Austrian Journal of Statistics
MMMMMM YYYY, Volume VV, 1–xx.
http://www.ajs.or.at/

Properties of the left-truncated two-parameter Weibull distributions

Ayşe Kızılersü
Department of Physics
University of Adelaide
5005, Adelaide, Australia

Markus Kreer
CAMPUSERVICE GmbH
Servicegesellschaft der
Johann Wolfgang
Goethe-Universität Frankfurt
60323, Frankfurt am Main, Germany

Anthony W. Thomas
Department of Physics
University of Adelaide
5005, Adelaide, Australia

Abstract

The left-truncated Weibull distribution is used in life-time analysis, it has many applications ranging from financial market analysis and insurance claims to the earthquake inter-arrival times. We present a comprehensive analysis of the left-truncated Weibull distribution when the shape, scale or both parameters are unknown and they are determined from the data using the maximum likelihood estimator. We demonstrate that if both the Weibull parameters are unknown then there are sets of sample configurations, with measure greater than zero, for which the maximum likelihood equations do not possess non trivial solutions. The modified critical values of the goodness-of-fit test from the Kolmogorov-Smirnov test statistic when the parameters are unknown are obtained from a quantile analysis. We find that the critical values depend on sample size and truncation level, but not on the actual Weibull parameters. Confirming this behavior, we present a complementary analysis using the Brownian bridge approach as an asymptotic limit of the Kolmogorov-Smirnov statistics and find that both approaches are in good agreement. A power testing is performed for our Kolmogorov-Smirnov goodness-of-fit test and the issues related to the left-truncated data are discussed. We conclude the paper by showing the importance of left-truncated Weibull distribution hypothesis testing on the duration times of failed marriages in the US, worldwide terrorist attacks, waiting times between stock market orders, and time intervals of radioactive decay.

Keywords: Maximum likelihood estimation; Kolmogorov-Smirnov goodness-of-fit test; left-truncated data; Monte Carlo simulations; asymptotic analysis; quantiles, Brownian bridge.

1. Introduction and preliminaries

The Weibull distribution with scale and shape parameters, $\alpha > 0$ and $\beta > 0$ respectively, is widely used in areas such as statistics, engineering, finance, insurance and biology (e.g. Weibull (1951), Balakrishnan and Cohen (1991), Rinne (2009)), mainly in the context of life-time analysis (survival analysis in medical studies and reliability analysis in engineering). In practical applications, very often truncated statistical distributions must be used: these truncated statistical distributions arise when a random variable τ follows a known distributional model, except that a portion of the sample space cannot be observed or is removed (for example in radioactive decay phenomena a Geiger-Müller counter does not permit detection of decays events within its dead time). An independent identically distributed (i.i.d.) left-truncated data set $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_n)$ of sample size n has the property that $\tau_L < \tau_i, i = 1, ..., n$ for a given non-negative parameter τ_L, the truncation point (Kendall and Stuart (1979),pp. 551, section 32.15). The left-truncated cumulative Weibull distribution function (cdf) is given by Wingo (1989)

$$F(\tau|\alpha, \beta, \tau_L) = 1 - \exp\left[\left(\frac{\tau_L}{\alpha}\right)^\beta - \left(\frac{\tau}{\alpha}\right)^\beta\right] \quad \text{for} \quad \tau > \tau_L \tag{1}$$

and the left-truncated probability density function (pdf) is

$$f(\tau|\alpha, \beta, \tau_L) = \frac{\beta}{\alpha}\left(\frac{\tau}{\alpha}\right)^{\beta-1} \exp\left[\left(\frac{\tau_L}{\alpha}\right)^\beta - \left(\frac{\tau}{\alpha}\right)^\beta\right] \quad \text{for} \quad \tau > \tau_L\,. \tag{2}$$

Putting $\tau_L = 0$ in Equation (1) and Equation (2), cdf and pdf of the complete Weibull distribution will be recovered, respectively. Throughout this paper we use the term complete Weibull distribution to refer to the untruncated Weibull distribution and in our investigation we assume that the truncation point τ_L is known or can be set. The literature on data analysis tends to focus either on complete or censored data, with much less attention paid to truncated data, moreover truncation formally defined as in Kendall and Stuart (1979), (pp. 551, section 32.15) is sometimes confused by censoring.

When dealing with a sample data obtained from observations one may wish to test the hypothesis that these data are drawn from a left-truncated Weibull distribution, even if the scale parameter α and shape parameter β are unknown. A common method for estimating the parameters of a pdf from a sample data set is maximum likelihood estimation (MLE). Note that the left-truncated Weibull pdf, Equation (2), is continuously differentiable in the argument τ and its two parameters, $0 < \alpha, \beta < \infty$, to any order and thus $f \in C^\infty((\tau_L, \infty) \times (0, \infty) \times (0, \infty))$. Also f and all its derivatives with respect to τ, α, β vanish for $\tau \to \infty$, at least like $\exp\left[-(\tau/\alpha)^{\beta'}\right]$ for $\alpha > 0$ and any $\beta' \in (0, \beta)$. These regularity conditions are essential for the "well-behaviour" of MLE.

To determine how well the sampled data fits the hypothesized distribution one must measure the goodness-of-fit (gof). Studies using Kolmogorov-Smirnov gof test to determine whether the sampled data belong to an untruncated Weibull distribution began in the late 1970s by Littell, McClave, and Offen (1979), Chandra, Singapurwalla, and Stephens (1981); Parsons and Wirsching (1982). In performing the hypothesis test it is crucial to use the correct critical values. When the Weibull parameters are estimated from the sample data, the standard Kolmogorov-Smirnov test tables Smirnov (1948); Miller (1956) for the case where the parameters are known cannot be used, because the probability integral transform using the estimated parameters destroys the independence of the transformed random variables as demonstrated by David and Johnson (1948).

In the literature there are very few studies dedicated to the left-truncated Weibull distributions (LTWD) is Wingo (1989), Balakrishnan and Mitra (2012). However, the MLE-approach in the first reference is rather heuristic level whereas the second reference is more concerned with a maximisation-expectation approach to handle left-truncation and right-censoring. For theoretical investigations of the Weibull distribution the reader is referred to Agostino and Stephens (1986) and Lehmann and Casella (1998).

For the left-truncated 2-parameter Weibull distribution we shall distinguish four cases throughout this article :

Case I: Both parameters, the scale parameter, $\alpha > 0$, and the shape parameter, $\beta > 0$, are known a-priori.

Case II: Both parameters, the scale parameter, $\alpha > 0$, and the shape parameter, $\beta > 0$, are unknown a-priori and need to be estimated from the sample data.

Case IIIa: The scale parameter, $\alpha > 0$, is unknown and needs to be estimated from the sample data, but $\beta > 0$ is known.

Case IIIb: The shape parameter, $\beta > 0$, is unknown and needs to be estimated from the sample data, but $\alpha > 0$ is known.

In the next section we briefly review the maximum likelihood estimation for Cases II - III and comment on the consistency, asymptotic normality and efficiency of the MLE when applied to data sampled from a left-truncated Weibull distribution. Details on these issues have been discussed in Kreer, Kizilersu, Thomas, and dos Reis (2015). In Section 3 we discuss and develop the Kolmogorov-Smirnov (KS) goodness-of-fit (gof) statistics for the left-truncated Weibull distribution to decide whether the sample data could belong to the hypothetical distribution. In Section 4 we present an asymptotic analysis exploiting the Brownian bridge character of the KS statistics following some prior work of Durbin (1973) and Stephens (1977) on untruncated distributions and give our results for the left-truncated Weibull distribution for all cases. The quantile analysis to determine the modified critical values using Monte Carlo simulations is given in Section 5, where we discuss our numerical algorithm and present our results on the left-truncated data for the four cases listed above. All the results obtained on modified critical values are discussed and analysed in Section 6. In Section 7 we give a procedure for interpreting the results and a power study for Case I and Case II. Section 8 discusses the application of the methods discussed throughout the paper to failed US marriages, worldwide terrorist attacks, a sample of stock market data from New York stock exchange, and the radioactive α-decay of Americium-241. All the results are discussed in the concluding section.

2. Maximum likelihood estimation of Left-Truncated Weibull parameters

The maximum likelihood estimates of the left-truncated Weibull parameters differ from the complete ones because the left-truncated pdf $f(\cdot)$ with left-truncation point $\tau_L > 0$ has an additional multiplicative factor $\exp\left(\frac{\tau_L}{\alpha}\right)^\beta$ in comparison to the complete one. From Equation (2) we determine that the likelihood function for the left-truncated Weibull distribution as

$$L_{trunc}(\tau_1, \tau_2, ..., \tau_n|\alpha, \beta, \tau_L) = \prod_{i=1}^{n} \frac{\beta}{\alpha}\left(\frac{\tau_i}{\alpha}\right)^{\beta-1} e^{\left(\frac{\tau_L}{\alpha}\right)^\beta - \left(\frac{\tau_i}{\alpha}\right)^\beta}, \tag{3}$$

and consequently the logarithm of the likelihood as

$$\begin{aligned} \log L_{trunc}(\tau_1, \tau_2, ..., \tau_n|\alpha, \beta, \tau_L) &= \sum_{i=1}^{n} \log\left[\frac{\beta}{\alpha}\left(\frac{\tau_i}{\alpha}\right)^{\beta-1} e^{-\left(\frac{\tau_i}{\alpha}\right)^\beta}\right] + n\left(\frac{\tau_L}{\alpha}\right)^\beta \\ &= n\log\beta - n\beta\log\alpha + (\beta-1)\sum_{i=1}^{n}\log\tau_i - \sum_{i=1}^{n}\left(\frac{\tau_i}{\alpha}\right)^\beta + n\left(\frac{\tau_L}{\alpha}\right)^\beta, \\ &= \log L\left(\boldsymbol{\tau}|\alpha, \beta, 0\right) + n\left(\frac{\tau_L}{\alpha}\right)^\beta \end{aligned} \tag{4}$$

where $L(\boldsymbol{\tau}|\alpha, \beta, 0)$ is the likelihood function for the untruncated distribution.

The Weibull parameters that maximize the likelihood function, Equation (3), are the same as those that maximise the log-likelihood function, Equation (4), and are obtained by calculating the partial derivatives with respect to α and β :

$$\frac{\partial}{\partial\alpha}\log L_{trunc}(\tau_1,\tau_2,...,\tau_n|\alpha,\beta,\tau_L) = \frac{\partial}{\partial\alpha}\log L(\tau_1,\tau_2,...,\tau_n|\alpha,\beta) + n\frac{\partial}{\partial\alpha}\left(\frac{\tau_L}{\alpha}\right)^{\beta} = 0\,,$$

$$\Longrightarrow -n\beta\frac{1}{\alpha} + \beta\sum_{i=1}^{n}\tau_i^{\beta}\alpha^{-\beta-1} - n\beta\left(\frac{\tau_L}{\alpha}\right)^{\beta}\frac{1}{\alpha} = 0\,. \tag{5}$$

$$\frac{\partial}{\partial\beta}\log L_{trunc}(\tau_1,\tau_2,...,\tau_n|\alpha,\beta,\tau_L) = \frac{\partial}{\partial\beta}\log L(\tau_1,\tau_2,...,\tau_n|\alpha,\beta) + n\frac{\partial}{\partial\alpha}\left(\frac{\tau_L}{\beta}\right)^{\beta} = 0\,,$$

$$\Longrightarrow \frac{n}{\beta} - n\log\alpha + \sum_{i=1}^{n}\log\tau_i - \sum_{i=1}^{n}\log\left(\frac{\tau_i}{\alpha}\right)\cdot\left(\frac{\tau_i}{\alpha}\right)^{\beta} + n\log\left(\frac{\tau_L}{\alpha}\right)\cdot\left(\frac{\tau_L}{\alpha}\right)^{\beta} = 0\,. \tag{6}$$

Rearranging Equation (5) we get

$$\alpha \;=\; \left(\frac{1}{n}\sum_{i=1}^{n}\left[\tau_i^{\beta} - \tau_L^{\beta}\right]\right)^{1/\beta}\,. \tag{7}$$

Note that Equation (7) is one of two MLE equations in Case II but is the only MLE equation in Case IIIa. There always exists a solution for α in Case IIIa for a given β. Rewriting Equation (6) we obtain the following

$$n\frac{1}{\beta} + \sum_{i=1}^{n}\log\left(\frac{\tau_i}{\alpha}\right) - \sum_{i=1}^{n}\left(\frac{\tau_i}{\alpha}\right)^{\beta}\log\left(\frac{\tau_i}{\alpha}\right) + \sum_{i=1}^{n}\left(\frac{\tau_L}{\alpha}\right)^{\beta}\log\left(\frac{\tau_L}{\alpha}\right) \;=\; 0\,. \tag{8}$$

Equation (8) is the second MLE equation in Case II but the only MLE equation in Case IIIb, where α is known. Eliminating α in Equation (8) using Equation (7), we obtain (after some algebraic manipulation) the following equation for β (for Case II) (Wingo (1989) and arxiv-version of Malevergne, Pisarenko, and Sornette (2005))

$$0 = \frac{1}{\beta} - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\tau_i}{\tau_L}\right)^{\beta}\log\frac{\tau_i}{\tau_L}}{\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{\tau_i}{\tau_L}\right)^{\beta} - 1\right]} + \frac{1}{n}\sum_{i=1}^{n}\log\frac{\tau_i}{\tau_L}\,. \tag{9}$$

Equations (7) and (9), reduce, in the limit $\tau_L \to 0$, to those given in Cohen (1965) for untruncated MLE equations. The solutions for α and β to the simultaneous Equations (7) and (9) are denoted by $\hat{\beta}_n = \hat{\beta}_n(\tau_1,\ldots,\tau_n|\tau_L)$ and $\hat{\alpha}_n = \hat{\alpha}_n(\tau_1,\ldots,\tau_n|\tau_L)$. For convenience we shall suppress the dependence on the sample $\tau_1,...,\tau_n$ and the left-truncation value τ_L and simply write $\hat{\alpha}$ and $\hat{\beta}$. The existence and uniqueness of a non-trivial MLE solution is almost trivial for Case IIIa, whereas the Case II and Case IIIb are dealt with the Lemma I given in Kreer *et al.* (2015). To assert the existence of a non-vanishing MLE-solution, the sample data need to satisfy the following inequality

$$2\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{\tau_i}{\tau_L}\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{\tau_i}{\tau_L}\right)^2 > 0\,. \tag{10}$$

If the condition given Equation (10) is not satisfied then the only solution to the MLE equation for β, Equation (9) is the trivial solution $\alpha = \beta = 0$. This can be shown by inserting $\alpha = \beta^{1/\beta}$ and taking the limit $\beta \to 0$ in Equation (3). Only in this case the likelihood Equation (3) is positive and non vanishing[1].

Table 1 was generated using a Monte Carlo Simulation and gives the percentage of left-truncated Weibull distributed random samples of size n satisfing Equation (10) and hence for which the MLE provides a non-trivial solution.

[1] An example of the violation of the second MLE equation Equation (9) is for $n = 30$ the random sample $\tau_i = \tau_L + \epsilon\cdot i$ for $i = 1, 2, ..., 25$, a sufficiently small $\epsilon > 0$ and $\tau_i = \ell\cdot\tau_L + \epsilon\cdot i$ for $i = 26, 27, ..., 30$ and some $\ell \gg 1$. In this case we see that Equation (10) is not satisfied.

Table 1: Percentage of left-truncated Weibull distributed random samples for which there exists a solution to the MLE equations for Case II.

Truncation Rate	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %
n=30	100±0	100±0	100±0	98 ±1	94±1	89±1	85±1	83±1	82 ±0	83±0
n=50	100±0	100±0	100±0	100±0	98±0	95±1	90±1	87±1	85±1	83±1
n=100	100±0	100±0	100±0	100±0	100±0	99±0	98±0	95±1	92±1	87±1

The consistency, asymptotic normality and efficiency of the MLE method for left-truncated Weibull distribution are discussed in Theorem 1 in Kreer *et al.* (2015) and the relevant proofs are provided as well. Key for the proof is the smoothness property of the left-truncated Weibull distribution. Denoting the true parameter vector by (α^0, β^0), we note in particular that all the asymptotic properties follow in this case from the asymptotic normality, i.e. $\sqrt{n}\left((\hat{\alpha}_n, \hat{\beta}_n) - (\alpha^0, \beta^0)\right)$ is asymptotically normal with vector mean zero and covariance matrix $[Z((\alpha^0, \beta^0))]^{-1}$ being the inverse of the Fisher information matrix

$$Z(\alpha^0, \beta^0) = -\mathbb{E}\begin{bmatrix} \frac{\partial^2 \log f(\tau|\alpha,\beta,\tau_L)}{\partial \alpha^2} & \frac{\partial^2 \log f(\tau|\alpha,\beta,\tau_L)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \log f(\tau|\alpha,\beta,\tau_L)}{\partial \beta \partial \alpha} & \frac{\partial^2 \log f(\tau|,\alpha,\beta,\tau_L)}{\partial \beta^2} \end{bmatrix}_{\alpha=\alpha^0, \beta=\beta^0} \tag{11}$$

The elements of the Fisher information matrix, Equation (11), are calculated as

$$\begin{aligned} \mathbb{E}\left(\frac{\partial^2}{\partial \alpha^2} \log f(\tau|\alpha,\beta,\tau_L)\right) &= -\frac{\beta^2}{\alpha^2}, \\ \mathbb{E}\left(\frac{\partial^2}{\partial \alpha \partial \beta} \log f(\tau|\alpha,\beta,\tau_L)\right) &= \frac{1}{\alpha}\left\{1 + [\log \eta + e^{\eta} E_1(\eta)]\right\}, \\ \mathbb{E}\left(\frac{\partial^2}{\partial \beta^2} \log f(\tau|\alpha,\beta,\tau_L)\right) &= -\frac{1}{\beta^2}\left\{1 + 2[\log \eta + e^{\eta} E_1(\eta)] + \left[(\log \eta)^2 + 2e^{\eta} E_2(\eta)\right]\right\}. \end{aligned}$$

where we have used the functions $E_1(s) = \int_s^\infty dy \exp(-y)/y$ (i.e. the exponential integral) and $E_2(s) = \int_s^\infty dy \exp(-y) \log(y)/y$.

3. Kolmogorov-Smirnov goodness-of-fit test for the left-truncated Weibull distribution

Let us test the following null hypothesis H_0: The i.i.d. sample $\tau_1, \tau_2, \cdots, \tau_n$ satisfying $\tau_L < \tau_i$ for $i = 1, 2, ..., n$ for some positive τ_L and some integer n, is drawn from a left-truncated Weibull distribution $F(\boldsymbol{\tau})$ as given in Equation (1) with estimated parameters $(\hat{\alpha}, \hat{\beta})$ obtained from MLE as discussed in the previous section[2]. Using the empirical distribution function $F_n(\boldsymbol{\tau})$, defined as the proportion of the values of the order statistics $\tau_{(1)}, \tau_{(2)}, ..., \tau_{(n)}$ smaller than $\tau \in (\tau_L, \infty)$, the Kolmogorov-Smirnov (KS) test statistic is given (e.g. Kendall and Stuart (1979), sect. 30.49 and Shorack and Wellner (2009)),

$$\begin{aligned} D_n &\equiv \sup_{-\infty < \tau < +\infty} \|F_n(\boldsymbol{\tau}) - F(\boldsymbol{\tau})\|, \qquad (12) \\ &= \sup_{\tau_L < \boldsymbol{\tau} < \infty} [F_n(\boldsymbol{\tau}) - F(\boldsymbol{\tau}), F(\boldsymbol{\tau}) - F_n(\boldsymbol{\tau})] \\ &= \max_{1 \le i \le n} \left[\frac{i}{n} - F(\tau_i), F(\tau_i) - \frac{i-1}{n}\right]. \qquad (13) \end{aligned}$$

Here D_n is the KS distance which is compared with a critical value $D_{cv}(n, p, 0.05)$, that depends on the sample size n, the truncation level p (the theoretical percentage removed

[2]From this point onwards we will drop the index n and use $\hat{\alpha}$ and $\hat{\beta}$.

from the untruncated distribution) and significance level 0.05 used throughout the paper. If the value of $\sqrt{n}D_n$ is greater than some critical value $D_{cv}(n,p,0.05)$ then the hypothesis that $F_n(\boldsymbol{\tau})$ and $F(\boldsymbol{\tau})$ come from the same distribution is rejected, i.e.,

> H_0 is the hypothesis that the set of values τ is sampled from a random distribution with a known cdf $F(\boldsymbol{\tau})$,

$$H_0 \quad \text{is not rejected if} \quad \sqrt{n}D_n < D_{cv}(n,p,0.05) \quad . \tag{14}$$

The critical values used in the hypothesis test, Equation (14), depend on whether the parameters, α, β, are known or unknown and are estimated from the data itself. The cases introduced earlier in section 1 can be grouped under two the categories for the purpose of KS statistics.

Out-sample KS statistics If the parameters of the distribution from which the sampled data is to be tested against are known precisely, i.e. if $F(\boldsymbol{\tau})$ in Equation (12) is known this referred to as an *out-sample* KS statistic. In this study this statistics is named as Case I where the critical values (CVs) of Kolmogorov and Smirnov are recovered. Moreover the CVs are independent of the distribution and the range of parameters.

In-sample KS statistics If the parameters of the distribution must be estimated from the sampled data to construct the theoretical cdf ($F(\boldsymbol{\tau})$ in Equation (12)), then D_n is referred as an *in-sample* KS statistic. It is well known, when the parameters are estimated from the sample and then the goodness-of-fit test is performed, that the probability integral transformation of the sample variables destroys their independence (see e.g. David and Johnson (1948)). Thus Kolmogorov's argument leading to Equation (13) for his universal critical values becomes invalid. We expect for each of our three cases to have different critical values, and in Case I we should recover Kolmorogorov's values.

Making use Equation (1) for F and of the τ_i's representation as given by Equation (30), Equation (13) can be written as :

$$\begin{aligned} D_n &= \max_{1\le i\le n}\left\{\frac{i}{n} - 1 + \exp\left[\left(\frac{\tau_L}{\hat\alpha}\right)^{\hat\beta} - \left(\frac{\tau_i}{\hat\alpha}\right)^{\hat\beta}\right], 1 - \exp\left[\left(\frac{\tau_L}{\hat\alpha}\right)^{\hat\beta} - \left(\frac{\tau_i}{\hat\alpha}\right)^{\hat\beta}\right] - \frac{i-1}{n}\right\} \\ &= \max_{1\le j\le n}\left\{\frac{i-n}{n} + \exp\left[\hat\eta - (\eta + y_i)^{\hat\beta/\beta^0}\left(\frac{\alpha^0}{\hat\alpha}\right)^{\hat\beta}\right], \frac{n+1-i}{n} - \exp\left[\hat\eta - (\eta+y_i)^{\hat\beta/\beta^0}\left(\frac{\alpha^0}{\hat\alpha}\right)^{\hat\beta}\right]\right\} \\ &= \max_{1\le i\le n}\left\{\frac{i-n}{n} + \exp\left[\left\{\eta^{\hat\beta/\beta^0} - (\eta+y_i)^{\hat\beta/\beta^0}\right\}\left(\frac{\alpha^0}{\hat\alpha}\right)^{\hat\beta}\right],\right. \\ &\qquad \left.\frac{n+1-i}{n} - \exp\left[\left\{\eta^{\hat\beta/\beta^0} - (\eta+y_i)^{\hat\beta/\beta^0}\right\}\left(\frac{\alpha^0}{\hat\alpha}\right)^{\hat\beta}\right]\right\}, \end{aligned} \tag{15}$$

where $\hat\alpha$ and $\hat\beta$ are the estimated parameters while α^0 and β^0 are the true ones, $\eta \equiv (\tau_L/\alpha^0)^{\beta^0}$ and likewise $\hat\eta \equiv (\tau_L/\hat\alpha)^{\hat\beta}$ and y_i's are standard exponential random variates, as described in Appendix A. Equation (15) describes the modified critical values for all four cases above.

The critical values in general are a function of the sample size n only when the untruncated data set is considered. But clearly, they also depend on the truncation parameters, such as the truncation level p or truncation parameter η, when truncated data is considered. However, for two cases we find simplified relations for D_n, which are independent of the truncation parameter τ_L or η and also independent of the true values of α^0 and β^0 :

Case I : ($\hat\eta = \eta$, $\hat\alpha = \alpha^0$ and $\hat\beta = \beta^0$)

$$D_n = \max_{1\le i\le n}\left\{\frac{i-n}{n} + \exp(-y_i), \ \frac{n+1-i}{n} - \exp(-y_i)\right\} . \tag{16}$$

Case IIIa : $(\hat{\beta} = \beta^0)$

$$D_n = \max_{1\le i\le n} \left\{ \frac{i-n}{n} + \exp\left[-y_i \left(\frac{\alpha^0}{\hat{\alpha}}\right)^{\beta^0}\right], \frac{n+1-i}{n} - \exp\left[-y_i \left(\frac{\alpha^0}{\hat{\alpha}}\right)^{\beta^0}\right]\right\}. \quad (17)$$

One observes that in Case IIIa when the shape parameter β^0 is known, Equation (15) simplifies to Equation (17) and becomes independent of truncation, τ_L (or η), because of $\hat{\beta}/\beta^0 = 1$ the η-terms cancel each other out.

To construct confidence tables without loss of generality one may assume $(\alpha^0, \beta^0) = (1, 1)$ and hence $\eta = \tau_L$. Following Thoman, Bain, and Antle (1969) we denote for general Weibull distributions with any positive (α^0, β^0) the random variables $(\alpha^0/\hat{\alpha})^{\hat{\beta}}$ and $\hat{\beta}/\beta^0$ as *pivotal functions*. Note that the KS distance D_n in Equation (15) depends on these pivotal functions, n and η. Consequently, D_n is "universal" for different combinations of (α^0, β^0) for the same n and η, provided the following holds true

$$\left(\frac{\alpha^0}{\hat{\alpha}_{(\alpha^0,\beta^0)}}\right)^{\hat{\beta}_{(\alpha^0,\beta^0)}} \overset{distrib.}{=} \left(\frac{1}{\hat{\alpha}_{(1,1)}}\right)^{\hat{\beta}_{(1,1)}} \quad , \quad \left(\frac{\hat{\beta}_{(\alpha^0,\beta^0)}}{\beta^0}\right) \overset{distrib.}{=} \hat{\beta}_{(1,1)} \quad (18)$$

where $\hat{\alpha}_{(1,1)}$ and $\hat{\beta}_{(1,1)}$ are the MLE estimates originating from the simplest choice of a Weibull distribution with $(\alpha^0, \beta^0) = (1, 1)$. Likewise $\hat{\alpha} = \hat{\alpha}_{(\alpha^0,\beta^0)}$ and $\hat{\beta} = \hat{\beta}_{(\alpha^0,\beta^0)}$ are the MLE estimates originating from a Weibull distribution for arbitrary positive (α^0, β^0). The latter equality in distribution, Equation (18), was demonstrated in Appendix 3 of Kreer *et al.* (2015). For untruncated data where $x_L = 0$ (thus η and p vanish), $D_{cv}(n, 0, 0.05)$ will only depend on n. This was observed by Thoman *et al.* (1969) and allowed Littell *et al.* (1979) and Parsons and Wirsching (1982) the production of confidence tables for in-sample KS tests with MLE equations solved for exponential random variates. Similarly, in Case IIIa when the shape parameter β is known, Equation (15) simplifies and becomes independent of truncation, x_L (and hence of the truncation level $p = 1 - e^{-\eta}$), because $\hat{\beta}/\beta^0 = 1$ and the η-terms cancel out. Only for Case II and Case IIIb do we need to investigate the dependence of $D_{cv}(n, p, 0.05)$ on the parameter η ($\eta = x_L$ if $(\alpha^0, \beta^0) = (1, 1)$) and n in greater detail.

4. Brownian bridge asymptotics for Kolmogorov-Smirnov goodness of fit tests

4.1. Brownian bridge and Donsker's theorem

As in the discussion of the MLE in section 2 it will be interesting to consider what happens to the KS test when $n \to \infty$. The asymptotic behaviour of the KS-test has been of interest from the 1940s onwards, Durbin (1973), Stephens (1977), and Shorack and Wellner (2009), in calculating the asymptotic critical values. For a random variable $\boldsymbol{\tau}$ distributed according to a theoretical Weibull distribution function $F(\boldsymbol{\tau}|\theta^0)$, one may define the difference between the theoretical (with or without estimated parameters $\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_n, \hat{\beta}_n)$) and empirical distributions as (Durbin (1973), Equation (2))

$$G_n(t) \quad = \quad \sqrt{n}\left[\hat{F}_n(t) - t\right] \quad (19)$$

where $\hat{F}_n(t)$ is the proportion of $\tau_1, \tau_2, \cdots, \tau_n$, i.i.d. for which $F(\tau_i|\hat{\theta}_n) \le t$, $t \in [0, 1]$, and $\hat{\boldsymbol{\theta}}_n$ is the MLE estimate for the true parameter $\boldsymbol{\theta^0} = (\alpha^0, \beta^0)$. Note that taking the absolute value of the supremum in Equation (19) would yield the KS-distance in Equation (13). Viewing Equation (19) as a stochastic process in $t \in [0, 1]$, Doob's Theorem (also known as the functional central limit theorem) asserts the convergence in distribution against a limiting stochastic process which is Gaussian with zero mean and the covariance structure of a

Brownian bridge (see Shorack and Wellner (2009)). For the case where the parameters are estimated from the sample itself a modification (due to Durbin (1973)) has to be made. We may apply Theorem 2 of Durbin (1973) (here $\boldsymbol{\theta} = (\alpha, \beta)$), where the limiting Gaussian process is denoted, in analogy from above, by $G_n(t)$ with mean of 0, i.e. $\mathbb{E}(G_n(t)) = 0$ and a covariance structure given by

$$C(s,t) = \mathbb{E}(G_n(s)G_n(t)) = \min(s,t) - s \cdot t - u^T(s)\, \Sigma\, u(t), \quad 0 \le s \le t \le 1 \tag{20}$$

where $\Sigma = Z^{-1}$ is the inverse of the Fisher information matrix $Z(\alpha, \beta)$ given in Equation (11), and $u(\cdot)$ are certain vector-valued functions given by Equation (21) below. Note that the supremum of this Gaussian process using only n points will converge to the asymptotic value of the KS-distance D_n, as given in Equation (13) of the previous section, when $n \to \infty$. This will be key in deriving the asymptotic values. We readily check that Durbin's assumptions (A2) and (A3) in Durbin (1973) are also satisfied for the truncated case with truncation point $\tau_L > 0$, so that Theorem 2 of Durbin (1973) may be applied. Stephens (1977) studies the Brownian bridge with the covariance structure given in Equation (20) for complete data, (i.e. $\tau_L = 0$). The vector-valued function $u(s)$ in Equation (20) for left-truncated Weibull distributions with $\tau_L > 0$ is

$$u(s) \equiv \begin{pmatrix} \frac{\partial s}{\partial \alpha} \\ \frac{\partial s}{\partial \beta} \end{pmatrix} = \begin{pmatrix} \frac{\beta}{\alpha} s \log s \\ -\frac{s}{\beta} \left\{ \eta \log \eta - (\eta - \log s) \log(\eta - \log s) \right\} \end{pmatrix}, \tag{21}$$

where $s = F(\tau) = F(\tau|\alpha, \beta, \tau_L)$. In the following calculations, without loss of generality, we may choose for convenience $(\alpha, \beta) = (1, 1)$. Using the covariance equations, Equation (20), together with Equation (21) and the matrix Σ as the inverse of Z, from Equation (11), we can now for any $m \in \mathbb{N}$ simulate a Brownian bridge with discrete values $t_i = i/m$ with the given discrete covariance structure $C_{i,j} = C(s = i/m, t = j/m)$, for $i, j = 0, 1, ..., m$.

4.2. Numerical implementation of the Brownian bridge

We perform the following procedure as described in Anderson and Stephens (1997) to calculate the critical values in the Brownian bridge approach :

1. Discretise the interval $[0,1]$ for given $m \in \mathbb{N}$, in discrete values $s = i/m, t = j/m$, $i, j = 0, 1, ..., m$.
2. The discrete covariance matrix $C^{(m+1)} = (C_{i,j})_{i,j}$ from Equation (20) now has entries
$$C_{i,j} = \min(i,j)/m - i/m \cdot j/m \, - \, u^T(i/m)\, \Sigma\, u(j/m) \tag{22}$$
and is symmetric and positive definite.
3. Calculate the Cholesky decomposition $C^{(m+1)} = BB^T$, where $B = B^{(m+1)}$ is a triangular matrix of dimension $(m+1) \times (m+1)$.
4. Draw $(\zeta_0, \zeta_1, ..., \zeta_m)$ standard normally distributed numbers (i.e. mean 0 and variance 1). Set $\zeta_0 = 0$ and $\zeta_m = 0$ and define the vector $z = (0, \zeta_1, ..., \zeta_{m-1}, 0)$.
5. The transformed $(m+1)$-vector Bz is a discrete representation of a Brownian bridge $G_m(t)$ starting at $t = 0$ with $G_m(0) = 0$ and ending at $t = 1$ with $G_m(1) = 0$. Find the following statistics
$D^+(m) = \max(Bz)$, $D^-(m) = -\min(Bz)$ and then set $D_m = \max\{D^+(m), D^-(m)\}$.
6. Keep D_m in a list and sort in ascending order. Take the 95% as a critical $D_m^{BB}(95\%)$.
7. Repeat procedure for $m = 30, 50, 100, 200, ...$ and fit D_m against the function $A + B/\sqrt{m}$ (see also Chandra *et al.* (1981)). The value A is the asymptotic value of the Kolmogorov-Smirnov statistic, $A = D_{cv}(\infty, 0.05)$.

4.3. Results: asymptotical critical values from Brownian bridge

We apply the Brownian Bridge (BB) approach to find the asymptotic critical values for the following cases and present the results in Table 2.

Case I Out-sample testing: Put $\Sigma = 0$ (because α and β are known precisely therefore Fisher information matrix is irrelevant here) and sample a pure Brownian bridge.

Case II In-sample testing for two unknown parameters (with truncation).

Case IIIa - IIIb In-sample testing with one-parameter known (with truncation): Get a one-dimensional Fisher Information matrix from Equation (11) with the unknown parameter and invert this element to obtain the corresponding Σ-matrix.

5. The quantile analysis for determining the critical values

5.1. The Monte-Carlo Algorithm

The quantile procedure to calculate the critical values is described below.

Algorithm 1: Procedure for calculating the mean and variance of the critical values of the KS-test

Input:
The values of α and β are both set to 1

Output: The mean and standard deviation of the critical values of the KS-test for a range of sample sizes n and truncation levels p, $\eta = \tau_L = \alpha\left(-\log\left(1-p\right)\right)^{\beta}$.

```
for p = 0 to 0.9 -STEP 0.1 do
    for n = 30, 50, 100, 200, 500, 1000, 10000 do
        for j = 1 to 100 do
            for k = 1 to 1000 do
                • Draw n random numbers u_i from a uniform distribution u_i ~ U(0,1). It follows
                  directly from the discussion in appendix A that the left-truncated Weibull
                  distributed random variables are   τ_i = τ_L − log u_i
                • Estimate α̂ and β̂ using MLE equations Equations (7) and (9).
                • Calculate the Kolmogorov-Smirnov statistic using Equation (13) and store it as
                  D(n, p, j, k)
                • Sort D(n, p, j, :)  ∀  k in ascending order. The 95% confidence interval, i.e.
                  (α_H = 0.05) is   D^q_cv(n, p, j) = 1/2 (D(n, p, j, 950) + D(n, p, j, 951))   .
            end
        end
        • Calculate the mean D^q_cv(n, p) and variance σ²_{D^q_cv(n,p)} from the 100 values.
    end
end
```

Table 2: The asymptotical critical values from BB approach for all cases.

Truncation Level	Truncation Parameter	Case I	Case II	Case IIIa	Case IIIb
p	η	D_{cv}^{BB}	D_{cv}^{BB}	D_{cv}^{BB}	D_{cv}^{BB}
0	0	1.356±0.008	0.901 ±0.007	1.093±0.005	1.317±0.006
0.1	0.1	1.359±0.004	0.862 ±0.002	1.094±0.003	1.329±0.006
0.2	0.2	1.358±0.005	0.860 ±0.007	1.095±0.003	1.321±0.006
0.3	0.35	1.358±0.008	0.860 ±0.006	1.094±0.003	1.291±0.005
0.4	0.5	1.359±0.005	0.874 ±0.003	1.095±0.005	1.260±0.006
0.5	0.7	1.358±0.003	0.880 ±0.003	1.094±0.005	1.234±0.006
0.6	0.9	1.361±0.004	0.879 ±0.006	1.094±0.002	1.198±0.003
0.7	1.2	1.357±0.005	0.892 ±0.007	1.094±0.002	1.183±0.004
0.8	1.6	1.358±0.006	0.900 ±0.007	1.093±0.001	1.163±0.004
0.9	2.3	1.359±0.005	0.909 ±0.007	1.093±0.006	1.142±0.003

5.2. Results: critical values from Monte-Carlo simulations

Our results obtained for the modified critical values using the quantile analysis (outlined in Algorithm 1) for each sample size $n = (30, 50, 100, 500, 1000, 10000)$ and truncation parameter

η are summarised in Table 3 for Case I, in Table 4 for Case II, in Table 5 for Case IIIa, and in Table 6 for Case IIIb.

Table 3: The critical values, D^q_{cv}, calculated from the quantile analysis for Case I.

p	η	$D^q_{n=30}$	$D^q_{n=50}$	$D^q_{n=100}$	$D^q_{n=200}$	$D^q_{n=500}$	$D^q_{n=1000}$	$D^q_{n=10000}$
0	0	1.322±0.025	1.329±0.024	1.336±0.024	1.343±0.024	1.346±0.024	1.346±0.023	1.354±0.027
0.1	0.1	1.321±0.024	1.333±0.023	1.339±0.026	1.345±0.026	1.348±0.027	1.351±0.024	1.352±0.021
0.2	0.2	1.321±0.025	1.327±0.024	1.339±0.025	1.345±0.023	1.344±0.025	1.352±0.027	1.351±0.026
0.3	0.35	1.322±0.023	1.335±0.028	1.341±0.025	1.349±0.025	1.349±0.024	1.350±0.025	1.359±0.026
0.4	0.5	1.319±0.026	1.330±0.027	1.338±0.026	1.345±0.026	1.347±0.024	1.356±0.025	1.352±0.026
0.5	0.7	1.322±0.024	1.331±0.024	1.334±0.024	1.345±0.028	1.349±0.022	1.356±0.025	1.353±0.028
0.6	0.9	1.322±0.025	1.331±0.024	1.340±0.023	1.343±0.024	1.349±0.028	1.352±0.026	1.357±0.026
0.7	1.2	1.322±0.027	1.330±0.023	1.339±0.027	1.345±0.024	1.346±0.023	1.350±0.025	1.359±0.025
0.8	1.6	1.319±0.029	1.330±0.025	1.338±0.021	1.345±0.029	1.348±0.024	1.351±0.025	1.355±0.026
0.9	2.3	1.323±0.024	1.328±0.023	1.340±0.026	1.348±0.024	1.346±0.023	1.349±0.024	1.354±0.026

Table 4: The critical values, D^q_{cv}, from the quantile analysis for Case II.

p	η	$D^q_{n=30}$	$D^q_{n=50}$	$D^q_{n=100}$	$D^q_{n=200}$	$D^q_{n=500}$	$D^q_{n=1000}$	$D^q_{n=10000}$
0	0	0.858±0.011	0.865±0.012	0.874± 0.012	0.881±0.013	0.887±0.012	0.890±0.015	0.893±0.015
0.1	0.1	0.817±0.012	0.829±0.011	0.838±0.012	0.843±0.013	0.850±0.013	0.851±0.013	0.857±0.012
0.2	0.2	0.815±0.012	0.824±0.011	0.838±0.012	0.842±0.012	0.847±0.013	0.852±0.012	0.856±0.011
0.3	0.35	0.818±0.013	0.830±0.012	0.840±0.010	0.848±0.013	0.854±0.013	0.856±0.012	0.859±0.012
0.4	0.5	0.821±0.012	0.832±0.011	0.846±0.011	0.853±0.012	0.857±0.012	0.862±0.013	0.866±0.013
0.5	0.7	0.824±0.013	0.840±0.013	0.852±0.013	0.860±0.012	0.863±0.011	0.868±0.014	0.872±0.012
0.6	0.9	0.830±0.012	0.844±0.013	0.857±0.012	0.866±0.012	0.873±0.011	0.876±0.013	0.881±0.011
0.7	1.2	0.835±0.012	0.853±0.012	0.864±0.013	0.873±0.012	0.878±0.013	0.882±0.013	0.888±0.012
0.8	1.6	0.839±0.012	0.855±0.012	0.871±0.013	0.880±0.013	0.886±0.014	0.890±0.015	0.894±0.014
0.9	2.3	0.843±0.012	0.864±0.011	0.880±0.014	0.890±0.014	0.897±0.014	0.897±0.013	0.904±0.014

Table 5: The critical values, D^q_{cv}, from the quantile analysis for Case IIIa.

p	η	$D^q{}_{n=30}$	$D^q{}_{n=50}$	$D^q{}_{n=100}$	$D^q{}_{n=200}$	$D^q{}_{n=500}$	$D^q{}_{n=1000}$
0	0	1.055±0.020	1.064±0.018	1.072±0.020	1.080±0.021	1.083±0.016	1.086±0.018
0.1	0.1	1.054±0.019	1.064±0.019	1.074±0.019	1.078±0.018	1.085±0.017	1.085±0.018
0.2	0.2	1.058±0.018	1.064±0.018	1.074±0.018	1.080±0.019	1.085±0.018	1.086±0.022
0.3	0.35	1.057±0.016	1.064±0.019	1.075±0.016	1.081±0.019	1.082±0.017	1.085±0.019
0.4	0.5	1.054±0.018	1.066±0.019	1.072±0.019	1.079±0.019	1.083±0.017	1.085±0.016
0.5	0.7	1.054±0.017	1.066±0.021	1.074±0.021	1.080±0.021	1.084±0.018	1.083±0.019
0.6	0.9	1.057±0.017	1.065±0.018	1.077±0.020	1.077±0.018	1.086±0.020	1.086±0.019
0.7	1.2	1.056±0.018	1.065±0.018	1.075±0.019	1.083±0.018	1.084±0.017	1.085±0.019
0.8	1.6	1.056±0.021	1.064±0.018	1.076±0.020	1.080±0.021	1.082±0.019	1.086±0.017
0.9	2.3	1.057±0.017	1.065±0.017	1.074±0.018	1.077±0.017	1.084±0.020	1.088±0.019

Table 6: The critical values, D^q_{cv}, from the quantile analysis for Case IIIb.

p	η	$D^q{}_{n=30}$	$D^q{}_{n=50}$	$D^q{}_{n=100}$	$D^q{}_{n=200}$	$D^q{}_{n=500}$	$D^q{}_{n=1000}$
0	0	1.281±0.024	1.289±0.022	1.301±0.024	1.302±0.028	1.310±0.025	1.310±0.023
0.1	0.1	1.301±0.026	1.307±0.023	1.314±0.023	1.320±0.024	1.322±0.025	1.323±0.027
0.2	0.2	1.283±0.026	1.293±0.023	1.299±0.026	1.308±0.025	1.306±0.026	1.307±0.026
0.3	0.35	1.255±0.023	1.262±0.025	1.270±0.023	1.273±0.024	1.284±0.023	1.281±0.024
0.4	0.5	1.224±0.022	1.233±0.021	1.241±0.022	1.246±0.024	1.250±0.027	1.257±0.022
0.5	0.7	1.194±0.021	1.203±0.022	1.212±0.020	1.214±0.021	1.223±0.022	1.227±0.023
0.6	0.9	1.171±0.020	1.177±0.019	1.189±0.023	1.193±0.020	1.194±0.023	1.198±0.023
0.7	1.2	1.144±0.020	1.154±0.021	1.162±0.023	1.169±0.021	1.174±0.021	1.179±0.022
0.8	1.6	1.122±0.019	1.136±0.021	1.142±0.024	1.148±0.020	1.153±0.022	1.154±0.021
0.9	2.3	1.100±0.019	1.110±0.021	1.125±0.022	1.124±0.020	1.131±0.018	1.136±0.019

6. Discussion of the results

The truncation in the analysis can be defined three equivalent ways: 1) τ_L, the value below which all data is removed/absent, 2) p the percentage of data removed/absent by the truncation procedure, and 3) the generalised truncation parameter $\eta \equiv \left(\frac{\tau_L}{\alpha}\right)^\beta$. These three parameters are related by the equations

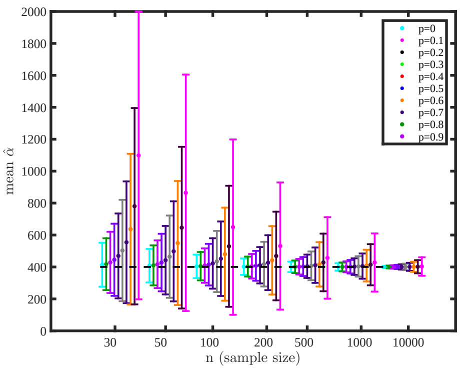
$$p = 1 - e^{-\left(\frac{\tau_L}{\alpha}\right)^\beta} = 1 - e^{-\eta} \quad . \tag{23}$$

All of these parameters will be used throughout this paper, depending on which is the most convenient.

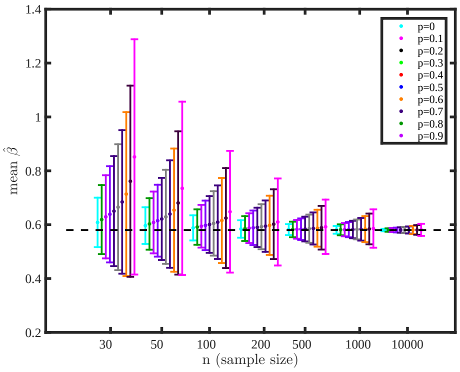
6.1. Estimation of the Weibull Parameters

Identifying and analyzing the distribution which represents the data set is our main focus, since it is the source of the predictability. Figure (1) is an errorbar plot of the MLE estimates of the parameters $\hat{\alpha}$ (Upper left, in Case II and Lower left in Case IIIa) and $\hat{\beta}$ (Upper right in Case II, Lower right in Case IIIb) for various sample sizes n and truncation levels p. Here the true values are taken as $\alpha^0 = 400$ and $\beta^0 = 0.58$. From these plots one can see that as the sample size increases the variance in the estimation of $\hat{\alpha}$ and $\hat{\beta}$ decreases in all cases. Furthermore as the truncation level increases the variance in estimation of $\hat{\alpha}$ and $\hat{\beta}$ increases continuously in Case II, while in Case IIIb it increases initially then decreases. Finally, in Case IIIa the estimation of the parameter is totally insensitive to the truncation, see Figure(1c).

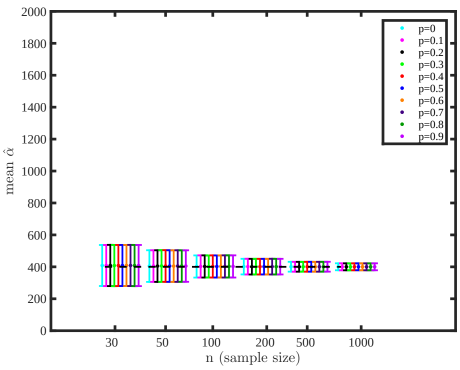
In summary, the estimation is better in Cases II and IIIb when the sample size is larger and the truncation is smaller. The single parameter estimates are far better than the double parameter estimates as expected. Case IIIa, where the shape parameter β is known, and the scale parameter α is unknown, is superior to Case IIIb with the unknown shape parameter β and known scale parameter α, since in Case IIIa the CVs are independent of truncation and the estimation of α is more precise, which is the optimum scenario. Comparisons on estimation of parameters show that the variance is reduced by 75% in α and by 50% in β between the two parameter and one parameter cases.



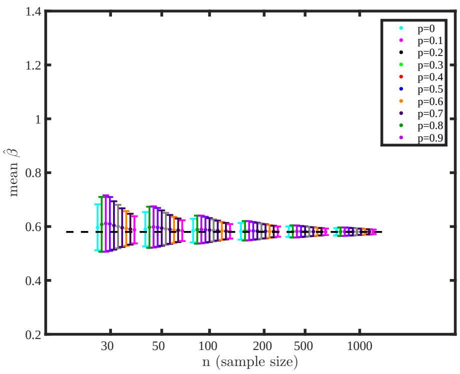
(a) $\hat{\alpha}$ – Case II.



(b) $\hat{\beta}$ – Case II.



(c) $\hat{\alpha}$ – Case IIIa.



(d) $\hat{\beta}$ – Case IIIb.

Figure 1: The mean value of the MLE of Weibull parameters α and/or β as a function of n and p (the percentage data removed by truncation). The error bars show one standard deviation in the estimated values of the parameters. The horizontal dashed line shows the true value of the parameters that was used to generate the data for $\alpha^0 = 400$ and $\beta^0 = 0.58$.

6.2. Critical Values as a Function of Sample Size n

Figure 2 depicts, for Cases II and IIIb, the dependence of the critical values, (given in Table 4 and Table 6) on n for a range of truncation levels, p (or truncation parameter η). For clarity the $x-$axis is plotted on a log scale. Both the cases show a distinctive separation between the lines for different truncation levels, indicating a dependence on the truncation level p.

On the other hand in Case I, the critical values are independent of truncation and only depend on n. As predicted by the theory, Equation (17), we also note that truncation has no noticeable effect on the critical values in Case IIIa as well. According to Miller's formula, Miller (1956), which was derived for the out sample, untruncated case namely Case I, the critical values are quadratic in $1/\sqrt{n}$

$$D_{cv}(n) = \sqrt{-\frac{1}{2}\log\frac{\alpha_H}{2}} - \frac{0.167}{\sqrt{n}} - \frac{\mathcal{A}}{n} \quad \text{for } n > 20 \quad (\alpha_H = 0.05,\ 95\% \text{ confidence level}) . \tag{24}$$

where the first term in above expression is Simirnov's asymptotic formula and calculated as 1.358 and

$$\begin{aligned}\mathcal{A} &\equiv 0.090\left(-\log_{10}\frac{\alpha_H}{2}\right)^{3/2} + 0.015\left(\log_{10}\frac{\alpha_H}{2}\right)^2 - 0.085\frac{\alpha_H}{2} - 0.111 \\ &= 0.109 .\end{aligned}$$

Although Miller's formula, Equation (24), is designed to be used for only Case I, where both the parameters are known a priori, we will however use it as a guide to investigate the functional dependence of the critical values on the sample size n for all cases. This can be achieved by fitting the critical values given in Tables 3 - 6 for each value of p to the function

$$D_{cv}(p|n) = \tilde{A}(p) + \frac{\tilde{B}(p)}{\sqrt{n}} + \frac{\tilde{C}(p)}{n} \quad . \tag{25}$$

Table 7: Results of fitting $D^q_{cv}(p|n)$ to quadratic and linear functions in $1/\sqrt{n}$. The critical values, D^q_{cv}, obtained as a function of sample size n from the quadratic fit on to left-truncated data of Case I for each truncation level p, truncation parameter η. $\tilde{A}(p|n)$, $\tilde{B}(p|n)$ and $\tilde{C}(p|n)$ are the fit parameters in Equation (25), $\tilde{A}_1(p|n)$, $\tilde{B}_1(p|n)$ in Equation (26).

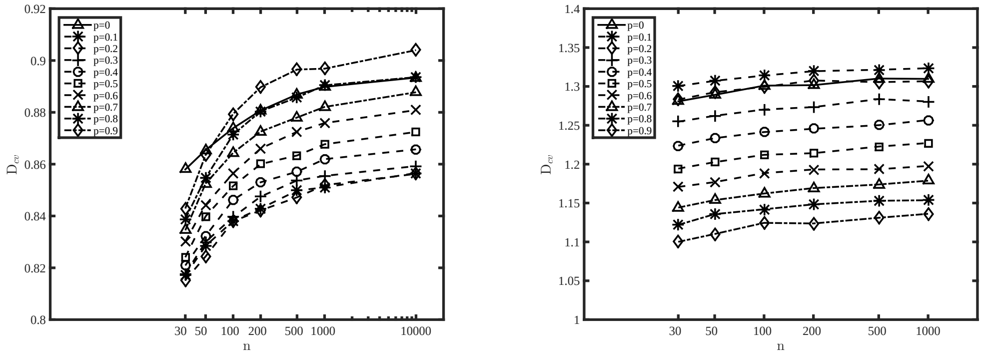
p	η	$\tilde{A}(p)$	$\tilde{B}(p)$	$\tilde{C}(p)$	$\tilde{A}_1(p)$	$\tilde{B}_1(p)$
0	0	1.355±0.004	-0.193±0.117	0.087±0.588	1.354±0.002	-0.177±0.024
0.1	0.1	1.353±0.003	-0.086±0.081	-0.484±0.408	1.356±0.003	-0.179±0.032
0.2	0.2	1.354±0.009	-0.128±0.223	-0.297±1.124	1.355±0.005	-0.185±0.048
0.3	0.35	1.357±0.008	-0.125±0.203	-0.354±1.026	1.359±0.005	-0.193±0.046
0.4	0.5	1.355±0.008	-0.116±0.221	-0.468±1.116	1.358±0.005	-0.206±0.052
0.5	0.7	1.358±0.010	-0.200±0.265	0.004±1.334	1.358±0.005	-0.199±0.054
0.6	0.9	1.359±0.004	-0.207±0.094	0.043±0.475	1.359±0.002	-0.199±0.019
0.7	1.2	1.359±0.007	-0.228±0.170	0.178±0.858	1.358±0.004	-0.194±0.036
0.8	1.6	1.356±0.002	-0.154±0.046	-0.266±0.234	1.358±0.002	-0.205±0.018
0.9	2.3	1.355±0.008	-0.128±0.218	-0.290±1.100	1.356±0.005	-0.184±0.047

The fit results are tabulated in Table 7 for Case I where the values of $\tilde{C}(p)$ are quite variable and the standard deviation in $\tilde{C}(p)$ is greater than the value itself. This suggests that $D^q_{cv}(p|n)$ is better approximated by a function that is linear in $1/\sqrt{n}$ instead of quadratic, i.e.,

$$D_{cv}(p|n) = \tilde{A}_1(p) + \frac{\tilde{B}_1(p)}{\sqrt{n}} . \tag{26}$$

Table 8: The critical values obtained from the quantile analysis fitted to the linear function for a range of truncation level p for Case II, Case IIIa and Case IIIb. $\tilde{A}_1(p|n)$, $\tilde{B}_1(p|n)$ are the fit parameters defined in Equation (26).

		Case II		Case IIIa		Case IIIb	
p	η	$\tilde{A}_1(p)$	$\tilde{B}_1(p)$	$\tilde{A}_1(p)$	$\tilde{B}_1(p)$	$\tilde{A}_1(p)$	$\tilde{B}_1(p)$
0	0	0.896±0.001	-0.211±0.012	1.093±0.002	-0.204±0.022	1.318±0.005	-0.197±0.047
0.1	0.1	0.859±0.002	-0.223±0.018	1.093±0.004	-0.207±0.034	1.329±0.002	-0.154±0.020
0.2	0.2	0.859±0.002	-0.238±0.024	1.093±0.002	-0.194±0.021	1.314±0.007	-0.160±0.065
0.3	0.35	0.864±0.002	-0.243±0.023	1.092±0.004	-0.188±0.034	1.288±0.005	-0.186±0.048
0.4	0.5	0.870±0.003	-0.262±0.029	1.093±0.004	-0.203±0.038	1.261±0.004	-0.201±0.034
0.5	0.7	0.877±0.004	-0.273±0.039	1.093±0.006	-0.200±0.051	1.232±0.004	-0.208±0.040
0.6	0.9	0.885±0.002	-0.295±0.021	1.093±0.005	-0.196±0.041	1.204±0.005	-0.179±0.047
0.7	1.2	0.892±0.004	-0.298±0.039	1.093±0.005	-0.200±0.050	1.185±0.002	-0.222±0.016
0.8	1.6	0.900±0.005	-0.322±0.047	1.093±0.005	-0.195±0.045	1.162±0.005	-0.205±0.047
0.9	2.3	0.911±0.006	-0.346±0.062	1.093±0.004	-0.200±0.033	1.143±0.007	-0.227±0.061



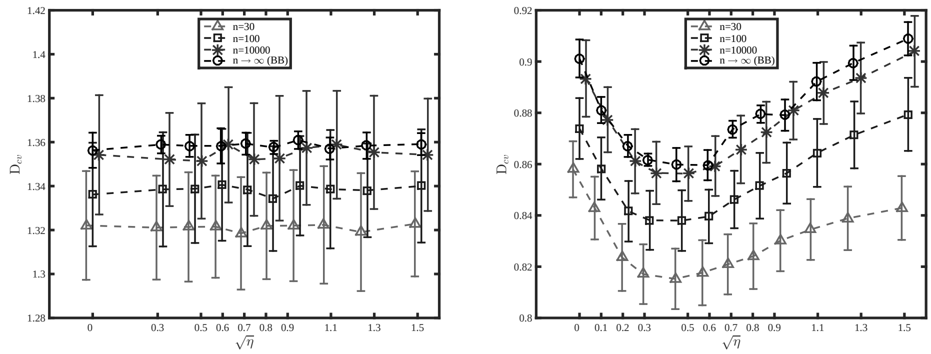
(a) Both α and β are unknown – Case II.

(b) β is unknown and α is known – Case IIIb.

Figure 2: Critical values as function of n for a range of truncation level p.

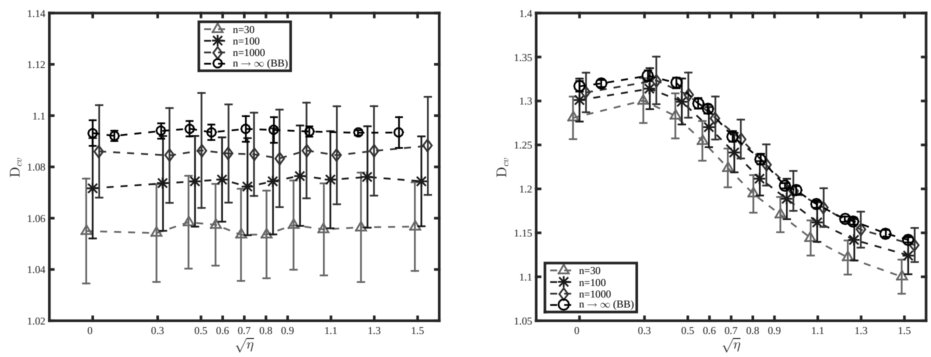
The linear function is a better fit to the data $D^q_{cv}(p|n)$, in the sense that there is no significant change in the *adjusted r-squared* goodness of fit statistic, but the standard deviation in $\tilde{B}$ over all values of p is an order of magnitude better when Equation (26) is used instead of Equation (25). The results for $\tilde{A}_1(p|n)$ and $\tilde{B}_1(p|n)$ given by fitting $D_{cv}(p|n)$ are in very good agreement with Miller's formula, Equation (24). The fit results are given in Table 8 for Case II, Case IIIa and Case IIIb, respectively.

6.3. Critical Values as a Function of Truncation Parameter eta



(a) Both α and β are known – Case I.

(b) Both α and β are unknown – Case II.



(c) β is known and α is not known – Case IIIa.

(d) β is unknown and α is known – Case IIIb.

Figure 3: The critical values as a function of $\sqrt{\eta}$ for a range of n values. The circled dashed line with the error bars are the Brownian Bridge calculation.

To determine the relationship between the critical values and the truncation parameter η, we plot in Figure 3 the critical values given in Tables 3-6, as a function of $\sqrt{\eta}$ for $n = (30, 100, 1000, 10000)$ for all cases. We have also included a plot of the Brownian Bridge results (with error bars) for all cases, since that provides an alternative way of estimating D_{cv} in the limit $n \to \infty$. For out-sample data the critical values are independent of truncation and this is verified in Figure 3a. We see that there is no variation in the critical values as a function of $\sqrt{\eta}$. On the other hand Figure 3b for Case II shows that the critical values initially decrease but then increase as the truncation level increases (boomerang shape), which is totally different behaviour from the out-sample case (Case I). In Figure 3c for Case IIIa the CV's do not change as η increases, similar to Case I in that for a fixed value of n the critical values are independent of the truncation. These results are consistent with the theory we outlined in Equation (17). Case IIIb in Figure 3d, on the other hand, shows that CV's initially slightly increase then decrease as the truncation level increases.

In summary, the CV's in Cases I and IIIa are truncation independent while in Cases II and IIIb they are not. For all cases the asymptotic critical value analysis from the Brownian Bridge confirms the same η dependence as we found in the quantile analysis. This section deals with formulating the critical values as a function of truncation parameter η. In both Case II and Case IIIb, the CV's are truncation dependent and among the many fit functions tried to describe the data we found that the quadratic ratio function

$$D^q_{cv}(\eta|n) \quad = \quad \frac{C(n) + B(n)\sqrt{\eta} + A(n)\eta}{E(n) + D(n)\sqrt{\eta} + \eta}, \; . \tag{27}$$

fited best. Its parameters are given in Tables 9 and plotted in Figs. 4c,g and 4d,h for $n = 30$ and $n = 10000$, respectively. In the figure the light shaded grey, tick band shows the error range on $D^q_{cv}(\eta|n)$ values whereas the darker shaded grey area between the dashed lines is the error band on the fit values. In addition the asymptotic critical values from the Brownian Bridge analysis (squares) are shown in the figures for only $n = 10,000$.

Table 9: The critical values obtained by fitting the ratio function to the data from the quantile analysis for various sample sizes, n. The fit parameters defined in Equation (27) are given for each n values and for Cases II and IIIb.

	Case II				
n	$A(n)$	$B(n)$	$C(n)$	$D(n)$	$E(n)$
30	0.870±0.008	-0.197±0.102	0.207±0.040	-0.182±0.131	0.241±0.046
50	0.902±0.017	-0.218±0.196	0.295±0.079	-0.184±0.252	0.340±0.091
100	0.933±0.026	-0.080±0.306	0.372±0.108	0.000±0.387	0.425±0.124
200	0.934±0.014	-0.279±0.150	0.359±0.060	-0.240±0.189	0.407±0.068
500	0.955±0.032	-0.118±0.340	0.407±0.126	-0.035±0.426	0.458±0.142
1000	0.940±0.017	-0.250±0.200	0.332±0.070	-0.209±0.247	0.374±0.079
10000	0.954±0.015	-0.207±0.169	0.383±0.062	-0.148±0.208	0.428±0.069
	Case IIIb				
n	$A(n)$	$B(n)$	$C(n)$	$D(n)$	$E(n)$
30	1.096±0.047	-1.079±0.395	0.896±0.156	-0.913±0.291	0.700±0.120
50	1.115±0.040	-1.150±0.353	0.910±0.146	-0.957±0.260	0.706±0.112
100	1.139±0.033	-1.274±0.306	0.955±0.137	-1.040±0.223	0.737±0.104
200	1.125±0.043	-1.135±0.380	0.918±0.154	-0.939±0.277	0.707±0.117
500	1.138±0.028	-1.209±0.252	0.952±0.107	-0.988±0.183	0.730±0.081
1000	1.137±0.023	-1.193±0.211	0.992±0.088	-0.973±0.153	0.759±0.067

6.4. The modified critical values as a function of n and eta

In this section both the sample size, n, and truncation dependence, η, are combined to give one formula for the critical values as a function of n and η.

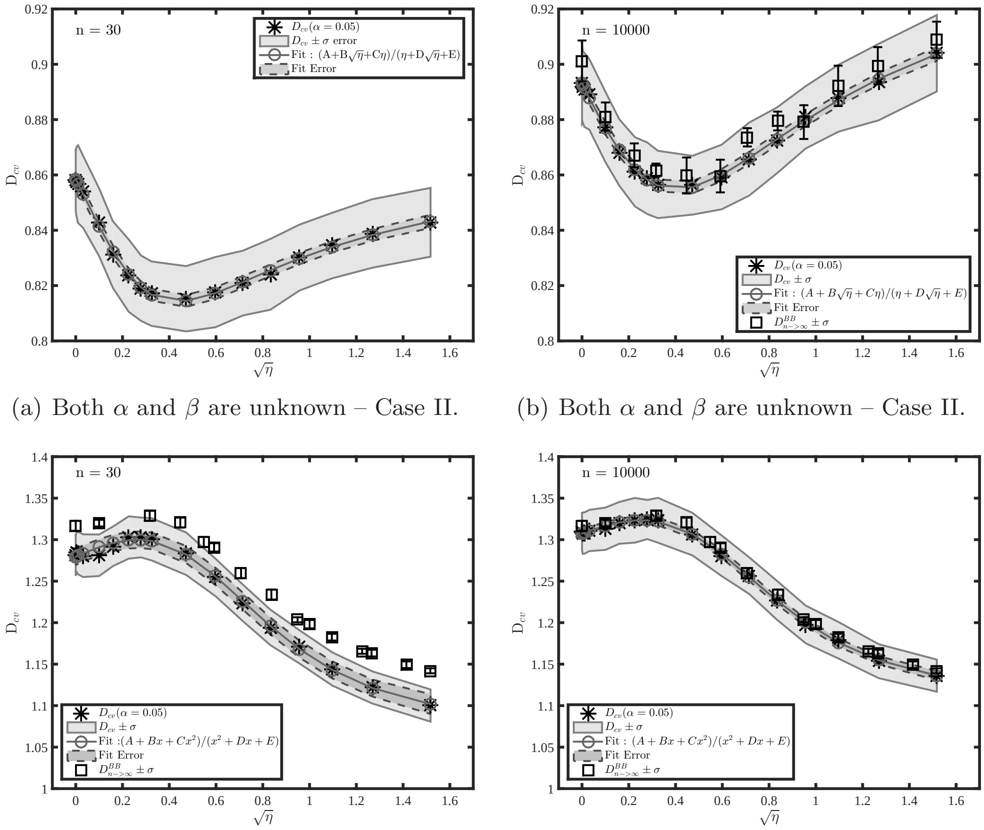
Case II and Case IIIb that both are sensitive to the truncation parameters. The critical values in Tables 4 and 6 can be fitted to the two dimensional function

$$D^q_{cv}(\eta, n) \quad = \quad A + \frac{B}{\sqrt{n}} + C\,\sqrt{\eta} + D\,\frac{\sqrt{\eta}}{\sqrt{n}} + E\,\eta + F\,\eta^{3/2}\,, \tag{28}$$

and the fit results are given in Table 10.

6.5. Exploring CV's for the Dependence of Weibull Parameter Ranges

This section numerically explores the effects of the range of the Weibull parameters on the critical values as discussed in section 3. For this purpose, we consider various combinations of the scale parameter $\alpha = 1, 400, 1000, 2000$ and shape parameter, $\beta = 0.2, 0.35, 0.58, 0.8, 1$.



(a) Both α and β are unknown – Case II. (b) Both α and β are unknown – Case II.

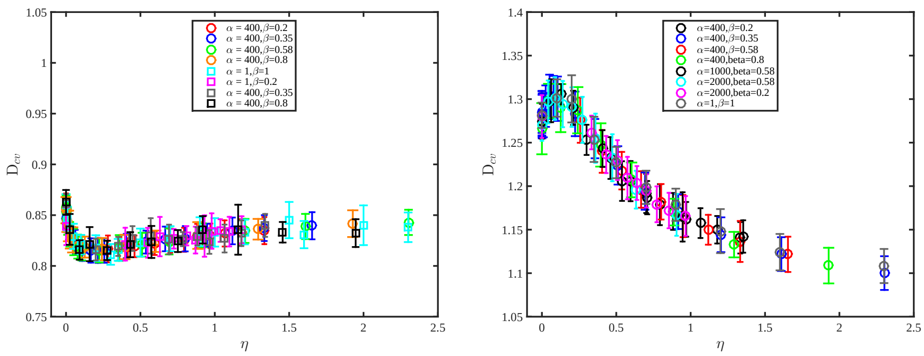
(c) β is unknown and α is known – Case IIIb. (d) β is unknown and α is known – Case IIIb

Figure 4: Critical values obtained from the quantile analysis and their fits are plotted as a function of $\sqrt{\eta}$ for a sample sizes $n = 30(left), 10,000(right)$.

Table 10: The fit parameters in Equation (28) are presented here for Cases II and IIIb.

Case II					
A	*B*	*C*	*D*	*E*	*F*
0.894±0.001	-0.196±0.01	-0.178±0.007	-0.096±0.02	0.263±0.012	-0.092±0.006
Case IIIb					
A	*B*	*C*	*D*	*E*	*F*
1.311± 0.003	-0.164± 0.028	0.187±0.17	-0.036±0.041	-0.495±0.028	0.198±0.013

The results are displayed in Figure 5, where the critical values plotted as a function η for sample sizes $n = 30(left)$ for Case II, Case IIIb. All the curves for different parameter combinations overlap with each other to show the insensitivity to different parameter values. In Case I and IIIa the CV's are independent of parameter, as is well known.



(a) n=30 – Case II. (b) n=30 – Case IIIb.

Figure 5: Critical values versus η for various combinations of α and β. $\alpha = 1, 400, 1000, 2000$, $\beta = 0.2, 0.35, 0.58, 0.8, 1$ ranges for $n = 30$.

6.6. Comparison of the results with literature

Comparison of our CV's with those already published are shown in Tables 11-14. We can see that there is excellent agreement. All of the previous studies in the literature only considered complete (untruncated) data, whereas our study considers a range of truncations, including the untruncated case. Therefore, we can only compare the complete case results with the literature. Also, we wish to remind the reader that the Weibull distribution is a special case of the generalised extreme value distribution.

Table 11: Comparison of our results with the available literature for Case I (α and β are known). To the best of our knowledge, no data is available for the CV's of left-truncated Weibull distribution. For *complete* data sets, ($\tau_L = 0 = \eta = 0 = p = 0$, our error is ± 0.025).

Authors	Estimation	Distribution	n	$D_{cv}(95\%)$	Our Results
Smirnov 1948 Smirnov (1948)	-	all	∞	1.36	1.356
Massey 1951 Massey (1951)	-	all	30	1.32	1.323
	-	all	∞	1.36	1.356
Birnbaum 1952 Birnbaum (1952)	-	all	30	1.3238	1.323
	-	all	50	1.3322	1.329
	-	all	100	1.3400	1.337
	-	all	∞	1.3581	1.356
Miller 1956 Miller (1956)	-	all	30	1.324	1.323
	-	all	50	1.332	1.329
	-	all	100	1.340	1.337
	-	all	∞	1.358	1.356

Table 12: Comparison of our results with the available literature for Case II (α and β are unknown). To the best of our knowledge, no CV's of left-truncated Weibull distributions are available. For *complete* data sets, ($\tau_L = 0 = \eta = 0 = p = 0$, our error is ± 0.015).

Authors	Estimation	Distribution	n	$D_{cv}(95\%)$	Our Results
Littell et al. 1979 Littell *et al.* (1979)	MLE	Weibull	30	0.854	0.858
Parsons & Wirsching 1982 Parsons and Wirsching (1982)	MLE	Weibull	30	0.854	0.858
	MLE	Weibull	∞	0.865	0.896
Chandra et al. 1981 Chandra *et al.* (1981)	MLE	extreme value	50	0.856	0.865
	MLE	extreme value	∞	0.874	0.896
D'Agostino & Stephens 1986Agostino and Stephens (1986)	MLE	extreme value	50	0.856	0.865
	MLE	extreme value	∞	0.874	0.896
Evans et al. 1989 Evans, Johnson, and Green (1989)	MLE	Weibull	30	0.8599	0.858
	MLE	Weibull	50	0.8697	0.865
	MLE	Weibull	100	0.8740	0.874
	MLE	Weibull	200	0.8796	0.881
	MLE	Weibull	∞	0.8982	0.896

Table 13: Comparison of our results with the available literature for Case IIIa (α unknown and β known). To the best of our knowledge, no CV's of left-truncated Weibull distributions are available. For *untruncated (complete)* data sets, ($\tau_L = 0 = \eta = 0 = p = 0$, our error is ± 0.020).

Authors	Estimation	Distribution	n	$D_{cv}(95\%)$	Our Results
Lilliefors 1969 Lilliefors (1969)	MLE	Exponential	30	1.052	1.055
	MLE	Exponential	∞	1.060	1.093
Durbin 1975 Durbin (1975)	MLE	Exponential	30	1.0580	1.055
	MLE	Exponential	50	1.0668	1.065
	MLE	Exponential	100	1.0753	1.073
Chandra et al. 1981 Chandra *et al.* (1981)	MLE	extreme value	50	1.067	1.064
	MLE	extreme value	∞	1.094	1.093
Woodruff et al 1983 Woodruff, Moore, Dunne, and Cortes (1983)	MLE	Weibull	30	1.057	1.055
D'Agostino & Stephens 1986 Agostino and Stephens (1986)	MLE	Exponential	50	1.061	1.065
	MLE	Exponential	100	1.072	1.073
	MLE	Exponential	∞	1.094	1.093
Shorack & Wellner p 239 Shorack and Wellner (2009)	MLE	Exponential	∞	1.094	1.093

Table 14: Comparison of our results with the available literature for Case IIIb (α known and β unknown). To the best of our knowledge, no CV's of left-truncated Weibull distributions are available. For *untruncated (complete)* data sets, ($\tau_L = 0 = \eta = 0 = p = 0$, our error is ± 0.025).

Authors	Estimation	Distribution	n	$D_{cv}(95\%)$	Our Results
D'Agostino & Stephens 1986 Agostino and Stephens (1986)	MLE	extreme value	50	1.29	1.289
	MLE	extreme value	∞	1.29	1.317

Table 16: Percentage pass rates in KS-test with and without $\hat{\eta}$-correction for 10000 simulations in Case IIIb (error is less than $\pm 0.5\%$).

p	η	n = 30		n=50		n=100		n=1000	
		$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected	$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected	$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected	$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected
0	0	95.6	95.6	95.2	95.2	95.0	95.0	95.2	95.2
0.1	0.1	94.7	94.7	95.0	95.0	95.0	95.0	95.2	95.2
0.2	0.2	94.8	94.8	94.6	94.6	94.9	94.9	94.5	94.5
0.3	0.35	95.1	95.0	95.0	94.9	95.1	95.0	94.9	94.9
0.4	0.5	95.7	95.6	95.0	95.0	95.4	95.4	95.0	95.0
0.5	0.7	95.2	95.2	95.4	95.4	95.5	95.5	95.4	95.4
0.6	0.9	95.3	95.3	95.5	95.5	95.2	95.2	95.4	95.4
0.7	1.2	95.2	95.2	95.1	95.1	95.3	95.3	94.9	94.9
0.8	1.6	94.6	94.6	95.0	95.0	94.2	94.2	94.8	94.8
0.9	2.3	95.4	95.4	94.9	94.9	95.5	95.5	95.2	95.2

7. Interpretation and evaluation of results

7.1. The eta-parameter in practical applications

If α and β are unknown then p and hence η are estimated from the sample set so that $\hat{\eta} = (\frac{\tau_L}{\hat{\alpha}})^{\hat{\beta}}$ and $\hat{p} = 1 - e^{-\hat{\eta}}$. As $\hat{\eta}$ is a non-linear function of $\hat{\alpha}$ and $\hat{\beta}$ then $\hat{\eta}$ will be a biased estimate of η. As discussed in Appendix B, for a sample size n the bias in η is defined as

$$\mathbb{E}\left[\Delta\hat{\eta}\right] = \mathbb{E}\left[\hat{\eta} - \eta\right] = \mathbb{E}\left[\hat{\eta}\right] - \eta\,, \tag{29}$$

so that an unbiased estimate of η is given in Appendix B by Equation (32) in conjunction with Equation (36). Estimated $\hat{\eta}$ and corrected (unbiased) $\hat{\eta}$ values for various sample sizes and truncation levels are given in the Tables 22 and 23 for Case II and Case IIIb respectively. Making use of these tables, we demonstrate the passing rates with and without the bias-correction of $\hat{\eta}$ in Tables 15 and 16 for Case II and Case IIIb, respectively. For large sample size the bias vanishes in accordance with Theorem 1 in Kreer *et al.* (2015). Furthermore, for small truncation parameters η the bias is of no relevance. Only for small sample sizes ($n = 30, 50, 100$) and truncation levels, p above 0.7 does the correction (unbiasing) formula need to be applied.

Table 15: Percentage pass rates in KS-test with and without $\hat{\eta}$-correction for 10000 simulations in Case II (error is less than $\pm 0.5\%$).

p	η	n = 30		n=50		n=100		n=1000	
		$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected	$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected	$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected	$\hat{\eta}$ uncorrected	$\hat{\eta}$ corrected
0	0	95.3	95.3	95.0	95.0	95.2	95.2	95.4	95.4
0.1	0.1	95.1	95.1	94.9	94.9	94.8	94.8	95.5	95.5
0.2	0.2	95.2	95.2	95.2	95.2	94.9	94.9	95.1	95.1
0.3	0.35	94.8	94.8	94.8	94.8	94.8	94.8	94.6	94.6
0.4	0.5	94.2	94.3	94.7	94.6	94.8	94.8	94.6	94.6
0.5	0.7	94.3	94.7	94.4	94.3	95.1	95.0	94.6	94.6
0.6	0.9	93.6	94.5	94.3	94.2	94.9	94.8	95.0	95.0
0.7	1.2	90.5	94.1	93.8	94.7	94.6	94.6	95.2	95.2
0.8	1.6	86.7	94.3	90.7	94.3	93.9	94.6	95.1	95.1
0.9	2.3	76.7	94.6	81.6	94.3	88.6	94.4	94.8	94.9

7.2. Power studies: comparison with other distributions in Case II

In order to answer the question " What is the chance that data drawn from some alternative distribution will pass the hypothesis test for a Weibull distribution?", the power test is employed.

We compare the power of our out-sample (Case I) and in-sample (Case II) tests by drawing the random numbers of our samples from alternative distributions commonly used in the literature for making goodness-of-fit comparisons. We follow Aho, Bain, and Engelhardt (1985) and consider as possible alternatives to the 2-parameter Weibull distribution, those distributions defined on the positive range. In particular, we consider the log-normal, log-Cauchy, Pareto (power law), log-double exponential, log-logistic and chi-square distributions with 1, 3 and 4 degrees of freedom (note that the chi-square distribution with 2 deegrees of freedom is the exponential and thus not in the scope here). We consider the chi-square distributions with 1, 3 and 4 degrees of freedom as academic only, as they only permit one to fit one single parameter, i.e. the degree of freedom k. As noted earlier by Aho *et al.* (1985), for the complete data set our test performs well for log-Cauchy, Pareto, log-double-exponential and log-logistic, namely one can rule out these distributions as candidates explaining the data set. On the other hand, we found that the power-testing does have problems ruling out χ^2-distributions with 1, 3 and 4 degrees of freedom and log-normal distributions. The latter can be ruled out by a likelihood ratio test. The results are summarized in Table 17 for the complete and the truncated Case I and Case II.

Table 17: Summary of in-sample KS-test, truncation rate $p = 0, 0.5$ for Case I and Case II, number of simulations, $N = 1000$.

		Case I		Case II	
		$p = 0$	$p = 0.5$	$p = 0$	$p = 0.5$
distribution	sample size n	pass rate %	pass rate %	pass rate %	pass rate %
Weibull2d	30	96	96	93	93
Weibull2d	100	95	96	93	93
Weibull2d	500	-	-	95	97
log-Cauchy	30	50	4	6	3
log-Cauchy	100	2	1	0	0
log-Cauchy	500	-	-	0	0
log-double exp.	30	57	42	39	62
log-double exp.	100	6	1	3	52
log-double exp.	500	-	-	0	43
log-logistic	30	46	1	63	85
log-logistic	100	1	0	16	85
log-logistic	500	-	-	0	56
log-normal	30	55	65	73	93
log-normal	100	4	17	30	93
log-normal	500	-	-	0	89
Pareto	30	0	1	1	42
Pareto	100	0	0	0	52
Pareto	500	-	-	0	44
chi-square(k=1)	30	56	78	92	95
chi-square(k=1)	100	8	40	81	96
chi-square(k=1)	500	-	-	43	94
chi-square(k=3)	30	0	3	93	89
chi-square(k=3)	100	0	0	92	95
chi-square(k=3)	500	-	-	81	95
chi-square(k=4)	30	0	0	92	90
chi-square(k=4)	100	0	0	87	88
chi-square(k=4)	500	-	-	63	87

8. Application of our modified KS test

8.1. US data on duration of ethnically mixed marriages

Data on the duration of marriages that end in divorce in the US is publicly available at (http://data.princeton.edu/wws509/datasets/#divorce). Most states in the United States require a minimum legal separation time prior to divorce, although not all do. The duration of marriages that ultimately end in a divorce in the database will therefore contain a mixture of those with a minimum duration (from 0 to 12 months). In order to determine the distribution that describes the duration of failed marriages in the US, it is therefore necessary to left-truncate the data.

We have taken a subset of 230 divorced couples where husband and wife belong to different ethnic groups. We then analyzed the duration of the marriages for a range of left-truncation values, specifically $\tau_L = 0.25$, 1, 5 and 10 years in Table 18. We observe from the data also that the smallest life time is bigger than 0.25 years. This is further evidence that the data is left-truncated. Truncation rates p are given as percentage of data which have been eliminated by the truncation procedure. From the estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ we get $\hat{\eta}$ as estimator for our critical value using Equation (28) and the KS distance D_n is calculated from the data using Equation (12). Due to moderate truncation levels we do not need to un-bias the value of $\hat{\eta}$. Hence, we cannot reject the hypothesis, that the data come from a left-truncated Weibull distribution for a wide range of truncation levels with $\beta = 0.25 \pm 0.07$ and $\alpha = 11.4 \pm 0.06$

years. The details of this analysis can be seen in the Table 18.

Table 18: Duration of ethnically mixed marriages ending in divorce in the US. y indicates year as a unit.

τ_L [y]	n	$\hat{\alpha}$ [y]	$\hat{\beta}$	$\hat{\eta}$	p[%]	D_n	$D_{cv}(n,\hat{\eta},0.05)/\sqrt{n}$	H_0
-	230	11.5 ± 0.6	1.29 ± 0.07	0.00 ± 0.00	0.0	0.0502	0.0581	Accept
0.25	230	11.3 ± 0.6	1.25 ± 0.07	0.01 ± 0.00	0.0	0.0437	0.0570	Accept
1	222	11.2 ± 0.7	1.24 ± 0.07	0.05 ± 0.01	3.5	0.0425	0.0572	Accept
5	157	11.4 ± 0.8	1.24 ± 0.08	0.36 ± 0.02	31.7	0.0551	0.0672	Accept
10	96	14.0 ± 0.9	1.50 ± 0.11	0.60 ± 0.02	58.3	0.0626	0.0861	Accept

8.2. Time between major terrorist attacks with minimum 10 casualties.

The worldwide probability distribution of terrorist attacks has been investigated by Clauset and Woodard (2013). We utilize the RAND-MIPT database (available at http://www.rand.org/nsrd/projects/ terrorism-incidents/download.html) containing 13,274 terrorist events worldwide from 1968 to 2007. Like Clauset and Woodard (2013) we are interested in "major attacks", defined as terrorist events with at least 10 casualties. We investigate the times between these major attacks and find that a large proportion of their tail can be described as left-truncated Weibull. From the estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ we get $\hat{\eta}$ as estimator for our critical value using Equation (28) and the KS distance D_n is calculated from the data using Equation (12). Results are given in Table 19. We note that the tail of the distribution can be described by a Weibull distribution with shape parameter $\beta \simeq 0.50$ whereas the short-end is described by something else and does not pass the Weibull hypothesis.

Table 19: Time between major terrorist attacks with minimum 10 casualties. d indicates day as a unit.

τ_L [d]	n	$\hat{\alpha}$ [d]	$\hat{\beta}$	$\hat{\eta}$	p[%]	D_n	$D_{cv}(n,\hat{\eta},0.05)/\sqrt{n}$	H_0
-	926	9.1 ± 0.5	0.61 ± 0.02	0.00 ± 0.00	0.0	0.2292	0.0290	Decline
10	204	12.7 ± 2.0	0.48 ± 0.03	0.89 ± 0.07	78.0	0.0491	0.0604	Accept
12	187	12.6 ± 2.0	0.48 ± 0.03	0.98 ± 0.07	79.8	0.0426	0.0632	Accept
14	173	12.5 ± 2.1	0.48 ± 0.03	1.06 ± 0.08	81.3	0.0447	0.0659	Accept
16	161	12.2 ± 2.1	0.47 ± 0.03	1.14 ± 0.08	82.6	0.0465	0.0684	Accept
18	148	15.6 ± 2.7	0.50 ± 0.03	1.08 ± 0.08	84.0	0.0526	0.0711	Accept
20	140	14.4 ± 2.6	0.49 ± 0.03	1.17 ± 0.08	84.9	0.0539	0.0733	Accept
22	132	14.2 ± 2.7	0.49 ± 0.03	1.24 ± 0.09	85.7	0.0557	0.0756	Accept
24	124	15.9 ± 3.0	0.51 ± 0.04	1.23 ± 0.09	86.6	0.0588	0.0779	Accept

8.3. Stock market data

We investigate the difference in arrival times between consecutive orders at the New York Stock Exchange (NYSE) for a given stock. The free data provided by www.tickdata.com comprises the entire trading day of shares of ITT Corp. on 11 January 2011, from 9:30 to 16:00 EST. For this example we only look at a snapshot from 12:00:00 to 12:00:21 EST, i.e. 21 seconds of data. The resolution of the arrival times is milliseconds.

Truncation of arrival time differences is the process of taking the differences between consecutive arrival times and keeping only those with differences greater than $\tau_L = 1, 2, 5, 10$ milliseconds. As we did in the previous examples we estimate the Weibull parameters and perform the hypothesis test; the results are given in Table 20.

Table 20: Arrival times of for ITT Corp. orders on NYSE on 11 Jan. 2011, 12:00:00-12:00:21.

τ_L [ms]	n	$\hat{\alpha}$ [ms]	$\hat{\beta}$	$\hat{\eta}$	p[%]	D_n	$D_{cv}(n,\hat{\eta},0.05)/\sqrt{n}$	H_0
-	100	-	-	0	0	-	0.0874	Decline
1	61	128 ± 38	0.46 ± 0.05	0.1073	39%	0.0684	0.1064	Accept
2	57	169 ± 46	0.51 ± 0.05	0.1041	43%	0.0734	0.1100	Accept
5	54	190 ± 49	0.55 ± 0.06	0.1352	46%	0.0797	0.1127	Accept
10	51	179 ± 50	0.53 ± 0.06	0.2168	49%	0.0839	0.1155	Accept

From Table 20 we see that we can not reject the hypothesis that our truncated samples come from a Weibull distribution. However when we analyse the complete (untruncated) sample we see by a similar computation that it leads to the rejection of the Weibull hypothesis as the zero-inflated data with arrival time differences below 1 millisecond prevent the MLE converging onto a solution. One millisecond truncation seems to corrupt the estimation of the Weibull parameters due to the error in time measurement of ± 1 millisecond. From 2 millisecond truncation onwards one finds consistent parameter estimation. Taking the weighted means and errors from the truncated data sets with truncations of 2, 5 and 10 milliseconds we find for the parameters $\hat{\alpha} = 179 \pm 37$ milliseconds and $\hat{\beta} = 0.53 \pm 0.04$.

8.4. Time intervals for radioactive decay of Americium-241

Since the pioneering work of Geiger and Rutherford (1910) the counting process of the particles arising from radioactive decay have been found to be described by a Poisson process. Due to the so-called "dead time" of the detection device, certain decay events might not be measured because the detector is still busy with "detecting" the previous event. Thus, the data set will be incomplete due to "truncation". This has given rise to certain corrections for the Poisson process. Only 60 years later it was possible to measure waiting times between radioactive decay events with acceptable accuracy using multichannel analyzers. Garfinkel and Mann (1968) did one of the first measurement using a probe of 0.2 μCi Americium-231 as a nearly pure α-source Their entire data set, comprising some 300'000 time intervals, was evaluated later by Berkson (1975) albeit under the assumption of a Poisson process and performing a χ^2-test on the bin-ed data. Here, we want to demonstrate our analysis of a smaller sample which is displayed in Garfinkel and Mann (1968) on page 709. We use the second, third and fourth block only because the first block contains some control measurements for calibration. Our data sample comprises 300 measurement points describing the time between subsequent α-particles. The dead time was estimated by the authors to be 2.54 T.U.(1 T.U. denotes a time unit and corresponds to the pulse frequency of 370 kHz). Our results are displayed in Table 21. We recover as expected a shape parameter $\beta = 1$ indicating that the waiting times are exponentially distributed giving rise to the Poisson process discovered in Geiger and Rutherford (1910).

Table 21: Time intervals for radioactive decay of Americium-241. *T.U.* indicates time unit.

τ_L [T.U.]	n	$\hat{\alpha}$ [T.U.]	$\hat{\beta}$	$\hat{\eta}$	p[%]	D_n	$D_{cv}(n, \hat{\eta}, 0.05)/\sqrt{n}$	H_0
-	300	15605 ± 947	1.00 ± 0.05	0.00 ± 0.00	0.0	0.0491	0.0510	Accept
3	300	15596 ± 947	1.00 ± 0.05	0.00 ± 0.00	0.0	0.0493	0.0508	Accept
10	300	15576 ± 948	1.00 ± 0.05	0.00 ± 0.00	0.0	0.0498	0.0507	Accept
100	298	15597 ± 951	1.00 ± 0.05	0.01 ± 0.01	0.6	0.0500	0.0504	Accept

9. Conclusion

The Weibull distributions with a shape parameter less than one is known as "heavy-tailed" because it has significant probabilities quite far from its mean. In insurance and other industries the cost of rare events due to "heavy tails" can be very high, so it is important to determine exactly how rare they actually are. This can only be done by taking the available data and testing it against hypothesized distributions.

Data obtained from real life examples are often left-truncated. To test the hypothesis that the data are sampled from a left-truncated Weibull distribution, one can perform a Kolmogorov-Smirnov goodness-of-fit test. If the shape and scale parameters are not known they must be estimated from the data itself. The commonly used maximum likelihood estimator does not always give a non-trivial solution to estimating the shape and scale parameters, especially for small sample sizes. For a small sample size there is a chance that the solution of the maximum likelihood estimate lie on the trivial boundaries where either one or both of the parameters vanish. A criterion for determining when non-vanishing solutions for the parameters exist was given in this paper. We demonstrated also that with increasing sample size non-trivial estimates exist with probability tending to one and these estimates are consistent, asymptotically normal, and efficient. Having obtained non-trivial estimates, a goodness-of-fit can be judged using a Kolmogorov-Smirnov test. If either the shape and/or scale parameters are unknown the critical values differ significantly from those when the parameters are known. If both the parameters or only the shape parameter are unknown the critical values depend on the truncation value as well the number of data.

The modified critical values presented here should be used to test if a set of data is sampled from a left-truncated Weibull distribution with a known truncation point but unknown shape and/or scale parameter. When both the parameters or only the shape parameter are unknown and the truncation level is greater than 10%, then the dependence of the critical value on the truncation level must be included, otherwise incorrect conclusions from the hypothesis tests will be drawn. We provided the modified CVs in Tables (3) - (6) for various sample sizes and truncation ranges and also formulas Equation (27) and Equation (28) where one can calculate them for any desired p (or η) for given n and for combination of (p (or η), n), respectively.

Although the results presented here on the left truncated Weibull distribution can be applied to a wide range of applications in many disciplines we are not aware of any other comprehensive studies that discuss the effects of truncation dependence on the critical values and parameter estimation. We are in the process of applying our techniques to investigate financial, insurance, and real estate data using our tables and models for the critical values which include the dependence on truncation and sample size.

10. Acknowledgement(s)

The authors are grateful to Ross Frick from University of South Australia for a careful reading of the manuscript which led to significant improvements. This work was supported by the University of Adelaide and by the Australian Research Council (AWT, FL0992247).

A. Left-truncated Weibull random variates and their representation by exponential variates

Let $u_i \in (0,1)$ denote the standard uniform random variable. Then from the cdf in equation (1) we obtain for the left-truncated Weibull random variable τ_i

$$\tau_i = \alpha \cdot \left[\left(\frac{\tau_L}{\alpha}\right)^{\beta} + \log\frac{1}{u_i}\right]^{1/\beta} = \alpha \cdot [\eta + y_i]^{1/\beta} \tag{30}$$

where y_i is a standard exponentially distributed random variable and $\eta \equiv (\tau_L/\alpha)^{\beta}$.

B. Bias in estimates of eta

The estimated value of η (i.e. $\hat{\eta} = (\tau_L/\hat{\alpha})^{\hat{\beta}}$), will have an estimation error $\delta\hat{\eta}$. In this sense, both $\hat{\eta}$ and $\Delta\hat{\eta}$ are random variables whereas η is a fixed real number:

$$\eta = \hat{\eta} - \Delta\hat{\eta} \quad \Rightarrow \quad \eta = \mathbb{E}[\hat{\eta}] - \mathbb{E}[\Delta\hat{\eta}]\,. \tag{31}$$

By definition $\eta \geq 0$ hence we use an un-biasing formula motivated by Equation (31)

$$\eta = \max\{0, \hat{\eta} - \mathbb{E}[\Delta\hat{\eta}]\} \tag{32}$$

where the individual $\hat{\eta}$ is unbiased by a correction term $\mathbb{E}[\Delta\hat{\eta}]$ subject to $\eta \geq 0$.

To derrive formally the correction term $\mathbb{E}[\Delta\hat{\eta}]$, recall that as discussed in section 2 for a large enough sample size the parameters $\hat{\alpha}$ and $\hat{\beta}$ are normally distributed

$$\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\theta^0}, \boldsymbol{Z}^{-1}\right) \quad , \tag{33}$$

where $\boldsymbol{\theta}^{\mathbf{0}} = [\alpha, \beta]$ and $\boldsymbol{Z}$ is the Fisher information matrix. To estimate the effect of errors in η due to errors in $\boldsymbol{\theta}$ we write similarly

$$\hat{\alpha} = \alpha + \Delta\alpha \qquad \hat{\beta} = \beta + \Delta\beta \tag{34}$$

$$\text{where} \qquad \mathbb{E}\left[\begin{bmatrix}\Delta\alpha \\ \Delta\beta\end{bmatrix}[\Delta\alpha \quad \Delta\beta]\right] = \begin{bmatrix}\sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2\end{bmatrix} = \boldsymbol{Z}^{-1}. \tag{35}$$

Here, we have used Equation (33) to calculate the expectation. Then the Taylor expansion of $\Delta\hat{\eta}$ gives

$$\begin{aligned}\Delta\hat{\eta} &= \frac{\partial\hat{\eta}}{\partial\hat{\alpha}}\Delta\hat{\alpha} + \frac{\partial\hat{\eta}}{\partial\hat{\beta}}\Delta\hat{\beta} + \frac{1}{2}\frac{\partial^2\hat{\eta}}{\partial\hat{\alpha}^2}\Delta\hat{\alpha}^2 + \frac{1}{2}\frac{\partial^2\hat{\eta}}{\partial\hat{\beta}^2}\Delta\hat{\beta}^2 + \frac{\partial^2\hat{\eta}}{\partial\hat{\alpha}\,\partial\hat{\beta}}\Delta\hat{\alpha}\,\Delta\hat{\beta} + \cdots \\ \mathbb{E}[\Delta\hat{\eta}] &= \frac{1}{2}\frac{\partial^2\hat{\eta}}{\partial\hat{\alpha}^2}\sigma_{\hat{\alpha}}^2 + \frac{1}{2}\frac{\partial^2\hat{\eta}}{\partial\hat{\beta}^2}\sigma_{\hat{\beta}}^2 + \frac{\partial^2\hat{\eta}}{\partial\hat{\alpha}\,\partial\hat{\beta}}\sigma_{\hat{\alpha}\,\hat{\beta}} + \cdots \\ &= \hat{\eta}\left\{\left[\frac{1}{2\hat{\alpha}^2}\hat{\beta}(1+\hat{\beta})\right]\sigma_{\hat{\alpha}}^2 + \left[\frac{1}{2\hat{\beta}^2}\log\hat{\eta}^2\right]\sigma_{\hat{\beta}}^2 - \frac{1}{\hat{\alpha}}\left[1+\log\hat{\eta}\right]\sigma_{\hat{\alpha}\,\hat{\beta}} + \cdots\right\}.\end{aligned} \tag{36}$$

Note that in Equation (36) we take the expectations only over the $\Delta\hat{\alpha}$ and $\Delta\hat{\beta}$ but not over the estimates $\hat{\alpha}$ or $\hat{\beta}$. Estimated $\hat{\eta}$ (uncorrected) values and corrected (unbias) $\hat{\eta}$ using Equation (36) for various sample sizes and truncation levels are given in Table 22 for Case II and in Table 23 for Case IIIb .

Table 22: Estimated $\hat{\eta}$ and unbiased $\hat{\eta}$ for 10,000 simulations for sample sizes $n = 30, 50, 100, 1000$ in Case II.

p	η	n = 30		n=50		n=100		n=1000	
		$\hat{\eta}$	$\hat{\eta} - unbias$	$\hat{\eta}$	$\hat{\eta} - unbias$	$\hat{\eta}$	$\hat{\eta} - unbias$	$\hat{\eta}$	$\hat{\eta} - unbias$
0	0	0	0	0	0	0	0	0	0
0.1	0.1	0.106	0.106	0.103	0.104	0.102	0.102	0.100	0.100
0.2	0.2	0.212	0.203	0.205	0.203	0.202	0.201	0.200	0.200
0.3	0.35	0.382	0.327	0.360	0.343	0.355	0.349	0.351	0.350
0.4	0.5	0.583	0.433	0.528	0.476	0.512	0.498	0.502	0.501
0.5	0.7	0.895	0.548	0.767	0.629	0.719	0.677	0.701	0.698
0.6	0.9	1.291	0.647	1.040	0.744	0.946	0.844	0.902	0.897
0.7	1.2	2.084	0.763	1.496	0.894	1.300	1.059	1.202	1.192
0.8	1.6	3.149	0.900	2.307	1.063	1.827	1.252	1.611	1.587
0.9	2.3	5.629	1.092	4.405	1.245	2.968	1.528	2.319	2.254

Table 23: Estimated $\hat{\eta}$ and unbiased $\hat{\eta}$ for 10,000 simulations for sample sizes $n = 30, 50, 100, 1000$ in Case IIIb.

p	η	n = 30		n=50		n=100		n=1000	
		$\hat{\eta}$	$\hat{\eta} - unbias$	$\hat{\eta}$	$\hat{\eta} - unbias$	$\hat{\eta}$	$\hat{\eta} - unbias$	$\hat{\eta}$	$\hat{\eta} - unbias$
0	0	0	0	0	0	0	0	0	0
0.1	0.1	0.097	0.104	0.097	0.101	0.099	0.101	0.100	0.100
0.2	0.2	0.191	0.201	0.195	0.201	0.197	0.200	0.200	0.200
0.3	0.35	0.034	0.347	0.342	0.348	0.346	0.349	0.350	0.350
0.4	0.5	0.488	0.497	0.493	0.497	0.497	0.499	0.500	0.500
0.5	0.7	0.693	0.697	0.696	0.698	0.698	0.699	0.700	0.700
0.6	0.9	0.898	0.899	0.898	0.899	0.899	0.900	0.900	0.900
0.7	1.2	1.205	1.202	1.203	1.201	1.201	1.201	1.200	1.200
0.7	1.6	1.608	1.603	1.608	1.604	1.604	1.602	1.600	1.600
0.9	2.3	2.288	2.281	2.297	2.292	2.302	2.298	2.301	2.301

References

Agostino R, Stephens MA (1986). *Goodness-of-fit techniques.* M. Dekker, New York. ISBN 0-8247-7487-6.

Aho M, Bain LJ, Engelhardt M (1985). "Goodness-of-fit tests for the weibull distribution with unknown parameters and heavy censoring." *Journal of Statistical Computation and Simulation.*, **21(3-4)**, 213–225.

Anderson T, Stephens M (1997). "The continuous and discrete Brownian bridges: Representations and applications." *Linear Algebra and its Applications*, **264 (6)**, 145–171.

Balakrishnan N, Cohen A (1991). *Order Statistics and Inference: Estimation Methods*. Academic Press, Boston.

Balakrishnan N, Mitra D (2012). "Left truncated and right censored Weibull data and likelihood inference with an illustration." *Computational Statistics & Data Analysis*, **56**, 4011 – 4025.

Berkson J (1975). "Do Radioactive Decay Events Follow a Random Poisson-Exponential?" *International Journal of Applied Radiation and Isotopes,*, **26(9)**, 543–549.

Birnbaum FJ (1952). "Numerical Tabulation of the Distribution of Kolmogorov's Statistic for Finite Sample Size." *Journal of the American Statistical Association*, **47**(259), 425–441.

Chandra M, Singapurwalla N, Stephens M (1981). "Kolmogorov statistics for tests of fit for the extreme value and Weibull distributions." *Journal of the American Statistical Association*, **76 (375)**, 729–731.

Clauset A, Woodard R (2013). "Estimating the historical and future probabilities of large terrorist events." *Annals of Applied Statistics*, **7 (4)**, 1838–1865.

Cohen CA (1965). "Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples." *Technometrics*, **7 (4)**, 579–588.

David F, Johnson N (1948). "The Probability Integral Transformation When Parameters are Estimated from the Sample." *Biometrika*, **35 (1/2)**, 182–190.

Durbin J (1973). "Weak convergence of the sample distribution function when parameters are estimated." *The Annals of Statistics*, **1 (2)**, 279–290.

Durbin J (1975). "Kolmogorov-Smirnov Tests when Parameters are Estimated with Applications to Tests of Exponentiality and Tests on Spacings." *Biometrica*, **62(1)**, 5–22.

Evans JW, Johnson RA, Green DW (1989). *Two-and three-parameter weibull goodness-of-fit tests*, volume 493. US Department of Agriculture, Forest Service, Forest Products Laboratory.

Garfinkel S, Mann W (1968). "A method for obtaining large numbers of measured time intervals in radioactive decay." *International Journal of Applied Radiation and Isotopes,*, **19(9)**, 707–709.

Geiger H, Rutherford E (1910). "The number of α particles emitted by uranium and thorium and by uranium minerals." *Philolophical Magazine*, **20(118)**, 691–698.

Kendall MG, Stuart A (1979). *The advanced theory of statistics - Inference and relationship*, volume 2. 4th revised edition edition. C. Griffin, London. ISBN 978—0-8-52-64-2 -255- 2.

Kreer M, Kizilersu A, Thomas AW, dos Reis AE (2015). "Goodness-of-fit tests and applications for left-truncated Weibull distributions to non-life insurance." *European Actuarial Journal*, **5 (1)**. doi:10.1007/s13385-015-0105-8.

Lehmann E, Casella G (1998). *Theory of Point Estimation* -. 2nd ed. 1998. corr. 4th printing 2003 edition. Springer, Berlin, Heidelberg. ISBN 978-0-387-98502-2.

Lilliefors H (1969). "On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown." *Journal of the American Statistical Association*, **64 (325)**, 387–389.

Littell RC, McClave JT, Offen WW (1979). "Goodness-of-fit tests for the two parameter Weibull distribution." *Communications in Statististics - Simulation and Computation*, **8 (3)**, 257–269.

Malevergne Y, Pisarenko V, Sornette D (2005). "Empirical distributions of stock returns: between the stretched exponential and the power law?" *Quantitative Finance*, **5 (4)**, 379–401.

Massey FJ (1951). "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association*, **46**(253), 68–78.

Miller LH (1956). "Table of Percentage Points of Kolmogorov Statistics." *Journal of the American Statistical Association*, **51**(273), pp. 111–121.

Parsons F, Wirsching PH (1982). "A Kolmogorov-Smirnov goodness-of-fit test for the two-parameter Weibull distribution when the parameters are estimated from the data." *Microelectronics Reliability*, **22(2)**, 163–167.

Rinne H (2009). *The Weibull Distribution: A Handbook*. 1st edition. CRC Press, Taylor & Francis Group, Boca Raton New York. ISBN 978-1-4200-8743-7.

Shorack GR, Wellner JA (2009). *Empirical processes with applications to statistics*. SIAM, Philadelphia. ISBN 978-0-898-71901-7.

Smirnov N (1948). "Table for Estimating the Goodness of Fit of Empirical Distributions." *The Annals of Mathematical Statistics*, **19(2)**, 279–281.

Stephens M (1977). "Goodness of Fit for the Extreme Value Distribution." *Biometrika*, **64 (3)**, 583–588.

Thoman DR, Bain LJ, Antle CE (1969). "Inferences on the parameters of the Weibull distribution." *Technometrics*, **11 (3)**, 445–460.

Weibull W (1951). "A statistical distribution function of wide applicability." *J. Appl. Mech.-Trans. ASME*, **73**, 293.

Wingo DR (1989). "The left-truncated Weibull distribution: theory and computation." *Statistical Papers*, **30**, 39–48.

Woodruff BW, Moore AH, Dunne EJ, Cortes R (1983). "A modified Kolmogorov-Smirnov test for Weibull distributions with unknown location and scale parameters." *IEEE Transactions on Reliability*, **R-32 (2)**, 209–213.

Affiliation:

Ayşe Kızılersü
Special Research Centre for the Subatomic Structure of Matter (CSSM)
Department of Physics
University of Adelaide
5005, Adelaide,Australia
E-mail: ayse.kizilersu@adelaide.edu.au
URL: http://www.physics.adelaide.edu.au/cssm/index.html

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume VV *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd