ESTIMATION OF THE WEIBULL DISTRIBUTION

WITH APPLICATIONS TO TORNADO

CLIMATOLOGY

by

MICHAEL B. MCCLELLAN

STAVROS A. BELBAS, COMMITTEE CHAIR

PU WANG
SEONGSIN KIM

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts
in the Department of Mathematics
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2012

ABSTRACT

Some general properties of the Weibull distribution are discussed. The mathematical development of the distribution is linked to the family of extreme value distributions, and the origins in science are found to be related to survival analysis. A selection of generalizations of the distribution are noted, and a limited discussion of its numerous applications undertaken.

One such application is the Weibull model of tornado intensity developed by Dotzek, Grieser, and Brooks (2003). In an attempt to improve this model, several methods for estimating the parameters of the Weibull distribution are discussed. Maximum likelihood estimation is found to be the best method of estimation for the two-parameter Weibull distribution with respect to certain asymptotic estimator properties and ease of implementation. An existing algorithm to locate the maximum likelihood estimator for the three-parameter Weibull distribution is described, and the complexities of the three-parameter case investigated.

It is known that the maximum likelihood estimates for the Weibull distribution display bias for small sample sizes. An equation is analytically derived to describe this small sample bias in the two-parameter case, and numerical unbiasing procedures discussed.

Simulated data are analyzed using the methods developed, and the asymptotic properties of the estimates detailed for the two-parameter case. The estimation procedures are then applied to actual tornado intensity data from the April 25th – 28th, 2011 tornado outbreak as well as the historic records for both Alabama and the United States as a whole. In all cases, the Weibull model is found to be appropriate as judged by the Chi-squared test at 5% significance.

DEDICATION

This thesis is dedicated to all those affected by the events of April 27th, 2011. In particular, I would like to acknowledge Johnny M. Hanna and Zackery J. Tavel, whose friendship and strength during following weeks motivated me to study the statistical methods of tornado climatology.

LIST OF ABBREVIATIONS AND SYMBOLS

$\Pr\{A\}$	Probability of the event A
3PW	Three-parameter Weibull distribution
2PW	Two-parameter Weibull distribution
1PW	One-parameter Weibull distribution
$E(\cdot)$	Expectation operator
μ	Mean: first moment about zero
σ^2	Variance: second moment about zero minus the square of the mean
$\boldsymbol{\theta}$	True parameter
$\hat{\boldsymbol{\theta}}$	Parameter estimate
x_p	$100p^{\text{th}}$ percentile
MLE	Maximum likelihood estimator
$\sim$	Approximately
$\psi(\cdot)$	Euler’s psi function: logarithmic derivative of the gamma function
$B(\hat{\boldsymbol{\theta}})$	Relative bias of the estimate: equal to $\left|\left(\hat{\theta}-\theta\right)/\theta\right|$
χ^2	Chi-squared goodness of fit value

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis advisor, Stavros A. Belbas, for his invaluable guidance and advice throughout my research. I would also like to express my appreciation of my committee members, Seongsin Kim and Pu Wang, for their input and support throughout the writing process.

In addition, I would like thank the University of Alabama community as a whole. From undergraduate freshmen to second year graduate student, the University has continuously supported me and provided me with a stimulating and challenging academic experience. In particular, I am indebted to the faculty, staff, and students of the Mathematics department. It has been a pleasure becoming acquainted with the department over the past two years, and I am honored to have worked and studied with its members.

Of course, none of my accomplishments would have been possible without the love and support of my family, who have always pushed me to strive for my best. To them, I am eternally grateful. I would also like to thank my fiancée, Bethany Bakane, for her input on the mentality of research and for encouraging me when I was frustrated with my progress.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The tornado outbreak of April 25th – 28th, 2011 was a devastating severe weather event for all affected areas. The events of April 27th had a particularly great impact on the University of Alabama community, as well as the state of Alabama as a whole. Of the 207 reported tornadoes on April 27th, a total of 62 were in Alabama, setting the record for most tornadoes in a single event in Alabama (see National Weather Service Weather Forecast Office, 2011). An untold number of people were injured with a reported 248 fatalities due to tornadoes on April 27th in Alabama alone. A total of 753 tornadoes made April 2011 the most active tornado month on record in the United States according to a 2011 report by the National Oceanographic and Atmospheric Administration. The impact of these events continues to be far reaching. The homes and places of business of many Alabamians were destroyed or damaged. A total of approximately $72 million in individual and household aid was approved according to the Joint Alabama – FEMA Situation Report No. 53 published in 2011. Due to great impact of such severe weather events, it is in the public interest to develop better prediction methods and risk assessment procedures. In Johnson and Holt (1997) argue that because weather information is largely subsidized by the public, efforts to justify the information's cost give it value. Economists have attempted to quantify this value of weather information both monetarily and socially (see references in Murthy, Xie, & Jiang, 2004). Therefore, it is natural to seek to improve the methods of collecting and interpreting weather information to increase this perceived value to the public good.

In an attempt to develop a statistical model for tornado climatology, Dotzek, Grieser, and Brooks (2003) proposed using the Weibull distribution to model tornado intensity due to intensity ratings. They showed that the Weibull distribution provided a better fit to the observed tornado intensity data than by the previously employed exponential distribution. However, this advantage comes at the cost of working with a more complicated distribution. Unlike the exponential distribution, equations for the parameters of the Weibull distribution are not always available in closed form. Furthermore, iterative methods used to estimate the parameters are not guaranteed to converge. Compared to the exponential distribution, whose single parameter can be readily estimated by a sufficient statistic, estimation of the Weibull distribution presents a more computationally intensive challenge. Therefore, work can be done to determine the comparative performance of different estimation procedures for the Weibull distribution and quantify the properties of the various estimates achieved.

In the first chapter of this thesis, an introduction to the Weibull distribution is given. Functional properties are derived, the historical development of the distribution is reviewed, and a limited selection of applications are discussed. The second chapter deals with the various methods of parameter estimation, with a focus on the method of maximum likelihood. An equation to estimate the bias of maximum likelihood estimate for the shape parameter is derived analytically, and numerical unbiasing procedures are reviewed. The third chapter details the numerical results of the estimation procedures applied to simulated data as well as data taken from historic tornado records. A discussion of the results can be found in chapter four.

THE WEIBULL DISTRIBUTION

The three parameter Weibull distribution (3PW) has probability distribution function

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left\{-\left(\frac{x-a}{b}\right)^{c}\right\}, \quad x \geq a \tag{1}$$

with location parameter a, scale parameter b, and shape parameter c. The location parameter can take all real values (i.e., $-\infty < a < \infty$) and determines the location of the distribution along the x-axis. The scale parameter can take positive values (i.e., $0 < b < \infty$) and determines the scale or dispersion of values along the x-axis. The shape parameter, fittingly, determines the shape or curvature the distribution takes, and can also assume all positive values.

If X is a random variable with density function $f(x)$, then the probability that X is less than or equal to the real number x is given by the cumulative distribution function defined by

$$\Pr\{X \leq x\} = F(x) \stackrel{\text{def}}{=} \int_{-\infty}^{x} f(x)\ \mathrm{d}x. \tag{2}$$

If X is a random variable distributed by the 3PW, integrating equation (1) from zero to x yields the cumulative distribution function of X,

$$F(x|a,b,c) = 1 - \exp\left\{-\left(\frac{x-a}{b}\right)^{c}\right\}, \quad x \geq a. \tag{3}$$

It is sometimes the case, in applications, that the location parameter is known, or equals zero. Performing an appropriate substitution then allows the 3PW to be reduced to the two parameter Weibull distribution (2PW) which has density function

$$f(x|b,c) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} \exp\left\{-\left(\frac{x}{b}\right)^{c}\right\} \tag{4}$$

and cumulative distribution function

$$F(x|b,c) = 1 - \exp\left\{-\left(\frac{x}{b}\right)^{c}\right\}, \quad x \geq 0. \tag{5}$$

An even simpler form of the Weibull distribution occurs when both the location and scale parameters are known. Another appropriate substitution then yields the density function of the one parameter Weibull distribution (1PW), which is given by

$$f(x|c) = c\, x^{c-1} \exp\{-x^{c}\}, \quad x \geq 0. \tag{6}$$

The moments of a distribution are important quantities in mathematical statistics, being frequently used as quantitative measures of the distributions shape. The r^{th} moment about zero of a random variable X with density function $f(x)$, also called the expectation of X^r, is given by

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x)\, dx. \tag{7}$$

Therefore, the r^{th} moment about zero of the 1PW is

$$E(X^r) = \int_0^\infty x^r \, c \, x^{c-1} \exp\{-x^c\} \, dx. \tag{8}$$

Performing the change of variables $t = x^c$, $dt = cx^{c-1}dx$ yields

$$E(X^r) = \int_0^\infty t^{r/c} \exp\{-t\} \, dt = \Gamma\left(\frac{r}{c} + 1\right) \tag{9}$$

where $\Gamma(\mathrm{z}) = \int_0^\infty t^{z-1} \exp\{-t\} \, dt$ is the widely studied Gamma function (see Artin, 1964/1931). Having calculated the expectation of X^r for the 1PW, the moments about zero for the 2PW and 3PW can be derived. This can be accomplished using straightforward transformations. In fact, if X is a random variable distributed by the 1PW, then bX is distributed by the 2PW and $a + bX$ is distributed by the 3PW for appropriate choices of a and b. The linearity of the expectation operator then yields

$$E((bX)^r) = b^r E(X^r) = b^r \Gamma\left(\frac{r}{c} + 1\right) \tag{10}$$

as the r^{th} moment about zero for a random variable following the 2PW. Again, the linearity of the expectation operator along with an application of the binomial theorem yields an equation for the r^{th} moment about zero for a random variable following the 3PW. Namely,

$$E((a+bX)^r) = \sum_{i=0}^{r} \binom{r}{i} a^i b^{r-i} E\left(X^{r-i}\right) = \sum_{i=0}^{r} \binom{r}{i} a^i b^{r-i} \Gamma\left(\frac{r-i}{c}+1\right). \tag{11}$$

Two important quantities arising from the concept of moments are the mean and variance of a distribution. The mean of a distribution, μ, is defined to be the first moment about zero. It is the weighted average of all possible values the random variable can take, and is analogous to the arithmetic mean when the probability is distributed uniformly. The means of the 1PW, 2PW, and 3PW can be found by letting $r = 1$ in equations (9), (10), and (11) respectively. The variance of a distribution, σ^2, is defined to be the second moment about zero subtracted by the square of the mean, i.e. $\sigma^2 = E(X^2) - \mu^2$. The variance can be interpreted as a measure of the width of the distribution, or how widely the data is dispersed about the mean. The second moment about zero for the three cases of the Weibull distribution can be found by setting $r = 2$ in equations (9), (10), and (11) respectively.

For the 3PW, the mean and variance are given by

$$\mu = a + b\Gamma\left(\frac{1}{c}+1\right) \tag{12}$$

and

$$\sigma^2 = b^2\left[\Gamma\left(\frac{2}{c}+1\right) - \Gamma\left(\frac{1}{c}+1\right)^2\right]. \tag{13}$$

Note that the variance increases or decreases as the scale parameter b increases or decreases. Thus, the interpretation of the scale parameter as a measure of the spread of values on the x-axis is appropriate.

ORIGINS AND APPLICATIONS OF THE DISTRIBUTION

The Weibull distribution owes its name to Swedish engineer Waloddi Weibull, who advocated its use across disciplinary boundaries during the mid-twentieth century. Weibull first proposed the distribution in a series of papers in 1939 on the breaking strength of certain materials (as cited by Rinne, 2009). It should be noted that Weibull arrived at his distribution through an analysis of the breaking strength data he had obtained through experiments, and not through a rigorous mathematical development. Therefore, the mathematical development of the distribution will be presented before Weibull's work is detailed.

Rinne (2009) traced the mathematical origin of the Weibull distribution to the development of the family of extreme value distributions. In mathematical statistics, it is common practice to study the smallest and largest values of a sample of size n. Symbolically, these extreme values are Y_1 and Y_n, called the first and n^{th} order statistics, respectively. The distributions of Y_1 and Y_n over all possible samples make up the family of extreme value distributions mentioned above. Of particular importance to the development of the Weibull distribution are the asymptotic distributions of Y_1 and Y_n, i.e., the distributions obtained as the sample size, n, increases without bound.

Although French mathematician Maurice René Fréchet arrived at the asymptotic distribution of Y_n in 1927, the English mathematicians Ronald Fisher and Leonard Tippett are more widely cited relating to extreme value theory (see Rinne, 2009). In their 1928 paper, Fisher and Tippett showed that the asymptotic probability distribution of Y_n must take one of the following three forms:

$$\text{I.} \quad f(x) = \exp\{-x - exp(-x)\}\, \mathrm{d}x \tag{14}$$

$$\text{II.} \quad f(x|k) = \left(\frac{k}{x^{k+1}}\right) \exp\{-x^{-k}\}\, \mathrm{d}x$$

$$\text{III.} \quad f(x|k) = k\,(-x)^{k-1} \exp\{-(-x)^{k}\}\, \mathrm{d}x$$

These three probability density functions are defined for positive x values, and have become known as the type I, II, and III extreme value distributions of Y_n. Upon closer inspection of the type III density function, similarities with that of the 1PW become apparent.

Performing the substitution $x \to -x$ leads to the type I, II, and III extreme value distributions of Y_1:

$$\text{I.} \quad f(x) = \exp\{x - exp\{x\}\}\, \mathrm{d}x \tag{15}$$

$$\text{II.} \quad f(x|k) = \left(\frac{k}{(-x)^{k+1}}\right) \exp\{-(-x)^{-k}\}\, \mathrm{d}x$$

$$\text{III.} \quad f(x|k) = kx^{k-1} \exp\{-x^{k}\}\, \mathrm{d}x$$

This is due to the fact that, for a sample of real numbers $\{x_1, x_2, \ldots, x_n\}$, $\max_{1 \le i \le n}\{x_i\} = -\min_{1 \le i \le n}\{-x_i\}$. It can now be seen that the type III PDF of Y_1 is exactly the PDF of the Weibull distribution with parameters $a = 0$, $b = 1$, and $c = k$.

Although Fisher and Tippett (1928) only derived the type I, II, and III extreme value distributions of Y_n in their paper, they were surely aware of the straightforward extension to the first order statistic. In fact, the title of their paper guarantees that they were. Starting with a random variable X taken from the type III extreme value distribution of Y_1, the 3PW can be reached through the substitution $a + bX$ for appropriate values of a and b.

The Weibull distribution is not the only celebrated distribution to arise out of the family of extreme value distributions. The type I distribution of Y_n, known as the Gumbel distribution, is named after German mathematician Emil Julius Gumbel. The type II distribution of Y_n, called the Fréchet distribution, is named after the aforementioned Fréchet. Like the Weibull distribution, both the Gumbel and Fréchet distributions have been dealt with extensively in the literature of mathematical statistics, allowing the study of extreme values to become an established topic of research.

Before returning to Weibull's development of the distribution, the topic of survival analysis will be discussed. Survival analysis uses random variables to model the time to failure of a manufactured device or time to death of a biological organism. Two functions that are of particular interest in survival analysis are the survival function and the hazard function. The survival function, S, for a random variable is defined to be the complement of its cumulative distribution function, i.e. $S(x) \stackrel{\text{def}}{=} 1 - F(x) = \Pr\{X > x\}$. Therefore, the Weibull distribution with parameters a, b, and c has the survival function

$$S(x) = 1 - F(x) = \exp\left\{-\left(\frac{x-a}{b}\right)^c\right\}, \quad x \geq a. \tag{16}$$

This function is useful in that it gives the probability of surviving through time x. In his 1951 paper, Weibull discussed how he originally arrived at the 2WP using a survival analysis of a chain consisting of n links. The probability that any one link survives through time x is given by its survival function, S. Weibull reasoned that the probability of the chain surviving through time x is the same as the probability of all links surviving through time x. He then assumed that the times to failure of the links were independent and identically distributed random variables. Letting Y denote the time to failure of the entire chain, the independence of the individual links yielded the survival function of Y

$$S_n(x) = \Pr\{Y > x\} = S(x)^n. \tag{17}$$

Weibull proposed that writing the cumulative distribution functions of the individual links in the form $F(x) = 1 - \exp\{-\varphi(x)\}$ leads to the advantage of having a simplified survival function for the entire chain. Namely, $S_n(x) = \exp\{-n\varphi(x)\}$. His last step was to specify the function $\varphi(x)$. The function must satisfy the general conditions of being positive, monotone increasing, and vanishing at some value a. Weibull argued that $\varphi(x) = [(x-a)/b]^c$ was the simplest function to satisfy these conditions, and thus arrived at the distribution function of the 3PW,

$$F(x) = 1 - S_n(x) = 1 - \exp\left\{-\left(\frac{x-a}{b}\right)^c\right\}. \tag{18}$$

Weibull applied this weakest link argument to failure in solids, using equation (18) to model the breaking strength data he had obtained through experiments and other sources.

As was mentioned, the hazard function is also widely used in survival analysis, and is defined to be the probability of failure in the next instant of time given survival through time x. For a random variable X, it is mathematically defined by

$$h(x) = \lim_{\Delta x \to 0} \frac{\Pr\{x < X \leq x + \Delta x \mid X > x\}}{\Delta x}. \tag{19}$$

If the density function and survival function of X are given by $f(x)$ and $S(x)$ respectively, then the definition of the conditional probability allows the hazard function to be written as

$$h(x) = \lim_{\Delta x \to 0} \frac{\Pr\{x < X \leq x + \Delta x\}}{\Delta x \Pr\{X > x\}} = \frac{f(x)}{S(x)}. \tag{20}$$

The conditional probability in the numerator of equation (14) should be interpreted as the probability of failing in the time interval Δx given survival through time x. Taking the limit as $\Delta x \to 0$ allows the hazard function to be viewed as the instantaneous failure rate of the device or organism at time x. Using equation (20) allows the hazard rate of the Weibull distribution with parameters a, b, and c to be written as

$$h(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}. \tag{21}$$

Rinne (2009) showed that the hazard rate of the Weibull distribution is a monotone function, meaning it is either non-decreasing (monotone increasing) or non-increasing (monotone decreasing) for a given value of the shape parameter. For $0 < c < 1$, the hazard rate is monotone decreasing. For $c = 1$, the hazard rate has a constant value of $1/b$. For $c > 1$, the hazard rate is monotone increasing.

In the context of survival analysis, a monotone decreasing hazard function can be interpreted as negative aging, meaning that the probability of instantaneous failure gets lower as the lifetime of the device or organism increases. A monotone increasing hazard function indicates the more intuitive property of positive aging, where the probability of instantaneous failure increases with age. A distribution with a constant hazard rate displays what is known as the memoryless property, which states that the instantaneous probability of failure does not depend on the time x. Mathematically, this can be expressed by

$$\Pr\{x < X \leq x + \Delta x \mid X > x\} = \Pr\{X \leq \Delta x\}. \tag{22}$$

The memoryless property is characteristic of the widely used exponential distribution (see Pinsky & Karlin, 2011), which, for $\lambda > 0$, has density function

$$f(x|\lambda) = \lambda \exp\{-\lambda x\}, \quad x \geq 0 \tag{23}$$

and survival function

$$S(x|\lambda) = \exp\{-\lambda x\}, \quad x \geq 0. \tag{24}$$

Applying equations (23) and (24) to equation (20) illustrates the memoryless property of the exponential distribution,

$$h(x) = \frac{f(x)}{S(x)} = \frac{\lambda \exp\{-\lambda x\}}{\exp\{-\lambda x\}} = \lambda. \tag{25}$$

The fact that the exponential distribution and the 2PW with $c = 1$ share this property highlights an important relationship between the two distributions. The density function of a Weibull distribution with parameters $a = 0$, b, and $c = 1$ is given by

$$f(x|a, b) = \frac{1}{b} \exp\left\{-\frac{x}{b}\right\}, \quad x \geq 0, \tag{26}$$

which is identical the density function of an exponential distribution with parameter $1/b$. Because the exponential distribution is a special case of the Weibull distribution, the 3PW can be thought of as a generalization of the exponential by the inclusion of two additional parameters. As another example, the Weibull distribution reduces to the Rayleigh distribution when $c = 2$. Modeling data with a distribution that is a generalization is useful because it can improve goodness of fit by the incorporation of additional parameters while still reducing to the simpler special cases when applicable.

Taking this concept a step further, the Weibull distribution can itself be imbedded into a larger family of probability distributions. Pham and Lai (2007) discussed recent generalizations of the Weibull distribution with applications to survival analysis. They paid particular attention to the different shapes the hazard functions of the generalizations take. As noted previously, the hazard function of the Weibull distribution is monotone, be it either increasing or decreasing. Some of the generalized Weibull distributions presented by Pham and Lai allow for non-monotone shapes. One such example is a generalization that displays a bathtub shaped hazard function where the hazard rate initially decreases, becomes constant, and then increases.

Hirose (1996) looked from the perspective of the generalized extreme value distribution. Maximum likelihood estimation methods were applied to the generalized extreme value distribution of the first order statistic, allowing an estimation procedure to determine which type of extreme value distribution best fits the data.

Stacy (1962) introduced the generalized gamma distribution, which has probability distribution function

$$f(x|\alpha,d,p) = \left(\frac{p}{\alpha^d}\right)\frac{x^{d-1}}{\Gamma(d/p)}\exp\left\{-\left(\frac{x}{\alpha}\right)^p\right\}, \quad x \geq 0 \tag{27}$$

where the parameters α, p, and d are strictly positive. Note that when $p = d$, this reduces to a two parameter Weibull distribution with scale parameter α and shape parameter p. A straightforward transformation yields the shifted generalized gamma function with location parameter a, which reduces to the 3PW when $p = d$. Wingo (1987) presented the maximum likelihood equations for the generalized gamma distribution and discussed a numerical root isolation procedure to estimate its parameters. This method is similar in spirit to the direct search method for the 3PW that will be discussed in Chapter 2.

Stacy and Mihram (1965) generalize the gamma distribution even further, altering the density function in equation (27) to allow for negative values of the parameter p. Gomes, Combes, and Dussauchoy (2008) discussed the estimation of this distribution and supplied references to several methods described in the literature. They went on to establish what they called an easy to implement algorithm to estimate the parameters. Before the estimation procedures are discussed, a survey of some of the applications of the Weibull distribution will be undertaken.

A review of the literature on the Weibull distribution reveals a wide variety of applications that have arisen since its development. As was noted, Weibull himself used the distribution to model breaking strength of materials.

Bartolucci, Singh, Bartolucci, and Bae (1999) used the 3PW to model time to failure in an adjuvant breast cancer therapy. In the study, two groups totaling 41 participants were given separate treatments and times to failure were observed. The authors assumed a Weibull fit for the data was appropriate and employed a method of weighted moments to estimate its parameters. Their estimation procedure will be discussed further in Chapter 2.

Dewanji, Krewski, and Goddard (2011) used a Weibull model to estimate tumorigenic potency in laboratory animals. In their paper, the index of tumorigenic potency was defined to be the "dose of a carcinogen which induces a specified excess tumor response rate following exposure for an extended period encompassing most of the expected lifespan of the animal" (Dewanji et al., 2011, p. 367). This index of potency can be interpreted as a dosage value above which the risk of tumor occurrence increases by a specified amount. They used 3PW random variables to model the time to tumor onset, time to death from tumor, and time to death from competing risks. The stochastic relationships between these three variables were investigated, and an estimation of tumorigenic potency for two carcinogens was performed based on data sets taken from previous studies. References to other work which used the Weibull distribution to model tumorigenic potency and tumor survival data were also provided.

In addition to applications in survival analysis, the Weibull distribution has found use modeling wind speed data. Celik (2003) discussed the wind energy applications of the 2PW and performed a statistical analysis of wind power density measurements taken in a region of southern Turkey using the Weibull and Rayleigh distributions. Both the Weibull and Rayleigh distributions have been used to model wind speed data and Celik provided references for both cases. The goal of the analysis was to determine if the particular location was suitable for a wind farm that could help Turkey meet its increasing energy demands. Continuous wind speed readings over one year were broken up into smaller samples based on month. The monthly data was averaged over hour long periods and stored as hourly figures. Weibull and Rayleigh fits were then applied to the hourly data to yield monthly wind speed distributions. Although Celik concluded that the site displayed poor characteristics for a wind farm location, it was determined that the Weibull distribution provided a better fit to the data than the Rayleigh distribution, displaying evidence for the usefulness of employing a generalized distribution.

Chang (2011) assumed a two-parameter Weibull distribution for wind speed data taken from three wind farm locations in Taiwan. An analysis of six different methods used to estimate the distribution parameters was undertaken using the wind speed data as well as Monte Carlo simulation. Chang (2011) came to the conclusion that the method of maximum likelihood performed best out of the six methods discussed.

As mentioned previously, the Weibull distribution has found use in the field of tornado climatology. Dotzek et al. (2003) gave an overview of tornado intensity based on the Fujita and

Beaufort scales developed by Tetsuya Fujita and Francis Beaufort respectively (see Meaden et al., 2005). The Fujita scale – and its updated counterpart, the Enhanced Fujita scale – give a tornadoes an integer rating from zero to five based on maximum wind speed, damage, and several other factors. A discussion of rating practices and the enhanced scale can be found in Storm Prediction Center (2011). Roughly speaking, an F0 tornado would be associated with wind speeds beginning around 18 miles per hour while an F5 would entail wind speeds in excess of 100 miles per hour.

Additionally, Dotzek et al. (2003) addressed the issue of determining a general shape for observed tornado intensity distributions. They noted that while Fujita scale, or F-scale, intensity distributions were usually modeled by the exponential distribution, the observed frequencies displayed a curvature to the right not captured by the exponential. The authors argued that this issue can be partly explained by the historic underreporting of weaker tornadoes as well as an apparent upper wind speed. Due to these effects, they proposed using a Weibull distribution to model the data. They also proposed extending the F-scale to take the additional subcritical values of F-1and F-2 to achieve the natural physical boundary condition of zero tornadoes with zero mile per hour wind speed. However, there are few reports for such subcritical tornadoes, making it difficult to obtain a meaningful sample. Therefore, the authors treated the number of subcritical tornadoes as a free parameter in their fitting procedure.

Feuerstein, Dotzek, and Grieser (2005) further developed the work of Dotzek et al. (2003) by attempting to account for the underreporting of weak tornadoes. This was achieved by including the number of F0 as well as subcritical tornadoes as a free parameter in the fitting procedure. They also investigated an apparent relationship between the scale and shape parameters for the Weibull fit of the F-scale data and attempted to explain it based on climatological arguments.

Both Dotzek et al. (2003) and Feuerstein et al. (2005) applied fits to the F-scale extended down to F-2. In their fits, the lower limit of the distribution was known. This translates to the case of having a known location parameter (in particular, $a = -2$). In this thesis, subcritical tornadoes will not be addressed. Instead, the F-scale beginning at F0 will be employed, which yields a zero location parameter. For either approach, it should be noted that a 2PW model will suffice. Furthermore, the curvature to the right reported by Dotzek et al. (2003) indicates a value of the shape parameter that would be greater than one.

CHAPTER 2

METHODS

There exist a large variety of methods which can be employed to estimate the parameters of a distribution given a particular sample. This is particularly true for the Weibull distribution. In fact, upon reviewing the literature on Weibull statistics, much of what has been written deals with estimation and estimator properties in some form. Each of the methods proposed has its own merits, and a method should be chosen based on the characteristics of the data being analyzed and the goals of the analysis. In this chapter, the method of percentiles, method of moments, and method of maximum likelihood will be discussed. An overview of the first two methods will be given, with the two parameter case illustrated in particular. The method of maximum likelihood will be developed in detail for both the two and three parameter cases. Before the methods are discussed, some mathematical terminology necessary for the discussion of estimators will be presented.

Estimators for a given parameter may be derived through different approaches. In an effort to distinguish the estimators and attempt to determine which is the most suitable for a particular application, several properties of estimators have been developed. Of interest in this thesis are the concepts of consistency, asymptotic normality, and asymptotic efficiency.

A sequence of estimators is said to be *consistent* if it converges in probability to the actual parameter. Mathematically, if $\boldsymbol{\theta}$ is the true parameter and $\widehat{\boldsymbol{\theta}}_n$ a consistent sequence of estimators, then for all $\varepsilon > 0$, $\lim_{n\to\infty} \Pr\{|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}| < \varepsilon\} = 1$. It is common for the sequence of estimators to be indexed by sample size. If this is the case, a consistent estimator becomes more likely to be close to the true parameter as the sample size increases. As an example, repeating measurements in an experiment would be a straightforward way to increase the sample size and thus increase the probability of accuracy for a consistent estimator.

An estimator is *asymptotically normal* if it is consistent and its asymptotic distribution around the true parameter value is normal. Complementary to the property of asymptotic normality is the concept of estimator bias. An estimator is said to be *biased* if its expected value is not equal to the true parameter value, i.e., $E(\widehat{\boldsymbol{\theta}}) \neq \boldsymbol{\theta}$. Therefore, an estimator that is asymptotically normal is said to be *asymptotically unbiased* because the expected value approaches the actual value as n increases. However, asymptotically unbiased estimators are still typically biased for small samples sizes.

The last estimator property to be discussed is asymptotic efficiency. An estimator is *asymptotically efficient* if it is asymptotically normal with variance that achieves the lower bound given by the Cramér-Rao inequality (see Hogg, McKean, and Craig, 2005), therefore having smaller variance than all other asymptotically normal estimators. Asymptotic efficiency guarantees the smallest possible variation in estimates, meaning the asymptotic distribution of the estimator is distributed as close as possible around the true parameter value.

Now that some properties of estimators have been discussed, notation for this section will be briefly mentioned. The random variables $X_1, X_2, \ldots, X_n$ will denote n independent and identically distributed Weibull random variables with true parameter $\boldsymbol{\theta}$. Explicitly, for the 3PW,

$\boldsymbol{\theta} = (a, b, c)$ and $\widehat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c})$. For the 2PW, $\boldsymbol{\theta} = (b, c)$ and $\widehat{\boldsymbol{\theta}} = (\hat{b}, \hat{c})$. The set of values $\{x_1, x_2, \dots, x_n\}$ will denote a realization of the random variables $X_1, X_2, \dots, X_n$.

PERCENTILE ESTIMATION

The first method of estimation to be discussed is percentile estimation, which uses the concept of percentiles to derive relationships between the parameters based on a given sample. For $0 < p < 1$, the $100p^{\text{th}}$ percentile of a random variable X is defined to be a value x_p such that $\Pr\{X \leq x_p\} \leq p$ and $\Pr\{X \geq x_p\} \leq 1 - p$. As an example, the $100(0.95)^{\text{th}}$ percentile of a random variable X would be the value $x_{0.95}$ such that $\Pr\{X < x_{0.95}\} \leq 0.95$ and $\Pr\{X \leq x_{0.95}\} \geq 0.95$, and is commonly used to determine the significance of statistical results.

For the 2PW random variable X, the $100p^{\text{th}}$ percentile can be obtained by inverting equation (3), yielding

$$x_p = b[-\ln(1 - p)]^{1/c}. \tag{28}$$

By equation (28), the $100(1 - e^{-1}) \cong 63.2^{\text{th}}$ percentile is then

$$x_{0.632} = b[-\ln e^{-1}]^{1/c} = b[1]^{1/c} = b. \tag{29}$$

Taking the ratio of equations (28) and (29) gives an expression that is independent of the scale parameter and can be solved for c, leading to the percentile estimate for the shape parameter

$$\hat{c} = \frac{\ln[-\ln(1-p)]}{\ln\left(x_p/x_{0.632}\right)}. \tag{30}$$

Equation (30) holds for $0 < x_p < x_{0.632}$. The percentiles of the given sample can be calculated and then substituted into equations (29) and (30) to obtain the percentile estimates for the 2PW. Wang and Keats (1995) employed Monte Carlo simulations over a wide range of data samples with different parameter values to find a value of $p = 0.15$ which minimized the bias of $\hat{c}$ over a wide range of parameter values and samples.

For the 3PW, Schmid (1997) introduced a percentile estimator for the scale parameter and summarized the work of Dubey (1967) and Zanakis (1979) who introduced percentile estimators for the location and shape parameters respectively. Schmid (1997) went on to examine the asymptotic behavior of each of the estimators, finding that they are consistent as well as asymptotically normal.

ESTIMATORS USING MOMENTS

The second estimation procedure to be discussed is the method of moments. The method of moments uses the equations for the moments of the distribution in question to derive relationships between the parameters and the sample to which the fit is being applied. These moment estimators have the appealing properties of being consistent as well as asymptotically normal (see Rinne, 2009). For the 2PW, two different moments must be used since there are two unknown parameters. The first two moments of the 2PW, obtained from equations (12) and (13) by letting $a = 0$, are

$$E(X) = b\Gamma\left(\frac{1}{c} + 1\right) \tag{31}$$

and

$$E(X^2) = b^2\Gamma\left(\frac{2}{c} + 1\right). \tag{32}$$

Therefore, the ratio

$$\frac{[E(X)]^2}{E(X^2)} = \frac{\left[\Gamma\left(\frac{1}{c} + 1\right)\right]^2}{\Gamma\left(\frac{2}{c} + 1\right)} \tag{33}$$

is a function of the shape parameter only. The first two moments of the sample can be substituted into equation (33), yielding an equation which can be solved for the moment estimator $\hat{c}$. This equation can be solved using graphical methods, an application of the Newton-Raphson method (see Rinne, 2009), or other techniques. Once the estimate for the shape parameter has been solved for, it can be inserted into equation (31), along with the sample mean, to achieve the moment estimate for the scale parameter. Rinne (2009) went on to use Monte Carlo simulation to reveal a negative bias in the estimate of the shape parameter, meaning estimates was typically smaller than the true parameter value. It was also found that the bias tends to decrease as the sample size increases. Methods using moments other than the first two have been investigated by other authors, as well as modified methods developed in an attempt to reduce the bias.

For the 3PW, an additional moment must be considered as there are now three parameters to be solved for. Using the first two moments as well as the coefficient of skewness, α_3, defined by

$$\alpha_3 = \frac{E[(X-\mu)^3]}{\{E[(X-\mu)^2]\}^{3/2}} \tag{34}$$

yields an equation independent of a and b which can be solved for the moment estimator $\hat{c}$ using iterative techniques. Once $\hat{c}$ has been determined, the other two estimators can be found in a straightforward manner. Bartolucci et al. (1999) discussed a method of probability-weighted moments in which the moments are weighted by terms involving the cumulative distribution function with the goal of reducing the bias of the estimates. Although the parameters of the 3PW cannot be expressed in terms of the moments about zero, they can be written as functions of the probability-weighted moments. Bartolucci et al. (1999) presented the equations for the estimates of the three parameters, and referenced consistent estimators of the weighted moments based on order statistics. They then applied the probability-weighted moment and maximum likelihood techniques to a sample of medical survival data and found that although both give reasonable estimates, the latter method is more accurate. They concluded that the probability-weighted moment method is most useful to provide initial estimates for the method of maximum likelihood.

MAXIMUM LIKELIHOOD ESTIMATORS

The method of maximum likelihood estimation seeks to determine the parameter $\widehat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c})$ that maximizes the likelihood function of the sample. This parameter is called the maximum likelihood estimator (MLE). The likelihood function is defined by

$$L(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{35}$$

and can be interpreted as the likelihood a distribution has parameter $\boldsymbol{\theta}$ given a particular sample. Before the method is discussed, some properties of the MLE are mentioned.

One useful property of the MLE, $\widehat{\boldsymbol{\theta}}$, of a distribution is functional invariance, meaning $f(\widehat{\boldsymbol{\theta}})$ is the MLE for $f(\boldsymbol{\theta})$ for some function f. Furthermore, under certain regularity conditions (Hogg, McKean, & Craig, 2005, Chapter 6), the MLE is a consistent, asymptotically normal, and asymptotically efficient estimator of the actual parameter, $\boldsymbol{\theta}$. Rinne (2009) gives an overview of imposing the regularity conditions on the Weibull distribution and reports that they are satisfied for the 2PW. One of the regularity conditions is that the density functions of the random variables must have common support. The 3PW violates this condition because the location parameter represents the lower bound of the support of the density function. Therefore, the MLE for the 3PW is not guaranteed to behave nicely like that of the 2PW. The derivation of the method of maximum likelihood for the 2PW will now be discussed.

The likelihood function for the 2PW is given by

$$L(\boldsymbol{\theta}) = \left(\frac{c}{b}\right)^{n} \prod_{i=1}^{n} \left(\frac{x_i}{b}\right)^{c-1} \exp\left\{-\sum_{i=1}^{n} \left(\frac{x_i}{b}\right)^{c}\right\}. \tag{36}$$

Due to the complexity of the previous equation, it is more convenient to maximize the natural logarithm of $L(\boldsymbol{\theta})$. Since the natural logarithm is a monotone function, the maximum of the log-likelihood function corresponds with the maximum of $L(\boldsymbol{\theta})$. The log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = n(\ln c - c \ln b) + (c-1)\sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \left(\frac{x_i}{b}\right)^{c}. \tag{37}$$

The first partials of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are then

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial b} = -\frac{nc}{b} + \frac{c}{b}\sum_{i=1}^{n} \left(\frac{x_i}{b}\right)^{c} \tag{38}$$

and

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial c} = \frac{n}{c} - nln(b) + \sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \left(\frac{x_i}{b}\right)^{c} \ln\left(\frac{x_i}{b}\right). \tag{39}$$

Setting equations (35) and (36) equal to zero with some rearrangement yields the equations

$$\hat{b} = \left[\frac{1}{n}\sum_{i=1}^{n} {x_i}^{\hat{c}}\right]^{1/\hat{c}} \tag{40}$$

and

$$\frac{1}{\hat{c}} + \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) - \frac{\sum_{i=1}^{n}(x_i)^{\hat{c}}\ln(x_i)}{\sum_{i=1}^{n}(x_i)^{\hat{c}}} = 0. \tag{41}$$

If the MLE of the shape parameter is known, functional invariance implies that $\hat{b}$ is the MLE of the scale parameter. Estimating the shape parameter is less straightforward because equation (36) is a non-linear equation which cannot be solved directly for $\hat{c}$. Instead, an iterative procedure can be employed. The solution to equation (41) is unique (see Rinne, 2009) and the equation satisfies the requirements for convergence of the Newton-Raphson method from any starting value (see Gupta, Gupta, & Lvin,1998). Therefore, the Newton-Raphson method has been chosen to obtain a solution to equation (41). It is concluded that the MLE is a superior choice of estimator for the 2PW due to the appealing asymptotic properties described and the relative ease of estimation. This sentiment is mirrored by Chang (2011), who concluded that the maximum likelihood approach outperformed other methods when applied to simulated samples as well as actual data.

The method of maximum likelihood applied to the 3PW is complicated by the need to estimate an additional parameter as well as the failure to satisfy the regularity conditions. For the 3PW, the log-likelihood equation is given by

$$\mathcal{L}(\boldsymbol{\theta}) = n(\ln c - c\ln b) + (c-1)\sum_{i=1}^{n}\ln(x_i - a) - \sum_{i=1}^{n}\left(\frac{x_i - a}{b}\right)^{c}. \tag{42}$$

Taking the first partial derivative of $\mathcal{L}(\boldsymbol{\theta})$ with respect to a yields

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial a} = (1-c)\sum_{i=1}^{n}\frac{1}{(x_i - a)} + \frac{c}{b}\sum_{i=1}^{n}(\frac{x_i - a}{b})^{c-1}. \tag{43}$$

Taking the first partial derivative of $\mathcal{L}(\boldsymbol{\theta})$ with respect to b yields

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial b} = -\frac{nc}{b} + \frac{c}{b}\sum_{i=1}^{n}\left(\frac{x_i - a}{b}\right)^{c}. \tag{44}$$

Finally, taking the first partial derivative of $\mathcal{L}(\boldsymbol{\theta})$ with respect to c yields

$$\frac{\partial\mathcal{L}(\theta)}{\partial c} = \frac{n}{c} - n\ln b + \sum_{i=1}^{n}\ln(x_i - a) - \sum_{i=1}^{n}\left(\frac{x_i - a}{b}\right)^{c}\ln\left(\frac{x_i - a}{b}\right). \tag{45}$$

Setting equations (38), (39), and (40) equal to zero yields the equations

$$(1-\hat{c})\sum_{i=1}^{n}\frac{1}{(x_i - \hat{a})} + \frac{\hat{c}}{\hat{b}}\sum_{i=1}^{n}(\frac{x_i - \hat{a}}{\hat{b}})^{\hat{c}-1} = 0, \tag{46}$$

$$-\frac{n\hat{c}}{\hat{b}}+\frac{\hat{c}}{\hat{b}}\sum_{i=1}^{n}\left(\frac{x_i-\hat{a}}{\hat{b}}\right)^{\hat{c}}=0, \tag{47}$$

and

$$\frac{n}{\hat{c}}-nln\left(\hat{b}\right)+\sum_{i=1}^{n}\ln(x_i-\hat{a})-\sum_{i=1}^{n}\left(\frac{x_i-\hat{a}}{\hat{b}}\right)^{\hat{c}}\ln\left(\frac{x_i-\hat{a}}{\hat{b}}\right)=0. \tag{48}$$

From equation (47), an estimator for b can be given explicitly in terms of $\hat{a}$ and $\hat{c}$ by

$$\hat{b}=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{a})^{\hat{c}}\right]^{1/\hat{c}}. \tag{49}$$

Again, due to functional invariance, $\hat{b}$ is indeed the MLE for b. Substituting this value for $\hat{b}$ into equations (41) and (43) yields the two equations

$$\frac{\hat{c}-1}{\hat{c}}\sum_{i=1}^{n}\frac{1}{(x_i-\hat{a})}-n\frac{\sum_{i=1}^{n}(x_i-\hat{a})^{\hat{c}-1}}{\sum_{i=1}^{n}(x_i-\hat{a})^{\hat{c}}}=0 \tag{50}$$

and

$$\frac{1}{\hat{c}}+\frac{1}{n}\sum_{i=1}^{n}\ln(x_i-\hat{a})-\frac{\sum_{i=1}^{n}(x_i-\hat{a})^{\hat{c}}\ln(x_i-\hat{a})}{\sum_{i=1}^{n}(x_i-\hat{a})^{\hat{c}}}=0. \tag{51}$$

Equations (50) and (51) represent a system of highly non-linear, coupled equations whose solution by iterative techniques is "usually considered a non-trivial problem because of its complexity" (Rinne, 2009, p. 416). Instead of using iterative techniques, a direct search method will be employed to look for the solutions. A form if this method, which is effective for non-negative location parameter, is presented by Rinne (2009). As cited by Rinne (2009), the development of this method relies mainly on the work of Panchang and Gupta (1989). In it, the minimum of the sample, x_{min}, is taken to be the natural upper bound of the location parameter. The interval $[0, x_{min})$ is partitioned into N subintervals of length Δa. The partition values are then $a_i = (i-1)\Delta a$ for $i = 1, \ldots, N+1$. For a particular a_i, the MLE for the 2PW is solved for the sample $\{x_i - a_i\}$, yielding the estimates b_i and c_i. The log-likelihood function is then evaluated at $\boldsymbol{\theta}_i = (a_i, b_i, c_i)$ for each i. The MLE is then taken to be $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_j$ such that $\mathcal{L}(\boldsymbol{\theta}_j) \geq \mathcal{L}(\boldsymbol{\theta}_i)$ for $i = 1, \ldots, N+1$. Although this method is not computationally efficient, it avoids the difficulties encountered when employing an iterative technique. To lessen unnecessary computation, a relatively small value of N can initially be chosen to determine the general location of the maximum. Once it is known, a larger value can be chosen to more accurately determine the location of the maximum in the restricted domain.

It should be noted that, in general, the solutions to the equation

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \tag{52}$$

are extreme values of the log-likelihood and not necessarily maxima. To address this issue for the 2PW, it will be shown that the second partials of the log-likelihood function evaluated at the relevant values are negative, insuring the maximization of the function.

Taking the second partial of equation (37) with respect to b yields the expression

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial b^2} = \frac{nc}{b}\left[1 - (1 + c)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{b}\right)^c\right]. \tag{53}$$

Evaluating equation (53) at $\hat{b}$ results in a value of $-c$. Since the shape parameter must be a positive quantity, the second partial evaluated at $\hat{b}$ is negative. Therefore, the likelihood function achieves its maximum with respect to the scale parameter at $\hat{b}$.

Taking the second partial of equation (37) with respect to c and evaluating the resulting expression at $\hat{b}$ yields

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial c^2} = \left(\sum x_i{}^c\right)\left(\sum \log x_i\right)^2 - 2\left(\sum x_i{}^c \log x_i\right)\left(\sum \log x_i\right) + \sum x_i{}^c (\log x_i)^2. \tag{54}$$

After lengthy algebraic manipulations not repeated here, it was determined that equation (54) is indeed negative, and so $\hat{c}$ is indeed the MLE of the scale parameter of the 2PW.

For the three parameter case, a numerical test to ensure the second partial of equation (42) evaluated at the estimated value is indeed negative can be easily included in the last steps of the procedure.

In practice, modeling a process or phenomenon with the 2PW is more appealing than using the 3PW because of the asymptotic properties of the two-parameter MLE and the computational ease of implementation. Therefore, before a three-parameter model is imposed on a given sample, it would be convenient to tell if the data set could suitably be modeled by the 2PW. McCool (1998) considered this point and proposed a hypothesis test to determine if a non-zero location parameter is appropriate to model a given set of data. Determining a Weibull model with zero location parameter was suitable before performing a fit could save time and computational complexity, as well as allow for the discussion of the asymptotic properties of the estimators.

ESTIMATOR BIAS

As discussed, the MLE for the 2PW is asymptotically normal. A consequence is that the bias of the estimates will approach zero as the sample size increases. However, in many applications, it is not cost effective or even possible to obtain more data. In such cases, it would be beneficial to attempt to describe the small sample bias through either numerical or analytic arguments.

Ross (1994) assumed an asymptotic behavior for the estimate of the shape parameter to determine an unbiasing factor. Using a very large number of Monte Carlo simulations over a wide range of parameters, Ross empirically derived a factor that, when multiplied by the original estimate, could typically reduce the relative bias of the estimates to less than 0.05%. Hirose (1999) proposed that the bias could be written as a power series in $1/n$ due to the fact it must converge to zero as n increases indefinitely. Like Ross (1994), Hirose(1999) used a large number of Monte Carlo simulations to obtain the coefficients for this presumed expansion. The coefficients were then listed in a series of tables for convenient reference. Montanari, Mazzanti, Cacciari, and Fothergill (1997) gave a review of the existing numerical bias correction methods, and found that the Ross unbiasing factor was highly effective.

Analytically estimating the bias of the MLE for the 2PW is complicated by the transcendental nature of equations (50) and (51). One way to achieve an estimate for the bias is to use arguments related to the consistency of the MLE. If convergence in probability of $\hat{b}$ to the true scale parameter is assumed to be carried through, the expression $\hat{c}$ can be reduced to

$$\frac{1}{\hat{c}} = \frac{\sum_{i=1}^{n} x_i{}^{\hat{c}} \ln x_i - \frac{1}{n}\sum_{j=1}^{n} \ln x_j \sum_{i=1}^{n} x_i{}^{\hat{c}}}{\beta}, \tag{55}$$

where $\beta = nb^c$. Taking the expectation of $\hat{c}$ would result in an integral with complicated sums in the denominator. To avoid this, the expectation of $1/\hat{c}$ will be investigated instead. Taking the expectation of $1/\hat{c}$ leads to an expression of the form

$$E\left(\frac{1}{\hat{c}}\right) \sim \frac{n}{\beta} I_1 - \frac{1}{\beta} I_2 \tag{56}$$

where

$$I_1 = \frac{c}{b^c} \int_0^\infty x^{2c-1} \exp\left\{-\left(\frac{x}{b}\right)^c\right\} \log x \, dx \tag{57}$$

and

$$I_2 = \iint_0^\infty \left(\frac{c}{b}\right)^2 \left(\frac{x}{b}\right)^{c-1} \left(\frac{y}{b}\right)^{c-1} \exp\left\{-\left(\frac{x}{b}\right)^c\right\} \exp\left\{-\left(\frac{y}{b}\right)^c\right\} x^c \log y \, dx \, dy. \tag{58}$$

Applying the substitution $u = (x/b)^c$ to I_1 results in the expression

$$I_1 = b^c \log b \int_0^\infty u \exp\{-u\} \; du + \frac{b^c}{c} \int_0^\infty u \exp\{-u\} \log u \; du. \tag{59}$$

The first integral is the gamma function evaluated at the value one, and is equal to one. The second integral is of a form documented by Gradshteyn and Ryzhik (1965). The authors provide the relation

$$\int_0^\infty x^{\nu-1} \exp\{-\mu x\} \log x \, dx = \frac{1}{\mu^\nu} \Gamma(\nu)[\psi(\nu) - \log \mu], \tag{60}$$

which leads to an expression for the first integral given by

$$I_1 = b^c \log b + \frac{b^c}{c} \; \psi(2). \tag{61}$$

Applying the substitutions $u = (x/b)^c$ and $v = (y/b)^c$ to I_2 results in

$$I_2 = b^c \log b \; \iint_0^\infty u \exp\{-(u+v)\} \; du \, dv + \frac{b^c}{c} \iint_0^\infty u \exp\{-(u+v)\} \log v \; du \, dv. \tag{62}$$

The first double integral of I_2 also integrates to one, and the second double integral can also be evaluated by equation (60). This leads to the equation

$$I_2 = b^c \log b + \frac{b^c}{c}\psi(1). \tag{63}$$

An estimate for the expectation of $1/\hat{c}$ in terms of the true parameters can then be obtained from equation (56) yielding

$$E\left(\frac{1}{\hat{c}}\right) \sim \frac{n}{\beta}\ b^c \log b + \frac{n}{\beta}\frac{b^c}{c}\ \psi(2) - \frac{1}{\beta}\ b^c \log b - \frac{1}{\beta}\frac{b^c}{c}\psi(1). \tag{64}$$

Substituting $\beta = nb^c$ back into equation (64) yields

$$E\left(\frac{1}{\hat{c}}\right) \sim \log b + \frac{1}{c}\ \psi(2) - \frac{1}{n}\ \log b - \frac{1}{nc}\psi(1). \tag{65}$$

Therefore, an estimate for the bias is found to be

$$E\left(\frac{1}{\hat{c}}\right) - \frac{1}{c} = \log b\ \left(1 - \frac{1}{n}\right) + \frac{1}{c}\left(\psi(2) - \frac{1}{n}\psi(1) - 1\right). \tag{66}$$

Based on this equation, it appears that the bias of $1/\hat{c}$ will increase as the scale parameter increases and decrease as the shape parameter increases.

CHAPTER 3

RESULTS

In order to judge the effectiveness of a particular estimation method, it is beneficial to have a data sample taken from a distribution with known parameters. This can be achieved by employing a form of Monte Carlo simulation where numbers are sampled randomly from a uniform distribution then transformed into random numbers from the desired distribution. In application, it is exceedingly difficult to attain truly random numbers. However, use of pseudo-random numbers, i.e. numbers that share statistical properties with random numbers but are actually generated by a deterministic process, will suffice. The cumulative distribution function of the Weibull distribution can be inverted, yielding the equation

$$x = a + b\left[\ln\left(\frac{1}{1-F(x)}\right)\right]^{1/c}. \tag{67}$$

If pseudo-random numbers from a uniform distribution on the interval $(0,1)$ are substituted for $F(x)$, the resulting value of x will be a pseudo-random number from a Weibull distribution with parameters a, b, and c.

SIMULATED DATA

As mentioned previously, a benefit of the asymptotic normality of the MLE for the 2PW is asymptotic unbiasedness. Therefore, taking more data in an experiment or study would directly increase the accuracy of the estimates. To see this effect illustrated, the Monte Carlo method described above was used to generate 2PW data for various parameter values and sample sizes. For each data set, a scale parameter value of $b = 2$ was chosen. The scale parameter was not varied systematically because its MLE is given in closed form in terms of the estimator of the shape parameter. Scale parameter values were taken to be 3, 7, and 12. For each set of parameters, samples of size 16, 32, 64, and 128 were generated. The Newton-Raphson method with error tolerance 10^{-4} was applied to each sample to obtain the MLE estimate $\hat{c}$. The estimate $\hat{b}$ was then calculated using equation (40). This process was repeated 1,000 times and the average of the resulting estimates for each sample computed. Tables (1) through (3) give the average parameter estimates for shape parameters $c = 3, 7,$ and 12 respectively. The biases listed are the percent differences of the average MLE and the actual parameter, and can therefore be interpreted as the relative bias of the estimators. Note that the relative bias decreases as sample size increases. As noted, this is a consequence of the asymptotic normality of the MLE in the two-parameter case.

Table 1: *Estimates of the 2PW data for* $b = 2, c = 3$

n	$\hat{b}$	$B(\hat{b})$	$\hat{c}$	$B(\hat{c})$
16	1.9902	0.0049	3.3204	0.1068
32	1.9943	0.0029	3.1265	0.0422
64	1.9964	0.0018	3.0614	0.0205
128	1.9967	0.0017	3.0204	0.0068

Table 2: *Estimates of the 2PW data for* $b = 2, c = 7$

n	$\hat{b}$	$B(\hat{b})$	$\hat{c}$	$B(\hat{c})$
16	1.9912	0.0044	7.7327	0.1047
32	1.9983	0.0009	7.376	0.0537
64	1.9984	0.0008	7.1768	0.0253
128	1.9993	0.0004	7.0755	0.0108

Table 3: *Estimates of the 2PW data for* $b = 2$, $c = 12$

n	$\hat{b}$	$B(\hat{b})$	$\hat{c}$	$B(\hat{c})$
16	1.9933	0.0034	13.0853	0.0904
32	1.9984	0.0008	12.5444	0.0454
64	1.9992	0.0004	12.2335	0.0195
128	1.9993	0.0004	12.1178	0.0098

The unbiasing factor for c described by Ross (1994) and the Hirose (1999) method to unbias b have been applied to the MLE of a sample of size five from the 2PW with parameters $b = 4$ and varying c. The unbiased values of $\hat{b}$ and $\hat{c}$ will be denoted $\hat{b}_u$ and $\hat{c}_u$ respectively. Table (4) shows the effect of the unbiasing procedure. The unbiasing of $\hat{c}$ by Ross's factor is particularly effective, as it reduces the relative bias by over 40% for each value of c. Not only do the Ross unbiasing factor and the Hirose method work well, they are easy to implement in the last stages of the iterative estimation procedure.

Table 4: *Biased and Unbiased Estimates for* $b = 4$, $n = 5$, and varying c

c	$\hat{b}$	$B(\hat{b})$	$\hat{b}_u$	$B(\hat{b}_u)$	$\hat{c}$	$B(\hat{c})$	$\hat{c}_u$	$B(\hat{c}_u)$
1	4.0453	0.0113	4.0234	0.0059	1.4345	0.4345	0.9962	0.0038
4	3.9696	0.0076	3.9837	0.0041	5.9231	0.4808	4.1133	0.0283
8	3.9729	0.0068	3.9816	0.0046	11.3841	0.4230	7.9056	0.0118

The direct search method detailed in the previous chapter has been employed to estimate the MLE for the 3PW. As cited by Rinne (2009), Panchang and Gupta (1989) determined a good initial estimate for c was $c_o = 1/(n \ln x_{max} - \sum \ln x_i)$, and this value was used in the present estimation. Monte Carlo simulation was used to generate 3PW data with parameters $a = 7$, $b = 2$, and varying c. The direct search method was then applied to obtain the MLE. This process was repeated 1,000 times and the average MLE value computed. The results are given in Tables (5) through (7). The estimator bias still tends to decrease as the sample size increases, but its behavior is more erratic than with the 2PW due to the lack of asymptotic efficiency. For all values of c, the estimates for the location parameter are quite accurate. In particular, for shape parameter values of 0.5 and 1.5, the relative bias of the location parameter estimates is less than one percent even for the smaller sample sizes.

Table 5: *Estimates for the 3PW with* $a = 7$, $b = 2$ *and varying* c

c	n	$\hat{a}$	$B(\hat{a})$	$\hat{b}$	$B(\hat{b})$	$\hat{c}$	$B(\hat{c})$
0.5	16	7.0130	0.0019	2.0915	0.0458	0.5224	0.0448
	32	7.0026	0.0004	2.0887	0.0444	0.5198	0.0396
	64	7.0001	0.0001	2.0925	0.0463	0.5185	0.0370
	128	6.9992	0.0001	2.0661	0.0331	0.5148	0.0296
1.5	16	7.0652	0.0093	1.8127	0.0937	1.4137	0.0575
	32	7.0933	0.0133	1.8434	0.0783	1.3601	0.0933
	64	7.0622	0.0089	1.8943	0.0529	1.3852	0.0765
	128	7.0367	0.0052	1.9557	0.0222	1.4348	0.0435
2	16	6.9578	0.0060	1.9551	0.0225	2.0492	0.0246
	32	7.0692	0.0099	1.8953	0.0524	1.8744	0.0628
	64	7.0718	0.0103	1.9061	0.0470	1.8370	0.0815
	128	7.0501	0.0072	1.9345	0.0328	1.8531	0.0735
7	16	5.3071	0.2418	3.6716	0.8358	13.403	0.9147
	32	5.7735	0.1752	3.2197	0.6099	11.5763	0.6538
	64	6.5848	0.0593	2.4098	0.2049	7.5315	0.0759
	128	6.6968	0.0433	2.2995	0.1498	7.5262	0.0752

TORNADO INTENSITY DATA

Tornado F-scale data has been obtained from National Oceanic and Atmospheric Administration's Storm Prediction Center (see Storm Prediction Center, 2012). Three samples in particular have been analyzed: the April 25th – 28th, 2011 tornado outbreak, the historic record of tornadoes in Alabama, and the national historic tornado record. As argued by Dotzek et al. (2003), reporting practices and historical bias have made the record of F0 tornadoes somewhat unrealistic. Therefore, the fits have been applied to only positive F-scale ratings.

The sample for the April 25th – 28th, 2011 tornado outbreak consists of the preliminary ratings of 234 tornadoes rated F1 through F5. The MLE method for the 2PW was applied with error tolerance of 10^{-4} and unbiasing, yielding estimates of $b = 1.8313$ and $c = 1.8499$ with $\chi^2 = 0.3744$. Consulting a table of Chi-squared values for three degrees of freedom in Hogg et al. (2005) showed that this is less than the critical value of 7.815 at 5% significance. Therefore, the Weibull fit is an appropriate model as judged by the Chi-squared test. Figure 1 is a bar graph of the frequency of tornadoes by F-scale rating. The solid line represents the density function of the estimated Weibull distribution.

The historic record of tornadoes in Alabama includes 1,144 tornadoes rated with positive F-scale dating back to the 1950. The sample size is large enough for the asymptotic normality of the MLE to ensure a small bias effective, and will therefore not be corrected for. The estimation procedure yields estimates $c = 2.1320$ and $b = 1.8584$, with $\chi^2 = 0.3204$. Again, the Weibull model is determined to be appropriate at 5% significance. Figure 2 is a bar graph of the frequency of tornadoes by F-scale rating. The solid line represents the density function of the estimated Weibull distribution.

Figure 1: *F-scale Values for April 25th – 28th, 2011 with Weibull fit*

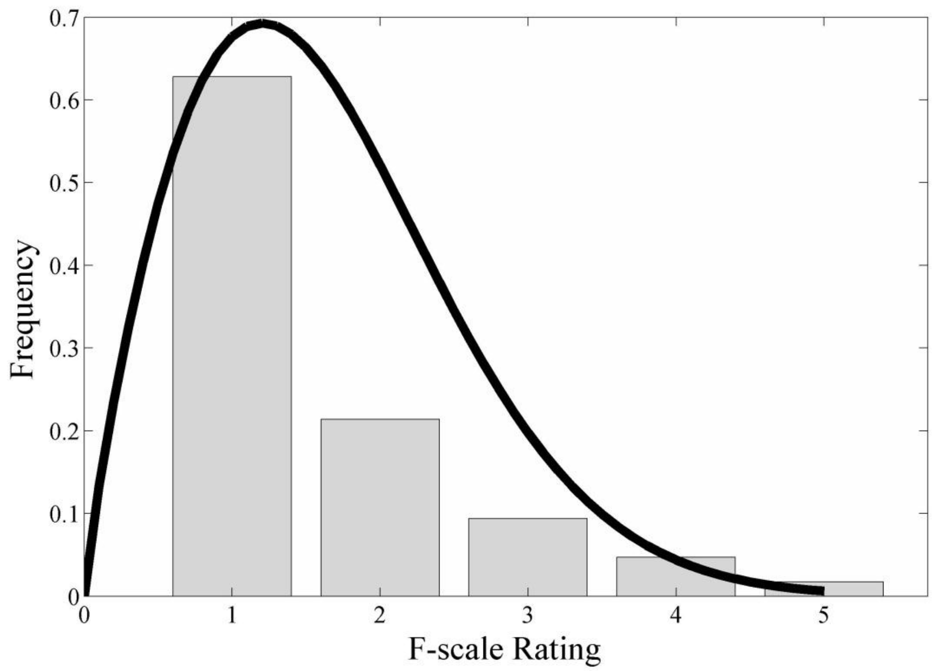
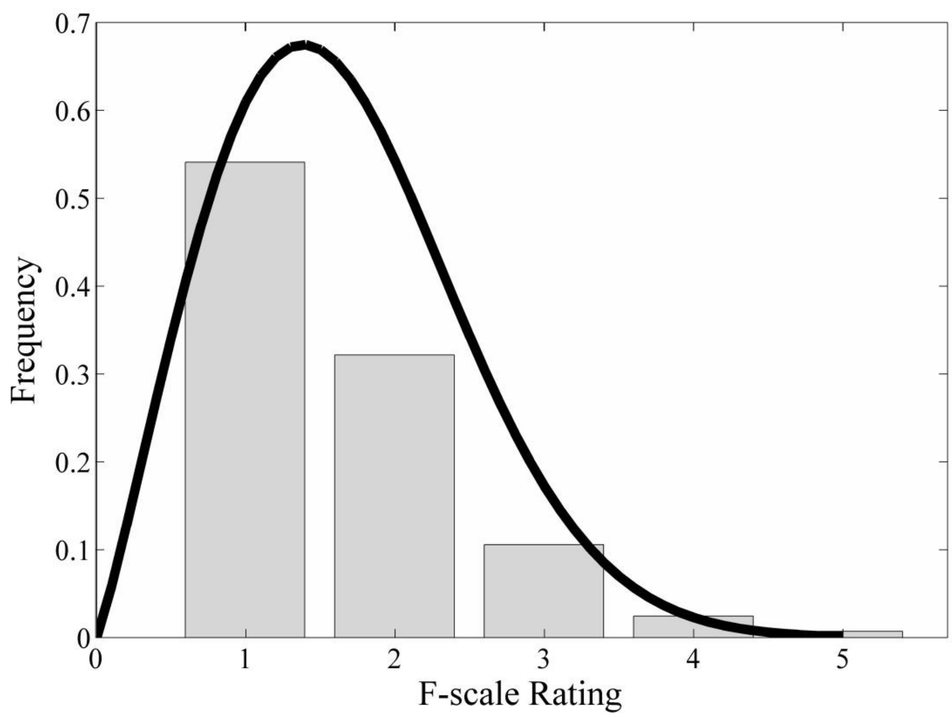
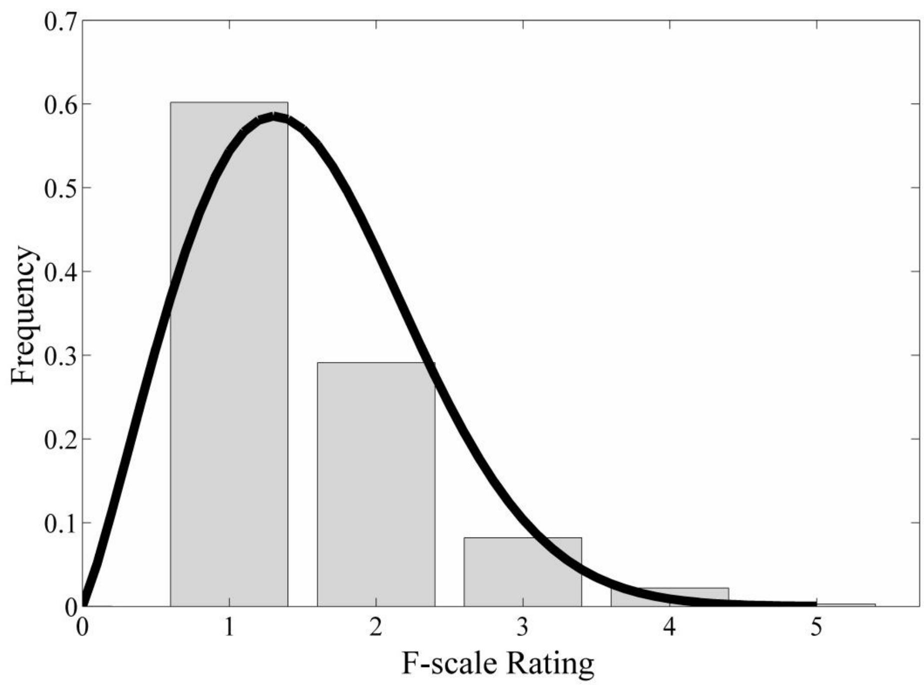


Figure 2: *Historic F-scale Values for the Alabama Record with Weibull fit*



The historic record of tornadoes in entire United States includes 29,609 tornadoes with positive F-scale rating dating back to 1950. At this sample size, convergence to the asymptotic distribution can certainly be assumed, and so bias will not be corrected for. The estimation procedure yields estimates $c = 2.1578$ and $b = 1.7416$ with $\chi^2 = 0.3148$. Therefore, the Weibull model is again determined to be appropriate at 5% significance. Figure 3 is a bar graph of the frequency of tornadoes by F-scale rating. The solid line represents the density function of the estimated Weibull distribution.

Figure 3: *Historic F-scale Values for the U.S. Record with Weibull fit*



CHAPTER 4

DISCUSSION

An overview of Weibull estimation was undertaken with a focus on the method of maximum likelihood. The maximum likelihood method for the two parameter Weibull distribution was shown to be the best choice of estimator based on the asymptotic properties discussed. An equation to analytically estimate the bias of the inverse of the shape parameter for the two-parameter MLE was also derived. Data sets were generated using Monte Carlo simulation and the maximum likelihood estimates for both the two and three parameter Weibull distributions obtained. As expected, the two parameter estimates showed evidence of asymptotic normality. The three parameter estimates were less well behaved, displaying an increased variance in estimates, but still had decreasing bias in many cases.

The two parameter Weibull distribution was fit to three samples of tornado F-scale data. Each fit was determined to be appropriate as judged by the Chi-squared test. The maximum likelihood estimates from the national tornado record were not compared with those obtained by the methods of Dotzek et al. (2003) and Feuerstein et al. (2005) because the estimates in this thesis were applied to positive F-scale ratings only.

Future work on this topic could involve further describing the estimator bias through analytic means. A more efficient procedure for the maximum likelihood estimation of three parameter Weibull distribution, such as the golden ratio algorithm, could be implemented. Such a procedure could possibly be modified to include estimates of subcritical and F0 tornadoes to achieve the realistic lower bound of zero tornadoes with zero mile per hour wind speed. The estimates achieved with such a method could then be compared with those obtained by Dotzek et al. (2003) and Feuerstein et al. (2005) to determine the relative effectiveness of the respective estimation procedures.

Additionally, estimation of the generalized gamma distribution could be developed and modified to account for subcritical and F0 tornadoes. Fits by the generalized gamma distribution could then be compared with those achieved by other methods to determine if there is any significant advantage to further generalizing the current tornado intensity model.

REFERENCES

Alabama Emergency Management Agency, & Federal Emergency Management Agency. (2011). *Joint Alabama – FEMA situation report No. 53*. Retrieved from the Alabama Emergency Management Agency website at http://ema.alabama.gov/

Artin, E. (1964). *The gamma function*. (M. Butler, Trans.). New York, NY: Holt, Rinehar and Winston, Inc. (Original work published in 1931).

Bartolucci, A.A., Singh, K.P., Bartolucci, A.D., & Bae, S. (1999). Applying medical survival data to estimate the three-parameter Weibull distribution by the method of weighted moments. *Mathematics and Computers in Simulation, 48*, 385-392.

Celik, A.N. (2003). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable Energy, 29*, 593-604. doi:10.1016/j.renene.2003.07.002

Chang, T.P., (2011). Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy, 88*, 272-282. doi:10.1016/j.apenergy.2010.06.018

Dewanji, A., Krewski, D., & Goddard, M.J. (1993). A Weibull model for the estimation of tumorigenic potency. *Biometrics, 49*, 367-377.

Dotzek, N., Grieser, J., & Brooks, H.E. (2003). Statistical modeling of tornado intensity distributions. *Atmospheric Research, 67-68*, 163-187. doi:10.1016/S0169-8095(03)00050-4

Dubey, S.D. (1967). On some permissible estimators of the location parameter of the Weibull distribution and certain other distributions. *Technometrics, 9*, 293-307.

Feuerstein, B., Dotzek, N., & Grieser, J. (2005). Assessing a tornado climatology from global tornado intensity distributions. *American Meteorological Society, 18*, 585-596.

Fisher, R.A., & Tippett , L.H.C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society, 24*, 180-190.

Gomes, O., Combes, C., & Dussauchoy, A. (2008). Parameter estimation of the generalized gamma distribution. *Mathematics and Computers in Simulation, 79*, 955-963.

Gradshteyn, I.S., & Ryzhik, I. (1965). *Tables of integrals series and products*. New York, NY: Academic Press, Inc.

Gupta, P.L., Gupta, R.C., & Lvin, S.J. (1998). Numerical methods for the maximum likelihood estimation of the Weibull parameters. *Journal of Statistical Computation and Simulation, 62*, 1-7.

Hirose, H. (1996). Maximum likelihood estimation in the 3-parameter Weibull distribution: A look through the generalized extreme-value distribution. *IEEE Transaction on Dielectrics and Electrical Insulation, 3 (1)*, 43-55.

Hirose, H. (1999). Bias correction for the maximum likelihood estimates in the two-parameter Weibull distribution. *IEEE Transaction on Dielectrics and Electrical Insulation, 6 (1)*, 66-68.

Hogg, R.V., McKean, J.W., & Craig, A.T. (2005). *Introduction to mathematical statistics* (6^{th} ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Johnson, S.R., & Holt, M.T., (1997). The value of weather information. In R.W. Katz & A.H. Murphy (Eds.), *Economic value of weather and climate forecasts* (pp. 75-108). Cambridge: Cambridge University Press.

McCool, J.I. (1998). Inference on the Weibull location parameter. *Journal of Quality Technology, 30*, 119-126.

Meaden, G.T., Kochev, S., Kolendowicz, L., Kosa-Kiss, A., Marcinoniene, I., Sioutas, M, Tooming, H., & Tyrrell, J. (2005). Comparing the theoretical versions of the Beaufort scale, the T-scale and the Fujita scale. *Atmospheric reseach, 83*, 446-449.

Montanari, G.C., Mazzanti, G., Cacciari, M., & Fothergill, J.C. (1997). In search of convenient techniques for reducing bias in the estimates of Weibull parameters for uncensored tests. *IEEE Transaction on Dielectrics and Electrical Insulation, 4*, 306-313.

Murthy, D.N.P., Xie, M., & Jiang, R. (2004). *Weibull Models*. Hoboken, NJ: John Wiley & Sons, Inc.

National Oceanographic and Atmospheric Administration. (2011). *2011 tornado information: Preliminary tornado statistics including records set in 2011*. Retrieved from the National Oceanographic and Atmospheric Administration website at http://www.noaanews.noaa.gov/2011_tornado_information.html

National Weather Service Weather Forecast Office. (2011). *April severe weather events set new tornado records for Alabama*. Retrieved from the National Oceanographic and Atmospheric Administration website at http://www.srh.noaa.gov/bmx/?n=climo_2011torstats

Panchang, V.G., & Gupta, R.C. (1989). On the determination of the three-parameter Weibull MLE's. *Communications in Statistics – Simulation and Computation, 18(3)*, 1037-1057. doi:10.1080/03610918908812805

Pham, H., & Lai, C. (2007). On recent generalizations of the Weibull distribution. *IEEE Transactions on Reliability, 56*, 454-458.

Pinsky, M.A., & Karlin, S. (2011). *An introduction to stochastic modeling (4th ed.)*. Burlington, MA: Elsevier, Inc.

Rinne, H. (2009). *The Weibull distribution: A handbook*. Boca Raton, FL: CRC Press.

Ross, R. (1994). Formulas to describe the bias and standard deviation of the ML-estimated Weibull shape parameter. *IEEE Transactions on Dielectrics and Electrical Insulation, 1(2)*, 247-253.

Schmid, U. (1997). Percentile estimators for the three-parameter Weibull distribution for use when all parameters are unknown. *Communications in Statistics – Theory and Methods, 26(3)*, 765-785. doi:10.1080/03610929708831948

Stacy, E.W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics, 33*, 1187-1192.

Stacy, E.W., & Mihram, G.A. (1965). Parameter estimation for a generalized gamma distribution. *Technometrics, 7*, 349-357.

Storm Prediction Center. (2011). *The enhanced Fujita scale*. Retrieved from the National Oceanographic and Atmospheric Administration's Storm Prediction Center website at http://www.spc.noaa.gov/efscale/

Storm Prediction Center. (2012). *Severe weather database files (1950-2010)* [Data files and format document]. Retrieved from http://www.spc.noaa.gov/wcm/

Wang, F.K., & Keats, J.B. (1995). Improved percentile estimation for the two-parameter Weibull distribution. *Microelectronics Reliability, 35*, 883-892.

Weibull, W. (1951). A statistical distribution of wide applicability. *Journal of Applied Mechanics, 18*, 293-297.

Wingo, D.R. (1987). Computing maximum-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation. *IEEE Transactions on Reliability, 36(5)*, 586-590.

Zanakis, S.H. (1979). Extended pattern search with Transformations for the three-parameter Weibull MLE problem. *Management Science, 25*, 1149-1161. Retrieved from http://www.jstor.org/stable/2630415