Summer 2015

Bayesian Inference of the Weibull-Pareto Distribution

James Dow

BAYESIAN INFERENCE OF THE WEIBULL-PARETO DISTRIBUTION

by

JAMES DOW

(Under the Direction of Arpita Chatterjee)

ABSTRACT

The Weibull distribution can be used to model data from many different subject areas such as survival analysis, reliability engineering, general insurance, electrical engineering, and industrial engineering. The Weibull distribution has been further extended by the Weibull-Pareto distribution. A desirable property of the Weibull-Pareto distribution is its ability to model skewed data. This is especially useful for developing models in human longevity and actuarial science. In this work a hierarchical Bayesian model was developed using the Weibull-Pareto distribution.

Key Words: Heavy-tailed Distribution, Weibull-Pareto, hierarchical Bayesian model, right censoring, survival model, MCMC

2009 *Mathematics Subject Classification*:

BAYESIAN INFERENCE OF THE WEIBULL-PARETO DISTRIBUTION

by

JAMES DOW

B.S. in Mathematical Sciences

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment

of the Requirement for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

2015

BAYESIAN INFERENCE OF THE WEIBULL-PARETO DISTRIBUTION

by

JAMES DOW

Major Professor: Arpita Chatterjee

Committee: Charles Champ

Stephen Carden

Electronic Version Approved:

July 2015

ACKNOWLEDGMENTS

I wish to acknowledge my committee members, Dr. Carden and Dr. Champ. I appreciate the time they took to help me with this thesis. I also would like to acknowledge my adviser, Dr. Chatterjee. It was with her guidance I was able to complete this document and learn an enormous amount about various topics of statistics. I also want to thank Dr. Oluyede who helped me gain valuable statistical consulting experience.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Weibull distribution is one of the most widely used distributions in reliability and survival analysis. The Weibull distribution is a versatile and can take on the characteristics of many different types of distributions, based on the value of its shape parameter. This has made it extremely popular among engineers and quality control practitioners, who have made it the most commonly used distribution for modeling reliability data. Various extensions of the Weibull distribution has been developed by many researchers, including the generalized Weibull distribution by Mudholkar and Kollia (1994)[8], the exponentiated-Weibull distribution by Mudholkar et al. (1995)[9], and the beta-Weibull distribution by Famoye et al. (2005)[3]. In recent past Alzaatreh et al. (2013)[2] introduced a new generalization of Weibull distribution based on a family of distributions called “Transformed-Transformer” family (T-X family), named Weibull-Pareto distribution. This distribution has higher skewness as compared to a Weibull distribution and therefore is more suitable to model a heavily skewed data often arise in reliability and survival analysis.

Alzaatreh et al. (2013)[1] mentioned that when the shape parameter of WPD is less than 1, the MLE for both the shape and scale parameters does not exist. Hence they introduced two alternative methods of parameter estimation, namely alternative maximum likelihood estimation (AMLE) and modified maximum likelihood estimation (MMLE). However, the AMLE procedure often results in large bias and the implementation of MMLE is computationally expensive. These methods are briefly discussed in Section 2.1.1. In this research we introduced a Bayesian counterpart of Weibull-Pareto model. Unlike frequentist approach, Bayesian approach assumes the model parameters to be randomly distributed following some probability distribution. This distribution is commonly referred as prior distribution and intend to capture the researcher's prior belief about the parameter, before collecting these relevant data. As we gather data we would like to update the prior belief in the face of reality, through the posterior distribution.

In recent years there have been several papers written on Bayesian estimation of model parameters. Some of these references are: Martin and Perez (2009)[7] which estimated the generalized lognormal distribution, Singh et al (2013)[14] estimated the inverse Weibull distribution, and Noor and Aslam(2013)[10] studied the inverse Weibull mixture distribution using Type-I censoring.

Weibull distribution and its counterparts have been used very effectively for analyzing lifetime data, particularly when the data are censored, which is very common in survival. For example, in many cancer studies, the main outcome is the time to an event of interest, which is commonly known as survival time. However, it can also be referred as to the time from complete remission to relapse or progression as equally as to the time from diagnosis to death. Moreover, it is usual that at the end of follow-up some of the individuals have not had the event of interest, and thus their true time to event is unknown. Hence these are considered as censored obervations. When we censor an observation, we truncate the observation since the event has not been observed due to random factors. The failure of an event to occur given the time interval is an example of this truncation. In Section 4.2 we implemented our proposed model to a cutaneous melanoma survival data.

The rest of this thesis is sectioned as follows. In Section 2 we will describe our Bayesian model and discuss the details of the Weibull-Pareto distribution. Our simulation study will be covered in Section 3. The application of the Bayesian model and survival model will be presented in Section 4 and we will conclude the paper in Section 5.

CHAPTER 2

BAYESIAN MODEL PARAMETER ESTIMATES

2.1 Weibull Pareto Distribution

Overall lifetimes are an example of positively skewed data that occur in survival and reliability analysis. For example, breast cancer data as studied by Khan et al. (2014)[5] is right skewed. The Weibull and lognormal distributions are commonly used to model this type of data. Depending on the choices of parameters these distributions can model right skewed data; however, when the data is heavily skewed these models fail to capture the heavy right tail as they drop-off at a much faster rate.

The Weibull-Pareto distribution was developed by Alzaatreh et al. (2013)[1]. It is derived from a family of distributions called “Transformed-Transformers” family (T-X family), defined as follows:

Definition 2.1. *(T-X family) Let F(x) be the cumulative distribution function (cdf) of any random variable X and r(t) be the probability density function (pdf) of a random variable T defined on* $[0, \infty]$*. The cdf of the generalized family is given by*

$$G(x) = \int_0^{-log(1-F(x))} r(t)dt.$$

If a random variable T follows Weibull distribution with parameters c and γ, $r(t) = (c/\gamma)(t/\gamma)^{c-1}e^{(t/\gamma)^c}, t \geq 0$ then we have Weibull-X family with pdf

$$g(x) = \frac{c}{\gamma}\frac{f(x)}{1-F(x)}\left\{\frac{-\log(1-F(x))}{\gamma}\right\}^{c-1}\exp\left\{-\left(\frac{-\log(1-F(x))}{\gamma}\right)^c\right\}.$$

If we let X be Pareto random variable with distribution with pdf $f(x) = k\theta/x^{k+1}, x > \theta$ we have

$$g(x) = \frac{kc}{\gamma x}\left\{\frac{k}{\gamma}\log\left(\frac{x}{\theta}\right)\right\}^{c-1}\exp\left\{-\left(\frac{k}{\gamma}\log\left(\frac{x}{\theta}\right)\right)^c\right\},\ x > \theta.$$

Letting $\beta = k/\gamma$, we have the pdf of the Weibull-Pareto distribution with parameters c, β, θ,

$$g(x) = \frac{\beta c}{x}\left\{\beta \log\left(\frac{x}{\theta}\right)\right\}^{c-1} \exp\left\{-\left(\beta \log\left(\frac{x}{\theta}\right)\right)^{c}\right\}, x > \theta, c, \beta, \theta > 0. \tag{2.1}$$

The Weibull-Pareto distribution is denoted by WPD(c, β, θ).

The cdf is given by

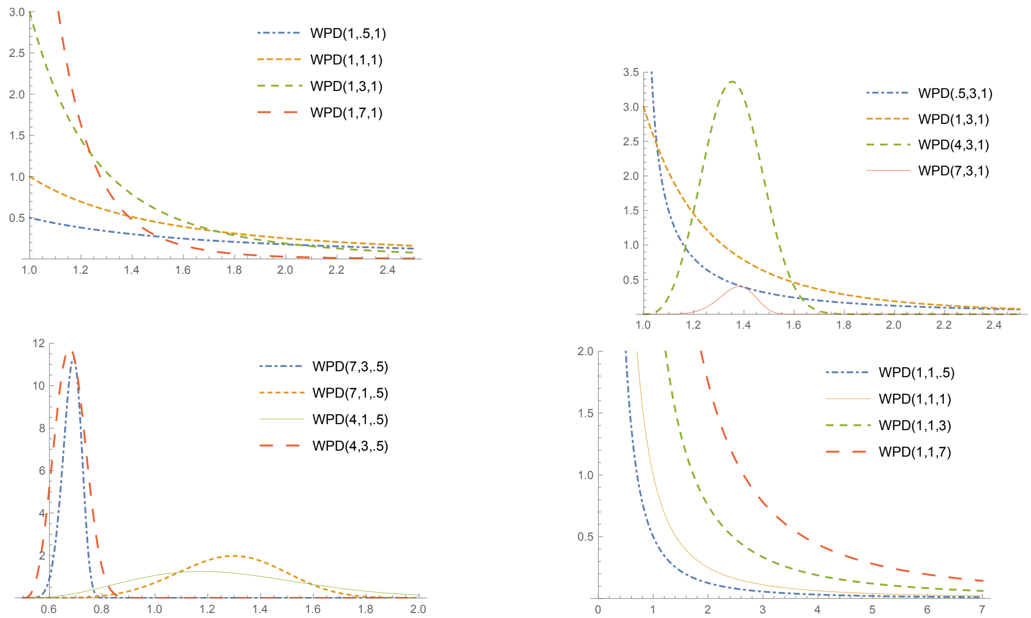
$$G(x) = 1 - exp\left\{-(\beta\ log(x/\theta))^{c}\right\}. \tag{2.2}$$

We found the following lemma to be very useful in Section 3.2. Lemma 2.2 shows the relation between WPD and Weibull distribution. The transformation can be implemented in generating random samples from WPD. Therefore, we can start with generating random observations from Weibull distribution and apply this transformation to get realizations from Weibull-Pareto distribution.

Lemma 2.2. *If a random variable Y follows the Weibull distribution with parameters c and $1/\beta$, the random variable $X = \theta e^{Y}$ follows WPD(c, β, θ).*

Figure 2.1 shows that the WPD is capable to cover a wide range of probability distributions depending on its parameter choices. The graph in the upper left of Figure 2.1 how the shape of the distribution changes if we let β varies while keeping c and θ fixed. We see that as β increases, the graph becomes more steep. The graph in the upper right shows β and θ are fixed and c is changing . We can see that the shape becomes more symmetric as c increases. Note that, in this research we are mainly interested in WPD with a small c, as it results in heavy right skewed data. The graph in lower left θ is fixed and c and β are changing. The graph in lower left reveals that as c increases $(c > 1)$ the WPD looks more and more symmetric. However, as the rate (β) decreases it becomes more and more flat. Similarly the graph on the lower right corner depicts that changing the value of θ shifts the probability distribution to the right.

Figure 2.1: PDF of WPD



2.1.1 Estimation

Our goal with any probability distribution is to estimate the model parameters. One obvious choice is the method of maximum likelihood estimation (MLE). However, Alzaatreh et al (2013)[1]. explains there are two issues we must consider in using this method. One is when $c < 1$ the likelihood function increases to infinity as θ approaches to the minimum observed value, which we will denote as $x_{(1)}$. When $c < 1$ and θ is estimated by $x_{(1)}$, the MLE does not exist for β and c. Smith (1985) [15] studied this issue and introduced alternative method called the alternative maximum likelihood estimation (AMLE), that assumes θ to be $x_{(1)}$. Even though this method reduces computational cost, it doesn't work well when $c >> 1$. Figure 2.1 shows that when $c < 1$, $x_{(1)}$ can give an accurate estimate for θ as pdf maximized at the minimum data value. However, as c increases, the WPD appears more symmetric and as a result $x_{(1)}$ fails to capture the estimate through AMLE. The modified maximum likelihood estimate proposed by Smith (1985) [15] is more applicable and will be introduced later in this section. Alzaatreh et. al. (2013)[1] implement AMLE to estimate the parameters of WPD. He considered probability densities of the form

$$f(x;\theta,\phi) = (x-\theta)^{c-1} q(x-\theta;\phi), \text{ for } x > \theta, \tag{2.3}$$

where θ and the parameter vector ϕ are unknown. For AMLE, $\phi = (\beta, c)$ and he proposed to let θ equal $x_{(1)}$ and them remove the sample minimum from the observed data. The AMLE function for WPD is given by

$$L_* = \sum_{x \neq x_{(1)}} \log\, g(x_i; c, \beta, x_{(1)})$$

$$= \sum_{x \neq x_{(1)}} \left\{ c \log \beta + \log c - \log x_i + (c-1) \log(\log(x_i/x_{(1)})) - (\beta \log(x_i/x_{(1)}))^c \right\}.$$

Taking the derivative with respect to β and c we have

$$\frac{\partial L_*}{\partial \beta} = \sum_{x_i \neq x_{(1)}} \left\{ c/\beta - c\beta^{c-1} \big(\log(x_i/x_{(1)})\big)^c \right\}$$

$$\frac{\partial L_*}{\partial c} = \sum_{x_i \neq x_{(1)}} \Big\{ c + \log \beta + \log(\log(x_i/x_{(1)}) - \log \beta (\beta \log(x_i/x_{(1)}))^c$$
$$-(\beta \log(x_i/x_{(1)}))^c \log(\log((x_i/x_{(1)})) \Big\}$$

Setting the previous equations to zero we have

$$\beta = \left\{ (n-n')/ \sum_{x_1 \neq x_{(1)}} (\log(x_i/x_{(1)}))^c \right\}^{1/c} \tag{2.4}$$

$$c^{-1} + \sum_{x_1 \neq x_{(1)}} \log\left(\log\left(\frac{x_i}{x_{(1)}}\right)\right) - \frac{\sum_{x_i \neq x_{(1)}} (\log(x_i/x_{(1)}))^c \log(\log(x_i/x_{(1)}))}{\sum_{x_i \neq x_{(1)}} (\log(x_i/x_{(1)}))^c} = 0. \tag{2.5}$$

n' in Equation (2.4) is the count of the occurrence minimum. AMLE is a consistent estimator, but it is unclear how it performs with relatively small samples. Alzaatreh et al.(2013)[1] conducted a simulation with 36 parameter combinations and two sample sizes. They let β and θ equal .5 , 1, and 3 and for c they chose .5, 1, 3, and 7. The two sample sizes were n=100 and n=500. For each parameter combination they used the previously discussed transformation; generated a random sample $y_1, y_2, ..., y_n$ from Weibull with parameters c and $1/\beta$. Then, $x_i = \theta exp(y_i)$, which is a random sample from Weibull-Pareto. The results of the simulation showed that the AMLE had large bias when $c > 1$. Also if $\hat{\theta} = x_{(1)}$ is greater than θ by a small amount, it gives large bias for c. The bias is explained by the fact that $\hat{c}$ is a solution for (2.5) . The term $\log[\log(x_i/\hat{\theta})]$ causes the large bias. An example described by Alzaatreh et al.(2013)[1] is as follows; suppose the true parameter $\theta = 1$, the estimated parameter $\hat{\theta} = 1.3$, and the observation $x_i = 1.3001$. Then $\log[\log((x_i/\hat{\theta})] = -9.4727$ while the actual value $\log[\log(x_i/\theta)] = -1.3377$. They note from their simulation that as c increases $\hat{\theta} = x_{(1)}$ overestimates θ.

The other issue is when $c >> 1$. When this occurs $x_{(1)}$ is a poor estimate for θ and AMLE produces large bias. Smith also developed modified maximum likelihood estimation (MMLE) for densities with form of Equation (2.3) , which can handle this second issue. Consider the following log-likelihood function

$$L_n(c, \beta, \theta) = \sum_{i=1}^{n} \log g(x_i; c, \beta, \theta),$$

which is defined for $\theta < x_{(1)}$. As with the traditional MLE method we take the derivatives with respect to each parameter. We need to show $\dfrac{\partial L_n(c, \beta, \theta)}{\partial \theta}$ exists whenever $\theta < x_{(1)}$. We have

$$\frac{\partial L_n(c, \beta, \theta)}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^{n} \frac{1-c}{\log(x_i/\theta)} + \frac{c\beta^c}{\theta} \sum_{i=1}^{n} (\log(x_i/\theta))^{c-1}.$$

Since the right hand side of the equation is continuous on the interval $0 < \theta < x_{(1)}$, $\dfrac{\partial L_n(c, \beta, \theta)}{\partial \theta}$ exists. Setting $\frac{\partial L_n(c,\beta,\theta)}{\partial \theta}$,$\frac{\partial L_n(c,\beta,\theta)}{\partial \theta}$, and $\frac{\partial L_n(c,\beta,\theta)}{\partial \theta}$ equal to zero we have,

$$\frac{1}{\theta}\sum_{i=1}^{n}\frac{1-c}{\log(x_i/\theta)}+\frac{c\beta^c}{\theta}\sum_{i=1}^{n}(\log(x_i/\theta))^{c-1}, \tag{2.6}$$

$$\beta=\left\{n\middle/\sum_{i=1}^{n}(\log(x_i/\theta))^c\right\}, \tag{2.7}$$

$$c^{-1}+\sum_{i=1}^{n}\log(\log(x_i/\theta)-\frac{\sum_{i=1}^{n}(\log(x_i/\theta))^c\log(\log(x_i/\theta))}{\sum_{i=1}^{n}(\log(x_i/\theta))^c}=0. \tag{2.8}$$

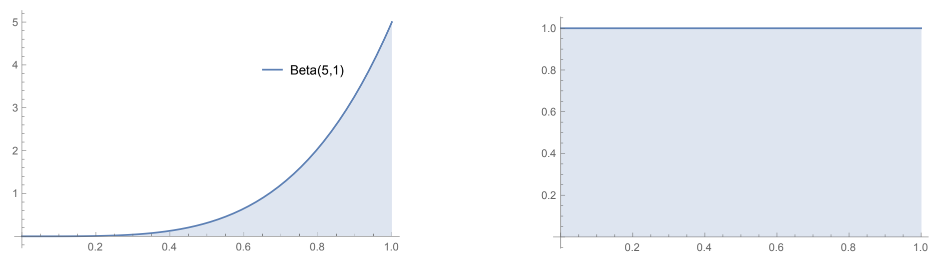
In order to solve Equation (2.6), (2.7), and (2.8) we need to apply numerical methods. Issues that may arise are slow convergence or our initial guess may cause the method to not converge. This makes parameter estimation difficult. We propose an alternative approach based on a hierarchical Bayesian model.

2.2 Introduction to Bayesian Statistics

The estimation procedure described in Section 2.1.1 is based on frequentist methodology, which assumes the parameter is fixed but unknown. However, the Bayesian approach considers the parameters to be random. The random behavior can be modeled through a known probability distribution, commonly known as a prior distribution.

In order to illustrate Bayesian methods, consider the following example. Suppose we are giving medication with a high success rate. The data given, which is binary, would be whether or not the medicine was effective. We can assume the data is coming from Bernoulli with probability of success p, which we know to be high. This prior information can be incorporated through a probability model, defined on the interval (0,1). One obvious choice is the Beta distribution, Beta(α_p, β_p). By changing α_p and β_p, we can cover a wide range of shapes. For example, Beta(5,1) can be used to model a negatively skewed density, an appropriate choice for our case. A distribution that models a parameter is known as a prior distribution as it models the researchers prior belief about model parameters. Another choice of prior distribution can be uniform distribution defined on the interval $(0, 1)$. Under uniform prior the parameter p can take on any value with equal chance with its range (0,1). However the Beta(5,1) is more informative in order to structure of our prior belief. Therefore, Beta(5,1) is known as informative prior, where as our Uniform(0,1) is commonly known as non-informative. Non-informative prior is a prior which is used when we have no assumptions about the parameter and is usually flat. An informative prior is a prior which reflects our assumption about the parameter, such as with the Beta distribution in our example. Figure 2.2 shows the density for prior choices.

Figure 2.2: Informative (left) and Non-informative (right) prior choices



As discussed in the earlier paragraph, the prior distribution reflects our prior belief about the parameter before observing the data. However, after collecting relevant data we should update our prior belief in the face of reality through a posterior distribution.

The model described in the above example can be written hierarchically as follows:

Likelihood:

$$Y_i \sim Bernoulli(n, p),\ \ i = 1, ..., n$$

Prior:

$$p \sim Beta(\alpha_p, \beta_p),$$

where α_p, β_p are known as hyperparameters and can be chosen appropriately to capture the prior information. Then we have

$$\Pi(p|data) = \frac{f(data|p)g(p|\alpha_p, \beta_p)}{h(data|\alpha_p, \beta_p)}, \tag{2.9}$$

where $f(data|p)$ and $g(p|\alpha_p, \beta_p)$ are the respective likelihood and prior distribution. By $data$ we denote the observation vector $\{Y_1, Y_2, ..., Y_N\}$. Moreover, $h(data|\alpha_p, \beta_p)$ is commonly known as the data marginal which can be obtained by,

$$h(data|\alpha_p, \beta_p) = \int_0^1 f(data|p)g(p|\alpha_p, \beta_p)\ dp. \tag{2.10}$$

For notational simplicity we will denote the data marginal $h(data|\alpha_p, \beta_p)$ as $h(data)$. In practice, for more complex models $h(data)$ is difficult to obtain, as it involves integrals on all model parameters. Therefore, the posterior distribution $\Pi(p|data)$ can be obtained through $\Pi(p|data) \propto f(data|p)g(p|\alpha_p, \beta_p)$ up to a normalizing constant. In words, this says the posterior is proportional to the likelihood and the prior. Using this form of Bayes', we can construct a hierarchical representation starting with the likelihood function and then gradually adding the priors to the model.

In general the Bayesian Theorem for probability distribution is given by:

$$\Pi(\eta|data) = \frac{f(data|\eta)g(\eta)}{h(data)}$$

where the $\Pi(\eta|data)$ is the posterior for the parameter η. The term $f(data|\eta)$ is the likelihood, $g(\eta)$ is the prior distribution, and as described earlier $h(data)$ is the data marginal. Finally, likelihood is given by

$$\prod_{i=1}^{n} f(x_i|\eta) = L(\eta|x_i), \tag{2.11}$$

The marginal probability acts as a normalizing constant. When the parameter space is continuous, $h(data) = \int_\eta f(data|\eta)g(\eta)\ d\eta$.

Chapter 4 of Lynch[6] lists the following steps for general Bayesian inference:

1. Establish the model and obtain a posterior distribution for the parameters.

2. Generate a sample from the posterior distributions.

3. Use discrete formulas applied to the samples from the posterior distribution to summarize knowledge of the parameters.

Ideally we want our priors to be conjugate priors, which results in a posterior distribution that falls in the same family of distribution as the prior distribution. We have not found a conjugate prior for the WPD. Since our model is complex, which we will show later in Section 2.2.1, we will need to employ Markov Chain Monte Carlo (MCMC) sampling. MCMC sampling allows us to handle multivariate densities by sampling through multiple dimensions of the posterior distribution and move through the entire support of the posterior distribution. More specifically a Gibbs sampler is the most basic method. It is described by Lynch[6] in the following steps:

0. Assign initial values, S, to the parameter vector $\eta^{j=0} = \{\eta_1, \eta_2, ..., \eta_k\} = S$. j indexes the iteration count.

1.Set j=j+1

2.Sample$(\eta_1^j|\eta_2^{j-1}, \eta_3^{j-1}, ..., \eta_k^{j-1})$

3.Sample$(\eta_2^j|\eta_1^j, \eta_2^{j-1}, ..., \eta_k^{j-1})$

$\vdots$

k.Sample$(\eta_k^j|\eta_1^j, \eta_2^j, ..., \eta_{k-1}^j)$

k+1. Return to **1**.

2.2.1 Bayesian WPD

Our choice of the likelihood for the Bayesian model is WPD. As in Equation (2.9) the likelihood depends on three parameters. Our choice of η is the parameter vector

(c, β, θ). The priors we selected for our model are exponential and gamma for the parameters.

Likelihood:

$$X|c, \beta, \theta \sim WPD(c, \beta, \theta)$$

Prior:

$$c|\alpha_c, \gamma_c \sim \Gamma(\alpha_c, \gamma_c)$$

$$\beta|\alpha_\beta, \gamma_\beta \sim \Gamma(\alpha_\beta, \gamma_\beta)$$

In the Bayesian version of AMLE we don't assume that θ has a prior distribution. Instead, we set $\theta = x_{(1)}$ where $x_{(1)}$ is the minimum of the data as discussed in Section 2.1. In this case there are two parameters two priors for our Gibbs sampler to estimate. We also consider our version of the MMLE where the prior on θ is specified and truncated at $x_{(1)}$,

$$\theta|\alpha_\theta, \gamma_\theta \sim \Gamma(\alpha_\theta, \gamma_\theta) I_{(0,x_1)}(\theta)$$

Prior selection for θ, β, and c could be any distribution defined on the $\mathbb{R}_+$.

If we were directly estimate the parameter c for example, we have:

$$\begin{aligned}
\Pi(c|x) &= \text{Ł}(WPD) * \Gamma(\alpha_c, \gamma_c) * \Gamma(\alpha_\beta, \gamma_\beta) * \Gamma(\alpha_\theta, \gamma_\theta) \\
= \frac{1}{\Gamma(\alpha_c)\gamma^{\alpha_c}} * c^{\alpha_c - 1} * e^{-c/\gamma_c} &\int_\beta \int_\theta \frac{\beta c}{x} \Bigg\{ \beta \log\Big(\frac{x}{\theta}\Big) \Bigg\}^{c-1} \exp\Bigg\{ - \Bigg(\beta \log\Big(\frac{x}{\theta}\Big) \Bigg)^c \Bigg\} \\
* \frac{1}{\Gamma(\alpha_\beta)\gamma^{\alpha_\beta}} * \beta^{\alpha_\beta - 1} * e^{-\beta/\gamma_\beta} &\frac{1}{\Gamma(\alpha_\theta)\gamma^{\alpha_\theta}} * \theta^{\alpha_\theta - 1} * e^{-\theta/\gamma_\theta} * I_{(0,x_{(1)})}(\theta) \; d\beta d\theta.
\end{aligned} \tag{2.12}$$

We would evaluate β and θ similarly. We can see Equation (2.12) is difficult to evaluate. For this reason we elected to use Gibbs sampler as described in Section 2.2.

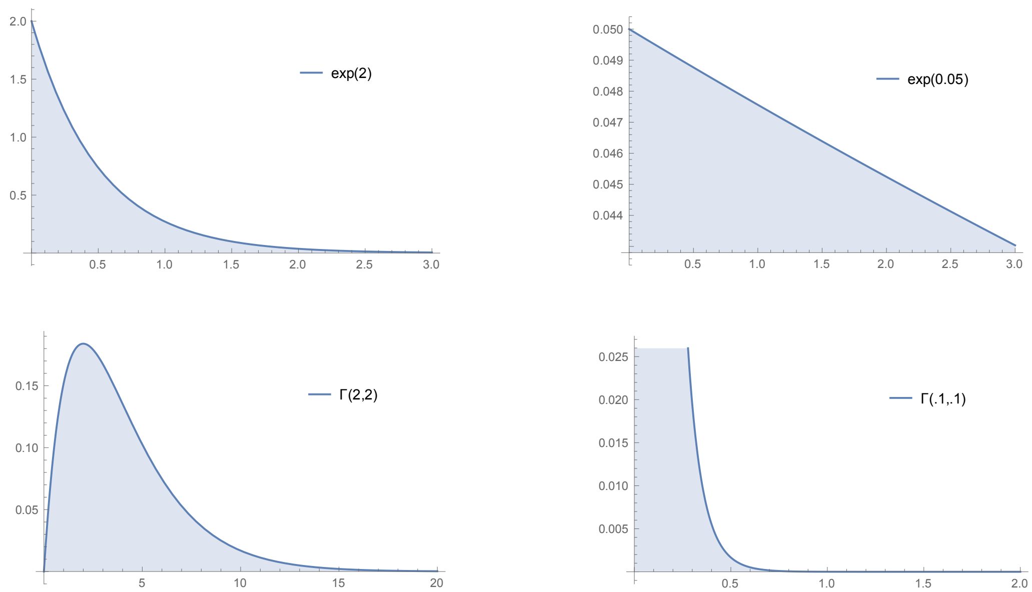
We used melanoma data Section 4.2 to test how well gamma and exponential prior distributions perform. We used both informative prior and non-informative priors. We ran the Gibbs Sampler using 100,000 iterations, we removed the first

50,000 of those 100,000, and then picked every 10^{th} iteration to form our estimates. Figure 2.3 displays the graphs of the priors used.

Table 2.1: Prior Choices

	Exponential Informative			Exponential Non-Informative			Gamma Informative			Gamma Non-Informative		
	mean	sd	MC Error	mean	sd	MC Error	mean	sd	MC Error	mean	sd	MC Error
$\hat{c}$	9.445	1.328	0.1038	12.91	1.135	0.1245	8.648	0.955	0.06645	17.97	1.501	0.1732
$\hat{\beta}$	0.1861	0.02227	0.001563	0.1411	0.01316	0.001301	0.2009	0.01918	0.001238	0.1031	0.009068	9.928E-4
$\hat{\theta}$	0.01792	0.01062	6.693E-4	0.003398	0.002748	2.494E-4	0.02519	0.0116	7.28E-4	2.951E-4	3.824E-4	6.492E-4

Figure 2.3: Prior Choices Graph

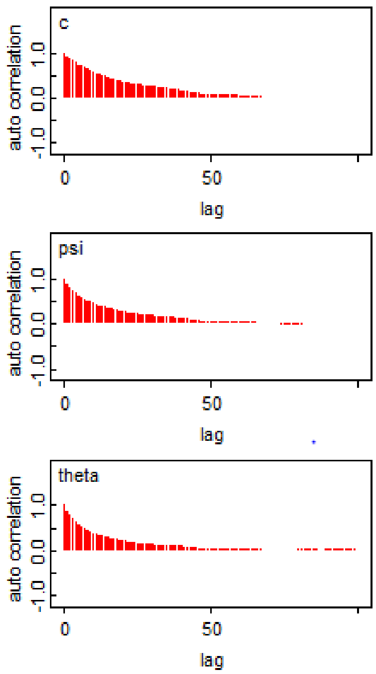


From Table 2.1 we can see both informative priors have similar means, standard deviations, and Monte Carlo error(MC error). We would desire for the MC error to be small. The non-informative priors performed the same. For the parameter $\hat{c}$ the exponential non-informative prior had less MC error than the gamma non-informative prior. Gamma non-informative had less MC error for $\hat{\theta}$.

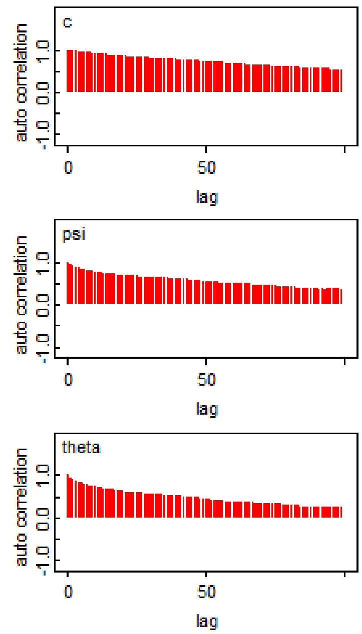
Figure 2.4 and 2.5 give the autocorrelation plots and the history of the prior choices.

To check the convergence of the Markov chain we can look at the history plot and the autocorrelation plot of the parameters. These are intuitive ways to check the convergence and can be easily implemented in OpenBUGS. In a history plot, we plot the parameter value as a function of sample number. If the model has converged, the plot will move up and down around the mode of the posterior distribution. A clear sign of non-convergence occurs when we observe some trending in the history plot. On the other hand, an autocorrelation plot will reveal whether or not there is correlation between successive samples. The presence of correlation indicates that the samples are not effective in moving through the entire posterior distribution, and may need large iteration or reconstruction of the Bayesian model. As described above, we take every 10^{th} iteration in order to reduce autocorrelation. Figure 2.4 and Figure 2.5 shows the non-informative priors of gamma and exponential have high autocorrelation and the history plots show a trend. Therefore the non-informative priors did not converge. The informative priors in these figures show little autocorrelation and the history plot appears random. Which suggests the informative priors converged.

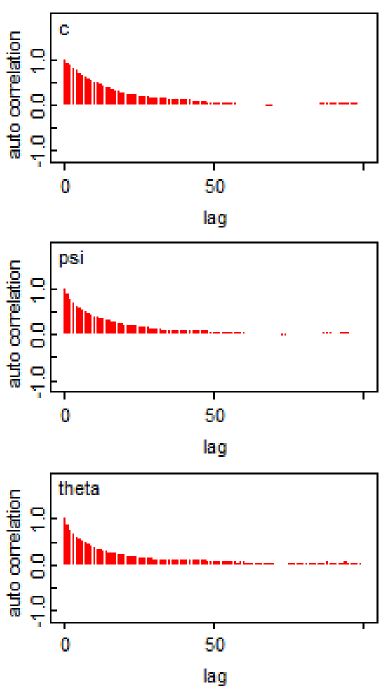
Figure 2.4: Auto-correlation Plots



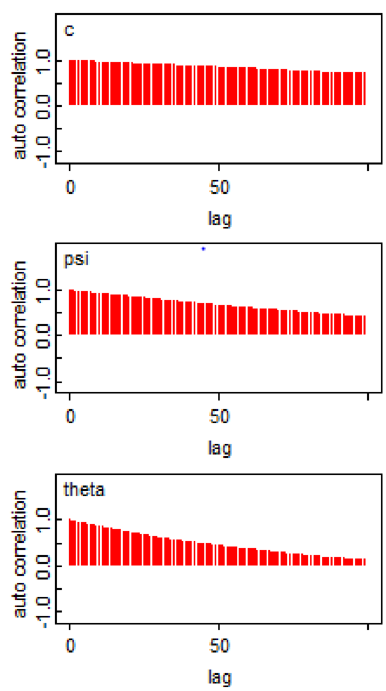
(a) Exponential Informative



(b) Exponential Non-Informative

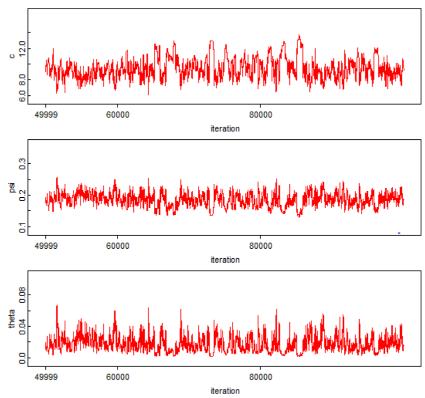


(c) Gamma Informative

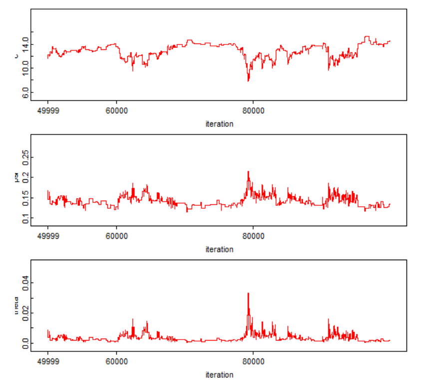


(d) Gamma Non-Informative

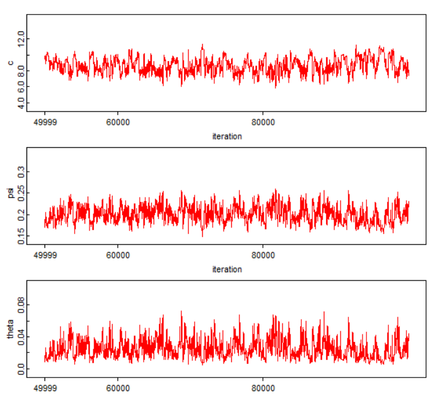
Figure 2.5: History



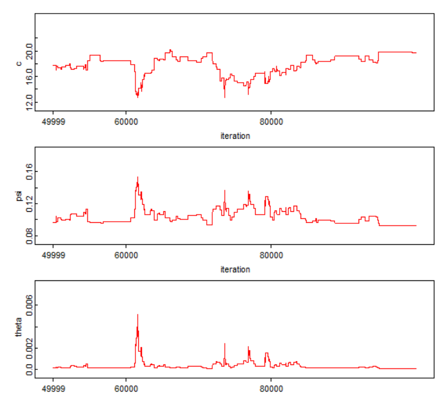
(a) Exponential Informative



(b) Exponential Non-Informative



(c) Gamma Informative



(d) Gamma Non-Informative

CHAPTER 3

SIMULATION

3.1 Simulation

In this section we will describe two simulation studies. The first one was designed to study the performance of the Bayesian WPD model in terms of bias and MSE of the model parameters. This study was repeated for various parameter values. For each parameter choice, we generated 100 data sets each with 100 observations, and 100 data sets with 500 observation. The second simulation study was performed to assess how well this model can perform while some percentage of the observed values are censored. In this study we are interested in capturing the true parameter value, for various censoring percentiles, through the 95% credible interval. The results given in Tables 3.3 and 3.4 were based on 500 pseudo data each with 100 observations. Because of the complexity of the WPD model we restrict our simulation study to 100 or 500 generated dated sets. However, in the future we will consider a large number of iterations. We chose exponential prior on both c and β. Moreover, as discussed in Section 2.2.1, we considered a truncated exponential prior on θ. Our prior choices are, $\theta \sim exp(0.05)T(0, x_{(1)}), \beta \sim exp(0.05)$, and $c \sim exp(0.05)$, where $x_{(1)}$ is the minimum of the sample.

3.2 Data Generation

In this section, we describe the algorithm to obtain pseudo data from the Weibull-Pareto distribution. We found Lemma 2.2 very useful as it relates the Weibull-Pareto to the Weibull distribution. We use the following steps in order to generate data.

1. Generate n observations from Weibull with parameters c and $1/\beta$.

2. Use the data in Step 1 and Lemma 2.2 to generate data of size n from the

Weibull-Pareto distribution.

3. Repeat steps 1 and 2 for a large number of iterations.

4. Repeat steps 1 through 3 to generate data sets with different parameter choices.

3.3 Simulation Study 1: Performance of Bayesian WPD model

This study was designed to examine how well the Bayesian Weibull-Pareto can perform in estimating the model parameters. In Table 3.1, we used $n = 100$ and Table 3.2 we used $n = 500$. We examine different sizes to see how increasing the number of observations effects bias and MSE. Our purpose for studying the WPD is to see how well it works with positively skewed data. In Figure 2.1, we saw that with small values of c WPD is positively skewed. In Table 3.1 and 3.2 we will see how well WPD works for small values of c for different sizes of n.

Table 3.1: Simulation Results for n=100

True value			Bias			MSE		
c	β	θ	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
0.5	0.5	0.5	0.002129811	0.073102786	-0.000002164	0.001761814	0.01895894	2.002612e-07
0.5	0.5	1	0.0099150100	0.0333320080	0.0001145717	0.001593829	0.01199942	2.988987e-06
0.5	0.5	3	0.0091700225	0.0772137927	-0.0002322325	0.002082500	0.01922018	5.380738e-06
1	0.5	0.5	0.006265250	0.013903652	-0.001628796	0.0071278183	0.0027844764	0.0001129739
1	0.5	1	0.015711468	0.010521510	-0.003227082	0.0074881016	0.0030497129	0.0004371173
1	0.5	3	0.0126477950	-0.0001631812	-0.0029130300	0.008015383	0.003231458	0.004849308
1	1	1	0.0181986190	0.0136460478	-0.0008130798	0.0083963957	0.0137898502	0.0001659547
1	1	3	0.01291250	0.01263585	-0.00881712	0.006798370	0.014594735	0.001290123
1	3	1	0.02482298	0.01923111	-0.00134432	0.009365873	0.1196966	1.080989e-05
1	3	3	0.009365939	0.062604910	-0.002626022	0.0077594085	0.1097467409	0.0001132329
1	7	1	0.0183454520	0.0439909200	-0.0006295968	0.008634790	0.6293276	2.560957e-06
4	0.5	0.5	0.15924688	0.02263368	0.03699240	0.786479570	0.008670353	0.026776585
4	0.5	1	0.05652267	0.02837874	0.09484234	0.612100354	0.007922381	0.103516462
4	0.5	3	0.14240777	0.02181913	0.21100991	0.736452697	0.007194806	0.804046570
4	1	0.5	0.4241059525	0.0006509685	-0.0163274190	1.292133717	0.033847635	0.007633421
4	3	1	0.86961504	-0.12338393	-0.04923811	2.702958484	0.403874856	0.009796301
4	3	3	1.1247682	-0.1955078	-0.1776196	4.3029700	0.4265535	0.1068641
4	7	1	1.24921152	-0.53722243	-0.03582712	5.096688635	2.582148315	0.003720049
7	0.5	0.5	-0.1472447	0.1036607	0.1803744	3.84196003	0.03695746	0.10242634
7	0.5	1	-0.30082735	0.09647315	0.33992266	2.77187225	0.02778827	0.33596013
7	0.5	3	-0.3593476	0.1062062	1.1067052	2.68143224	0.03265477	3.49608159
7	1	0.5	0.67829400	0.10219010	0.01170808	5.46684825	0.08569965	0.01383927
7	1	1	0.615419020	0.088010525	0.009819396	5.00692051	0.08421026	0.05392994
7	1	3	0.3153701	0.1493958	0.1614523	4.6895699	0.1284626	0.6313439
7	3	0.5	2.12247436	-0.04913447	-0.03192561	12.255698323	0.673126395	0.003397868
7	3	1	1.22950478	0.30888657	-0.03063851	9.42094087	1.22883087	0.01234657
7	3	3	1.7346162	0.1015389	-0.1577227	12.9595889	0.9568738	0.1449145

The bias was calculated by subtracting the mean estimate from the true parameter (i.e $c - \hat{c}$). The MSE was calculated as variation plus the bias squared (i.e $Var(\hat{c}) + bais_c^2$). When $c \leq 1$, we see that the model has small bias and MSE throughout all parameter choices. Suggesting the model works well when we have positively skewed data. When $c > 1$, the bias and MSE for c increased greatly. When $c > 1$ and $\beta > 1$, the bias and MSE for β also increases. The parameter θ performed rela-

Table 3.2: Simulation for results n=500

True value			Bias			MSE		
c	β	θ	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
0.5	0.5	0.5	5.596911e-03	1.384806e-02	-7.328075e-05	1.339933e-03	1.417710e-02	8.451127e-08
0.5	0.5	1	7.531968e-03	7.531968e-03	-4.508175e-05	1.409118e-03	1.573573e-02	8.861059e-07
0.5	0.5	3	0.008797122	0.051014747	-0.000357250	1.929890e-03	1.826296e-02	4.885888e-06
1	0.5	0.5	-0.0012497210	-0.0007515835	0.0002451115	1.223116e-03	6.237983e-04	4.646336e-06
1	0.5	1	0.0028293625	0.0030992490	0.0001162338	1.313297e-03	6.088439e-04	2.161319e-05
1	0.5	3	-0.0017509243	-0.0017509243	-0.0004026875	0.0008745643	0.0006008416	0.0001518166
1	1	1	0.0010709212	-0.0037865287	-0.0000217805	1.377850e-03	2.654785e-03	4.211163e-06
1	1	3	0.006507584	-0.0015396895	-0.0000382000	1.201734e-03	2.535635e-03	4.326752e-05
1	3	3	-0.0030374823	0.0095009975	-0.0001363425	1.187428e-03	1.834478e-02	3.489241e-06
1	3	1	8.154675e-03	-9.874600e-04	1.338901e-03	1.338901e-03	2.154246e-02	1.080989e-05
1	7	1	4.355002e-03	-5.608451e-02	-6.797575e-05	1.184372e-03	1.133858e-01	1.023097e-07
4	0.5	0.5	0.038499715	0.006342355	0.010955393	0.167494824	0.002181525	0.007539611
4	0.5	1	0.077152303	0.077152303	-0.001557034	0.212408305	0.002319323	0.030607729
4	0.5	3	0.030607729	0.007371553	0.071236146	0.170972268	0.001918046	0.243701035
4	1	0.5	0.129761378	-0.012261333	-0.009609649	0.162493926	0.006097347	0.001409241
4	3	1	0.21013405	-0.06566586	-0.01246237	0.33505571	0.09393229	0.00136805
4	3	3	0.15553516	-0.04756925	-0.03447667	0.228879979	0.069509991	0.008803241
4	7	1	0.11262364	-0.07919613	-0.00400735	0.178763771	0.340733447	0.000148442
7	0.5	0.5	0.03657421	0.02086180	0.03820247	1.278143427	0.005969986	0.005969986
7	0.5	1	0.245844248	0.009642056	0.036645576	0.036645576	0.00523846	0.07440450
7	0.5	3	0.11888068	0.01589822	0.17984952	1.612725718	0.006773238	0.849629403
7	1	0.5	0.120854155	0.050013526	0.009339225	1.784286979	0.034474075	0.007168955
7	3	0.5	0.77232087	-0.08332107	-0.01415032	2.807819288	0.207978787	0.001045332
7	1	1	0.67104516	-0.01030971	-0.04057308	3.74455055	0.03429284	0.03429284
7	1	3	0.17691185	0.04321077	0.03092214	2.44122483	0.03696885	0.28205981
7	3	1	0.80810035	-0.04121598	-0.02746845	4.656942717	0.319094556	0.006507584
7	3	3	0.9278334	-0.1150892	-0.1022471	3.7579442	0.2773772	0.0523737

tively well. The highest bias and MSE for θ occurred when both c and β were large. We can see from Table 3.2, compared to Table 3.1, when we increase the number of observations the bias and MSE decrease significantly. If we were to compare the last three rows of Table 3.1 and 3.2 (WPD(7,1,3), WPD(7,3,1), and WPD(7,3,3)), we see that the bias and MSE decreases significantly as expected.

3.4 Simulation Study 2: Bayesian WPD under censoring

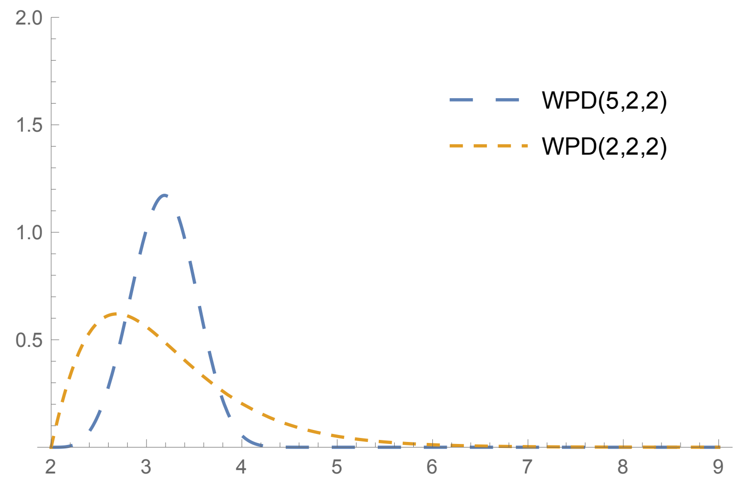
We will introduce the survival model as follows:

$$\mathcal{S}(x,t) = [g(x)]^t * [1 - G(x)]^{t-1} \tag{3.1}$$

where $t = \{0, 1\}$ is the censoring variable, $g(x)$ is the pdf and $G(x)$ is the cdf of the WPD given by Equations (2.1) and (2.2) respectively. Note that the survival function $S(x) = 1 - G(x)$. From Equation 3.1, if an observation is censored, $t = 0$, the value comes from the survival function; if it is not, $t = 1$, it is coming from the pdf, $g(x)$. This model will later be used in Section 4.2. We used right censoring for the simulation. We generated 100 observations in the way described in Section 3.2. We produce the censoring variable by ordering the observations, taking the j^{th} observation, and censoring everything above it. In this way we have exactly $(100 - j)\%$ censored. We generated 500 different data sets. We wanted to see how both a symmetric and skewed distribution performed under the censoring model which is shown in Figure 3.1.

For Table 3.3 and Table 3.4 bias and MSE were calculated as described in Section 3.3. %C is the censoring percentage. The % True Parameter of CI is the number of times the 95% credible interval captured the true parameter divided by the number of iteration (iter=500). From Table 3.3, we see that $\hat{c}$ has relatively high bias and MSE. For $\hat{\beta}$ the bias and MSE were relatively low for 10% and 20% censoring. However,

Figure 3.1: PDF for parameter choices



there was a large increase in the 35% and 50% censoring. $\hat{\theta}$ performed well since both the MSE and bias where relatively low. For all parameters the bias and MSE increased as the censoring percentage increased. The proportion of the true parameter captured by the credible interval decreased as the censoring percentage increased.

As we can see in Table 3.4, $\hat{c}$ had both high bias and MSE. $\hat{\beta}$ had large bias and MSE throughout all censoring choices. $\hat{\theta}$ also had large bias and MSE. Again, the bias and MSE increased for all parameters as the censoring increased. The proportion of the true parameter captured by the credible interval decreased as the censoring percentage increased, except for the parameter β, where the proportion for β increased. Table 3.5 provides the parameter estimates for 35% and 50%. We see β at 35% is higher than β at 50% , and therefore the censoring may have caused the credible interval to over estimate the parameter β.

Table 3.3: Right Censoring for WPD(2,2,2)

Bias(MSE)				%True Parameter Captured by CI		
%C	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
10	-0.26939507	0.01421504	0.02632264	0.744	0.946	0.87
	(0.11441356)	(0.03343525)	(0.00457954)			
20	-0.49162637	-0.07831612	0.04394210	0.0302	0.946	0.804
	(0.269685739)	(0.028299786)	(0.004787532)			
35	-0.75643641	-0.34672244	0.06169448	0.026	0.502	0.634
	(0.59125417)	(0.13150875)	(0.00665626)			
50	-0.95687543	-0.75339821	0.06558217	0	0.006	0.534
	(0.927656797)	(0.574439426)	0.007724839			

Table 3.4: Right Censoring for WPD(5,2,2)

Bias(MSE)				%True Parameter Captured by CI		
%C	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
10	-1.3463562	0.4909197	0.1709286	0.662	0.802	0.778
	2.43496554	0.45633039	0.05693664			
20	-1.9829797	0.5566727	0.2347656	0.302	0.696	0.622
	4.29508220	0.48576893	0.07609605			
35	-2.6791479	0.4747181	0.3001660	0.04	0.578	0.38
	7.4101037	0.3472820	0.1083566			
50	-3.13417537	0.09953852	0.31507083	0.008	0.95	0.3
	9.99747159	0.06539004	0.11810655			

Table 3.5: WPD(5,2,2) Parameter Estimate for 35% and 50%

	Mean			Sd		
%C	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
35	2.320852	2.474718	2.300166	0.4540593	0.2550084	0.0914429
50	1.865825	2.099539	2.315071	0.37754154	0.23091182	0.07767302

CHAPTER 4

APPLICATION

4.1 Tribolium Confusum and Tribolium Casteneum

In this section,we use the WPD to model data from two different studies on adult numbers of Tribolium Confusum. We used these data sets to investigate how well WPD performs for different shapes of applied data. The data in Table 4.1 is approximately symmetric and the data in Table 4.2 has a very long tail with left tail characteristics.

The data is from Park et al. (1964)[12] and Park (1954)[11]. One sample was kept at a temperature of 29° and the other is at 24°. We will compare how the WPD compares with generalized Weibull (GW), exponentiated-Weibull (EW), and the Weibull distribution. Table 4.1 and Table 4.2 contain the observed and expected frequency of the two sets of data. Figure 4.1 and Figure 4.2 correspond to Table 4.1 and Figure 4.3 and Figure 4.4 correspond to Table 4.2.

For our data, we should expect our models to capture the tails of the data. In Table 4.1, we see the first observed value is five and the expected frequency for Weibull is 11.3131. The WPD comes closest for x values 35-40 with expected frequency 5.12015. GW and EW also estimated the first value well. All models seem to perform well for the middle of the data distribution, which is expected. For the last three observations, Weibull does not do well in capturing the tail of the data. The values for Weibull dropped off faster compared to the other models. WPD and EW performed about the same, capturing the tail of the data. GW values dropped faster than WPD and EW, but not as quickly as Weibull.

Table 4.2 shows Weibull provided poor estimates poorly for all x-values. EW and WPD seem to estimate the tails of the data about the same while GW did not seem to capture the tails. EW, WPD, and GW seem to perform about the same for the middle x-values.

We constructed histograms of the observed data with overlays of the curves for both tables to provide a clear picture of how they performed. Figure 4.1 demonstrates how well each model estimates the data. In the upper right of Figure 4.1 we can see the Weibull tail drops of quickly compared to the other models. In the rest of the graphs in Figure 4.1 we see the models capture the tails of the data. Figure 4.2 shows the overlay of the curves; we do not see a clear distinction between EW,GW, and WPD.

Again, Figure 4.3 shows that Weibull right tail drops off quickly and does not provide an precise estimate of the data. GW provides better estimates of the middle x-values than WPD and EW. Figure 4.4 shows the models perform about the same in capturing the tails of the data.

Figure 4.1: Histogram and Curves

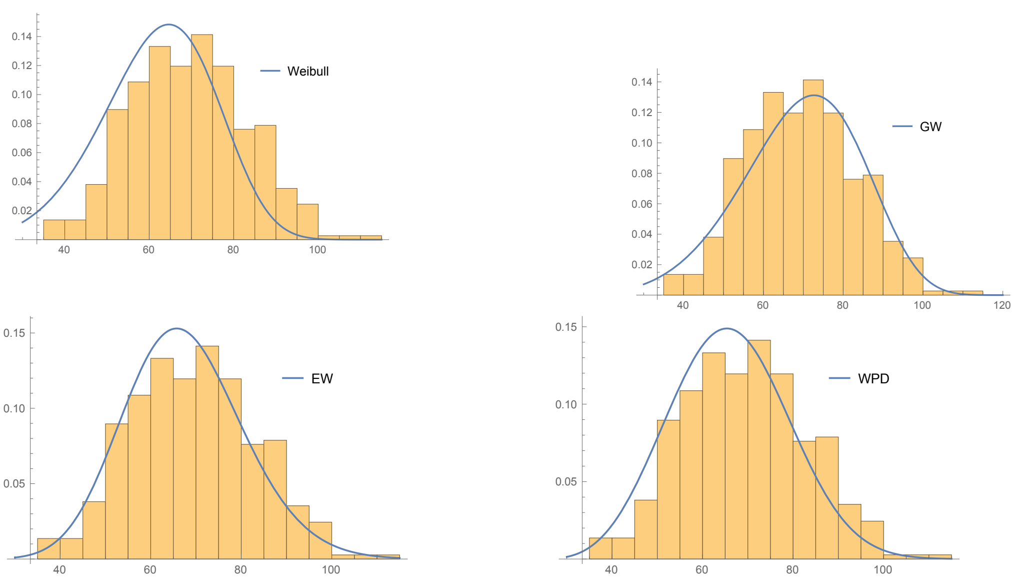


Table 4.1: Observed and expected frequencies of Confusum 29°

x-value	Observed	Weibull	Generalized Weibull	Exponentiated Weibull	Weibull-Pareto
35-40	5	11.3121	6.58432	3.12327	5.12015
40-45	5	18.5436	10.8356	8.54702	12.2578
45-50	14	27.8444	16.5835	18.3373	22.734
50-55	33	38.2767	23.726	31.7221	35.0143
55-60	40	47.8314	31.7438	45.2372	46.2852
60-65	49	53.6348	39.5699	54.1866	53.3184
65-70	44	52.9734	45.6122	55.4484	53.8479
70-75	52	44.9804	48.0683	49.2368	47.7321
75-80	44	31.8667	45.5942	38.4942	37.0826
80-85	28	18.1731	38.1322	26.8443	25.1793
85-90	29	7.99975	27.3789	16.8824	14.8929
90-95	13	2.5899	16.3023	9.65709	7.64642
95-100	9	0.583621	7.69315	5.054	3.39628
100-105	1	0.0860152	2.70823	2.42802	1.30088
105-110	1	0.00772955	0.654471	1.07209	0.42846
110-115	1	0.000391392	0.0964364	0.434938	0.121039
Total	368	356.704	361.284	366.706	366.358
Parameter Estimates		v=5.316	$\alpha = .1928$	$\alpha = 2.795$	$c = 6.694$
		$\lambda = 1.93 * 10^{-10}$	$\lambda = 0.395$	$\sigma = 52.74$	$\beta = .798$
			$\phi = 141$	$\theta = 4.502$	$\theta = 20.06$

Table 4.2: Observed and expected frequencies of Confusum 24°

x	Observed	Weibull	Generalized Weibull	Exponentiated Weibull	Weibull-Pareto
20-30	0	4.08727	3.81726	0.0706614	0
30-40	0	9.95255	9.05459	0.871671	0
40-50	3	19.2467	17.1896	4.93156	6.16691
50-60	9	32.1385	28.3509	16.6664	21.8221
60-70	39	48.2017	42.2456	38.9595	45.9356
70-80	53	66.232	58.0376	69.0166	73.1826
80-90	77	84.1816	74.2854	98.5644	97.3012
90-100	105	99.3362	89.0104	118.655	113.191
100-110	135	108.814	99.9526	124.529	118.255
110-120	114	110.347	105.026	116.965	112.746
120-130	113	103.112	102.891	100.38	99.159
130-140	92	88.2405	93.4732	80.0228	81.0757
140-150	59	68.6524	78.187	60.0404	61.9983
150-160	54	48.1565	59.6824	42.8357	44.5553
160-170	38	30.1769	41.1235	29.2925	30.2142
170-180	22	16.7261	25.2413	19.3165	19.4014
180-190	17	8.11328	13.5805	12.3391	11.8332
190-200	6	3.40556	6.27936	7.66069	6.87419
200-210	10	1.22242	2.4348	4.63371	3.81306
210-220	3	0.370604	0.767591	2.7354	2.02419
220-230	2	0.0936773	0.189087	1.57792	1.03055
230-240	0	0.0194784	0.0345429	0.890258	0.50416
240-250	1	0.00328556	0.00436008	0.491594	0.237433
250-260	0	0.000443128	0.000344343	0.265816	0.107824
260-270	0	0.0000470792	0.000014711	0.140805	0.0472914
Total	952	931.583	950.858	951.852	952.476
Parameter Estimates		v=3.711	$\alpha = 0.2757$	$\alpha = 1.75$	c=5.324
		$\lambda = 1.827 * 10^{-8}$	$\lambda = 0.03658$	$\sigma = 70.2$	$\beta = 0.6845$
			$\phi = 318$	$\theta = 5.935$	$\theta = 28.12$

Figure 4.2: Overlay of Graphs

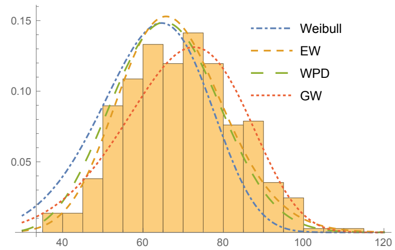


Figure 4.3: Histogram and Curves

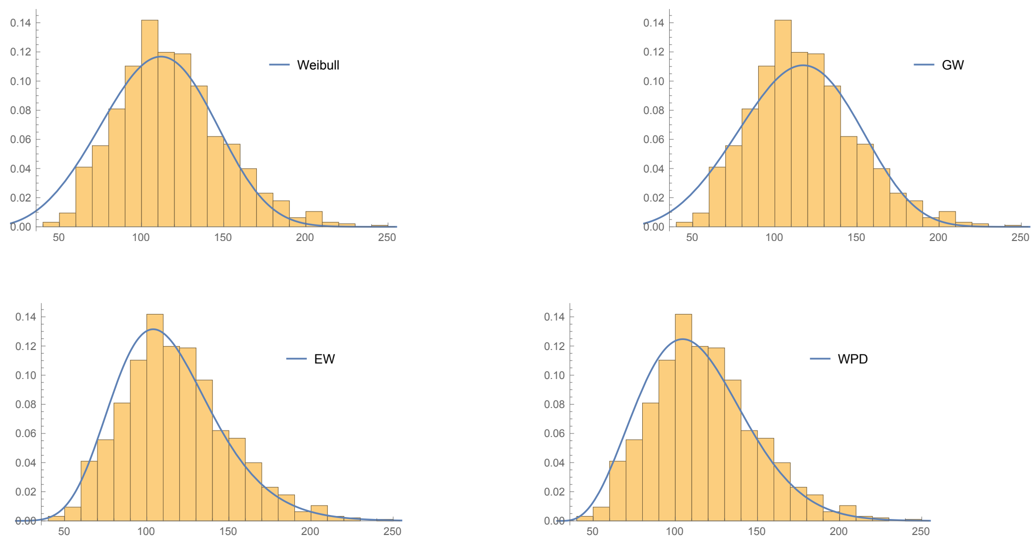
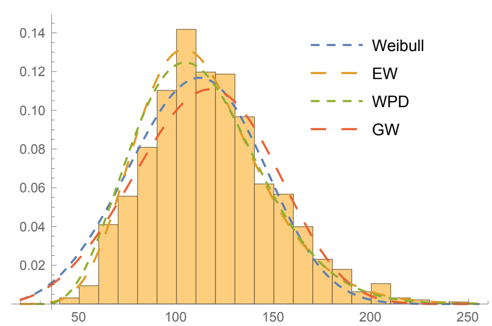


Figure 4.4: Overlay of Graphs



4.2 Melanoma data with Censoring

In this section, the data used was part of a study on cutaneous melanoma in order to evaluate a certain drug administered post operation. The study was from 1991 to

1995, and followups were conducted until 1998. The sample contains 417 patients; the observed time is the patient time of death or the time of censoring. The patients age is included[13]. We constructed a survival model using the Weibull-Pareto distribution and applied it to right censored data. From Equation (2.2), our survival function is given by :

$$S(x) = 1 - F(x) = exp\Big\{ - \Big(\beta log\Big(\frac{x}{\theta}\Big)\Big)^c\Big\},$$

where t is the censoring indicator. Then following Section 3.4 we can write the likelihood as:

Likelihood:

$$[f(x)]^t * [S(x)]^{1-t}.$$

In this example we would like to model one of the Weibull-Pareto parameters through the available covariate age. One obvious choice is to model the $log(scale)$ through a linear function involving the covariate or in other words: $1\backslash\beta = e^{\psi_{age} * x_{age}}$ (used ψ as regression coefficient to prevent confusion). We chose β to model the co- variate since c and θ would be difficult to interpret. Moreover, in the Bayesian model the prior on θ is truncated at $x_{(1)}$, this restriction may not hold if we model the covariate through θ. On the other hand, the model will become computationally expensive if we chose the parameter c instead. The next step involves the prior specification on the model parameters c, θ and ψ_{age}. For the priors on c and θ, we used a flat gamma prior. For ψ_{age} we used a normal prior. Our priors are as follows:

Prior:

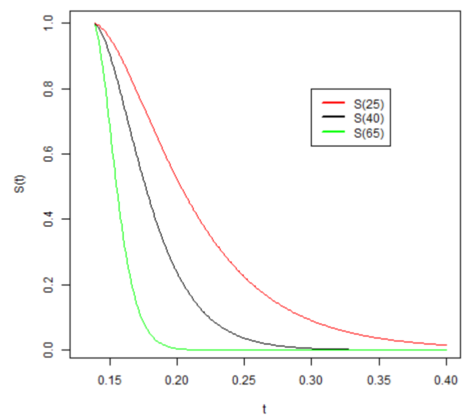
$$c \sim \Gamma(0, 0.1)$$

$$\theta \sim \Gamma(1, 0.1) I_{(0,x_{(1)})}(\theta)$$

$$\psi_{age} \sim N(0, 0.1).$$

Our estimation for the parameters are given:

Figure 4.5: Graph of Survival Function



- $\hat{c} = 1.778$
- $\hat{\psi}_{age} = 0.0299$
- $\hat{\theta} = 0.1381$

Finally we turn our attention to the role of the covariate in estimation the survival function. The positive estimate of the regression parameter shows that as age increases the rate parameter of WPD (β) decreases, i.e, the tail of the distribution will drop at a much smaller rate results in a flat curve. Therefore the survival function will drop at a much faster rate. This fact can be supported by Figure 4.5 which shows our survival curves for ages 25, 40, and 65. We can see someone at age 25 has an overall better chance of surviving than someone at age 65. For example, when $t = 0.25$, the survival curve for 65 is zero and the survival curve for 25 there is about 20% chance of surviving.

CHAPTER 5

CONCLUSION

We explored how Bayes estimates performed for the parameters of the Weibull-Pareto Distribution through simulations. The simulation showed the model performed well for $c \leq 1$, which is when the distribution is positively skewed. Issues arise when both c and $\beta > 1$, there is large bias among the parameters. The right censoring simulation gave us an idea how precisely the model works for censored data for a symmetrically shaped and skewed shaped Weibull-Pareto distribution. As the censoring increased, the bias increased and decreased number of times our credible interval captured the true parameter for WPD(2,2,2). For WPD(5,2,2), we saw how β was overestimated at 35% censoring which caused the credible interval to fail to capture β. For the application of the Tribolium Confusum, WPD Bayesian estimates were comparable to EW and GW. In the application to survival data we see that the covariate age shows that as age increases the chance of survival decreases. The software used for data analysis were R 3.1.2, OpenBUGS, and Wolfram Mathematica 10.

For our future research we will investigate other loss functions such as the Linex loss. We will also investigate WPD under other censoring methods such as Type I, Type II, and progressive. Additionally, we will apply the model to more relevant data sets with a heavy tail. Further we will see the performance of Bayesian WPD under small data sizes.

REFERENCES

[1] Alzaatreh, A., Lee, C., Famoye, F. (2013). Weibull-Pareto Distribution and It's Applications. *Communications in Statistics-Theory and Methods*, 42: 1673-1691.

[2] Alzaatreh, A., Lee, C., Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron: International Journal of Statistics*. 71:6379

[3] Famoye, F., Lee, C., Olumolade, O. (2005).The beta-weibull distribution.*J. Statist. Theor. Appl.*, 4(2):121136.

[4] Hosmer,D., Lemeshow, S., May, S., *Applied Survival Analysis*,2nd ed., John Wiley & Sons, Inc., (New Jersey)(2008).

[5] Khan, H., M., R., Saxena, A., Vera, V., Abdool-Ghany, F. Gabbidon, K., Perea, N., Stewart, T. S., Ramamoorthy, V. (2014). Black Hispanic and Black Non-Hispanic Breast Cancer Survival Data Analysis with Half-normal Model Application. *Asian PAcific Journal of Cancer Prevention* 15 9453-9458.

[6] Lynch, Scott. *Introduction to Applied Statistics and Estimation for Social Scientists*, Springer, (New York) (2007).

[7] Martin,J., Perez,C.J (2009),Bayesian analysis of a generalized lognormal distribution,*Computational Statistics and Data Analysis*, 53: 1377-1387.

[8] Mudholkar, G. S., Kollia, G. D. (1994). Generalized Weibull family: A structural analysis. *Commun. Statist. Theor. Meth.*, 23(4):11491171.

[9] Mudholkar, G. S., Srivastava, D. K., Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics* 37(4):436445.

[10] Noor,F. and Aslam M. (2013), Bayesian inference of the inverse Weibull mixture distribution using type-I censoring,*Journal of Applied Statistics*, Vol.40, No.5, 1076-1089.

[11] Park, T. (1954). Experimental studies of interspecies competition II. Temperature, humidity, and competition in two species of Tribolium. *Physiological Zoology* 27:177238.

[12] Park,T.,Leslie, P.H., Mertz., D. B. (1964). Genetic strains and competition in population of Tribolium.*Physiological Zoology* 37:97-162.

[13] Rodrigues, J., Concho, V.G, Castro, M., and Balakrishnan,N. (2010). A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data. *Statistical Methods in Medical Research* 21(6) 585-597.

[14] Singh, S.K, Singh, U., and Kumar, D.(2013), Bayesian estimation of parameters of inverse Weibull distribution,*Journal of Applied Statistics*, Vol.40, No.7, 1597-1607.

[15] Smith, L.R. (1985). Maximum likelihood in a class of nonregular cases. *Biometrika*, 72 (1) : 67-90.