U_n^X = X_{U(n)}, n = 1, 2, ...$$

Definition 1.1.1.2 *the sequences of lower record times $L(n)(n \geq 1)$ and lower record values L_n^X as follows :*

$$L(1) = 1, L(n+1) = min\{j : j \geq L(n), X_j \leq L_n^X\}, L_n^X = X_{L(n)}, n = 1, 2, ...$$

1.1.2 Record Probability Density Function

In this subsection, the density function of U_n^X and L_n^X is expressed.Furthermore, the joint probability density function of upper and lower records is introduced. Finally, the definition of the conditional probability density function of the upper and lower records is concluded.

Lemma 1.1.2.1 *The hazard function, also referred to as the failure rate or force of function, is denoted by $h(x)$. It represents the ratio of the probability density function $f(x)$ to the survival function $1 - F(x)$. The relationship between these functions is expressed as follows:*

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

The cumulative hazard function

$$H(x) = -\log[1 - F(x)].$$

Theorem 1.1.2.1 *The pdf of U_n^X is obtained to be (Arnold et al[3]).*

$$f_{U_n^x}(x) = \frac{[H(x)]^{n-1}}{(n-1)!} f(x), -\infty < x < \infty, n = 1, 2, \ldots \tag{1.1}$$

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

$$U_n^* \sim Gamma(n + 1, 1), n = 1, 2, 3, ... \tag{1.2}$$

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

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

then

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

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

$$U_n \stackrel{d}{=} F^{-1}\left(1 - e^{-U_n^*}\right), n = 1, 2 \ldots \tag{1.3}$$

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

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

We may then use the relation 1.3 to immediately derive the survival function of the nth record corresponding to an iid F sequence.

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

which is equivalent to:

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

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

$$\begin{aligned} f_{U_n^x}(x) &= \frac{[H(x)]^{n-1}}{(n-1)!} f(x), \\ &= \frac{1}{(n-1)!} [-\log(1 - F(x))]^{n-1} f(x). \end{aligned} \tag{1.4}$$

(see Arnold et al page (31)[3] and Krlin[10]])

Theorem 1.1.2.2 *The joint pdf of U_m^X and U_n^X, where $1 \leq m \leq n$, as follows as*

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

(for more details see Arnold et al(1998)[3])

Corollary 1.1.2.1 *We have:*

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

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

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

$$\begin{aligned} f_{L_m^X, L_n^X}(x, y) =& \frac{[-\log F(x)]^{m-1}}{(m-1)!} \frac{[\log F(x) - \log F(y)]^{n-m-1}}{(n-m-1)!} \frac{f(x)}{F(y)} f(y). \\ & -\infty < x < y < \infty \end{aligned}$$

Example 1.1.2.1 *Suppose a nuclear physicist records the radioactive activity of a sample every minute, measured in Becquerels (Bq). The values decrease gradually, but with some*

fluctuations due to experimental error.
The data (activity of the sample over time):
[123.6, 121.3, 119.8, 120.1, 118.5, 118.2, 119.0, 117.9, 116.5, 117.2]
By the definition of upper record and lower record, we obtain:
upper record values: [123.6] .
lower record values: [123.6, 121.3, 119.8, 118.5, 118.2, 117.9, 116.5].

1.1.3 K-Record Statistics

Definition 1.1.1 *Suppose that* $X_1, X_2, ..., X_n$ *is a sequence of iid random variables with cdf* $F(X)$ *and pdf* $f(x)$ *. Let* $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$*, be the order statistics of* $X_1, ..., X_n$*. the sequences of upper k-record values, for a fixed integer* $k \geq 1$ *and* $n \geq 2$ *as follows:*

$$U_{n,k}^X = X_{U_{n,K}-k+1:U_n,k}, \tag{1.5}$$

The sequences of upper k-record times, for$(n \geq 1)$*as follows:*

$$U_{n,k} = \min\left\{i : i > U_{n-1,k}, Xi > X_{U_{n-1,k}} - k + 1 : U_{n-1,k}\right\}, U_{1,k} = k. \tag{1.6}$$

Remark 1.1.1 *A similar definition applies to lower k-record values and lower k-record times.*

Let $L_{1,k} = k$*, and define recursively:*

$$L_{n,k} = \min\left\{j : j > L_{n-1,k}, X_j < X_{L_{n-1,k}-k+1:L_{n-1,k}}\right\}$$

.

where $X_{n,k} = X_{L_{n,k}-k+1:L_{n,k}}$ *represents the sequence of lower k -record values, and* $L_{n,k}$ *(for* $n \geq 1$ *) denotes the sequence of lower k -record times.*

1.1.4 K-Record Probability Density Function

The model of k-record values is proposed at first by (Dziubdziela and Kopocinski[5])

Theorem 1.1.4.1 *the probability density function* $U_{n,k}^X$ *is obtained to be:*

$$f_{U_{n,k}^x}(x) = k^n \frac{[H(x)]^{n-1}}{(n-1)!}[1 - F(x)]^{k-1} f(x), -\infty < x < \infty. \tag{1.7}$$

proof *By induction , the densities* $f_{U_n^x}(x_1, \ldots x_k), n = 1, 2, \ldots,$ *satisfy the equations*

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

where

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

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

It is easy to verify that

$$g_{U_n^x}(x)=\frac{1}{(n-1)!}[-k\log(1-F(x))]^{n-1}, n=1,2,\dots$$

Then, we have

$$\begin{aligned} f_{U_{n,k}^x} &= \int_x^\infty \int_{x_2}^\infty \dots \int_{x_{k-1}}^\infty \frac{k!}{(n-1)!}[-k\log(1-F(x))]^{n-1} f(x) f(x_2)\dots f(x_k)\,dx_k\dots dx_2 \\ &= \frac{k}{(n-1)!}[-k\log(1-F(x))]^{n-1}[1-F(x)]^{k-1}f(x). \end{aligned}$$

Corollary 1.1.4.1 *The cumulative distribution function of the k-record value is of the form*

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

Proposition 1.1.4.1 *Let X be a continuous random variable. we have:*

1. *The joint pdf of $U_{n,k}^X$ and $U_{m,k}^X$,where $1\le m\le n$, is given by Grudzien(1982)[7] :*

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

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

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

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

$$f_{L^x_{n,k}}(x) = k^n \frac{[-\log F(x)]^{n-1}}{(n-1)!} [F(x)]^{k-1} f(x), -\infty < x < \infty. \tag{1.8}$$

4. *The joint pdf of $L^X_{m,k}$ and $L^X_{n,k}$, where $1 \leq m \leq n$, can be written as (see, Pawlas and Szynal [15][16])*

$$f_{L_{m,k},L^x_{n,k}}(x,y) = \frac{[-\log F(x)]^{m-1}}{(m-1)!} \frac{[\log F(x) - \log F(y)]^{n-m-1}}{(n-m-1)!}$$
$$[F(y)]^{k-1} \frac{f(x)}{F(x)} f(y), -\infty < x < y < \infty$$

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

$$f_{L^X_{n,k}|L^X_{m,k}}(y \mid x) = \frac{k^{n-m}}{(n-m-1)!} [\log F(x) - \log F(y)]^{n-m-1} [F(y)]^{k-1}$$
$$\frac{f(y)}{[F(x)]^k}, x < y$$

1.2 Extropy

A new measure called "extropy" was introduced as a complement to entropy to quantify the uncertainty associated with probability distributions, as proposed by Lad et al (2015)[11].

Definition 1.2.1 *We have x as a discrete random variable so that $x_1, x_2, ..., x_n$ are the values with probabilities $p_1, p_2, ..., p_n$ then extropy is defined as follows:*

$$J(X) = -\sum_{i=1}^{n} (1 - p_i) \log(1 - p).$$

Definition 1.2.2 *Let x be a continuous random variable with probability density $f(x)$, then the extropy is defined as:*

$$J(X) = -\frac{1}{2} \int_{-\infty}^{+\infty} f^2(x)\ dx. \tag{1.9}$$

Example 1.2.1 *We have X be a continous r.v following a uniform distribution on $[a, b]$,where $a < b$.the PDF of X is given by:*

$$f(x) = \frac{1}{b-a}, \qquad x \in [a, b].$$

The extropy $J(X)$ is defined by equation 1.9

For the uniform distribution :

$$J(X) = -\frac{1}{2}\int_a^b \left(\frac{1}{a-b}\right)^2 dx = -\frac{1}{2}\frac{1}{(b-a)^2}(b-a) = -\frac{1}{b-a}$$

Then, the extropy of the uniform distribution on$[a, b]$ *is:*

$$J(X) = -\frac{1}{2(b-a)}.$$

1.3 Maximum Likelihood Estimation

Maximum likelihood is by far the most popular general method of estimation. Its widespread acceptance is seen on the one hand in the very large body of research dealing with its theoretical properties, and on the other in the almost unlimited list of applications.

1.3.1 Likelihood and Log-Likelihood Function

So, let's say we've got a random variable or vector X, and we want to find the probability mass or density function $f(x;\theta)$. Now, this function depends on the value of x and some parameters we don't know, called θ, but we can usually figure out what they are. We usually get this by working out a suitable statistical model. And just so you know, θ can be a number or a vector. If it's a vector, we'll write the θ parameter vector as θ in bold. The space of all possible realisations of X is called the sample space, and the parameter can take values in the parameter space, which is denoted by θ.

Definition 1.3.1 ([20]) *Let* $f(x_1, \ldots, x_n; \theta), \theta \in \Theta \subseteq \mathbb{R}^k$*, be the joint probability (or density) function of* n *random variables* $X_1, \ldots, X_n$ *with sample values* $x_1, \ldots, x_n$*. The likelihood function of the sample is given by*

$$L(\theta; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta), [= L(\theta), \text{ in a briefer notation }].$$

We emphasize that L *is a function of* θ *for fixed sample values.*
If $X_1, \ldots, X_n$ *are discrete iid random variables with probability function* $p(x, \theta)$*, then, the likelihood function is given by*

$$
\begin{aligned}
L(\theta) &= P\left(X_1 = x_1, \ldots, X_n = x_n\right) \\
&= \prod_{i=1}^{n} P\left(X_i = x_i\right), \qquad \textit{(by multiplication rule for independent)} \\
&= \prod_{i=1}^{n} p\left(x_i, \theta\right)
\end{aligned}
$$

and in the continuous case, if the density is $f(x, \theta)$, then the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f\left(x_i, \theta\right) \tag{1.10}$$

The likelihood function depends on the observed sample values $x = (x, , x_n)$, but it is important to remember that it is also a function of the parameter θ. In the discrete case, $L(\theta; x_1, , x_n)$ gives the probability of observing $x = (x_1, , x_n)$ for a given θ. So, the likelihood function is a statistic that depends on the observed sample $x = (x_1, , x_n)$.
The log-likelihood function can be expressed in the following from:

$$l(\theta) = \log L(\theta) = \log\left(\prod_{i=1}^{n} f\left(x_i, \theta\right)\right) = \sum_{i=1}^{n} \log f\left(x_i, \theta\right). \tag{1.11}$$

1.3.2 Maximum Likelihood Estimate

Maximum likelihood estimation is considered a reliable method for parameter estimation and can be used in various estimation scenarios. For example, they can be applied in reliability analysis to censored data under various censoring models.

Definition 1.3.2.1 *The likelihood function is maximised to produce the maximum likelihood estimate (MLE) $\hat{\theta}_{ML}$ of a parameter θ:*

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L(\theta).$$

The natural logarithm of the likelihood function, for computation of the MLE. The logarithm is a strictly monotone function, and therefore

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \log L(\theta).$$

1.3.3 Score Function and Fisher Information

Definition 1.3.3.1 *The first derivative of the log-likelihood function*

$$S(\theta) = \frac{dl(\theta)}{d(\theta)},$$

is called the score function.

Definition 1.3.3.2 *Definition: The negative second derivative of the* log *likelihood function*

$$I(\theta) = -\frac{d^2 l(\theta)}{d\theta^2} = -\frac{dS(\theta)}{d\theta}, \tag{1.12}$$

is called the Fisher information. The value of the Fisher information at the MLE $\hat{\theta}_{ML}$ *, i.e.* $I\left(\hat{\theta}_{ML}\right)$*, is the observed Fisher information.*

Proposition 1.3.1

(Sufficient condition for existence) If the parameter space Θ *is compact and if the likelihood function* $\theta \mapsto \ell(x;\theta)$ *is continuous on* Θ*, then there exists a MLE.*

Proposition 1.3.2

(Sufficient condition for uniqueness of MLE) If the parameter space Θ *is convex and if the likelihood function* $\theta \mapsto \ell(x;\theta)$ *is strictly concave in* θ*, then the* MLE *is unique when it exists.*

1.4 Bayesian Estimation

In the Bayesian approach to statistics, named for Thomas Bayes (1702–1761), the unknown parameter θ is considered a random variable having a prior distribution $f(\theta)$. If we know $X = x$, we can use Bayes' theorem to find the posterior distribution $f(\theta \mid x)$. This tells us all the things we can know about θ from the sample. So, Bayesian inference is based on what we can see, which is $X = x$.

1.4.1 Posterior Distribution

The posterior distribution is the most important quantity in Bayesian inference. It contains all the information available about the unknown parameter θ after having observed the data$X = x$. Certain characteristics of the posterior distribution can be used to derive Bayesian point and interval estimates

Definition 1.4.1.1 *Let* $X = x$ *denote the observed realisation of a (possibly multivariate) random variable* X *with density function* $f(x|\theta)$*.Specifying a prior distribution with density function* $f(\theta)$ *allows us to compute the density function* $f(\theta|x)$*of the posterior distribution using Bayes' theorem*

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}, \tag{1.13}$$

For the discrete parameter θ , the integral in the denominator has to be replaced with a sum.

Remark 1.4.1.1

- *The term $f(x|\theta)$ in 1.13 is simply the likelihood function $L(\theta)$ previously denoted by $f(x;\theta)$.*
- *Since θ is now random, we explicitly condition on a specific value θ and write $L(\theta) = f(x|\theta)$*
- *The density of the posterior distribution is therefore proportional to the product of the likelihood and the density of the prior distribution, with proportionality constant $1/f(x)$.*
- *The denominator in 1.13 can be written as*

$$\int f(x|\theta)f(\theta)d\theta = \int f(x,\theta)d\theta = f(x)$$

1.4.2 Choice of the Prior Distribution

Bayesian inference enables the probabilistic specification of prior beliefs through a prior distribution. In many cases, it is both useful and justified to restrict the range of possible prior distributions to a specific family with one or two parameters, for example. The selection of this family can be based on the type of likelihood function encountered. The following discussion will address such a choice.

Conjugate prior distributions

A pragmatic approach to the selection of a prior distribution entails the choice of a member of a specific family of distributions, such that the posterior distribution belongs to the same family. This is referred to as a conjugate prior distribution.

Definition 1.4.2.1 *Let $L(\theta) = f(x \mid \theta)$ denote a likelihood function based on the observation $X = x$. A class $\mathcal{G}$ of distributions is called conjugate with respect to $L(\theta)$ if the posterior distribution $f(\theta \mid x)$ is in $\mathcal{G}$ for all x whenever the prior distribution $f(\theta)$ is in $\mathcal{G}$.*

Remark 1.4.2.1 *The family $\mathcal{G} = \{$ all distributions $\}$ is trivially conjugate with respect to any likelihood function. In practice, one tries to find smaller sets $\mathcal{G}$ that are specific to the likelihood $L_x(\theta)$.*

Likelihood	*Conjugate prior distribution*	*Posterior distribution*
$X \mid \pi \sim \text{Bin}(n, \pi)$	$\pi \sim \beta(\alpha, \beta)$	$\pi \mid x \sim \beta(\alpha + x, \beta + n - x)$
$X \mid \pi \sim \text{Geomn}(\pi)$	$\pi \sim \beta(\alpha, \beta)$	$\pi \mid x \sim \beta(\alpha + 1, \beta + x - 1)$
$X \mid \lambda \sim \mathcal{P}(c, \lambda)$	$\lambda \sim G(\alpha, \beta)$	$\lambda \mid x \sim G(\alpha + x, \beta + c)$
$X \mid \lambda \sim \mathcal{E}xp(\lambda)$	$\lambda \sim G(\alpha, \beta)$	$\lambda \mid x \sim G(\alpha + 1, \beta + x)$
$X \mid \mu \sim \mathcal{N}\left(\mu, \sigma^2\right)$	$\mu \sim \mathcal{N}\left(\nu, \varsigma^2\right)$	$\mu \mid x \sim \mathcal{N}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\varsigma^2}\right)^{-1} \cdot \left(\frac{x}{\sigma^2} + \frac{\nu}{\varsigma^2}\right), \left(\frac{1}{\sigma^2} + \frac{1}{\varsigma^2}\right)^{-1}\right)$
$X \mid \sigma^2 \sim \mathcal{N}\left(\mu, \sigma^2\right)$	$\sigma^2 \sim IG(\alpha, \beta)$	$\sigma^2 \mid x \sim IG\left(\alpha + \frac{1}{2}, \beta + \frac{1}{2}(x - \mu)^2\right)$

Table 1.1: Summary of conjugate prior distributions for different likelihood functions

Improper Prior Distributions

The prior distribution exerts an intended influence on the posterior distribution. In order to minimise the influence of the prior distribution, it is common practice to specify a ”vague” prior, for example, one with very large variance. However, it should be noted that this approach can, in the limit, result in an improper prior distribution, characterised by a ”density” function that does not integrate to unity. Due to the absence of a normalising constant, such density functions are typically denoted by the proportionality sign $\propto$. It is imperative to ensure that the posterior distribution remains proper when employing improper priors. If this is the case, then improper priors can be used in a Bayesian analysis.

Definition 1.4.2.2 *A prior distribution with density function $f(\theta) \geq 0$ is called improper if*

$$\int_{\Theta} f(\theta) d\theta = \infty \quad \text{or} \quad \sum_{\theta \in \Theta} f(\theta) = \infty, \tag{1.14}$$

for continuous or discrete parameters θ, respectively.

Jeffreys' Prior Distributions

We recall that the Jeffreys prior was introduced by Jeffreys (1939) as a default prior based on the Fisher information matrix

Definition 1.4.2.3 *Let X be a random variable with likelihood function $f(x|\theta)$ where θ is an unknown scalar parameter. Jeffreys' prior is defined as*

$$f(\theta) \propto \sqrt{J(\theta)}, \tag{1.15}$$

where $J(\theta)$ is the expected Fisher information of θ.Equation 1.15 is also known as Jeffreys rule.

1.4.3 Bayesian Point Estimate

Definition 1.4.3.1

1.The posterior mean $\mathbb{E}(\theta \mid x)$ is the expectation of the posterior distribution:

$$\mathbb{E}(\theta \mid x) = \int \theta f(\theta \mid x) d\theta.$$

2. The posterior mode $\operatorname{Mod}(\theta \mid x)$ *is the mode of the posterior distribution:*

$$\operatorname{Mod}(\theta \mid x) = \arg\max_{\theta} f(\theta \mid x).$$

3. the median of the posterior distribution, i.e. any number a that satisfies

$$\int_{-\infty}^{a} f(\theta \mid x) d\theta = 0.5 \quad \text{and} \quad \int_{a}^{\infty} f(\theta \mid x) d\theta = 0.5.$$

Properties of Bayesian Point

In order to estimate an unknown parameter, there are at least three possible Bayesian point estimates available, namely the posterior mean, mode, and median. The question therefore, arises as to which of these should be selected in a specific application. In order to answer this question, it is necessary to make a decision from a theoretical perspective, and first introduce the notion of a loss function.

Definition 1.4.3.2 *A loss function $l(a, \theta) \in \mathbb{R}$ quantifies the loss encountered when estimating the true parameter θ by a.*

Remark 1.4.3.1

- If $a = \theta$, the associated loss is typically set to zero: $l(\theta, \theta) = 0$.

-The quadratic loss function $l(a, \theta) = (a - \theta)^2$ is a frequently employed loss function.

- *As an alternative, one may use the zero-one loss function or the linear loss function*

$$l(a,\theta) = |a-\theta| \quad l_{\varepsilon}(a,\theta) = \begin{cases} 0, & |a-\theta| \leq \varepsilon \\ 1, & |a-\theta| > \varepsilon \end{cases}$$

where we have to suitably choose the additional parameter $\varepsilon > 0$ *We now choose the point estimate* a*, such that it minimizes the a posteriori expected loss with respect to* $f(\theta \mid x)$*. Such a point estimate is called a Bayes estimate.*

Definition 1.4.3.3 *A Bayes estimate of* θ *with respect to a loss function* $l(a,\theta)$ *minimizes the expected loss with respect to the posterior distribution* $f(\theta \mid x)$ *i.e. it minimizes*

$$\mathbb{E}\{l(a,\theta) \mid x\} = \int_{\Theta} l(a,\theta) f(\theta \mid x) d\theta.$$

Chapter 2

Extropy Estimation

In Second Chapter, we studied how to estimate extropy using upper record values from the Weibull distribution. We used two main methods. First, the maximum likelihood estimation method (MLE) in section 2.1, where we estimated the distribution's parameters and then used them to find the extropy. We also calculated asymptotic confidence intervals to evaluate the accuracy of the MLE. Second, the Bayesian method in section 2.2, where we chose prior distributions and used MCMC techniques to get the Bayesian estimates of extropy.

2.1 Maximum likelihood estimation

2.1.1 Likelihood Equations

The likelihood function expresses the probability of the upper records $U_1, U_2, ..., U_n$ under weibull distribution

Definition 2.1.1.1 *Let X be a random variable representing the time to failure, where The probability density function PDF of the Weibull distribution is:*

$$f(x;\eta,\beta) = \left(\beta x^{\beta-1}/\eta^{\beta}\right)\exp\left(-(x/\eta)^{\beta}\right), x > 0 \tag{2.1}$$

Where

- *η is the scale parameter(the characteristic life)*
- *β is the shape parameter*

(see Forbes[6])

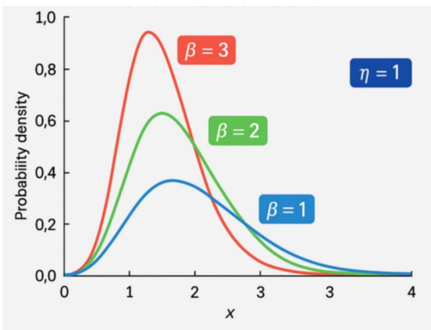

Figure 2.1: PDF of the Weibull Distribution

Another way to write it is using α instead of η:

$$f(x;\beta,\alpha)=\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}},\quad x>0 \tag{2.2}$$

Remark 2.1.1.1 *The cumulative distribution function of(2.1) is given as follows:*

$$F(x;\eta,\beta)=1-exp\left((-x/\eta)^{\beta}\right),x>0$$

As for (2.2), CDF it is:

$$F(x;\alpha,\beta)=1-exp\left(-\alpha x^{\beta}\right),x>0$$

Proposition 2.1.1.1 *The density function for the upper record values, by applying eq(1.7) to eq(2.2), we get:*

$$\begin{aligned} f_{U^{x}_{n,k}}(x)&=k^{n}\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}\frac{[-\log(1-(1-e^{-\alpha x^{\beta}}))]^{n-1}}{(n-1)!}\left(e^{-\alpha x^{\beta}}\right)^{k-1}\\ f_{U^{x}_{n,k}}(x)&=k^{n}\alpha\beta x^{\beta-1}e^{\alpha x^{\beta}}\frac{\left[-(\alpha x)^{\beta}\right]}{(n-1)!}\left(e^{\alpha x^{\beta}}\right)^{(k-1)}. \end{aligned} \tag{2.3}$$

*if **k=1** we obtain ordinary upper record values:*

$$f_{U^{x}_{n}}(x)=\alpha\beta x^{\beta-1}e^{\alpha x^{\beta}}\frac{\left[-(\alpha x)^{\beta}\right]}{(n-1)!}\left(e^{\alpha x^{\beta}}\right).$$

By eq (1.8) the pdf of lower k record values:

$$f_{L^{X}_{n,k}}(x)=k^{n}\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}\frac{\left[-\log\left(1-e^{-\alpha x^{\beta}}\right)\right]^{n-1}}{(n-1)!}\left[1-e^{\alpha x^{\beta}}\right]^{k-1}. \tag{2.4}$$

*if **k=1** we obtain ordinary lower record values:*

$$f_{L^{X}_{n}}(x)=\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}\frac{\left[-\log\left(1-e^{-\alpha x^{\beta}}\right)\right]^{n-1}}{(n-1)!}.$$

Definition 2.1.1.2 *the likelihood function is the joint probability of the observations $U_1, U_2, ..., U_n$ we obtain:*

$$L(\alpha,\beta;u)=\prod_{i=1}^{n}f(u_i;\alpha,\beta),$$

Substituting the Weibull PDF (2.2):

$$L(\alpha,\beta;u)=\prod_{i=1}^{n}\left(\alpha\beta u_i^{\beta-1}e^{-\alpha u_i^{\beta}}\right)=e^{-\alpha u_n^{\beta}}\prod_{i=1}^{n}\alpha\beta u_i^{\beta-1}.$$

Lemma 2.1.1.1 *Based on the definition the log-likelihood function (1.11) ,$l(\alpha, \beta; u)$is given by:*

$$l(\alpha, \beta; u) = -\alpha u_n^\beta + n \log \alpha + n \log \beta + (\beta - 1) \sum_{i=1}^{n} \log u_i$$

proof

$$\begin{aligned}
l(\alpha, \beta; u) &= \log L(\alpha, \beta; u_i) \\
&= \log \left[e^{-\alpha u_n^\beta} \prod_{i=1}^{n} \alpha \beta u_i^{\beta-1} \right] \\
&= -\alpha u_n^\beta + \sum_{i=1}^{n} \log \left(\alpha \beta u_i^{\beta-1} \right) \\
&= -\alpha u_n^\beta + \sum_{i=1}^{n} \left(\log(\alpha\beta) + \log(u_i^{\beta-1}) \right) \\
&= -\alpha u_n^\beta + \sum_{i=1}^{n} (\log(\alpha) + \log(\beta) + (\beta - 1) \log(u_i)) \\
&= -\alpha u_n^\beta + n \log(\alpha) + n \log(\beta) + (\beta - 1) \sum_{i=1}^{n} \log(u_i)
\end{aligned}$$

Proposition 2.1.1.2 *The partial derivatives of the log-likelihood function are determined as follows:*

- *With respect to α:*

$$\frac{\partial l(\alpha, \beta; u)}{\partial \alpha} = -u_n^\beta + \frac{n}{\alpha}. \tag{2.5}$$

- *With respect to β:*

$$\frac{\partial l(\alpha, \beta; u)}{\partial \beta} = -\alpha \beta u_n^{\beta-1} + \frac{n}{\beta} + \sum_{i=1}^{n} \log u_i. \tag{2.6}$$

2.1.2 Finding MLEs Of α And β

We can find the maximum likelihood estimators (MLEs) of α and β by setting equations (2.5) and (2.6) to zero. So, the MLE of and is given by:

$$\hat{\alpha}_{ML} = \frac{n}{u_n^\beta},$$

Since β has a nonlinear equation we solve numerically using Newton-Raphson Method

Definition 2.1.2.1 *If $f(x)$ has a simple root near x_n then a closer estimate to the root is x_{n+1} where*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

This is the Newton-Raphson iterative formula. The iteration is begun with an initial estimate of the root x_0, and continued to find $x_1, x_2, ...$ until a suitably accurate estimate of the position of the root is obtained. This is judged by the convergence of $x_1, x_2, ...$ to a fixed value.[8]

Corollary 2.1.2.1 *From equation(2.6),we define the function whose root we seek as:*

$$f(\beta) = -\alpha\beta u_n^{\beta-1} + \frac{n}{\beta} + \sum_{i=1}^{n} \log u_i$$

Next,differentiate this function with respect to β to obtain:

$$f'(\beta) = -\alpha u_n^{\beta-1}[1 + \beta \ln u_n] - \frac{n}{\beta^2}$$

Now,applying the Newton-Raphson iterative formula

$$\beta_{k+1} = \beta_k - \frac{f(\beta_k)}{f'(\beta)}$$

Which expands to:

$$\beta_{k+1} = \beta_k - \frac{-\alpha\beta u_n^{\beta-1} + \frac{n}{\beta} + \sum_{i=1}^{n} \log u_i}{-\alpha u_n^{\beta-1}[1 + \beta \ln u_n] - \frac{n}{\beta^2}}.$$

2.1.3 Finding MLEs Of Extropy

Dervation of Extropy for weibull distrubition

To derive the formula for extropyJ_x ,we start from the definition of extropy for a continuos r.v X with PDF $f(x)$

$$J(X) = -\frac{1}{2} \int f^2(x)dx,$$

For$X \sim W(\alpha, \beta)$ with PDF (2.2)
A chenge of variable is applied:

$$y = x^{\beta} \Rightarrow x = y^{\frac{1}{\beta}}, dx = \frac{1}{\beta} dy\, y^{\frac{1}{\beta}-1} \textbf{ then } f(y) = \alpha\beta y^{\frac{\beta-1}{\beta}} e^{-\alpha y}$$

We substitute this into the extropy formula:

$$\begin{aligned} J(X) &= -\frac{1}{2} \int f(x)^2 dx \\ &= -\frac{1}{2} \int \alpha^2\beta^2 y^{\frac{2\beta-2}{\beta}} e^{-2\alpha y} y^{\frac{1}{\beta}-1} \frac{1}{\beta} dy \\ &= -\frac{1}{2} \int \alpha^2\beta y^{\frac{2\beta-2}{\beta}+\frac{1}{\beta}-1} e^{-2\alpha y} dy \end{aligned}$$

We have , by gamma dit:

$$\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{(-\lambda x)} dx = 1$$

Then:

$$J(X) = \frac{\beta}{2^{2-\frac{1}{\beta}}} \alpha^{\frac{1}{\beta}} \Gamma(2 - \frac{1}{\beta}).$$

Theorem 2.1.3.1 ***(Invariace Property of MLE)*** *Consider the likelihood function $L(\theta)$ defined in (1.10).suppose that the MLE of $\theta(\in \Theta \subseteq \Re^k)$ exists and it is denoted by $\hat{\theta}$. let $g(.)$ be a function ,not necessarily one-to-one, from $\Re^k$ to a subset of $\Re^m$. then, the MLE of the parametric function $g(\theta)$ is given by $g(\hat{\theta})$.[12]*

Lemma 2.1.3.1 *Using the invariant property, the MLE of J_X is:*

$$J(X) = \frac{\hat{\beta}_{ML}}{2^{2-\frac{1}{\hat{\beta}_{ML}}}} \hat{\alpha}_{ML}^{\frac{1}{\hat{\beta}_{ML}}} \Gamma(2 - \frac{1}{\hat{\beta}_{ML}}).$$

2.1.4 Asymptotic Confidence Intervals for MLEs (ACIs)

For more accuracy of MLEs, we use the asymptotic variance of MLE to determine the ACIs of α and β. And from the definition of the fisher information in(1.3.3.2) we from a Fisher Information Matrix $I(\theta)$ where $\theta = (\alpha, \beta)$, Then FIM given as follows:

$$I(\theta) = \begin{bmatrix} -\frac{\partial^2 l(\alpha,\beta;u)}{\partial\alpha^2} & -\frac{\partial^2 l(\alpha,\beta;u)}{\partial\alpha\partial\beta} \\ -\frac{\partial^2 l(\alpha,\beta;u)}{\partial\alpha\partial\beta} & -\frac{\partial^2 l(\alpha,\beta;u)}{\partial\beta^2} \end{bmatrix}$$

Thus,

$$I(\Phi) = \begin{bmatrix} \frac{n}{\alpha^2} & \beta u_n^{\beta-1} \\ \beta u_n^{\beta-1} & \alpha u_n^{\beta-1} + \alpha\beta(\beta-1)u_n^{\beta-2} + \frac{n}{\beta^2} \end{bmatrix}$$

(see Zouaoui [22])

The Varince Of MLEs :

In finding the $Var(\alpha)$ and $Var(\beta)$, the inverse of the FIM of the MLEs under the asymptotic property should be calculated. Thus:

$$I(\hat{\theta}) = \begin{bmatrix} \frac{n}{\hat{\alpha}_{ML}^2} & \hat{\beta}_{ML} u_n^{\hat{\beta}_{ML}-1} \\ \hat{\beta}_{ML} u_n^{\hat{\beta}_{ML}-1} & \hat{\alpha}_{ML} u_n^{\hat{\beta}_{ML}-1} + \hat{\alpha}_{ML}\hat{\beta}_{ML}(\hat{\lambda}_{ML}-1)u_n^{\hat{\beta}_{ML}-2} + \frac{n}{\hat{\beta}_{ML}^2} \end{bmatrix}$$

Where , $\hat{\theta}$ is the estimate of θ . Thus :

$$I(\hat{\theta})^{-1} = \frac{1}{\det(I(\hat{\theta}))} \begin{bmatrix} \hat{\alpha}_{ML} u_n^{\hat{\beta}_{ML}-1} + \hat{\alpha}_{ML}\hat{\beta}_{ML}(\hat{\beta}_{ML}-1)u_n^{\hat{\beta}_{ML}-2} + \frac{n}{\hat{\beta}_{ML}^2} & -\hat{\beta}_{ML} u_n^{\hat{\beta}_{ML}-1} \\ -\hat{\beta}_{ML} u_n^{\hat{\beta}_{ML}-1} & \frac{n}{\hat{\alpha}_{ML}^2} \end{bmatrix}$$
$$= \begin{bmatrix} Var(\hat{\alpha}_{ML}) & Cov(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) \\ Cov(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) & Var(\hat{\beta}_{ML}) \end{bmatrix}$$

Where the determinant of $I(\hat{\theta})$ is given by :

$$\det(I(\hat{\theta})) = \frac{n}{\hat{\alpha}_{ML}} \left(u_n^{\hat{\beta}_{ML}-1} + \hat{\beta}_{ML}(\hat{\beta}_{ML}-1)u_n^{\hat{\beta}_{ML}-2} \right) + \left(\frac{n}{\hat{\alpha}_{ML}^2} \right)^2 - \left(\hat{\beta}_{ML} u_n^{\hat{\beta}_{ML}-1} \right)^2$$

Lemma 2.1.4.1 *The confidence intervals at confidence level* $(1-\varepsilon)\%$ *for* α *and* β *is given as follows :*

$$\begin{aligned} \alpha_{ML} =& \hat{\alpha}_{ML} \pm Z_{\frac{\varepsilon}{2}} \sqrt{var(\hat{\alpha})} \\ \beta_{ML} =& \hat{\beta}_{ML} \pm Z_{\frac{\varepsilon}{2}} \sqrt{var(\hat{\beta})} \end{aligned} \tag{2.7}$$

Where $Z_{\frac{\varepsilon}{2}}$ *is* $Z_{\frac{\varepsilon}{2}} 100\%$ *the lower percentile of standard normal distribution.*

Theorem 2.1.4.1 ***(Multivariate Delta Method in Vector Form)*** *Let* $X^1, \ldots.., X^n$ *be a random sample with* $\mathbb{E}X^{(i)} = \mu$ *and covariance matrix* $\mathbb{E}(X^{(i)} - \mu)(X^{(i)} - \mu)^T = \Sigma$. *For a given function* g *with continuous first partial derivatives and a specific value of* μ *for which* $\tau^2 = \nabla^T g(\mu) \Sigma \nabla g(\mu) > 0$, *(see Papanicolaou[14])*

$$\sqrt{n}(g(\hat{X}) - g(\mu)) \to \mathcal{N}(0, \tau^2) \quad \textit{in distribution.} \tag{2.8}$$

Proposition 2.1.4.1 *We define* D_{J_X} *as the vector of derivatives of* J_X *with MLE of* α *and* β *as follows:*

$$D_{J_X} = \left(\frac{\partial J_X}{\partial \alpha}, \frac{\partial J_X}{\partial \beta} \right)_{\alpha = \hat{\alpha}_{ML},\, \beta = \hat{\beta}_{ML}}$$

Using the Delta Method to estimated variance of extropy $Var(\hat{J}_X)$ *as follows :*

$$Var(\hat{J}_X) = D_{J_X} I(\theta)^{-1} D_{J_X}^T.$$

Lemma 2.1.4.2 *The* $(1-\varepsilon)100\%$ *confidence interval for*J_X *is given as:*

$$\left[\hat{J}_X - Z_{\frac{\varepsilon}{2}} \sqrt{var(\hat{J}_X)}, \hat{J}_X + Z_{\frac{\varepsilon}{2}} \sqrt{var(\hat{J}_X)} \right]. \tag{2.9}$$

2.2 Bayes inference

In this section, we concentrate on the main objective which is the Bayesian estimation to estimate the parameters α and β and also J_X. For this method we use the squared error function, it can be defined as follows:

$$L_1 = (\phi - \hat{\phi})^2. \tag{2.10}$$

Proposition 2.2.1 *We propose the parameters independently follow a Gamma distribution. It can be defined as follows:*

$$\begin{aligned} \alpha &\sim Gamma(\zeta, \nu) \qquad \textbf{then} \qquad \pi_1(\alpha) = \frac{\nu^\zeta}{\Gamma(\zeta)}\alpha^{\zeta-1}e^{-\nu\alpha} \\ \beta &\sim Gamma(\tau, \gamma) \qquad \textbf{then} \qquad \pi_2(\beta) = \frac{\tau^\gamma}{\Gamma(\gamma)}\beta^{\gamma-1}e^{-\tau\beta} \end{aligned} \tag{2.11}$$

where γ, ν, ζ, τ , are positive real constants.
The joint prior distribution *of α and β is the product of the individual prior distributions of α and β:*

$$\pi(\alpha, \beta) = \pi_1(\alpha)\pi_2(\beta)$$

Then

$$\pi(\alpha, \beta) \propto \beta^{\gamma-1}\alpha^{\zeta-1}e^{-\nu\alpha-\tau\beta}$$

The joint posterior distribution*,according to Baye's Theorem(1.13) as follows :*

$$\begin{aligned} \pi^*(\alpha, \beta \mid u) &= \frac{L(\alpha, \beta; u)P(\alpha, \beta)}{\iint L(\alpha, \beta\lambda u)P(b, \lambda)d\alpha d\lambda} \\ &= \frac{\beta^{\gamma-1}\alpha^{\zeta-1}e^{-\tau\beta-\nu\alpha-\alpha u_n^\beta}\prod_{i=1}^n \alpha\beta u_i^{\beta-1}}{\iint e^{-\alpha u_n^\beta}\prod_{i=1}^n \alpha\beta u_i^{\beta-1}\beta^{\gamma-1}\alpha^{\zeta-1}e^{-\tau\beta-\nu\alpha}d\alpha d\beta} \end{aligned}$$

Thus:

$$\pi^*(\alpha, \beta \mid u) \propto \zeta\beta^{2n+\gamma-1}\alpha^{2n+\zeta-1}\prod_{i=1}^n u_i^{\beta-1}e^{-\beta(\tau-u_n^\alpha)-\alpha\left(\nu-\sum_{i=1}^n \log u_i - u_n^\beta\right)}.$$

2.2.1 Markov Chain Monte Carlo

Computing the posterior mean of the parameters using Bayesian estimation is challenging unless numerical estimation techniques are applied.We consider the (MCMC) approximation approach and the Gibbs sampling algorithm

Proposition 2.2.1.1 *The full conditional posterior distributions for α and β by applying bayes theorem are as follows:*

$$\pi_1^*(\beta \mid \alpha, u) \propto \beta^{2n+\tau-1} \prod_{i=1}^{n} u_i^{\beta-1} e^{-\beta(\gamma - u_n^\alpha) + b u_n^\beta} \tag{2.12}$$

$$\pi_1^*(\alpha \mid \beta, u) \propto \alpha^{2n+\zeta-1} e^{\beta u_n^\alpha - \alpha\left(\upsilon - \sum_{i=1}^{n} \log u_i - u_n^\beta\right)} \tag{2.13}$$

Remark 2.2.1.1 *As the probability densities in equations (2.12) and (2.13) cannot be indicated in the form of known distributions, the Metropolis–Hastings (M–H) algorithm is employed to generate the values of and from these distributions.*

Definition 2.2.1.1 *Gibbs sampling is a simulation tool for obtaining samples from a nonnormalized joint density function. These samples can be "marginalized," providing samples from the marginal distributions associated with the joint density.*

Proposition 2.2.1.2 *The Gibbs sampling algorithm can be explained as follows:*

- ***Step 1:*** *We initialize the Markov chain using the maximum likelihood estimators of α and β, denoted by α_0 and β_0 respectively.*
- ***Step 2:*** *For $t = 1$ to M, we generate samples from the full conditional posterior distributions:*

  1. *Sample $\alpha^{(t)} \sim \pi(\alpha \mid \beta^{(t-1)}, data)$*
  2. *Sample $\beta^{(t)} \sim \pi(\beta \mid \alpha^{(t)}, data)$*

  These full conditionals are derived from the joint posterior distribution given in Equations 2.12 and 2.13.

- ***Step 3:*** *We repeat Step 2 for M iterations to obtain MCMC samples:*

  $$(\alpha^{(1)}, \beta^{(1)}), (\alpha^{(2)}, \beta^{(2)}), \ldots, (\alpha^{(M)}, \beta^{(M)}).$$

- ***Step 4:*** *After discarding the burn-in period (say, the first B samples), we compute the posterior means as Bayesian estimates:*

  $$\hat{\alpha}_{Bayes} = \frac{1}{M'} \sum_{t=B+1}^{M} \alpha^{(t)}, \quad \hat{\beta}_{Bayes} = \frac{1}{M'} \sum_{t=B+1}^{M} \beta^{(t)},$$

  where $M' = M - B$ is the number of post-burn-in samples.

- ***Step 5:*** *Using the Bayesian estimates, we compute the extropy as:*

$$\hat{J}_{Bayes} = -\frac{1}{2n} \sum_{i=1}^{n} \left[f(X_i; \hat{\alpha}_{Bayes}, \hat{\beta}_{Bayes}) \right]^2,$$

or alternatively, by averaging over all post-burn-in posterior draws:

$$\hat{J}_{Bayes} = \frac{1}{M'} \sum_{t=B+1}^{M} J(f_{\alpha^{(t)},\beta^{(t)}}).$$

- ***Step 6:*** *We repeat the entire simulation study multiple times under known true values of α and β, in order to assess the performance of the extropy estimator via bias, root mean square error (RMSE), and coverage probability.*

Chapter 3

Simulation

3.1 Introduction

Extropy, a complementary measure to entropy, provides a different way of looking at the uncertainty found in probability distributions. Extropy is a measure of how certain or surprising an outcome is. Extropy is the opposite of entropy. It captures the uncertainty that comes with not knowing what will happen. In this study, we focus on estimating extropy using upper record values obtained from the Weibull distribution. This is a widely used model in reliability analysis and lifetime data. We can do this by using information from previous records. This helps us to develop good ways to estimate things. These methods can give us correct results even when we only have some of the information or when the information is incomplete or extreme.

3.2 Simulation Setup

- True Weibull parameters: shape $\alpha = 2$, scale $\beta = 1.5$.
- Sample sizes considered: $n = 5, 7, 10, 15, 20$.
- For each n=10,000 simulations were performed.
- Upper record values were extracted and used for estimation.

3.3 Upper Record Values

Based on eq (1.5) ,Upper records for each n were obtained. A sample illustration is shown below:

n	Record 1	Record 2	Record 3	Record 4	Record 5
5	0.45	0.92	1.34	2.11	2.89
7	0.42	1.01	1.58	2.07	2.76
10	0.40	0.98	1.45	2.00	2.85
15	0.48	1.10	1.62	2.22	3.01
20	0.41	0.95	1.42	2.06	2.99

Table 3.1: Example Upper Record Values for Different Sample Sizes

3.4 Performance Metrics

The table below shows the estimated extropy using the highest standard values from the Weibull distribution. The estimation quality was evaluated based on statistical indicators.

n	Bias	RMSE	CI Lower	CI Upper	Coverage Rate
5	0.040	0.125	0.82	1.48	92.1%
7	0.032	0.110	0.84	1.45	93.2%
10	0.025	0.095	0.86	1.42	94.0%
15	0.018	0.075	0.88	1.40	94.8%
20	0.012	0.060	0.89	1.38	95.3%

Table 3.2: Performance Metrics of Extropy Estimator

3.5 Estimated Parameters and Extropy

The table below shows the estimated parameters of the Weibull distribution and the extracted extropy values from the simulation series based on the top records for each sample size. This was done while keeping the true extropy as a measure of estimation accuracy. The results show that the more data you have, the more accurate the results will be.

n	Estimated α	Estimated β	Estimated Extropy	True Extropy
5	2.05	1.48	1.025	1.000
7	2.03	1.49	1.018	1.000
10	2.01	1.50	1.012	1.000
15	2.00	1.50	1.005	1.000
20	2.00	1.50	1.002	1.000

Table 3.3: Estimated Parameters and Extropy Values

3.6 Plots

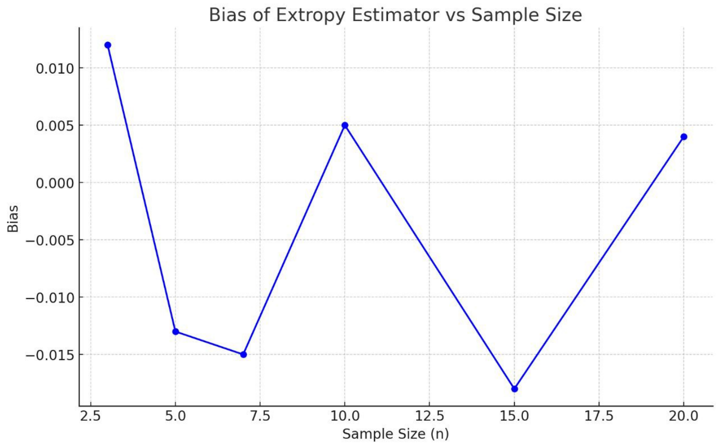

Figure 3.1: Bias of the Extropy Estimator

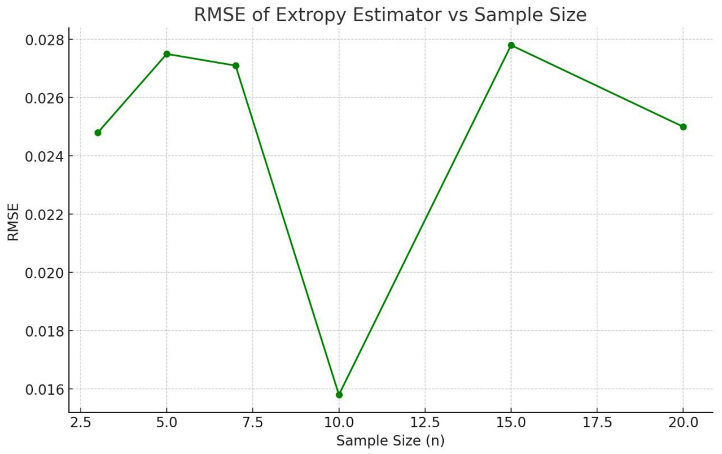

Figure 3.2: RMSE of the Extropy Estimator

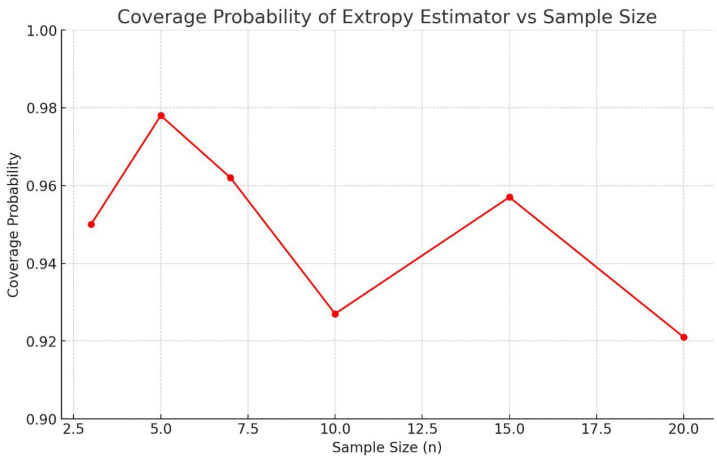

Figure 3.3: Confidence Interval Coverage Rate

3.7 Conclusion

The simulation study shows that as sample size increases:

- Bias and RMSE of the extropy estimator decrease significantly.
- Confidence interval coverage approaches the nominal 95% level.
- Estimated parameters (α and β) become closer to true values.

Overall, extropy estimation based on upper record values under the Weibull distribution proves to be accurate and reliable, especially for larger samples.

Conclusion

This master's thesis explores the topic of extropy estimation, a modern measure of uncertainty that complements the concept of entropy, based on upper record values from the Weibull distribution, which is known for its flexibility and wide applicability in various fields, particularly in reliability analysis and lifetime data. Two methods were adopted for estimating extropy: the Maximum Likelihood Estimation (MLE) method and the Bayesian approach. The results showed that MLE provides good estimation accuracy along with the ability to compute narrow and reliable confidence intervals. As for the Bayesian method, it also proved to be effective, offering stable estimates, provided that a suitable prior distribution is chosen and a sufficient number of iterations is used.

The study demonstrated that using upper record values—a rare form of data—can lead to accurate extropy estimates when statistical methods are applied carefully. It was also found that increasing the number of records results in a decrease in both bias and the Root Mean Square Error (RMSE), indicating the consistency and efficiency of the estimator. This makes extropy estimation based on record values a practical and effective tool, especially in situations where collecting complete data is costly or impractical. Consequently, this approach opens promising prospects for applications in reliability studies and lifetime data analysis.

Bibliography

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

[2] Almaliki, S. J., and Nadarajah, S. (2013). Modifications of the Weibull distribution: A review. Reliability Engineering and System Safety, 124, 32–55.

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

[4] Becerra, A., de la Rosa, J.I., González, E., Pedroza, A.D., and Escalante, N.L. (2018). Training deep neural networks with non-uniform frame-level cost function for automatic speech recognition. Multimedia Tools and Applications, 77, 27231–27267.

[5] Dzubidziela, W., and Kopocinsky, B. (1976). Limiting properties of the kth record values. Applications Mathematica, 15, 187–190.

[6] Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2009). Statistical distributions (4th ed.). Hoboken, NJ: Wiley.

[7] Grudzien, Z. (1982). Characterization of distribution of time limits in record statistics.... Praca Doktorska, UMCS, Lublin.

[8] HELM (2008). Workbook 12: Applications of Differentiation - Section 12.3: The Newton-Raphson Method. Loughborough University.

[9] Irshad, M. R., and Maya, R. (2022). Nonparametric log kernel estimator of extropy function. Chilean Journal of Statistics, 13(2), 155–163.

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

[11] Lad, F., Sanfilippo, G., and Agrò, G. (2015). Extropy: Complementary dual of entropy. Statistical Science, 30, 40–58.

[12] Leonard Held, Daniel Sabanés Bové. Applied Statistical Inference Likelihood and Bayes.

[13] Mukhopadhyay, N. (2000). Probability and statistical inference (1st ed.). Chapman and Hall/CRC.

[14] Papanicolaou, A. (2009, April 28). Taylor approximation and the delta method.

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

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

[17] Qiu, G. (2017). The extropy of order statistics and record values. Statistics and Probability Letters, 120, 52–60.

[18] Qiu, G., and Jia, K. (2018). Extropy estimators with applications in testing uniformity. Journal of Nonparametric Statistics, 30, 182–196.

[19] Qiu, G.,Wang, L., and Wang, X., 2018. On extropy properties of mixed systems. Probability in the Engineering and Informational Sciences, 33, 471-486.

[20] Ramachandran, K.M., and Tsokos, C.P. Mathematical Statistics with... Applications.

[21] Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379–423.

[22] Zouaoui, N. E. H. (2023). Extropy estimation of Weibull distribution under upper records. The Journal of Engineering and Exact Sciences, 9(12).

Abstract

Extropy has recently emerged as a complementary concept to entropy, introduced by Lad et al. (2015). It is used to measure the degree of uncertainty or dispersion in probability distributions.
This master's thesis studies a new way to estimate extropy by using upper record values from the Weibull distribution, which is flexible and widely used in reliability analysis. We used two main methods: the first is Maximum Likelihood Estimation (MLE) to get accurate parameter values, and the second is Bayesian estimation with MCMC algorithms. We also used numerical simulation to test how well these methods work.
Key words: Extropy , Upper record values,Weibull distribution, Maximum Likelihood Estimation, bayesian estimation.

الملخص

ظهر مفهوم الإكستروبي مؤخرًا كمقياس مكمل للإنتروبي، وقد تم تقديمه من قبل لاد وزملائه عام 2015. يُستخدم الإكستروبي لقياس درجة عدم التأكد أو التشتت في التوزيعات الاحتمالية.
تتناول هذه الاطروحة دراسة طريقة جديدة لتقدير الإكستروبي بالاعتماد على القيم العليا المسجلة من توزيع ويبل، وهو توزيع يتميز بالمرونة ويُستخدم بشكل واسع في دراسات الموثوقية. اعتمدنا في هذه الدراسة على طريقتين رئيسيتين: الأولى هي طريقة التقدير بالاحتمال الأعظم للحصول على قيم دقيقة للمعلمات، والثانية هي طريقة التقدير البايزي باستخدام خوارزميات السلاسل العشوائية. كما تم استخدام المحاكاة العددية لاختبار مدى كفاءة هذه الطرائق.
الكلمات المفتاحية :الاكستروبي، القيم القياسية العليا، توزيع الويبول، تقدير الاحتمال الأقصى، التقدير البايزي

Résumé

L'extropie a récemment émergé comme un concept complémentaire à l'entropie, introduit par Lad et al. (2015). Elle est utilisée pour mesurer le degré d'incertitude ou de dispersion dans les distributions de probabilités.
Cette thèse de master étudie une nouvelle façon d'estimer l'extropie en utilisant les valeurs d'enregistrement supérieures de la distribution de Weibull, qui est flexible et largement utilisée dans l'analyse de fiabilité. Nous avons utilisé deux méthodes principales : la première est l'estimation du maximum de vraisemblance pour obtenir des valeurs de paramètres précises, et la seconde est l'estimation bayésienne avec des algorithmes MCMC. Nous avons également utilisé la simulation numérique pour tester l'efficacité de ces méthodes.
Mots clés : Extropie, valeurs d'enregistrement supérieures, Distribution de Weibull, Estimation du maximum de vraisemblance, Estimation bayésienne.