A MODIFIED METHOD FOR BAYESIAN PREDICTION OF FUTURE ORDER STATISTICS FROM GENERALIZED POWER FUNCTION

ALMUTAIRI ANED OMAR

UNIVERSITI SAINS MALAYSIA

2015

A MODIFIED METHOD FOR BAYESIAN PREDICTION OF FUTURE ORDER STATISTICS FROM GENERALIZED POWER FUNCTION

By

ALMUTAIRI ANED OMAR

Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

April 2015

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deep gratitude to those who have contributed throughout the duration of my research work. First and foremost, I would like to extend my gratitude to my supervisor, Professor Dr. Low Heng Chin, for her guidance, invaluable advice and continuous concern throughout this thesis. Sincere thanks to the Dean of the School of Mathematical Sciences and the staff in Universiti Sains Malaysia especially the staff in the School of Mathematical Sciences for their assistance and cooperation. Finally, I would like to thank all my family members, my colleagues and friends for their support and sharing. Last but not least, my sincere thanks and appreciation to my father, my mother and my sister for their encouragement, understanding and support throughout the period spent in completing this thesis.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

a	location parameter of standard generalized power function
b	scale parameter of standard generalized power function
(c, d)	parameters of informative prior function
$d(y)$	decision function
$d(Y)$	Bayes estimator
d^*	Bayesian decision function
$E(p\|y)$	mean of the posterior distribution for p
$E(b\|y)$	mean posterior function for b
e	informative experiment
f	future experiment
$f(y)$	probability density function
$F(y)$	cumulative distribution function
(g, h)	parameters of informative prior function
Group A	action space
Group Θ	parameter space
Group R_y	space of Y
Group d	decision space

$H(p\|y)$	posterior distribution for informative prior function
$H_1(p\|y)$	posterior distribution for non-informative prior function
$H_2(p\|y)$	posterior distribution of informative prior function for multi sample
$H_3(p\|y)$	posterior distribution of non-informative prior function for multi sample
$H_4(p\|y)$	posterior distribution for non- informative prior function
$H_5(p\|y)$	posterior distribution for informative prior function
$h(y)$	failure rate function
$h(\theta)$	prior density function
I	sufficient statistic for p and multi samples prediction
k_n	complete sufficient statistic for p and the complete sample
$L(X)$	lower limit for Bayesian prediction interval
L	likelihood function
$l(\theta, a)$	squared error loss function
m	size of future sample
p	shape parameter of standard generalized power function
$P(X_k\|y)$	predictive density function of X_k
$P(Y_{r+s}\|y)$	predictive density function of Y_{r+s}

$P(Y_s\|y)$	predictive density function of Y_s
$R(\theta, d)$	risk function
$r(d)$	Bayesian risk
$S(y)$	Reliability distribution function
S_t	survival function of pareto distribution
S_t^*	Bayesian estimator of S_t
$T(y)$	complete sufficient statistic for p and one-two sample prediction
$U(X)$	upper limit for Bayesian prediction interval
U	sufficient statistic for p and one-two sample prediction
X_k	k^{th} order statistic from the future sample for two samples prediction
X_s	s^{th} order statistic from the future sample for one sample prediction
Y_1	the first order statistic
Y_n	the order statistic n
Y_i	the order statistic i
$(y, y+\delta y)$	interval of failure rate function
y_f	future response from two-parameter exponential distribution
Y_{r+s}	$(r+s)^{th}$ order statistic of future sample for one sample prediction

$1-\alpha$	confidence prediction coefficient
μ	location parameter of two-parameter exponential distribution
σ	scale parameter of two-parameter exponential distribution
ε	complete sufficient statistic for p and multi samples prediction
δ	sufficient statistic for p and the complete sample
$\hat{a}$	Bayesian estimator of a
$\hat{p}_1$	Bayesian estimator of p for non-informative prior function
$\hat{p}_2$	Bayesian estimator of p for informative prior function
$\hat{b}$	Bayesian estimator of b

LIST OF ABBREVIATIONS

Cdf	cumulative density function
CFR	constant failure rate
DFR	decreasing failure rate
IFR	increasing failure rate
LINEX	linear exponential loss function
MMAE	minimum mean absolute error
MMSE	minimum mean squared error
PD	highest predictive density
pdf	probability density function

LIST OF PUBLICATIONS

1. **Al Mutairi Aned O.** and Heng Chin Low, Bayesian Estimate for Shape Parameter from Generalized Power Function Distribution, Journal of Mathematical Theory and Modeling. (2012), Vol.2, No.12, ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online).

2. **Al Mutairi Aned O.** and Heng Chin Low, Bayesian Estimates of Parameters for Life Distributions based on Different Loss Functions, Advances in Applied Mathematical Analysis Journal. (2013), Vol. 8 Number 1, ISSN 0973-5313 p.11-17.

3. **Al Mutairi Aned O.** and Heng Chin Low, Concepts in Order Statistics and Bayesian Estimation, Journal of Mathematical Theory and Modeling. (2013), Vol.3, No.4, ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online).

SATU KAEDAH DIUBAHSUAI UNTUK RAMALAN BAYESIAN BAGI STATISTIK TERTIB MASA DEPAN DARIPADA FUNGSI KUASA TERITLAK

ABSTRAK

Statistik Bayesian adalah kaedah statistik yang digunakan secara meluas dalam pelbagai bidang seperti perubatan, sains sosial dan ekonomi. Ramalan Bayesian adalah salah satu kaedah statistik Bayesian. Ia bekerja dengan pelbagai kaedah. Kajian ini turut membincangkan tiga kaedah ramalan Bayesian. Terdapat satu sampel, dua sampel dan sampel pelbagai ramalan. Tiga kaedah bertindak balas terhadap motivasi praktikal yang memerlukan data sampel kurang, dalam kebanyakan kajian statistik digunakan. Oleh itu, sampel yang akan datang adalah istilah yang penting dalam tesis ini. Pendekatan Bayesian menggunakan ramalan untuk statistik tertib masa depan berdasarkan data tertib yang diperhatikan dan fungsi ketumpatan ramalan memberi selang ramalan Bayesian untuk statistik tertib masa depan. Taburan fungsi kuasa teritlak piawai adalah dasar untuk tiga kaedah dengan menggunakan teori Bayes untuk mencapai had rendah dan had atas yang sempit bagi selang ramalan Bayesian 95% dan selang ramalan Bayesian 99%. Selang yang dicadangkan menyumbang dalam peningkatan ketepatan untuk nilai ramalan. Prestasi tiga kaedah dinilai oleh had rendah dan had atas dari sampel yang diperhatikan dan untuk statistik tertib dari sampel masa depan. Dua jenis fungsi priori digunakan dalam ramalan Bayesian: prior bermaklumat dan prior bukan bermaklumat. Analisis berangka menunjukkan had rendah dan had atas bagi selang

ramalan Bayesian 95% dan selang ramalan Bayesian 99% untuk tiga kaedah, dan set data dihasilkan daripada taburan fungsi kuasa teritlak piawai.

Anggaran Bayesian digunakan untuk parameter bentuk, parameter skala dan parameter lokasi. Penganggar Bayesian dicadangkan sebagai min taburan posterior berdasarkan fungsi prior bermaklumat atau fungsi prior bukan bermaklumat. Kedua-dua fungsi prior menggunakan rumus bagi taburan posterior dari teori Bayes untuk menggabungkan fungsi kebolehjadian dan fungsi prior. Penganggar Bayesian yang dicadangkan ialah min taburan posterior berdasarkan taburan fungsi kuasa teritlak piawai dan fungsi kehilangan ralat kuasa dua. Selain teknik ini, kriteria Bayesian digunakan. Prestasi penganggar bentuk, skala dan lokasi dinilai oleh beberapa jenis taburan prior dan digunakan dalam masa yang sama dengan ramalan Bayesian, yang jika dibandingkan, mengesahkan kewajaran dan kelebihan dari beberapa jenis taburan prior untuk anggaran atau ramalan menggunakan kaedah Bayesian. Analisis berangka menunjukkan penganggar yang dicadangkan oleh set data yang dihasilkan daripada taburan fungsi kuasa teritlak piawai.

A MODIFIED METHOD FOR BAYESIAN PREDICTION OF FUTURE ORDER STATISTICS FROM GENERALIZED POWER FUNCTION

ABSTRACT

Bayesian statistics is a statistical method that is widely used in many fields, including medicine, social and applied sciences. These fields occasionally have little or limited information about their populations. Therefore, using new techniques that require fewer samples while providing the same quality as the case of available samples is necessary. Bayesian prediction is a commonly used tool in Bayesian statistics. This study modifies three Bayesian prediction methods: one-, two- and multi-sample predictions. Bayesian prediction modified method does not require the many samples. Therefore, a future sample is a significant term in this thesis. Our Bayesian prediction modified method used a prediction for the future order statistics based on the observed ordered data, and predictive densities provided the Bayesian prediction intervals for the future order statistics. The standard generalized power function distribution serves as the basis for the three modified methods by applying Bayes' theory to achieve close lower and upper limits for the 95% and 99% Bayesian prediction intervals. The proposed intervals contributed to increasing the precision for the predictive value. The performance of the three modified methods is evaluated using the lower and upper limits from the observed sample and for the order statistic from the future sample. Two types of prior functions are used in Bayesian prediction: informative and non-informative priors, both of which use Bayes' theory. The numerical analysis illustrates lower and upper limits for the 95% and 99% Bayesian

prediction intervals for the three modified methods, and the data set generated from the standard generalized power function distribution. Bayesian estimation is used to determine the shape, scale and location parameters. Bayesian estimators are suggested as the mean of the posterior distribution based on an informative or non-informative prior function. Both prior functions use a formula for the posterior distribution from Bayes theory to combine the likelihood function and prior function. The proposed Bayesian estimator is the mean of the posterior distribution based on the standard generalized power function distribution and a squared error loss function. In addition to this technique, a Bayesian criterion is used. The performance of the shape, scale and location estimators are evaluated with some types of prior distributions and used simultaneously with the Bayesian prediction, which, when compared, confirms the suitability and advantage of some types of prior distributions for estimation or prediction using the Bayesian method. The numerical analysis illustrates the proposed estimators derived from the data set generated from the standard generalized power function distribution.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

The Reliability distribution function analysis $S(y)$ is used for estimation and prediction. It was formulated during the early research in the science of insurance, life schedules and mortality. Modern analysis of the reliability distribution $S(y)$ began approximately 50 years ago in engineering applications. The incentives to study this probability function increased in World War II when studying the failure rates of war machines led to a marked improvement in them. After World War II, the progress made in increasing the lifetimes of war machines was used for machines in civil applications. Researchers initially focused on the parametric approach for random variables that follow a known distribution, such as the normal, exponential, Weibull or gamma distributions; thereafter, they focused on the prediction and testing of hypotheses for parameters of these distributions. When interest of the prediction and testing of hypotheses in medical and biological research increased, so did interest in using a non-parametric approach. Both non-parametric and parametric approaches concerning positive random variables are used in estimation and prediction.

1.1.1 Future sample and present sample

Statistical inference from a Bayesian perspective usually requires less sample data to achieve the same quality of inference compared with sampling theory-based methods. In many cases, this is the practical motivation for using a future sample. Aitchison and Dunsmore (1975) presented a medical example to explain the concept of future sample.

Example 1.1

Table 1.1. Survival times (weeks) of 20 carcinoma patients medical diagnosis

25	45	238	194	16	23	30	16	22	123
51	412	45	162	14	72	5	35	30	91

The data in Table 1.1 are the survival times (weeks) of 20 patients suffering from a certain type of carcinoma and receiving treatment of preoperative radiotherapy followed by radical surgery. On the basis of this information, what can appropriately be said about the future of a new patient with this type of carcinoma and assigned to this form of treatment? Clearly any rational statement would regard 100 weeks survival as much more plausible than 500 weeks survival, but how should such views be summarized and quantified? What is a reasonable assessment of the probability that the patient will survive 100 weeks?

In this example the informative experiment e consists of recording the survival times of the 20 patients already treated. The future experiment f consists of treating a new patient similarly and recording his survival time. If no change in the treatment has been made since conducting e, then e and f consist of 20 replicates and a single replicate of the same basic trial (record the survival time of a treated patient), respectively and are independent.

Attempts to quantify medical prognoses are of vital importance when similar information on an alternative treatment, for example, radical surgery followed by postoperative radiotherapy, is available and a choice has to be made between treatments for a particular patient.

In life testing, it is possible to predict age of survival of observations or age of survival for all systems because we have the sample space, random variables and distribution functions. But sometimes the information of the experiment is not complete. Therefore, there is a necessity to use new methods that provide some new information, especially when the sample is not found.

Bayesian prediction can treat the problem of absent sample (future sample) by using the present sample and can assume some assumptions which relate between random variable X from present sample and random variable Y from future sample as an interval which is given as follows:

$$Pr\big(L(X) < Y < U(X)\big) = 1 - \alpha, \tag{1.1}$$

where $L(x)$ is the lower prediction bound and $U(x)$ is the upper prediction bound from the present sample.

$(L(X)\ ,\ U(X)\ )$ is a $100(1-\alpha)\%$ Bayesian prediction interval for a future random variable Y if $1-\alpha$ is called the confidence prediction coefficient (Alamm *et al.,* 2007). However, Gianola and Fernando (1986) presented the concept of $100(1-\alpha)\%$ Bayesian confidence interval from Bayesian prediction interval which plays an essential role in the posterior function if we have the following:

$$Pr(\theta \in R|x) = \int f(\theta|x)d\theta, \tag{1.2}$$

where R is a region of the space of parameter θ. Fixing the probability in Equation (1.2) at say $(1-\alpha)$, for a given α, it is possible to obtain an interval for θ such that its probability content is $(1-\alpha)$.

When we choose a specific distribution, we can obtain estimates of parameters or Bayesian prediction intervals. Kamps (1995) and Kamps and Gather (1997) introduced the concept of generalized order statistics as a unified approach to various ordered random variable models, such as last order statistics, sequential order statistics, and last observed values. After that emerged the concept of censoring schemes. Many authors have used censored order statistics in their works (Burkschat *et al.,* 2006, Howlader and Hossain, 2002). In addition, Soliman (2000) used the order statistics censoring with Rayleigh distribution and presented its estimators. Then the researchers started thinking about future order statistics as Fernandez (2000a) and Alamm *et al.* (2007).

1.1.2 Order statistics

Order statistics have a major role in statistical inference, especially in laboratory methods. If $Y=(Y_1,Y_2,\dots,Y_n)$ is a random sample selected from a population, then the probability density function is $f(y)$ and its cumulative distribution function is $F(y)$. If we arrange these observations in increasing order, we obtain the following:

$$Y_1<Y_2<\cdots<Y_i<\cdots<Y_n \tag{1.3}$$

where Y_1 is the smallest observation and is called the first order statistic,

Y_n is the largest observation and is called the n^{th} order statistic,

and Y_i is the i^{th} order observation and is called the i^{th} order statistic.

Thus, $(Y_1, Y_2, \ldots, Y_n)$ are the order statistics and they yield the probability density function for the order statistics as follows:

$$g(y_i) = \frac{n!}{(i-1)!\,(n-i)!}[F(y)]^{i-1}[1-F(y)]^{n-i}f(y) \tag{1.4}$$

If $i = 1$, the probability density function for the (smallest) first observation is

$$g(y_1) = n\,[1-F(y)]^{n-1}f(y) \tag{1.5}$$

However, if $i = n$, the probability density function for the largest observation is

$$g(y_n) = n\,[F(y)]^{n-1}f(y) \tag{1.6}$$

(Hogg and Craig, 2013).

1.1.3 Aging

The concept of aging is a basic concept in statistics, especially in reliability, and it has many applications in engineering, physics, and biology. Aging has three types:

i. Positive aging with time, where the unit age reduces with the progress of time. Over time, this type of aging may lead to the erosion of industrial units, which requires plans for maintenance or replacement. In addition, the passage of time may result in living organisms, including humans reaching old age, and this possibility requires the development of plans for remediation**.**

ii. Fixed aging, which is not affected by time. An example of fixed aging is electronic industrial units that follow the exponential model of a failure rate that is fixed in time**.**

iii. Negative aging with time, which leads to the improvement of an industrial unit after the beginning of the operation of new units. In that case, the proper units will remain and defective units will fail at the beginning of their industrial operation. As for the biological side, children who are healthy at birth live and become stronger over time, whereas unhealthy children die after birth (Hendi and Sultan, 2004).

Some basic functions that accompany aging are defined as follows:

Definition 1.1

If Y is a positive random variable, the reliability distribution function $S(y)$ after age Y is defined as follows:

$$S(y) = \overline{F}(y) = 1 - F(y) = P(Y > y) \tag{1.7}$$

Definition 1.2

If Y is a positive random variable, then the cumulative distribution $F(y)$ is defined as:

$$F(y) = Pr(Y \leq y) \tag{1.8}$$

Definition 1.3

If Y is a positive random variable, the probability density function $f(y)$ is defined as

$$f(y) = \frac{d}{dy}F(y) \tag{1.9}$$

(Hogg and Craig, 2013).

Definition 1.4

If Y is a positive random variable, the failure rate function $h(Y)$ during the interval $(y, y + \delta y)$ is defined as:

$$h(Y) = \lim_{\delta \to 0} Pr(y < Y < y + \delta y | Y > y) = \frac{f(y)}{S(y)}, \quad Y > 0 \tag{1.10}$$

The failure rate function, $h(Y)$, has three important cases in all applications (whether engineering or biological) through the following Equation:

$$h(Y) = kc\lambda Y^{c-1} + (1-k)bY^{b-1}\theta e^{Y^{k\theta}},$$

$$\theta, b, c, \lambda > 0, \quad 0 \leq k \leq 1, \quad Y > 0 \qquad (1.11)$$

The previous Equation (1.11) takes a U-shaped curve with the Y-axis of time as follows:

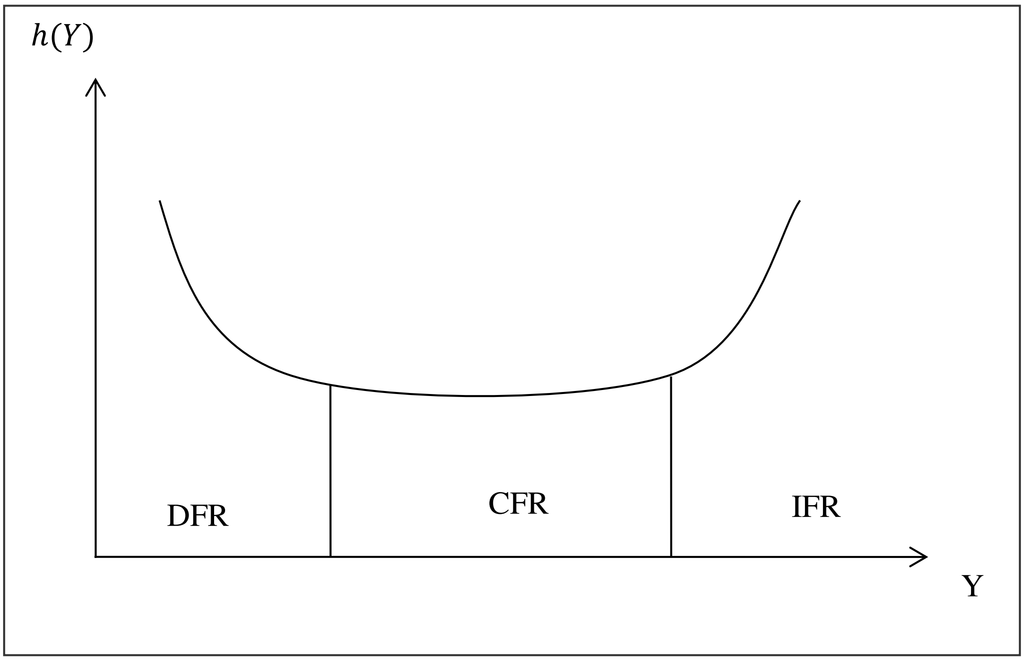

Figure 1.1. Curve of the failure rate function

From Figure 1.1, the curve of the failure rate function is described in $h(Y)$ by Equation (1.11).

i. The first part, DFR, of the U-shaped curve indicates that the failure rate function, $h(Y)$, is decreasing with time Y. This process is symbolized by the Decreasing Failure Rate (DFR), as the value of the failure rate function, $h(Y)$, is high at the beginning of the operation of industrial units because of the presence of manufacturing defects in raw materials, machines, workers, the quality of production, or in terms of biology (including human) because of the presence of birth defects after birth.

ii. The second part, CFR, of the U-shaped curve of the failure rate function, $h(Y)$, is a constant function over time Y and is equal to a fixed amount c, which is not affected by time and is symbolized by the Constant Failure Rate (CFR) and explained practically by uncontrolled reasons. Uncontrolled reasons include accidents representing sudden high-voltage power in the electricity source that is due to a default in the main power station; such accidents result in the destruction of houses or infant mortality, which can occur when infants drink poisonous substances that are intended for cleaning or chase a ball into a street in front of a speeding car.

iii. The third part, IFR, of the U-shaped curve shows that the failure rate function, $h(Y)$, is increasing over time Y, which is symbolized by the Increasing Failure Rate (IFR). Indeed, the value of the failure rate function, $h(Y)$, is small at the beginning with respect to time Y, and then, this function increases over time. This phenomenon is explained practically for reasons of aging, antiquity, and corrosion of industrial machines or in terms of biological (including human) factors for reasons of aging (Hendi and Sultan, 2004).

1.1.4 Bayesian statistic

In some statistical applications, it is preferred to use statistical inferences from the perspective of decision science through the study of loss associated with the sequence of possible decisions, e.g., patterns of decision of a buyer, investor, or institution.

The formulation of decision-making using a probabilistic model requires a mathematical structure in a general form. Bayesian statistical inferences and the corresponding mathematical structure are formulated as follows:

a) Group A represents action space where A contains every method possible $a \epsilon A$.

b) Group Θ represents parameter space where Θ contains every level possible $\theta \in \Theta$.

c) Group R_y is a partial group from real numbers set R representing space of Y for every value $Y = y$ to ensure $y \in R_y$; thus, y has probability density function related to the distribution family $\{f(y, \theta); \theta \in \Theta\}$.

d) Group d represents the decision space and contains every possible decision $d \epsilon D$.

(Hendi and Sultan, 2004)

The decision maker, when choosing the action a, should note some connotations from the population that provide him with information on the unknown parameter θ. Furthermore, the decision maker studies the random variable Y, where $Y = y$ and y is a real value from R, as well as the probability density function of the random variable Y, namely, $f(y, \theta)$, which provides him with some guidance on the parameter θ that will assist him in the selection of the action a. Decision theory focuses on the properties and methods that demonstrate how to make such a choice. The parameter θ, where $\theta \in \Theta$, is found in nature without any intervention from the decision maker and without any knowledge of its value having been estimated by the decision maker with the action $a \in A$. A loss arises from this estimation, and the loss's function is denoted by the symbol $l(\theta, a)$. It is appropriate for the decision maker to choose the random variable, Y, which has a density function $f(y, \theta)$, to provide him with some information about the nature of $\theta \in \Theta$. The selection of the decision function $d(y)$ represents the plan followed by the decision maker in determining this nature.

The Bayesian estimate and tests of hypotheses are only special cases of issues of public decisions. In the Bayesian estimate, we set $A = \Theta$ (in some cases, it is the real number set or an interval from it). The required step is to select the appropriate value for the parameter θ.

(Hendi and Sultan, 2004)

Definition 1.5

The linear loss function $l(\theta, a)$ is defined as follows (Berger, 1985):

$$l(\theta, a) = \begin{cases} c_1(\theta - a), & if\ \theta - a \geq 0 \\ c_2(a - \theta), & if \theta - \ a < 0, \end{cases} \tag{1.12}$$

where (c_1, c_2) are constants, a is any possible action, and θ is a parameter.

The loss function depends on the nature of the problem under study. The values of (c_1, c_2) are selected for positive real numbers set (the selection is conducted to reflect an overhead cost or a lower cost). If $c_1 = c_2 = 1$, the linear loss function becomes $l(\theta, a) = |\theta - a|$. In this case, l is defined as the absolute error loss function.

The squared error loss function is defined, which logically represents the loss and is frequently used in other researches (Berger, 1985).

Definition 1.6

The squared error loss function can be defined as follows:

$$l(\theta, a) = (\theta - a)^2 \qquad (1.13)$$

where a is any possible action and θ is a parameter.

Although most of the deduction operations of the Bayes estimators are performed by using the squared error loss function, the estimators themselves can be derived by what is known as the linear exponential loss function (LINEX) (Berger, 1985).

Definition 1.7

The LINEX loss function is defined as follows:

$$l(\Delta) = [e^{a\Delta} - a\Delta - 1], \qquad a \neq 0 \qquad (1.14)$$

where $\Delta = \frac{\hat{\theta}}{\theta} - 1$, $\hat{\theta}$ is the estimator of θ and a is a constant.

The positive value of a is used when overestimation is more serious than under estimation, whereas the negative value of a is used in the reverse situation. (Soliman, 2000)

1.1.4.1 Prior distributions

The prior distribution function $h(\theta)$ is used to explain the information about the parameter θ before taking the sample $Y = (Y_1, Y_2, \dots, Y_n)$ from a population that has a distribution $f(y, \theta)$. The prior distribution function $h(\theta)$ is classified into two types:

i. The non-informative prior distribution function: This function occasionally represents vague information, which means cases where the prior information about the parameter θ is relatively scant (or very limited), Therefore, we may use the quasi-density prior in the form:

$$h(\theta) \propto \frac{1}{\theta^d}\,, \quad \theta > 0, \qquad d > 0 \tag{1.15}$$

If $d = 1$, we obtain a non-informative prior function

$$h(\theta) \propto \frac{1}{\theta}\,, \quad \theta > 0 \tag{1.16}$$

Also, if $d = 3$, we obtain the asymptotically invariant prior

$$h(\theta) \propto \frac{1}{\theta^3}\,, \qquad \theta > 0 \tag{1.17}$$

(Soliman, 2000, and Yang and Berger, 1998)

ii. The informative prior distribution function: This function is used when the information about θ is not vague, and this function is occasionally called the conjugate prior when the prior information $h(\theta)$ is comparable with the probability function $l(y,\theta)$ for the population taken from the sample $Y = (Y_1, Y_2, \dots, Y_n)$. For example:

a) The Bernoulli distribution $B(\theta, n)$ is comparable with the beta prior distribution.

$$h(\theta) = \frac{1}{\beta(g,d)}\theta^{g-1}(1-\theta)^{d-1}\,, \qquad g,d > 0, \qquad \theta \in [0,1] \tag{1.18}$$

b) In the case of a Poisson distribution $P(\theta)$ and a standard generalized power function the gamma prior distribution is proportional.

$$h(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\,, \quad 0 < \theta \tag{1.19}$$

c) In the case of a normal distribution $N(\mu, \sigma)$, with unknown mean μ and known variance σ^2 this is to be proportional to a prior distribution of the following normal distribution family:

$$h(\theta) = \frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{-1}{2\sigma^2}\right)(\theta-\mu)^2}, \qquad \theta \in R, -\infty < \mu < \infty, \ \sigma^2 > 0 \tag{1.20}$$

(Lee, 2012)

1.1.4.2 Posterior distributions

First, consider a case that has continuous values and is related to some of the previous cases.

If (Y, θ) has a joint probability density function $f(y, \theta)$, we can define the joint density function from standard probability theory as follows:

$$f(\theta, y) = f(y|\theta)f(\theta) \tag{1.21}$$

$$f(\theta, y) = f(\theta|y)f(y) \tag{1.22}$$

where $f(\theta)$ and $f(y)$ represent the marginal density functions of θ and y, respectively. From Equations (1.21) and (1.22),

$$f(\theta|y) = f(y|\theta)f(\theta)/f(y) \tag{1.23}$$

Note that:

$$f(y) = \int f(y, \theta)\, d\theta$$

$$= \int f(y|\theta)f(\theta)d\theta \tag{1.24}$$

In Bayesian terminology, $f(\theta)$ is known as the prior density of θ, which reflects the relative uncertainty about the possible values of θ before the data vector $\boldsymbol{Y}$ is realized. The density $f(y|\theta)$ is the likelihood function which represents the contribution of y to knowledge about θ. If we consider $H(\theta|y)$ as a posterior density function and $h(\theta)$ as a prior density function, we can write the posterior density function as follows:

$$H(\theta|y) = \frac{f(y|\theta)h(\theta)}{\int f(y|\theta)h(\theta)d\theta} \tag{1.25}$$

(Gianola and Fernando, 1986).

1.1.4.3 The Bayes risk function

At first glance, it appears that no problem results in the choice of the ideal decision; instead, we choose the decision function d that enables us to obtain the minimum loss regardless of the real value of the parameter level of θ. Unfortunately, choosing the ideal decision is impossible without knowing the real value of the parameter level of θ, but if the real value of the level of θ is known, there is no problem. To demonstrate this fact, we will discuss parameter estimation using θ with an error loss function $l(\theta, a)$ that takes the quadratic form $l(\theta, a) = (\theta - a)^2$. Suppose that $Y = y$, $a = d(y)$ is the chosen estimation, and the real value of the parameter is θ. Then, we suffer from error loss that equals $l(\theta, a) = (\theta - d(y))^2$. If $\theta = \theta_0$, then $d(y) = \theta_0$ is necessary for an ideal decision, but if $\theta = \theta_1$, then we must have $d(y) = \theta_1$. However, because we do not know the value of θ, we cannot choose $d(y)$ to obtain the minimum error loss; thus, mathematicians have disregarded this problem.

The available way to measure the quality of the decision function comes from searching how to improve the average of the error loss function $l(\theta, d(y))$ using the risk function $R(\theta, d)$, whose domain is $\Theta \times D$ and whose range is the real numbers set R. The risk function is defined as follows:

A. In the case of a continuous random variable Y,

$$R(\theta, d) = \int l\big(\theta, d(y)\big) f(y, \theta) dy \tag{1.26}$$

B. In the case of a discrete random variable Y,

$$R(\theta, d) = \sum_{y \in R_y} l\big(\theta, d(y)\big) f(y, \theta) \tag{1.27}$$

where $f(y,\theta)$ is the probability density function and R_y is the space of the random variable Y.

Previous cases means the risk function $R(\theta,d)$ represents the expected error loss for the probability density function $f(y,\theta)$ for the decision maker when he chooses a real level for θ and a decision d. If we symbolize the expected error loss by the risk function, then it can be written in the given form as follows:

$$R(\theta,d) = E_\theta\left(l\big(\theta,d(y)\big)\right) \tag{1.28}$$

The risk function $R(\theta,d)$ expresses the expected error loss from making the decision d provided we know the parameter θ, and if we assume that $R(\theta,d)$ is the distribution function of θ, we know d. Thus, we can calculate the expectation value for the posterior distribution $H(\theta|y)$, and the Bayesian risk $r(d)$ can be defined for the decision d as follows:

A. In the case that θ has continuous values:

$$r(d) = \int R(\theta,d)H(\theta|y)d\theta \tag{1.29}$$

B. In the case that θ has discrete values:

$$r(d) = \sum_{\Theta\in\theta} R(\theta,d)H(\theta|y) \tag{1.30}$$

It is natural to seek the decision function d that provides the minimum value around all error loss averages. For example, we may choose d^* to demonstrate the following rule:

$$r(d^*) = min_{d\in D}\big(r(d)\big) \tag{1.31}$$

The function d^*, if applicable, is called the Bayesian decision function, and $r(d)$ is the Bayesian risk of the decision function d. When given a decision problem, there is no single Bayesian decision function, and furthermore, the obtained answer depends on the choice of the posterior distribution function $H(\theta|y)$ if d is a Bayesian decision function. The function $d(Y)$ is called the Bayes estimator, and $d(y)$ is called Bayes estimate (Berger, 1985).

1.1.5 Bayesian prediction

In statistics, attention is often focused on confidence intervals, or prediction intervals. In the issues of life testing, either the age of survival for the units involved in the test or the age of survival for the system as a whole can be predicted. Moreover, in conventional statistics or Bayesian statistics, the issue of prediction has specific formulas for the distributions of tests of life. Prediction has contributed to the development of statistics. In particular, Bayesian statistical prediction has been useful for its association with Bayesian estimation. For example, future sample, minimum or maximum values, the arithmetic mean, or the range of the future sample can be predicted by the present sample.

1.1.5.1 Types of Bayesian prediction problems

1. One-sample Bayesian prediction

Suppose that $Y_1 < \cdots < Y_r$, where r is the number of units that first failed in a random sample that has size n. Because the failure intervals are identical in distribution with a random variable Y that has the probability density function $f(y|\theta)$, we can predict the intervals of failure in the test for some or all of the remaining order statistics $(n-r)$ to find the ages of survival $Y_{r+1} < \cdots < Y_n$.

2. Two-sample Bayesian prediction

Suppose that $Y = (Y_1, Y_2, \dots, Y_n)$ are random variables in a sample that has size n and probability density function $f(y|\theta)$. Furthermore, suppose that $X = (X_1, X_2, \dots, X_m)$ is an independent random sample that has size m and has future units of the same distribution. Then, the prediction order statistics X_s will depend on the known sample $Y = (Y_1, Y_2, \dots, Y_n)$.

3. Multi-sample Bayesian prediction

Suppose that we have an independent sequence $n_1, n_2, \dots, n_\omega$ of random samples and the probability density function is $f(y|\theta)$. The multi-sample prediction can be obtained from the order statistics of a future sample based on previous samples.

1.1.5.2 Bayesian prediction intervals

The Bayesian prediction interval for the order statistic X_s of the future sample $X = (X_1, X_2, \dots, X_m)$ using the known sample $Y = (Y_1, Y_2, \dots, Y_n)$ can be determined using other methods than the Bayesian methods. Suppose that $f(y, \theta)$ is a probability density function of a random sample with a size of n units and has a random variable Y, where $Y = (Y_1, Y_2, \dots Y_n)$ and $X = (X_1, X_2, \dots, X_m)$ is another random sample of size m that represents the future units of the same distribution and that $f(x_s|\theta)$ is the probability density function of the Bayesian prediction for the order statistic X_s, where $1 \le s \le m$. Then, the function has the following form:

$$P(x_s|y) = \int f(x_s|\theta) H(\theta|y) d\theta, \tag{1.32}$$

where $H(\theta|y)$ is the posterior density function of the parameter θ given Y.

The lower $L(y)$ and upper $U(y)$ Bayesian prediction limits for the order statistic X_s can be calculated by using the survival function $Pr[L(y) < X_s < U(y)|y]$. This function can be defined as follows:

$$Pr[L(y) < X_s < U(y)|y] = \gamma \tag{1.33}$$

First, we find a solution for Equation (1.33) for the lower limit $L(y)$:

$$Pr[X_s > L(y)|y] = \frac{1+\gamma}{2} \tag{1.34}$$

Second, we find a solution to Equation (1.33) for the upper limit $U(y)$:

$$Pr[X_s > U(y)|y] = \frac{1-\gamma}{2} \tag{1.35}$$

Both $100\ (1-\alpha)\%$ of the lower limit and the upper limit of the Bayesian prediction interval can be obtained by solving Equation (1.33), where $\gamma = 1-\alpha$ and the values of α are $\alpha = 0.01$ and 0.05 (Alamm *et al.,* 2007).

1.2 Problem identification and the importance of the study

In life testing, it is typically postulated that there are n units identical and independent under testing in a specific experiment and that the respective times of failure for these units are recorded. In this case, the continuation in this experiment until the failure of all units would not be practical, particularly if the sample was large or units are expensive. Thus, it is best to stop the experiment after obtaining partial information, and this fact distinguishes the field of survival analysis from other fields in statistics, which uses censoring.

It was found that using partial information yields precise estimations that are as accurate as the complete sample of the random variable. But sometimes the information and the partial information of the experiment do not exist. Therefore, there is a necessity to use a new method that provides some new information, especially when the sample is not found. We call this type of sample, future sample.

1.3 Objectives of the study

Main objective

1. To develop a method for Bayesian prediction of future order statistics from generalized power function distribution for one sample, two samples and multi samples.

Sub-objectives

i. To obtain the posterior distribution function for generalized power function distribution.

ii. To obtain Bayesian estimates for shape parameter, location parameter and scale parameter from generalized power function distribution.

iii. To compare between lower and upper limits of 95% Bayesian prediction intervals, and lower and upper limits of 99% Bayesian prediction intervals for future order statistics from generalized power function distribution.

1.4 Organization of thesis

This thesis contains six chapters. Chapter 2 gives an overview of the literature in the field of Bayesian prediction and Bayesian estimation. A new construction for Bayesian prediction based on the observed ordered data from the predictive densities is provided in Chapter 3. This is used to determine Bayesian prediction intervals for future order statistics for one-sample, two-samples and multi-samples from generalized power function by informative prior distribution and non-informative prior distribution. Chapter 4 discusses the proposed Bayesian estimators for shape parameter, scale parameter and location parameter by using squared error loss function in the case of informative prior distribution and non-informative prior distribution. A numerical comparison between the Bayesian prediction confidence intervals on 95% and 99% is also performed and it is presented in Chapter 5. Chapter 6 gives the summary and conclusions of the study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this Chapter, the Pearson's system is first introduced. The derivation of the generalized power function distribution from one of its phases is then explained. Different forms of the probability density function for the generalized power function distribution are considered.

Furthermore, this chapter provides an overview of important works on Bayesian prediction intervals in the context of several distributions, multiple prior distributions and additional confidence intervals. In particular, Bayesian prediction intervals by one-sample, two-samples and multi-samples in two cases using an informative prior distribution and non- informative prior distribution are discussed.

Finally, an overview on the Bayesian estimation by different loss functions is provided for some parameters and functions, as well as how to obtain good estimators in place of original parameters.

2.2 Pearson's system

The Pearson's system is one in which the random variable is continuous, and all of the elements of its probability density function satisfy the differential equation $df/dx = -(x+a)f(x)/(b_0 + b_1 x + b_2 x^2)$. The shape of the differential equation depends on the values of the parameters b_0, b_1, b_2 and a. If $b_1 = b_2 = 0$, the differential equation becomes $d \log f(x)/dx = -(x+a)/b_0$, hence $f(x) = Q \exp\left[-\frac{(x+a)^2}{2b_0}\right]$, where Q is a constant chosen so that $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (Johnson *et al.,* 1994).

The Pearson's system contains 12 phases known as density functions of the well known distributions. Table 2.1 shows the Pearson phases for the well known distributions.

Table 2.1. Phases of the Pearson's system

Distribution name	Pearson phase	Density function
Beta	I	$(1+x)^{m_1}(1-x)^{m_2}$
	II	$(1-x^2)^m$
Gamma	III	$x^m \exp(-x)$
	IV	$(1+x^2)^{-m} \exp(-\nu \tan^{-1} x)$
Inverse Gamma	V	$x^{-m} \exp(-x^{-1})$
Inverse Beta (F)	VI	$x^{m_2}(1+x)^{-m_1}$
T	VII	$(1+x^2)^{-m}$
	VIII	$(1+x)^{-m}$
	IX	$(1+x)^m$
Exponential	X	$\exp(-x)$
Pareto	XI	x^{-m}
	XII	$\left[\frac{(1+x)}{(1-x)}\right]^m$

(Source: Johnson *et al.,* 1994).

The most important items in Table 2.1 are Pearson phases I and II, which take the form of Beta distributions. The family of Beta distributions consists of all the distributions that have the probability density function given as:

$$f(y) = \frac{1}{\beta(p,q)} \frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}}, a \le y \le b, a, b > 0, p > 0, q > 0 \quad (2.1)$$

A Beta distribution is denoted by the symbol $\beta\ (p,q)$.

Using $x = (y - a)/(b - a)$, Equation (2.1) will take the standard formula of the Beta density function as follows:

$$y = (b - a)x + a$$

Substituting y value in Equation (2.1)

$$f(x) = \frac{1}{\beta(p,q)} \frac{((b-a)x + a - a)^{p-1}(b - ((b-a)x + a))^{q-1}}{(b-a)^{p+q-1}}$$

$$= \frac{1}{\beta(p,q)} \frac{((b-a)x)^{p-1}((b-a) - (b-a)x)^{q-1}}{(b-a)^{p+q-1}}$$

$$= \frac{1}{\beta(p,q)} \frac{x^{p-1}(1-x)^{q-1}}{(b-a)}$$

If $b = 1$ and $a = 0$, then the standard Beta density function is:

$$f(x) = \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1}, \quad 0 \le x \le 1 \tag{2.2}$$

If $q = 1$ in Equation (2.2), then the standard power probability density function has the following formula:

$$f(x) = px^{p-1}, \quad 0 \le x \le 1, \quad p > 1 \tag{2.3}$$

setting $q = 1$ in Equation (2.1) representing the Beta distribution, a distribution known as the generalized power function distribution, and its probability density function takes the following formula:

$$f(x) = \frac{p}{(b-a)^p}(x-a)^{p-1}, \; p > 1, \quad a \le x \le b, \quad a, b > 0 \tag{2.4}$$

(Johnson *et al.*, 1994)

2.2.1 Different forms of the probability density function for the generalized power function distribution

(a) The first form:

The generalized power function distribution was identified as the generalized geometric distribution

$$f(y) = \frac{p}{\sigma b^p}\left(\frac{y-\theta}{\sigma} + a\right)^{p-1}, \quad p \geq 1, \qquad \theta - a\sigma \leq y \leq \theta + (b-a)\sigma \qquad (2.5)$$

where a and b are defined as:

$$b = (p+1)\sqrt{\frac{p+2}{p}}, \qquad a = \sqrt{p(p+2)}, \qquad p \geq 1 \qquad (2.6)$$

(Sultan *et al.*, 2002)

By setting $\theta = 0, \sigma = 1$ in Equation (2.5), the probability density function for the standard generalized power function distribution is obtained. This form was discussed in Sultan *et al.* (2002).

$$f(y) = \frac{p}{b^p}(y+a)^{p-1}, \qquad -a \leq y \leq b-a, \qquad p > 1, \qquad a, b \geq 1 \qquad (2.7)$$

Figure 2.1 illustrates the probability density function for the standard generalized power function with a known scale parameter $b = 1$ for several values of the shape parameter p and location parameter a of standard generalized power function.

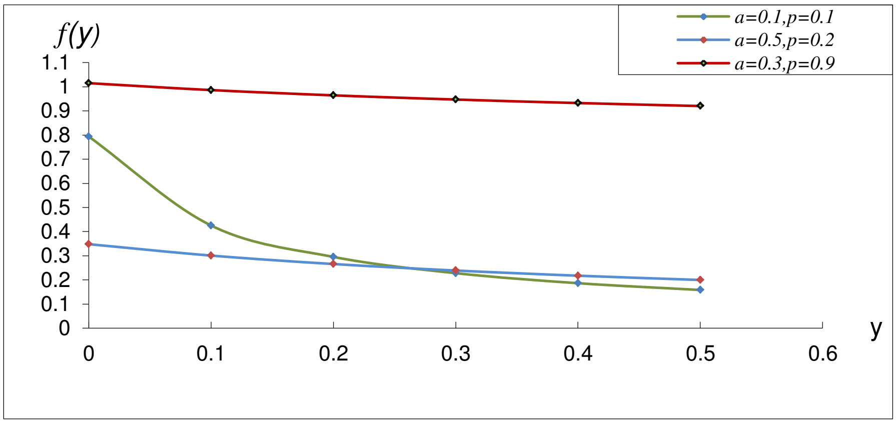

Figure 2.1. Probability density functions for the standard generalized power function distribution with a known scale parameter ($b = 1$).

(b) The second form:

The second form is seen in Equation (2.4), which expresses the generalized power function distribution. It can be proven that this Equation is equivalent to Equation