République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

Université de Saida - Dr Moulay Tahar.
Faculté des Sciences.
Département de Mathématiques.

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Filière : Mathématiques

Spécialité: Analyse stochastique, statistique des processus et applications (ASSPA)

par

Khaled Bouchra Manel [1]

Sous la direction de

Dr. Rajaa Hazeb

Thème:

STATISTICAL INFERENCES BASED ON RECORD DATA

Soutenue le 10/06/2023 devant le jury composé de

Dr. H. Hennoun	Université de Saïda Dr. Moulay Tahar	Présidente
Dr. R. Hazeb	Université de Saïda Dr. Moulay Tahar	Encadreur
Pr. S. Rahmani	Université de Saïda Dr. Moulay Tahar	Examinatrice

Année universitaire: 2022/2023

[1] e-mail: khaledbouchra99@gmail.com

Dedication

On this occasion, I dedicate my graduation, my success, and all that is beautiful in my life:

To the owner of fragrant biography and enlightened thought... ***My beloved father***

To the one who placed the Lord-Glory to Her, Paradise under her feet, and venerated her in the Holy Quran... ***My beloved mother****,*

To my second mom who supports me in all the little ones and big ones...

My maternal aunt*,*

To those who have never left my side, given me the strength in my life, to the well of my secrets... ***My brothers and sisters****,*

To the ones who always support me in whatever endeavor I undertake, you inspire me to reach higher than my goals... ***My uncles***

To ***My family****,* ***friends and colleagues****... To everyone who has contributed through all my walks of life... To them all: I dedicate this work, which I sincerely ask Almighty God to accept...*

Acknowledgments

I want to start by thanking my parents and all my family.
Above all, I would like to thank my supervisor, Dr. R. Hazeb, for the time she has set aside for me, for her seemingly endless patience, and for the wealth of knowledge that she has shared with me during the preparation of this master thesis.
Also, I would like to thank the committee members, Pr. S. Rahmani and Dr. H. Hennoun, for examining my work, also Pr. A. Kandouci
Then I want to expand my thanks to my classmates for their encouragement and support throughout this dissertation.

Contents

Acknowledgments **3**

Notations And Symbols **8**

Introduction **10**

1 Backgrounds **13**

1.1 Maximum Likelihood Estimation 13

1.1.1 Likelihood and Log-Likelihood Functions 13

1.1.2 Maximum Likelihood Estimate 14

1.1.3 Score Function and Fisher Information 17

1.1.4 Existence and Uniqueness of MLE 18

1.2 Confidence Interval 20

1.3 Bayesian Estimation 20

1.3.1 Posterior Distribution 21

1.3.2 Choice of the Prior Distribution 22

1.3.3 Bayesian Point Estimate 24

1.4 Record Data 27

1.4.1 Standard Record Value Processes 27

1.4.2 Basic Distributional Results 28

1.4.3 Record from Kies Distribution 31

1.4.4 Distributional Properties of Records from Kies Distribution 32

2 Classical and Bayesian Estimation Based on Records from Kies Distribution **42**

2.1 Classical Estimation 42

2.1.1 Maximum Likelihood Estimation 42

2.1.2 Asymptotic Confidence Interval 45

2.1.3 Bootstrap Method 47

2.2 Bayesian Estimation 48
2.3 Comparison between obtained estimators 51

3 Prediction of Records from Kies Distribution **56**
3.1 Classical Point Prediction 56
3.1.1 Maximum Likelihood Predictor 57
3.1.2 Modified Maximum Likelihood Predictor 59
3.1.3 Conditional Median Predictor 59
3.2 Bayesian Point Prediction 60
3.2.1 One-Sample Prediction Problem 60
3.3 Prediction Intervals 63
3.3.1 Pivotal Method 63
3.3.2 The Shortest Length Prediction Intervals 65
3.4 Comparison between Proposed Prediction Methods 67

Conclusion **72**

Refrences **73**

List of Tables

1.1 Summary of conjugate prior distributions for different likelihood functions 23
1.2 Expected values and variances of records from $K(\lambda, \beta)$ with $\lambda = 1, 2$ and $\beta = 0.75, 2$ . . . 35

2.1 Average and MSE Values of the MLEs and Bayes estimates when $\lambda = 1$ and $\beta = 2$ . . . 53
2.2 Average and MSE Values of the MLEs and Bayes estimates when $\lambda = 2$ and $\beta = 1$ . . . 54
2.3 ALs and CPs of 95% CIs of $\lambda = 1$ and $\beta = 2$ . . . 55
2.4 ALs and CPs of 95% CIs of $\lambda = 2$ and $\beta = 1$ . . . 55

3.1 MSPEs and Average Bias from simulations of $\lambda = 1$ and $\beta = 2$ . . . 68
3.2 MSPEs and Average Bias from simulations of $\lambda = 2$ and $\beta = 1$ . . . 69
3.3 CPs and ALs from simulations of $\lambda = 1$ and $\beta = 2$ . . . 70
3.4 CPs and ALs from simulations of $\lambda = 2$ and $\beta = 1$ . . . 70

List of Figures

1.1 pdf of the m^{th} record plots with $\lambda = 1$ and $\beta = \{1, 2, 3, 5\}$ for Y_m , $m = 3$. 36
1.2 pdf of the m^{th} record plots with $\lambda = \{1, 2, 3, 5\}$ and $\beta = 1$ for Y_m , $m = 3$. 36
1.3 pdf of the m^{th} record plots with $\lambda = 1$ and $\beta = \{1, 2, 3, 5\}$ for Y_m , $m = 7$. 37
1.4 pdf of the m^{th} record plots with $\lambda = \{1, 2, 3, 5\}$ and $\beta = 1$ for Y_m , $m = 7$. 37
1.5 pdf of the m^{th} record plots with $\lambda = \{1, 2, 3, 5\}$ and $\beta = 1$ for Y_m , $m = 10$ 38
1.6 pdf of the m^{th} record plots with $\lambda = 1$ and $\beta = \{1, 2, 3, 5\}$ for Y_m , $m = 10$ 38
1.7 pdf of the m^{th} record plots with $\lambda = \{0.5, 0.7, 0.8, 0.9\}$ and $\beta = 0.7$ for Y_m , $m = 3$ 39
1.8 pdf of the m^{th} record plots with $\lambda = \{0.5, 0.7, 0.8, 0.8\}$ and $\beta = 0.7$ for Y_m , $m = 7$ 39
1.9 pdf of the m^{th} record plots with $\lambda = \{0.5, 0.7, 0.8, 0.8\}$ and $\beta = 0.7$ for Y_m , $m = 10$ 40
1.10 pdf of the m^{th} record plots with $\lambda = \{8, 10, 15, 20\}$ and $\beta = 0.5$ for Y_m , $m = 3$ 40
1.11 pdf of the m^{th} record plots with $\lambda = \{8, 10, 15, 20\}$ and $\beta = 0.5$ for Y_m , $m = 7$ 41
1.12 pdf of the m^{th} record plots with $\lambda = \{8, 10, 15, 20\}$ and $\beta = 0.5$ for Y_m , $m = 10$ 41

List Of Notations And Symbols

ACIs :	Asymptotic confidence intervals
ALs :	Average lengths
$\mathbf{Boot-p}$:	Percentile Bootstrap
cdf :	cumulative distribution function
CIs :	Confidence intervals
CMP :	Conditional median predictor
CPs :	Coverage probabilities
LINEX :	linear exponential
log :	Logarithm function
MLE :	Maximum Likelihood Estimation
MLP :	Maximum likelihood predictor
MMLP :	Modified maximum likelihood predictor
MSEs :	Mean square errors $\mathbb{E}\left(\hat{\theta}-\theta\right)^2$
MSPEs :	Mean square prediction errors
PCDF :	Predictive cumulative distribution function
pdf :	probability density function
pmf :	probability mass function
PIs :	Prediction intervals
PSF :	Predictive survival function
SE :	Squared error
SL :	Shortest length
rs :	random sample
rv :	random variable
$\mathcal{B}$:	Binomial Distribution
β :	Beta Distribution
$\mathcal{E}\mathbf{xp}$:	Exponential Distribution
G :	Gamma Didtribution
Hyp Geo :	Hypergeometric Distribution
$\mathcal{N}$:	Normal Distribution
$\mathcal{P}$:	Poisson Distribution
U :	Uniform distribution
$\mathcal{X}^{\mathbf{2}}_{\mathbf{2(s-m)}}$:	Chi square Distribution

$\mathbf{A_i}$:	$\dfrac{y_i}{1-y_i}$
$\mathbf{B}(.,.)$:	Beta function
$\mathbb{E}(.)$:	Expected value
$\mathbf{f_s^p}(.)$:	Posterior predictive density
$\hat{\mathbf{f}}_{\mathbf{s}}^{*}(.)$:	Approximate Posterior predictive density
$\mathbf{H}(.)$:	Cumulative hazard rate function
$\mathbf{h}(.)$:	Hazard rate function
$\mathbf{H}$:	Hessian matrix
$\mathbf{I}(.,.)$:	Observed information matrix
$\mathbf{IG}$:	Invers Gamma
$\mathbf{K}\,(\lambda,\beta)$:	Two-parameter Kies distribution
$\mathbf{L}(.)$:	Likelihood function
$\mathbf{m}$:	Number of records
$\mathbb{P}$:	Probability
$\mathbf{S}(.)$:	Survival function
$\mathbf{S^p}(.)$:	Survival prediction function
$\mathbf{Var}(.)$:	variance
$\mathbf{W_{c_1,c_2}}(.)$:	Whittaker function
$\pi(.)$:	Joint posterior density
$\mu^{\mathbf{k}}$:	k^{th} Moment
ν :	The shape parameter of LINEX loss function
$\mathbf{\Gamma}(.,.)$:	Upper incomplete gamma function
$\gamma(.,.)$:	Lower incomplete gamma function
$(.)_{\mathbf{k}}$:	Pochhammer symbol

Introduction

All of as know the famous book(Guinness), the longest winning streak, the largest and the most powerful earthquake in the world, the temperature of the hottest day ever,the tallest skyscraper,the deepest dive without using an aqualung, the lowest stock market figure, the two highest months for record reporting are January and October and the highest rainfall level on the earth, these all are called records values, so we can't resist record in our daily life. The literature on records theory is insufficient and scanty for two main reasons:

- When studying the basic record model involving i.i.d. observations that was introduced by Chandler (1952) [9], in a sense is too easy to analyze and as soon as we add bells and whistles to better model reality, it suddenly becomes too hard.

- The study of records deters anyone want to develop inferential techniques because of the lack of records phenomena, So the obtained sample are small.

Record value theory has its own existence (see Arnold et all [5]).
The study of record values has been carried out by a relatively and highly talented group of individuals. Chandler (1958) [9] was the founder of records. Stuart(1954) [13] was a pioneer in discovering record counting statistics. Record statistics are defined as model for successive extremes in a sequence of independently identically distributed random variables. For instance, it may be helpful for modeling the successive largest insurance claims studied by an insurance company, for highest water levels or highest temperatures. For more details about records and their applications, one may refer to Nagaraja (1988) [25], Arnold et al. (1998) [5] and Ahsanuallah (2004) [2].

In many aspects, the study of record values is adjacent to the study of order statistics; indeed, the two are intrinsically linked. This means that, in general, things that are doable for order statistics are also doable for records. Things that were challenging for order statistics are now equally or even more complex for records. Specific distribution findings for the n^{th} record are usually only available for friendly distributions. Fortunately, there are numerous interrelationships, boundaries, and approximations for distributions and moments.
For its importance in modeling lifetimes in many practical fields, the Weibull distribution has received the attention of several authors in the literature. Moreover, many modified versions of the Weibull distribution were developed in the literature, one of the modified versions of the Weibull distribution is known as Kies distribution that was firstly proposed by Kies (1958) [17], Kies distribution has an increasing, decreasing and bathtub shape for

hazard rate function like the other extended Weibull models. Due to its bathtub shaped hazard rate function, Kies distribution becomes a better alternative to the Weibull distribution than the other extended versions including the generalized Weibull (GW), the modified Weibull (MW), the beta Weibull (BW), the beta generalized Weibull (BGW) distributions. The Kies distribution, indicated by $K(\lambda, \beta)$, has a bounded range, making it suitable model for fitting real-life data sets having a finite range. However, there are many cases in which observations can only take values within a specific range, such as proportions, percentages or fractions. Variables in many economic applications, such as the fraction of total weekly hours spent working, the proportion of income spent on non-durable consumption, industry market shares, and a fraction of land area allocated to agriculture, are all bounded between zero and one, according to Papke and Wooldridge (1996) [26]. Furthermore, Genc (2013) [16] stated that when measuring reliability as a percentage or ratio, it is critical to employ models established on the unit interval.
Recently, several authors in the literature have paid attention to Kies distribution. The reduced Kies (RK) distribution, a specific example of the one-parameter Kies distribution that Kumar and Dharmaja (2013) [18] investigated, is demonstrated to have several distinctive characteristics that are similar to those of the Weibull distribution. The generalized Weibull (GW), modified Weibull (MW), beta Weibull (BW), and beta generalized Weibull (BGW) distributions are some versions of the extended Weibull distributions. In another paper, Kumar and Dharmaja (2014) [19] studied some of the Kies distribution's statistically important aspects and demonstrated that it possesses increasing, decreasing, and bathtub hazard rate functions. The prediction of future events based on past and present data is a fundamental topic in statistics. Extensive work on prediction may be found in the literature in the context of record statistics. Bayesian predictive distribution of future records based on observed records for two parameters exponential distribution was derived by Dunsmore (1983) [11]. Awad and Raqab (2000) [6] obtained prediction intervals for the future records based on observed records from one parameter exponential distribution. Basak and Balakrishnan (2003) [7] discussed the problem of predicting a future record statistics using maximum likelihood prediction and the conditional median predictor.

The main objective of this study is divided into two folds. Firstly, we consider the estimation of the two unknown parameters (shape and scale) of the Kies distribution, $K(\lambda, \beta)$, based on upper record data using classical and Bayesian methods. Second, we study the prediction problem of future records from $K(\lambda, \beta)$ distribution using classical and one-sample Bayesian prediction.
The rest of my dissertation work is organized as follows :
We set some of needed background, definitions, properties of record values from the Kies $K(\lambda, \beta)$ distribution. Also, some basic definitions of the considered estimation methods

in Chapter 1.
In Chapter 2, the scale parameter λ and the shape parameter β of $K(\lambda, \beta)$ are estimated using classical and Bayesian techniques, such as the maximum likelihood and confidence interval. The Bayesian approach takes into account both symmetric and asymmetric loss functions of λ and β. Prediction based on record samples from $K(\lambda, \beta)$ is considered in Chapter 3, with different point predictors including maximum likelihood, conditional median and one-sample Bayesian predictors are obtained. Further, to construct prediction intervals of future records, the pivotal quantity and shortest length interval is investigated. For comparison purposes we have considered some simulation results on the behavior of the proposed estimation and prediction methods.

Chapter 1

Backgrounds

This chapter is divided into three important sections. The first one contains some definitions of maximum likelihood estimation. Then, some of definitions and properties on bayesian estimation are also provided. Finally, record data from kies distribution with distributional properties are defined .

1.1 Maximum Likelihood Estimation

Maximum likelihood estimation is a probabilistic method for addressing the density estimation issue. It entails finding the probability distribution and parameters that best describe the given data by maximising a likelihood function.

1.1.1 Likelihood and Log-Likelihood Functions

Let $X = x$ represent the realisation of a random variable (rv) or vector X with probability mass or density function (pmf or *pdf*) $f(x;\theta)$. The function $f(x;\theta)$ is considered to be known but depends on the realisation x and often unknown parameter θ. It is usually the result of the development of an appropriate statistical model.
Note that θ can be a scalar or a vector. The space Ω of all possible realisations of X is called sample space, whereas the parameter θ can take values in the parameter space Θ.
Definition 1.1.1. The likelihood function $L(\theta)$ is the pmf or pdf of the observed data x, viewed as a function of the unknown parameter θ. i.e:

$$L(\theta; x) = f(x;\theta)$$

The log-likelihood function can be rewritten as $logL(\theta)$.

When X is a random simple (rs), as a generalization of previous definition, we assume $x_1, ..., x_n$ to be n observations of the random vector $\underline{X} = (X_1, ..., X_n)$ such that $X_i \overset{iid}{\sim} f(x;\theta), i = 1....n.$

The *pdf* of $\underline{X}$ is:

$$f(\underline{x};\underline{\theta}) = \prod_{i=1}^{n} f(x_i;\underline{\theta}).$$

So, the likelihood function based on a random sample (rs) can therefore be written as the product of the individual likelihood contributions as

$$L(\underline{\theta};\underline{x}) = \prod_{i=1}^{n} L(\underline{\theta};x_i) = \prod_{i=1}^{n} f(x_i;\underline{\theta}).$$

The log-likelihood is hence the sum of the individual log-likelihood contributions as

$$logL(\underline{\theta};\underline{X}) = log\left(\prod_{i=1}^{n} f(x_i;\underline{\theta})\right) = \sum_{i=1}^{n} logf(x_i;\underline{\theta}).$$

Remark. There are terms that are not related to the parameter, which are called multiplicative constants and the remaining terms are called the likelihood kernel.

1.1.2 Maximum Likelihood Estimate

Maximum likelihood provides a consistent approach to parameter estimation problem. This means that maximum likelihood estimates can be developed for a large variety of estimation situations. For example, they can be applied in reliability analysis to censored data under various censoring models.

Definition 1.1.2. 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 log-likelihood function is frequently numerically suitable $logL(\theta)$ it can be used to calculate the MLE, since the logarithm is a strictly monotone function, So:

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

Example 1.1.1. Let X be an rv, that represent the observed number of babies with Klinefelter's syndrome among n male newborns. The number of male newborns n is hence known, while the true prevalence θ of Klinefelter's syndrome among male newborns is unknown. $X \sim Bin\,(n, \theta)$ denote a binomially distributed rv

The corresponding likelihood function is

$$L\,(\theta) = \mathcal{C}_n^x \theta^x \,(1-\theta)^{n-x}\,.$$

with unknown parameter $\theta \in (0,1)$ and sample space $\Omega = \{0,1,...,n\}$. The multiplicative term $\mathcal{C}_n^x$ does not depend on θ and can therefore be ignored, i.e. it is sufficient to consider the log-likelihood $\theta^x\,(1-\theta)$. The log-likelihood turns out to be

$$logL\,(\theta) = log\left(\mathcal{C}_n^x \theta^x \,(1-\theta)^{n-x}\right)$$

$$logL\,(\theta) = log\,(\mathcal{C}_n^x) + xlog\,(\theta) + (n-x)\,log\,(1-\theta)$$

with derivative

$$\frac{\partial logL\,(\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

Setting this derivative to zero:

$$\frac{\partial logL\,(\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0$$

$$\frac{x\,(1-\theta) - (n-x)\,\theta}{\theta\,(1-\theta)} = 0$$

gives

$$x - n\theta = 0$$

so the MLE is:

$$\hat{\theta}_{ML} = \frac{x}{n}.$$

Example 1.1.2. Let $\underline{X}$ denote a rs from an exponential distribution $\mathcal{E}xp\,(\theta)$.Then:

$$\begin{aligned} L\,(\theta) &= \prod_{i=1}^{n}\{\theta e^{(-\theta x_i)}\} \\ &= \theta^n e^{\left(-\theta \sum\limits_{i=1}^{n} x_i\right)} \end{aligned}$$

is the likelihood function of $\theta \in \mathbb{R}^+$. The log-likelihood function is therefore:

$$logL(\theta) \quad = \quad nlog(\theta) - \theta \sum_{i=1}^{n} x_i$$

with derivative:

$$\frac{\partial logL(\theta)}{\partial\theta} \quad = \quad \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

Setting the derivative to zero, we easily obtain the MLE $\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$ is the mean observed survival time.

Theorem 1.1.1. (Asymptotic Normality of MLE)[14]

Let $\hat{\theta}$ be the MLE for an unknown parameter θ. Then, we have

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

As we can see, the asymptotic variance dispersion of the estimate around true parameter will be smaller when Fisher information is larger.

Definition 1.1.3. (Delta Method)

The delta method is a result concerning the approximate probability distribution for a function of an asymptotically normal statistical estimator from knowledge of the limiting variance of that estimator. The delta method generalizes easily to a multivariate setting, careful motivation of the technique is more easily demonstrated in univariate terms. Roughly, if there is a sequence of random variables X_n satisfying

$$\sqrt{n}\,[X_n - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where θ and σ^2 are finite valued constants and $\xrightarrow{d}$ denotes convergence in distribution, then

$$\sqrt{n}\,[g(X_n) - g(\theta)] \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 . \left[g'(\theta)\right]^2\right)$$

for any function g satisfying the property that $g'(\theta)$ exists and is non-zero valued.

1.1.3 Score Function and Fisher Information

Definition 1.1.4. The first derivative of the log-likelihood function

$$S(\theta) = \frac{\partial logL(\theta)}{\partial \theta}$$

is called the **score function**.

Remark. Computation of the MLE is typically done by solving the score equation $S(\theta) = 0$.

Definition 1.1.5. The negative second derivative of the log-likelihood function

$$I(\theta) = -\frac{\partial^2 logL(\theta)}{\partial \theta^2} \tag{1.1}$$

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

Example 1.1.3. The score function of a binomial observation $X = x$ with $x \sim Bin(n, \theta)$ is:

$$S(\theta) = \frac{\partial logL(\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta} \tag{1.2}$$

and has been derived already in Example(1.1.1). Taking the derivative of $S(\theta)$ gives the Fisher information

$$\begin{aligned} I(\theta) &= -\frac{\partial^2 logL(\theta)}{\partial \theta^2} = -\frac{\partial S(\theta)}{\partial \theta} \\ &= \frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \\ &= n\left\{\frac{x/n}{\theta^2} - \frac{(n-x)/n}{(1-\theta)^2}\right\} \end{aligned}$$

Plugging in the MLE $\hat{\theta}_{ML} = \frac{x}{n}$ we finally obtain the observed Fisher information

$$I\left(\hat{\theta}_{ML}\right) = \frac{n}{\hat{\theta}_{ML}\left(1 - \hat{\theta}_{ML}\right)} \tag{1.3}$$

This result is plausible if we take a frequentist point of view and consider the MLE

as a random variable. Then

$$\begin{aligned} Var\left(\hat{\theta}_{ML}\right) &= Var\left(\frac{x}{n}\right) = \frac{1}{n^2}.Var\left(x\right) \\ &= \frac{1}{n^2}n\theta\left(1-\theta\right) \\ &= \frac{\theta\left(1-\theta\right)}{n} \end{aligned}$$

Remark. • Note that the MLE $\hat{\theta}_{ML}$ is a function of the observed data, which explains the terminology "observed" Fisher information for $I\left(\hat{\theta}_{ML}\right)$.

• the variance of $\hat{\theta}_{ML}$ has the same form as the inverse observed Fisher information; the only difference is that the MLE $\hat{\theta}_{ML}$ is replaced by the true (and unknown) parameter θ. The inverse observed Fisher information is hence an estimate of the variance of the MLE.

1.1.4 Existence and Uniqueness of MLE

In this section we will illustrates un example that treats both cases of existence and uniqueness of MLE. The uniqueness of the MLE is not guaranteed, and in certain examples there may exist at least two parameter values $\hat{\theta}_1 \neq \hat{\theta}_2$ with $L\left(\hat{\theta}_1\right) = L\left(\hat{\theta}_2\right) = \arg\max_{\theta\in\Theta} L\left(\theta\right)$. In other situations, the MLE may not exist at all.

To acheive this goels we consider the capture-recapture method. This approach attempts to estimate the size of an individual population, such as the number N of fish in a lake.
To do this, a sample of M fishes is collected from the lake, all of which are marked, and then put back into the lake.After a suitable amount of time has passed, a second sample of size n is collected, and the number x of marked fish in that sample is recorded.
The target now is to estimate N from M, n and x.
By correlating the proportion of marked fish in the lake with the equivalent proportion in the sample, an ad hoc estimate can be obtained:

$$\frac{M}{N} \approx \frac{x}{n}$$

As a result, the number N of fish in the lake is approximated to be $\hat{N} \approx M\frac{n}{x}$.

When we put $X = x$, The suitable statistical model for X is therefore a hypergeometric distribution:

$$X \sim HypGeo\,(n, N, M)$$

with *pdf*

$$\mathbb{P}\,(X = x) = f\,(x; \theta = N) = \frac{\mathcal{C}_M^x \mathcal{C}_{N-M}^{n-x}}{\mathcal{C}_N^n}$$

for $x \in \tau = \{\max\{0, n - (N - M)\}, ..., \min\,(n, M)\}$. The likelihood function for N is therefore

$$L\,(N) = \frac{\mathcal{C}_M^x \mathcal{C}_{N-M}^{n-x}}{\mathcal{C}_N^n}$$

for $N \in \Theta = \{\max\,(n, M + n - x)\,, \max\,(n, M + n - x) + 1, ...\}$, where we could have ignored the multiplicative constant $\mathcal{C}_M^x \dfrac{n!}{(n-x)!}$.

The likelihood function is maximised at $\hat{N}_{ML} = \lfloor M \cdot \dfrac{n}{x} \rfloor$,where $\lfloor y \rfloor$denotes the largest integer not greater than y. For example, for $M = 26$, $n = 63$ and $x = 5$ we obtain $\hat{N}_{ML} = \lfloor 26 \cdot 63/5 \rfloor = \lfloor 327.6 \rfloor = 327$.
Sometimes the MLE is not unique, and the likelihood function attains the same value at $\hat{N}_{ML} - 1$. Like when we put $M = 13$, $n = 10$ and $x = 5$, we have $\hat{N}_{ML} = \lfloor 13 \cdot \dfrac{10}{5} \rfloor = 26$, but $\hat{N}_{ML} = 25$ also attains exactly the same value of $L\,(N)$.

For existence if we take $x = 0$, the MLE does not exist because the likelihood function $L\,(N)$ is then monotonically increasing.

The next theorem provide the existence and uniqueness of MLE, was introduced by Mkelinen.T et al. (1981)[24]
Theorem 1.1.2. [24] Let $L(\underline{\theta}) \in \mathcal{C}^2$ with θ varying in a connected open subset $\Theta \subset R^p$. Suppose that

$$\lim_{\theta \to \Theta} L(\underline{\theta}) = 0 \tag{1.4}$$

and that the Hessian matrix :

$$H(\underline{\theta}) = \left\{ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} (\underline{\theta}) \right\} \tag{1.5}$$

of second partial derivatives is negative definite at every point $\theta_i \in \Theta$ for which the gradient vector

$$\nabla L = \left\{ \frac{\partial L}{\partial \theta_i} \right\} \tag{1.6}$$

vanishes. Then, there is a unique maximum likelihood estimate $\underline{\hat{\theta}} \in \Theta$.

1.2 Confidence Interval

When you make an estimate in statistics, whether it is a summary statistic or a test statistic, there is always uncertainty around that estimate because the number is based on a sample of the population you are studying. The confidence interval is the range of values that you expect your estimate to fall between a certain percentage of the time if you run your experiment again or re-sample the population in the same way with a confidence level which is the percentage of times you expect to reproduce an estimate between the upper and lower bounds of the confidence interval.

Definition 1.2.1. For fixed $\gamma \in (0, 1)$, a $\gamma.100\%$ confidence interval for θ is defined by two statistics $T_l = h_l\left(\underline{X}\right)$ and $T_u = h_u\left(\underline{X}\right)$ based on a rs $\underline{X}$, which fulfill

$$\mathbb{P}\left(T_l \leq \theta \leq T_u\right) = \gamma \tag{1.7}$$

for all $\theta \in \Theta$. The statistics T_l and T_u are the limits of the confidence interval, and we assume $T_l \leq T_u$ throughout. The confidence level γ is also called coverage probability.

1.3 Bayesian Estimation

Frequentist inference considers the data X to be random. To get frequentist characteristics of the estimates, point estimate of the parameter θ are seen as functions of the data X. The parameter θ is unknown, but it is viewed as fixed rather than random.

In the Bayesian method to statistical reasoning, named after Thomas Bayes, things are simply the other way around. The unknown parameter θ is now a random variable with the proper prior distribution $f\left(\theta\right)$. The posterior distribution $f\left(\theta|x\right)$, determined by Bayes' theorem, summarises the information to θ after observing the data $X = x$. In contrast to frequentist inference, Bayesian inference is conditional on the observation $X = x$.

Theorem 1.3.1. [14] Let A and B denote two events A,B with $0 < \mathbb{P}\left(A\right) < 1$ and $\mathbb{P}\left(B\right) > 0$.then

$$\mathbb{P}\left(A|B\right) = \frac{\mathbb{P}\left(B|A\right)\mathbb{P}\left(A\right)}{\mathbb{P}\left(B\right)} \tag{1.8}$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B|A)\,\mathbb{P}(A) + \mathbb{P}(B|A^c)\,\Delta\mathbb{P}(A^c)} \tag{1.9}$$

For a general partition $A_1, A_2, ..., A_n$ with $\mathbb{P}(A_i) > 0$ for all $i = 1, ..., n$ we have that

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\,\mathbb{P}(A_j)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\,\mathbb{P}(A_i)} \tag{1.10}$$

for each $j = 1, ..., n$.

1.3.1 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 estimate.

Definition 1.3.1. Let $X = x$ denote the observed realization of a (possibly multivariate) rv 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.

$$f(\theta|x) = \frac{f(x|\theta)\,f(\theta)}{\int f(x|\theta)\,f(\theta)\,d\theta} \tag{1.11}$$

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

Remark. • The term $f(x|\theta)$ in (1.11) is simply the likelihood function $L(\theta)$ denoted previously.

• Since θ is now random, we explicitly condition on a specific value θ and write $L(\theta) = f(x|\theta)$.

• The denominator in (1.11) can be written as

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

• The density function of the posterior distribution can be obtained through multiplication of the likelihood function and the density function of the prior distribution with subsequent normalisation.

Example 1.3.1. Inference for a proportion is based on a rs of size n and determines the number $X = x$ of individuals in this sample with a certain event of interest. It is often reasonable to assume that $X \sim Bin\,(n, \theta)$ where $\theta \in (0, 1)$ is the unknown probability of the event. It is tempting to select a beta distribution as a prior distribution for θ because the support of a beta distribution equals the parameter space $(0, 1)$. So let the prior be $\theta \sim \mathcal{B}\,(\alpha, \beta)$ with suitably chosen parameters $\alpha, \beta > 0$. Then

$$f\,(x|\theta) = \mathcal{C}_n^x \theta^x\,(1-\theta)^{n-x}\,. \qquad\qquad x = 0, 1, ..., n$$

$$f\,(\theta) = \frac{1}{B\,(\alpha, \beta)}\theta^{\alpha-1}\,(1-\theta)^{\beta-1}\,. \qquad\qquad 0 < \theta < 1$$

where $B\,(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. So, the posterior distribution is

$$\begin{aligned} f\,(\theta|x) &\propto f\,(x|\theta) \cdot f\,(\theta) \\ &\propto \theta^x\,(1-\theta)^{n-x} \cdot \theta^{\alpha-1}\,(1-\theta)^{\beta-1} \\ &= \theta^{\alpha+x-1}\,(1-\theta)^{\beta+n-x-1} \end{aligned}$$

This can be easily identified as yet another $\mathcal{B}$ distribution with parameters $\alpha + x$ and $\beta + n - x$:

$$\theta|x \sim \mathcal{B}\,(\alpha + x, \beta + n - x) \tag{1.12}$$

1.3.2 Choice of the Prior Distribution

Through a prior distribution, Bayesian inference enables the probabilistic specification of prior beliefs. It is usually helpful and reasonable to limit the set of potential prior distributions to a certain family with, say, one or two parameters .
The choice of this family can be based on the type of likelihood function encountered. We will now discuss such a choice.

Conjugate prior distributions

A pragmatic method to selecting a prior distribution is to choose a member of a given family of distributions so that the posterior distribution also belongs to that family. This is referred to as a conjugate prior distribution.

Definition 1.3.2. Let $L(\theta) = f(x|\theta)$ 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 $f(\theta|x)$ is in $\mathcal{G}$ for all x whenever the prior distribution $f(\theta)$ is in $\mathcal{G}$.
Remark. The family $\mathcal{G} = \{\text{all distributions}\}$ 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 $L_x(\theta)$.

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

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

Improper Prior Distributions

The prior distribution has an intended influence on the posterior distribution. If one wants to minimize the influence of the prior distribution, then it is common to specify a vague prior, for example one with very large variance. In the limit this may lead to an improper prior distribution, with a "density" function that does not integrate to unity. Due to the missing normalizing constant, such density functions are usually specified using the proportionality sign "$\propto$". If one uses improper priors, then it is necessary to check that at least the posterior distribution is proper. If this is the case, then improper priors can be used in a Bayesian analysis. We now add a formal definition of an improper prior distribution.
Definition 1.3.3. A prior distribution with density function $f(\theta) \geq 0$ is called improper if

$$\int_{\Theta} f(\theta)\, d\theta = \infty \quad or \quad \sum_{\theta\in\Theta} f(\theta) = \infty \tag{1.13}$$

for continuous or discrete parameters θ, respectively.

Jeffreys Prior Distributions

It turns out that a particular choice of prior distribution is invariant under reparametrisation. This is Jeffreys prior (after Sir Harold Jeffreys, 1891-1989).

Definition 1.3.4. Let X be a rv with likelihood function $f(x|\theta)$ where θ is an unknown scalar parameter. Jeffreys prior is defined as

$$f(\theta) \propto \sqrt{J(\theta)} \tag{1.14}$$

where $J(\theta)$ is the expected Fisher information of θ. Equation (1.14) is also known as Jeffreys rule.

1.3.3 Bayesian Point Estimate

Definition 1.3.5.

1. The posterior mean $\mathbb{E}(\theta|x)$ is the expectation of the posterior distribution:
$$\mathbb{E}(\theta|x) = \int \theta f(\theta|x)\, d\theta$$

2. The posterior mode Mod$(\theta|x)$ is the mode of the posterior distribution:
$$Mod(\theta|x) = \arg\max_{\theta} f(\theta|x)$$

3. The posterior median Med$(\theta|x)$ is the median of the posterior distribution, i.e. any number a that satisfies
$$\int_{-\infty}^{a} f(\theta|x)\, d\theta = 0.5 \qquad and \qquad \int_{a}^{\infty} f(\theta|x)\, d\theta = 0.5 \tag{1.15}$$

Properties of Bayesian Point Estimation

To estimate an unknown parameter θ , there are at least three possible Bayesian point estimates available, the posterior mean, mode and median. Which one should we choose in a specific application ? To answer this question, we take a decision−theoretic view and first introduce the notion of a loss function.

Definition 1.3.6. Loss function

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

Remark. • 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 $l(a,\theta) = |a-\theta|$

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

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 $f(\theta|x)$. Such a point estimate is called a **Bayes estimate**.

Definition 1.3.7. (Bayes estimate)

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

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

Theorem 1.3.2. [14]

1. The Bayes estimate with regard to quadratic loss is the posterior mean.

2. The posterior median is the Bayes estimate with respect to linear loss.

3. The posterior mode is the Bayes estimate with respect to zero$-$one loss, as $\varepsilon \to 0$.

To prove theorem (1.3.2) we need the next integral rule.

Definition 1.3.8. (Leibniz Integral Rule)

Let a, b and f be real-valued functions that are continuously differentiable in t .Then the Leibniz integral rule is

$$\frac{\partial}{\partial t}\int_{a(t)}^{b(t)} f(x,t)\, dx = \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f(x,t)\, dx - f\{a(t),t\}.\frac{\partial}{\partial t} b(t) \tag{1.16}$$

This rule is also known as differentiation under the integral sign.

Proof. We first derive the posterior mean $\mathbb{E}(\theta|x)$ as the Bayes estimate with respect to quadratic loss. The expected quadratic loss is

$$\begin{aligned} \mathbb{E}\{l(a,\theta)|x\} &= \int l(a,\theta) f(\theta|x)\, d\theta \\ &= \int (a-\theta)^2 f(\theta|x)\, d\theta \end{aligned}$$

Setting the derivative with respect to a to zero leads to

$$2\int (a-\theta)\, f(\theta|x)\, d\theta = 0 \Leftrightarrow a - \int \theta f(\theta|x)\, d\theta = 0 \tag{1.17}$$

It immediately follows that $a = \int \theta f(\theta|x)\, d\theta = \mathbb{E}(\theta|x)$. Consider now the expected linear loss

$$\begin{aligned}
\mathbb{E}\{l(a,\theta)\,|x\} &= \int l(a,\theta)\, f(\theta|x)\, d\theta = \int |a-\theta| f(\theta|x)\, d\theta \\
&= \int_{\theta\le a} (a-\theta)\, f(\theta|x)\, d\theta + \int_{\theta>a} (\theta-a)\, f(\theta|x)\, d\theta
\end{aligned}$$

The derivative with respect to a can be calculated using Leibniz′s integral rule (1.3.8):

$$\begin{aligned}
\frac{\partial}{\partial a}\mathbb{E}\{l(a,\theta)\,|x\} &= \frac{\partial}{\partial a}\int_{-\infty}^{a} (a-\theta)\, f(\theta|x)\, d\theta + \frac{\partial}{\partial a}\int_{a}^{\infty} (\theta-a)\, f(\theta|x)\, d\theta \\
&= \int_{-\infty}^{a} f(\theta|x)\, d\theta - (a-(-\infty))\, f(-\infty|x)\,.0 + (a-a)\, f(a|x)\,.1 \\
&- \int_{a}^{\infty} f(\theta|x)\, d\theta - (a-a)\, f(a|x)\,.1 + (\infty-a)\, f(\infty|x)\,.0 \\
&= \int_{-\infty}^{a} f(\theta|x)\, d\theta - \int_{a}^{\infty} f(\theta|x)\, d\theta
\end{aligned}$$

Setting this equal to zero yields the posterior median $a = Med(\theta|x)$ as the solution for the estimate. Finally, the expected zero-one loss is

$$\begin{aligned}
\mathbb{E}\{l(a,\theta)\,|x\} &= \int l_{\varepsilon}(a,\theta)\, f(\theta|x)\, d\theta \\
&= \int_{-\infty}^{a-\varepsilon} f(\theta|x)\, d\theta + \int_{a+\varepsilon}^{+\infty} f(\theta|x)\, d\theta \\
&= 1 - \int_{a-\varepsilon}^{a+\varepsilon} f(\theta|x)\, d\theta
\end{aligned}$$

This will be minimised if the integral $\int_{a-\varepsilon}^{a+\varepsilon} f(\theta|x)\, d\theta$ is maximised. For small ε the integral is approximately $2\varepsilon f(a|x)$, which is maximised through the posterior mode $a = Mod(\theta|x)$.

□

1.4 Record Data

Chandler (1952)[9] pioneered the study of record values and outlined many of their fundamental characteristics. Record values are used in many statistical applications, including statistical modeling and inference with data from studies on the weather, sports, economics, and other topics. The fastest period to recite the periodic table of the elements, the shortest tennis matches in terms of the number of games and time, or the fastest indoor marathon are a few examples from Guinness World Records. There are many efforts attempted to break records, but records are only broken when an attempt is successful. The vast majority of efforts to break the record go unreported. The main findings in this section were discovered between 1952 and 1983.

1.4.1 Standard Record Value Processes

There are two types of records, upper and lower records. Let $X_1, X_2, ...$ be a sequence of identical and independent rv from a continuous distribution.
An observation X_j is called an upper record (or simply a record) if $X_j > X_i$ for all $j > i$ i.e an upper record value is a value that is larger than all the previous observations and a lower record value is a value that is smaller than all the previous observations i.e if $X_j < X_i$ for all $j > i$, X_j an observation is called a lower record. The times at which records appear are of interest. For convenience, let us assume that X_j is observed at time j. Then the record time sequence $\{T_n, n \geq 0\}$ is defined in the following manner:

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

and, for $n \geq 1$,

$$T_n = \min\{j : X_j > X_{T_{n-1}}\}. \tag{1.18}$$

The record value sequence $\{R_n\}$ is then defined by

$$R_n = X_{T_n}, \quad n = 0, 1, 2, ... \tag{1.19}$$

Here R_0 is referred to as the reference value or the trivial record. The rest are treated as nontrivial records. The record sequence described above implicitly presupposes that the cdf F won't result in any unbreakable records. That is, if the X_j can obtain the largest real value with a high degree of probability. For example, By rolling a dice repeatedly, the result of $X_j = 6$ is an unbreakable record. X_j stands for the number received on the j-th throw.

1.4.2 Basic Distributional Results

The distributions of the record values R_n are predictably affected by F. These observations suggest the desirability of carefully selecting a common distribution for the X_j's, to make derivations as simple as possible. Thus, therefore Arnold et al. (1998)[5] have considered the classical model, as a strong argument that can be made in favor of studying i.i.d. exponentially distributed X_j's.
The exponential distribution has the lack of memory property and consequently $\{J^*\} = \left\{R_n^* - R_{n-1}^*; n \geq 1\right\}$ are i.i.d. $\mathcal{E}xp(1)$ random variables. It follows that for the n_{th} record, R_n^* corresponding to an i.i.d. $\mathcal{E}xp(1)$ sequence, we have

$$R_n^* \sim G(n+1, 1) \qquad n = 0, 1, 2, ... \tag{1.20}$$

We may use this result to obtain the distribution of the n^{th} record corresponding to an i.i.d. sequence of random variables $\{X_j\}$ with common continuous cdf F for more information see Arnold and all(1998)[5]. If X has a continuous cdf F, then

$$H(X) = -log\,(1 - F(X)) \tag{1.21}$$

has a standard exponential distribution. Consequently, $X \overset{\mathrm{d}}{=} F^{-1}(1-e^{-X})$ where, as usual, X^* is $\mathcal{E}xp(1)$. Since X is a monotone function of X^*, the n^{th} record of the $\{X_j\}$ sequence is expressible as a simple function of the n^{th} record of an $\{X^*\}$ (standard exponential) sequence. Specifically we have

$$R_n \overset{\mathrm{d}}{=} F^{-1}(1 - e^{-R_n^*}), \qquad n = 0, 1, 2, ... \tag{1.22}$$

Then using the relation (1.22), we derive the survival function of the n^{th} record corresponding to an i.i.d. F sequence.

$$\mathbb{P}(R_n > r) = [1 - F(r)] \sum_{k=0}^{n} \frac{[-log\,(1 - F(r))]^k}{k!} \tag{1.23}$$

and the joint pdf is obtained as follows :

$$f_{R_0^*, R_1^*, ..., R_n^*}(r_0^*, r_1^*, ..., r_n^*) = e^{-r_n^*}, \qquad 0 < r_0^* < r_1^* < ... < r_n^* \tag{1.24}$$

Definition 1.4.1. The pdf of the n^{th} record R_n is given by :

$$f_{R_n}(r) = f(r)\frac{[-log\,(1 - F(r))]^n}{n!} \tag{1.25}$$

the joint pdf of the set of records $R_0, R_1, ..., R_n$ it is given by:

$$f_{R_0,R_1,...,R_n}(r_0, r_1, ..., r_n) \;=\; f(r_n)\prod_{i=0}^{n-1} h(r_i) \tag{1.26}$$

$-\infty < r_0 < r_1 < ... < r_n < \infty$, where $h(r) = \dfrac{\partial H(r)}{\partial r} = \dfrac{f(r)}{[1 - F(r)]}$ represents the failure rate function.

Record Values from Specific Distributions

The record value sequence admits a clear and concise description for specific cdf F choices. We have already remarked on the simplicity encountered when $F\,(x) = 1 - e^{-x}$. In this case the n^{th} record has a $G\,(n + 1, 1)$ distribution. This is included as a special case of the Weibull distribution. The list of parent distributions for which the record sequence may be represented simply is not too long. We will now go to defined some specific distributions.

Weibull Records

Suppose that the distribution of the population random variable X is Weibull with cdf

$$F\,(x) = \begin{cases} 0, & x \leq 0; \\ 1 - e^{-\left(\frac{x}{\sigma}\right)^k}, & x > 0. \end{cases} \tag{1.27}$$

where the parameters are $\sigma, k > 0$, Such a random variable admits the representation

$$X \stackrel{\text{d}}{=} \sigma X^{*\frac{1}{k}} \tag{1.28}$$

where X^*, as usual, is an $\mathcal{E}xp(1)$ random variable. Consequently, for $n = 0, 1, 2, ...$

$$R_n \stackrel{\text{d}}{=} \sigma \left(\sum_{i=0}^{n} X_i^*\right)^{\frac{1}{k}} \tag{1.29}$$

Power Function Distribution Records

If the population *cdf* is of the form

$$F(x) = \begin{cases} 0, & x \leq 0; \\ x^{\alpha}, & 0 < x < 1; \\ 1, & x \geq 1. \end{cases} \tag{1.30}$$

where $\alpha > 0$, a simple description of the record value sequence is possible, for $n = 0, 1, 2, ...$ we have

$$R_n \stackrel{\mathrm{d}}{=} \left[1 - \prod_{j=0}^{n}(1 - U_j)\right]^{\frac{1}{\alpha}} \tag{1.31}$$

where $\{U_j, j \geq 0\}$ is a sequence of i.i.d. $Uniform(0, 1)$ random variables. The lower record sequence, $\tilde{R}_n$, admits a slightly simpler representation

$$\tilde{R}_n = \left[\prod_{j=0}^{n} U_j\right]^{\frac{1}{\alpha}} \tag{1.32}$$

Pareto Records

Suppose that the observations are taken from a *cdf*

$$F(x) = \begin{cases} 0, & x < \sigma; \\ 1 - \left(\frac{x}{\sigma}\right)^{-}\alpha, & x \geq \sigma. \end{cases} \tag{1.33}$$

where $\alpha > 0$ and $\sigma > 0$. If (1.33) holds, we say the population *rv*, X, is Pareto(σ, α). Observe that for such a random variable we have

$$X \stackrel{\mathrm{d}}{=} \sigma U^{-\frac{1}{\alpha}} \tag{1.34}$$

where U is $Uniform(0, 1)$. Consequently for $n = 0, 1, 2, ...$

$$R_n \stackrel{\mathrm{d}}{=} \sigma \left(\prod_{j=0}^{n} U_j\right)^{-\frac{1}{\alpha}} \tag{1.35}$$

where the U_j's are i.i.d. $Uniform(0, 1)$ random variables.

1.4.3 Record from Kies Distribution

This section introduces the Kies distribution $K(\lambda, \beta)$ records as well as some of its distributional characteristics. In this section, the lower and upper incomplete gamma functions are required:

$$\int_0^z t^{\alpha-1} e^{-\mu t} dt = \mu^{-\alpha} \gamma(\alpha, \mu z), \tag{1.36}$$

$$\int_z^\infty t^{\alpha-1} e^{-\mu t} dt = \mu^{-\alpha} \Gamma(\alpha, \mu z), \tag{1.37}$$

respectively.
Additionally,

$$\int_z^\infty t^{-\alpha} e^{-t} dt = z^{\frac{-\alpha}{2}} e^{\left(\frac{-\alpha}{2}\right)} \mathbf{W}_{-\frac{\alpha}{2}, \left(\frac{1-\alpha}{2}\right)}(z) \tag{1.38}$$

where $\mathbf{W}_{c_1,c_2}(q)$ is the Whittaker function which is defined, for $|\arg(-q)| < \frac{3\pi}{2}$, as

$$\mathbf{W}_{c_1,c_2}(q) = \frac{\Gamma(-2c_2)}{\Gamma\left(\frac{1}{2} - c_2 - c_1\right)} M_{c_2,c_2}(q) + \frac{\Gamma(2c_2)}{\Gamma\left(\frac{1}{2} + c_2 - c_1\right)} M_{c_2,-c_2}(q), \tag{1.39}$$

in which

$$M_{c_1,c_2}(q) = e^{-\frac{\alpha}{2}} q^{c_2+\frac{1}{2}} \sum_{k=0}^{\infty} \left\{ \frac{\left(\frac{1}{2} - c_1 + c_2\right)_k}{(1+2c_2)_k} \frac{q^k}{k!} \right\} \tag{1.40}$$

The series given in Eq.(1.40) converges for all finite values of q. Also the Pochhammer symbol is defined as follows:

$$(a)_k = a(a+1)(a+2)\ldots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)} = \prod_{i=1}^{k} (a+i-1), \tag{1.41}$$

where $(a)_0 = 1$ and $(1)_k = k!$.

Kies Distribution

One of the modified versions of the Weibull distribution is known as Kies Distribution and was firstly proposed by Kies (1958). The Kies probability model was proposed as an alternative to the extended Weibull models as it provides a more efficient fit to some real-life data sets in comparison to the aforementioned model.
A random variable X is said to follow a two-parameter Kies distribution with shape and scale parameters, β and λ, respectively.

- if its cumulative distribution function *cdf* is given by:

$$F\left(x;\lambda,\beta\right)=1-e^{-\lambda\left(\frac{x}{1-x}\right)^{\beta}} \tag{1.42}$$

- if its probability density function *pdf* is given by:

$$f\left(x;\lambda,\beta\right)=\frac{\beta\lambda x^{\beta-1}}{\left(1-x\right)^{\beta+1}}e^{-\lambda\left(\frac{x}{1-x}\right)^{\beta}} \tag{1.43}$$

- if its hazard rate function is given by:

$$h\left(x;\lambda,\beta\right)=\frac{\beta\lambda x^{\beta-1}}{\left(1-x\right)^{\beta+1}} \tag{1.44}$$

- if its cumulative hazard rate functions is given by:

$$H\left(x;\lambda,\beta\right)=\lambda\left(\frac{x}{1-x}\right)^{\beta} \tag{1.45}$$

1.4.4 Distributional Properties of Records from Kies Distribution

Many properties of records can be obtained in terms of the cumulative hazard function, $H\left(x\right)=-log\left(1-F\left(x\right)\right)$. The pdf of the m^{th} record value is given by:

$$f_m\left(y\right)=\frac{\left[H\left(y\right)\right]^{m-1}}{\Gamma\left(m\right)}f\left(y\right) \tag{1.46}$$

The joint pdf of the m^{th} and s^{th} records is given by:

$$f_{m,s}\left(y,z\right)=\frac{\left[H\left(y\right)\right]^{m-1}}{\Gamma\left(m\right)}\frac{f\left(y\right)}{1-F\left(y\right)}\frac{\left[H\left(z\right)-H\left(y\right)\right]^{s-m-1}}{\Gamma\left(s-m\right)}f\left(z\right) \tag{1.47}$$

which is equivalent to

$$f_{m,s}\left(y,z\right)=\frac{\left[H\left(y\right)\right]^{m-1}}{\Gamma\left(m\right)}h\left(y\right)\frac{\left[H\left(z\right)-H\left(y\right)\right]^{s-m-1}}{\Gamma\left(s-m\right)}f\left(z\right), \tag{1.48}$$

where $-\infty<y<z<\infty$, $h\left(.\right)$ is the hazard rate function. The conditional density of y_s given y_m, where $y_m<y_s$, is given by

$$f\left(y_s|y_m;\lambda,\beta\right)=\frac{\left[H\left(y_s\right)-H\left(y_m\right)\right]^{s-m-1}}{\Gamma\left(s-m\right)}\frac{f\left(y_s|\lambda,\beta\right)}{1-F\left(y_m|\lambda,\beta\right)} \tag{1.49}$$

- Using Eq's (1.43), (1.44) and (1.45), the pdf of the m^{th} record and the joint pdf of the m^{th} and s^{th} records from $K(\lambda, \beta)$ given in Eq's.(1.46) and (1.48), respectively, become:

$$f_m(y) = \frac{\beta\lambda^m}{\Gamma m}\left(\frac{y}{1-y}\right)^{m\beta}\frac{1}{y(1-y)}e^{-\lambda\left(\frac{y}{1-y}\right)^{\beta}} \tag{1.50}$$

$$f_{m,s}(y,z) = \frac{\lambda\beta^2}{\Gamma(m)}\frac{y^{m\beta-1}}{(1-y)^{m\beta+1}}\frac{\left[\left(\frac{z}{1-z}\right)^{\beta}-\left(\frac{y}{1-y}\right)^{\beta}\right]^{s-m-1}}{\Gamma(s-m)}\left(\frac{z^{\beta-1}}{(1-z)^{\beta+1}}\right)e^{-\lambda\left(\frac{z}{1-z}\right)^{\beta}} \tag{1.51}$$

where $0 < y < z < 1$ and $\lambda, \beta > 0$.

- Using Eq's. (1.51) and (1.36) the cdf F_m of the m^{th} record value from Kies distribution is given by:

$$F_m\left(y\right) = \frac{\gamma\left(m, \lambda\left(\frac{y}{1-y}\right)^{\beta}\right)}{\Gamma\left(m\right)}, \qquad m \geq 1 \tag{1.52}$$

where $0 < y < 1$ and $\lambda, \beta \geq 0$

Proposition 1.4.1. [4]

Suppose that the random variable X follows a Kies distribution. Then, one can prove that

$$X \stackrel{\mathrm{d}}{=} \frac{\left(\frac{1}{\lambda}X^*\right)^{\frac{1}{\beta}}}{1+\left(\frac{1}{\lambda}X^*\right)^{\frac{1}{\beta}}} \tag{1.53}$$

where D means converges in distribution and $X^* = -log(1-U)$ where U is $Uniform(0,1)$. It is obvious that X^* follows a standard exponential distribution. Consequently, using the result, (A.4.10), Page(174) of Houchens (1984)[**?**],
the corresponding sequence of records can be described by

$$Y_m \stackrel{\mathrm{d}}{=} \frac{\left(\frac{1}{\lambda}\sum_{i=1}^{m}X_i^*\right)^{\frac{1}{\beta}}}{1+\left(\frac{1}{\lambda}\sum_{i=1}^{m}X_i^*\right)^{\frac{1}{\beta}}} \tag{1.54}$$

where $\{X_i^*\}_{i=1}^{m}$ is a sequence of i.i.d. $\mathcal{E}xp(1)$ random variables.

Proposition 1.4.2. [4] If the random variable X has a Kies distribution, then k^{th} moment

$\mu_m^{(k)} = \mathbb{E}(Y_m^{(k)})$ for the m^{th} record from the Kies distribution is given by

$$\begin{aligned}
\mu_m^{(k)} &= \frac{1}{\Gamma(m)}\sum_{j=0}^{\infty}(-1)^j\frac{(k)_j}{j!}\lambda^{-\left(\frac{k+j}{\beta}\right)}\gamma\left(m+\frac{k+j}{\beta},\lambda\right) \\
&+ \frac{1}{\Gamma(m)}\sum_{j=0}^{\beta(m-1)}(-1)^j\frac{(k)_j}{j!}\lambda^{\frac{j}{\beta}}\Gamma\left(m-\frac{j}{\beta},\lambda\right) \\
&+ \frac{1}{\Gamma(m)}\left[\sum_{j=\beta(m-1)+1}^{\infty}(-1)^j\frac{(k)_j}{j!}\lambda^{\frac{m}{2}+\frac{j}{2\beta}-\frac{1}{2}}e^{\frac{m}{2}-\frac{j}{2\beta}-\frac{1}{2}} * \mathbf{w}_{\frac{m}{2}-\frac{j}{2\beta}-\frac{1}{2},\left(\frac{m}{2}-\frac{j}{2\beta}\right)}(\lambda)\right]
\end{aligned} \tag{1.55}$$

Proof. By using Eq (1.50), the k^{th} moment for the m^{th} record from the Kies distribution is given by:

$$\mathbb{E}(Y_m^k) = \int_0^1 \frac{\beta\lambda^m}{\Gamma(m)}\left(\frac{y_m}{1-y_m}\right)^{m\beta}\frac{y_m^k}{y_m(1-y_m)}e^{-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}}dy_m \tag{1.56}$$

On substituting $\left(\frac{y_m}{1-y_m}\right)^{\beta} = t$ in Eq (1.56), we get

$$\mathbb{E}(Y_m^k) = \frac{\lambda^m}{\Gamma(m)}\int_0^{\infty}\left(\frac{t^{\frac{1}{\beta}}}{1+t^{\frac{1}{\beta}}}\right)^k t^{m-1}e^{-\lambda t}dt \tag{1.57}$$

On splitting the integral and expanding $\left(1+t^{\frac{1}{\beta}}\right)^{-k}$ using Newton's generalization of the binomial theorem, we get the following:

$$\begin{aligned}
\mathbb{E}(Y_m^k) &= \frac{\lambda^m}{\Gamma(m)}\int_0^1\frac{t^{m+\frac{k}{\beta}-1}}{\left(1+t^{\frac{1}{\beta}}\right)^k}e^{-\lambda t}dt + \frac{\lambda^m}{\Gamma(m)}\int_0^{\infty}\frac{t^{m+\frac{k}{\beta}-1}}{t^{\frac{k}{\beta}}\left(t^{-\frac{1}{\beta}}+1\right)^k}e^{-\lambda t}dt \\
&= \frac{\lambda^m}{\Gamma(m)}\sum_{j=0}^{\infty}\frac{(-1)^j(k)_j}{j!}\int_0^1\left(t^{\frac{k+j+m\beta}{\beta}-1}\right)e^{-\lambda t}dt \\
&+ \frac{\lambda^m}{\Gamma(m)}\sum_{j=0}^{\infty}\frac{(-1)^j(k)_j}{j!}\int_0^{\infty}\left(t^{\frac{m\beta-j}{\beta}-1}\right)e^{-\lambda t}dt
\end{aligned}$$

where $(.)_j$ is the Pochhammer symbol given by (1.41). If we put $\mu = \lambda t$, we get

$$\begin{aligned}
\mathbb{E}(Y_m^k) &= \frac{1}{\Gamma(m)}\sum_{j=0}^{\infty}\frac{(-1)^j(k)_j}{j!}\lambda^{-\frac{k+j}{\beta}}\int_0^{\lambda}\left(\mu^{\frac{k+j+m\beta}{\beta}-1}\right)e^{-\mu}d\mu \\
&+ \frac{1}{\Gamma(m)}\sum_{j=0}^{\infty}\frac{(-1)^j(k)_j}{j!}\lambda^{\frac{j}{\beta}}\int_{\lambda}^{\infty}\left(\mu^{\frac{m\beta-j}{\beta}-1}\right)e^{-\mu}d\mu
\end{aligned}$$

since the exponent $\left(m - \frac{j}{\beta}\right)$ in the second integral carries positive and negative values.
Therefore, on splitting the second summation we get the following:

$$\begin{aligned}
\mathbb{E}(Y_m^k) &= \frac{1}{\Gamma(m)} \sum_{j=0}^{\infty} \frac{(-1)^j (k)_j}{j!} \lambda^{-\frac{k+j}{\beta}} \int_0^{\lambda} \left(\mu^{\frac{k+j+m\beta}{\beta}-1}\right) e^{-\mu} d\mu \\
&+ \frac{1}{\Gamma(m)} \sum_{j=0}^{\beta(m-1)} \frac{(-1)^j (k)_j}{j!} \lambda^{\frac{j}{\beta}} \int_{\lambda}^{\infty} \mu^{m-\frac{j}{\beta}-1} e^{-\mu} d\mu \\
&+ \frac{1}{\Gamma(m)} \sum_{j=\beta(m-1)+1}^{\infty} \frac{(-1)^j (k)_j}{j!} \lambda^{\frac{j}{\beta}} \int_{\lambda}^{\infty} \mu^{-\left(1+\frac{j}{\beta}-m\right)} e^{-\mu} d\mu
\end{aligned}$$

which leads to (1.55) in the light of (1.36) (1.37) (1.38).

□

The expected value of the m^{th} record $[\mathbb{E}(Y_m)]$ is the first moment which is denote $\mu_m^{(1)}$

In addition, the variance of the m^{th} record is

$$var(y_m) = \mu_m^{(2)} - [\mu_m^{(1)}]^2$$

For illustrative purposes, $\mathbb{E}(Ym)$ and variance of some records of Kies distribution, namely $3^{rd}, 5^{th}, 7^{th}$ and 10^{th}, are computed and summarized in the next table assuming different values of λ and β.

Table 1.2: Expected values and variances of records from $K(\lambda, \beta)$ with $\lambda = 1, 2$ and $\beta = 0.75, 2$

m	$\lambda = 1$				$\lambda = 2$			
	$\beta = 0.75$		$\beta = 2$		$\beta = 0.75$		$\beta = 2$	
	$\mathbb{E}(y_m)$	Variance	$\mathbb{E}(y_m)$	Variance	$\mathbb{E}(y_m)$	Variance	$\mathbb{E}(y_m)$	Variance
3	0.74800	0.02350	0.61100	0.00549	0.57000	0.03240	0.52800	0.00585
5	0.86600	0.00625	0.67800	0.00267	0.73200	0.01470	0.59900	0.00317
7	0.95500	0.00201	0.71700	0.00161	0.81600	0.00659	0.64200	0.00203
10	0.97800	0.00021	0.75400	0.00092	0.88200	0.00228	0.68500	0.00124

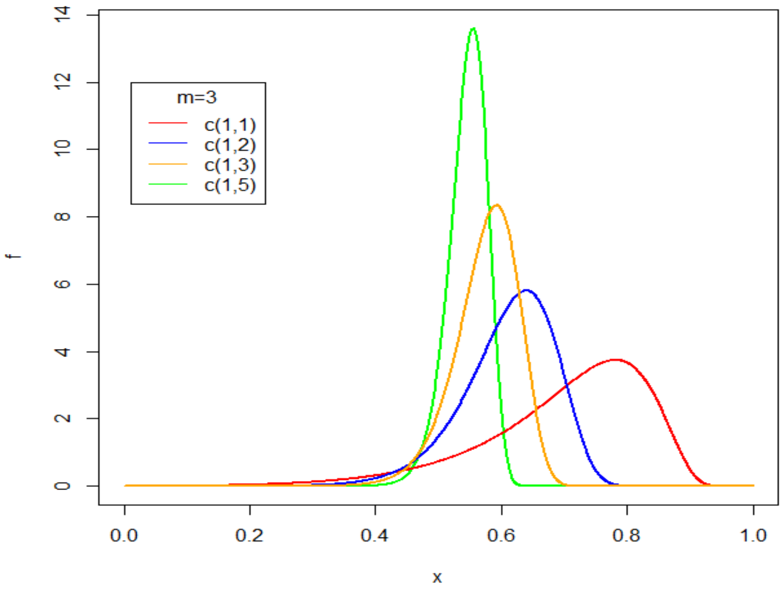

Figure 1.1: pdf of the m^{th} record plots with $\lambda = 1$ and $\beta = \{1, 2, 3, 5\}$ for Y_m , $m = 3$

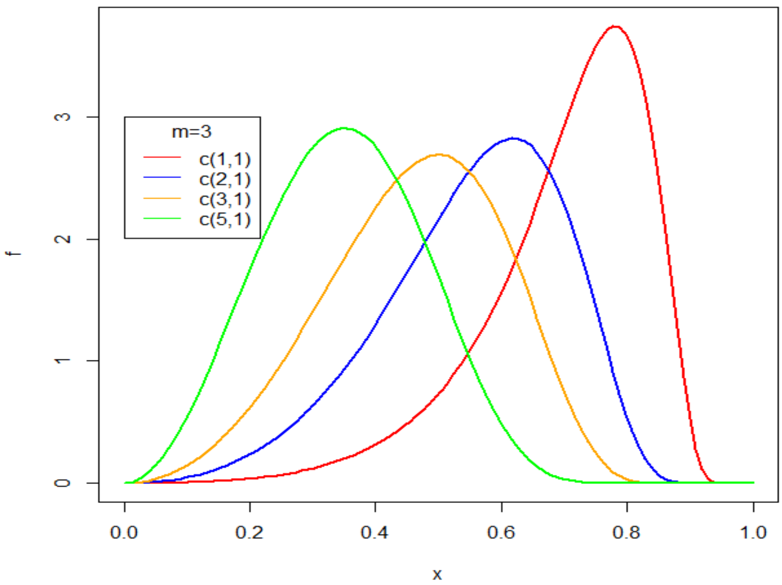

Figure 1.2: pdf of the m^{th} record plots with $\lambda = \{1, 2, 3, 5\}$ and $\beta = 1$ for Y_m , $m = 3$

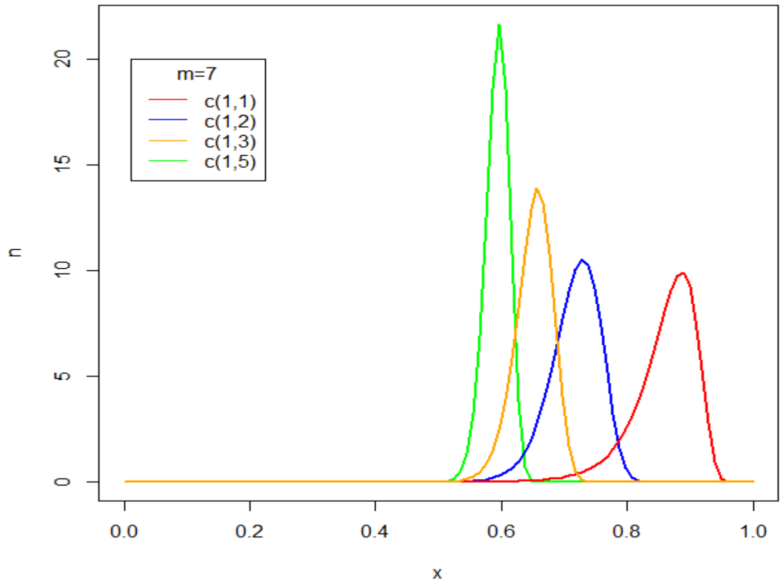

Figure 1.3: pdf of the m^{th} record plots with $\lambda = 1$ and $\beta = \{1, 2, 3, 5\}$ for Y_m , $m = 7$

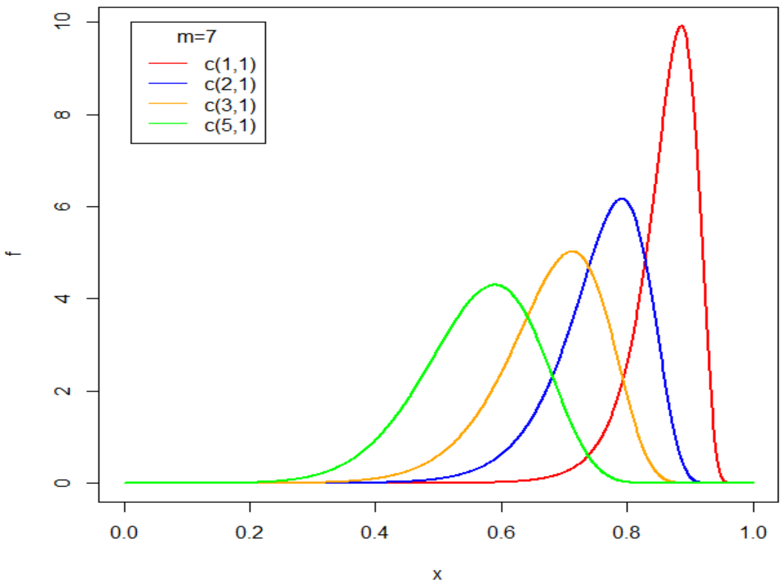

Figure 1.4: pdf of the m^{th} record plots with $\lambda = \{1, 2, 3, 5\}$ and $\beta = 1$ for Y_m , $m = 7$

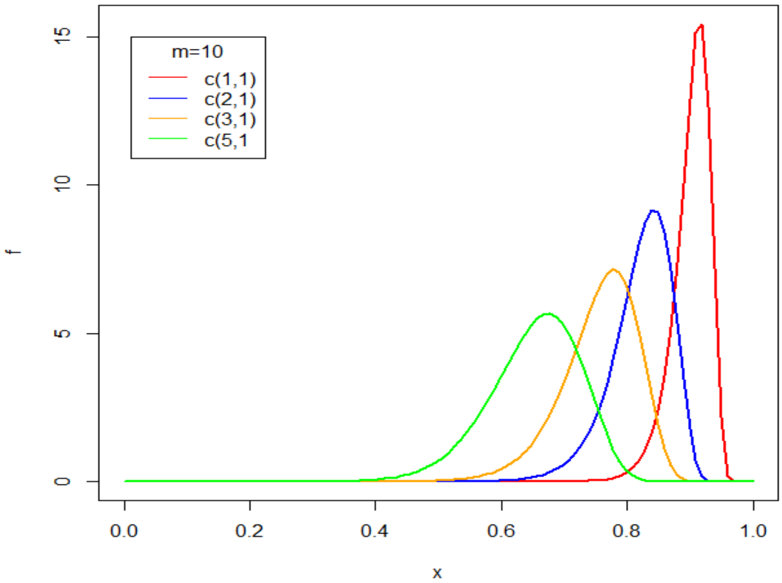

Figure 1.5: pdf of the m^{th} record plots with $\lambda = \{1, 2, 3, 5\}$ and $\beta = 1$ for Y_m , $m = 10$

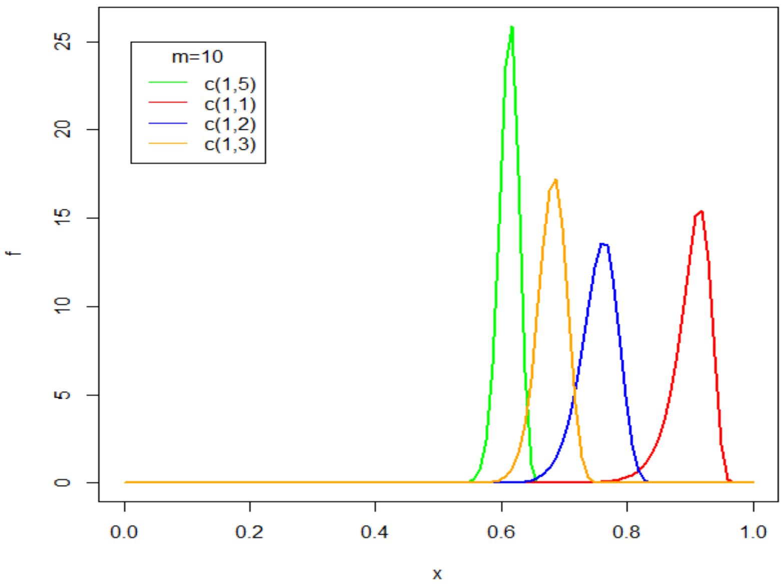

Figure 1.6: pdf of the m^{th} record plots with $\lambda = 1$ and $\beta = \{1, 2, 3, 5\}$ for Y_m , $m = 10$

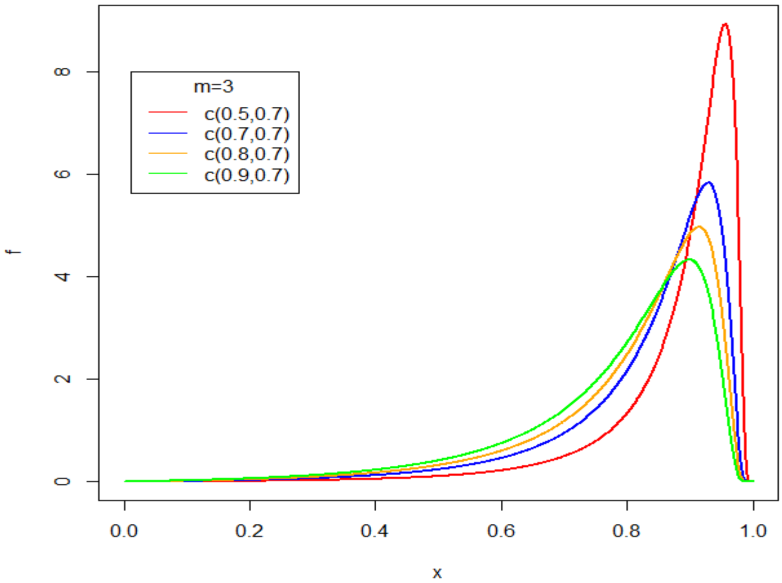

Figure 1.7: pdf of the m^{th} record plots with $\lambda = \{0.5, 0.7, 0.8, 0.9\}$ and $\beta = 0.7$ for Y_m , $m = 3$

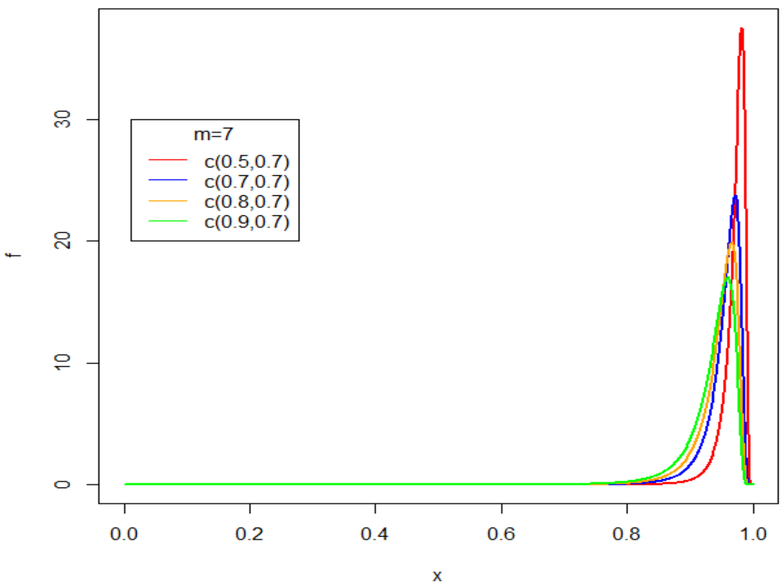

Figure 1.8: pdf of the m^{th} record plots with $\lambda = \{0.5, 0.7, 0.8, 0.8\}$ and $\beta = 0.7$ for Y_m , $m = 7$

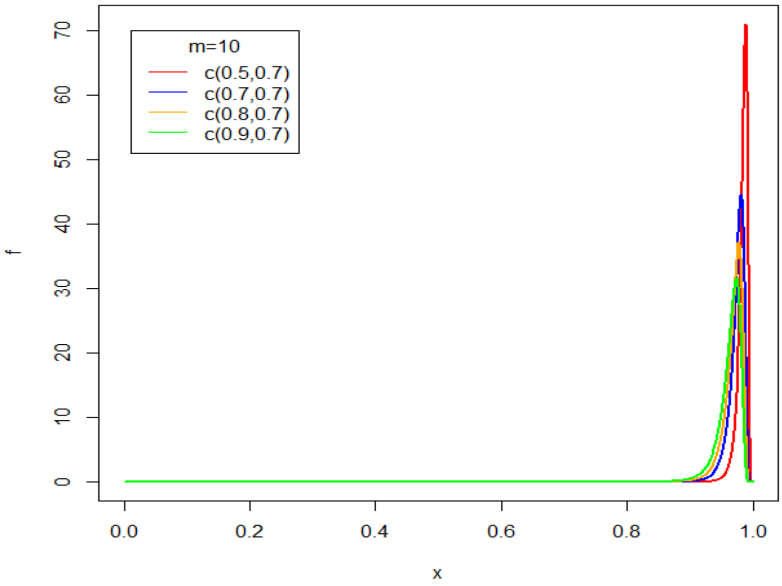

Figure 1.9: pdf of the m^{th} record plots with $\lambda = \{0.5, 0.7, 0.8, 0.8\}$ and $\beta = 0.7$ for Y_m , $m = 10$

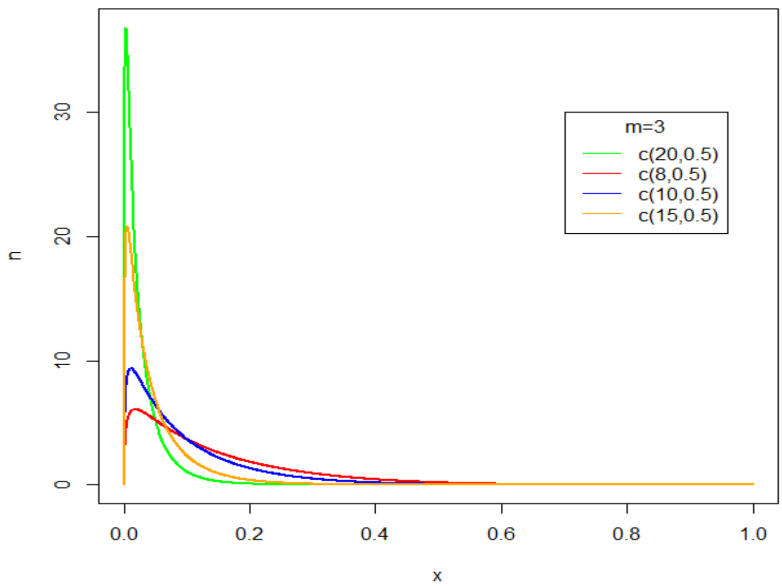

Figure 1.10: pdf of the m^{th} record plots with $\lambda = \{8, 10, 15, 20\}$ and $\beta = 0.5$ for Y_m , $m = 3$

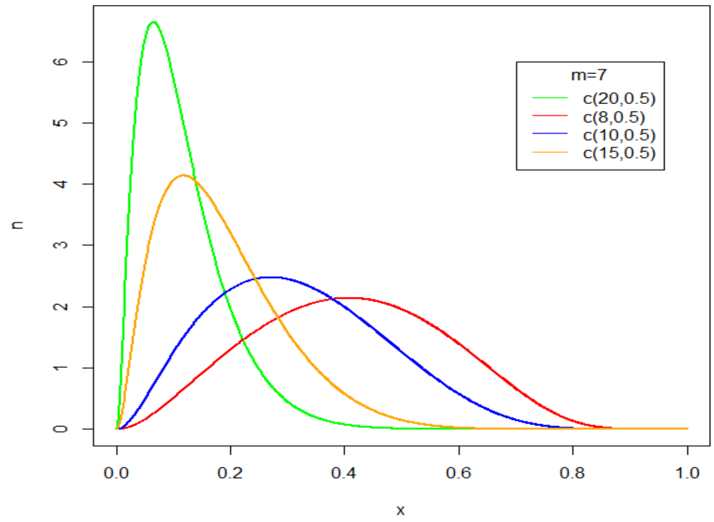

Figure 1.11: pdf of the m^{th} record plots with $\lambda = \{8, 10, 15, 20\}$ and $\beta = 0.5$ for Y_m , $m = 7$

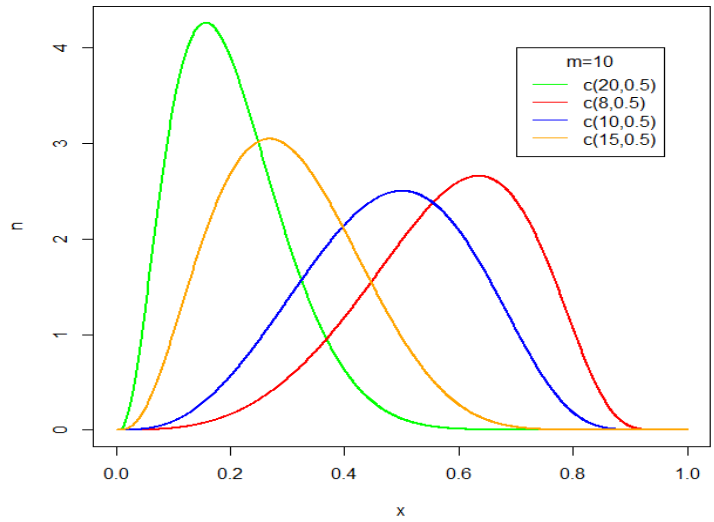

Figure 1.12: pdf of the m^{th} record plots with $\lambda = \{8, 10, 15, 20\}$ and $\beta = 0.5$ for Y_m , $m = 10$

Figures from Figure 1.1 to Figure 1.12 show the distribution curves of the m^{th} record from $K(\lambda, \beta)$ distribution for different values of λ, β at different values of m. From these graphs, one can notice the effects of the parameters λ, β on the distribution curves. For instance, it is also obvious that the distributions of the random variable y_m are unimodal. One also can notice that the distribution is almost symmetric, skewed to the left or to the right under specific values of the parameters.

Chapter 2

Classical and Bayesian Estimation Based on Records from Kies Distribution

The Kies distribution, indicated by $K(\lambda, \beta)$, has a restricted range, making it an ideal model for modeling actual data sets. Recently, it has received the attention of different authors in the literature. In this chapter we will address the attention to the estimation problem of the two unknown parameters of the Kies distribution, $K(\lambda, \beta)$, based on upper record data. Classical and Bayesian approaches are considered. The maximum likelihood estimation along with the associated asymptotic and bootstrap-p confidence intervals are obtained. Furthermore, Bayes estimates based on the square error and the linear exponential loss functions are computed using gamma priors for the unknown parameters.

2.1 Classical Estimation

2.1.1 Maximum Likelihood Estimation

Let $\underline{y} = \{y_1, y_2, ..., y_m\}$ be the first m upper record values arising from a sequence of iid random variables from a Kies distribution $K(\lambda, \beta)$ with *cdf*, *pdf* and hazard rate given in Eq's.(1.42), (1.43) and (1.44), respectively.
The likelihood function of the $\underline{y}$ is given by, (1.26)

$$\begin{aligned} L\left(\underline{y};\lambda,\beta\right) &= f\left(y_m;\lambda,\beta\right)\prod_{i=1}^{m-1} h\left(y_i;\lambda,\beta\right) \\ &= \beta^m\lambda^m e^{-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}}\prod_{i=1}^{m}\frac{y_i^{\beta-1}}{\left(1-y_i\right)^{\beta+1}} \end{aligned} \quad (2.1)$$

The log-likelihood function is:

$$\begin{aligned} logL\left(\underline{y};\lambda,\beta\right) &= log\left(\beta^m\lambda^m e^{-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}}\prod_{i=1}^{m}\frac{y_i^{\beta-1}}{\left(1-y_i\right)^{\beta+1}}\right) \\ &= log\left(\beta^m\right)+log\left(\lambda^m\right)+log\left(e^{-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}}\right)+\sum_{i=1}^{m} log\left(\frac{y_i^{\beta-1}}{\left(1-y_i\right)^{\beta+1}}\right) \\ &= mlog\beta+mlog\lambda-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}+\sum_{i=1}^{m}\left(log\left(y_i^{\beta-1}\right)-log\left(\left(1-y_i\right)^{\beta+1}\right)\right) \\ &= mlog\lambda\beta-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}+\left(\beta-1\right)\sum_{i=1}^{m} log\left(y_i\right)-\left(\beta+1\right)\sum_{i=1}^{m} log\left(1-y_i\right) \\ &= mlog\beta+mlog\lambda-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}+\beta\sum_{i=1}^{m} log\left(y_i\right)-\sum_{i=1}^{m} log\left(y_i\right) \\ &- \beta\sum_{i=1}^{m} log\left(1-y_i\right)-\sum_{i=1}^{m} log\left(1-y_i\right) \\ &= mlog\beta+mlog\lambda-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta}+\beta\sum_{i=1}^{m} log\left(\frac{y_i}{1-y_i}\right)-\sum_{i=1}^{m} log\left(y_i-y_i^2\right) \end{aligned}$$

where $0 < y_1 < y_2 < ... < y_m < 1$, $\beta > 0$ and $\lambda > 0$.

$$\frac{\partial logL\left(\underline{y};\lambda,\beta\right)}{\partial\lambda} = \frac{m}{\lambda}-\left(\frac{y_m}{1-y_m}\right)^{\beta} \quad (2.2)$$

$$\frac{\partial logL\left(\underline{y};\lambda,\beta\right)}{\partial\beta} = \frac{m}{\beta}-\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right)+\sum_{i=1}^{m} log\left(\frac{y_i}{1-y_i}\right) \quad (2.3)$$

We set the partial derivatives to zero we obtian the $MLEs$:

$$\hat{\lambda} = \frac{m}{\left(\frac{y_m}{1-y_m}\right)^{\hat{\beta}}} \quad (2.4)$$

$$\hat{\beta} = \frac{m}{\sum_{i=1}^{m-1} log\left(\frac{y_m(1-y_i)}{(1-y_m)y_i}\right)} \tag{2.5}$$

The following proposition shows existence and uniqueness of the $MLEs$ of λ and β see, Xia et al.(2009)[31].

Proposition 2.1.1. [4] The log-likelihood function $logL(\underline{y}|\lambda,\beta)$ is unimodal function of λ and β.

Proof. Note that $logL(\underline{y}|\lambda,\beta)$ is continuous in λ and β , and the Hessian matrix of the log-likelihood function is negative definite, which is given by:

$$H = \begin{pmatrix} \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda^2} & \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda\partial\beta} \\ \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta\partial\lambda} & \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta^2} \end{pmatrix}$$

where

$$\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda^2} = -\frac{m}{\lambda^2} \tag{2.6}$$

$$\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda\partial\beta} = \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta\partial\lambda} = -\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right) \tag{2.7}$$

$$\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta^2} = -\left(\frac{m+\lambda\beta^2\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)}{\beta^2}\right) \tag{2.8}$$

Thus, $logL(\underline{y}|\lambda,\beta)$ is unimodal of λ and β. This immediately proves existence and uniqueness of the $MLEs$ of the unknown parameters λ and β of the Kies distribution. In order to prove that the Hessian matrix is negative definite for λ and β, sufficient conditions are:

1. Determine of the upper left 1 -by- 1 corner $\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda^2}$ of H need to be negative;
2. Determinant of H, $det(H)$, needs to be positive for λ and β.

Based on (2.11), condition one is satisfied. In order to prove the second condition,

$$det(H) = \frac{m}{\lambda^2}\left(\frac{m}{\beta^2} + \lambda\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)\right) - \left(\lambda\beta^2\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right)\right)^2 \tag{2.9}$$

if we replace λ in (2.9) by its corresponding MLE, $\hat{\lambda}$, we get $\left(\frac{\left(\frac{y_m}{1-y_m}\right)^{\beta}}{\beta}\right)^2 > 0$. $\square$

2.1.2 Asymptotic Confidence Interval

Since it is not easy to derive the exact distribution of the $MLEs$ in Eq's (2.4) and (2.5), we cannot obtain the exact confidence intervals (CIs) for the parameters λ and β Consequently, asymptotic CIs (ACIs) of the parameters are derived using the asymptotic distribution of $MLEs$. To this end, we need to find the variance-covariance matrix of the $MLEs$. The observed information matrix of $\underline{\theta} = (\lambda, \beta)$ is given by:

$$I(\underline{\theta}) = -\begin{pmatrix} \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda^2} & \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda\partial\beta} \\ \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta\partial\lambda} & \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta^2} \end{pmatrix} \tag{2.10}$$

where

$$\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda^2} = -\frac{m}{\lambda^2} \tag{2.11}$$

$$\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda\partial\beta} = \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta\partial\lambda} = -\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right) \tag{2.12}$$

$$\frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta^2} = -\left(\frac{m+\lambda\beta^2\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)}{\beta^2}\right) \tag{2.13}$$

Hence, the inverse of the observed information matrix is given by:

$$I^{-1}(\theta) = -\begin{pmatrix} \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda^2} & \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\lambda\partial\beta} \\ \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta\partial\lambda} & \frac{\partial^2 logL(\underline{y}|\lambda,\beta)}{\partial\beta^2} \end{pmatrix}^{-1} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \tag{2.14}$$

where

$$\begin{aligned}
V_{11} &= \frac{m+\lambda\beta^2\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)}{\frac{m}{\lambda^2}\left(m+\lambda\beta^2\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)\right)-\beta^2\left(\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right)\right)^2}\\
V_{12}=V_{21} &= \frac{-\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right)}{\frac{m}{\lambda^2\beta^2}\left(m+\lambda\beta^2\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)\right)-\left(\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right)\right)^2}\\
V_{22} &= \frac{1}{\frac{m}{\beta^2}+\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta} log^2\left(\frac{y_m}{1-y_m}\right)-\frac{\left(\lambda\left(\frac{y_m}{1-y_m}\right)^{\beta} log\left(\frac{y_m}{1-y_m}\right)\right)^2}{m}}
\end{aligned}$$

The asymptotic joint distribution of the $MLEs$ $\hat{\lambda}$ and $\hat{\beta}$ is approximated by a bivariate normal distribution, and is given by:

$$\begin{pmatrix}\hat{\lambda}\\ \hat{\beta}\end{pmatrix} \overset{\text{D}}{\sim} N\left[\begin{pmatrix}\lambda\\ \beta\end{pmatrix} \quad , \quad \begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix}\right] \tag{2.15}$$

Hence, by replacing λ and β by their $MLEs$, we get an estimate of $I^{-1}(\theta)$, which is called the approximate variance-covariance matrix for the $MLEs$ $\hat{\theta}=(\hat{\lambda},\hat{\beta})$ as follows:

$$I^{-1}(\hat{\theta}) = \begin{pmatrix} \frac{m}{\left((A_m)^{\hat{\beta}}\right)^2}\left(1+\hat{\beta}^2 log^2\left(A_m\right)\right) & \frac{-\hat{\beta}^2 log\left(A_m\right)}{(A_m)^{\hat{\beta}}}\\ \frac{-\hat{\beta}^2 log\left(A_m\right)}{(A_m)^{\hat{\beta}}} & \frac{\hat{\beta}^2}{m}\end{pmatrix} = \begin{pmatrix} var(\hat{\lambda}) & cov(\hat{\lambda},\hat{\beta})\\ cov(\hat{\beta},\hat{\lambda}) & var(\hat{\beta})\end{pmatrix} \tag{2.16}$$

Where $A_m = \dfrac{y_m}{1-y_m}$ Consequently, asymptotic $100(1-\alpha)$ % (CIs) for the parameters λ and β are, respectively, given by:

$$(L_\lambda, U_\lambda) = \left(\hat{\lambda}-z_{1-\frac{\alpha}{2}}\sqrt{var(\hat{\lambda})}, \hat{\lambda}+z_{1-\frac{\alpha}{2}}\sqrt{var(\hat{\lambda})}\right) \tag{2.17}$$

and

$$(L_\beta, U_\beta) = \left(\hat{\beta}-z_{1-\frac{\alpha}{2}}\sqrt{var(\hat{\beta})}, \hat{\beta}+z_{1-\frac{\alpha}{2}}\sqrt{var(\hat{\beta})}\right) \tag{2.18}$$

where z_α is $100\alpha^{th}$ percentile of the standard normal distribution. However, some cases provide negative lower bounds of the asymptotic CI while the parameters λ and β are positive. In order to avoid this, we propose using delta method and log-transformation for the parameters in order to construct a modified asymptotic confidence intervals for λ and β, following the lines of Ren and Gui (2020)[29]. The asymptotic distribution of $log(\hat{\theta}_j)$, $j = 1, 2$, is given by:

$$\left(log(\hat{\theta}_j) - log(\theta_j)\right) \overset{D}{\sim} N\left(0, var(log(\hat{\theta}_j))\right) \tag{2.19}$$

where $var\left(log(\hat{\theta}_j)\right) = \dfrac{var(\hat{\theta}_j)}{\hat{\theta}_j^2} = \dfrac{I^{-1}(\hat{\theta}_j)}{\hat{\theta}_j^2}$, hence

$$\frac{\hat{\theta}_j}{\sqrt{I^{-1}(\hat{\theta}_j)}}\left(log(\hat{\theta}_j) - log(\theta_j)\right) \overset{D}{\sim} N(0,1) \tag{2.20}$$

Therefore, modified asymptotic $(1-\alpha)100\%(0 < \alpha < 1)$ CIs for λ and β can be easily obtained, respectively, as follows:

$$\left(\frac{\hat{\lambda}}{e^{\frac{z_{1-\frac{\alpha}{2}}}{\hat{\lambda}}\sqrt{var(\hat{\lambda})}}}, \hat{\lambda}e^{z_{1-\frac{\alpha}{2}}\sqrt{var(\hat{\lambda})}}\right) \; and \; \left(\frac{\hat{\beta}}{e^{\frac{z_{1-\frac{\alpha}{2}}}{\hat{\beta}}\sqrt{var(\hat{\beta})}}}, \hat{\beta}e^{z_{1-\frac{\alpha}{2}}\sqrt{var(\hat{\beta})}}\right) \tag{2.21}$$

2.1.3 Bootstrap Method

In this section, the percentile Bootstrap method, also known as Boot-p, is introduced to create estimated confidence intervals (CIs) for λ and β using the following methodology because asymptotic CIs results do not perform well for a small sample size. For an illustration, see [1, 12].

step (1) From the records $y_1, y_2, ..., y_m$, compute the $MLEs$.$\hat{\lambda}_{ML}$ and $\hat{\beta}_{ML}$.

step (2) Using $\hat{\lambda}_{ML}$ and $\hat{\beta}_{ML}$ that are obtained in Step(1), generate a randomsample of records from $K(\lambda, \beta)$, called a bootstrap sample.

step (3) Based on the Bootstrap sample that is obtained in Step(2), compute the corresponding $MLEs$ $\hat{\lambda}^*$ and $\hat{\beta}^*$ of λ and β, respectively

step (4) Repeat Steps(2) and (3) B-times to obtain $\left\{\hat{\lambda}_1^*, \hat{\lambda}_2^*, ..., \hat{\lambda}_B^*\right\}$ and $\left\{\hat{\beta}_1^*, \hat{\beta}_2^*, ..., \hat{\beta}_B^*\right\}$.

step (5) Arrange $\left\{\hat{\lambda}_1^*, \hat{\lambda}_2^*, ..., \hat{\lambda}_B^*\right\}$ and $\left\{\hat{\beta}_1^*, \hat{\beta}_2^*, ..., \hat{\beta}_B^*\right\}$ in ascending order and obtain $\left\{\hat{\lambda}_{(1)}^*, \hat{\lambda}_{(2)}^*, ..., \hat{\lambda}_{(B)}^*\right\}$ and $\left\{\hat{\beta}_{(1)}^*, \hat{\beta}_{(2)}^*, ..., \hat{\beta}_{(B)}^*\right\}$.

step (6) The approximate $100(1-\alpha)\%$ Boot-p CIs for λ and β are given by $\left(\hat{\lambda}^*_{(B_{\frac{\alpha}{2}})}, \hat{\lambda}^*_{(B_{(1-\frac{\alpha}{2})})}\right)$ and $\left(\hat{\beta}^*_{(B_{\frac{\alpha}{2}})}, \hat{\beta}^*_{(B_{(1-\frac{\alpha}{2})})}\right)$, respectively.

2.2 Bayesian Estimation

Bayesian approach is one of the most popular techniques for making inferences about unknown parameters of a probability distribution. In Bayesian approach, the unknown parameters are assumed to be random variables with distributions called prior distribution. The second factor affecting Bayesian estimation is the loss function, which represents the losses associated with errors committed while estimating the parameters. We confine our interest to two kinds of loss functions, the symmetric square error (SE) and asymmetric linear exponential ($LINEX$) loss functions of the parameter θ_i and an estimate $\hat{\theta}_i$ which are given, respectively, by:

$$L_{SE}\left(\hat{\theta},\theta\right) = \left(\hat{\theta}-\theta\right)^2 \tag{2.22}$$

and

$$L_{LINEX}\left(\hat{\theta},\theta\right) = b\left[e^{\nu\left(\hat{\theta}-\theta\right)} - \nu\left(\hat{\theta}-\theta\right) - 1\right] \tag{2.23}$$

where $b > 0$ is the scale of the loss function. In our study, we assume $b = 1$. The parameter $v \neq 0$ indicates the shape parameter of the loss function. The $LINEX$ loss function is affected by ν, the sign of ν indicates the direction of the asymmetry, and the magnitude of ν indicates the degree of the asymmetry. It is known that if $\nu > 0$ then overestimation is considered to be more serious than underestimation, while if $\nu < 0$ then the reverse situation, while when ν is close to zero, the $LINEX$ loss function is almost symmetric and is approximately equal to the SE loss function. Thus for small values of ν, estimation results obtained by both $LINEX$ and SE are close. For more details about the $LINEX$ loss function, readers may refer to Zellner, A. (1986) [32].

A natural choice of priors for λ and β would be to assume that the two quantities are independent with gamma distributions; namely $G(a_1, b_1)$ and $G(a_2, b_2)$, respectively, where hyper-parameters a_1, a_2, b_1 and b_2 are nonnegative numbers chosen to reflect prior knowledge about the parameters λ and β.

The joint prior distribution of λ and β is obtained as follows:

$$g(\lambda,\beta) \propto \lambda^{a_1-1}e^{-b_1\lambda}\beta^{a_2-1}e^{-b_2\beta} \tag{2.24}$$

Using the upper record $\underline{y} = y_1, y_2, ..., y_m$, the joint posterior distribution of λ and β is obtained as follows:

$$\pi(\lambda, \beta|\underline{y}) \propto L(\underline{y}|\lambda, \beta)g(\lambda, \beta) \tag{2.25}$$

where $L(\underline{y}|\lambda, \beta)$ is the likelihood function given in Eq (2.1) and $g(\lambda, \beta)$ is the joint prior density given in Eq (2.24) By substituting Eq's (2.24) and (2.1) in Eq (2.25), the joint posterior density of λ and β is given by:

$$\pi(\lambda, \beta|\underline{y}) \propto \lambda^{m+a_1-1}\beta^{m+a_2-1}e^{-\beta b_2}e^{-\lambda\left(b-1+A_m^{\beta}\right)}\prod_{i=1}^{m}\left(\frac{y_i}{1-y_i}\right)^{\beta} \tag{2.26}$$

It can be seen that the joint posterior distribution in Eq (2.26) can be represented as follows:

$$\pi(\lambda, \beta|\underline{y}) \propto \pi_1(\beta|\underline{y})\pi_2(\lambda|\beta, \underline{y}) \tag{2.27}$$

where

$$\pi_1(\beta|\underline{y}) \propto \frac{\beta^{m+a_2-1}e^{-\beta b_2}\prod_{i=1}^{m}\left(\frac{y_i}{1-y_i}\right)^{\beta}}{\left(b_1 + A_m^{\beta}\right)^{m+a_1}} \tag{2.28}$$

and $\pi_2(\lambda|\beta, \underline{y})$ is a gamma density with shape and scale parameters $m + a_1$ and $\left[b_1 + A_m^{\beta}\right]^{-1}$, respectively.

Lemma 2.2.1. [4] The conditional distribution of β given the observed records, $\pi_1(\beta|\underline{y})$, is log concave.

Proof. The log likelihood of conditional distribution of β given the observed records, Eq (2.28), is given by:

$$log\pi_1(\beta|\underline{y}) \propto -(m+a_1)log\left(b_1 + A_m^{\beta}\right) + (m+a_2-1)log(\beta) - \beta\left(b_2 - \sum_{i=1}^{m} log\left(\frac{y_i}{1-y_i}\right)\right) \tag{2.29}$$

By differentiating $log\pi_1(\beta|\underline{y})$ twice with respect to β, we get:

$$\frac{\partial\pi_1(\beta|\underline{y})}{\partial\beta} = -(m+a_1)\frac{A_m^{\beta}logA_m}{b_1 + A_m^{\beta}} + \frac{m+a_2-1}{\beta} - b_2 + \sum_{i=1}^{m} log\left(\frac{y_i}{1-y_i}\right) \tag{2.30}$$

$$\frac{\partial^2 \pi_1(\beta|\underline{y})}{\partial \beta^2} \quad = \quad -(m+a_1)\left(\frac{A_m^\beta\left(logA_m\right)^2\left(b_1+A_m^\beta\right)-\left(A_m^\beta logA_m\right)^2}{\left(b_1+A_m^\beta\right)^2}\right)-\frac{m+a_2-1}{\beta^2} \tag{2.31}$$

Since $\frac{\partial^2 \pi_1(\beta|\underline{y})}{\partial \beta^2} < 0$, this follows that $\pi_1(\beta|\underline{y})$ is log-concave density. □

Subsequently, the Bayes estimate of any function of λ and β, say $\eta(\lambda,\beta)$, under $LINEX$ and $LINEX$ loss functions separately are given by:

$$\hat{\theta}_{BS} = \frac{\int_0^\infty \int_0^\infty \eta(\lambda,\beta)\pi_1(\beta|\underline{y})\pi_2(\lambda|\beta,\underline{y})d\beta d\lambda}{\int_0^\infty \int_0^\infty \pi_1(\beta|\underline{y})\pi_2(\lambda|\beta,\underline{y})d\beta d\lambda} \tag{2.32}$$

$$\hat{\theta}_{BL} = -\frac{1}{\nu}log\left(\frac{\int_0^\infty \int_0^\infty e^{-\nu\eta(\lambda,\beta)}\pi_1(\beta|\underline{y})\pi_2(\lambda|\beta,\underline{y})d\beta d\lambda}{\int_0^\infty \int_0^\infty \pi_1(\beta|\underline{y})\pi_2(\lambda|\beta,\underline{y})d\beta d\lambda}\right) \tag{2.33}$$

respectively.
Unfortunately, Bayes estimates in Eq's (2.32) and (2.33) cannot be derived in explicit forms. Therefore, the importance sampling technique was proposed by Chen and Shao (1999)[10] to approximate the Bayes estimates. Similar methods were employed, for instance, by Pradhan and Kundu (2009)[21], and Bayoud (2016)[8].

It can be easily seen that the marginal posterior of β in Eq (2.28) can be rewritten as follows:

$$\pi_1(\beta|\underline{y}) \propto g_1(\beta|\underline{y})g_2(\beta) \tag{2.34}$$

where $g_1(\beta|\underline{y})$ is a gamma density with shape and scale parameters $(m+a_2)$ and $\frac{1}{b_2}$, respectively, and

$$g_2(\beta) = \frac{\prod_{i=1}^{m}\left(\frac{y_i}{1-y_i}\right)^\beta}{\left(b_1+A_m^\beta\right)^{m+a_1}} \tag{2.35}$$

In order to compute the approximate Bayes estimates for the parameters β and λ, we now suggest the following approach, which is similar to that proposed by Kundu and Pradhan (2009)[21]. Let $\underline{y} = y_1, y_2, ..., y_m$ be a set of m upper records and let a_i and b_i, $(i = 1, 2)$ be pre-assumed hyper-parameters chosen based on prior information about the underlying parameters β and λ.

Step (1) Generate a random sample of size M from a gamma prior distribution with *pdf* $g_1(\beta|\underline{y})$, say $\beta_1, \beta_2, ..., \beta_M$;

Step (2) For each β_j , generate λ_j from the gamma density function $\pi_2(\lambda|\beta_j, \underline{y})$, say $\lambda_1, \lambda_2, ..., \lambda_M$;

Step (3) Compute $g_2(\beta_i)$, for $j = 1, 2, ..., M$;

Step (4) Under the SE loss function, the approximate Bayes estimate of $\eta(\lambda, \beta)$ can be obtained using the importance sampling technique as:

$$\hat{\eta}_{BS}(\lambda, \beta) = \frac{\sum_{j=1}^{M} \eta(\lambda_j, \beta_j) g_2(\beta_j)}{\sum_{j=1}^{M} g_2(\beta_j)} \tag{2.36}$$

Hence, $\hat{\beta}_{BS} = \frac{\sum_{j=1}^{M} \beta_j g_2(\beta_j)}{\sum_{j=1}^{M} g_2(\beta_j)}$ and $\hat{\lambda}_{BS} = \frac{\sum_{j=1}^{M} \lambda_j g_2(\beta_j)}{\sum_{j=1}^{M} g_2(\beta_j)}$.

Step (5) Under the $LINEX$ function, the approximate Bayes estimate of $\eta(\lambda, \beta)$ can be obtained using the importance sampling technique as:

$$\hat{\theta}_{BL} = \hat{\eta}(\lambda, \beta) - \frac{1}{\nu} log \frac{\sum_{j=1}^{M} e^{-\nu\eta(\lambda_j, \beta_j)} g_2(\beta_j)}{\sum_{j=1}^{M} g_2(\beta_j)} \tag{2.37}$$

Hence, $\hat{\beta}_{BL} = -\frac{1}{\nu} log \frac{\sum_{j=1}^{M} e^{-\nu\beta_j} g_2(\beta_j)}{\sum_{j=1}^{M} g_2(\beta_j)}$ and $\hat{\lambda}_{BL} = -\frac{1}{\nu} log \frac{\sum_{j=1}^{M} e^{-\nu\lambda_j} g_2(\beta_j)}{\sum_{j=1}^{M} g_2(\beta_j)}$.

2.3 Comparison between obtained estimators

In this section we will illustrate the performance of the so obtained estimators, to end this Al-Olaimat et al. (2021) [4] have conducted a simulation study to compare between the classical, bayesian and confidence interval estimation methods of the unknown parameters λ and β. Simulation is performed under the assumption of two sets of parameter values ($\lambda = 1$,$\beta = 2$), ($\lambda = 2$,$\beta = 1$) a given number m of upper records are generated from $K(\lambda, \beta)$. The MLEs and the approximate Bayes estimates are computed using the

importance sampling procedure. Bayes estimates are computed under the SE and LINEX loss functions assuming the following priors which are assumed based on the values of the true parameters under study: Prior 0: $a_1 = 0$, $b_1 = 0$, $a_2 = 0$, $b_2 = 0$.
For $\lambda = 1$, $\beta = 2$:
Prior 1: $a_1 = 2$, $b_1 = 2$, $a_2 = 16$, $b_2 = 8$ and prior 2: $a_1 = 4$, $b_1 = 4$, $a_2 = 8$, $b_2 = 4$.
For $\lambda = 2$, $\beta = 1$:
Prior 3: $a_1 = 4$, $b_1 = 2$, $a_2 = 8$, $b_2 = 8$ and prior 4: $a_1 = 8$, $b_1 = 4$, $a_2 = 16$, $b_2 = 16$.
These priors are assumed so as λ has the same mean but with different variances, and β has the same mean but different variances. The main purpose of this is to reflect the sensitivity of this inferences to the choice of the hyper-parameters as they said in this study. The shape parameter of LINEX loss function ν is assumed to equal $\{-0.01, 0.5, 2\}$, separately.
Simulations are performed with $M = 1000$ iterations used Mathematica 9. The mean squared errors (MSEs) of the proposed MLEs and Bayes estimates are computed. The point estimation results are reported in tables Tables 2.1 and 2.2 and for the performance of the proposed classical CIs it was carried out in terms of the AL and the CP. The ALs and CPs of the 95% ACIs and Boot-p CIs for λ and β assuming $m = \{5, 6, 7, 8\}$. are resumed in tables Tables 2.3 and 2.4.

Table 2.1: Average and MSE Values of the MLEs and Bayes estimates when $\lambda = 1$ and $\beta = 2$

m	Parameter	Criterion	MLE	Bayes Estimates										
				Prior 0				Prior 1				Prior 2		
				LINEX			SE	LINEX			SE	LINEX		
				$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$		$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$		$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$
m=5	β	Average	4.285	3.730	3.005	2.813	2.143	2.144	2.097	1.971	2.199	2.201	2.129	1.952
		MSE	12.258	9.9641	3.0226	0.6083	0.0504	0.0507	0.0370	0.0236	0.1420	0.1431	0.1062	0.0650
	λ	Average	0.550	0.491	0.672	0.808	0.941	0.961	0.913	0.804	1.015	1.038	0.987	0.866
		MSE	0.670	0.834	0.598	0.412	0.1192	0.1099	0.1005	0.1019	0.0905	0.0911	0.0785	0.0746
m=6	β	Average	2.776	2.995	2.671	2.324	2.092	2.093	2.052	1.941	2.160	2.162	2.101	1.951
		MSE	4.9816	5.1280	2.0570	0.5392	0.0350	0.0352	0.0271	0.0221	0.1177	0.1183	0.0938	0.0642
	λ	Average	0.939	0.936	0.950	0.982	1.048	1.053	1.004	0.893	1.015	1.038	0.987	0.866
		MSE	0.6109	0.6132	0.5151	0.3858	0.1100	0.1049	0.0887	0.0749	0.0905	0.0911	0.0785	0.0746
m=7	β	Average	2.492	2.501	2.237	2.101	1.984	1.984	1.949	1.857	2.155	2.157	2.1050	1.979
		MSE	2.4238	2.4667	1.2455	0.4004	0.0304	0.0305	0.0228	0.0210	0.093	0.094	0.0737	0.0503
	λ	Average	1.158	1.205	1.121	1.108	1.048	1.073	1.023	0.899	0.955	0.998	0.949	0.833
		MSE	0.5639	0.5658	0.4832	0.3647	0.0802	0.0841	0.0694	0.0549	0.0828	0.0831	0.0751	0.0727
m=8	β	Average	2.376	2.421	2.181	2.115	2.0922	2.0828	2.051	1.968	2.114	2.115	2.074	1.977
		MSE	0.7529	0.8120	0.6131	0.3737	0.0153	0.0154	0.0103	0.0072	0.0501	0.0504	0.0383	0.0218
	λ	Average	1.100	1.112	1.081	1.072	1.051	1.086	1.037	0.921	0.993	1.050	1.000	0.880
		MSE	0.5416	0.5430	0.4717	0.3597	0.0787	0.0745	0.0624	0.0541	0.0791	0.0606	0.0688	0.0683

Table 2.2: Average and MSE Values of the MLEs and Bayes estimates when $\lambda = 2$ and $\beta = 1$

m	Parameter	Criterion	MLE	Bayes Estimates										
				Prior 0				Prior 3				Prior 4		
				LINEX			SE	LINEX			SE	LINEX		
				$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$		$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$		$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$
m=5	β	Average	1.626	1.703	1.405	1.357	1.068	1.069	1.049	0.998	1.046	1.046	1.034	1.002
		MSE	1.5740	1.5929	0.9497	0.3181	0.0285	0.0286	0.0246	0.0181	0.0123	0.0123	0.0109	0.0087
	λ	Average	1.784	1.762	1.708	1.798	2.003	2.014	1.888	1.617	2.039	2.044	1.960	1.760
		MSE	1.7627	1.7681	1.5471	1.4310	0.2154	0.2184	0.1907	0.2612	0.1028	0.1164	0.1006	0.1278
m=6	β	Average	1.444	1.398	1.208	1.198	1.0785	1.079	1.062	1.017	1.026	1.026	1.016	0.988
		MSE	0.5791	0.5814	0.3471	0.2047	0.0250	0.0251	0.0241	0.0148	0.0105	0.0105	0.0097	0.0085
	λ	Average	1.944	1.925	1.901	1.899	2.083	2.106	1.980	1.702	2.019	2.021	1.941	1.747
		MSE	1.7591	1.752	1.4231	1.390	0.2113	0.2168	0.1888	0.2477	0.0892	0.0971	0.0893	0.1277
m=7	β	Average	1.315	1.297	1.2577	1.189	1.038	1.038	1.024	0.986	1.040	1.041	1.031	1.006
		MSE	0.2623	0.2541	0.1914	0.1310	0.0172	0.0173	0.0154	0.0124	0.0094	0.0094	0.0085	0.0068
	λ	Average	1.536	1.621	1.781	1.812	1.914	1.977	1.866	1.618	1.988	1.985	1.910	1.729
		MSE	1.5867	1.5973	1.2577	1.0010	0.1879	0.1387	0.1543	0.2381	0.0881	0.0969	0.0853	0.1197
m=8	β	Average	1.174	1.179	1.128	1.118	1.044	1.044	1.033	1.003	1.032	1.0321	1.024	1.001
		MSE	0.1703	0.1715	0.1128	0.0891	0.0131	0.0132	0.0116	0.0087	0.0093	0.0093	0.0082	0.0066
	λ	Average	1.771	1.684	1.702	1.779	1.898	1.899	1.818	1.604	2.027	2.046	1.969	1.780
		MSE	1.0664	1.0981	0.9087	0.8727	0.16697	0.1272	0.1394	0.2344	0.0879	0.0884	0.0810	0.1136

Table 2.3: ALs and CPs of 95% CIs of $\lambda = 1$ and $\beta = 2$

Cases		ACI		Boot-p	
m		β	λ	β	λ
m=5	CP	0.80	0.93	0.70	0.81
	AL	3.8612	6.0170	7.5359	12.7930
m=6	CP	0.88	0.94	0.76	0.85
	AL	2.6151	5.8688	4.3292	7.0043
m=7	CP	0.88	0.95	0.82	0.87
	AL	2.2432	5.6800	2.9700	4.9216
m=8	CP	0.88	0.95	0.83	0.94
	AL	2.0205	5.4132	2.5839	4.9142

Table 2.4: ALs and CPs of 95% CIs of $\lambda = 2$ and $\beta = 1$

Cases		ACI		Boot-p	
m		β	λ	β	λ
m=5	CP	0.84	0.93	0.71	0.77
	AL	3.3471	6.4309	6.2641	28.2670
m=6	CP	0.86	0.93	0.73	0.83
	AL	2.6966	5.8499	4.7969	7.5577
m=7	CP	0.86	0.95	0.75	0.83
	AL	2.3408	5.6125	4.0902	6.7176
m=8	CP	0.90	0.97	0.79	0.86
	AL	1.9024	5.5002	2.7860	4.6248

Table Tables 2.1 and 2.2 show that the performances of the Bayes estimates are better than MLE for both parameters in terms of MSE. It can be also seen that the informative Bayes estimates under LINEX loss function with positive ν outperform the other estimates in most considered cases. some prior assumptions produce better Bayes estimates than other priors. For example, the MSEs of the Bayes estimates under Prior 4 are getting smaller than their counterparts under Prior 3. Clearly, the MSE of the proposed estimates decreases as m increases for both λ and β.

In view of interval estimation, Table Tables 2.3 and 2.4 summarize the ALs and CPs of ACIs and Boot-p CIs of λ and β when $(\lambda, \beta) = (1, 2)$. ACIs are superior to the Boot-p CIs as they produce higher CPs with less ALs.

Chapter 3

Prediction of Records from Kies Distribution

Prediction of future events on the basis of past and present information is a fundamental topic in statistics. Many real applications can be found. For example, in industrial production, a manufacturer would use the known previous imperfection and faults to predict an idealistic quality of the product. In economics, one would like to predict the next highest closing price of a particular stock.
The prediction problem is different from estimation. In prediction, the predictor uses a given data to guess about random value that is not a part of the data set, while an estimator uses the sample data to guess the value of the parameter.
In this chapter, we deal with the prediction problem of future records based on observed records from two-parameter Kies distribution. We derive different point predictors; namely: Maximum Likelihood, the modified maximum likelihood, conditional median and one-sample Bayesian predictors. Procedures of prediction intervals are also discussed such as pivotal method, shortest length prediction interval.

3.1 Classical Point Prediction

We deal with a variety of predictor types while discussing non-Bayesian point predictors, including maximum likelihood predictor, modified maximum likelihood predictor and conditional median predictor, which will be covered in the sections that follows.
let $y_1 < y_2 < ... < y_m$ be the first m upper records, let y_s be the s^{th} future record where $s > m$. Prediction of y_s based on the first m observed records, $\underline{y} = \{y_1, y_2, ..., y_m\}$, depends mainly on the conditional predictive density function of y_s given the observed records $\underline{y} = \{y_1, y_2, ..., y_m\}$. Using the Markovian property of records, the conditional

distribution of y_s given $\underline{y}$ is just the conditional distribution of y_s given y_m, see Arnold et al. (1998)[5], with a pdf given by:

$$f\left(y_s|y_m;\lambda,\beta\right)=\frac{\left[H\left(y_s\right)-H\left(y_m\right)\right]^{s-m-1}}{\Gamma\left(s-m\right)}\frac{f\left(y_s|\lambda,\beta\right)}{1-F\left(y_m|\lambda,\beta\right)} \tag{3.1}$$

Hence, using Eq's (1.42) and (1.43), Eq (3.1) reduces to

$$f(y_s|y_m)=\lambda^{s-m}\beta\frac{\left(\frac{y_s}{1-y_s}\right)^{\beta}}{y_s(1-y_s)}\frac{\left[\left(\frac{y_s}{1-y_s}\right)^{\beta}-\left(\frac{y_m}{1-y_m}\right)^{\beta}\right]^{(s-m-1)}}{\Gamma(s-m)}e^{-\lambda\left[\left(\frac{y_s}{1-y_s}\right)^{\beta}-\left(\frac{y_m}{1-y_m}\right)^{\beta}\right]} \tag{3.2}$$

where $0<y_m<y_s<1$.

3.1.1 Maximum Likelihood Predictor

Predicting the unobserved value of Y based on an observed random Simple $\underline{X}$ with joint pdf $f(\underline{x},y|\underline{\theta})$ is the main challenge in non-Bayesian prediction. Hence, let

$$L(y,\underline{\theta}|\underline{x})=f(y|\underline{x},\underline{\theta})f(\underline{x};\underline{\theta}),$$

be the predictive likelihood function (PLF) of y and $\underline{\theta}$ given $\underline{x}$, where $f(y|\underline{x},\underline{\theta})$ is the conditional density of y given the observed value of $\underline{x}$ and $f(\underline{x};\underline{\theta})$ is the joint pdf of the sample data $\underline{x}$.

Now, Assume $Y^{\star}=u(\underline{x})$ and $\theta^{\star}=w(\underline{x})$ are statistics. Then, $Y^{\star}$ is called a maximum likelihood predictor (MLP) of Y, and $\underline{\theta}^{\star}$ is the predictive maximum likelihood estimates (PMLEs) of $\underline{\theta}$, if

$$L(y^{\star},\theta^{\star}|\underline{x})=\sup_{y,\underline{\theta}}L(y,\underline{\theta}|\underline{x})$$

Now for our case $\underline{y}=\{y_1,y_2,...,y_m\}$ is a sequence of observed records from kies distribution where (λ,β) are the two unknown parameters of kies distribution. The PLF of y_s,λ and β is

$$L\left(y_s,\underline{\theta}|\underline{y}\right)=\prod_{i=1}^{m}h(y_i,\underline{\theta})\left(\frac{\left[H(y_s,\underline{\theta})-H(y_m,\underline{\theta})\right]^{s-m-1}}{\Gamma(s-m)}f(y_s,\underline{\theta})\right) \tag{3.3}$$

in general, if $\hat{y}_{MLP}=u(\underline{y}),\hat{\lambda}=v(\underline{y})$ and $\hat{\beta}=w(\underline{y})$ are statistics for which

$$L(\hat{y}_{MLP},\hat{\lambda},\hat{\beta}|\underline{y})=\sup_{y_s,\lambda,\beta}L(y_s,\lambda,\beta|\underline{y}) \tag{3.4}$$

then $u(\underline{y})$ is said to be the MLP of $y_s, 1 < m < s$, and $v(\underline{y})$ and $w(\underline{y})$ are the predictive maximum likelihood estimates (PMLEs) of λ and β, respectively.
Using Eq's (1.43), (1.44) and (1.45). Eq(3.3) is simplified to:

$$L(y_s, \lambda, \beta|\underline{y}) = \lambda^s \beta^{m+1} \prod_{i=1}^{m} \frac{A_i^\beta}{y_i(1-y_i)} \left[\frac{\left(A_s^\beta - A_m^\beta\right)^{s-m-1}}{\Gamma(s-m)} \frac{A_s^\beta}{y_s\,(1-y_s)} e^{-\lambda A_s^\beta} \right] \tag{3.5}$$

where $A_i = \dfrac{y_i}{1-y_i}$, so the predictive log-likelihood function is given by:

$$\begin{aligned} log\left(L(y_s, \lambda, \beta, |\underline{y})\right) &\propto slog(\lambda) + (m+1)log(\beta) + \beta \sum_{i=1}^{m} log(A_i) \\ &+ (s-m-1)log\left(A_s^\beta - A_m^\beta\right) \\ &+ (\beta-1)log\left(\frac{A_s}{1+A_s}\right) - (\beta+1)log\left(\frac{1}{1+A_s}\right) - \lambda A_s^\beta \end{aligned} \tag{3.6}$$

By using Eq (3.6) the predictive likelihood equations (PLEs) for y_s, λ and β are derived and presented, respectively, as follows:

$$\frac{\partial log(L(y_s, \lambda, \beta|\underline{y}))}{\partial \lambda} = \frac{s}{\lambda} - A_s^\beta = 0, \tag{3.7}$$

$$\begin{aligned} \frac{\partial log(L(y_s, \lambda, \beta))}{\partial \beta} &= \frac{m+1}{\beta} + \sum_{i=1}^{m} logA_i + (s-m-1)\frac{A_s^\beta logA_s - A_m^\beta logA_m}{A_s^\beta - A_m^\beta} \\ &+ (1-\lambda A_s^\beta)logA_s = 0 \end{aligned} \tag{3.8}$$

$$\begin{aligned} \frac{\partial log(L(y_s, \lambda, \beta))}{\partial y_s} &= (s-m-1)\beta \frac{A_s^\beta}{y_s(1-y_s)(A_s^\beta - A_m^\beta)} + \frac{\beta + 2y_s - 1}{y_s(1-y_s)} \\ &- \lambda\beta \frac{y_s^{\beta-1}}{(1-y_s)^{\beta+1}} = 0 \end{aligned} \tag{3.9}$$

The PMLE of λ is obtained from Eq (3.7) and it is given by:

$$\hat{\lambda} = \frac{s}{A_s^\beta} \tag{3.10}$$

The PMLE of β, say $\hat{\beta}$, and MLP of y_s, say $\hat{y}_{MLP}$, can be obtained by substituting Eq (3.10) in Eq (3.8) and Eq (3.9),respectively. Then, the obtained equations need to be solved with respect to β and y_s using numerical methods(Newton Raphson)

3.1.2 Modified Maximum Likelihood Predictor

In practical situations, experimenters are often interested in obtaining a simple and quick predictor. Al-olaimat et al. (2021) [4] proposed a modified maximum likelihood predictor (MMLP) by substituting the parameters λ and β in Eq (3.9) by their MLEs $\hat{\lambda}$ and $\hat{\beta}$ which are stated in Chapter 2. MMLP of y_s is obtained as the solution of the following equation:

$$\frac{1}{y_s(1-y_s)}\left[(s-m-1)\hat{\beta}\frac{A_s^{\hat{\beta}}}{A_s^{\hat{\beta}}-A_m^{\hat{\beta}}}+2y_s-\hat{\lambda}\hat{\beta}A_s^{\hat{\beta}}+\hat{\beta}-1\right]=0 \tag{3.11}$$

where $y_s > y_m$.
The MMLP of $y_s, \hat{y}_{MMLP}$ must be computed using a numerical method because Eq (3.11) cannot be solved analytically.

3.1.3 Conditional Median Predictor

Another potential predictor is the conditional median predictor (CMP), it was suggested by Raqab (1992) [28]. $\hat{Y}_s$ is the conditional median predictor of y_s given the observed data $\underline{y}$, if it is the median of its the conditional distribution. That is,

$$P_{\underline{\theta}}(y_s < \hat{Y}_s|\underline{y}) = P_{\underline{\theta}}(y_s > \hat{Y}_s|\underline{y}) = \frac{1}{2}.$$

Let $\hat{Y}_{CMP}$ is the conditional median predictor of y_s. Put $\hat{Y}_{CMP} = k(y_m;\lambda,\beta)$ which is a function of y_m, so

$$P((y_s|y_m;\lambda,\beta) \le k(y_m,\lambda,\beta)) = \frac{1}{2} = P((y_s|y_m;\lambda,\beta) \ge k(y_m,\lambda,\beta)),$$

Consequently from Eq (3.2), we have:

$$\int_{y_m}^{k(y_m,\lambda,\beta)} \lambda^{s-m}\beta\frac{A_s^{\beta}}{y_s(1-y_s)}\frac{(A_s^{\beta}-A_m^{\beta})^{s-m-1}}{\Gamma(s-m)}e^{-\lambda(A_s^{\beta}-A_m^{\beta})}dy_s = \frac{1}{2} \tag{3.12}$$

Setting $b = A_s^{\beta} - A_m^{\beta}$, we get :

$$\int_0^{\left[\frac{(k(A_m,\lambda,\beta))^{\beta}}{1-(k(A_m,\lambda,\beta))^{\beta}}-A_m^{\beta}\right]} \frac{\lambda^{s-m}}{\Gamma(s-m)}b^{s-m-1}e^{-\lambda b}db = \frac{1}{2} \tag{3.13}$$

Thus

$$\hat{Y}_{CMP} = \frac{\left(Med(W) + A_m^{\beta}\right)^{\frac{1}{\beta}}}{1 + \left(Med(W) + A_m^{\beta}\right)^{\frac{1}{\beta}}} \tag{3.14}$$

where $W \sim G\left(s - m, \dfrac{1}{\lambda}\right)$.

Assume that we want to predict the first future record, hence we'll use the formula $s = m + 1$ and $W \sim \mathcal{E}xp(\lambda)$ with $Med(W) = \dfrac{1}{\lambda} log2$, therefore

$$\hat{Y}_{CMP} = \frac{\left(log2^{\frac{1}{\lambda}} + A_m^{\beta}\right)^{\frac{1}{\beta}}}{1 + \left(log2^{\frac{1}{\lambda}} + A_m^{\beta}\right)^{\frac{1}{\beta}}} \tag{3.15}$$

3.2 Bayesian Point Prediction

The posterior prediction density of unobserved data based on observed ones is the main focus of Bayesian prediction. The posterior predictive distribution, $f_s^P(y_s|\underline{y})$, is defined as the distribution of the unobserved values y conditional on the observed data $\underline{x}$. It is given by:

$$\begin{aligned} f_s^P(y_s|\underline{y}) &= \mathbb{E}_{posterior}\left(f(Y_s|\underline{y}, \lambda, \beta)\right) \\ &= \int_{\underline{\theta}} f(y|\underline{\theta}, \underline{y})\pi(\underline{\theta}|\underline{y})d\underline{\theta} \end{aligned} \tag{3.16}$$

where $\pi(\underline{\theta}|\underline{x})$ is the posterior distribution of $\underline{\theta}$ given observed data.

3.2.1 One-Sample Prediction Problem

The objective of this section is to determine the Bayes predictive estimate of the s^{th} record y_s, $s > m$ based on observed record sample $\underline{y} = \{y_1, y_2, ..., y_m\}$ from kies distribution under the assumption of SE and LINEX loss functions.
Given the values of $\underline{y} = \{y_1, y_2, ..., y_m\}$ the posterior predictive density of y_s is given by:

$$f_s^P(y_s|\underline{y}) = \int_0^{\infty}\int_0^{\infty} f(y_s|\underline{y}, \lambda, \beta)\pi(\lambda, \beta|\underline{y})d\lambda d\beta \tag{3.17}$$

where $f(y_s|\underline{y},\lambda,\beta)$ and $\pi(\lambda,\beta|\underline{y})$ are given in Eq's (3.2) and (2.26) respectively. Substituting these equations in Eq(3.17), then the posterior predictive density function $f_s^P(y_s|\underline{y})$ becomes

$$\begin{aligned} f_s^P(y_s|\underline{y}) &= \frac{1}{C}\int_0^\infty\int_0^\infty \frac{\lambda^{s-m}\beta}{\Gamma(s-m)}\frac{A_s^\beta}{y_s(1-y_s)}(A_s^\beta - A_m^\beta)^{s-m-1}e^{-\lambda(A_s^\beta - A_m^\beta)} \\ &\times \; \lambda^{m+a_1-1}\, e^{-\lambda(b_1+A_m^\beta)}\beta^{m+a_2-1}\, e^{-\beta\left(b_2-\sum\limits_{i=1}^{m} log(A_i)\right)} d\lambda d\beta \end{aligned} \tag{3.18}$$

Since

$$\int_0^\infty \lambda^{s+a_1-1}e^{-\lambda(b_1+A_s^\beta)}d\lambda = \frac{\Gamma(s+a_1)}{(b_1+A_s^\beta)^{s+a_1}} \tag{3.19}$$

and

$$\begin{aligned} C &= \int_0^\infty\int_0^\infty \lambda^{m+a_1-1}e^{-\lambda(b_1+A_m^\beta)}\beta^{m+a_2-1}e^{-\beta\left(b_2-\sum\limits_{i=1}^{m} log(A_i)\right)}d\lambda d\beta \\ &= \frac{\Gamma(m+a_1)\Gamma(m+a_2)}{\left(b_2-\sum\limits_{i=1}^{m} log(A_i)\right)^{m+a_2}} \times \mathbb{E}_{\pi_\beta^*}\left[J(\beta)\right] \end{aligned} \tag{3.20}$$

then Eq (3.18) reduces to the form

$$\begin{aligned} f_s^P(y_s|\underline{y}) &= \frac{\left(b_2-\sum\limits_{i=1}^{m} log(A_i)\right)^{m+a_2}}{\Gamma(m+a_1)\Gamma(m+a_2)} \frac{\Gamma(s+a_2)\Gamma(m+a_2+1)}{\Gamma(s-m)\left(b_2-\sum\limits_{i=1}^{m} log(A_i)\right)^{m+a_2+1}} \\ &\times \; \frac{1}{y_s(1-y_s)}\frac{\mathbb{E}_{\pi_1^*}\left[I(y_s,\beta)\right]}{\mathbb{E}_{\pi_2^*}\left[J(\beta)\right]} \end{aligned} \tag{3.21}$$

where

$$\begin{aligned} I(y_s,\beta) &= \frac{(A_s^\beta - A_m^\beta)^{s-m-1}A_s^\beta}{(A_s^\beta+b_1)^{s+a_1}}, \qquad J(\beta) = \frac{1}{(b_1+A_m^\beta)^{m+a_1}} \\ \pi_1^* &\propto Gamma\left(m+a_2+1, \frac{1}{b_2-\sum\limits_{i=1}^{m} log(A_i)}\right) \\ \pi_2^* &\propto Gamma\left(m+a_2, \frac{1}{b_2-\sum\limits_{i=1}^{m} log(A_i)}\right) \end{aligned}$$

One can notice that Eq (3.21) cannot be computed explicitly. Therefore, an approximate can be proposed for $f_s^P(y_s|\underline{y})$ which is denoted by $\hat{f}_s^*(y_s|\underline{y})$, by replacing the parameter

β by its Bayes estimate obtained in Chapter 2, using two Bayes estimates under SE and LINEX.

$$\begin{aligned}\hat{f}_s^*(y_s|\underline{y}) &= \frac{1}{Q(y_m)}\frac{\Gamma(s+a_2)(m+a_2)}{\Gamma(m+a_1)\Gamma(s-m)\left(b_2-\sum\limits_{i=1}^{m}log(A_i)\right)} \\ &\times \frac{1}{y_s(1-y_s)}\frac{I(y_s,\hat{\beta})}{J(\hat{\beta})}\end{aligned} \tag{3.22}$$

where $Q(y_m)=\displaystyle\int_{y_m}^{1}\hat{f}_s^*(y_s|\underline{y})dy_s$. If $\hat{Y}$ is a predictor of y_s, $0<y_m<y_s<1$, then the Bayes predictive estimates of y_s under SE loss function given by:

$$\begin{aligned}\hat{Y}_{SEP} &= \mathbb{E}_{\hat{f}_s^*}\left(Y_s|\underline{y}\right) \\ &= \int_{y_m}^{1} y_s\hat{f}_s^*(y_s|\underline{y})dy_s \\ &= \frac{1}{Q(y_m)}\frac{\Gamma(s+a_2)(m+a_2)}{J(\hat{\beta})\Gamma(m+a_1)\Gamma(s-m)\left(b_2-\sum\limits_{i=1}^{m}log(A_i)\right)} \\ &\times \int_{y_m}^{1}\frac{I(y_s,\hat{\beta})}{(1-y_s)}dy_s\end{aligned} \tag{3.23}$$

and here for the LINEX loss function:

$$\begin{aligned}\hat{Y}_{LEP} &= \frac{-1}{\nu}log\mathbb{E}_{\hat{f}_s^*}\left(e^{-\nu Y_s}|\underline{y}\right) \\ &= \frac{-1}{\nu}log\int_{y_m}^{1} e^{-\nu y_s}\hat{f}_s^*(y_s|\underline{y})dy_s \\ &= \frac{-1}{\nu}log\left[\frac{1}{Q(y_m)}\frac{\Gamma(s+a_2)(m+a_2)}{J(\hat{\beta})\Gamma(m+a_1)\Gamma(s-m)\left(b_2-\sum\limits_{i=1}^{m}log(A_i)\right)}\right] \\ &+ \frac{-1}{\nu}log\left[\int_{y_m}^{1} e^{-\nu y_s}\frac{I(y_s,\hat{\beta})}{y_s(1-y_s)}dy_s\right]\end{aligned} \tag{3.24}$$

Additionally, since it is a common goal of mine to predict the initial value of the unobserved record, y_{m+1}. In Eq's (3.23) and (3.24), we substitute $s=m+1$. Therefore, by applying the binomial expansion to $(A_s^{\hat{\beta}}-A_m^{\hat{\beta}})^{s-m+1}$, we immediately obtain the following predictors:

$$\hat{Y}_{m+1}^{SEP} = \frac{1}{Q(y_m)} \frac{(b_1 + A_s^{\hat{\beta}})^{m+a_1} \Gamma(s+a_2)(m+a_2)}{\Gamma(m+a_1)\left(b_1 - \sum_{i=1}^{m} log(A_i)\right)} \times \int_{y_m}^{1} \frac{1}{1-y_s} \frac{A_s^{\hat{\beta}}}{(b_1 + A_s^{\hat{\beta}})^{s+a_1}} dy_s \tag{3.25}$$

and

$$\hat{Y}_{m+1}^{LEP} = \frac{-1}{\nu} log \left[\frac{1}{Q(y_m)} \frac{(b_1 + A_s^{\hat{\beta}})^{m+a_1} \Gamma(s+a_2)(m+a_2)}{\Gamma(m+a_1)\left(b_1 - \sum_{i=1}^{m} log(A_i)\right)} \right] + \frac{-1}{\nu} log \left[\int_{y_m}^{1} e^{-\nu y_s} \frac{1}{y_s(1-y_s)} \frac{A_s^{\hat{\beta}}}{(b_1 + A_s^{\hat{\beta}})^{s+a_1}} dy_s \right] \tag{3.26}$$

3.3 Prediction Intervals

One would want to forecast a future observation in many practical situations using historical data from the same population. Making a prediction interval, which is an interval that will contain the unobserved value with a given probability, is one technique to achieve this. We consider the prediction interval problem using two classical methods: the shortest length and pivotal quantity methods.

3.3.1 Pivotal Method

Define the random variable T as

$$T = A_s^{\beta} - A_m^{\beta}$$

it can be easily seen that $T|y_m \sim G\left(s-m, \frac{1}{\lambda}\right)$ by using transformation of $T|y_m$ in Eq (3.2).

Therefore, when the parameters β and λ are known and y_m is given then the pivotal quantity $2\lambda T|y_m \sim \mathcal{X}^2_{(2(s-m))}$. The exact $(1-\alpha)100\%$ PI of y_s is therefore $(L_1(y_m), U_1(y_m))$, where $L_1(y_m)$ is given by :

$$L_1(y_m) = \frac{\left(\frac{\mathcal{X}^2_{(2(s-m))}\left(\frac{\alpha}{2}\right)}{2\lambda} + A_m^{\beta}\right)^{\frac{1}{\beta}}}{1 + \left(\frac{\mathcal{X}^2_{(2(s-m))}\left(\frac{\alpha}{2}\right)}{2\lambda} + A_m^{\beta}\right)^{\frac{1}{\beta}}} \tag{3.27}$$

and $U_1(ym)$ is given by:

$$U_1(y_m) = \frac{\left(\frac{\mathcal{X}^2_{(2(s-m))}\left(1-\frac{\alpha}{2}\right)}{2\lambda} + A_m^\beta\right)^{\frac{1}{\beta}}}{1+\left(\frac{\mathcal{X}^2_{(2(s-m))}\left(1-\frac{\alpha}{2}\right)}{2\lambda} + A_m^\beta\right)^{\frac{1}{\beta}}} \tag{3.28}$$

When λ and β are unknown, the parameters in Eq's (3.27) and (3.28) must be calculated, meaning by their MLEs. Consequently, the following is how a rough $(1-\alpha)100\%$ PI is obtained:

$$\hat{L}_1(y_m) = \frac{\left(1+\frac{\mathcal{X}^2_{(2(s-m))}\left(\frac{\alpha}{2}\right)}{2m}\right)^{\frac{1}{\beta}} A_m}{1+\left(1+\frac{\mathcal{X}^2_{(2(s-m))}\left(\frac{\alpha}{2}\right)}{2m}\right)^{\frac{1}{\beta}} A_m} \tag{3.29}$$

and

$$\hat{U}_1(y_m) = \frac{\left(1+\frac{\mathcal{X}^2_{(2(s-m))}\left(1-\frac{\alpha}{2}\right)}{2m}\right)^{\frac{1}{\beta}} A_m}{1+\left(1+\frac{\mathcal{X}^2_{(2(s-m))}\left(1-\frac{\alpha}{2}\right)}{2m}\right)^{\frac{1}{\beta}} A_m} \tag{3.30}$$

respectively.

Since we are usually interested in the first prediction, $s = m+1$. Then, by using the pivotal quantity $\lambda T|y_m \sim \mathcal{E}xp(1)$, the $(1-\alpha)100\%$ exact PI of y_{m+1} are given by:

$$L_2(y_m) = \frac{\left(A_m^\beta - \frac{1}{\lambda} log\left(1-\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}}{1+\left(A_m^\beta - \frac{1}{\lambda} log\left(1-\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}} \tag{3.31}$$

$$U_2(y_m) = \frac{\left(A_m^\beta - \frac{1}{\lambda} log\left(\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}}{1+\left(A_m^\beta - \frac{1}{\lambda} log\left(\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}} \tag{3.32}$$

and approximate PI of y_{m+1} are obtained by:

$$\hat{L}_2(y_m) = \frac{A_m\left(1 - \frac{1}{m} log\left(1-\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}}{1+A_m\left(1 - \frac{1}{m} log\left(1-\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}} \tag{3.33}$$

$$\hat{U}_2(y_m) = \frac{A_m\left(1 - \frac{1}{m} log\left(\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}}{1+A_m\left(1 - \frac{1}{m} log\left(\frac{\alpha}{2}\right)\right)^{\frac{1}{\beta}}} \tag{3.34}$$

3.3.2 The Shortest Length Prediction Intervals

The shortest length prediction interval (SLPI) is an other methods related PI. Using the realization that

$$V = \frac{2m\left(A_s^{\hat{\beta}} - A_m^{\hat{\beta}}\right)}{A_m^{\hat{\beta}}} \sim \mathcal{X}^2_{(2(s-m))}.$$

We select the variables c_1 and c_2 where $c_1 < c_2$ matching the criteria listed below:

$$\mathbb{P}\left(c_1 < \mathcal{X}^2_{(2(s-m))} < c_2\right) = 1 - \alpha \tag{3.35}$$

it is equal to

$$\mathbb{P}\left(c_1 < \frac{2m\left(A_s^{\hat{\beta}} - A_m^{\hat{\beta}}\right)}{A_m^{\hat{\beta}}} < c_2\right) = 1 - \alpha \tag{3.36}$$

or equivalently

$$\mathbb{P}\left(\frac{A_m\left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1 + A_m\left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}} < y_s < \frac{A_m\left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1 + A_m\left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}}\right) = 1 - \alpha \tag{3.37}$$

Thus, a $(1-\alpha)100\%$ PI for y_s can be obtained as $(L_3(y_m), U_3(y_m))$, where:

$$L_3(y_m) = \frac{A_m\left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1 + A_m\left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}} \tag{3.38}$$

and

$$U_3(y_m) = \frac{A_m\left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1 + A_m\left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}} \tag{3.39}$$

The best possibilities for c_1 and c_2 are the ones that minimise $U_3(y_m) - L_3(y_m)$ the width of PI. By imposing (3.35) on the Lagrangian multipliers function, it is possible to derive the $(1-\alpha)100\%$ SL PI as follows:

$$L(c_1, c_2, \omega) = U_3(y_m) - L_3(y_m) - \omega\left[\int_{c_1}^{c_2} g(v)d(v) - (1-\alpha)\right] \tag{3.40}$$

it's equal to

$$L(c_1, c_2, \omega) = \frac{A_m \left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1+A_m \left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}} - \frac{A_m \left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1+A_m \left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}} - \omega \left[\int_{c_1}^{c_2} g(v)d(v) - (1-\alpha)\right] \tag{3.41}$$

where g is the pdf of $\mathcal{X}^2_{(2(s-m))}$ distribution and ω is a Lagrangian multiplier. The constants c_1 and c_2 can be derived by equating the partial derivative of $L(c_1, c_2, \omega)$, with respect to c_1, c_2 and ω, to zero as follows:

$$\frac{\partial L}{\partial c_1} = \frac{\frac{-A_m}{2m\hat{\beta}} \left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}-1}}{\left(1+A_m \left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}\right)^2} + \omega g(c_1) = 0 \tag{3.42}$$

$$\frac{\partial L}{\partial c_2} = \frac{\frac{A_m}{2m\hat{\beta}} \left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}-1}}{\left(1+A_m \left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}\right)^2} - \omega g(c_2) = 0 \tag{3.43}$$

$$\frac{\partial L}{\partial \omega} = -\left[\int_{c_1}^{c_2} g(v)d(v) - (1-\alpha)\right] = 0 \tag{3.44}$$

After some algebraic computations on Eq's (3.42) and (3.43) We get to

$$\left(\frac{c_2}{c_1}\right)^{s-m-1} e^{-\frac{1}{2}(c_2-c_1)} = \left(\frac{2m+c_2}{2m+c_1}\right)^{\frac{1}{\hat{\beta}}-1} \times \left[\frac{1+A_m \left(1+\frac{c_1}{2m}\right)^{\frac{1}{\hat{\beta}}}}{1+A_m \left(1+\frac{c_2}{2m}\right)^{\frac{1}{\hat{\beta}}}}\right]^2 \tag{3.45}$$

also from Eq (3.44), we reach

$$\int_{c_1}^{c_2} g(v)d(v) = (1-\alpha) \tag{3.46}$$

Now, c_1 and c_2 of the shortest PI can be computed simultaneously by solving Eq's (3.45) and (3.46) numerically.

We now consider the case where $s = m+1$. In this case g is decreasing function with $g(0) = \dfrac{1}{2}$ and $g_(\infty) = 0$. Consequently, the lower endpoint of the shortest PI can be chosen simply as $L_3(y_m) = y_m$, this leads that the $(1-\alpha)100\%$ PI is $(y_m, U_3(y_m))$ as a modified SL PI for y_{m+1}.

3.4 Comparison between Proposed Prediction Methods

We present a simulation analysis in this section to evaluate the efficiency of the suggested prediction techniques that were covered in this chapter. Mean square prediction errors (MSPEs) and average biases of predictor are used to quantify their performance. We contrast the PIs in terms of their estimated CPs and ALs from the previous sections as well.
According to Al-Olaimat et al. (2021) [4], the following priors are assumed on the instances being evaluated while computing Bayes predictors under the SE and LINEX loss functions:
Prior 0:$a_1 = 0, b_1 = 0, a_2 = 0, b_2 = 0$.
For $\lambda = 1, \beta = 2$:
Prior 1:$a_1 = 20, b_1 = 20, a_2 = 16, b_2 = 8$.
For $\lambda = 2, \beta = 1$:
Prior 1:$a_1 = 1, b_1 = 0.5, a_2 = 20, b_2 = 20$.
Depending to the values of λ and β that are set in the last section , simulation is conducted using a variety of records from the Kies distribution. they calculate the point predictor's value in each case using both classical and Bayesian methods. Additionally, 95% PIs using the pivotal quantity and SL methods are simulated. MSPEs and prediction biases of the predictors are reported. Moreover, the CPs and ALs of the PIs are computed, the obtained results are summarized in Tables 3.1 and 3.2 and in Tables 3.3 and 3.4, respectively.

Table 3.1: MSPEs and Average Bias from simulations of $\lambda = 1$ and $\beta = 2$

m	Y_s	Criterion	MLP	MMLP	CMP	Prior 0				Prior 1			
						BE_{SE}	BE_{LE}			BE_{SE}	BE_{LE}		
							$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$		$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$
m=5	Y_7	MSP E	0.001492	0.001276	0.000465	0.001538	0.001541	0.001432	0.00127	0.000724	0.000725	0.000699	0.000649
		Bias	-0.025749	-0.020434	-0.002417	-0.014546	-0.014617	-0.011204	-0.002874	-0.006494	-0.006517	-0.005386	-0.002143
	Y_8	MSP E	0.002023	0.001688	0.000459	0.002091	0.002096	0.001909	0.001649	0.000846	0.000847	0.000811	0.000743
		Bias	-0.029308	-0.021296	-0.001470	-0.018323	-0.018420	-0.013771	-0.002445	-0.007216	-0.007246	-0.005708	-0.001262
m=6	Y_8	MSP E	0.000849	0.000745	0.000258	0.000459	0.000460	0.000436	0.000426	0.000333	0.000334	0.000328	0.000320
		Bias	-0.019653	-0.015785	-0.001638	-0.004980	-0.005027	-0.002732	0.003049	-0.001814	-0.001830	-0.001048	0.001246
	Y_9	MSP E	0.001666	0.001460	0.000426	0.000616	0.000617	0.000566	0.000529	0.000383	0.000384	0.000372	0.000353
		Bias	-0.025196	-0.019210	-0.003844	-0.007818	-0.007883	-0.004729	0.003205	-0.003305	-0.003325	-0.002780	0.000744
m=7	Y_9	MSP E	0.000674	0.000589	0.000313	0.000388	0.000389	0.000378	0.000375	0.000286	0.000286	0.000283	0.000277
		Bias	-0.017886	-0.015077	-0.004944	-0.007485	-0.007520	-0.005830	-0.003474	-0.005342	-0.005355	-0.004736	-0.002978
	Y_{10}	MSP E	0.000716	0.000604	0.000296	0.000539	0.000539	0.000522	0.000528	0.000367	0.000368	0.000365	0.000363
		Bias	-0.017047	-0.012547	-0.001886	-0.005522	-0.005561	-0.003252	0.002697	-0.001536	-0.001553	-0.000724	0.001703
m=8	Y_{10}	MSP E	0.000529	0.000478	0.000215	0.000332	0.000332	0.000322	0.000311	0.000264	0.000264	0.000261	0.000252
		Bias	-0.014086	-0.011842	-0.002672	-0.003652	-0.003677	-0.002404	0.000932	-0.002214	-0.002224	-0.001756	-0.000378
	Y_{11}	MSP E	0.000866	0.000762	0.000299	0.000551	0.000552	0.000529	0.000501	0.000419	0.000419	0.000412	0.000395
		Bias	-0.017215	-0.013541	-0.003970	-0.005774	-0.005810	-0.004032	0.000622	-0.003266	-0.003271	-0.002619	0.002931

Table 3.2: MSPEs and Average Bias from simulations of $\lambda = 2$ and $\beta = 1$

m	Y_s	Criterion	MLP	MMLP	CMP	Prior 2				Prior 3			
						BE_{SE}	BE_{LE}			BE_{SE}	BE_{LE}		
							$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$		$\nu = -0.01$	$\nu = 0.5$	$\nu = 2$
m=5	Y_7	MSP E	0.006668	0.005880	0.001870	0.004541	0.004554	0.004346	0.003902	0.002720	0.002720	0.002717	0.002713
		Bias	-0.048639	-0.038692	-0.004069	-0.026404	-0.026460	-0.023726	-0.016903	-0.013969	-0.013973	-0.013787	-0.013326
	Y_8	MSP E	0.007759	0.006576	0.001536	0.005861	0.005868	0.005525	0.004803	0.002859	0.002859	0.002854	0.002844
		Bias	-0.052920	-0.039338	-0.001672	-0.030690	-0.030766	-0.027083	-0.017871	-0.013394	-0.013401	-0.013067	-0.012201
m=6	Y_8	MSP E	0.004181	0.003571	0.001082	0.001787	0.001788	0.001747	0.001676	0.0012774	0.001277	0.001278	0.001271
		Bias	-0.043753	-0.036647	-0.009032	-0.017547	-0.017584	-0.015776	-0.011111	-0.008661	-0.008673	-0.008453	-0.007864
	Y_9	MSP E	0.004781	0.003882	0.000991	0.002315	0.002317	0.002224	0.002049	0.001393	0.001394	0.001391	0.001385
		Bias	-0.044832	-0.034681	-0.005064	-0.019082	-0.019132	-0.016684	-0.010374	-0.006791	-0.006807	-0.006462	-0.005522
m=7	Y_9	MSP E	0.002482	0.002163	0.000964	0.000728	0.000729	0.000705	0.000658	0.000553	0.000553	0.000552	0.000541
		Bias	-0.032610	-0.027503	-0.005911	-0.009754	-0.009778	-0.008530	-0.005255	-0.005776	-0.005778	-0.005645	-0.005315
	Y_{10}	MSP E	0.002956	0.002462	0.001034	0.000945	0.000946	0.000903	0.000818	0.000582	0.000582	0.000581	0.000580
		Bias	-0.034864	-0.027463	-0.004354	-0.013391	-0.013434	-0.011725	-0.007236	-0.007706	-0.007710	-0.007472	-0.006851
m=8	Y_{10}	MSP E	0.001291	0.001143	0.000553	0.000965	0.000965	0.000939	0.000871	0.000771	0.000771	0.000768	0.000760
		Bias	-0.021682	-0.018040	-0.005155	-0.010264	-0.010283	-0.009383	-0.006975	-0.007069	-0.007072	-0.006903	-0.006431
	Y_{11}	MSP E	0.001623	0.001382	0.000528	0.001115	0.001116	0.001075	0.000982	0.000826	0.000826	0.000822	0.000811
		Bias	-0.024281	-0.018911	-0.006263	-0.011421	-0.011455	-0.010221	-0.006917	-0.007066	-0.007070	-0.006832	-0.006185

Table 3.3: CPs and ALs from simulations of $\lambda = 1$ and $\beta = 2$

m	Y_s	Criterion	Pivot	SL
m=5	Y_7	CP	0.94	0.68
		AL	0.081834	0.054774
	Y_8	CP	0.95	0.63
		AL	0.08628	0.068958
m=6	Y_8	CP	0.96	0.76
		AL	0.067859	0.046175
	Y_9	CP	0.98	0.64
		AL	0.072816	0.058322
m=7	Y_9	CP	0.93	0.67
		AL	0.060137	0.037941
	Y_{10}	CP	0.97	0.67
		AL	0.065109	0.045184
m=8	Y_{10}	CP	0.99	0.84
		AL	0.049566	0.038217
	Y_{11}	CP	0.99	0.76
		AL	0.054562	0.045384

Table 3.4: CPs and ALs from simulations of $\lambda = 2$ and $\beta = 1$

m	Y_s	Criterion	Pivot	SL
m=5	Y_7	CP	0.95	0.69
		AL	0.138920	0.093530
	Y_8	CP	0.95	0.65
		AL	0.141140	0.116990
m=6	Y_8	CP	0.95	0.68
		AL	0.108720	0.076062
	Y_9	CP	0.96	0.72
		AL	0.112680	0.107510
m=7	Y_9	CP	0.95	0.76
		AL	0.092416	0.066912
	Y_{10}	CP	0.95	0.754
		AL	0.096927	0.091193
m=8	Y_{10}	CP	0.95	0.79
		AL	0.078396	0.058354
	Y_{11}	CP	0.95	0.77
		AL	0.083206	0.080934

From Tables 3.1 and 3.2, and by considering the prediction average biases as an optimality criterion, there is a clear evident that the CMP are the most preferred classical point predictors. When comparing among the classical methods, one can see that the prediction average biases of the CMP are less than those of the MLP and MMLP for all the considered cases. Further, we note the prediction average biases of the MMLP are less

than those of MLP for all the considered cases. When comparing between Bayesian and frequentist methods, we can observe that Bayes predictors perform well under different error loss functions and priors in the sense of bias compared with MLPs and MMLPs.
By considering MSPEs as an optimality criterion, it is observed that the Bayes predictors perform better than MLPs and MMLPs. Further, it can be observed that Bayes predictors under the informative priors Prior 1, Prior 2 are more efficient than the corresponding Bayes predictors under Prior 0.
From Tables 3.3 and 3.4 , and by considering the AL as an optimality criterion, we can see the SL method is more efficient than the other method for obtaining PIs.

Conclusion

In this dissertation, classical and Bayesian inferences (estimation and prediction) for the two-parameter Kies distribution based on upper records were proposed. Based on records, some distributional features of the Kies distribution were investigated. The existence and uniqueness of MLEs are discussed. Asymptotic and bootstrap confidence intervals are computed. Bayesian estimates based on the SE and LINEX loss functions are proposed. Exact Bayes estimations of the parameters are not possible. In general, the suggested informative Bayes estimates outperform the classical estimates in all scenarios evaluated. However, for small ν, non-informative Bayesian and classical estimate approaches perform approximately equivalent under SE and LINEX, while Bayesian methods perform better under LINEX for other positive values ν. For the considered confidence intervals. The ACIs outperform Boot-p CIs in all cases. We also investigated future record prediction for the two-parameter Kies distribution. To construct point and interval predictors of future records, both classical and Bayesian techniques were established. The MLP and the MMLP are close to each other. It has been noticed that Bayes predictors perform better than the MLP and MMLP in terms of bias and MSPEs, under SE and LINEX loss functions. In the context of prediction intervals, it was observed that the SL method is the most appropriate technique for obtaining PIs of the unobserved future records when adopting ALs as the optimality criterion. When adopting the CPs as the optimality criterion, it was noticed that the pivotal quantity method is an efficient technique for obtaining PIs in most the considered cases.

Bibliography

[1] Ahmed, E. A., (2014). Bayesian estimation based on progressive Type-II censoring from two parameter bathtub-shaped lifetime model: an Markov chain Monte Carlo approach, Journal of Applied Statistics, 41:4, 752-768.

[2] Ahsanullah, M. (2004). *Record values-theory and applications.* University Press of America.

[3] Akgül, A. (2018). A novel method for a fractional derivative with non-local and non-singular kernel. Chaos, Solitons and Fractals, 114, 478-482.

[4] Al-Olaimat, N. M., Bayoud, H. A., Raqab, M. Z. (2021). Record data from Kies distribution and related statistical inferences. *Statistics in Transition new series*, 22(4), 153-170.

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

[6] Awad, A. M., and Raqab, M. Z. (2000). Prediction intervals for the future record values from exponential distribution: comparative study. Journal of Statistical Computation and Simulation, 65(1-4), 325-340.

[7] Basak, P., and Balakrishnan, N. (2003). Maximum likelihood prediction of future record statistic. In Mathematical and statistical methods in reliability (pp. 159-175).

[8] Bayoud, H. A. (2016). Admissible minimax estimators for the shape parameter of Topp-Leone distribution. Communications in Statistics-Theory and Methods, 45(1), 71-82.

[9] Chandler, K. (1952). The distribution and frequency of record values. Journal of the Royal Statistical Society: Series B (Methodological), 14(2), 220-228.

[10] Chen, M. H., Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. Journal of computational and Graphical Statistics, 8(1), 69-92.

[11] Dunsmore, I. R. (1983). The future occurrence of records. Annals of the Institute of Statistical Mathematics, 35, 267-277.

[12] Efron, B., Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.

[13] Foster, F. G., and Stuart, A. (1954). Distribution-Free Tests in Time-Series Based on the Breaking of Records Division of Research Techniques, London School of Economics. Journal of the Royal Statistical Society: Series B (Methodological), 16(1), 1-13.

[14] Held, L., Daniel, S. B (2016) .*Applied statistical inference*. Springer, Berlin Heidelberg. 16, 978-3 .

[15] Houchens, R. L. (1984). Record value theory and inference. University of California, Riverside.

[16] Genc, A. I., (2013). Estimation of $\mathbb{P}(X > Y)$ with Topp-Leone distribution, Journal of Statistical Computation and Simulation, 83 (2), 326-339.

[17] Kies, J. A. (1958). The strength of glass. Naval Research Laboratory.

[18] Kumar, C. S., and Dharmaja, S. H. S. (2013). On reduced Kies distribution. Collection of recent statistical methods and applications, 111-123.

[19] Kumar, C. S., and Dharmaja, S. H. S. (2014). On some properties of Kies distribution. Metron, 72, 97-122.

[20] Kumar, C. S., and Dharmaja, S. H. S. (2017). The exponentiated reduced Kies distribution: Properties and applications. Communications in Statistics-Theory and Methods, 46(17), 8778-8790.

[21] Kundu, D., and Pradhan, B. (2009). Bayesian inference and life testing plans for generalized exponential distribution. Science in China Series A: Mathematics, 52(6), 1373-1388.

[22] Lehmann, E. L., and Casella, G. (1998). Preparations. Theory of Point Estimation, 1-81.

[23] Lehmann, E. L., and Casella, G. (2006). Theory of point estimation. Springer Science and Business Media.

[24] Mkelinen, T., Schmidt, K., Styan, G. P. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. The Annals of Statistics, 758-767.

[25] Nagaraja, H. N. (1988). Record values and related statistics-a review. Communications in Statistics-Theory and Methods, 17(7), 2223-2238.

[26] Papke, L. E., Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of applied econometrics, 11(6), 619-632.

[27] Pradhan, B. and Kundu, D., (2011). Bayes estimation and prediction of the two-parameter gamma distribution, Journal of Statistical Computation and Simulation, 81(9), pp. 1187-1198.

[28] Raqab, M. Z. (1992). Predictors of future order statistics from Type II censored samples. The Ohio State University.

[29] Ren, J. and Gui, W., (2020). Inference and optimal censoring scheme for progressively Type-II censored competing risk model for generalized Rayleigh distribution, Computational Statistics, 1-35.

[30] Runcorn, K. (1989). Sir Harold Jeffreys (1891-1989). Nature, 339(6220), 102-102.

[31] Xia, J., Mi, J., Zhou, Y. (2009). On the existence and uniqueness of the maximum likelihood estimators of normal and lognormal population parameters with grouped data. Journal of Probability and Statistics.

[32] Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. Journal of the American Statistical Association, 81(394), 446-451.