AN ANALYSIS OF THE WEIBULL FAMILY OF DISTRIBUTIONS
FROM A QUANTILE BOX- PLOT PERSPECTIVE

By J. Michael White
Institute of Statistics, Texas A&M University

February 1979

Texas A & M Research Foundation
Project No. 5061

"Maximum Robust Likelihood Estimation and Non-parametric Statistical Data Modeling"

Sponsored by the U.S. Army Research Office

Professor Emmanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

The Weibull family of distributions has been the topic of a number of papers that attempt to define its properties and investigate techniques to detect data distributed as Weibull. This paper discusses the properties of the one parameter Weibull family, with distribution function,

\[ F(x, \theta) = 1 - \exp(-x^{1/\theta}), \quad x > 0, \]

for a range of values of \( \theta \) from a Quantile Box Plot perspective. The values of \( \theta \) used are \( 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 \).

The findings in this report are not to be construed as an official Department of the Army position, unless designated by other authorized documents.

Top Secret/SCI (If the document is to be placed in the secure box, identify by black marker)
AN ANALYSIS OF THE WEIBULL FAMILY OF DISTRIBUTIONS FROM A QUANTILE BOX-PLOT PERSPECTIVE

By J. Michael White

Abstract

The Weibull family of distributions has been the topic of a number of papers that attempt to define its properties and investigate techniques to detect data distributed as Weibull. This paper discusses the properties of the one parameter Weibull family, with distribution function,

\[ F(x, \beta) = 1 - \exp(-x^{1/\beta}) \quad x > 0, \]

for a range of values of \( \beta \) from a Quantile Box Plot perspective. The values of \( \beta \) used are \([1, .667, .5, .333, .25, .2, .167, .125, .1]\). Section 1 of the paper discusses the features of Quantile Box-Plots and the associated diagnostics. Section 2 presents the results of analyzing the Weibull family using the Quantile Box-Plot technique applied to true quantile functions. The results of simulation studies using samples from the Weibull family are presented in Section 3. Comments and conclusions are stated in Section 4.
1. Quantile Box-Plots and Diagnostic Measures

One approach used to display and graphically analyze a batch of data is the Box-Plot introduced by Tukey (1977), as modified by Parzen (1978) under the name Quantile-Box-Plot. Five values from a set of data are usually used to construct the plots -- the extremes, the upper and lower quartiles (called hinges or H values), and the median (M-value). The basic configuration of Tukey's Box-Plot display is a box of arbitrary width and of length H equal to the upper H minus the lower H (H-spread) with a solid horizontal line drawn within the box passing through the median. The distance from the median to the lower H is called MH. Dashed lines extend vertically from the H-values connecting the H-values with the extremes. When the Box-Plot is superimposed on the sample quantile function, \( \hat{Q}(u) \), \( 0 \leq u \leq 1 \), it is called a Quantile Box-Plot.

The quantile function, \( Q(u) = F^{-1}(u) \), \( 0 \leq u \leq 1 \) of a given distribution function \( F \) can be estimated for a batch of data, \( X_1, \ldots, X_n \), by

\[
\hat{Q}(u) = \hat{F}^{-1}(u) = \inf \{ x : \hat{F}(x) \geq u \},
\]

where \( \hat{F}(x) \) is the sample distribution function; \( \hat{Q}(u) \) can be computed in terms of the order statistics \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \) of \( X_1, X_2, \ldots, X_n \) by

\[
\hat{Q}(u) = X_{(j)}, \quad \frac{1}{n} \leq u \leq \frac{1}{n}
\]

The above definition of \( \hat{Q}(u) \) gives a piecewise constant function. It is more convenient to define \( \bar{Q}(u) \) as a piecewise linear function and divide the interval \( (0, 1) \) into \( 2n \) subintervals. Define

\[
\bar{Q}_{\left(\frac{j-1}{2n}\right)} = X_{(j)}, \quad j = 1, 2, \ldots, n.
\]

For \( u \in \left(\frac{2j-1}{2n}, \frac{2j+1}{2n}\right) \), \( j = 1, 2, \ldots, n-1 \), define \( \bar{Q}(u) \) using linear interpolation by

\[
\bar{Q}(u) = n\left(1 - \frac{2j-1}{2n}\right) X_{(j+1)} + n\left(1 + \frac{2j+1}{2n} - u\right) X_{(j)}.
\]

In particular, \( \bar{Q}(\frac{1}{2}) = \frac{1}{2} (X_{(j+1)} + X_{(j)}) \). This definition has the merit that \( \bar{Q}(.5) \) equals the sample median as usually defined:

\[
\bar{Q}(.5) = X_{(m+1)} \quad \text{if } n = 2m + 1 \text{ is odd},
\]

\[
\bar{Q}(.5) = \frac{1}{2} (X_{(m)} + X_{(m+1)}) \quad \text{if } n = 2m \text{ is even}.
\]

The Quantile Box-Plot obtained by superimposing the Box-Plot on \( \bar{Q}(\cdot) \) consists of a box of width extending from \( .25 \) to \( .75 \) and

\[
79 04 26 032
\]
length extending from \( Q_l(25) \) to \( Q_l(75) \). A horizontal line is
drawn through \( Q_l(5) \) from .25 to .75. This box so drawn is more
aptly labeled the H-box. An E-box can also be drawn using the upper eighth
= \( Q_l(875) \) (upper E-value) and the lower eighth = \( Q_l(125) \) (lower E-value) as
the upper and lower bounds; .125 and .875 are the left and right bounds of
the box. A D-box can be drawn as well using the upper sixteenth = \( Q_l(9375) \)
(upper D-value) and the lower sixteenth = \( Q_l(0625) \) (lower D-value) as the upper
and lower bounds; .0625 and .9375 are the left and right bounds of the box.

An approximate 95% confidence interval for the population median
can be drawn using a vertical line of length \( HI/\sqrt{n} \) centered on \( Q_l(5) \)
(see McGill, Tukey, and Larsen (1978) and Parzen (1978)).

\( Q_l(u) \) is useful for detecting the presence of outliers, modes,
and the existence of two populations. Flat slots in \( Q_l(u) \) indicate modes.
Sharp rises in \( Q_l(u) \) for \( u \) near 0 or 1 suggest outliers; sharp
rises in \( Q_l(u) \) within the boxes lead one to suspect the existence of
two (or more) populations. The location of the box is useful for
detecting symmetry and long or short tailedness. One can readily spot
a symmetric batch of data if the median is centered within each of the
boxes. One can detect data from a long tailed or a short tailed popu-
lation by comparing the ratio of the lengths of the boxes, using as ideal
values the ratios obtained from a normal population.

While the Quantile Box-Plot is a useful graphical tool for
exploratory data analysis, accompanying diagnostics provide analytical
tools to help one to interpret the plots more objectively. A seven point

summary provided by the \( M, H, E \) and \( D \) values of a batch of data
is the basis for the following diagnostic measures.

The five mid-summaries of interest are of the form

\[
\tilde{\mu}(p) = \frac{1}{2} [Q_l(1 - p) + Q_l(p)] \quad \text{for} \quad p = .5, .25, .125, .0625 ,
\]

giving us

\[
\begin{align*}
\text{Med} &= \tilde{\mu}(.5) = \text{sample median}, \\
\text{Mid H} &= \tilde{\mu}(.25) = 1/2 \quad \text{(upper H \ lower H)} \\
\text{Mid E} &= \tilde{\mu}(.125) = 1/2 \quad \text{upper E + lower E} \\
\text{Mid E} &= \tilde{\mu}(.0625) = 1/2 \quad \text{(upper D + lower D)} \\
\tilde{\mu}(0) &= 1/2 \quad \text{(upper extreme + lower extreme)}
\end{align*}
\]

= mid extreme

\[
\bar{X} = \sum X_i / n \quad \text{is computed as well.}
\]

A "quick and dirty" estimator of \( \mu \) in the model \( H_0 : Q(u) = \mu + \sigma Q_0(u) \),
when \( Q_0(u) \) is symmetric in the sense that \( Q_0(u) = -Q_0(1 - u) \),
is provided by \( \tilde{\mu}^* = (1/4) (\text{Med} + \text{Mid H} + \text{Mid E} + \text{Mid D}) \).

The scale summaries of interest are of the form

\[
\tilde{\sigma}(p) = \frac{Q_l(1 - p) - \tilde{\mu}(p)}{Q_0(1 - p) - Q_0(p)} \quad \text{for} \quad p = .25, .125, .0625 .
\]

When \( H_0 \) holds, \( \tilde{\sigma}(p) \) is an approximately unbiased estimator for \( \sigma \).

Two basic quantile functions \( Q_0 \) are \( Q_0 = \theta^{-1} \) (normal) and
\( Q_0 = -\log (1 - u) \) (exponential). The resulting scale summaries are:
HH/HHNor = \sigma^*_{\text{exp}}(.25)
HH/HHexp = \sigma^*_{\exp}(.)
EE/EENor = \sigma^*_{\text{exp}}(.125)
EE/EEexp = \sigma^*_{\exp}(.)
DD/DDNor = \sigma^*_{\text{exp}}(.0625)
DD/DDexp = \sigma^*_{\exp}(.)

A quick and dirty estimator of \sigma under H_0 is provided by
\hat{\sigma}^* = 1/ [\hat{\sigma}(.) + \hat{\sigma}(.) + 2\hat{\sigma}(.)] giving us the two estimators
\hat{\sigma}^*_{\text{exp}} and \hat{\sigma}^*_{\exp}. We also compute
\bar{s} = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n - 1}}.

The skewness diagnostics are of the form
skew(p) = [\hat{\sigma}(p) - \hat{\sigma}(p)]/[\hat{\sigma}(1-p) - \hat{\sigma}(p)] for \ p = .25, .125, .0625

giving us
MH/HH = skew(.25)
ME/EE = skew(.125)
MD/DD = skew(.0625)

For data from a symmetric population, we expect skew(p) to be close
to .5.

The tail diagnostocs computed are of the form
Tail(p, q) = log[\hat{\sigma}(1-p) - \hat{\sigma}(p)]/[\hat{\sigma}(1-q) - \hat{\sigma}(q)]
for \ p, q = .25, .125, .0625.

log (HH/EE) = Tail(., .25)
log (EE/DD) = Tail(., .125)

These values are to be compared with
Tail_{0}(p, q) = log[Q_{0}(1-p) - Q_{0}(p)]/[Q_{0}(1-q) - Q_{0}(q)]
and Tail_{exp}(p, q) = log[\hat{\sigma}(p)/\hat{\sigma}(q)]

Using the values obtained for the normal distribution as the ideal
situation, if a batch of data passes the test for symmetry, it is checked
for normality by comparing Tail(p, q) with Tail_{0}(p, q). Tail(p, q)
significantly larger than Tail_{0}(p, q) indicates a long tailed distribution
while Tail(p, q) significantly smaller than Tail_{0}(p, q) indicates a short tailed
(e.g. uniform) or bimodal distribution.

The true values which we compare the sample values to are:
Tail_{0}(., .25) = -5339
Tail_{0}(., .125) = -2879
Tail_{exp}(., .25) = -5717
Tail_{exp}(., .125) = -3305

2. Quantile Box-Plots of Theoretical Weibull Distributions

Our purpose in undertaking this study is twofold. First we desired
to discover what the Quantile Box-Plots would look like and what values of
the diagnostics we would obtain for the Weibull family of distributions.

We also wanted to discover for which values of the Beta parameter we
might expect data from a Weibull distribution to look like data from more standard distributions, in particular, the Normal and Exponential distributions.

The values of the diagnostics for the family of Weibull distributions analyzed is summarized in Table I. Figures 1 to 9 display the Quantile Box-Plots for the distributions analyzed. The results were obtained by evaluating \[ Q_j(u) = (-\log (1-u))^\beta \] for \( u = \frac{j}{200}, \ j = 1, \ldots, 199 \). Thus there is slight approximation error for some statistics.

When analyzing the diagnostics, bear in mind that one of our purposes is to discover values that conform to what we would expect for the normal and/or exponential distributions. Hence, in analyzing the midsummaries we are not only interested in trends in the diagnostics but also in comparing the values to the accepted location diagnostic for normal data, \( \overline{X} \). For \( \beta = 1 \), we notice a great trend upwards in the value of the midsummaries indicating sharp skewness in the distribution. The mid-summary average \( \overline{\mu}^* \) is much greater than \( \overline{X} \).

For \( \beta = .667 \) and .5 this trend becomes more gradual as the distribution loses its skewness and \( \overline{\mu}^* \) approaches \( \overline{X} \). For \( \beta = .333 \) the upward trend is slight and \( \overline{\mu}^* \) has converged to \( \overline{X} \) reflecting the symmetry of the distribution.

We see that a gradual downward trend has developed for \( \beta = .25 \) and .2 but \( \overline{\mu}^* \) still approximately equals \( \overline{X} \). The downward trend continues for \( \beta = .167, .125 \) and .1 with \( \overline{\mu}^* \) becoming slightly smaller than \( \overline{X} \). Remember that \( Q(u) \) takes on only positive values so the progressively larger left tail as \( \beta \rightarrow 0 \) is balanced by a few large values in computing \( \overline{X} \).

For the scale summaries we are interested in detecting trends in the diagnostics and also in comparing \( \overline{\sigma}^* \) to \( \sigma = \text{standard error} \) and \( \sigma = \int_0^1 f_0(u)dum \) where \( f_0(u) = \phi(t^{-1}(u)) \). For a discussion of the use of \( \sigma_0 \) as an estimate of scale see Parzen (1977), (1978).

Of course for \( \beta = 1 \) the Weibull distribution is the same as the Exponential \( (\lambda = 1) \) and so, as expected, the ratio of the \( \overline{\sigma}(p) \) to the expected exponential value is 1. Notice the influence of the approximation error in computing \( DD/\text{DDexp} \). We notice an upward trend in the values of the scale summaries when compared to the expected normal values. The value of \( \sigma \) is larger than \( \sigma_0 \) which is above the range of \( \overline{\sigma}_4^* \).

For \( \beta = .667 \) the scale summaries for the normal case are also approximately equal and \( \overline{\sigma}_4^* \) is slightly less than \( \sigma_0 \) but considerably less than \( \sigma \). For the exponential case, there is a downward trend with \( \overline{\sigma}_4^* \) larger than \( \sigma_0 \).

For \( \beta = .5 \) and .333, there is a downward trend in the values of \( \overline{\sigma} \) for both the normal and exponential case with \( \overline{\sigma}_4^* \) in the same range but slightly larger than \( \sigma_0 \) or \( \sigma \); \( \overline{\sigma}_4^* \) is larger than \( \sigma_0 \).
in both cases. There is no trend in the values of the scale summaries for the normal case for $\beta = .25$ and .2 indicating symmetry; $\sigma^2$ approximately equals $\sigma_0$ and $\sigma$.

In analyzing the skewness measures one can see a well defined pattern for the Weibull distribution. A symmetric distribution forces the values of the diagnostics to equal .5. A large right tail yields small values of the skewness measures and a large left tail yields values close to 1. For $\beta \geq .333$, there is a downward trend in the values of the diagnostics with very small values for $\beta = 1$ and values close to .5 with only slight trending for $\beta = .333$. For $\beta < .25$ there is an upward trend in the values of the skewness measures with a general tendency to get larger values and a steeper trend as $\beta$ gets smaller. This reflects the same information that one obtains from the Quantile Box-Plots.

The tail measure diagnostics are significant when interpreted in conjunction with the skewness diagnostics. If we have symmetric data we should be interested in how the tails compare to the tails of a normal distribution. For $\beta = 1$, the values of the tail measures are close to what we would expect from the exponential distribution. For $\beta = .667$ log (HH/EE) is far from the exponential value and close to the normal value; log (EE/DD) is closer to the normal value also. Likewise, for $\beta = .5$ the values are closer to the normal values than to the exponential values. For $\beta \leq .333$ there is a general trend downward in the value of log (HH/EE) and log (EE/DD). Values of log (HH/EE) are close to the normal and far from the exponential for $\beta = .333, .25, .2$ and .167; values of log (EE/DD) are between the normal and exponential values but closer to the normal values. For $\beta = .125$ and .1, the values of both diagnostics are between the normal values and exponential values with log (HH/EE) closer to the normal values.

The Quantile Box-Plots tell the same story as the diagnostics. Q(u) is skewed and has a large upper tail for $\beta = 1$. This trend is modified progressively for $\beta = .667$ and .5. Q(u) is symmetric for $\beta = .333$ and .25 and looks very much like $q^{-1}(u)$. For $\beta = .2$ we see the development of a longer lower tail with the tail very pronounced for $\beta = .1$. The Quantile Box Plot for $\beta = .1$ looks almost the opposite as that for $\beta = .667$. 


Table 1
Values of Diagnostic Measures for Theoretical Weibull Distributions

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.667</th>
<th>0.500</th>
<th>0.333</th>
<th>0.250</th>
<th>0.200</th>
<th>0.167</th>
<th>0.125</th>
<th>0.100</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-Point Summary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper D</td>
<td>2.817</td>
<td>1.994</td>
<td>1.678</td>
<td>1.412</td>
<td>1.295</td>
<td>1.230</td>
<td>1.188</td>
<td>1.138</td>
<td>1.109</td>
</tr>
<tr>
<td>Upper E</td>
<td>2.079</td>
<td>1.629</td>
<td>1.442</td>
<td>1.276</td>
<td>1.201</td>
<td>1.158</td>
<td>1.130</td>
<td>1.096</td>
<td>1.076</td>
</tr>
<tr>
<td>Upper H</td>
<td>1.386</td>
<td>1.243</td>
<td>1.177</td>
<td>1.115</td>
<td>1.085</td>
<td>1.068</td>
<td>1.056</td>
<td>1.042</td>
<td>1.033</td>
</tr>
<tr>
<td>Median</td>
<td>0.693</td>
<td>0.783</td>
<td>0.833</td>
<td>0.885</td>
<td>0.912</td>
<td>0.929</td>
<td>0.941</td>
<td>0.955</td>
<td>0.964</td>
</tr>
<tr>
<td>Lower H</td>
<td>0.288</td>
<td>0.436</td>
<td>0.536</td>
<td>0.660</td>
<td>0.732</td>
<td>0.779</td>
<td>0.812</td>
<td>0.856</td>
<td>0.883</td>
</tr>
<tr>
<td>Lower E</td>
<td>0.134</td>
<td>0.261</td>
<td>0.365</td>
<td>0.511</td>
<td>0.604</td>
<td>0.669</td>
<td>0.715</td>
<td>0.777</td>
<td>0.818</td>
</tr>
<tr>
<td>Lower D</td>
<td>0.062</td>
<td>0.156</td>
<td>0.249</td>
<td>0.395</td>
<td>0.498</td>
<td>0.573</td>
<td>0.629</td>
<td>0.706</td>
<td>0.757</td>
</tr>
<tr>
<td>Mid-Summary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(L_a) Median</td>
<td>0.693</td>
<td>0.783</td>
<td>0.833</td>
<td>0.885</td>
<td>0.912</td>
<td>0.929</td>
<td>0.941</td>
<td>0.955</td>
<td>0.964</td>
</tr>
<tr>
<td>(L_b) Mid H</td>
<td>0.837</td>
<td>0.840</td>
<td>0.857</td>
<td>0.888</td>
<td>0.909</td>
<td>0.923</td>
<td>0.934</td>
<td>0.949</td>
<td>0.958</td>
</tr>
<tr>
<td>(L_h) Mid E</td>
<td>1.106</td>
<td>0.954</td>
<td>0.904</td>
<td>0.894</td>
<td>0.903</td>
<td>0.913</td>
<td>0.922</td>
<td>0.937</td>
<td>0.947</td>
</tr>
<tr>
<td>(L_d) Mid D</td>
<td>1.439</td>
<td>1.075</td>
<td>0.963</td>
<td>0.904</td>
<td>0.897</td>
<td>0.901</td>
<td>0.908</td>
<td>0.922</td>
<td>0.933</td>
</tr>
<tr>
<td>(L_s) Average</td>
<td>1.019</td>
<td>0.911</td>
<td>0.889</td>
<td>0.893</td>
<td>0.905</td>
<td>0.917</td>
<td>0.926</td>
<td>0.941</td>
<td>0.950</td>
</tr>
<tr>
<td>(\bar{X}) X Bar</td>
<td>0.997</td>
<td>0.902</td>
<td>0.886</td>
<td>0.893</td>
<td>0.906</td>
<td>0.918</td>
<td>0.928</td>
<td>0.942</td>
<td>0.951</td>
</tr>
<tr>
<td>(H_{\bar{X}}) bar</td>
<td>0.110</td>
<td>0.081</td>
<td>0.064</td>
<td>0.045</td>
<td>0.035</td>
<td>0.029</td>
<td>0.024</td>
<td>0.019</td>
<td>0.015</td>
</tr>
<tr>
<td>Scale Summary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH/HHser</td>
<td>0.815</td>
<td>0.599</td>
<td>0.475</td>
<td>0.337</td>
<td>0.262</td>
<td>0.214</td>
<td>0.180</td>
<td>0.138</td>
<td>0.111</td>
</tr>
<tr>
<td>EE/EEser</td>
<td>0.846</td>
<td>0.595</td>
<td>0.468</td>
<td>0.333</td>
<td>0.259</td>
<td>0.213</td>
<td>0.180</td>
<td>0.138</td>
<td>0.112</td>
</tr>
<tr>
<td>DD/DDser</td>
<td>0.898</td>
<td>0.599</td>
<td>0.466</td>
<td>0.331</td>
<td>0.260</td>
<td>0.214</td>
<td>0.182</td>
<td>0.141</td>
<td>0.115</td>
</tr>
<tr>
<td>(z) Average</td>
<td>0.853</td>
<td>0.597</td>
<td>0.470</td>
<td>0.334</td>
<td>0.260</td>
<td>0.213</td>
<td>0.181</td>
<td>0.139</td>
<td>0.113</td>
</tr>
<tr>
<td>HH/HHexp</td>
<td>1.000</td>
<td>0.735</td>
<td>0.584</td>
<td>0.414</td>
<td>0.321</td>
<td>0.262</td>
<td>0.222</td>
<td>0.169</td>
<td>0.137</td>
</tr>
<tr>
<td>EE/EEexp</td>
<td>1.000</td>
<td>0.703</td>
<td>0.553</td>
<td>0.393</td>
<td>0.306</td>
<td>0.251</td>
<td>0.213</td>
<td>0.164</td>
<td>0.133</td>
</tr>
<tr>
<td>DD/DDexp</td>
<td>1.017</td>
<td>0.679</td>
<td>0.528</td>
<td>0.376</td>
<td>0.294</td>
<td>0.243</td>
<td>0.207</td>
<td>0.160</td>
<td>0.130</td>
</tr>
<tr>
<td>(\bar{z}) Average</td>
<td>1.006</td>
<td>0.706</td>
<td>0.555</td>
<td>0.394</td>
<td>0.307</td>
<td>0.252</td>
<td>0.214</td>
<td>0.164</td>
<td>0.133</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>.9225</td>
<td>6.004</td>
<td>.4625</td>
<td>.3268</td>
<td>.2557</td>
<td>2.108</td>
<td>1.796</td>
<td>1.388</td>
<td>1.132</td>
</tr>
</tbody>
</table>

Skewness

| HH/HH | 0.3690 | 0.4302 | 0.4620 | 0.4943 | 0.5105 | 0.5203 | 0.5268 | 0.5349 | 0.5398 |
| EE/EE | 0.2876 | 0.3816 | 0.4339 | 0.4886 | 0.5164 | 0.5331 | 0.5443 | 0.5583 | 0.5666 |
| DD/DD | 0.2291 | 0.3411 | 0.4083 | 0.4816 | 0.5195 | 0.5424 | 0.5577 | 0.5767 | 0.5881 |

Tail

| Log(HH/EE) | -5.715 | -5.270 | -5.184 | -5.201 | -5.250 | -5.294 | -5.328 | -5.378 | -5.412 |

Weibull Beta=1.000

Weibull Beta=.667

Weibull Beta=.500

Weibull Beta=.333
Quantile Box-Plots of Samples from Weibull Distributions

Simulation samples from Weibull distributions, for various values of the $\beta$-parameter, were analyzed via the Quantile Box-Plot procedure. Six simulations with sample size $n = 100$ were conducted for each of the nine values of $\beta$ used in the study. While more simulations of different sample sizes are desirable and necessary for further analysis, it is hoped that the simulations conducted will give some notion as to what kind of values and deviation we might expect from real data coming from a Weibull distribution.

Uniform (0,1) random variables were generated using the subroutine RANDU, available on the TIMESBOARD library; these were transformed by $Y = (-\log(1-u))^\beta$ giving us our simulated data from the Weibull distribution. The diagnostics were computed and Quantile Box-Plots were drawn for each of the samples. We then computed the sample mean and standard error of each diagnostic using the six estimates thus obtained. An approximate 95% confidence interval for the true value, $\mu$, of the diagnostic could be formed using $\bar{X} \pm 1.96s$ where $\bar{X}$ is the sample mean of our estimates and $s$ is the standard error of our estimates with $\sqrt{n}(\bar{X} - \mu)/s$ approximately distributed as a Student's $t$ with $5$ d.f. The results of these computations are presented in Table II. The first number in each entry in the table is $\bar{X}$ followed by $s$.

Upon examining the midsummaries one sees about the same trends as in Table I with the trending least evident for $\beta = .337, .25$ and .20. One also notices that as $\beta$ goes to 0, the standard error of the estimates generally decreases. One notices that the length of the confidence interval $\bar{H}/\sqrt{n}$ also decreases as $\beta$ goes to 0. A further trend evident in Table I is that the values of the midsummary statistics generally increase as $\beta$ goes to 0 probably due to the elimination of a constant in the computation of $Q(u)$ . One notices that $\bar{H}$ is closer to $\bar{X}$ for $\beta = .333$ than for $\beta = .5$. One might expect the standard error to increase as one goes from Med to Mid D but this is not evident.

The scale summary diagnostics reveal the same kind of trending in the standard error of the estimates. The standard error of the estimates decreases as $\beta$ goes to zero. One detects a surprisingly high standard error in the values of the scale summaries in the exponential case for $\beta = 1$. When the scale summaries in the normal case are compared to the sample standard error and $\sigma$, one sees the closeness of the statistics for all values of $\beta < .667$ except for $\beta = .25$ which was unexpected. It is interesting to note the consistency of the estimates even for quite small values of $\beta$.

The skewness diagnostics reveal one surprising result. There seems to be an absence of trending in the standard error of the estimates as $\beta$ goes to 0 and the standard error is greatest for $\beta$ in the
range (.667, .167) which includes the normal case. One also sees a wider range of values of the diagnostic than one expects when the results in Table II are compared to those of Table I. As expected, one finds values of the diagnostic closest to .5 for $\beta$ in the range (.333, .167). However, the lack of symmetry in the tails of the Weibull distribution with $\beta = .167$ is more obvious when using MD/DD.

Upon examining the tail diagnostic, one again notices no trending in the standard error of the estimates as $\beta$ goes to 0. One would probably not use this diagnostic for $\beta = 1$ and it is not surprising that the greatest standard error is for $\beta = 1$. The trending in the values of the diagnostic that was evident in Table I is also absent. The values for $\beta = 1$ are not close to the expected exponential values of -.5717 and -.3305. However, one does notice values of -.5339 and -.2879 for $\beta$ in the range (.333, .20). It is curious that for $\beta = .1$ we also obtain values close to the expected normal values but one would disregard these values, having already disregarded normality based on the skewness diagnostics. The values of the diagnostics in Table II show wide deviations from their true value as found in Table I and one sees little trending in the over or under-estimations.

A selected number of Quantile Box-Plots are included in the Appendix to give some idea as to what kind of deviations from the ideal one can expect.

<table>
<thead>
<tr>
<th>Table II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mid E</td>
</tr>
<tr>
<td>HH/$\bar{N}$</td>
</tr>
<tr>
<td>$\bar{\delta}$ Average</td>
</tr>
<tr>
<td>H/H/HHexp</td>
</tr>
<tr>
<td>DD/DDexp</td>
</tr>
<tr>
<td>$\delta_{exp}$ Average</td>
</tr>
<tr>
<td>$\sigma_0$</td>
</tr>
<tr>
<td>ME/EE</td>
</tr>
<tr>
<td>log(HH/EE)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mid H</td>
</tr>
<tr>
<td>Mid E</td>
</tr>
<tr>
<td>Mid D</td>
</tr>
<tr>
<td>\bar{z} Average \bar{X}</td>
</tr>
<tr>
<td>HH/\bar{H}</td>
</tr>
<tr>
<td>DD/DDH</td>
</tr>
<tr>
<td>\bar{z} \bar{z} Average \bar{z}</td>
</tr>
<tr>
<td>HH/HHexp</td>
</tr>
<tr>
<td>EE/EEx</td>
</tr>
<tr>
<td>DD/DDexp</td>
</tr>
<tr>
<td>\bar{z} \bar{z} Average \bar{z}</td>
</tr>
<tr>
<td>S. E.</td>
</tr>
<tr>
<td>\bar{g}_0</td>
</tr>
<tr>
<td>MD/DD</td>
</tr>
<tr>
<td>log(HH/EE)</td>
</tr>
<tr>
<td>log(EE/DD)</td>
</tr>
</tbody>
</table>

4. **Comments and Conclusions**

This study is lacking in one outstanding aspect; more simulations are needed for different sample sizes. The distribution theory of the skewness and tail diagnostics can be computed with difficulty using the distribution of functions of order statistics but more work in this direction is also needed. One should be cautious about drawing conclusions from a limited Monte Carlo study. However, it seems that one can draw some conclusions from this analysis of the Weibull distribution.

First one notices a surprising predictability in the values of the midsummary and scale summary diagnostics for rather small values of \( \beta \). It appears that in the simulations \( \mu(p) \) is closer to its true value than \( \bar{X} \) to its true value for almost the entire range of \( \beta \) values analyzed. The scale summary diagnostics are not quite so predictable but are in general at least as consistent as the \( \bar{z} \) statistic. It seems that the skewness diagnostics are susceptible to be influenced by mild aberrations in the data. The same seems to be true of the tail diagnostics. This could account somewhat for the unexpected fluctuations in the value of the statistics.

While one can very easily detect from looking at the diagnostics obtained from the true quantile function that the Weibull distribution with \( \beta = 1 \) is the same as the Exponential (\( \lambda = 1 \)) distribution, it is not so obvious to detect a batch of data from the Weibull (\( \beta \neq 1 \))
distribution. However it seems that data from a Weibull distribution with $\beta$ in the range (.33, .20) could easily be classified as Normal data when analyzed from a Quantile Box-Plot perspective. The analysis of the true Quantile function supports this contention with all evidence pointing to the approximate equivalence of the two distributions for $\beta$ in the range (.33, .20). It has been suggested that the Weibull distribution as $\beta$ goes to zero closely resembles an extreme value distribution. The Quantile Box plots point to this equivalence. However, the extreme value distribution has not been analyzed from a Quantile Box-Plot perspective; here is another area for further exploration.

The values of the diagnostics as given in Table I have inherent worth in that they are the true values of the statistics for the Weibull distribution for the range of $\beta$ values analyzed. The corresponding Quantile Box-Plots also have value in that they represent the true Quantile function with the associated boxes. Table II and the Quantile Box-Plots associated with the simulations are interesting in that they reveal the deviations from the ideal that we can expect.

References


