Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

COMSTA 2581
pp: 1-12 (col.fig.: nil)
PROD. TYPE: COM
ED: JS
PAGN: Usha.N – SCAN: Bindu

ARTICLE IN PRESS

ELSEVIER Computational Statistics & Data Analysis ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Transformation approaches for the construction of Weibull prediction interval

Zhenlin Yang[a,*], Stanley P. See[b], M. Xie[c]

[a] *Department of Economics, National University of Singapore, Singapore 117570, Singapore*
[b] *Department of Statistics and Applied Probability, National University of Singapore, Singapore*
[c] *Department of Industrial and System Engineering, National University of Singapore, Singapore*

Received 31 July 2002

Abstract

Two methods of transforming the Weibull data to near normality, namely the Box–Cox method and Kullback–Leibler (KL) information method, are discussed and contrasted. A simple prediction interval (PI) based on the better KL information method is proposed. The asymptotic property of this interval is established. Its small sample behavior is investigated using Monte Carlo simulation. Simulation results show that this simple interval is close to the existing complicated PI where the percentage points of the reference distribution have to be either simulated or approximated. The proposed interval can also be easily adjusted to have the correct asymptotic coverage.

Keywords: Box–Cox transformation; Coverage probability; Kullback–Leibler information; Prediction interval; Weibull distribution

1. Introduction

Weibull distribution has been shown to be useful for modelling and analysis of lifetime data in medical and biological sciences, engineering, etc. Many statistical methods have been developed for this distribution. However, simple inference methods without requiring further approximation or simulation as those of normal distribution for simple

* Corresponding author: School of Economics and Social Sciences, Federal Building, 469 Bukit Timah Road, Singapore Management University, 259756 Singapore, Singapore. Tel.: +65-6822-0852; fax: +65-6822-0833.

E-mail addresses: ecsyzl@nus.edu.sg, zlyang@smu.edu.sg (Z. Yang).

0167-9473/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0167-9473(02)00232-3

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

problems such as predicting a single future observation do not seem to exist. In this article, we explore the transformation approach for this prediction problem.

A continuous random variable X is said to follow a Weibull distribution and denoted by $X \sim \mathrm{WB}(\alpha, \beta)$, if its probability density function (pdf) takes the form:

$$f(x;\beta,\alpha) = (\beta/\alpha^{\beta})x^{\beta-1}\exp[-(x/\alpha)^{\beta}], \quad \beta > 0 \text{ and } \alpha > 0. \tag{1.1}$$

The parameter α is called the scale parameter and β is the shape parameter. It is well known that $\log(X)$ follows an extreme value distribution, a member of the location-scale family.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a sample of past observations from a $\mathrm{WB}(\alpha, \beta)$ population, and X^0 be a single future observation from the same population. The present article concerns with the problem of constructing a prediction interval (PI) for X^0 based on the observed value of $\mathbf{X}$. Engelhardt and Bain (1979) developed a pivotal quantity from which a one- or two-sided PI can be constructed. However, the process of constructing this PI is rather complicated as the true distribution of the pivotal quantity is unknown. Hence, its percentage points have to be either simulated (Fertig et al., 1980; Mee and Kushary, 1994) or approximated (Engelhardt and Bain, 1982). Dellaportas and Wright (1991) studied Weibull prediction problem based on the Bayesian approach. Yang (1999b) employed the Box–Cox transformation (Box and Cox, 1964) and developed a unified PI for all the lifetime distributions, including the Weibull. It is shown that this unified PI often meets or outperforms the corresponding frequentist PIs for specified distributions. When compared with the Weibull PI with approximated percentage points, his simulation results showed that the unified PI has slightly higher coverage (closer to the nominal level), but with slightly longer length than the frequentist PI. Hence, the two PIs are about the same.

When it is known that the data are from the Weibull distribution, a simpler and more stable method than that of Box and Cox for determining the transformation parameter is available. Hence, the resulted PI should behave better than the unified PI of Yang (1999b). Also, for a known family of distribution, it is possible to derive the exact asymptotic coverage and to adjust the interval for a correct asymptotic coverage. These ideas are realized in this paper for the Weibull distribution, using the method based on the Kullback–Leiber (KL) information number (see Kullback, 1968). The resulted PI is shown to be about equivalent to the existing one, but is much easier to implement: once the maximum-likelihood estimate (MLE) of β is obtained, the interval can simply be calculated by a calculator.

This paper is organized as follows. Section 2 outlines and compares the two methods of transforming the Weibull data. Section 3 presents the transformation-based PI and studies its large sample property upon which a method of adjusting the interval to give correct asymptotic coverage is discussed. Furthermore, results of Monte Carlo simulation are presented for the evaluation of the small sample properties. A numerical example is given in Section 4 for illustration and further comparison. A general discussion is given in Section 5.

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

2. Methods for transforming the Weibull data

If it is known that the observations are positive and continuous, such as the lifetime observations, the Box–Cox procedure (Box and Cox, 1964) can be applied to transform a nonnormal distribution to near normality, so that further analysis can easily be carried out based on normality assumption. When the exact distribution is known, the KL information can be used to obtain the relationship between the transformation parameter and the parameter(s) of the distribution (at least this is the case for the Weibull distribution). The former procedure is unified, but should be less precise as more parameters need to be estimated. These two methods are outlined and compared in this section.

2.1. The Box–Cox method

For a nonnegative random variable X, the Box–Cox power transformation is defined as

$$X(\lambda) = \begin{cases} (X^{\lambda} - 1)/\lambda, & \lambda \neq 0, \\ \log X, & \lambda = 0. \end{cases} \tag{2.1}$$

Let $\mathbf{X}(\lambda) = \{X_1(\lambda),\ X_2(\lambda), \ldots, X_n(\lambda)\}^{\mathrm{T}}$ denote the vector of transformed past observations and $X^0(\lambda)$ denote the transformed future observation. Box and Cox (1964) assumed the existence of λ such that $X_i(\lambda) \sim N(\mu, \sigma^2)$ for some μ and σ. This assumptions leads to the Box–Cox estimate $\hat{\lambda}_{\mathrm{BC}}$ of λ being the solution of $S(\lambda) = 0$, where

$$S(\lambda) = -n \frac{\sum_{i=1}^{n} [X_i(\lambda) - \bar{X}(\lambda)][X_i'(\lambda) - \bar{X}'(\lambda)]}{\sum_{i=1}^{n} [X_i(\lambda) - \bar{X}(\lambda)]^2} + \sum_{i=1}^{n} \log X_i, \tag{2.2}$$

where $X_i'(\lambda)$ is the derivative of $X_i(\lambda)$ and $\bar{X}'(\lambda)$ is the corresponding average. The reason that $\hat{\lambda}_{\mathrm{BC}}$ is called the Box–Cox estimator instead of MLE is that $X_i(\lambda)$ cannot be exact normal unless $\lambda = 0$. More about the Box–Cox transformation can be found in Sakia (1992).

2.2. The method based on Kullback–Leibler information

Knowing which population that the observations come from, an optimal transformation can be found based on a measure of "closeness" between two pdfs, called the KL information

$$I(f, \phi) = \int f(y; \alpha, \beta, \lambda) \log \left\{ \frac{f(y; \alpha, \beta, \lambda)}{\phi(y; \mu, \sigma)} \right\} \mathrm{d}y,$$

where $f(y; \alpha, \beta, \lambda)$ denotes the pdf of $X_i(\lambda)$ and $\phi(y; \mu, \sigma)$ the pdf of a normal distribution with mean μ and standard deviation σ. By minimizing $I(f; \phi)$ with respect to μ, σ and λ, we get an optimal $f(y; \alpha, \beta, \lambda)$ which offers the best approximation to

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

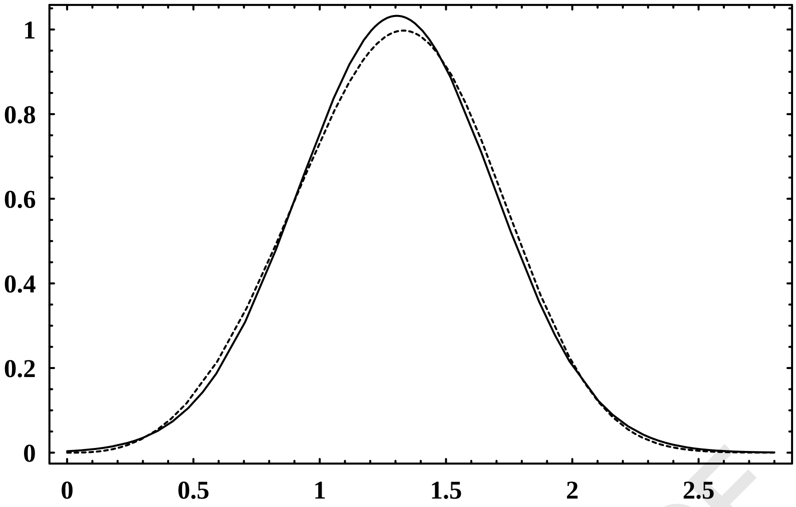

Fig. 1. Plot of the normal (solidline) and transformed Weibull pdfs.

$\phi(y;\mu,\sigma)$ within the power family. Hernandez and Johnson (1980) showed that the best normalizing transformation for the Weibull is

$$\lambda = 0.2654\beta.$$

The corresponding mean and standard deviation for a transformed observation can easily be seen to be $\mu = (0.9034\alpha^{\lambda} - 1)/\lambda$ and $\sigma = 0.2675\alpha^{\lambda}/\lambda$. The minimum value of $I(f;\phi)$ is 0.00278, a very small number that is independent of the parameters α and β. This means that the Weibull distribution can be transformed very closely to normal, but cannot be further improved by changing the α and β values.

This result says that the Weibull is closest to normal when the shape parameter $\beta = \frac{1}{0.2654}$, regardless of values of the scale parameter α. In fact, Weibull distribution can be considered as a power transformation of an exponential distribution and an exponential variable can be transformed to near normality by a power of 0.2654. To illustrate the closeness between the two distributions, we give a plot in Fig. 1.

The Box–Cox transformation by Hernandez and Johnson (1980) can only be used when the shape parameter is known, which is normally not the case. However, using the above relationship, one can easily obtain the MLE of the transformation parameter

$$\hat{\lambda}_{\rm ML} = 0.2654\hat{\beta},$$

where $\hat{\beta}$ is the MLE of β defined as the solution of

$$\frac{1}{\hat{\beta}} = \left(\sum_{i=1}^{n} X_i^{\hat{\beta}} \log X_i\right)\left(\sum_{i=1}^{n} X_i^{\hat{\beta}}\right)^{-1} - \frac{1}{n}\sum_{i=1}^{n} \log X_i. \tag{2.3}$$

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

Table 1
Simulation results for the bias and MSE of $\hat{\lambda}_{BC}$ and $\hat{\lambda}_{ML}$

		Bias		MSE	
α, β	n	$\hat{\lambda}_{BC}$	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{BC}$	$\hat{\lambda}_{ML}$
1, 2	20	−0.04937	0.042316	0.113826	0.014040
	30	−0.03318	0.024965	0.064531	0.007592
	50	−0.02171	0.014489	0.033722	0.004080
	100	−0.00971	0.007797	0.014709	0.001885
1, 0.5	20	−0.01269	0.010224	0.007209	0.000836
	30	−0.00869	0.006503	0.004099	0.000484
	50	−0.00523	0.003749	0.002172	0.000258
	100	−0.00248	0.001844	0.000915	0.000118

Hence, the problem of estimating the transformation parameter reduces to finding the MLE of the shape parameter, which is often a very simple task.

2.3. A comparison of the two transformation estimates

Obviously, when β is known, one would use the KL information method to determine the transformation. When β is unknown, one could choose between $\hat{\lambda}_{ML}$ and $\hat{\lambda}_{BC}$. Because the process of obtaining $\hat{\lambda}_{ML}$ involves estimating only one parameter, whereas the process of obtaining $\hat{\lambda}_{BC}$ involves estimation of three parameters, $\hat{\lambda}_{ML}$ is expected to have a smaller variance than $\hat{\lambda}_{BC}$. Using a result of Yang (1999a, p. 175) and the asymptotic variance formula for $\hat{\beta}$ (Johnson et al., 1994, p. 657), the difference between the two estimators in terms of variability is as follows: $n\,\mathrm{Var}(\hat{\lambda}_{ML}) \approx 0.0428\beta^2$ and $n\,\mathrm{Var}(\hat{\lambda}_{BC}) \approx 0.5429\beta^2$. This clearly shows that $\hat{\lambda}_{BC}$ is much more variable than $\hat{\lambda}_{ML}$. Although the conditions of Yang's (1999a) formula are not exactly satisfied by the Weibull variable, this approximation should still be very informative in seeing the larger variability of $\hat{\lambda}_{BC}$ than $\hat{\lambda}_{ML}$. To confirm this conclusion and to see the difference when sample size is not large, a Monte Carlo comparison between these two estimators is carried out and the results are summarized in Table 1. The results in each row are based on 10 000 random samples.

The results are clearly in favor of the $\hat{\lambda}_{ML}$ as it has consistently lower bias and MSE than $\hat{\lambda}_{BC}$, hence should be recommended for use for developing the PI. The simulation results are consistent with the large sample results. The results of this section are useful as it gives a clear indication on which method to choose when one considers transforming a Weibull data set for statistical analysis.

3. The transformation-based prediction interval

In this section, we propose the transformation-based PI for the Weibull distribution and investigate its asymptotic as well as small sample properties.

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

Table 2
A summary of the limiting coverage probabilities of PI (3.1)

Nominal Level $1-\delta$	0.9000	0.9250	0.9500	0.9750	0.9800	0.9900
Actual Coverage	0.8986	0.9256	0.9526	0.9792	0.9843	0.9939
Adjusted Level $1-\delta^*$	0.9013	0.9244	0.9475	0.9710	0.9758	0.9858
Actual Coverage	0.9000	0.9250	0.9500	0.9750	0.9800	0.9900

3.1. The prediction interval and its large sample property

The close approximation to normality of $X_i(\lambda)$ allows the PI for $X^0(\lambda)$ be constructed in the usual manner. Since

$$T(\lambda) = \frac{X^0(\lambda) - \bar{X}(\lambda)}{s(\lambda)\sqrt{1+n^{-1}}} \tag{3.1}$$

is approximately distributed as t_{n-1}, where $\bar{X}(\lambda)$ and $s(\lambda)$ are the sample mean and standard deviation of the $X_i(\lambda)$'s, an inverse transformation of the resulted PI for $X^0(\lambda)$ gives an approximate $100(1-\delta)\%$ PI for X^0

$$\{1 + \lambda[\bar{X}(\lambda) \pm t_{n-1}(\delta/2)s(\lambda)\sqrt{1+n^{-1}}]\}^{1/\lambda}, \tag{3.2}$$

which becomes $\exp\{\bar{X}(0) \pm t_{n-1}(\delta/2)s(0)\sqrt{1+n^{-1}}\}$ when $\lambda = 0$, where $t_{n-1}(\delta/2)$ is the upper $100(\delta/2)$ percentage point of t_{n-1}. In the case of unknown λ, it is replaced by its estimator $\hat{\lambda}$ and the resulted transformation-based PI takes the final form:

$$\{1 + \hat{\lambda}[\bar{X}(\hat{\lambda}) \pm t_{n-1}(\delta/2)s(\hat{\lambda})\sqrt{1+n^{-1}}]\}^{1/\hat{\lambda}}. \tag{3.3}$$

Denote the upper $100(\delta/2)\%$ point of the standard normal by $Z_{\delta/2}$. The theoretical property of interval (3.3) is summarized in the following theorem:

Theorem 3.1. *Let $\hat{\lambda}$ be the MLE or any consistent estimator of λ, $L_\delta(\mathbf{X})$ and $U_\delta(\mathbf{X})$ be, respectively, the lower and upper bound of PI (3.3) with the nominal level $1-\delta$, and X^0 be the future observation. Then as $n \to \infty$, we have,*

$$P\{L_\delta(\mathbf{X}) \leqslant X^0 \leqslant U_\delta(\mathbf{X})\} \to 1 - \delta^* = \mathrm{e}^{-\ell_\delta} - \mathrm{e}^{-u_\delta},$$

where $\ell_\delta = (0.9034 - 0.2675 Z_{\delta/2})^{1/0.2654}$ and $u_\delta = (0.9034 + 0.2675 Z_{\delta/2})^{1/0.2654}$.

The proof of the Theorem 3.1 is lengthy and is put in Appendix A. To have some idea on how close the limiting coverage probability is to the nominal level, we list a few values in the following table (the upper part of Table 2).

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

Sometimes one may wish to have a PI that is at least asymptotically correct. This can easily be achieved, based on the result of Theorem 3.1, by adjusting the δ value in (3.3). For example, for a 95% PI, let $e^{-\ell_\delta} - e^{-u_\delta} = 0.95$, solve for $Z_{\delta/2}$ and hence the corresponding δ. This can easily be accomplished by a MATHEMATICA function called *RootFind*. In this case, the value $\delta = 0.05$ should be adjusted to $\delta^* = 0.0525$ and the $t_{n-1}(\delta/2)$ value in (3.3) should be replaced by $t_{n-1}(\delta^*/2)$. The value of δ^* for common levels are listed in the lower part of Table 2. From the results (upper part) of Table 2 we see that the limiting coverage probabilities are all quite close to the corresponding nominal levels, especially around the 90–95% levels. The discrepancy between the limiting coverage and the nominal level reflects the effect of nonnormality of the transformed Weibull random variable. The results in Table 2 show that it may only be necessary to adjust the 99% PI. We will further examine this point in the next subsection.

3.2. Small sample property of the interval

Theorem 3.1 summarizes the large sample property of the PI (3.1). In this Section, we investigate the small sample property of this interval using Monte Carlo simulation. The performance of the interval is also compared with the one reported in Engelhardt and Bain (1982). The results are reported in Table 2.

First, the simulation results for the new interval closely agree with the conclusion of Theorem 3.1: the interval is a bit conservative at high coverage levels. For the 90% and 95% levels, simulation results indicate that the two intervals are almost equivalent in the overall sense with the new interval being slightly longer but with a higher coverage that is closer to the nominal level. For the 99% PIs, the new interval is a bit conservative whereas the existing one is liberal especially when n is not large (e.g., 0.9793 vs. 0.99). This results in the new interval being longer than the existing one. However, this is significant only when population is very skewed, and when this happens one may consider to adjust the interval for having a correct coverage and hence a shorter length. The simulation results for the adjusted interval that corresponds to the last row of Table 3 are: 49.8513 0.9904, 40.2908 0.9905, 34.7752 0.9901, 31.2983 0.9899, and that corresponds to the last column of the 99% PI: 1.0733 0.9897, 2.2799 0.9903, 5.5259 0.9902, 11.6466 0.9901, 31.2983 0.9899. This indicates that if we adjust coverage level of the new interval to be the same as that of the existing interval, the lengths of the two intervals would be about the same. The simulation results corresponding to each combination of α, β and n values are based on 10 000 samples.

From the simulation results, we see that the new PI can be significantly longer than the existing one when 99% confidence level is used. The more skewed the population is, the longer it is the new PI relative to the existing one. As the skewness of a Weibull population depends only on the value of the shape parameter β, following rule of thumb can be followed: apply the adjusted PI only when (estimated) $\beta < 1$ and a 99% PI is desired.

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

Table 3
A summary of simulation results: upper entries for new PI

		$n=20$		30		50		100	
(α,β)	ω^{a}	A.L.	C.prob.	A.L.	C.prob.	A.L.	C.prob.	A.L.	C.prob.
		90% Prediction intervals							
(1,5)	−0.25	0.7361	0.8976	0.7181	0.8949	0.7067	0.8998	0.6985	0.9032
		0.7030	0.8816	0.6984	0.8843	0.6974	0.8951	0.6900	0.8976
(1,2)	0.63	1.5653	0.8938	1.5479	0.8943	1.5321	0.8999	1.5195	0.8952
		1.5060	0.8789	1.5073	0.8898	1.5054	0.8972	1.4887	0.8925
(1,1)	2.0	3.2116	0.8938	3.1323	0.8905	3.0649	0.8913	3.0286	0.8971
		2.9693	0.8863	2.9579	0.8851	2.9442	0.8888	2.9006	0.8927
(1,0.7)	3.5	5.8023	0.8944	5.4291	0.8984	5.2085	0.9018	5.0442	0.9032
		4.9593	0.8776	4.8559	0.8870	4.8368	0.8955	4.7057	0.8976
(1,0.5)	6.6	11.4222	0.8930	10.5220	0.8917	10.0063	0.8925	9.6121	0.9002
		9.7602	0.8845	9.4158	0.8865	9.2472	0.8912	8.8245	0.8953
		95% Prediction intervals							
(1,5)	−0.25	0.9007	0.9548	0.8718	0.9521	0.8550	0.9542	0.8406	0.9546
		0.8315	0.9337	0.8268	0.9367	0.8254	0.9442	0.8153	0.9472
(1,2)	0.63	1.8931	0.9512	1.8610	0.9520	1.8338	0.9528	1.8128	0.9528
		1.7613	0.9329	1.7632	0.9419	1.7614	0.9467	1.7422	0.9441
(1,1)	2.0	4.2241	0.9491	4.0720	0.9483	3.9453	0.9500	3.8762	0.9520
		3.7013	0.9355	3.6847	0.9387	3.6630	0.9429	3.6042	0.9452
(1,0.7)	3.5	8.4843	0.9559	7.8801	0.9511	7.3934	0.9513	7.0692	0.9494
		6.6615	0.9363	6.5965	0.9362	6.5143	0.9426	6.3125	0.9395
(1,0.5)	6.6	19.5860	0.9511	17.5046	0.9490	16.3039	0.9500	15.4300	0.9516
		15.0230	0.9358	14.3997	0.9373	14.0955	0.9431	13.3677	0.9463
		99% Prediction intervals							
(1,5)	−0.25	1.2888	0.9935	1.2160	0.9937	1.1699	0.9934	1.1355	0.9937
		1.0708	0.9798	1.0604	0.9853	1.0578	0.9870	1.0523	0.9878
(1,2)	0.63	2.5793	0.9927	2.4989	0.9939	2.4352	0.9924	2.3884	0.9933
		2.2242	0.9793	2.2280	0.9840	2.2274	0.9863	2.2078	0.9885
(1,1)	2.0	7.0463	0.9922	6.5645	0.9931	6.1901	0.9931	5.9829	0.9940
		5.3682	0.9796	5.3363	0.9841	5.2925	0.9857	5.1944	0.9888
(1,0.7)	3.5	18.0705	0.9933	15.5408	0.9934	14.1481	0.9937	13.1202	0.9940
		11.3955	0.9811	11.0975	0.9823	11.0236	0.9850	10.6166	0.9863
(1,0.5)	6.6	56.0809	0.9929	45.8804	0.9942	40.2587	0.9951	36.5485	0.9932
		32.0789	0.9805	30.2839	0.9843	29.4231	0.9885	27.5576	0.9870

[a] The skewness ω is defined as (third central moment)/(second central moment)$^{3/2}$.

4. A numerical example

In this section, we consider a real life example to illustrate the interval and further compare it with the existing one. The data considered are the test results (millions of revolutions before failure) on the endurance of deep-groove ball bearings:

17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84, 51.96, 54.12, 55.56,

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

Table 4
Prediction limits for the ball bearings data

	The PI with $\hat{\lambda}_{BC}$	The new PI	The existing PI
90%	(23.55, 152.6)	(19.07, 142.91)	(17.96, 143.06)
95%	(18.55, 179.26)	(12.33, 161.38)	(12.23, 158.20)
99%	(10.85, 247.07)	(2.67, 203.09)	(4.90, 188.22)

67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92,

128.04, 173.40.

The data set was originally given by Lieblein and Zelen (1956) and has been used by numerous authors for illustrating the applications of the Weibull distribution. The sample skewness is 0.9206, indicating that the data is moderately skewed to the right. The transformation estimators are $\hat{\lambda}_{BC} = 0.1905$ and $\hat{\lambda}_{ML} = 0.5579$, showing that the two estimators can differ substantially. The 90%, 95% and 99% PIs are calculated using the existing method as well as the new method based on $\hat{\lambda}_{ML}$ or $\hat{\lambda}_{BC}$. The results are summarized in Table 4.

The new 90% PI is shorter, 95% PI is about the same as the existing one and the 99% PI is longer than the existing one. However, after the adjustment the new 99% PI becomes (4.16, 194.03), very close to the existing one in interval length. The new PI after adjustment has a confidence level about 99%, but the existing one has a confidence level lower than 99%. It is also interesting to compare the new PI with the PI using $\hat{\lambda}_{BC}$. The results show that the latter is much wider, showing the gains by using a better transformation estimator. The observations from this real data example closely agree with the theory and simulation results given in the earlier sections.

5. Discussion

A simple prediction interval for the Weibull distribution is given in this paper. It is obtained by first transforming the Weibull observations to near normality, constructing a prediction interval (PI) for a transformed future observation in the usual way, and then inverting this interval to give a PI for the original future observation. The normalizing transformation is a simple power transformation with the estimated ‘power’ parameter being simply a constant times the MLE of the shape parameter. The interval is seen to meet the existing one in terms of combined consideration of the coverage probability and interval length, but its simplicity makes it attractive.

It should be noted that the idea of constructing PI based on normalizing transformation can be extended to work for any other distribution with domain being positive half-real line. This is specifically meaningful for distributions where no exact methods are available, such as the popular gamma distribution (see Yang, 1999b).

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

When PIs of high confidence levels (such as 99%) are of interest, use of the Box–Cox estimate $\hat{\lambda}_{\mathrm{BC}}$ may give the quantity inside the curling brackets of (3.3) being negative due to the large variability of $\hat{\lambda}_{\mathrm{BC}}$. As a result, the PI becomes undefined (see Yang, 1999b, p. 274). This problem does not show up when $\hat{\lambda}_{\mathrm{ML}}$, the MLE of the transformation parameter, is used.

When it is known which distribution the data came from, the PI can be adjusted to have an asymptotically correct coverage. This is not possible with the general Box–Cox procedure under no distributional assumption. However, when it is not clear which lifetime distribution that the data come from, the more robust Box–Cox procedure should be used.

Although it makes it possible to use the simple normal method to predict the future Weibull lifetime, the transformation approach has a limitation: no straightforward generalization of the above result to censored data while keeping its simplicity. However, it is in practice not uncommon to have complete life data and in this case the above result greatly simplifies the predictive inference for the Weibull distribution.

Acknowledgements

The authors would like to thank the Co-Editor, Professor Erricos John Kontoghiorghes, the Associate Editor, and the three referees for the helpful comments.

Appendix A. Proof of Theorem 3.1.

First we show $\bar{X}(\hat{\lambda}) \xrightarrow{\mathrm{p}} \mu$ and $s^2(\hat{\lambda}) \xrightarrow{\mathrm{p}} \sigma^2$. A Taylor series expansion gives

$$\bar{X}(\hat{\lambda}) = \bar{X}(\lambda) + (\hat{\lambda} - \lambda)[\bar{X}'(\lambda) + R_n],$$

where

$$\bar{X}'(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathrm{d}X_i(\lambda)}{\mathrm{d}\lambda} = \frac{1}{n\lambda}\sum_{i=1}^{n} [X_i^{\lambda}\log(X_i) - X_i(\lambda)].$$

Now, $EX_i(\lambda)$ is finite. Letting $Y_i = (X_i/\alpha)^{\beta}$ gives

$$E(X_i^{\lambda}\log X_i) = \int_0^{\infty} \alpha^{0.2654\beta}(\log\alpha + \beta^{-1}\log y_i)y_i^{0.2654}\mathrm{e}^{-y_i}\,\mathrm{d}y_i,$$

which is finite as well since it is a weighted sum of gamma and digamma functions. The law of large number thus ensures that $\bar{X}'(\lambda)$ converges in probability, hence $R_n \xrightarrow{\mathrm{p}} 0$ as $\hat{\lambda} \xrightarrow{\mathrm{p}} \lambda$, so is the second term. Also, $\bar{X}(\lambda) \xrightarrow{\mathrm{p}} \mu$, which gives $\bar{X}(\hat{\lambda}) \xrightarrow{\mathrm{p}} \mu$. Now,

$$s^2(\hat{\lambda}) = \frac{1}{n-1}\sum_{i=1}^{1}[X_i(\hat{\lambda}) - \bar{X}(\hat{\lambda})]^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2(\hat{\lambda}) - \frac{n}{n-1}\bar{X}^2(\hat{\lambda}).$$

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

The last term converges in probability to μ^2 and the first term becomes by a Taylor expansion

$$\frac{1}{n-1}\sum_{i=1}^{n} X_i^2(\hat\lambda) = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2(\lambda) + (\hat\lambda - \lambda)\left[\frac{2}{n-1}\sum_{i=1}^{n} X_i(\lambda)X_i'(\lambda) + R_n^*\right].$$

Again letting $Y_i = (X_i/\alpha)^\beta$, we have

$$E(X_i^{2\lambda}\log X_i) = \int_0^\infty \alpha^{0.5308\beta}(\log\alpha + \beta^{-1}\log y_i)\,y_i^{0.5308}\mathrm{e}^{-y_i}\,\mathrm{d}y_i,$$

which is finite as it is a weighted sum of gamma and digamma functions. This shows the finiteness of $E[X_i(\lambda)X_i'(\lambda)]$. Hence, the second term in the above Taylor expansion converges in probability to zero and

$$\frac{1}{n-1}\sum_{i=1}^{n} X_i^2(\hat\lambda) \sim \frac{1}{n-1}\sum_{i=1}^{n} X_i^2(\lambda) \xrightarrow{\mathrm{p}} \sigma^2 + \mu^2,$$

which gives $s^2(\hat\lambda) \xrightarrow{\mathrm{p}} \sigma^2$.

Second, we show that $X^0(\hat\lambda) \xrightarrow{\mathrm{D}} X^0(\lambda)$, which is trivial by the Taylor expansion

$$X^0(\hat\lambda) = X^0(\lambda) + (\hat\lambda - \lambda)[\mathrm{d}X^0(\lambda)/\mathrm{d}\lambda + R].$$

Finally, an application of Slutsky's theorem gives

$$T(\hat\lambda) = [X^0(\hat\lambda) - \bar X(\hat\lambda)]/s(\hat\lambda) \xrightarrow{\mathrm{D}} [X^0(\lambda) - \mu]/\sigma,$$

which implies that

$$\begin{aligned} P\{L_\delta(\mathbf{X}) \leqslant X^0 \leqslant U_\delta(\mathbf{X})\} &= P\{-t_{n-1}(\delta/2) \leqslant T(\hat\lambda) \leqslant t_{n-1}(\delta/2)\} \\ &\to P\{\mu - Z_{\delta/2}\sigma \leqslant X^0(\lambda) \leqslant \mu + Z_{\delta/2}\sigma\}. \end{aligned}$$

The relations $\mu = (0.9034\alpha^\lambda - 1)/\lambda$ and $\sigma = 0.2675\alpha^\lambda/\lambda$ give the result of the theorem.

References

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Statist. Soc. Ser. B 26, 211–252.

Dellaportas, P., Wright, D.E., 1991. Numerical prediction for the two-parameter Weibull distribution. The Statistician 40, 365–372.

Engelhardt, M., Bain, L.J., 1979. Prediction limits and two-sample problems with complete or censored Weibull data. Technometrics 21, 233–237.

Engelhardt, M., Bain, L.J., 1982. On prediction limits for samples from a Weibull or extreme-value distribution. Technometrics 24, 147–150.

Fertig, K.W., Meyer, M.E., Mann, N.R., 1980. On constructing prediction intervals for samples from a Weibull or extreme value distribution. Technometrics 22, 567–573.

Hernandez, F., Johnson, R.A., 1980. The large-sample behavior of transformations to normality. J. Amer. Statist. Assoc. 75, 855–861.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions. Wiley, New York.

Kullback, S., 1968. Information Theory and Statistics. Dover Publications, New York.

Lieblein, J., Zelen, M., 1956. Statistical investigation of the fatigue life of deep-groove ball bearings. J. Res. Nat. Bur. Standards 47, 273–317.

Appeared in: Computational Statistics and Data Analysis, 2003, 43, 357-368.

Mee, R.W., Kushary, D., 1994. Prediction limits for the Weibull distribution utilizing simulation. Comput. Statist. Data Anal. 17, 327–336.

Sakia, R.M., 1992. The Box–Cox transformation technique—a review. The Statistician 41 (2), 169–178.

Yang, Z.L., 1999a. Estimating a transformation and its effect on Box–Cox T-ratio. Test 8, 167–190.

Yang, Z.L., 1999b. Predicting a future lifetime through Box–Cox transformation. Lifetime Data Anal. 5, 265–279.