BAYESIAN ANALYSIS OF WEIBULL DISTRIBUTION AND ITS APPLICATIONS

Dissertation Submitted

In partial fulfillment for the award of degree of
Master of Philosophy
in
STATISTICS

By

KAISAR AHMAD

Under the Supervision of
Dr. SHEIKH PARVAIZ AHMAD
Sr. Assistant Professor

DEPARTMENT OF STATISTICS
UNIVERSITY OF KASHMIR
Hazratbal, Srinagar, J&K, India
NAAC Accredited Grade "A"

2013

DEPARTMENT OF STATISTICS

University of Kashmir,
Hazratbal, Srinagar 190006

Dr. Sheikh Parvaiz Ahmad
(Ph.D., UoK)
Sr. Assistant Professor
Web Add:kashmiruniversity.net

E-mail: sprvz@yahoo.com
Tel.:0194-2416589(Res.)
Telefax:0194-2421357

Dated: _____

CERTIFICATE

This is to certify that the scholar **Mr. Kaisar Ahmad,** has carried out the present dissertation entitled **"Bayesian Analysis of Weibull Distribution and its applications"** under my supervision and the work is suitable for submission for the award of degree of Master of Philosophy in Statistics. It is further certified that the work has not been submitted in part or full for the award of M.Phil or any other degree.

Dr. Sheikh Parvaiz Ahmad
Supervisor

Dedicated

To My

Beloved Parents

ACKNOWLEDGEMENT

In the completion of this dissertation, there are countless debits incurred that can never be adequately repaid, yet the most appropriate thing is to acknowledge. First of all my thanks goes to my esteemed supervisor ***Dr. Sheikh Parvaiz Ahmad*** *whose keen interest, clear thinking, sympathetic nature and insistence on better quality inspired me to work hard.*

I wish to express deep sense of gratitude to ***Prof. Aquil. Ahmad*** *(Head, Department of statistics). It is all because of his open mindedness, objective way of thinking, deep insight in research that I came to analyze, interpret in a proper perspective. Without his sincere devotion, constant encouragement, valuable guidance and sustained interest in my work, the study would not have been completed.*

I wish to place my affectionate thanks to Dr. M.A.K Baig and Dr. Tariq Rashid Jan for their help and valuable suggestions.

No words would be sufficient to express my gratitude to all my fellow scholars Javaid Ahmad Reshi, Kowsar Fatima, Raja Sultan Ahmad Reshi, Hummara Sultan, Adil Rashid, Adil Hamid Khan, Javeed Ahmad, Shabir Ahmad Chopan and Afaq Ahmad for their constant support, encouragement and memorable company.

I feel a deep sense of gratitude for my parents, who taught me the value of hard work by their own example. I would like to share this moment of happiness with my father, mother, brother and sister. They rendered me enormous support during the whole tenure of my research.

Kaisar Ahmad

Preface

The Bayesian method of reasoning is currently riding a high tide of popularity in virtually all areas of statistical application. A distinctive feature of Bayesian inferences is that it takes explicit account of prior information in the analysis. Classical statistical inference, based on sampling theory, usually does not consider information beyond the sample data. The Bayesian use of relevant past experience, which is quantified by a prior distribution, produces more informative inferences in those cases where the prior distribution accurately reflects the variation in the unknown parameter. In addition, the Bayesian method usually requires less sample data to achieve the same quality of inferences than the method based on sampling theory. In many cases this is the practical motivation for using a Bayesian approach and represents the practical advantage in the use of prior information. In Bayesian Statistics, the posterior distribution summarizes the current state of knowledge about all the uncertain quantities including unobservable parameters. The dissertation is organized in the form of five chapters and bibliography.

Chapter I *is introductory in nature and presents introduction to classical & Bayesian statistics. Bayes theorem, sequential nature of Bayes theorem, marginal and conditional inferences, prior and different types of priors, loss function and various types of loss functions have been discussed.*

Chapter II *of this dissertation gives the brief introduction and the structural properties of Weibull distribution.*

Chapter III *deals with the different estimation techniques for estimating parameters of Weibull distribution. The objective of this chapter is to compare maximum likelihood estimation, method of moments and least square estimation methods.*

Chapter IV *deals with the Bayesian estimation of Scale parameter of Weibull distribution under various priors by using different loss functions. This chapter also includes Bayesian credible intervals for the parameters of Weibull distribution. The methods are illustrated with the help of examples.*

Chapter V *considers an application of Weibull distribution in Survival analysis. The objective of this chapter is to compare the two therapies based on survival functions of the patients assuming Weibull model for each therapy.*

CONTENTS

Chapter No.s	Description	Page No.s
1.	Introduction to Bayesian Statistics	1-14
	Introduction	1-3
	Bayes Theorem	3-5
	Sequential Nature Of Bayes Theorem	5-6
	Marginal and Conditional Inferences	6-7
	Prior distribution	7-12
	Loss function	12-14
	Risk function	14
2.	Characterization of Weibull Distribution	15-24
	2.1 Introduction	15-17
	2.2 Different methods of derivation of Weibull distribution	17-20
	2.3 Moments of two parameter Weibull distribution	20-22
	2.4 Standard deviation	22
	2.5 Coefficient of variation	23
	2.6 Skewness and kurtosis	23-24
3.	Estimation of Weibull Distribution	25-39
	3.1 Introduction	25-26
	3.2 Methods of estimation	26-29
	3.3 Estimation of parameters of Weibull distribution using different methods	29-36
	3.4 Numerical Illustration	36-39

4.	Bayesian Analysis of Weibull Distribution	40-55
	4.1 Introduction	40-41
	4.2 Bayesian estimation of Weibull distribution under Jeffrey's prior by using different loss functions	41-45
	4.3 Bayesian estimation of Weibull distribution under extension of Jeffrey's prior by using different loss functions	45-50
	4.4 Bayesian credible regions for parameters of Weibull distribution	50-52
	4.5 Simulation Study	52-55
5.	Application of Weibull Distribution	56-71
	5.1 Introduction	56-59
	5.2 Model Formulation	59-62
	5.3 Model Compatibility and comparison	62-64
	5.4 Bayes Information criterion	64
	5.5 Numerical Illustration	64-71
	Bibliography	72-77

1.1 Introduction

Statistics is the science that relates data to specific questions of interest. This includes devising method to gather data relevant to the question, methods to summarize and display the data to shed light on the question, and methods that enables us to draw answers to the question that are supported by the data. Data almost always contain uncertainty. This uncertainty may arise from selection of the items to be measured, or it may arise from variability of the measurement process. Drawing conclusions from data is the basis for increasing knowledge about the world, and is the basis for all rational scientific inquiry. Statistical inference gives us methods and tools for doing this despite the uncertainty in the data. The methods used for analysis depend on the way the data were gathered. It is vitally important that there is a probability model explaining how the uncertainty gets into the data.

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

a) ***Frequentist Approach To Statistics*:** It is based on the following ideas:
   - Parameters, the numerical characteristics of the population, are fixed but unknown constants.
   - Probabilities are always interpreted as long run relative frequency.
   - Statistical procedures are judged by how well they perform in the long run over an infinite number of hypothetical repetitions of the experiment.

Probability statements are only allowed for random quantities. The unknown parameters are fixed, not random, so probability statements cannot be made about their value. Instead, a sample is drawn from the population, and a sample statistic is calculated. The probability distribution of the statistic over all possible random samples from the population is determined, and is known as the sampling distribution of the statistic. The parameter of the population will also be a parameter of the sampling distribution. The probability statement that can be made about the statistic based on its sampling distribution is converted to a confidence statement about the parameter. The confidence is based on the average behavior of the procedure under all possible samples.

b) ***Bayesian Approach to Statistics*:** (The Reverend Thomas Bayes first discovered the theorem that now bears his name/. It was written up in a paper An Essay Towards Solving a Problem in the Doctrine of Chances. This paper was found after his death by his friend Richard Price, who had it published posthumously in the Philosophical Transactions of the Royal Society in 1763). Bayes showed how inverse probability could be used to calculate probability of antecedent events from the occurrence of the consequent event. His methods were adopted by Laplace and other scientists in the 19th century, but had largely fallen from favor by the early 20th century. By mid 20th century interest in Bayesian methods was renewed by De Finetti, Jeffreys, Savage, and Lindley, among others. They developed a complete method of statistical inference based on Bayes' theorem. The ideas that form the basis of Bayesian approach are:

- Since we are uncertain about the true value of the parameters we will consider them a random variable.
- The rules of probability are used directly to make inferences about the parameters.
- Probability statements about parameters must be interpreted as "degree of belief." The prior distribution must be subjective. Each person can have his/her own prior, which contains the relative weights that person gives to every possible parameter value. It measures how "plausible" the person considers each parameter value to be before observing the data.
- We revise our beliefs about parameters after getting the data by using Bayes' theorem. This gives our posterior distribution which gives the relative weights we give to each parameter value after analyzing the data. The posterior distribution comes from two sources: the prior distribution and the observed data.

This has a number of advantages over the conventional frequentist approach. Bayes' theorem is the only consistent way to modify our beliefs about the parameters given the data that actually occurred. This means that the inference is based on the actual occurring data, not all possible data sets that might have occurred, but didn't! Allowing the parameter to be a random variable lets us make probability statements about it, posterior to the data. This contrasts with the conventional approach where inference probabilities are based on all possible data sets that could have occurred for the fixed parameter value. Given the actual data there is nothing random left with a fixed parameter value, so one can only make confidence statements, based on what could have occurred. Bayesian statistics also has a general way of dealing with a

nuisance parameter. A nuisance parameter is one which we don't want to make inference about, but we don't want them to interfere with the inferences we are making about the main parameters. Frequentist statistics does not have a general procedure for dealing with them. Bayesian statistics is predictive, unlike conventional frequentist statistics. This means that we can easily find the conditional probability distribution of the next observation given the sample data.

1.2 Bayes' Theorem

The foundation of Bayesian statistics is Bayes' theorem. Bayesian probability or Bayesian theory is named after Thomas Bayes, who proved a special case of what is called Bayes' theorem. Consider an unobservable vector θ and observable vector x of length k and n, respectively with their density $p(\theta , x)$. From standard probability theory, we have

$$p(\theta, x) = p(\theta \mid x)\, p(x) \qquad (1.2.1)$$

$$p(\theta, x) = p(x \mid \theta)\, p(\theta) \qquad (1.2.2)$$

From (1.2.1) and (1.2.2), we get

$$p(\theta|x) = \frac{p(x|\theta)\, p(\theta)}{p(x)} \qquad (1.2.3)$$

Note that

$$p(x) = \int p(\theta, x)\, d\theta$$

$$= \int p(x|\theta)\; p(\theta)\, d\theta$$

$$= E\,[\, p(x|\theta)\,] \qquad (1.2.4)$$

where E indicates averaging with respect to distribution of θ (e.g., Box and Tiao, 1973; Gelman, Carlin, Stern and Rubin, 1995; Lee, 1997 and Carlin and Louis, 2000). It is clear that $p(y)$ is not a function of θ. As a result, (1.2.3) can be rewritten as

$$p(\theta| x) \quad \alpha \quad p(x|\theta)\; p(\theta) \qquad (1.2.5)$$

This is well known Bayes' theorem. In the Bayesian terminology, $p(\theta)$ is a prior density of θ, which tells us what is known about θ without knowledge of data. The density $p(x|\theta)$ is likelihood function of θ, which represents the contribution of x(*data*) to knowledge about θ (e.g., Berger, 1985 and Zellener, 1971). Finally, $p(\theta|x)$ is the posterior density, which tells us what is known about θ given knowledge of data x.

Inferences about θ are made from the posterior density and solutions to the statistical problem is completed with

$$pr(\theta \in R|x) = \int_R p(\theta|x)\, d\theta \qquad (1.2.6)$$

where R is the region of the space of θ. Fixing the probability in (1.2.6) at, say 1-α, for a given α, it is possible to obtain an interval for θ such that its probability content is 1-α. Note that R in (1.2.6) is conceptually different from the usual *confidence interval* when repeated sampling of *y* is envisaged (e.g., Gianola and Fernando, 1986).

Posterior estimates of θ can also be obtained from the posterior density $p(\theta|x)$. For example, posterior mean and posterior mode defined in equations (1.2.7) and (1.2.8) can serve as point estimates of θ, i.e.

$$E(\theta|x) = \int \theta\ p(\theta|x)\, d\theta \qquad (1.2.7)$$

and $$\hat{\theta} = \sup_{\theta \in R} p(\theta|x) \qquad (1.2.8)$$

where R is the space of θ.

It may be noted that contrary to sampling theory approach to statistical analysis, point and interval estimations (or hypothesis testing) are not two different issues in Bayesian scenario. That is, we can say that Bayesian statistical analysis problem is regarded as solved as soon as we supply a posterior distribution $p(\theta|x)$ which shows what can be inferred about the vector parameter θ from the data y given a relevant prior state of knowledge represented by $p(\theta)$ (e.g., Khan, 1997).

The origin of Bayes' theorem has a fascinating history. It is named after the Rev. Thomas Bayes, a priest who never published a mathematical paper in his lifetime. The paper in which the theorem appears was posthumously read before the Royal Society by his friend Richard Price in 1764. Stigler suggests it was first discovered by Nicolas Saunderson, a blind mathematician / optician who, at age 29, became Lucasian Professor of Mathematics at Cambridge (the position earlier held by Isaac Newton). More details are discussed in Stigler, 1983. The term *Bayesian*, however, came into use only around 1950, and in fact it is not clear that Bayes would have endorsed the very broad interpretation of probability now called "Bayesian". Laplace independently proved a more general version of Bayes' theorem and put it to

good use in solving problems in celestial mechanics, medical statistics and, by some accounts, even jurisprudence.

1.3 Sequential Nature of Bayes' Theorem

The theorem in (1.2.5) is appealing because it provides a mathematical formulation of how previous knowledge may be combined with new knowledge. Indeed, the theorem allows us to continually update information about a set of parameters θ as more observations are taken. Thus, suppose we have an initial sample of observations x_1, then Bayes' formula gives

$$p(\theta|x_1) \ \alpha \ \ p(x_1 \,|\, \theta)\ p(\theta) \qquad (1.3.1)$$

Now, suppose we have a second sample of observations x_2 distributed independently of the first sample, then

$$p(\theta|x_2, x_1) \alpha \ \ p(\theta)\ \ p(x_1|\theta)\ \ p(x_2|\theta)$$

$$p(\theta|x_2, x_1) \alpha \ \ p(\theta|x_1)\ \ p(x_2|\theta) \qquad (1.3.2)$$

The expression (1.3.2) is precisely of the same form as (1.3.1), except that $p(\theta|x_1)$, the posterior distribution for θ given x_1, plays the role of the prior distribution for the second sample. Similarly, if we have *n* independent observations, then posterior can be recalculated after each new observation i.e.

$$p(\theta|x_1, x_2, \ldots, x_n) \ \alpha \ \ p(\theta|x_1, x_2, \ldots, x_{n-1})\ p(x_n|\theta) \qquad (1.3.3)$$

Thus, Bayes' theorem describes, in a fundamental way, the process of learning from experience, and shows how knowledge about the state of nature represented by θ is continually modified as new data becomes available (e.g., Box and Tiao, 1973).

1.4 Marginal and Conditional Inferences

Often, only a subset of the unknown parameters is really of concern to us, the rest being nuisance parameters that are really of no concern to us. A very strong feature of Bayesian analysis is that we can remove the effects of the nuisance parameters by simply integrating them out of the posterior distribution to generate a

marginal posterior distribution for the parameters of interest. For example, if θ is partitioned as (θ_1, θ_2), with θ_1 a *p* dimensional vector and θ_2 as *(k-p)* dimensional vector, then the marginal posterior density for θ_1 is given by

$$p(\theta_1 \mid x) = \frac{\int_{R_2} p(x \mid \theta)\, p(\theta)\, d\theta_2}{\int_R p(x \mid \theta)\, p(\theta)\, d\theta} \qquad (1.4.1)$$

Similarly, the marginal posterior density for θ_2 is given by

$$p(\theta_2 \mid x) = \frac{\int_{R_1} p(x \mid \theta)\, p(\theta)\, d\theta_1}{\int_R p(x \mid \theta)\, p(\theta)\, d\theta} \qquad (1.4.2)$$

The requirement of orthogonality between nuisance parameter and the parameter of interest is not required in this framework (e.g., Cox and Reid, 1987). Moreover, marginal posterior densities are better substitutes of conditional profile likelihoods of Cox and Reid (1987).

Conditional inferences for θ_1 given θ_2; and θ_2 given θ_1 can also be made using the posteriors

$$p(\theta_1 \mid x, \theta_2) = \frac{p(x \mid \theta_1, \theta_2)\, p(\theta_1 \mid \theta_2)}{\int_{R_1} p(x \mid \theta_1, \theta_2)\, p(\theta_1 \mid \theta_2)\, d\theta_1} \qquad (1.4.3)$$

and

$$p(\theta_2 \mid x, \theta_1) = \frac{p(x \mid \theta_1, \theta_2)\, p(\theta_2 \mid \theta_1)}{\int_{R_2} p(x \mid \theta_1, \theta_2)\, p(\theta_2 \mid \theta_1)\, d\theta_2} \qquad (1.4.4)$$

Marginal and conditional inference procedures are two entirely different things. In the former, we ignore one of the components of θ by integrating it out from the joint posterior $p(\theta \mid y)$, while in the later we control (or adjust) one of the components of θ (e.g., Khan, 1997 and Ahmad, 2006).

1.5 Prior Distribution:

The fundamental part of any Bayesian analysis is the prior distribution. The prior distribution P(θ) represents all that is known or assumed about the parameter θ usually the prior information is subjective and is based on a person's own experience and judgment, a statement of one's degree of belief regarding the parameter, design information and personal opinions. The other critical feature of the Bayesian analysis is the choice of a prior. The key here is that when the data have sufficient signal, even

a bad prior will still not greatly influence the posterior. In a sense, this is an asymptotic property of Bayesian analysis in that all but pathological priors will be overcome by sufficient amounts of data. We can check the impact of the prior by seeing how stable to posterior distribution is to different choices of priors. If the posterior is highly dependent on the prior, then the data (the likelihood function) may not contain sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contains significant information.

Prior distribution may be categorical in different ways. One common classification is a dichotomy that separated "proper" and "improper" priors. A prior distribution is proper if it does not depend on the data and the value of integral $\int_{-\infty}^{\infty} P(\theta)d\theta$ or summation $\sum P(\theta)$ is one. If the prior does not depend on the data and the distribution does not integrate or sum to one then we say that the prior is improper. Other classification of prior is either based on properties or on distributional forms as under:

a) Uniform prior

In a state of ignorance the prior distribution is accepted as being uniform. It appears that great minds like Gauss, Bernoulli and Laplace used the principle in some form or other in their work. It is claimed that Bayes himself used uniform prior in his revolutionary work.

The apparent success with uniform prior subscribed to the senore's idea that perhaps the uniform prior is the final answer. Jeffery's (1961) makes an interesting comment that there is no more need for such an idea than to suggest that an oven which cooked roast beef once cannot cook anything other than roast beef. One should be cautious before invoking the uniform prior theory, for a careless and mechanical use of this principal may lead to contradiction and confusion.

b) Non- informative prior (NIP)

One class of prior distribution is called non-informative prior and as the name suggests, it is prior that contains no information about Θ. Non informative priors are also called priors of ignorance Box and Tiao (1973) provides a thorough discussion of non informative priors for one or more parameters.

Rather than a state of a complete ignorance, the non informative prior refers to the case when relatively little (or very limited) information is available a priori. In

other words, a priori information about the parameter is not substantial relative to the information expected to be provided by the sample of empirical data. A prior probability distribution that represents perfect ignorance or indifference would produce a posterior probability distribution that represents what one should need about the parameter Θ on the basis of the evidence (data) Y alone. Such a prior is called "neutral" or non informative priors by Royall (1997). According to Jeffery's (1983), non -informative priors provide a formal way of expressing ignorance of the value of the parameter over the permitted range.

If the prior is non informative, we should assign the same density to each $\theta \in \Omega$, which of course implies that prior $P(\theta)$ is uniform given by $P(\theta = k, \theta \in \Omega)$.

The non informative prior often leads to a class of improper prior, improper in the sense that $\int_{\Omega} P(\Theta) d\Theta \neq 1.$ The derivation of non informative prior is mathematically very closely associated with variance stabilizing transformations (Bartlett, 1937) and Fishers information (Fisher, 1922).

c) Natural conjugate prior (NCP)

Raiffa and Schlaifer (1961) presented a formal development of conjugate prior distribution, intuitively, a conjugate prior distribution; say $P(\Theta)$ for given sampling distribution, say $f(x|\theta)$ is such that the posterior distribution $P(\theta|x)$ and the prior $P(\theta)$ are members of the same family of distributions.

Let $\underline{x} = (x_1, x_2, ..., x_n)$ be a data from some family of distribution $f(x|\theta)$ which combines basic information. Such a function is known as sufficient statistic. Sufficient statistic exists for a number of standard distributions.

As in frequencies frame work, sufficient statistic plays an important role in Bayesian interference in constructing a family of prior distribution known as natural conjugate prior (NCP) .The family of prior distribution $P(\theta), \Theta \in \Omega$ is called a natural conjugate family if the corresponding posterior distribution belongs to the same family as $P(\theta)$.

The below given table has shown the conjugate priors for several common likelihood functions.

Table 1.1: Conjugate prior for common likelihood functions

Likelihood	Conjugate priors
Binomial	Beta
Multinomial	Dirichlet
Poisson	Gamma
NormalμUnknown, σ^2 known	Normal
Normalμ Known, σ^2 unknown	Inverse chi-square
Multivariate normal μ unknown, ν known	Multivariate normal
Multivariate normal μ known, ν unknown	Inverse Wishart

d) Jeffrey's Invariant Prior (1946, 1961)

In situations where we only have limited data available and we have no expert knowledge available. We should be able in such situations to choose a suitable prior which should obey the invariant property under parameter transformation. The Jeffery prior was designed to solve the invariance under the parameter transformation problem. According to the Jeffery principal the following equation should hold:

$$P(\Phi) = P(\theta)\left|\frac{d\theta}{d\phi}\right| = P(\theta)\,\left|h'(\theta)\right|^{-1}$$

where $\Phi = h(\theta)$ is a one to one parameter transformation. This states that a rule for determining a prior should yield an equivalent transformed. From the above formulation we can derive the general formula of the Jeffery prior, which is given as

$$P(\theta) \propto \sqrt{I(\theta)} \propto \left[-E\left\{\frac{\partial^2 \log L(\theta \mid x)}{\partial \theta^2}\right\}\right]^{1/2}$$

where $I(\theta)$ is the Fisher information for the parameter θ. When there are multiple parameters I is the Fisher information matrix, the matrix of the expected second partials

$$I(\boldsymbol{\theta}) = E\left\{\frac{\partial^2 \log L(\theta \mid x)}{\partial \theta_i \partial \theta_j}\right\}$$

In this case, the Jeffery prior becomes

$$P(\boldsymbol{\theta}) \propto \sqrt{\det[I(\boldsymbol{\theta})]}$$

Table 1.2: Jeffery prior for the common probability distribution

Probability Distribution	Jeffery's prior
Normal μ Unknown, σ^2 known	$P(\mu)$ = constant
Normal μ Known, σ^2 unknown	$P(\sigma) \propto \frac{1}{\sigma}$
Normal μ, σ^2 Both unknown	$P(\mu,\sigma) = P(\sigma)P(\mu \mid \sigma) = P(\sigma)P(\mu)$
Exponential Distribution	$P(\theta) \propto \frac{1}{\theta}$
Binomial Distribution with n independent draws	$P(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$
Weibull (α, β)	$P(\beta,\alpha) = P_1(p)P_2(\alpha \mid \beta)$ $P(\beta,\alpha) \propto \frac{1}{\beta^{\alpha}}$
Negative Binomial Distribution	$P(\theta) \propto \theta^{-1}(1-\theta)^{-1/2}$
Uniform Distribution i.e. $X \sim U(0,\theta)$	$P(\theta) \propto \frac{1}{\theta}$

e) **Maximal information prior (MIP)**

Zellner (1977) used the information theoretic approach to define maximal information prior. Let $I_x(\theta) = \int f(x \mid \theta) \log f(x \mid \theta) d\theta$ be a measure of information in the pdf $f(x \mid \theta)$. The prior average information is defined as

$$\bar{I}_x(\theta) = \int I_x(\theta) P(\theta) d\theta$$

where $P(\theta)$ is a prior density of θ and $\int P(\theta) \log P(\theta) d\theta$ measures the information in prior $P(\theta)$.

$$G = \bar{I}_x(\theta) - \int P(\theta) \log P(\theta) d\theta$$

$$G = \int I_x(\theta) P(\theta) d\theta - \int P(\theta) \log P(\theta) d\theta$$

is defined as a measure of gain in information, the maximal information prior is the one that maximizes G for varying $P(\theta)$ subject to the condition $\int P(\theta) d\theta = 1$

f) **Asymptotically locally invariant prior (ALIP)**

Hartigan (1964) derived a family of prior densities to represent our ignorance about Θ using invariance techniques similar to those suggested by Jeffery's (1946). He named this asymptotically locally invariant (ALI) prior. The ALI priors are easy to derive for exponential family of distributions.

Hartigan (1964) point out that in some instances, the posterior distribution based on the ALI prior may lead to a chi-square having a degree of freedom contrary to the usual rule of assigning the degree of freedom to chi-square.

g) Dirichlets prior (DP)

Dirichlets prior distribution is

$$P(p_1, p_2, ..., p_k) = \frac{\Gamma\theta}{\Gamma\theta_1.\Gamma\theta_2.....\Gamma\theta_k} p_1^{\theta_1-1}.p_2^{\theta_2-1}.....p_k^{\theta_k-1}$$

where $\theta = \sum_{i=1}^{k} \theta_i$, $\sum_{i=1}^{k} p_i = 1,; 0 < p_i < 1; \theta_i > 0$ is a generalization of the beta – prior.

$$P(\theta) = \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1}, 0 < \theta < 1; a, b > 0$$

h) Haldane's prior (1931):

Halden's prior is given as

$P(\theta) \propto \theta^{-1}(1-\theta)^{-1}$, $\theta \in [0,1]$

which is an improper density. We get if we put $\alpha = \beta = 0$ in Beta prior.

1.6 Loss Function (Lf):

The word "loss" is used in place of "error" and the loss function is used as a measure of the error or loss. Let Θ be an unknown parameter of some distribution $f(x|\theta)$ and suppose that θ is estimated by some statistics $T(x) \cong T$. Let $L(\theta, T)$ represents the loss incurred when the true value of the parameter is Θ where Θ is estimated by the statistics T.

Loss function is a measure of the error and presumably would be greater for large error than for small error. We would want the loss to be small or we want the estimate to be close to what it is estimating. Our objective is to select an estimator T= L(x_1, x_2,…,x_n) that makes this error or loss small. Loss depends on sample and we cannot hope to make the loss small for every possible sample but can try to make the

loss small on the average. Our objective is to select an estimator that makes the average loss (risk) small and ideally select an estimator that has the small risk.

Some Important Loss Functions are as under:

a) Squared-Error Loss Function:

The squared error loss function (SELF) was proposed by Legendre and Gauss (1805) to develop least squares theory. Later, it was used in estimation problems when unbiased estimators of θ were evaluated in terms of the risk function $R(\theta, T)$ which becomes nothing but the variance of the estimator. It was also observed that SELF is a convex loss function and therefore, restricts the class of estimators by excluding randomized estimator. The SELF is given as

$$L(\theta, T) = (\theta - T)^2 .$$

b) Weighted SELF :

A generalization of squared-error loss, which is of interest, is

$$L(\theta, T) = W(\theta)(\theta - T)^2$$

This loss is called weighted squared-error loss and has the attractive feature of allowing the squared error, $(\theta - T)^2$ to be weighted by a function of $-\Theta$.

c) Quadratic SELF:

Other variant of square error loss is quadratic SELF. If $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$ is a vector to be estimated by $T = (t_1, t_2, ..., t_p)'$, and Q is pxp positive definite matrix, and then $L(\boldsymbol{\theta}, T) = (\boldsymbol{\theta} - T)'Q(\boldsymbol{\theta} - T)$ is called quadratic loss. When Q is diagonal, this reduces to

$$L(\boldsymbol{\theta}, T) = \sum_{i=1}^{k} Q_i (\theta_i - t_i)^2$$

and is natural extension of squared-error loss to themultiparametricsituation.

d) Linear Loss:

When utility function is approximately linear (as is often the case over a reasonable segment of the reward space), the loss function will tend to be linear. Thus of interest is the linear loss

$$L(\theta, T) = C_1(\theta - T), \quad \theta \geq T$$

and $L(\theta, T) = C_2(T - \theta); \quad \theta < T$

The constants C_1 and C_2 reflect the effect of over and over estimating θ. By suitably choosing C_1, C_2 any fractile of the posterior distribution will be a Bayes estimator (Box and Tiao, 1973).

If C_1 and $C_2$are functions of θ, the above loss function is called weighted linear loss function.

e) Absolute Loss:

$$L(\theta,T)=|\theta-T|$$

is called the absolute loss function. As per De Groot (1970) for such a loss function, Bayesian estimator is the posterior median.

f) Zero –One loss:

$$L(\theta,T)=0 \quad \text{iff} \ |\theta-T|\le C$$

and

$$L(\theta,T)=1 \quad \text{iff} \ |\theta-T|> C$$

where c is the small positive constant.

As per Raiffa and Schlaifer (1961), Bayes estimator for such a loss function is mode of posterior distribution. The risk function $R(\theta,T)$, associated with the estimator T is defined as the expected value of the loss function. The loss is Zero if the decision is made correct about T and the loss is one if the decision about T is made incorrect.

$$R(\theta,T)=E_x[L(\theta,T)]=\int L(\theta,T)\,f(x\mid\theta)dy$$

$$=P\left[|\theta-T|>C\right]$$

=P[incorrect decision about T]

1.7 Risk Function:

The risk function $R(\theta,T)$ associated with an estimator T is defined as the expected value of the loss function and is given by

$$R(\theta,T)=E_x\left[L(\theta,T)\right]=\int L(\theta,T)\,f(x\mid\theta)dx$$

Bayes risk associated with an estimator T is defined as the expected value of the risk function $R(\theta,T)$ with respect to the prior distribution $p(\theta)$ of θ and is given by

$$R(\theta,T) = E_\theta\left[R(\theta,T)\right]$$

$$= \int R(\theta,T)P(\theta)d\theta$$

$$= \int E_x\left[L(\theta,T)\right]P(\theta)d\theta$$

$$= \int\left[\int L(\theta,T)\, f(x\mid\theta)\right]P(\theta)d\theta$$

Bayesian risk of an estimator is an average risk, which is a real number. Risk can be used as a guide. A good decision would be that minimizes the risk for all values of θ in Ω.For two estimators $T_1 = t_1(x_1, x_2, ..., x_n)$ and $T_2 = t_2(x_1, x_2, ..., x_n)$ estimator T_1 is defined to be better estimator than T_2 if

$Rt_1(\theta) \le Rt_2(\theta), \forall\, \theta \in \Omega$

Thus, risk and loss functions are used to assess the goodness of estimators.

2.1 Introduction

The Weibull distribution was introduced by the Swedish Physicist Waloddi Weibull in 1939. This distribution has been extensively used in life time and reliability problem. The Weibull distribution is perhaps the most widely used life time distribution model. Its application in connection with lifetimes of many types of manufactured items has been widely advocated (e.g., Weibull, 1951; Berrentoni, 1964), and it has been used as a model with diverse types of items such as vacuum tubes (Kao, 1959), ball bearings (Lieblein and Zelen, 1956), and electrical insulation. It is also widely used in biomedical applications, for example, in studies of the time to the occurrence of tumors in human population (Whittemore and Altschuler, 1976) or in laboratory animals (Pike, 1966 and Peto et al., 1972) and in many other situations. A comprehensive review of this model is available in Johnson, Kotz and Balakrishnan (1995). Mudholkar *et al.* (1996) described a certain generalization of the Weibull distribution and applied it to survival data. Hirose and Lai (1997) constructed confidence internals for the parameters, including a location parameter, for the case of grouped data. Marshall and Olkin (1997) introduced a method for adding new parameters to an existing two parameters Weibull distribution. Their distribution is known as the Marshall-Olkin extended distribution. Nordman and Meeker (2002) evaluated exact coverage probabilities of approximate prediction intervals for the number of failures to be observed in a future inspection of a sample. Xie *et al*. (2002) developed a model named modified Weibull extension with three parameters. This model allows for increasing, bathtub-shaped or decreasing failure rate function and the resulting Weibull probability plot is concave. Lai *et al*.(2003) proposed a modified Weibull distribution involving three parameters. This distribution has increasing or bathtub-shaped failure rate function and its probability plot is concave. Further, Tang *et al*. (2003) have carried out the statistical analysis of the extension. Ghitany *et al*. (2005) showed that the Marshall–Olkin extended Weibull distribution could be obtained as compound distribution with mixing exponential distribution Nadarajah and Kotz (2007) discussed products and ratios of Weibull random variables.

The Weibull distribution has pdf of the form

$$f(x;\alpha,\beta)=\frac{\beta}{\alpha}x^{\beta-1}\exp\left(-\frac{x^{\beta}}{\alpha}\right) \qquad ,x\geq 0 \qquad (2.1)$$

where $\alpha>0$ *and* $\beta>0$ are shape and scale parameters respectively.

The mean and variance of this distribution are

$$\alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right) \qquad \textit{and} \qquad \alpha^{\frac{1}{\beta}}\left[\Gamma\left(\frac{2}{\beta}+1\right)-\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^{2}\right]$$

respectively, and *pth* quantitile is $x_p=\left[-\alpha\log(1-p)\right]^{\frac{1}{\beta}}$. The Weibull distribution is monotone increasing if $\beta>1$, decreasing if $\beta<1$ and constant for $\beta=1$. The model is fairly flexible and has been found to provide a good description of many types of lifetime data. The Weibull distribution arises as an asymptotic extreme value distribution, and in some instances can be used to provide motivation for it as model (Weibull, 1951; Peto et al., 1972). A plot of Weibull distribution with different shape parameters is shown below in figure 2.1.

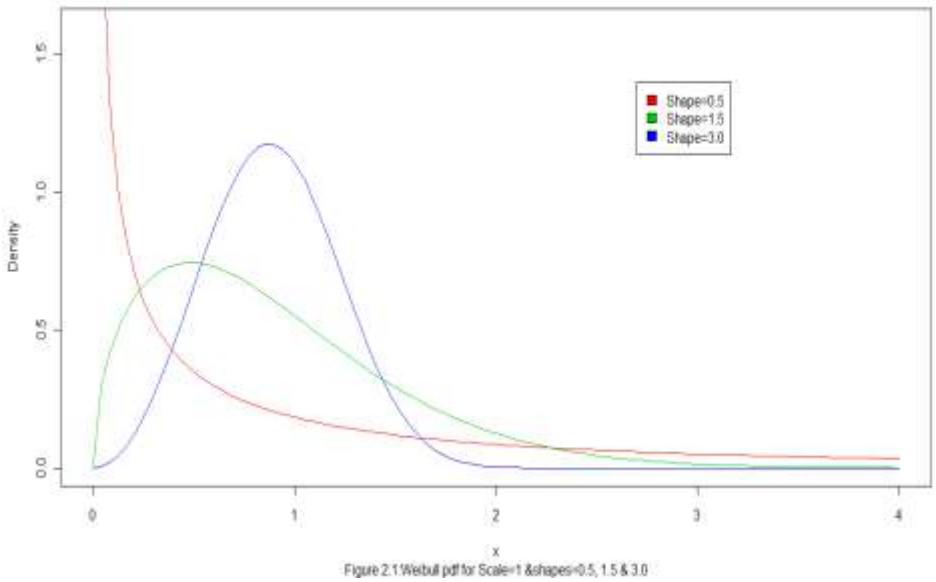

Figure 2.1 Weibull pdf for Scale=1 &shapes=0.5, 1.5 & 3.0

The shape of the Weibull pdf depends upon the value of β, in fact β is sometimes called the shape parameter for the distribution. "Typical" β values vary

from application to application, but in many situations distributions with β in the range from 1 to 3 seem appropriate.

2.2 Different methods of Derivation of Weibull distribution

The Weibull distribution arises from the exponential distribution in the following way.

a) ***I –method:*** Suppose that instead of assuming that the failure time is distribution exponentially, we assume some power (say, pth) of the failure time is distributed exponentially. Thus if X is a random variable representing the failure time, then we assume that Y = X^P has the p.d.f. given by

$$f(y) = \frac{1}{\alpha}\exp\left(\frac{-y}{\alpha}\right) \qquad y \geq 0, \alpha > 0 \qquad (2.2.1)$$

Since Y = X^p

$$\therefore dy = PX^{p-1}\,dx$$

$$\frac{dy}{dx} = PX^{p-1}$$

Now make a Jacobian transformation

$$f(x) = f(y)\,|J| \qquad (2.2.2)$$

Where $|J| = \left|\frac{dy}{dx}\right|$

$$J = PX^{p-1}$$

In our case P = β

$$|J| = \beta X^{\beta-1} \qquad (2.2.3)$$

Thus using eq. (2.2.1) and eq. (2.2.3) in eq. (2.2.2), we get

$$f(x) = \frac{1}{\alpha}\exp\left(-\frac{x^\beta}{\alpha}\right)\beta x^{\beta-1}$$

$$\Rightarrow \qquad f(x) = \frac{\beta}{\alpha}x^{\beta-1}\exp\left(-\frac{x^\beta}{\alpha}\right) \quad x \geq 0, \alpha, \beta > 0 \qquad (2.2.4)$$

Which is the pdf of two parameter Weibull distribution.

b) ***II Method:*** For the two parameter Weibull distribution, the distribution function is given by

$$F(x) = \int_o^x f(x)dx$$

$$\Rightarrow \quad F(x) = \int_o^x \frac{\beta}{\alpha} x^{\beta-1} \exp\left(\frac{-x^\beta}{\alpha}\right) dx \qquad (2.2.5)$$

Put $\frac{x^\beta}{\alpha} = z$ (2.2.6)

$$\frac{\beta}{\alpha} x^{\beta-1} dx = dz \qquad (2.2.7)$$

Thus by using eq. (2.2.6) and eq. (2.2.7) in eq. (2.2.5), we have

$$F(x) = \int_o^z e^{-z} dz$$

$$\Rightarrow \quad \mathrm{F(x)} = \left[\frac{e^{-z}}{-1}\right]_0^z = -[e^{-z} - e^o]$$

$$\Rightarrow \quad \mathrm{F(x)} = (e^{-z} - 1) = 1 - \exp(-\mathrm{z})$$

$$\Rightarrow \quad F(x) = 1 - \exp\left(-\frac{x^\beta}{\alpha}\right)$$

Also we know that the reliability function is given by

$$R(x) = 1 - F(x)$$

$$\Rightarrow \quad R(x) = 1 - \left\{1 - \exp\left(-\frac{x^\beta}{\alpha}\right)\right\}$$

$$\Rightarrow \quad R(x) = \exp\left(-\frac{x^\beta}{\alpha}\right) \qquad (2.2.8)$$

and the instantaneous failure rate or hazard rate is given by

$$\mu(x) = \frac{f(x)}{R(x)} \qquad (2.2.9)$$

Now by using eq. (2.2.4) and eq. (2.2.8) in eq.(2.2.9),we have

$$\mu(x) = \frac{\frac{\beta}{\alpha} x^{\beta-1} \exp\left(-\frac{x^\beta}{\alpha}\right)}{\exp\left(-\frac{x^\beta}{\alpha}\right)}$$

$$\mu(x) = \frac{\beta}{\alpha} x^{\beta-1} \qquad (2.2.10)$$

If β>1, then μ (x) is increasing function.

If β <1, then μ (x) is decreasing function.

and If β = 1, then μ(x) is constant and leads to exponential distribution.

This observation lends another way of deriving Weibull distribution; namely assuming that the hazard rate μ(x) is proportional to power of x. Thus, if we assume that $\mu(x) = Cx^k$, then

Since $f(x) = \mu(x)\, exp\left\{-\int_o^x \mu(w)dw\right\}$

then $f(x) = Cx^k \exp\left\{-\int_0^x Cx^k dx\right\}$

$$f(x|c,k) = Cx^k \exp\left\{-\frac{Cx^{k+1}}{K+1}\right\} \qquad (2.2.11)$$

If we take $C = \frac{\beta}{\alpha}$ and $k = \beta - 1$ in eq. (2.2.12), we have

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left\{-\frac{x^\beta}{\alpha}\right\} \; x \geq 0,\, \alpha,\beta > 0 \qquad (2.2.12)$$

This is the required pdf of two parameter Weibull distribution.

If β = 1, then (2.2.12) becomes $f(x,\alpha) = \frac{1}{\alpha}\exp\left(-\frac{x}{\alpha}\right)$, which is the pdf for an exponential distribution.

If α = β = 1, then pdf of (2.2.12) becomes $f(x,\alpha) = \exp(-x)$, which is the pdf of standard exponential distribution. Thus, we see that the exponential distribution is a special case of Weibull distribution.

2.3 Moments of two parameter Weibull distribution

Moments are the expected values of certain functions of a random variable. They serve to numerically describe the variable with respect to given characteristics for e.g. location, variation, skewness and kurtosis.

The expected value of X^r is termed as rth moment about origin of the random variable X:

$\mu_r' = E(X)^r$

Since $X \sim W(\alpha, \beta)$

$$f(x, \alpha, \beta) = \frac{\beta}{\alpha} \, x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha}\right) \quad x \geq 0 \, , \alpha, \beta > 0 \tag{2.3.1}$$

$\therefore \quad \mu_r' = E(X)^r = \int_o^{\infty} x^r \, f(x, \, \alpha. \beta) \, dx$

$$\Rightarrow \quad \mu_r' = \int_o^{\infty} x^r \, \frac{\beta}{\alpha} \, x^{\beta-1} \, \exp\left(-\frac{x^{\beta}}{\alpha}\right) dx \tag{2.3.2}$$

Put $\frac{x^{\beta}}{\alpha} = z$

$$\Rightarrow \quad \frac{\beta}{\alpha} \, x^{\beta-1} dx = dz$$

Thus after substituting these values in eq. (2.3.2), we get

$$\mu_r' = \int_0^{\infty} (\alpha \, z)^{\frac{r}{\beta}} \exp(-z) dz$$

$$\Rightarrow \quad \mu_r' = (\alpha)^{\frac{r}{\beta}} \int_0^{\infty} (z)^{\frac{r}{\beta}+1-1} \exp(-z) dz$$

$$\Rightarrow \quad \mu_r' = (\alpha)^{\frac{r}{\beta}} \Gamma\left(\frac{r}{\beta}+1\right) \tag{2.3.3}$$

Put r = 1 in eq. (2.3.3) then we get

$$\mu_1' = (\alpha)^{\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta}+1\right) \tag{2.3.4}$$

Which is the mean of two parameter Weibull distribution.

Put r = 2 in eq.(2.3.3) then we get

$$\mu_2' = (\alpha)^{\frac{2}{\beta}} \Gamma\left(\frac{2}{\beta}+1\right)$$

Then variance σ^2 is given by

$$\sigma^2 = \mu_2 = \mu_2' - \mu'^{\,2}$$

$$\sigma^2 = (\alpha)^{\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right) - \left((\alpha)^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right)^2$$

$$\Rightarrow \quad \sigma^2 = (\alpha)^{\frac{2}{\beta}}\left[\Gamma\left(\frac{2}{\beta}+1\right) - \Gamma\left(\frac{1}{\beta}+1\right)^2\right] \tag{2.3.5}$$

If r = 3 in eq. (2.3.3) then we have

$$\mu_3' = (\alpha)^{\frac{3}{\beta}}\Gamma\left(\frac{3}{\beta}+1\right)$$

Also, $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

$$\therefore \quad \mu_3 = (\alpha)^{\frac{3}{\beta}}\Gamma\left(\frac{3}{\beta}+1\right) - 3\left(\alpha^{\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right)\right)\left(\alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right) + 2\left(\alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right)^3$$

$$\Rightarrow \quad \mu_3 = (\alpha)^{\frac{3}{\beta}}\left[\Gamma\left(\frac{3}{\beta}+1\right) - 3\left(\Gamma\left(\frac{2}{\beta}+1\right)\right)\left(\Gamma\left(\frac{1}{\beta}+1\right)\right) + 2\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^3\right] \tag{2.3.6}$$

If r = 4 in eq. (2.3.3), then we have

$$\mu_4' = (\alpha)^{\frac{4}{\beta}}\Gamma\left(\frac{4}{\beta}+1\right)$$

Also, $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

$$\therefore \mu_4 = \left((\alpha)^{\frac{4}{\beta}}\Gamma\left(\frac{4}{\beta}+1\right)\right) - 4\left((\alpha)^{\frac{3}{\beta}}\Gamma\left(\frac{3}{\beta}+1\right)\right)\left((\alpha)^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right) + 6\left((\alpha)^{\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right)\right)\left((\alpha)^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right)^2 - 3\left((\alpha)^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right)^4$$

$$\Rightarrow \mu_4 = (\alpha)^{\frac{4}{\beta}}\left[\left(\Gamma\left(\frac{4}{\beta}+1\right)\right) - 4\Gamma\left(\frac{3}{\beta}+1\right)\Gamma\left(\frac{1}{\beta}+1\right) + 6\Gamma\left(\frac{2}{\beta}+1\right)\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2 - 3\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^4\right] \tag{2.3.7}$$

2.4 Standard Deviation

The positive square root of the variance is called standard deviation. The idea of standard deviation was first given by Karl Pearson in 1983.

Symbolically $\sigma = \sqrt{\sigma^2}$

Since from equation (2.3.5), we have the variance of two parameter Weibull distribution given as

$$\sigma^2 = (\alpha)^{\frac{2}{\beta}}\left[\Gamma\left(\frac{2}{\beta}+1\right)-\Gamma\left(\frac{1}{\beta}+1\right)^2\right]$$

$$\Rightarrow \quad \sigma = (\alpha)^{\frac{1}{\beta}}\sqrt{\Gamma\left(\frac{2}{\beta}+1\right)-\Gamma\left(\frac{1}{\beta}+1\right)^2} \qquad (2.4.1)$$

$$\Rightarrow \quad \sigma = (\alpha)^{\frac{1}{\beta}}\sqrt{\sigma_2-\sigma_1^2}$$

Where $\sigma_k = \Gamma\left(\frac{k}{\beta}+1\right)$

2.5 Coefficient of variation

It is the ratio of standard deviation and mean. Usually it is denoted by C.V. and is given by

$$C.V. = \frac{\sigma}{\mu} \qquad (2.5.1)$$

Now by using eq. (2.4.1) and eq.(2.3.4) in eq.(2.5.1),we have

$$C.V. = \frac{(\alpha)^{\frac{1}{\beta}}\sqrt{\Gamma\left(\frac{2}{\beta}+1\right)-\Gamma\left(\frac{1}{\beta}+1\right)^2}}{(\alpha)^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)}$$

$$\Rightarrow \quad C.V. = \frac{\sqrt{\sigma_2-\sigma_1^2}}{\sigma_1}$$

Where $\sigma_k = \Gamma\left(\frac{k}{\beta}+1\right)$

2.6 Skewness and Kurtosis

The most popular way to measure the Skewness and kurtosis of a distribution function rests upon ratios of moments. Lack of symmetry of tails (about mean) of frequency distribution curve is known as skewness. The formula for measure of skewness given by Karl Pearson in term of moments of frequency distribution is given by

$$\beta_1 = \frac{\mu_4^3}{\mu_2^3} \qquad (2.6.1)$$

Now by using eq. (2.3.4) and (2.3.5) in eq. (2.6.1), we have

$$\Rightarrow \qquad \beta_1 = \left(\alpha^{3/\beta}\right)^2 \left[\Gamma\left(\frac{3}{\beta}+1\right) - 3\left(\Gamma\frac{2}{\beta}+1\right)\Gamma\left(\Gamma\frac{1}{\beta}+1\right) + 2\Gamma^3\left(\frac{1}{\beta}+1\right)\right]^2$$

$$\Rightarrow \qquad \beta_1 = \left[\Gamma\left(\frac{3}{\beta}+1\right) - 3\left(\Gamma\frac{2}{\beta}+1\right)\Gamma\left(\frac{1}{\beta}+1\right) + 2\left(\Gamma^3\frac{1}{\beta}+1\right)\right]^2$$

And

$$\Rightarrow \qquad \beta_1 = \frac{\left(\sigma_3 - 3\sigma_2\sigma_1 + 2\sigma_1^3\right)^2}{\left(\sigma_2 - 2\sigma_1^2\right)^2}$$

Where $\sigma_k = \Gamma\left(\frac{k}{\beta}+1\right)$

and $\gamma_1 = \sqrt{\beta_1}$

$$\Rightarrow \qquad \gamma_1 = \frac{\left(\sigma_3 - 3\sigma_2\sigma_1 + 2\sigma_1^3\right)}{\left(\sigma_2 - 2\sigma_1^2\right)^{3/2}}$$

If $\gamma_1 < 0, then\ the\ frequency\ curve\ is\ negatively\ skewed$

If $\gamma_1 \geq 0\ then\ the\ frequency\ curve\ is\ positively\ Skwwed$

Kurtosis

Kurtosis meaning bulginess. The formula for measure of Kurtosis is given by

$$\beta_{2=}\ \frac{\mu_4}{\mu_2^2} \qquad (2.6.2)$$

Thus by using eq. (2.3.4) and eq.(2.3.6) in eq.(2.6.2),we have

$$\Rightarrow \qquad \beta_2 = \alpha^{4/\beta} \frac{\left[\Gamma\left(\frac{4}{\beta}+1\right)-4\left(\frac{3}{\beta}+1\right)\Gamma\left(\frac{1}{\beta}+1\right)+6\left(\frac{2}{\beta}+1\right)\Gamma^2\left(\frac{1}{\beta}+1\right)-3\Gamma^4\left(\frac{1}{\beta}+1\right)\right]}{\alpha^{4/\beta}\left[\Gamma\left(\frac{2}{\beta}+1\right)-\Gamma^2\left(\frac{1}{\beta}+1\right)\right]^2}$$

$$\beta_2 = \frac{\sigma_4 - 4\sigma_3\,\sigma_1 + 6\sigma_2\sigma_{1^2} - 3\sigma_{1^4}}{(\sigma_2 - \sigma_{1^2})^2}$$

When $\sigma_k = \Gamma\left(\frac{k}{\beta}+1\right)$

$\therefore \qquad \gamma_2 = \beta_2 - 3$

$$\Rightarrow \qquad \gamma_{2=} \frac{\sigma_4 - 4\sigma_3\,\sigma_1 + 6\sigma_2\sigma_{1}{}^2 - 3\sigma_{1}{}^4}{(\sigma_2 - \sigma_{1}{}^2)^2} - 3$$

If $\gamma_2 > 0$, then the curve is leptokurtic

If $\gamma_2 < 0$, then the curve is platykurtic

And is $\gamma_2 = 0$, then the curve is mesokurtic or we can say, there is no kurtosis.

3.1 Introduction

There are a few occasions when population is studied as a whole. As a matter of fact, generally a sample is drawn from the population and population constants are determined on the basis of sample values. Population parameters are usually those constant which occur in probability density or mass function or the moments or some other constants of the population like median.

We know that various sampling procedures do exist and also there are many techniques to determine the value of population constants through sample values. The constant determined through sample observations which stands for population parameter θ or a function $f(\theta)$ though $f(\theta)$ in many cases is equal to θ.

The choice of a technique depends on the type of the estimator vis-a-vis estimate and the purpose of study. The goodness of an estimator is governed by certain properties. An estimator possessing the maximum properties will be considered as a good estimator.

So in estimation theory we are concerned with the properties of estimators and methods of estimation. The merits of an estimator are judged by the properties of the distribution of estimates obtained through estimators i.e. by the properties of the sampling distribution. Further, it is emphasized that estimation is possible only if there is a random sample.

The theory of estimation was founded by Prof. R. A. Fisher in a series of fundamental papers round about 1930 and is divided into two groups (i) point estimation and (ii) Internal estimation. In point estimation, a sample statistic (numerical values) is used to provide an estimate of the population parameter whereas in Interval Estimation, probable range is specified within which the true value of the parameter might be expected to lie.

The word estimator stands for the function, and the word, estimate means a value of that function. In estimator we take a random sample $(x_1, x_2, \ldots, x_n)$ from the distribution to draw out some information about unknown parameter θ. That is, we repeat the experiment n independent times, observe the sample, and we try to estimate the value of θ. The function of $(x_1, x_2, \ldots, x_n)$ used to estimate θ, say the statistic $T(x_1, x_2, \ldots, x_n)$ called an estimator of θ . we want it to be such that the computed estimate$(x_1, x_2, \ldots, x_n)$ is usually close to θ.

Thus any statistic those values are used to estimate s(θ) where S(.) is some function of the parameter θ, is defined to be an estimator r(θ). An estimator is always a statistic which is both a random variable and a function.

3.2 Methods of estimation:

A variety of methods to estimate the unknown parameters have been proposed. The common used methods are:

a) Method of maximum likelihood estimation,
b) Method of moment
c) Method of least square estimation
d) Method of minimum variance,
e) Method of least square estimation,
f) Method of minimum chi-square, and
g) Bayesian estimation.

The various methods used in this chapter are as follows:

a) Method of maximum likelihood estimation (MLE):

The most general method of estimation known is the method of maximum likelihood estimators (MLE) which was initially formulated by C.F. Gauss but as a general method of estimation was first introduced by Professor. R. A. Fisher in the early (1920) and later on developed by him in a series of papers. He demonstrated the advantages of this method by showing that it yields sufficient estimators, which are asymptotically MVUES. Thus the essential feature of this method is that we look at the value of the random sample and then choose our estimate of the unknown population parameter, the value of which the probability of obtaining the observed data is maximum. If the observed data sample values are $(x_1, x_2, \ldots, x_n)$, we can write in the discrete case

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1, x_2, \ldots, x_n)$$

which is just the value of joint probability distribution of the random values $(x_1, x_2, \ldots, x_n)$ at the sample point $(x_1, x_2, \ldots, x_n)$ since the sample values has been observed and are therefore fixed numbers, we regard $f(x_1, x_{2,}, \ldots, x_n; \theta)$ as the value of a function of the parameter Θ, referred to as the likelihood function. A similar definition applies when the random sample comes from a continuous

population but in that case $f(x_1, x_{2,}, \ldots, x_n; \theta)$ is the value of joint pdf at the sample point $(x_1, x_2, \ldots, x_n)$ i.e.; the likelihood function at the sample value $(x_1, x_2, \ldots, x_n)$.

$$L(x \mid \theta) = \prod_{i=1}^{n} f(x_i, \theta) \qquad (3.2.1)$$

Since the principle of maximum likelihood consists in finding an estimator of the parameter which maximizes L for variation in the parameter. Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ of the sample values which maximizes $L(x \mid \theta)$ for variation in Θ, then $\hat{\theta}$ is to be taken as the estimator of Θ. $\hat{\theta}$ is usually called ML estimators. Thus $\hat{\theta}$ is the solution if and only if

$$\frac{\partial L(x \mid \theta)}{\partial \theta} = 0 and \frac{\partial^2 L(x \mid \theta)}{\partial \theta^2} < 0 \qquad (3.2.2)$$

Since $L(x \mid \theta)$ >0, so $\log L(x \mid \theta)$ which shows that $L(x \mid \theta)$ and $\log L(x \mid \theta)$ attains their extreme values at the $\hat{\theta}$. Therefore, the equation becomes

$$\frac{1}{L(x \mid \theta)} \frac{\partial L(x \mid \theta)}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L(x \mid \theta)}{\partial \theta} = 0 \qquad (3.2.3)$$

a form which is more convenient from practical point of view.

b) Method of moments (substitution principle) (MM):

One of the simplest and oldest methods of estimation is the substitution principle. The method of moments was discovered and studied in detail by Karl Pearson. The method of moments is special case when we need to estimate some known function of finite number of unknown moments.

Let $f(x; \theta_1, \theta_2, \ldots\ldots, \theta_k)$ be density function of the parent population with k parameters $\theta_1, \theta_2, \ldots\ldots, \theta_k$. If μ_r' denotes the rth moment about origin, then

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x; \theta_1, \theta_2, \ldots., \theta_k), r = 1,2,\ldots., k \qquad (3.2.4)$$

In general $\mu_1', \mu_2', \ldots., \mu_k'$ will be functions of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Let $x_i, i = 1,2,\ldots, n$ n be a random sample of size n from the given population. The method of moments consists in solving the k-equation (i) for $\Theta_1, \Theta_2, \ldots\ldots, \Theta_k$ in terms of $\mu_1', \mu_2', \ldots., \mu_k'$ and then replacing these moments.

$\mu_r'; r = 1,2,3,\ldots., k$ by the sample moments

e.g. $\hat{\theta}_i = \hat{\theta}(\hat{\mu}_1', \hat{\mu}_2', \ldots., \hat{\mu}_k') = \theta_i(m_1', m_2', \ldots., m_k'); i = 1,2,\ldots., k \qquad (3.2.5)$

where m_i is the ith moment about origin in the sample.

Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$ are the estimators of respectively.

c) Method of least square estimation (LSE):

The idea of least square estimation emerges from the method of maximum likelihood itself. Consider the ML estimation of μ on the basis of a sample of size n from a normal population N (μ, σ^2)

$$f(y,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \quad Y \geq 0, \sigma > 0$$

The likelihood function is given by

$$L(y,\mu,\sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y-\mu)^2}$$

$$LogL = -\frac{n}{2}\log\sqrt{2\pi}\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y-\mu)^2$$

Maximizing logL implies that $\sum_{i=1}^{n}(y-\mu)^2$ must be minimized. i.e., sum of squares $\sum_{i=1}^{n}(y_i-\mu)^2$ must be least square.

The method of least square estimation is mostly used in estimating the parameters of linear function. This very idea can be translated by considering μitself a linear function of certain parameters βj (j=1, 2… k)

i.e. μ= $\sum_{i=1}^{k} x_j\beta_j$

Where x_j's are some unknown coefficient of β_j's forming a linear function of β_j. In this situation estimation of βj's is in the offing. For estimating βj's, we have to minimize

$$E = \sum_{j=1}^{k}\left(Y_i - \sum_{j=1}^{k} x_j\beta_j\right)^2$$ with respect to β_j.

i.e. We have to differentiate the above equation with respect to βj's and equating to zero.

i.e. $\frac{\partial E}{\partial \beta_j} = 0$; j=1,2,…,k.

3.3 Estimation of Parameters of Weibull Distribution using different methods

a) MLE Method:

Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed random variable from two parameter Weibull distribution with the probability density function given by

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha}\right) \quad , x \geq 0, \quad \alpha, \beta > 0 \qquad (3.1.1)$$

The likelihood function of equation (3.3.1) is given by

$$L(x/\alpha,\beta) = \frac{\beta^n}{\alpha^n} \prod_{i=1}^{n} x_i^{\beta-1} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha}\right)$$

The log likelihood function is given by

$$\log L(x/\alpha,\beta) = n\log\beta - n\log\alpha + \sum_{i=1}^{n} x_i^{\beta} - \frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \qquad (3.3.2)$$

For determining the MLE of α we differentiate eq. (3.3.2) with respect to α and equating to zero, we get

$$-\frac{n}{\alpha} + \frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha^2} = 0 \qquad \Rightarrow \hat{\alpha} = \frac{\sum_{i=1}^{n} x_i^{\beta}}{n} \qquad (3.3.3)$$

b) ***Method of moments:***

Although in many cases the method of moments estimator (MME) is superseded by Fisher's MLE concerning asymptotic unbiasedness and minimal variance, the method of moments estimators can, in many cases and quite accurately, be derived by hand.

If the numbers $(x_1, x_2, \ldots, x_n)$ represents a set of data, then an unbiased estimator for the kth moment about origin is

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (3.3.4)$$

where $\hat{m}_k$ stands for the estimate of m_k. In two parameter Weibull distribution, the kth moment about origin is given by

$$\mu_k' = e(x^k) = \int_o^{\infty} x^k f(x, \alpha, \beta) dx$$

$$\Rightarrow \qquad \mu_k' = \frac{\beta}{\alpha}\int_o^{\infty} x^{(k+\beta)-1} \exp\left(-\frac{x^\beta}{\alpha}\right) dx$$

$$\Rightarrow \qquad \mu_k' = \alpha^{\frac{k}{\beta}}\Gamma\left(\frac{k}{\beta}+1\right) \qquad (3.3.5)$$

Where Γ. represents the gamma function and $\Gamma s = \int_0^{\infty} x^{s-1}e^{-x}dx, \quad s > 0$ (Casella and Berger 2002).

If k = 1, then equation (3.3.5) becomes

$$\mu_1' = \alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)$$

If k = 2, then equation (3.3.5) becomes

$$\mu_2' = \alpha^{\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right)$$

Thus the variance is given by

$$\sigma^2 = \mu_2 = \mu_2' - \mu_1'^2$$

$$\Rightarrow \qquad \sigma^2 = \alpha^{\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right) - \left(\alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right)^2$$

$$\Rightarrow \qquad \sigma^2 = \alpha^{\frac{2}{\beta}}\left[\Gamma\left(\frac{2}{\beta}+1\right) - \left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2\right]$$

When we divide σ^2 by $\mu_1'^2$ (e.g., Murthy, Xie, and Jiang (2004)), we get an expression which is a function of β only

$$\frac{\sigma^2}{\mu_1'^2} = \frac{\alpha^{\frac{2}{\beta}}\left[\Gamma\left(\frac{2}{\beta}+1\right)-\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2\right]}{\left[\alpha^{\frac{2}{\beta}}\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2\right]}$$

$$\Rightarrow \quad \frac{\sigma^2}{\mu_1'^2} = \frac{\Gamma\left(\frac{2}{\beta}+1\right)-\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2}{\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2} \qquad (3.3.6)$$

On taking the square roots of (3.3.6), we have the coefficient of variation

$$\frac{\sigma}{\mu_1'} = \frac{\sqrt{\left(\Gamma\left(\frac{2}{\beta}+1\right)-\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^2\right)}}{\left(\Gamma\left(\frac{1}{\beta}+1\right)\right)} \qquad (3.3.7)$$

Now, we can form a table for various CV by using (3.3.7) for different β values. In order to estimate β and α, we need to calculate the coefficient of variation (CV)d of the data on hand. Having done this, we compare (CV)d with CV using the table. The corresponding β is the estimated one $\hat{\beta}$. The scale parameter (α) can then be estimated using the following equation.

i.e. $m_1' = \mu'_1$

If k= 1, in equation (3.3.4), then $\mu_1' = \bar{x}$.

and if k = 1 in eq. (3.3.5), then $\mu_1' = \alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)$

Thus, $m_1' = \mu'_1$

$$\bar{x} = \alpha^{\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)$$

Where $\bar{x}$ is the mean of the data.

$$\therefore \qquad \hat{\alpha} = \left(\frac{\bar{x}}{\Gamma\left(\frac{1}{\beta}+1\right)} \right)^{\beta}$$

c) Least square Estimation:

The last estimator is computed by least square estimation. This method is the simplest of the three in this chapter. Although it has neither the asymptotic properties of the MLE nor the accuracy of the MME estimator, it is quick, simple, and fairly accurate. This method was specifically used in the past as a method that could be done by hand. Thus, it is safe to assume that this method's accuracy will be mediocre in comparison to the previous methods. Let $(x_1, x_2, \ldots, x_n)$ be a random sample of size n from the Weibull distribution with probability density function

$$f\left(x, \alpha, \beta\right) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha} \right) \alpha, \beta > 0, x \geq 0$$

Its distribution function is given by

$$F(x) = \int_o^x f(x)dx$$

$$\Rightarrow \qquad F(x) = \frac{\beta}{\alpha} \int_o^x x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha}\right) dx \qquad (3.3.8)$$

Put $\qquad \frac{x^{\beta}}{\alpha} = \mathrm{Z} \qquad (3.3.9)$

$$\Rightarrow \qquad \beta \frac{x^{\beta-1}}{\alpha} dx = d\mathrm{Z} \qquad (3.3.10)$$

Now using equation (3.3.9) and (3.3.10) in equation (3.3.8), we get

$$F(x) = \int_o^z e^{-z}\, d\mathrm{Z}$$

$$\Rightarrow \qquad F(x) = \left[\frac{e^{-z}}{-1}\right]_o^z = -(e^{-z} - e^{-0})$$

$$\Rightarrow \qquad F(x) = 1 - \exp(-\mathrm{Z})$$

$$\Rightarrow \qquad F(x) = 1 - \exp\left(-\frac{x^{\beta}}{\alpha}\right)$$

Apply natural logarithm by on b/s we have

$$\ln(1-F(x)) = \ln\left(\exp\left(-\frac{x^{\beta}}{\alpha}\right)\right)$$

$$\Rightarrow \qquad \ln(1-F(x)) = \left(-\frac{x^{\beta}}{\alpha}\right)$$

$$\Rightarrow \qquad \ln\left(\frac{1}{1-F(x)}\right) = \frac{x^{\beta}}{\alpha}$$

Consequently

$$\Rightarrow \qquad \ln\ln\left(\frac{1}{1-F(x)}\right) = \ln\left(\frac{x^{\beta}}{\alpha}\right)$$

$$\Rightarrow \qquad \ln\ln\left(\frac{1}{1-F(x)}\right) = \beta\ln(x) - \ln(\alpha) \qquad (3.3.11)$$

Now let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ represent the order statistics of $X_1, X_2 \ldots, X_n$.

That is $X_{(i)}$= the i^{th} smallest of $X_1, X_2, \ldots X_n$ for i = 1, 2, ….,n and the data results in $X_{(i)} = x_{(i)}$. If we were able to approximate the quantities $\ln\ln\left(\frac{1}{1-F(x)}\right)$, we could establish the values of

$$y_i \approx \ln x_{(i)} - \ln\alpha \qquad , i = 1,2,\ldots,n$$

Subsequently, we could choose α and β to minimize the sum of squares due to error.

That is α and β are chosen to minimize

$$E = \sum_{i=1}^{n}\left(y_i - \beta\ln x_{(i)} + \ln\alpha\right)^2 \qquad (3.3.12)$$

In order to estimate parameters α and β we differentiate eq. (3.3.12) w.r.t to α and β respectively and equate to zero.

i.e. $$\frac{\partial E}{\partial \beta} = 0 \qquad \Rightarrow \frac{\partial}{\partial \beta}\sum_{i=1}^{n}(y_i - \beta\ln x_{(i)} + \ln\alpha)^2 = 0$$

$$\Rightarrow \qquad 2\sum_{i=1}^{n}(y_i - \beta\ln x_{(i)} + \ln\alpha)(-\ln x_{(i)}) = 0$$

Since -2 ≠0

$$\Rightarrow \quad \sum_{i=1}^{n}(y_i - \beta \ln x_{(i)} + \ln \alpha)(-\ln x_{(i)}) = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n}(y_i \ln x_{(i)}) - \beta \sum_{i=1}^{n}(\ln x_{(i)})^2 + \ln \alpha \sum_{1=1}^{n} \ln x_{(i)} = 0$$

$$\Rightarrow \quad \beta = \frac{\sum_{i=1}^{n}(y_i \ln x_{(i)}) + \ln \alpha \sum_{1=1}^{n} \ln x_{(i)}}{\sum_{i=1}^{n}(\ln x_{(i)})^2} \qquad (3.3.13)$$

Also, $\frac{\partial E}{\partial \alpha} = 0 \qquad \Rightarrow \frac{\partial}{\partial \alpha} \sum_{i=1}^{n}(y_i - \beta \ln x_{(i)} + \ln \alpha)^2 = 0$

$$\Rightarrow \quad 2\sum_{i=1}^{n}(y_i - \beta \ln x_{(i)} + \ln \alpha)(\frac{1}{\alpha}) = 0$$

Since 2≠0

$$\Rightarrow \quad \sum_{i=1}^{n}(y_i - \beta \ln x_{(i)} + \ln \alpha)(\frac{1}{\alpha}) = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n}(y_i - \beta \ln x_{(i)} + \ln \alpha) = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n}(y_i - \beta \ln x_{(i)}) = -n \ln \alpha$$

$$\Rightarrow \quad n \ln \alpha = \sum_{i=1}^{n}(\beta \ln x_{(i)} - y_i) = n\beta \ln \mathbf{x} - n\bar{y}$$

where $\ln x = \frac{1}{n}\sum_{i=1}^{n} \ln x_{(i)}$ *and* $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$\Rightarrow \quad \ln \alpha == \beta \ln \mathbf{x} - \bar{y} \qquad (3.3.14)$$

$$\Rightarrow \quad \hat{\alpha} = \exp(\beta \ln x - \bar{y}) \qquad (3.3.15)$$

Also for $\hat{\beta}$ we use eq. (3.3.14) in eq. (3.3.13) we get

$$\hat{\beta}\sum_{i=1}^{n}\left(\ln x_{(i)}\right)^2 = \sum_{i=1}^{n} y_i \ln x_{(i)} + \left(\beta \ln x - \bar{y}\right)\sum_{i=1}^{n}\ln x_{(i)}$$

$$= \sum_{i-1}^{n} y_i \ln x_{(i)} + \beta \ln x \sum_{i=1}^{n}\ln x_{(i)} - \bar{y}\sum_{i=1}^{n}\ln x_{(i)}$$

$$\Rightarrow \qquad \beta\sum_{i=1}^{n}\left(\ln x_{(i)}\right)^2 - n\beta(\ln x)^2 = \sum_{i=1}^{n} y_i \ln x_{(i)} - n \ln x \bar{y}$$

Where $\ln x = \frac{1}{n}\sum_{i=1}^{n}\ln x_{(i)}$ *and* $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$\therefore \qquad \hat{\beta} = \frac{\sum_{i=1}^{n} y_i \ln x_i - n \ln x \bar{y}}{\sum_{i=1}^{n}(\ln x_i)^2 - n\left(\ln x_{(i)}\right)^2} \qquad (3.3.16)$$

We conclude that the preceding minimum (equation (3.3.12) is attained when (α ,β) = $\left(\hat{\alpha}, \hat{\beta}\right)$

To apply the foregoing, we need to determine the values of y_i that approximate

$$\Rightarrow \qquad \ln\ln\left(\frac{1}{1-F(x_{(i)})}\right) = \ln\left(-\ln\left(1-F(x_{(i)})\right)\right) \qquad ,i = 1,2,\ldots,n$$

We now present one approach for doing this. Using the fact that

$$\mathrm{E}[F(x_i)] = \frac{i}{(n+1)}$$

And then approximate $F\left(x_{(i)}\right) by\, E\left[F\left(x_{(i)}\right)\right]$.

Therefore, this approach calls for using

$$\Rightarrow \qquad y_i = \ln\left(-\ln\left(1-E(F(x_{(i)}))\right)\right) = \ln\left(-\ln\left(\left(1-\frac{i}{n+1}\right)\right)\right)$$

$$\Rightarrow \qquad y_i = \ln\left(-\ln\left(\left(\frac{n-i+1}{n+1}\right)\right)\right)$$

Substituting these y_i (i=1,2,…., n) into equations (3.3.15) and (3.3.16).we easily obtain $\left(\hat{\alpha}\, and\, \hat{\beta}\right)$. The $\left(\hat{\alpha}\, and\, \hat{\beta}\right)$is called the least square estimates of (α, β).

3.4 Numerical Illustrations

In order to illustrate and compare the methods described earlier, we have coded the three analytical methods MLE, MOM and LSM in R Software. For quantative comparisons of different estimators, mean square error (MSE) was used to test the estimators of three methods. MSE is a measure of accuracy of the estimator. MSE can be calculated as below

$$MSE = \sum_{i=1}^{n} \left\{ \hat{F}(x_i) - F(x_i) \right\}^2$$

Where $\hat{F}(x_i) = 1 - \exp\left(- \frac{x_i^{\hat{\beta}}}{\hat{\alpha}} \right)$ $\qquad$ $f(x_i) = \frac{i - 0.3}{\alpha + 0.4}$

Example: Consider the following example where x_i represents the i^{th} failure time

Table5.1

i	x_i	i	x_i
1	0.438	9	4.508
2	2.413	10	4.981
3	3.073	11	5.115
4	3.079	12	5.592
5	3.137	13	5.848
6	3.198	14	5.958
7	3.918	15	6.013
8	4.287		

Table 5.2 below shows the complete results of the estimates of the shape and scale parameters and the mean squared error (MSE) for each method.

Table 5.2

Method	B	α	MSE
MLE	2.923	4.552	3.57×10^{-3}
MOM	2.941	4.587	3.17×10^{-3}
LSM	1.8515	4.756	8.4×10^{-3}

Since MOM has the minimum MSE, then $\hat{\beta}$=2.941 and $\hat{\alpha}$=4.587.

The objective of our experiments is to compare the three methods namely, LSM, MLE and MOM. We have generated random samples with known parameters. For each sample, we have varied the size from 20 to 100. To be able to compare, we calculated the total deviation (TD) for each method as follows:

$$TD = \left|\frac{\hat{\beta}-\beta}{\beta}\right| + \left|\frac{\hat{\alpha}-\alpha}{\alpha}\right|$$

Where β and α are the known parameters, and $\hat{\alpha}$ and $\hat{\beta}$ are the estimated parameters by any method.

Table 5.3 shows the complete results. The last column of the table shows the best method which yields the minimum total deviation. Notice that the maximum *TD* is 0.55 for all methods. This means that at the worst case the estimated parameters are within ± 50% of their actual values.

Table 5.4 shows a summary of the results of Table 5.3. As it is obviously seen, MOM is the best method. MOM achieves the best estimate 12 times out of 25 which is approximately 50% of the time. Its average deviation from the actual values is 17% with a standard deviation of 14% which is quite good. MOM achieves good results because it involves more calculations and requires more computation time than LSM or MLE. However, for a sample of size 100, MOM takes only few seconds.

Table 5.3 Comparison between LSM, MLE and MOM

β	α	Sizes	LSM			MLE			MOM			Best
			β	A	TD	B	α	TD	β	α	TD	
1	10	20	1.34	7.08	0.63	1.22	7.3	0.5	1.1	7.09	0.4	MOM
		40	0.85	10.96	0.24	0.94	10.74	0.13	0.98	10.8	0.1	MOM
		60	0.84	12.17	0.37	0.926	11.87	0.26	0.94	11.9	0.25	MOM
		80	1.07	10.83	0.16	1.11	10.8	0.21	1.15	10.9	0.24	LSM
		100	0.87	11.18	0.24	0.93	10.97	0.16	0.93	10.8	0.15	MOM
2.3	145	20	2.36	180.3	0.27	2.7	178.4	0.41	2.7	178.1	0.4	LSM
		40	3.4	155.1	0.55	3.3	155.8	0.51	3.33	155.2	0.52	MLE
		60	2.19	132.5	0.13	2.38	131.8	0.13	2.35	131.6	0.11	MOM
		80	2.43	139.06	0.09	2.25	140.5	0.05	2.25	140.1	0.05	MLE
		100	2.44	156.9	0.14	2.62	155.6	0.21	2.6	155.6	0.2	MOM
2.9	357	20	2.91	358.7	0.01	3.35	354.9	0.16	3.33	354.1	0.15	LSM
		40	3.32	365.7	0.14	3.6	364.1	0.25	3.4	363.6	0.19	LSM
		60	2.63	373.8	0.13	2.95	370.5	0.05	2.9	370.4	0.04	MOM
		80	2.62	373.04	0.14	3.03	367.8	0.07	3.0	367.8	0.06	MOM
		100	2.7	367.6	0.1	3.02	363.7	0.06	3.0	367.4	0.05	MOM
3.5	1270	20	3.77	1276.4	0.08	4.4	1270.3	0.27	4.1	1264.6	0.19	LSM
		40	3.12	1200	0.16	3.57	1192.4	0.08	3.4	1189.6	0.09	MLE
		60	3.07	1250.7	0.13	3.3	1248.5	0.06	3.2	1245.3	0.1	MLE
		80	3.09	1304.5	0.14	3.8	1282.8	0.1	3.6	1284.2	0.04	MOM
		100	3.46	1230.9	0.04	3.52	1228.2	0.03	3.57	1223.5	0.05	MLE
1.9	872	20	1.5	888.8	0.22	1.83	867.7	0.03	1.9	870.6	0.01	MOM
		40	2.1	949.6	0.19	2.6	920.9	0.47	2.8	926.7	0.53	LSM
		60	2.3	957.3	0.32	2.25	962.3	0.3	2.25	959.9	0.28	MOM
		80	1.64	907.8	0.17	1.95	884.7	0.04	2.0	890.2	0.08	MLE
		100	1.91	827.3	0.06	1.96	824.8	0.08	1.98	824.8	0.1	LSM

Table 5.4 Summary of Results

Method	No. of times the method gives the best estimate	Average percentage of deviation from actual value (μ)	Standard deviation of the deviations (σ)
LSM	7	21 %	15 %
MLE	6	18 %	15 %
MOM	12	17 %	14 %

In this chapter, we have used three methods for estimating the Weibull distribution parameters. It has been shown from the computational results that the method which gives the best estimates is the method of moments.

4.1. Introduction

The Weibull distribution is one of the most widely used distributions for analyzing lifetime data. It is found to be useful in diverse fields ranging from engineering to medical sciences (see Lawless (2002), Martz and Waller (1982)). The Weibull family is a generalization of the exponential family and can model data exhibiting monotone hazard rate behavior, i.e. it can accommodate three types of failure rates, namely increasing, decreasing and constant. The probability density function of the Weibull distribution is given by:

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha}\right) \qquad , x \geq 0 \qquad (4.1.1)$$

where $\alpha > 0$ and $\beta > 0$ are the scale and shape parameters of the distribution. In Weibull lifetime analysis it is frequent case that the value of shape parameter is known. For example, the exponential and Rayleigh distributions are obtained when β=1 and β=2 respectively. Soland (1968) gives a justification for this situation. The Weibull distribution was studied by Weibull (1951) in connection with the strength of materials; Lieblein and Zelen (1956); Kao (1959) considerd application in reliability and Pike (1966) applications in medicine. Malik (1975) and Frank (1988) have assigned meaning and interpretations for the Weibull distribution. Hallinan (1993) has recently provided an excellent review of the Weibull distribution by presenting historical facts, and the many different forms of this distribution as used by practitioners and possible confusions, errors that arise due to this non-uniqueness.

Maximum Likelihood Estimation has been the most widely used method for estimating the parameters of the Weibull distribution. Recently Bayesian estimation approach has received great attention by most researchers among them are, Ahmed et al. (2011). They considered Bayesian Survival Estimator for Weibull distribution with censored data while Al-Aboud (2009) studied Bayesian estimation for the extreme value distribution using progressive censored data and asymmetric loss. Bayes estimator for exponential distribution with extension of Jeffreys' prior information was considered by Al-Kutubi (2009). Others including, Pandey et a. (2011), Al-Atari (2011), and Hossain and Zimmer (2003) did some comparative studies on the estimation of Weibull parameters using complete and censored samples and Lye et al. (1993) determined Bayes estimation of the Extreme-Value Reliability Function.

4.2 *Bayesian estimation of Weibull distribution under Jeffrey's prior by using different Loss Functions*

Let $(x_1, x_2, \ldots, x_n)$ be a random sample of size n having the probability density function as

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha}\right) \qquad , x \geq 0 \tag{4.2.1}$$

The likelihood function is given by

$$L(x \mid \alpha) = \frac{\beta^n}{\alpha^n} \prod_{i=1}^{n} x_i^{\beta-1} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha}\right)$$

Prior Distribution:

Quite often, the derivation of the prior distribution based on information other than the current data is impossible or rather difficult. Moreover, the statistician may be required to employ as little subjective inputs as possible, so that the conclusion may appear solely based on sampling model and the current data.

Jeffrey's (1946) proposed a formal rule for obtaining a non-informative prior as

$$g(\theta) \quad \alpha \quad \sqrt{\det(I(\theta))}$$

Where θ is k-vector valued parameter and I(θ) is the Fisher's information matrix of order $k \times k$. In particular if θ is a scalar parameter, Jeffrey's non-informative prior for θ is $\sqrt{\det(I(\theta))}$ Thus, in our problem we consider the prior distribution of α to be

$$g(\alpha) \quad \alpha \quad \sqrt{\det(I(\alpha))}$$

$$\Rightarrow \qquad g(\alpha) = k\frac{1}{\alpha}$$

where k is a constant.

The posterior distribution of α is given by

$$p(\alpha \mid x) \quad \alpha \quad L(x \mid \alpha) g(\alpha)$$

$$p(\alpha \mid x) \;\alpha\; \frac{\beta^n}{\alpha^{n+1}} \prod_{i=1}^{n} x_i^{\beta-1} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \right)$$

$$\Rightarrow \qquad p(\alpha \mid x) = \frac{k}{\alpha^{n+1}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \right) \tag{4.2.2}$$

Where k is independent of α.

and $\qquad k^{-1} = \int_0^{\infty} \frac{1}{\alpha^{n+1}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \right) d\alpha$

$$\Rightarrow \qquad k^{-1} = \frac{\Gamma n}{\left(\sum_{i=1}^{n} x_i^{\beta} \right)^n}$$

$$\Rightarrow \qquad k = \frac{\left(\sum_{i=1}^{n} x_i^{\beta} \right)^n}{\Gamma n}$$

After using the value of k in (4.2.2) we get the posterior distribution as

$$\Rightarrow \qquad p(\alpha \mid x) = \frac{\left(\sum_{i=1}^{n} x_i^{\beta} \right)^n}{\alpha^{n+1} \Gamma n} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \right) \tag{4.2.3}$$

a)Estimation under Linex Loss

For determining the Bayes estimate of scale parameter α we will introduce a very useful asymmetric linex loss function given by

$$L(\sigma) = \exp(a\sigma) - a\sigma - 1$$

Where $\sigma = \frac{\hat{\alpha}}{\alpha} - 1, \qquad a \neq 0$

To obtain the Bayes estimator, we minimize the posterior expected loss given by

$$\rho = E\{L(\sigma)\} = \int_0^\infty L(\sigma)\, p(\alpha \mid x) d\alpha$$

$$\Rightarrow \qquad \rho = \int_0^\infty \{\exp(a\sigma) - a\sigma - 1\} \frac{\exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \right) \left(\sum_{i=1}^{n} x_i^{\beta} \right)^n}{\alpha^{n+1} \Gamma n} d\alpha$$

$$\Rightarrow \qquad \rho = \int_0^\infty \left\{ \exp\left(a\left(\frac{\hat{\alpha}}{\alpha} - 1 \right) \right) - a\left(\frac{\hat{\alpha}}{\alpha} - 1 \right) - 1 \right\} \frac{\exp\left(-\frac{t}{\alpha} \right) (t)^n}{\alpha^{n+1} \Gamma n} d\alpha$$

Where $\sigma = \frac{\hat{\alpha}}{\alpha} - 1$ and $t = \sum_{i=1}^{n} x_i^{\beta-1}$

$$\Rightarrow \qquad \rho = \frac{t^n}{\Gamma n} \int_0^\infty \frac{1}{\alpha^{n+1}} \exp\left(-\left(\frac{t - a\hat{\alpha}}{\alpha} \right) \right) \exp(-a) d\alpha - \frac{t^n}{\Gamma n} a\hat{\alpha} \int_0^\infty \frac{1}{\alpha^{(n+1)+1}} \exp\left(-\frac{t}{\alpha} \right) d\alpha$$
$$+ \frac{at^n}{\Gamma n} \int_0^\infty \frac{1}{\alpha^{n+1}} \exp\left(-\frac{t}{\alpha} \right) d\alpha - \frac{t^n}{\Gamma n} \int_0^\infty \frac{1}{\alpha^{n+1}} \exp\left(-\frac{t}{\alpha} \right) d\alpha$$

$$\Rightarrow \qquad \rho = \frac{t^n \exp(-a)}{(t - a\hat{\alpha})^n} - \frac{t^n}{\Gamma n} a\hat{\alpha} \frac{\Gamma(n+1)}{(t)^{n+1}} + \frac{at^n}{\Gamma n} \frac{\Gamma n}{t^n} - \frac{t^n}{\Gamma n} \frac{\Gamma n}{t^n}$$

$$\Rightarrow \qquad \rho = \frac{t^n \exp(-a)}{(t - a\hat{\alpha})^n} - \frac{na\hat{\alpha}}{t} + a - 1 \qquad (4.2.4)$$

Now solving $\frac{\partial \rho}{\partial \hat{\alpha}} = 0$ in eq. (4.2.4), we obtain the Bayes estimator as

$$\frac{\partial \rho}{\partial \hat{\alpha}} = 0$$

$$\Rightarrow \qquad t^n \exp(-a)(-n(t - a\hat{\alpha})^{-n-1})(-a) - \frac{na}{t} = 0$$

$$\Rightarrow \qquad nat^n \exp(-a)(t - a\hat{\alpha})^{-n-1} = \frac{na}{t}$$

$$\Rightarrow \qquad t^{n+1} \exp(-a) = (t - a\hat{\alpha})^{n+1}$$

$$\Rightarrow \qquad t(\exp(-a))^{\frac{1}{n+1}} = (t - a\hat{\alpha})$$

$$\Rightarrow \qquad a\hat{\alpha} = (t - t(\exp(-a))^{\frac{1}{n+1}})$$

$$\Rightarrow \qquad \hat{\alpha} = \frac{t}{a}(1 - (\exp(-a))^{\frac{1}{n+1}})$$

$$\Rightarrow \qquad \hat{\alpha}_{BL} = \frac{\sum_{i=1}^{n} x_i^{\beta}}{a}\left\{1 - \exp\left(-\frac{a}{n+1}\right)\right\} \qquad (4.2.5)$$

b) Estimation under squared error loss function

The squared error loss function (SELF) was proposed by legendere (1805) and Gauss to develop least square theory. Later, it was used in estimation problems when unbiased estimations of θ were evaluated in terms of the risk function $R(\theta, a)$ which becomes nothing but the variance of the estimator.

In our problem SELF is given by

$$l(\hat{\alpha} - \alpha) = c(\hat{\alpha} - \alpha)^2 \qquad (4.2.6)$$

By using the squared error loss function $l(\hat{\alpha} - \alpha) = c(\hat{\alpha} - \alpha^2)$, The Risk function is given by

$$\rho = R(\hat{\alpha} - \alpha) = El(\hat{\alpha} - \alpha)^2$$

$$\rho = E(L(\alpha)) = \int_0^{\infty} l(\hat{\alpha} - \alpha)\, p(\alpha \mid x) d\alpha$$

$$\Rightarrow \qquad \rho = \int_0^{\infty} c(\hat{\alpha} - \alpha)^2 \frac{\left(\sum_{i=1}^{n} x_i^{\beta}\right)^n}{\Gamma n} \frac{1}{\alpha^{n+1}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha}\right) d\alpha$$

$$\Rightarrow \qquad \rho = \frac{c\hat{\alpha}^2 t^n}{\Gamma n}\int_0^{\infty} \frac{1}{\alpha^{n+1}} \exp\left(-\frac{t}{\alpha}\right) d\alpha - 2\frac{c\hat{\alpha} t^n}{\Gamma n}\int_0^{\infty} \frac{1}{\alpha^{n}} \exp\left(-\frac{t}{\alpha}\right) d\alpha + \phi(\alpha)$$

Where $\phi(\alpha) = \frac{ct^n}{\Gamma n}\int \frac{1}{\alpha^{n-1}} \exp\left(-\frac{t}{\alpha}\right) d\alpha$ and $t = \sum_{i=1}^{n} x_i^{\beta-1}$

$$\Rightarrow \qquad \rho = \frac{c\hat{\alpha}^2 t^n}{\Gamma n}\frac{\Gamma n}{t^n} - 2\frac{c\hat{\alpha} t^n}{\Gamma n}\frac{\Gamma(n-1)}{t^{n-1}} + \phi(\alpha)$$

$$\Rightarrow \qquad \rho = c\hat{\alpha}^2 - 2\frac{c\hat{\alpha} t}{n-1} + \phi(\alpha) \qquad (4.2.7)$$

Now solving $\frac{\partial \rho}{\partial \hat{\alpha}} = 0$ in eq. (4.2.7), we obtain the Bayes estimator as

$$\frac{\partial \rho}{\partial \hat{\alpha}} = 0$$

$$\Rightarrow \quad \frac{\partial}{\partial \hat{\alpha}}(c\hat{\alpha}^2 - 2\frac{c\hat{\alpha}t}{n-1} + \phi(\alpha)) = 0$$

$$\Rightarrow \quad 2c\hat{\alpha} - 2\frac{ct}{n-1} = 0$$

$$\Rightarrow \quad \hat{\alpha}_{BS} = \frac{t}{n-1}$$

$$\Rightarrow \quad \hat{\alpha}_{BS} = \frac{\sum_{i=1}^{n} x_i^{\beta}}{n-1} \qquad (4.2.8)$$

4.3 ***Bayesian estimation of Weibull distribution under Extension of Jeffrey's prior by using different Loss Functions***

Let $(x_1, x_2, \ldots, x_n)$ be a random sample of size n having the probability density function as

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha} x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\alpha}\right) \qquad , x \geq 0 \qquad (4.3.1)$$

The likelihood function is given by

$$L(x \mid \alpha) = \frac{\beta^n}{\alpha^n} \prod_{i=1}^{n} x_i^{\beta-1} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha}\right)$$

Thus, in our problem we consider the prior distribution of α to be

$$g(\alpha) \;\; \alpha \;\; [\det | I(\alpha) |]^c \quad , c \in R^+$$

$$\Rightarrow \quad g(\alpha) = k\frac{1}{\alpha^{2c}}$$

where k is a constant.

The posterior distribution of α is given by

$$p(\alpha \mid x) \quad \alpha \quad L(x \mid \alpha) g(\alpha)$$

$$\therefore \qquad p(\alpha \mid x) \ \alpha \ \frac{\beta^n}{\alpha^{n+2c}} \prod_{i=1}^{n} x_i^{\beta-1} \exp\left(- \frac{\sum_{i=1}^{n} x_i^{\beta-1}}{\alpha} \right)$$

$$\Rightarrow \qquad p(\alpha \mid x) = \frac{k}{\alpha^{n+2c}} \exp\left(- \frac{\sum_{i=1}^{n} x_i^{\beta-1}}{\alpha} \right) \tag{4.3.2}$$

Where k is independent of α.

And $$k^{-1} = \int_0^{\infty} \frac{1}{\alpha^{n+2c}} \exp\left(- \frac{\sum_{i=1}^{n} x_i^{\beta-1}}{\alpha} \right) d\alpha$$

$$\Rightarrow \qquad k^{-1} = \int_0^{\infty} \frac{1}{\alpha^{n+2c-1+1}} \exp\left(- \frac{\sum_{i=1}^{n} x_i^{\beta-1}}{\alpha} \right) d\alpha$$

$$\Rightarrow \qquad k^{-1} = \frac{\Gamma(n+2c-1)}{\left(\sum_{i=1}^{n} x_i^{\beta} \right)^{n+2c-1}}$$

$$\Rightarrow \qquad k = \frac{\left(\sum_{i=1}^{n} x_i^{\beta} \right)^{n+2c-1}}{\Gamma(n+2c-1)}$$

After using the value of k in (4.3.2) we get the posterior distribution as

$$\Rightarrow \qquad p(\alpha \mid x) = \frac{\left(\sum_{i=1}^{n} x_i^{\beta} \right)^{n+2c-1}}{\alpha^{n+2c} \Gamma(n+2c-1)} \exp\left(- \frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha} \right) \tag{4.3.3}$$

a) Estimation under Linex Loss

For determining the Bayes estimate of scale parameter α we will introduce a very useful asymmetric linex loss function given by

$$L(\sigma) = \exp(a\sigma) - a\sigma - 1$$

Where $\sigma = \frac{\hat{\alpha}}{\alpha} - 1, \qquad a \neq 0$

To obtain the Bayes estimator, we minimize the posterior expected loss given by

$$\rho = E\{L(\sigma)\} = \int_0^\infty L(\sigma)\, p(\alpha \mid x) d\alpha$$

$$\Rightarrow \qquad \rho = \int_0^\infty \{\exp(a\sigma) - a\sigma - 1\} \frac{\exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta-1}}{\alpha} \right) \left(\sum_{i=1}^{n} x_i^{\beta} \right)^{n+2c-1}}{\alpha^{n+2c}\Gamma(n+2c-1)} d\alpha$$

$$\Rightarrow \qquad \rho = \int_0^\infty \left\{ \exp\left(a\left(\frac{\hat{\alpha}}{\alpha} - 1 \right) \right) - a\left(\frac{\hat{\alpha}}{\alpha} - 1 \right) - 1 \right\} \frac{\exp\left(-\frac{t}{\alpha} \right)(t)^{n+2c-1}}{\alpha^{n+2c}\Gamma(n+2c-1)} d\alpha$$

Where $\sigma = \frac{\hat{\alpha}}{\alpha} - 1$ and $t = \sum_{i=1}^{n} x_i^{\beta}$

$$\Rightarrow \rho = \frac{t^{n+2c-1}}{\Gamma(n+2c-1)} \int_0^\infty \frac{1}{\alpha^{(n+2c-1)+1}} \exp\left(-\left(\frac{t - a\hat{\alpha}}{\alpha} \right) \right) \exp(-a) d\alpha - \frac{t^{n+2c-1}}{\Gamma(n+2c-1)} a\hat{\alpha} \int_0^\infty \frac{1}{\alpha^{n+2c+1}} \exp\left(-\frac{t}{\alpha} \right) d\alpha$$
$$+ \frac{at^{n+2c-1}}{\Gamma(n+2c-1)} \int_0^\infty \frac{1}{\alpha^{(n+2c-1)+1}} \exp\left(-\frac{t}{\alpha} \right) d\alpha - \frac{t^{n+2c-1}}{\Gamma(n+2c-1)} \int_0^\infty \frac{1}{\alpha^{(n+2c-1)+1}} \exp\left(-\frac{t}{\alpha} \right) d\alpha$$

$$\Rightarrow \qquad \rho = \frac{t^{n+2c-1}\exp(-a)}{(t - a\hat{\alpha})^{n+2c-1}} - \frac{(n+2c-1)a\hat{\alpha}}{t} + a - 1 \qquad (4.3.4)$$

Now solving $\frac{\partial \rho}{\partial \hat{\alpha}} = 0$ in eq. (4.3.4), we obtain the Bayes estimator as

$$\frac{\partial \rho}{\partial \hat{\alpha}} = 0$$

$$\Rightarrow \qquad t^{n+2c-1}\exp(-a)(-(n+2c-1))(t - a\hat{\alpha})^{-n-2c})(-a) - \frac{(n+2c-1)a}{t} = 0$$

$$\Rightarrow \qquad t^{n+2c-1}\exp(-a)(t - a\hat{\alpha})^{-n-2c}(n+2c-1)a = \frac{(n+2c-1)a}{t}$$

$$\Rightarrow \quad t^{n+2c}\exp(-a) = (t - a\hat{\alpha})^{n+2c}$$

$$\Rightarrow \quad t\left(\exp(-a)\right)^{\frac{1}{n+2c}} = (t - a\hat{\alpha})$$

$$\Rightarrow \quad a\hat{\alpha} = (t - t\left(\exp(-a)\right)^{\frac{1}{n+2c}})$$

$$\Rightarrow \quad \hat{\alpha}_{be} = \frac{t}{a}(1 - \left(\exp(-a)\right)^{\frac{1}{n+2c}})$$

$$\Rightarrow \quad \hat{\alpha}_{be} = \frac{t}{a}\left(1 - \left(\exp\left(-\frac{a}{n+2c}\right)\right)\right)$$

$$\Rightarrow \quad \hat{\alpha}_{be} = \frac{\sum_{i=1}^{n} x_i}{a}\left(1 - \left(\exp\left(-\frac{a}{n+2c}\right)\right)\right) \qquad (4.3.5)$$

Which is the required bayes estimate of scale parameter α under extension of Jeffrey's prior.

b) Estimation under squared error loss function

In our problem SELF is given by

$$l(\hat{\alpha} - \alpha) = c(\hat{\alpha} - \alpha)^2$$

By using the squared error loss function $l(\hat{\alpha} - \alpha) = c(\hat{\alpha} - \alpha^2)$, The Risk function is given by

$$\rho = \ R(\hat{\alpha} - \alpha) = El(\hat{\alpha} - \alpha)^2$$

$$\Rightarrow \quad \rho = E\left(L(\alpha)\right) = \int_0^{\infty} l\left(\hat{\alpha} - \alpha\right)\, p\left(\alpha \mid x\right) d\alpha$$

$$\Rightarrow \quad \rho = \int_0^{\infty} c\left(\hat{\alpha} - \alpha\right)^2 \frac{\left(\sum_{i=1}^{n} x_i^{\beta}\right)^{n+2c-1}}{\Gamma\left(n+c-1\right)} \frac{1}{\alpha^{n+2c}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha}\right) d\alpha$$

$$\Rightarrow \quad \rho = \frac{c\hat{\alpha}^2 t^{n+2c-1}}{\Gamma(n+2c-1)} \int_0^{\infty} \frac{1}{\alpha^{n+2c}} \exp\left(-\frac{t}{\alpha}\right) d\alpha - 2\frac{c\hat{\alpha} t^{n+2c-1}}{\Gamma(n+2c-1)} \int_0^{\infty} \frac{1}{\alpha^{n+2c-1}} \exp\left(-\frac{t}{\alpha}\right) d\alpha + \psi(\alpha)$$

Where $\psi(\alpha) = \frac{ct^{n+2c-1}}{\Gamma(n+2c-1)} \int_0^\infty \frac{1}{\alpha^{n+2c-2}} \exp\left(-\frac{t}{\alpha}\right) d\alpha$ and $t = \sum_{i=1}^{n} x_i^{\beta}$

$$\Rightarrow \qquad \rho = c\hat{\alpha}^2 - 2\frac{c\hat{\alpha}t}{n+2c-2} + \psi(\alpha) \tag{4.3.6}$$

Now solving $\frac{\partial \rho}{\partial \hat{\alpha}} = 0$ in eq. (4.3.6), we obtain the Bayes estimator as

$$\frac{\partial \rho}{\partial \hat{\alpha}} = 0$$

$$\Rightarrow \qquad \frac{\partial}{\partial \hat{\alpha}}\left(c\hat{\alpha}^2 - 2\frac{c\hat{\alpha}t}{n+2c-2} + \psi(\alpha)\right) = 0$$

$$\Rightarrow \qquad 2c\hat{\alpha} - 2\frac{ct}{n+2c-2} = 0$$

$$\Rightarrow \qquad \hat{\alpha}_{BL} = \frac{t}{n+2c-2}$$

$$\Rightarrow \qquad \hat{\alpha}_{BL} = \frac{\sum_{i=1}^{n} x_i^{\beta}}{n+2c-2} \tag{4.3.7}$$

Which is required Bayes estimator of α under extension of Jeffrey's prior.

4.4 Bayesian Credible Regions for Parameters of Weibull Distribution

Consider the likelihood of Weibull distribution

$$p(y \mid \alpha, \beta) = \left(\frac{\beta}{\alpha}\right)^n \lambda^{\beta-1} \exp\left(-\frac{\sum_{i=1}^{n} y_i^{\beta}}{\alpha}\right) \tag{4.4.1}$$

where $\lambda = \prod_{i=1}^{n} y_i$.

Using Jeffrey's prior $p(\alpha, \beta) = \frac{1}{\alpha\beta}$, the joint posterior of α and β is given by

$$p(\alpha,\beta \mid y) \quad \alpha \quad \beta^{n-1}\lambda^{\beta-1}\alpha^{-(n+1)} \exp\left(-\frac{\sum_{i=1}^{n} y_i^{\beta}}{\alpha}\right) \qquad (4.4.2)$$

Credible region β: Integrating out α in (3.3.2), we have the marginal posterior of β given out by

$$p_1(\beta \mid y) = K_\beta \frac{\beta^{n-1}\lambda^{\beta-1}}{\left(\sum_{i=1}^{n} y_i^{\beta}\right)^n} \quad , \quad \beta > 0 \qquad (4.4.3)$$

where

$$K_\beta^{-1} = \int_0^\infty \frac{\beta^{n-1}\lambda^{\beta-1}}{\left(\sum_{i=1}^{n} y_i^{\beta}\right)^n} d\beta \qquad (4.4.4)$$

Using the squared-error loss function, we can compute Bayes posterior mean

$$E(\beta \mid y) = K_\beta \int_0^\infty \frac{\beta^{n-1}\lambda^{\beta-1}}{\left(\sum_{i=1}^{n} y_i^{\beta}\right)^n} d\beta \qquad (4.4.5)$$

At this point we resort to numerical integration of (4.4.4) and (4.4.5) since the integrals involved do not exist in closed forms. We may also use (4.3.3) to tabulate the posterior p_1 and hence find its mode as well as various intervals for β using a numerical integration computer routine.

We may find *100(1-α)%* equal-tail credible interval $[c_1^{(\beta)}, c_2^{(\beta)}]$ such that

$$\int_o^{c_1^{(\beta)}} p_1(\beta \mid y)\, d\beta = \int_{c_2^{(\beta)}}^\infty p_2(\beta \mid y)\, d\beta = \frac{\alpha}{2}$$

We may further find *100(1-α)%* shortest credible interval $[h_1^{(\beta)}, h_2^{(\beta)}]$ where $h_1^{(\beta)}$ *and* $h_2^{(\beta)}$ satisfy

$$p_1(h_1^{(\beta)} \mid y) = p_1(h_2^{(\beta)} \mid y)$$

as well as

$$\int_{h_1^{(\beta)}}^{h_2^{(\beta)}} p(\beta \mid y)\, d\beta = 1-\alpha$$

We know that if $p(\beta|y)$ is unimodal, the shortest credible interval $[h_1^{(\beta)}, h_2^{(\beta)}]$ is also called the HPD-interval.

Credible region for α : We can obtain marginal posterior of α by integrating (4.3.1) with respect to β

$$p_2(\alpha \mid y) = K_\alpha\, \alpha^{-(n+1)} \int \beta^{n-1} \lambda^{\beta-1} \exp\left(-\frac{\sum_{i=1}^{\infty} y_i^{\beta}}{\alpha} \right) d\beta \ , \alpha > 0 \quad (4.4.6)$$

where

$$K_\alpha^{-1} = \Gamma n \int_0^\infty \frac{\beta^{n-1} \lambda^{\beta-1}}{\left(\sum_{i=1}^{n} y_i^{\beta} \right)} d\beta \qquad (4.4.7)$$

The posterior mean is given by

$$E(\alpha \mid y) = K_\alpha \Gamma(n-1) \int_0^\infty \frac{\beta^{n-1} \lambda^{\beta-1}}{\left(\sum_{i=1}^{n} y_i^{\beta} \right)^{n-1}} d\beta \qquad (4.4.8)$$

Numerical integration allows us to tabulate $p_2(\alpha \mid y)$ and $E(\alpha \mid y)$ so that we may also numerically determine the mode, compute $[c_1^{(\alpha)}, c_2^{(\alpha)}]$, an equal tail credible interval of confidence (1-α) where

$$\int_o^{c_1^{(\alpha)}} p_1(\alpha \mid y)\, d\alpha = \int_{c_2^{(\alpha)}}^{\infty} p_2(\alpha \mid y)\, d\alpha = \frac{\alpha}{2}$$

and the shortest credible (or HPD, if the posterior of α is unimodal) interval $[h_1^{(\alpha)}, h_2^{(\alpha)}]$ where $h_1^{(\alpha)}$ *and* $h_2^{(\alpha)}$ are such that

$$p_2(h_1^{(\alpha)} \mid y) = p_2(h_2^{(\alpha)} \mid y)$$

and

$$\int_{h_1^{(\alpha)}}^{h_2^{(\alpha)}} p_2(\alpha \mid y)\, d\alpha = 1-\alpha$$

4.5 Simulation Study

In our simulation study, we chose a sample size of n=25, 50 and 100 to represent small, medium and large data set. The scale parameter is estimated for Weibull distribution with Maximum Likelihood and Bayesian using Jeffrey's & extension of Jeffrey's prior methods. For the scale parameter we have considered α= 0.5 and 1.5. The Shape parameter β has been fixed at 0.8 and 1.2. The values of Jeffrey's extension were c = 0.4 and 1.4. The value for the loss parameter a = ± 0.6 and ±1.6. This was iterated 1000 times and the scale parameter for each method was calculated. The results are presented in tables for different selections of the parameters and c extension of Jeffrey's prior.

Table 4.1: Mean Squared Error for $(\hat{\alpha})$ under Jeffrey's Prior.

n	α	B	$\hat{\alpha}_{ML}$	$\hat{\alpha}_{BL}$ $a=0.6$	$\hat{\alpha}_{BL}$ $a=-0.6$	$\hat{\alpha}_{BL}$ $a=1.6$	$\hat{\alpha}_{BL}$ $a=-1.6$	$\hat{\alpha}_{BS}$
25	0.5	0.8	0.06714	0.06382	0.06531	0.06261	0.06658	0.06994
	0.5	1.2	0.05943	0.05649	0.05781	0.05542	0.05894	0.06191
	1.5	0.8	0.08946	0.08503	0.08702	0.08343	0.08872	0.09319
	1.5	1.2	0.05943	0.05649	0.05781	0.05781	0.05542	0.06191
50	0.5	0.8	0.04461	0.04347	0.04399	0.04305	0.04442	0.04551
	0.5	1.2	0.03797	0.03701	0.03746	0.03665	0.03782	0.03875
	1.5	0.8	0.10740	0.10467	0.10591	0.10369	0.10696	0.10959
	1.5	1.2	0.18145	0.17685	0.17895	0.17513	0.18072	0.18516
100	0.5	0.8	0.05302	0.05234	0.05266	0.05209	0.05292	0.05356
	0.5	1.2	0.04451	0.04393	0.04419	0.04371	0.04441	0.04495
	1.5	0.8	0.13590	0.13415	0.13495	0.13349	0.13562	0.13727
	1.5	1.2	0.14073	0.13893	0.13975	0.13824	0.14045	0.14215

ML= Maximum Likelihood, BL= LINEX Loss Function, BS=Squared Error Loss Function

Table 5.2: Mean Squared Error for $(\hat{\alpha})$ under extension of Jeffrey's prior.

n	α	c	β	$\hat{\alpha}_{ML}$	$\hat{\alpha}_{BL}$ $a=0.6$	$\hat{\alpha}_{BL}$ $a=-0.6$	$\hat{\alpha}_{BL}$ $a=1.6$	$\hat{\alpha}_{BL}$ $a=-1.6$	$\hat{\alpha}_{BS}$
25	0.5	0.4	0.8	0.06714	0.06431	0.06582	0.06308	0.06712	0.07052
	0.5	0.4	1.2	0.05943	0.05693	0.05827	0.05584	0.05941	0.06243
	0.5	1.4	0.8	0.06714	0.05973	0.06103	0.05867	0.06215	0.06506
	0.5	1.4	1.2	0.05943	0.05287	0.54033	0.05194	0.05502	0.05759
	1.5	0.4	0.8	0.08946	0.08569	0.08770	0.08405	0.08943	0.09397
	1.5	0.4	1.2	0.13369	0.12805	0.13106	0.12561	0.13364	0.14043
	1.5	1.4	0.8	0.08946	0.07959	0.08132	0.07818	0.08281	0.08669
	1.5	1.4	1.2	0.13369	0.11893	0.12153	0.11683	0.12375	0.12954
50	0.5	0.4	0.8	0.04461	0.04364	0.04416	0.04322	0.04461	0.04571
	0.5	0.4	1.2	0.03797	0.037161	0.03761	0.03679	0.03797	0.03891
	0.5	1.4	0.8	0.04461	0.042003	0.04248	0.04161	0.04288	0.04391
	0.5	1.4	1.2	0.03797	0.03576	0.03616	0.03542	0.03651	0.03737
	1.5	0.4	0.8	0.10741	0.10508	0.10633	0.10406	0.10739	0.11004
	1.5	0.4	1.2	0.18145	0.17755	0.17966	0.17581	0.18144	0.18592
	1.5	1.4	0.8	0.10741	0.101131	0.10228	0.10018	0.10326	0.10571
	1.5	1.4	1.2	0.18145	0.17086	0.17861	0.16925	0.17446	0.17861
100	0.5	0.4	0.8	0.05302	0.05245	0.05276	0.05219	0.05302	0.05367
	0.5	0.4	1.2	0.04451	0.04401	0.04428	0.04379	0.04450	0.04504
	0.5	1.4	0.8	0.05302	0.05143	0.05173	0.05118	0.05198	0.05261
	0.5	1.4	1.2	0.04451	0.04316	0.04341	0.04295	0.04362	0.04414
	1.5	0.4	0.8	0.13590	0.13442	0.13522	0.13376	0.13590	0.13755
	1.5	0.4	1.2	0.14073	0.13920	0.14244	0.13851	0.14073	0.14244
	1.5	1.4	0.8	0.13590	0.13181	0.13258	0.131178	0.13323	0.13482
	1.5	1.4	1.2	0.14073	0.13650	0.13730	0.13584	0.13797	0.13962

ML= Maximum Likelihood, BL= LINEX Loss Function, BS= Squared Error Loss Function

In table 4.1, Bayes estimation with LINEX Loss function under Jeffrey's prior provides the smallest values in most cases especially when loss parameter a is1.6. Similarly, in table 4.2, Bayes estimation with LINEX Loss function under extension of Jeffrey's prior provides the smallest values in most cases especially when loss parameter a is1.6 whether the extension of Jeffrey's prior is 0.4 or 1.4.

In this chapter, we have addressed the problem of Bayesian estimation for the Weibull distribution, under asymmetric and symmetric loss functions and that of Maximum Likelihood Estimation. A simulation study was conducted R-software to examine and compare the performance of the estimates for different sample sizes with different values for the Extension of Jeffreys' prior and the loss functions.

From the results, we observe that in most cases, Bayesian Estimator under Linear Exponential Loss function (LINEX) has the smallest Mean Squared Error values for both prior's i.e, Jeffrey's and an extension of Jeffrey's prior information.

5.1 Introduction

Carcinoma is an important disease that causes large number of deaths around the world. The biggest problem with the disease is that it is often not diagnosed at an early stage precluding the chances of cure of the patients. At advanced stages of the disease, the medical practitioners often have limited scope to relieve the patient. Earlier the patients were used to be treated with radiotherapy but later on they were given chemotherapy followed by radiotherapy. The interest therefore centres among the medical practitioners that which of the two therapies provide better results in terms of the survival of the patients. Such studies have been considered earlier by a number of researchers when patient is suffering from advanced stages of carcinoma in different organs. The studies were mainly carried out by medical practitioners or statisticians working with the medical data and often focused on randomized, prospective, retrospective, multi centre clinical trials. Broadly speaking the analyses used both classical and Bayesian methodologies although the former dominate the latter paradigm. Since the studies are numerous covering a variety of cancerous forms, it is not possible to provide an exhaustive list of references and, therefore, we shall be focusing primarily on the studies based on survival data. It is to be noted that the data for such comparisons often come in the form of survival times and, among the various approaches, a better therapy can be suggested to practitioners on the basis of comparison between the corresponding survival functions.

Among the earlier classical developments, one can refer to Kaplan and Meier (1958), Mantel (1966), etc. for estimating and comparing survival functions based on Kaplan-Meier estimate or log rank test. Cox (1972) proposed the concept of proportional hazards model for quantifying the effects of covariates on the survival times. A systematic review of literature can be had from Qian (1994), Lawless (2002), Kalbfleisch and Prentice (2002) among others. On the Bayesian front Sinha and Dey (1997) is an important review article that provides a number of practical problems and correspondingly provides an updated list of related developments. The other important references include Ibrahim et al. (2001), Gelman (2004), etc. Most of these references are intended towards comparing survival functions obtained on the basis of various modelling assumptions for the data.

The present Chapter considers a comparison of survival functions obtained from two therapies assuming Weibull models for the corresponding survival times so

that one can be in a position to say that which therapy is better. The Weibull distribution has a wide range of applicability especially in lifetime data analysis perhaps because of its virtue of versatility or flexibility. The p.d.f. of two-parameter Weibull distribution can be written as

$$f(x,\alpha,\beta) = \frac{\beta}{\alpha}x^{\beta-1}\exp\left(\frac{-x^{\beta}}{\alpha}\right) \qquad (5.1.1)$$

Where α and β determines the scale and shape parameter of the distribution. The corresponding survival function can be given by

$$S(x,\alpha,\beta) = exp\left\{\frac{-x^{\beta}}{\alpha}\right\} \qquad (5.1.2)$$

The Weibull distribution encompasses monotonically increasing (for β >1), decreasing (for β <1), and constant (for β =1) failure rate and, as such, the model has been successfully used to describe both initial failures as well as the failures due to remission or aging (see Lawless (2002)). One of the biggest advantages with the Weibull model is the availability of closed form survival function, which makes the inferences related to the model quite easy although the non-availability of sufficient statistics poses some problem in comparison to those situations where the existence of the same is guaranteed.

The Weibull model is perhaps the richest one as far as the inferential developments are concerned both with regard to classical and Bayesian paradigms. Lawless (2002) is an important text which systematically describes the classical developments based on the model both in the context of engineering and medical applications. A few other important references include Kalbfleisch and Prentice (2002), Lee and Wang (2003), etc. On the Bayesian front, a systematic accountability can be seen in Martz and Waller (1982) and, more recently, in Singpurwalla (2006). The other important references include Gelman et al. (2003), Ibrahim et al. (2004), Arora et al. (2008) etc. although a number of research papers on Weibull distribution appeared regularly in various journals.

The distribution is, in general, not too straightforward to deal with. The classical developments on the model mostly relied on large sample approximations or empirical results. The problem with the Bayesian inference lies in the involvement of integrals in the posterior based inferences, which are difficult to solve analytically and, as such, require specialized techniques of Bayesian computation (see, for example, Upadhyay et al. (2001), and, more recently. Gamerman and Lopes (2006)). This last reference

advocated the use of sample based approaches in Bayesian computation because of their several inherent advantages. A few such advantages may include the straightforwardness of the procedures to deal with censored data problems and routine inferential development for some nonlinear functions of the model parameters.

The Weibull distribution becomes straightforward if one is confronted with a situation where shape parameter β can be taken to be unity. The resulting distribution becomes one-parameter exponential and inferential developments based on it are routine. This is equivalent to say that an experimenter tests β against unity for the given data set and goes for the exponential model if the hypothesis is accepted. Such problems have been considered earlier by a number of authors in both classical and Bayesian paradigms. The most frequently used classical tool for testing β against unity is based on the likelihood ratio test. The earlier cited references do provide enough material on testing β against unity for the Weibull model.

For Bayesians, the obvious technique can be based on the evaluation of Bayes factor which is a bit difficult when the priors are non-informative and the data are compounded with censoring mechanism (see, for example, Upadhyay and Mukherjee (2008)). The problem of testing β =l can also be visualized as that of model comparison where one can use, for example, the Bayes information criterion (BIC) (see, for example, Schwarz (1978)) for drawing the necessary conclusion. No doubt, this measure is comparatively easy and provides answers parallel to that based on Bayes factor.

A model comparison is justified among the compatible models only where compatibility is referred to mean that all the models under consideration do provide an adequate representation to the given data. Therefore, we first propose a compatibility study of the exponential and Weibull models and then provide a model comparison study to pick up an appropriate model. For studying the compatibility, we have used the posterior Bayes factor based on Bayesian version of chi-square discrepancy measure.

5.2 Model Formulation

TO begin with let us consider a group of *n* patients who have undergone treatments for certain disease. These *n* patients can also be considered prospectively especially when one is indulged with experiments involving clinical trials. It is to be noted that here we primarily considers a group of advanced stage cancer patients treated with either radiotherapy (RT) or chemotherapy followed by radiotherapy (CT+RT). We further assume that out of *n* patients receiving a particular therapy, survival times

$X_i (i = 1,2,...,r)$ for r patients are observed completely whereas for remaining *n-r* patients, we simply have the information that j[th] patient left the study at the censoring time, say C_j (j=r+l,...,n). If we consider x's to follow Weibull model with parameters (α, β), the corresponding likelihood function (LF) can be written as

$$l(x|\alpha,\beta) \propto \frac{\beta^r}{\alpha^r} \prod_{i=1}^{r} {x_i}^{\beta-1} \, exp\left[-\sum_{i=1}^{r} \frac{{xi}^{\beta}}{\alpha}\right] \prod_{j=r+1}^{n} \Pr(x_j > c_j), \quad \alpha, \beta > 0 \qquad (5.2.1)$$

where x is used to denoted the available information on survival times. The term $\Pr(x_j > c_j)$ is the survival function at C_j corresponding to the Weibull model and it is available in a nice closed form. This last term occurs in (5.2.1) because of the censored data and it can be completely removed if there had been no censoring and survival times for all the n patients are observed completely, that is, *r* itself becomes *n*.

The Bayesian formulation next requires appropriate priors for the parameters. If we have enough information that can help us to go for informative prior, it may certainly be preferred over all other choices. Otherwise it is better to stick to non-informative or vague priors. Here, we prefer taking a vague choice on the lines of Upadhyay et al. (2001) (see also Singpurwalla (2006)). The priors considered by the authors are

$$g(\alpha,\beta) \propto (\alpha.\beta)^{-1} \qquad (5.2.2)$$

Combining the LF with the prior via Bayes theorem yields the posterior that can be written upto proportionality as

$$p(\alpha,\beta|x) \propto \frac{\beta^{r-1}}{\alpha^{r+1}} \prod_{i=1}^{r} (x_i)^{\beta-1} \, exp\left\{-\sum_{i=1}^{r} \frac{{x_i}^{\beta}}{\alpha}\right\} \prod_{j=r+1}^{n} exp\left\{\frac{-(c_j)^{\beta}}{\alpha}\right\} \qquad (5.2.3)$$

The posterior given in (5.2.3) can be analyzed by any of the various available techniques (see, for example, Upadhyay et al. (2001)). The solution is not that difficult, as we often require solving only one-dimensional integral whether the interest focuses on joint posterior or the marginal posterior. There are several other approximate techniques (see, for example, Gamerman and Lopes (2006)) which can equally well be applied to obtain the desired inferences from the posterior given in (5.2.3). We, however, advocate the use of sample based approaches, in particular the Gibbs sampler, simply because of its inherent ease.

The Gibbs sampler algorithm is a Markovian updating scheme that proceeds by generating from various full conditionals specified upto proportionality from the joint

posterior, the latter also needs to be specified upto proportionality only. In order to run the algorithm some initial values are assigned at the beginning to the generating variates and then the chain proceeds in a cyclic order using the most recent values of all other variates. The details about the algorithm, its necessary implementation, and the convergence diagnostic issues can be found in Smith and Roberts (1993) and Upadhyay et al. (2001) among others. The algorithm can be implemented either by means of a single long run of the chain or by means of multiple chains of long run and then outcomes can be picked up once the convergence is assured in the generating chain. In a single long run of the chain the outcomes can be picked up from equidistant positions to avoid serial correlation among the generating variates. Similarly, for parallel chains the outcomes can be taken from the same relative positions after the convergence is assured (see, for example, Smith and Roberts (1993)). The final selected outcomes can be regarded as random samples from the joint posterior with components as the random samples from the corresponding marginal posteriors.

The Gibbs sampler algorithm has an apparent advantage when one is interested in the posterior of some non-linear function of the original variates. The analytical derivation of this posterior is often difficult. The Gibbs sampler algorithm suggests that samples from such a posterior can be easily obtained by replacing each parameter in the nonlinear function with the corresponding sample. Thus sample-based estimates can be easily derived once the final samples are made available from the corresponding posteriors. In case of censoring Gibbs sampler can be routinely extended without any extra burden. We apply the scheme on the concerned posterior in a usual way, treating the censored observations as further unknowns. The rest of the developments are same except that new full conditionals are introduced corresponding to the unknown censored data. That is, the full conditionals corresponding to unknown parameters will be same as would have been obtained had there been no censoring. The full conditionals corresponding to independent censored data are, however, the parent sampling distributions truncated in the appropriate regions (see, for example, Upadhyay et al. (2001)). Thus the unknown censored data can be generated as independent draws from the truncated parent sampling distribution.

The implementation of the algorithm for the posterior (5.2.3) is, therefore, quite straightforward. We simply need to think for the full conditionals of α and β. We also need to think for the full conditionals corresponding to unknown censored data. It can be shown that the full conditional of α reduces to gamma distribution after a simple

transformation whereas that of β can be shown to be log-concave. The full conditional corresponding to censored data, say $X_j(>c_j)$, is truncated Weibull distribution in the region (C_J,∞). An apparent advantage of this scheme is that it can be used to assess the unknown censored data exactly the way it does provide information on unknown α and β. That is, once the convergence monitoring is done on all the unknowns, the samples from the generated chains can be used to study the desired features of interest.

As already mentioned, the Weibull distribution reduces to one-parameter exponential distribution when the shape parameter β becomes unity. In this case if we consider the prior for α proportional to α^{-1}, the corresponding posterior can be easily reduced from (5.2.3) by putting β =1. The posterior after a simple reciprocal transformation can be written as

$$p_e(\lambda|x) \propto (\lambda)^{r-1} exp\left[-\lambda\left(\sum_{i=1}^{r}(x_i) + \sum_{j=r+1}^{n}(c_j)\right)\right] \qquad (5.2.4)$$

where $\lambda = \alpha^{-1}$. Obviously, (5.2.4) is gamma density with shape parameter r and scale parameter$\left(\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n}(c_j)\right)^{-1}$. Therefore, the posterior (5.2.4) corresponding to exponential model can be easily managed for any desired inferences.

5.3 Model Compatibility and Comparison

Model compatibility study is meant to see if a model under consideration does provide a good fit to the data in hand and, therefore, provides a valid reason for considering a model. A number of tools have been suggested for studying compatibility of a model in both classical and Bayesian frameworks. An important approach in classical paradigm is to use tail area probability or better known as the p-value based on a goodness of fit test and to replace the unknown parameter(s), if any, involved in the process by some good estimates usually the maximum likelihood (see, for example, Lawless (2002)). Bayesian paradigm offers a number of possibilities for checking model compatibility, the most important being the one based on predictive simulation ideas. The idea suggests that if the observed data and the data predicted from the model exhibit some kind of similarities, the model under consideration can be considered compatible with the observed data (see, for example, Gelman et al. (1996)). Bayesians have also defined a number of versions of p-values analogously to the classical approach but they have suggested integrating out the unknown parameter(s) by some of its possible distributions. These versions are referred to as the prior, posterior, conditional, or partial posterior predictive p-values. Each of these measures has their own merits or demerits but we do not go it to the details of these various aspects due to

space restriction. Gelman et al. (1996), Bayarri and Berger (1998), and Upadhyay and Mukherjee (2008), etc. are some important references for a detailed discussion of these ideas. For the purpose of our illustration, Arora et. Al (2008) considered the use of posterior predictive p-value based on an important classical discrepancy measure in spite of the fact that the measure has invited a few shortcomings too. They simply use it because of its ease and also because of the fact that our compatibility study requires only a tentative answer and the final answer will be based on the result of model comparison. Moreover, as pointed out by a number of authors, the posterior predictive p-value can be used at least for a preliminary check of model compatibility (see, for example, Upadhyay and Peshwani (2008)).

Let the observed data be denoted by x and the predictive data by y, D is the measure of discrepancy between the samples and population values and f (.|θ) be assumed model for the data. Then the Bayesian posterior predictive p- value can be defined as

$$p = Pr\left[(D_2 \geq D_1 | f, x)\right] = \int Pr(D_2 \geq D_1 | f, \theta) p(\theta | f, x) d\theta \qquad (5.3.1)$$

where $p(\theta |$ f, x) is the posterior distribution of θ under the model f, D_1 and D_2 are the measures of discrepancy corresponding to the observed and the predictive data, respectively. Equation (5.3.1) can be regarded as the classical p-value averaged over the posterior distribution of θ under the model f. If we assume, for example, chi- square as a measure of discrepancy (Gelman et al. (1996)), we can write

$$D = \sum_{i=1}^{n} \frac{(d_1 - E(d_i|\theta))^2}{V(d_i|\theta)}, \qquad (5.3.2)$$

where d_i (i=l,..., n) is the i^{th} observation in the considered data set and n is the corresponding sample size. It is to be noted that chi-square discrepancy measure is arbitrarily chosen for illustration only; one can similarly define p-values based on other discrepancy measures as well. Thus using (5.3.2) in (5.3.1), the posterior predictive p-values corresponding to chi-square discrepancy measure can be easily obtained. Our conclusion based on the evaluated p-value will simply be 'larger the p- value, better is the compatibility of the considered model with the observed data' (see also Upadyay and Peshwani (2008)). The integration in (5.3.1) can often be a major difficulty in evaluating the p-value and the situation worsens with the increasing dimensionality of θ. The situation can, however, be easily managed if one resorts to sample based approaches for simulating the posterior p(θ|f, x) and then evaluates the sample-based

estimate of the corresponding p-value. Upadhyay et al. (2001) have provided details of the various steps involved in the evaluation of (5.3.1) using sample based approaches.

In performing compatibility study, it is often seen that a number of models are found compatible with the data in hand. The question, therefore, arises which model should be finally considered for the data. The question, although difficult, can only be answered if one performs some sort of comparison among the competitive models and then accordingly recommends a model. It is to be noted that the results of model compatibility study can never be used for comparing the models rather it can be used only to check if the assumed model is compatible with the data or not. Arora et al. (2008) make a simple comment on parsimony principle which recommends a model which is simplest. Undoubtedly, this principle is quite useful and advocated by a number of authors but sometimes, while recommending a model according to this principle, the experimenter may loose some of the important inferential aspects (see, for example, Upadhyay and Mukherjee (2008)). Here, we shall focus on BIC although a number of other sophisticated tools can also be used for the desired comparison.

5.4 Bayes Information Criterion

The BIC also known as Schwarz criterion is a well-known criterion for comparing the models. According to this criterion, a model is recommended if it minimizes the term given by

$$\text{BIC} = -2(\log(L(\hat{\theta})) + p \log(n \qquad (5.4.1)$$

where $L(\hat{\theta})$ denotes the maximized likelihood function corresponding to a model indexed with the parameter θ, n denotes the total number of observations and p is the dimension of the concerned model. First term supports the more complex model and second term supports a simpler model having low dimensions. It is obvious from (5.4.1) that BIC is free from any prior information and it penalizes the complexity of the model according to its dimension. It is a consistent measure in the sense that the probability of selecting the correct model tends to unity as the number of observations approaches to infinity although it suffers from a disadvantage that it is a valid measure only for a well-behaved model. The quantity $\hat{\theta}$ in (5.4.1) can be replaced by posterior mode if the prior is vague. Similarly, an extension of BIC to censored data problems is routine if one employs sample-based approaches, in particular the Gibbs sampler, and replaces the corresponding censored data with their estimates obtained through Gibbs

run.

5.5 Numerical Illustration

For numerical illustration, we considered a real data set on survival times of patients with stage III non-small cell lung cancer (NSCLC). The data were the results of phase III clinical trial conducted by Cancer and Leukemia group B (CALGB) in United States of America from May 1984 to May 1987 in the form of five interim analyses. The objective of the study included the comparison of two cancer therapies, that is, CT+RT on one hand and RT alone on the other. These clinical trials mostly used log rank tests and Kaplan-Meier plot (see, for example, Lawless (2002)) for the comparison of two therapies at each interim analysis. After the trial stopped enrolling the new patients, enrolled patients were followed up until the summer of 1992 and this data set was finally analyzed by Li (1994), Qian et al. (1996), among others. The complete description of the entire study is given in Li (1994) (see also Qian (1994)). The data set in an ordered form is summarized in Table 5.1 where asterisk with an observation denotes the censored value.

Li (1994) and Qian et al. (1996) (see also Qian (1994)) assumed exponential and Weibull models, respectively, for analyzing the data corresponding to RT. For data corresponding to combined therapy CT+RT, they however assumed the same models but with a restrictive assumption on the scale parameters. The authors assumed that the logarithm of the ratio of the scale parameters for the models corresponding to CT+RT and those corresponding to RT is constant. This assumption makes sense with regard to the exponential model as the ratio becomes simply the hazard ratio but it is certainly not appealing with the assumption of Weibull model. The authors finally considered a comparison of two therapies in a Bayesian framework based on estimated survival functions and concluded that the combined therapy CT+RT does provide a significant improvement over RT in terms of the survival of the patients.

Table 5.1: Survival times of NSCLC patients receiving two different therapies

Therapy	Survival times in days
RT	0.27, 0.37, 1.23, 1.27, 1.27, 2.27, 2.30, 2.40, 2.60, 2.73, 2.87, 2.93, 2.97, 3.37, 3.57, 3.63, 4.23, 4.40, 4.50, 4.83, 5.33, 6.00, 6.10, 6.10, 6.77, 6.87, 6.90, 7.17, 7.50, 7.57, 7.63, 7.67, 8.13, 8.30, 8.53, 8.57, 8.90, 9.50, 9.67, 10.13, 10.27, 10.47, 10.53, 10.67, 10.67, 10.83, 12.63, 12.67, 12.77, 13.10, 13.23, 14.20, 15.00, 15.20, 15.33, 15.83, 16.10, 16.23, 16.87, 17.50, 18.10, 19.73, 19.77, 19.93, 21.43, 23.30, 23.40, 31.20*, 31.93, 32.90, 42.47*, 44.13, 45.40, 62.50*. 64.87*, 73.43*, 83.77* (n=77, r=71)
CT+RT	0.20, 1.83, 2.70, 3.13, 3.90, 3.97, 4.03, 4.50, 5.03, 5.20, 5.93, 6.07, 6.27, 6.33, 6.47, 6.57, 6.70, 7.00, 7.00, 7.20, 7.47, 7.53, 7.97, 8.33, 8.73, 9.03, 9.43, 9.47, 9.50, 9.80, 10.03, 10.10, 10.97, 11.40, 11.67, 12.03, 12.83, 13.30, 13.73, 14.07, 14.57, 15.57, 16.40, 16.53, 16.53, 16.87, 17.23, 17.47, 18.13, 18.53, 18.93, 19.03, 19.07, 20.47, 20.67, 21.20, 23.00, 23.43, 28.83, 39.47*, 40.27, 46.90, 47.83, 48.07, 52.60*, 52.67*, 55.03*, 55.73*, 55.77*, 56.67*, 57.43*, 59.03*, 62.37, 62.40, 66.07*, 66.33*, 69.13* 73.93* (n=78, r=65)

In order to formalize the analysis, we first simulated the posterior (5.2.3) separately for RT and CT+RT data using the Gibbs sampler algorithm. The details of the implementation of the Gibbs sampler algorithm for censored data situations can be seen in Section 5.2. For the initial values of α and β, we considered maximum likelihood estimates using the corresponding data sets whereas for the initial values of unknown censored observations X_j, we used the corresponding truncation points C_j (j= r+1,..., n). We next considered a single chain of long run through Gibbs algorithm and the convergence monitoring was done for both (α, β) and unknown censored observations using the ergodic averages. Finally, samples of size 1000 corresponding to each variate were picked up from the generating chain using equidistant (every 10^{th}) outcomes. The gaps were chosen to make serial correlation negligibly small. These samples can be regarded as random samples from the distributions of the corresponding unknowns. Thus the sample-based inferences can be easily drawn once the samples for the corresponding unknowns are made available. Sample based estimates in the form of posterior modes are shown in Table 5.2. The table also provides the estimated modal values of the corresponding unknown censored observations. All these estimates are based on samples of size 1000 from each of the unknowns and will be used for further

inferences. The estimated survival curve corresponding to each therapy is shown in Figure 5.1 when the underlying model is Weibull distribution. These estimated curves are based on the modal values of the corresponding estimates and have been drawn using R software.

Table 5.2: Estimates based on sample of size 1000 corresponding to RT and CT+RT data when the underlying modeling assumption is Weibull

Variate	Estimates corresponding to	
	RT	CT+RT
α	15.421	26.669
β	0.968	0.926
Censored observations (in an ordered form)	38.554, 44.401, 69.817, 70.358, 77.294, 106.001	46.851, 64.093, 70.201, 71.579, 72.375, 72.624, 72.807, 72.955, 74.980, 77.516, 84.064, 84.382, 90.600

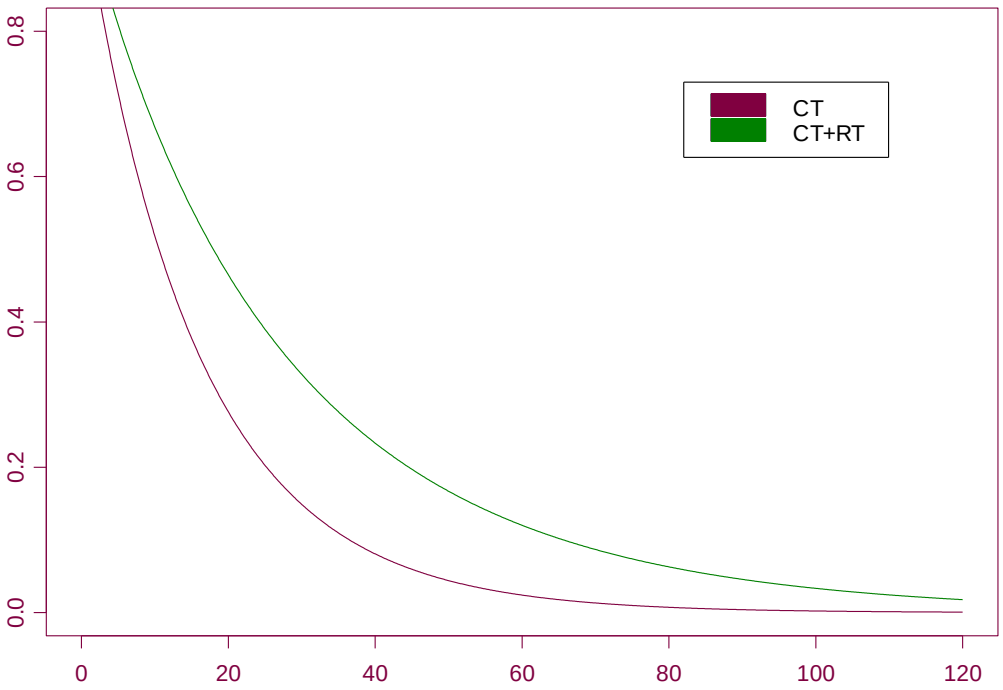

Figure 5.1: Estimated Survival functions when the underlying modelling assumption is Weibull

A number of results can be reported likewise once the samples are made available but we shall concentrate on two important findings based on Table 5.2 and Figure 5.1. First, the estimated posterior mode of β is quite close to unity for both RT

and CT+RT which, in turn, provides an impression that exponential model is a strong candidate for both the data sets. Second, the survival curve corresponding to CT+RT is, in general, higher than the corresponding curve for RT which shows that the combined therapy provides a better survival to the patients suffering from NSCLC (see Figure 5.1).

Guided by our first conclusion above, we propose to consider the exponential modelling assumption (that is β=l in the previous formulation) as well for the proposed analysis and then intend to provide a comparison of the two models so that a better one can be recommended. The corresponding posterior is given in (5.2.4). We separately implemented our strategy for exponential distribution (see Section 5.2) on both RT and CT+RT data and generated a single chain of long run. It is to be noted that exponential distribution has a single parameter but several unknowns in the form of censored observations. We picked up samples of size 1000 from the corresponding distributions of each of the unknowns in a way similar to what has been discussed for Weibull modelling assumption. Sample based estimates of X in the form of posterior modes are shown in Table 5.3 for both the data sets. The table also provides the estimated modal values of the corresponding unknown censored observations.

Table 5.3: Estimates based on sample of size 1000 corresponding to RT and CT+RT data when the underlying modelling assumption is exponential

Variate	Estimates corresponding to	
	RT	CT+RT
α	14.703	26.371
Censored observations (in an ordered form)	37.974, 44.312, 69.438, 70.155, 76.840,85.800	51.221, 62.705, 66.262, 68.203, 68.514, 69.725, 69.771, 70.025, 73.482, 76.764, 77.194, 77.668, 85.597

Figure 5.2 presents the estimated survival curves for the two therapies using the estimated modal values given in Table 5.2. These curves are more or less similar to those shown in Figure 5.1 and provide exactly the same conclusion that was drawn using the Weibull model for the two data sets. That is, combined therapy CT+RT provides significant improvement in the survival of patients in comparison to those who are treated with RT alone. Another important finding is based on the estimates

reported for censored observations. It is to be noted that these observations correspond to the patients who failed to report during follow up and, as such, their actual survival times could not be recorded. Based on the estimated modal values of these censored observations, one can at least get an idea of actual survival times for these patients. It is to be noted that exponential model, in general, provides smaller estimated values for the highest ordered censored observations than those based on Weibull modelling assumption. This may not be a striking finding but an underestimated value is certainly a good indicative for deciding improved therapy. Moreover, we should not expect enough survival for such category of patients who left the study (or treatment) after surviving for a longer duration of time (see Tables 5.1-5.3).

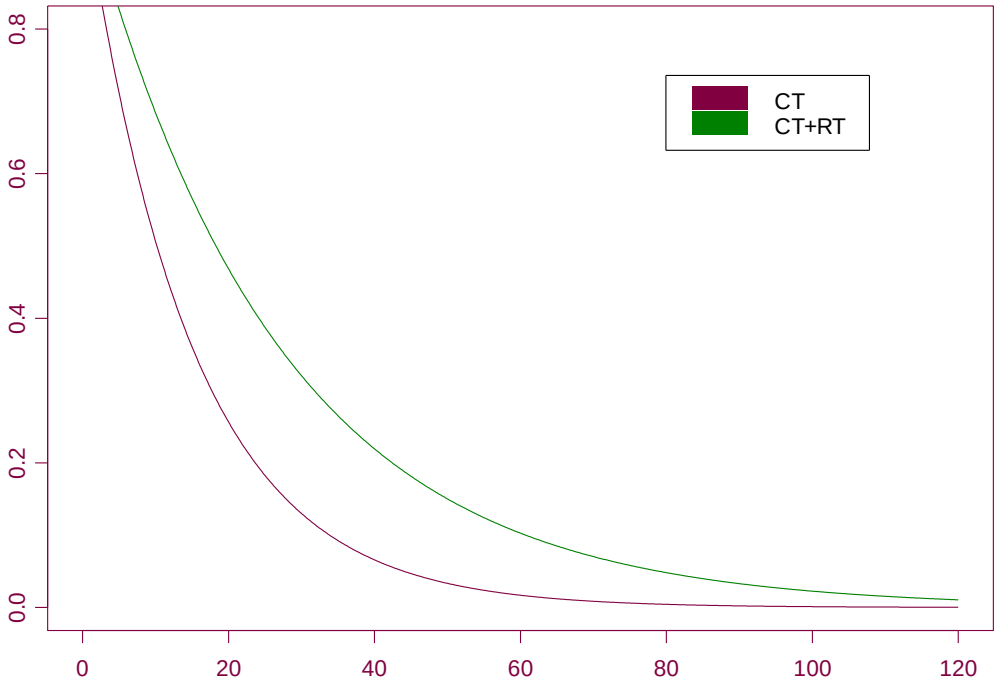

Figure 5.2: Estimated Survival functions when the underlying modelling assumption is exponential

The conclusion in favour of combined therapy CT+RT was noted by earlier authors as well who considered the two data sets and the above two modelling assumptions though their approaches were slightly different (see, for example, Li (1994) and Qian et al. (1996)).Moreover, none of the authors tried comparing the models or testing β against unity. They simply considered either exponential or Weibull model without giving any justification of the fact that why they are using these models. Besides, they took some unrealistic assumptions especially when they used Weibull

model for the reported data.

Arora et al. (2008) advocated in favour of the exponential model simply because the estimated β, when Weibull model was considered to be a true model, was found to be close to unity. This was obviously a vague criterion and, therefore, we propose to consider a comparison of the two models based on BIC. Before we begin, we shall however study compatibility of the models using posterior predictive p-value obtained by considering the Bayesian version of chi-square discrepancy measure. In order to obtain the same, we first considered the sample-based output, each of size 1000, of the unknown parameters involved in both exponential and Weibull models separately for the two data sets. We replaced the censored observations with the corresponding estimated modal values as mentioned before. Using each observation of the sample-based output in the first step; we then generated 1000 predictive samples with sizes equal to those of observed data and correspondingly obtained D2 and D1 based on predictive and observed data sets, respectively. Posterior predictive p-values based on chi-square discrepancy measure were then obtained using (5.3.1) as proportion of times D2 exceeds D1 (see also Section 5.3). The values were found to be 0.458(0.711) and 0.248(0.649), respectively, when Weibull and exponential models were considered to be the true models. The bracketed values correspond to CT+RT data.

It is obvious from the results that both the models are compatible for the two data sets and none can be rejected. It is, however, important to mention here a few things before we close our discussion. First, we are aware with the fact that model compatibility study based on posterior predictive p-values has invited a few criticisms (see, for example, Bayarri and Berger (1998)) especially the fact that it incorporates double use of data, once in simulating posteriors and second, in obtaining the p-value. This can be an important demerit but in either case it can be used as a preliminary tool as mentioned earlier (see also Upadhyay and Mukherjee (2008)). Second, once exponential model is justified for the data in hand, the Weibull model being a more complex generalization is certainly justified. Therefore, we do not need to consider the compatibility of latter but we have done simply for the sake of completeness of our study. A word of remark: the model compatibility study or the p-values should not be taken as model selection tool so we are not recommending any particular model at this stage. The parsimony principle, however, suggests that since both the models are compatible, we should go with the simple exponential model.

To complete the study for recommending a model, we evaluated BIC for the two models. These values were found to be 576.36(658.53) and 583.57(668.38), respectively, for exponential and Weibull models where the bracketed values correspond to those based on CT+RT data. Since the values corresponding to exponential model are, in general, smaller to those corresponding to Weibull model, we may safely recommend exponential model for both RT and CT+RT data although the values corresponding to the two models are not wide apart from each other. The same conclusion was drawn by parsimony principle as well, which we advocated earlier when both the models were found compatible with the data but making a conclusion after comparing the two models provides an added safety.

Conclusion :Advanced stage cancer patients are usually treated with RT or combined CT+RT. It has been a long and continuous debate among the medical practitioners that which therapy actually provides a better survival. A number of studies are performed earlier but most of these studies do not provide any convincing way for dealing with censored data although the studies have shown that CT+RT does provide better survival. The present study provides a similar conclusion based on Weibull modelling assumption, deals systematically with censored data, and successfully obtains the estimated survival times for such censored data situations.

Weibull distribution is quite flexible and perhaps because of the same reason it was used earlier by Qian et al. (1996) for the data in hand although he did not provide any convincing argument for considering this model. Our proposed study not only examines the compatibility of the Weibull model with the data but also examines the suitability of exponential model so that the resulting inferences become easy to draw if the same is recommended. It has been successfully shown after comparing the two models that unnecessary complication by assuming Weibull model can be avoided. This is what parsimony principle also suggests after getting compatibility of the two models.

Bibliography:

- *Ahmad, S.P. (2006).Optimization Tools In Bayesian Data Analysis. Ph.D. Thesis submitted to the Dept. of Statistics, University of Kashmir, Srinagar.*
- *Ahmed, A. O. M., Ibrahim, N. A., Arasan, J. and Adam, M.B. (2011). Extension of Jeffrey's Prior Estimate for Weibull Censored data using Lindley's approximation. Australian Journal of Basic and Applied Sciences, 5(12), pp 884-889.*
- *Al-Aboud, F. M. (2009). Bayesian estimations for the extreme value distribution using progressive censored data and asymmetric loss. International Mathematical Forum. vol. 4, no. 33, pp. 1603–1622.*
- *Al-Athari, F. M. (2011). Parameter estimation for the double-pareto distribution. Journal of Mathematics and Statistics, vol. 7, no. 4, pp. 289–294.*
- *Al-Kutubi, H. S. and Ibrahim, N. A. (2009) "Bayes estimator for exponential distribution with extension of Jeffery prior information. Malaysian Journal of Mathematical Sciences, vol. 3, no. 2, pp. 297–313.*
- *Arora, P. and Upadhyay, S.K.(2008). A Bayes Comparison of two Different cancer therapies under the assumption of Weibull Survival model or its Subfamily. Journal of Reliability and Statistical Studies. Vol. 1, Issue1, 6-17.*
- *Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. Proc. R. Soc. London A, 160, 268-282.*
- *Bayarri, M.J. and Berger, J.O. (1998). Quantifying surprise in the data and model verification, Bayesian Statistics 6, Bernardo, J.M., Berger, J.O., David, A.P., and Smith, A.F.M., Eds, Oxford University Press, p.53-82.*
- *Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, 2nd edition. New York: Springer-Verlag.*
- *Berrettoni, J. N. (1964) : Practical Applications of the Weibull distribution, Industrial Quality Control, 21, 71-79.*
- *Box, G.E.P., and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading.*
- *Carlin, B.P., and Louis, T.A. (2000). Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall, London*
- *Casella, G., and Berger, R. L. (2002), Statistical Inference (2nd ed.), Duxbury.*
- *Cox, D.R. (1972). Regression models and life tables, J. Roy. Statist. Soc. B. 34, p. 187-220.*

- *Cox, D.R., and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society, B, 49, 1-39.*
- *De Groot, M.H.(1970). Optimal Statistical decisions, Mc Graw Hills, New York.*
- *Fisher, R. A.(1922): On The Mathematical Foundations Of Theoretical Statistics, Phil.Trans,Roy. Soc. Series A. Vol.222.Pp-308-358.*
- *Franck, J.R. (1988). A simple explanation of the Weibull distribution and its applications, Reliability Review, 8, No. 3, 93-116.*
- *Gamerman, D. and Lopes, H.F. (2006). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd ed., Chapman & Hall/CRC.*
- *Gelman, A. (2004). Parameterization and Bayesian modeling, J. Amer. Statist. Assoc., 99, p.537-545.*
- *Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). Bayesian Data Analysis. Chapman and Hall, London.*
- *Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2003). Bayesian Data Analysis. 2nd ed., Chapman & Hall, London.*
- *Gelman, A., Meng, X.L. and Stern, H.S. (1996). Posterior predictive assessment of model fitness via realized discrepancies, Statistical Sinicia, 6, 733-803.*
- *Ghitany, M. E., Al-Hussaini, E. K. & Al-Jarallah, R. A. (2005). Marshall-Olkin extended weibull distribution and its application to censored data, Journal of Applied Statistics, 32, 1025 – 1034.*
- *Gianola, D., and Fernando, R.L. (1986). Bayesian methods in animal breeding theory. J. Animal. Sci., 63, 217-244.*
- *Haldane, J. B. S.(1931): A Note On Inverse Probability Proceedings Of Cambridge Philosophical Society 28,55- 61.*
- *Hallinan, A.J. (1993). A review of the Weibull distribution, Journal of Quality Technology, 25, 282-284.*
- *Hartigan, J.A. (1964). The Asymptotic Unbiased Prior Distribution. Annals of Mathematical Statistics, 36, 1137-1152.*
- *Hirose, H. & Lai, T. L. (1997). Inference from grouped data in three-parameter Weibull models with applications to breakdown-voltage experiments, Technometrics, 39, 199-210.*
- *Hossain, A. and Zimmer, W. (2003).“Comparison of estimation methods for Weibull parameters: complete and censored samples,” Journal of Statistical Computation and Simulation, vol. 73, no. 2, pp. 145–153.*

- *Ibrahim, J.G., Ming-Hui Chen and Sinha, D. (2001). Bayesian Survival Analysis. SpringerVerlag, New York.*
- *Ibrahim,J.G., Ming-HuiChen and Sinha,D.(2004). Bayesian methods for joint modelling of longitudinal and survival*
- *Jefferys, H. (1946). An invariant form for the prior probability in estimation problems; proceedings of the royal Society of London, Series. A. 186, 453-461.*
- *Jefferys, R. C. (1983). The logic of Decision, II Edition. University of Chicago press.*
- *Jeffreys, H. (1961). Theory of Probability. Oxford Press.*
- *Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2. John Wiley & Sons, New York.*
- *Kalbfleisch, J.D. and Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data. Wiley, NewYork.*
- *Kao, J.H.K. (1959). A graphical estimation of mixed Weibull parameters in life testing of electron tubes. Technometrics 1, 389-407.*
- *Kaplan, E.L., and Meier, P. (1958).Non parametric estimation from incomplete observations, J. Amer. Statist. Assoc.,53, p. 457-81.*
- *Khan, A.A. (1997). Asymptotic Bayesian Analysis in Location-Scale Models. Ph.D. Thesis submitted to the Dept. of Mathematics & Statistics, HAU, Hisar.*
- *Lai, C.D., Xie, M., and Murthy, D.N.P. (2003). Modified Weibull model, IEEE Transactions on Reliability, 52, 33–37.*
- *Lawless, J.F. (2002). Statistical Models and Methods for Lifetime Data. 2 ed., Wiley, New York.*
- *Lee, E.T. and Wang, J.W. (2003). Statistical Mehods for Survival Data Analysis, Wiley,New York.*
- *Lee, P.M. (1997). Bayesian Statistics: An Introduction, 2nd ed. Arnold, London.*
- *Legendre, A.(1805): Nouelles Methods Pour La Determination Des Orbites Dis Cometes, Coucier, Pairs.*
- *Li, C. (1994). Comparing Survival Data for Two Therapies: Nonhierarchical and Hierarchical Bayesian Approaches, Unpublished Ph.D. thesis, Department of Statistics, Duke University, USA.*
- *Lieblein, J., and Zelen, M. (1956). Statistical investigation of the fatigue life of deep groove ball bearing. J. Res. Nat. Bur. Stand. 57, 273-316.*

- *Lye, L. M., Hapuarachchi, K. P. and Ryan, S. (1993). "Bayes estimation of the extreme-value reliability," IEEE Transactions on Reliability, vol. 42, no. 4, pp. 641–644.*
- *Malik, M.A.K. (1975). A note on the physical meaning of the Weibull distribution, IEEE Transactions on Reliability, 24, 95.*
- *Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising ill its consideration, Cancer Chemotherapy Reports, 50, p.163-70.*
- *Marshall, A. W. & Olkin, I. (1997). A new method for adding a parameter to a family of distribution with application to the Exponential and Weibull families, Biometrika, 84, 641-652.*
- *Martz, H.F. and Waller, R.A. (1982). Bayesian Reliability Analysis. Wiley, New York.*
- *Mudholkar, G. S., Srivastava, D. K. & Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data, Journal of the American Statistical Association, 91, 1575-1583.*
- *Murthy, D. P., Xie, M., and Jiang, R. (2004), Weibull Models, John Wiley and Sons, Inc. ReliaSoft Corportation (1996{2006), \Estimation of Weibull Parameters,"*
- *Nadarajah, S. & Kotz, S. (2005). On some recent modifications of Weibull distribution, IEEE Transactions on Reliability, 54, 561-562.*
- *Nadarajah, S. & Kotz, S. (2007). On the Moment of the Maxima of Weibull Random Variables, IEEE Transactions on Vehicular Technology, 56, 1467-1468.*
- *Nordman, D.J. and Meeker, W.Q. (2002). Weibull prediction intervals for a future number of failures, Technometrics, 44, 15–23.*
- *Pandey, B. N., Dwividi, N. and Pulastya, B. (2011). "Comparison between Bayesian and maximum likelihood estimation of the scale parameter in Weibull distribution with known shape under linex loss function," Journal of Scientific Research, vol. 55, pp. 163–172.*
- *Peto, R., and Peto, J. (1972). Asymptotically efficient rank invariant procedures (with discussion). Journal of the Royal Statistical Society A 135, 185-206.*
- *Pike, M.C. (1966). A method of analysis of certain classes of experiments in carcinogenesis. Biometrics 22, 142-161.*
- *Qian, J. (1994). A Bayesian Weibull Survival Model, Unpublished Ph.D. thesis, Department of Statistics, Duke University, USA.*

- *Qian, J., Stangl, D., and George, S. (1996). A Weibull model for survival data: using predictions to decide when to stop a clinical trial. Bayesian Biostatistics, Berry, D. and Stangl, D., Eds., Marcel Dekker, p.187-205.*
- *Raffia, H., and Schlaifer, R. (1961). Applied Statistical Decision Theory. Boston, Massachusetts: Harvard Business School.*
- *Royall, R. M. (1997). Statistical evidence – A likelihood Paradigm, Chapman & Hall.*
- *Schwartz, G. (1978). Estimating the dimension of a model. Ann. Statist., 6, p.461-68.*
- *Singpurwalla, N.D. (2006). Reliability and Risk: A Bayesian Perspective. Wiley, New York.*
- *Sinha, D. and Dey, D.K. (1997). Semiparametric Bayesian Analysis of Survival Data, J.Amer. Statist. Assoc., 92, p. 1195-1212.*
- *Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo Methods, J. Roy. Statist. Soc. B, 55, p. 2-23.*
- *Soland, R.M. (1968). Bayesian analysis of the Weibull process with unknown scale parameter and its application to acceptance sampling. IEEE Transactions on Reliability 17, 84-90.*
- *Stigler, S.M. (1983). Who discovered Baye's theorem? American Statistics 37: 290-296.*
- *Tang, Y., Xie, M. & Goh, T. N. (2003). Statistical analysis of a Weibull extension model, Communications in Statistics – Theory and Methods, 32, 913-928.*
- *Upadhyay, S.K. and Mukherjee, B. (2008). Assessing the value of the threshold parameters in the Weibull distribution using Bayes paradigm, IEEE Trans. Reliab., 57. p. 489-497.*
- *Upadhyay, S.K. and Peshwani, M. (2008). Posterior analysis of lognormal regression models using the Gibbs sampler, Statistical Papers, 49, p. 59-85.*
- *Upadhyay, S.K., Vasistha, N. and Smith, A.F.M. (2001). Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation, Sankhya A, 63, p. 15-40.*
- *Weibull, W. (1951). A statistical distribution function of wide applicability. Appl. Mech., 18, 293-297.*

- *Whittemore, A., and Altschuler, B. (1976). Lung cancer incidence in cigarette smokers: Further analysis of Doll and Hill's data for British physicians, Biometrics, 32, 805-816.*
- *Xie, M., Goh, T.N. and Tang, Y. (2002): A modified Weibull extension with bathtub–shaped failure rate function; Reliability Engineering and Systems Safety 76, 279–285*
- *Zellener, A. (1971). An Introduction to Bayesian Inference in Econometrics. Wiley, New York.*
- *Zellner, A.(1977): Maximal Data Information Prior Distribution: In: New Developments In The Application Of Bayesian Methods , A.Aykac & C, Brumat (Eds.), North Holland Publishing Co.,201-15.*