An Investigation of Transformation-based Prediction Interval for the Weibull Median Life

Zhenlin Yang*, Stanley P. See
Department of Statistics and Applied Probability, National University of Singapore

Min Xie
Department of Industrial and System Engineering, National University of Singapore

Abstract

Statistical inference based on the Weibull distribution, a distribution widely used in reliability and survival analysis, is usually difficult as it often involves numerical computation and approximation. However, this distribution can be transformed to near-normality by a simple power transformation. Based on this transformation, a prediction interval (PI) for its median can be easily constructed through an inverse transformation. The procedure for selecting the best power transformation through minimizing Kullback-Leibler information is described. The property of this transformation-based PI is investigated. Simple correction factors are also proposed. It is shown that the transformation-based PI with corrections performs well, irrespective of the sample size and parameter values. Simulation results show that the new PI generally outperforms the existing PI. Numerical examples are given for illustration.

Keywords : Kullback-Leibler information, Weibull median life, coverage probability, prediction interval, variance correction.

* To whom correspondence should be addressed. Department of Statistics and Applied Probability, National University of Singapore, 3 Science Drive 2, Singapore 117543. Tel: (65) 874-6829, Fax: (65) 872-3919, email: stayzl@nus.edu.sg.

1. Introduction

The median of a lifetime distribution is usually interpreted as the 'typical' life or the 'characteristic' life of a population. Hence inference concerning the median is often an interesting study in the fields of reliability, quality control, medical and biological sciences, etc. The Weibull distribution (Weibull, 1951) is one of the most popular lifetime distributions upon which numerous research articles have been published and active research is still going on (Bain and Engelhardt, 1991 and Johnson *et al.*, 1994), especially in relation to engineering and medical applications. However, simple and accurate statistical methods for basic problems such as prediction interval for the median do not seem to exist, as the statistical inference for the Weibull distribution is generally difficult. Nelson (1982, p232) describes an approximate method that is rough unless the sample size is larger than 100. Lawless (1974, 1978) gives a method for exact conditional confidence limits, but the method requires a special computer program for its implementation.

On the other hand, most lifetime distributions are transformable to near-normality (Hernandze and Johnson, 1980; Yang 1999b), hence certain statistical intervals can be constructed through an inverse transformation if the quantity of interest is invertable (Hahn and Meeker, 1991, p72-74), such as the median or general percentiles. This approach is attractive for its simplicity hence should be recommended for the cases where the existing methods are too complicated to be implemented in practice. It usually works well if the data can be transformed to exact normality and the transformation is known. Often in practice, however, the transformation may be known only up to a certain functional form. Certain transformation parameter(s) have to be decided based on the data (Box and Cox, 1964). Also, in many situations, even the 'best' transformation may only be able to transform the data to near-normality. Hence for the transformation-based prediction or prediction intervals, there are two general issues that require rigorous examination, namely, the effect of nonnormality and the effect of estimating the transformation, which are often ignored by practitioners.

In this article, we explore a transformation approach for the construction of a simple prediction interval for the Weibull median and compare it with the one described in Nelson (1982). The Weibull distribution can be transformed to a near-normal distribution, and many statistical methods for normal distribution can then be applied. Using the transformation approach, one first transforms the data by some monotonic transformation so that the transformed data become closely normally distributed. Then a prediction interval for the median of the transformed future observation can be derived. Finally, the interval for the transformed median can be inverted to give a prediction interval for the original median. In this article, a simple power transformation is considered. The effect of nonnormality and the effect of estimating transformation are quantified.

The paper is organised as follows. In Section 2, the transformation-based PI for the Weibull median is outlined, and its asymptotic property is discussed. Section 3 presents some theoretical results that quantify the large sample effect of nonnormality and the effect of estimating transformation. Based on this theory, simple correction factors are proposed. Section 4 presents simulation results for the small sample behaviour of the proposed and existing PIs. Two numerical examples are given in Section 5 for further comparisons and for illustrations. Finally, a general discussion is given in Section 6.

The simulation results show that the corrected transformation-based PI performs very well in general, irrespective of the sample size and values of the parameters. When sample size is not large the new PI outperforms the existing PI, particularly in terms of the coverage probability. This means that the new PI is not only simple but also accurate.

The results shed much light on the application of the transformation approach to more complicated Weibull prediction problems concerning percentiles or reliabilities, etc., and to the prediction problems under a Weibull regression model.

2. The Transformation-based Prediction interval

Let **X** = $(X_1, X_2, \ldots, X_n)$ be a sample of past observations from a Weibull population and X^0 a single future observation from the same population with the cumulative distribution function (CDF)

$$F(x;\alpha,\beta) = 1 - \exp[-(x/\alpha)^{\beta}],$$

where α is the scale parameter and β is the shape parameter. Let θ_0 be the median of X^0. We are interested in constructing a prediction interval for θ_0 using the transformation approach and compare it with the existing one.

2.1. Power transformation of Weibull to near-normality

Let $X(\lambda)$ be a monotonic transformation of a Weibull random variable X and $f(y;\alpha,\beta,\lambda)$ be its probability density function (pdf). Let $\phi(y;\mu,\sigma)$ be the pdf of a normal random variable with a mean μ and a standard deviation σ. Clearly, we want to find a transformation such that the two pdfs $f(y;\alpha,\beta,\lambda)$ and $\phi(y;\mu,\sigma)$ are closest in a certain sense for some chosen λ, μ and σ. A popular measure of discrepancy between two pdfs is called *Kullback-Leibler (KL) information* (Kullback, 1968) defined as

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

When $X(\lambda)$ is the simple power transformation $X(\lambda) = X^{\lambda}$, Hernandez and Johnson (1980) showed that the best normalizing transformation for the Weibull distribution in the sense of a minimized KL information has the power parameter given as

$$\lambda_0 = \ell_0 \beta,$$

where ℓ_0 is the solution of the following equation

$$\gamma - \frac{1}{\ell_0} + \frac{\Gamma(1+2\ell_0)\Psi(1+2\ell_0) - \Gamma^2(1+\ell_0)\Psi(1+\ell_0)}{\Gamma(1+2\ell_0) - \Gamma^2(1+\ell_0)} = 0, \tag{2.1}$$

where γ is the Euler constant, $\Gamma(\alpha)$ is the gamma function and $\Psi(\alpha)$ is the di-gamma function that is defined as the derivative of $\log\Gamma(\alpha)$. All the functions can be easily calculated using some statistical software such as MATHEMAICA, and the equation (2.1) can be easily solved, which gives a value 0.2654 for ℓ_0 (up to four decimal points).

The corresponding mean and standard deviation for the transformed random variable $X(\lambda_0)$ are $\mu_0 = \alpha^{\lambda_0}\,\Gamma(1+\ell_0)$ and $\sigma_0 = \alpha^{\lambda_0}\,[\Gamma(1+2\ell_0) - \Gamma^2(1+\ell_0)]^{1/2}$. The minimum value of $I(f;\phi)$ is 0.00278, independent of α and β, indicating that $f(y;\alpha,\beta,\lambda_0)$ and $\phi(y;\mu_0,\sigma_0)$ are very close and that the closeness cannot be further improved using the power transformation. The simple relationship between the transformation parameter and the Weibull shape parameter allows us to estimate the transformation parameter in a simple way: $\hat{\lambda} = \ell_0\,\hat{\beta}$ where $\hat{\beta}$ is the MLE of β defined as the solution of

$$\frac{1}{\hat{\beta}} = \left(\sum_{i=1}^{n} X_i^{\hat{\beta}} \log X_i\right)\left(\sum_{i=1}^{n} X_i^{\hat{\beta}}\right)^{-1} - \frac{1}{n}\sum_{i=1}^{n}\log X_i\,. \tag{2.2}$$

Alternatively, an estimator of the transformation parameter can be obtained based on the general procedure given in Box and Cox (1964). This procedure works for all nonnegative continuous observations and the Box-Cox estimator of transformation is defined as

$$\hat{\lambda}_{\mathrm{B}} = \arg\min_{\ell} \dot{X}^{-\ell} s(\ell)$$

where $\dot{X}$ is the geometric mean of the X's and $s(\ell)$ is the standard deviation of the transformed sample in ℓ scale.

Authors who have contributed to the applications of the Box-Cox transformation technique to prediction problems include: Carroll and Ruppert (1981, 1991), Duan (1983), Taylor (1985), Collins (1991), Hamilton and Taylor (1993), Yang (1999b), among others. Specifically, Carroll and Rupper (1981) studied the effect of estimating transformation on the estimation of the median in original scale and concluded that this effect is not large. Sakia (1992) gave a review on the Box-Cox transformation technique.

Intuitively our approach should be simpler and more efficient as extra information regarding the distribution is used. In fact, using the asymptotic results by Yang (1999a, p176) and Bain and Engelhardt (1991, p217), one can easily see that $\hat{\lambda}_{\mathrm{B}}$ could be as much as about ten times more variable than $\hat{\lambda}$. For this reason, we adopt the estimator $\hat{\lambda}$ for the development of a prediction interval for the Weibull median using transformation approach.

2.2. The transformation-based prediction interval

For a set of past observations, the transformation-based prediction interval is developed by first transforming the data, obtaining a prediction interval for the transformed data, and then invert the prediction interval back to the original scale.

Let $X_1(\lambda_0)$, $X_2(\lambda_0)$, ..., $X_n(\lambda_0)$ be the transformed past sample and $X^0(\lambda_0)$ be the transformed future observation. Let $\overline{X}(\lambda_0)$ and $s(\lambda_0)$ be the mean and standard deviation of the transformed past sample. Since $X_i(\lambda_0)$'s are approximately normal and are independent, we have $E[X^0(\lambda_0)] \approx \text{Med}[X^0(\lambda_0)]$ and

$$T(\lambda_0) = \frac{\bar{X}(\lambda_0) - \text{Med}[X^0(\lambda_0)]}{s(\lambda_0)/\sqrt{n}} \overset{approx.}{\sim} t_{n-1}.$$

An approximate 100(1-δ) % prediction interval (PI) for $\text{Med}[X^0(\lambda_0)]$ is obtained as:

$$\bar{X}(\lambda_0) \pm t_{n-1}(\delta/2)\, s(\lambda_0)/\sqrt{n},$$

where $t_{n-1}(\delta/2)$ is the upper 100($\delta/2$) percentage value of a t distribution with $n-1$ degrees of freedom. Since the power transformation is monotonic, it is clear that $\text{Med}[X^0(\lambda_0)] = [\text{Med}(X^0)]^{\lambda_0} = \theta_0^{\lambda_0}$. An approximate 100(1-$\delta$) % PI for θ_0 is obtained by a simple inverse transformation:

$$\{\bar{X}(\lambda_0) \pm t_{n-1}(\delta/2)\, s(\lambda_0)/\sqrt{n}\}^{1/\lambda_0}.$$

Using the central limit theorem and the laws of large numbers, one can easily see that as long as the Weibull observations can be transformed to have the same mean and median, $T(\lambda_0)$ converges to the standard normal, and hence the above interval has a correct asymptotic coverage. However, this interval assumes that the true transformation parameter λ_0 is known, which clearly is not a realistic assumption. When λ_0 is unknown, a common practice is to replace it by its estimator $\hat{\lambda}$ and give a PI for θ_0 as

$$\{\bar{X}(\hat{\lambda}) \pm t_{n-1}(\delta/2)\, s(\hat{\lambda})/\sqrt{n}\}^{1/\hat{\lambda}}. \tag{2.3}$$

The PI given by Equation (2.3) is very simple, especially when $\hat{\lambda}$ is determined by the MLE procedure outlined in the earlier subsection. It, however, ignores two things: one is the effect of nonnormality, in particular, the equality $\text{Med}[X^0(\lambda_0)] = \text{E}[X^0(\lambda_0)]$ does not hold exactly, and the other is the effect of estimating the transformation. We will study these issues theoretically by providing some asymptotic results and numerically for small samples using Monte Carlo simulation in the subsequent sections. It should be pointed out that a common erroneous impression about the interval (2.3) is that, similar to the case of λ_0 known, it also possesses a correct asymptotic coverage. Our result given in next section indicates that it is not true and hence the interval needs some corrections.

3. Prediction Interval with Correction Factor

Clearly, for the PI (2.3) to have good asymptotic property, it is necessary that the following pivotal quantity

$$T(\hat{\lambda}) = \frac{\overline{X}(\hat{\lambda}) - \theta_0^{\hat{\lambda}}}{s(\hat{\lambda})/\sqrt{n}}$$

is approximately t distributed with n–1 degrees of freedom or at least converges to the standard normal as n becomes large. A simple Taylor expansion shows that $T(\hat{\lambda})$ does not converge to the standard normal, hence the interval (2.3) does not have a correct asymptotic coverage. This is definitely an undesirable property for any statistical interval, hence necessary corrections need to be considered. The following theory quantifies of effect of estimating transformation and the effect of nonnormality and provides a theoretical base for the introduction of the correction factors.

Theorem 3.1. Let $\hat{\lambda}$ be MLE of λ_0, i.e., $\hat{\lambda} = \ell_0 \hat{\beta}$ with $\hat{\beta}$ being the solution of (2.2). We have

$$T^*(\hat{\lambda}) = \frac{\overline{X}(\hat{\lambda}) - c_b\theta_0^{\hat{\lambda}}}{s(\hat{\lambda})/\sqrt{n}} \xrightarrow{D} N(0, 1 + c_v^2)$$

where the bias correction factor c_b and the variance correction factor c_v^2 are given as

$$c_b = \Gamma(1+\ell_0)\exp(-u_0\ell_0) \quad \text{and} \quad c_v^2 = k_1^2 + 2k_1k_2$$

with $k_1 = \ell_0\, k_0\, \Gamma(1+\ell_0)\,[\Psi(1+\ell_0) - u_0\,][\Gamma(1+2\ell_0) - \Gamma^2(1+\ell_0)]^{-1/2}$,

$k_0^2 = \lim_{n\to\infty} Var\left[\sqrt{n}\,(\hat{\beta}-\beta)/\beta\right]$,

$u_0 = \log[-\log(0.5)]$, and

$k_2 = \sigma_0^{-1}\lambda_0^{-1}\lim_{n\to\infty} nE\left[(\overline{X}(\lambda_0)-\mu_0)(\hat{\lambda}-\lambda_0)\right]$.

The proof of Theorem 3.1 is lengthy but straightforward, hence is put in the Appendix. Theorem 3.1 tells that estimating transformation inflates the variance of the pivotal quantity to be used in the PI construction. It suggests that in order for the PI (2.3) to perform well, at least asymptotically, two corrections are needed: one is c_b, related to the bias of estimation and called *bias correction factor* and the other is c_v^2, related to the variance of $T(\hat{\lambda})$ hence called the *variance correction factor*. The PI after corrections thus takes the form:

$$\left\{c_b^{-1}\left[\overline{X}(\hat{\lambda}) \pm t_{n-1}(\delta/2)\, s(\hat{\lambda})\sqrt{(1+c_v^2)/n}\right]\right\}^{1/\hat{\lambda}}. \tag{3.1}$$

The value for c_b up to four decimal places is **0.9957** and the value of c_v^2 is approximately **0.1168**, where k_0^2 = 0.6079 (from Bain and Engelhardt, 1991, p219) and k_2 = 0.4767 (obtained by Monte Carlo simulation). The constant k_2 is a pure number with its value being difficult to be obtained in an analytical way, hence Monte Carlo simulation is employed. It seems that the bias correction factor may be negligible, but the variance correction factor is not. Simple calculation gives $(1+c_v^2)^{1/2} \approx$ 1.0568, which would affect the interval performance significantly.

4. Monte Carlo Simulation and Comparison

In the previous section, the transformation-based PI (3.1) is derived and is shown to possess a nice asymptotic property. However, it is important to investigate the performance of this PI when the sample size is not large. In this section, a simulation study is carried out to investigate the small sample property of the new PI and to compare it with the existing PI that is reported in Nelson (1982, p232). To help seeing the gains of introducing correction factors, the results for the PI (2.3) are also reported.

The simulation process can be simply described as follows. For each combination of the parameter values (α, β), the sample size n and the nominal level $(1-\delta)$, 10,000 Weibull random samples are generated and the three PIs (transformation-based, corrected transformation-based and Nelson's) are calculated for each sample. The proportions of the PIs that cover the median are used as Monte Carlo estimates of the coverage probability of the intervals and the average lengths of these intervals are served as Monte Carlo estimates of the expected lengths of the intervals. The performance of the PIs are not affected by the scale parameter α, hence it is fixed at a value 1. Three difference values of β are considered, resulted in population skewness from small to large. Four difference sample sizes (small to large) and three nominal levels are considered. The simulated average length (A.L.) and coverage probability (C. Prob.) for the three PIs are summarized in Table 4.1.

The simulation results show that the transformation-based PI with corrections performs very well, irrespective of the choices of sample size, parameter value and nominal level. All the simulated coverage probabilities are very close to the corresponding nominal levels. The lengths are all comparable to the existing PI. However, the existing PI has rather poor coverage when n is small and it seems that it deteriorates further as population skewness ω increases. For examples, when n = 10, the simulated coverage probabilities could be as low as 83%, 89% and 95%, respectively, with the corresponding nominal levels 90%, 95% and 99%. Notice that the transformation-based PI without corrections also performs better than the existing one in terms of coverage probability when n is small. Simulation results also show that the gains of introducing the correction factors are generally quite significant, especially for the 90% and 95% PIs.

Table 4.1. A summary of simulation results for PI (3.1) (first row), PI (2.3) (second row) and the Nelson (1982) (third row).

		n = 10		20		30		50	
(β, α)	ω	A.L.	C.Prob.	A.L.	C.Prob	A.L.	C.Prob.	A.L.	C.Prob.
		90% Prediction Intervals							
(2,1)	.63	0.5225	0.8933	0.3631	0.8907	0.2966	0.8978	0.2290	0.8996
		0.4914	0.8712	0.3421	0.8754	0.2787	0.8771	0.2152	0.8778
		0.4716	0.8429	0.3514	0.8732	0.2922	0.8840	0.2291	0.8933
(1,1)	2.0	0.9137	0.8890	0.6208	0.8946	0.5015	0.9004	0.3851	0.8984
		0.8445	0.8702	0.5780	0.8731	0.4662	0.8783	0.3580	0.8778
		0.8350	0.8314	0.6053	0.8720	0.4965	0.8840	0.3868	0.8926
(.5,1)	6.6	1.6666	0.8957	0.9872	0.8958	0.7591	0.8982	0.5619	0.9009
		1.4856	0.8712	0.9017	0.8725	0.6904	0.8718	0.5144	0.8821
		1.5594	0.8378	0.9844	0.8717	0.7606	0.8858	0.5708	0.8996
		95% Prediction Intervals							
(2,1)	.63	0.6450	0.9466	0.4395	0.9459	0.3570	0.9471	0.2745	0.9501
		0.6066	0.9303	0.4141	0.9335	0.3354	0.9343	0.2579	0.9350
		0.5659	0.8969	0.4202	0.9226	0.3490	0.9357	0.2734	0.9445
(1,1)	2.0	1.1429	0.9419	0.7555	0.9492	0.6057	0.9538	0.4625	0.9502
		1.0549	0.9286	0.7031	0.9336	0.5629	0.9349	0.4299	0.9314
		1.0219	0.8885	0.7312	0.9245	0.5971	0.9382	0.4635	0.9387
(.5,1)	6.6	2.2321	0.9460	1.2387	0.9482	0.9352	0.9481	0.6829	0.9533
		1.9730	0.9316	1.1270	0.9328	0.8486	0.9323	0.6242	0.9378
		2.0445	0.8939	1.2337	0.9238	0.9382	0.9342	0.6947	0.9441
		99% Prediction Intervals							
(2,1)	.63	0.9271	0.9880	0.6007	0.9894	0.4811	0.9896	0.3661	0.9895
		0.8720	0.9859	0.5660	0.9845	0.4520	0.9831	0.3439	0.9849
		0.7568	0.9573	0.5570	0.9719	0.4613	0.9804	0.3605	0.9841
(1,1)	2.0	1.7107	0.9888	1.0479	0.9912	0.8237	0.9906	0.6199	0.9915
		1.5726	0.9859	0.9739	0.9871	0.7648	0.9864	0.5758	0.9841
		1.4339	0.9513	0.9935	0.9759	0.8025	0.9829	0.6174	0.9835
(.5,1)	6.6	4.0522	0.9897	1.8608	0.9892	1.3378	0.9894	0.9428	0.9902
		3.5005	0.9855	1.6765	0.9850	1.2066	0.9858	0.8586	0.9834
		3.3893	0.9534	1.8308	0.9739	1.3403	0.9807	0.9613	0.9842

5. Numerical Examples

In this section, we use two real life examples to illustrate the transformation-based PIs and further compare them with the existing one. Both data sets have been extensively used for testing the statistical techniques developed for certain lifetime model including the Weibull.

Example 5.1: the vehicle failure data. This data set was reported by Bilikan *et al.* (1979) and used again by Cheng and Iles (1990) to illustrate their methods for fitting a three parameter lifetime distribution. The data represent the mileages to failure of a type of vehicle: 164, 250, 439, 440, 450, 478, 487, 524, 688, 850, 1048, 1280, 1364, 1488, 1513, 1860, 1947, 1991, 2200, 2446.

Example 5.2: the repair time data: This data set, taken from Chhikara and Folks (1989, p139) where a 0.5 was missed from the original data set, contains the repair times (in hours) for an airborne communication transceiver. The data is given as: 0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2, 2.5, 2.7, 3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5.

Table 5.1. The PIs based on real life data.

	PI (3.1)		PI (2.3)		The existing PI	
	Lower	Upper	Lower	Upper	Lower	Upper
Vehicle Failure Data : n = 20, $\hat{\lambda}$ = 0.4292, $\hat{\omega}$ = 0.4032						
90%	714.6624	1272.8924	720.0704	1242.9814	748.3878	1255.1046
95%	666.3705	1342.4290	674.3180	1307.5682	712.2276	1318.8270
99%	570.7345	1496.1442	583.3476	1450.0315	646.5259	1452.8498
Repair Time Data : n = 46, $\hat{\lambda}$ = 0.2385, $\hat{\omega}$ = 2.8568						
90%	1.6701	3.0448	1.6690	2.9458	1.8000	2.9771
95%	1.5653	3.2182	1.5701	3.1049	1.7154	3.1241
99%	1.3690	3.5910	1.3842	3.4460	1.5612	3.4326

The calculated prediction intervals are summarized in Table 5.1. The results show that the existing PI (Nelson, 1982, p232) and the transformation-based PI without corrections given in (2.3) are shorter than the transformation-based PI with corrections given in (3.1), indicating the two intervals are a bit too tight, especially when n is small. When n is large (the case of second data set), the difference in interval lengths is not significant. These results are consistent with the simulation results given in the earlier section.

6. Discussions

The problem of obtaining a prediction interval for the median of a future Weibull observation is studied. The approach we followed is to first transform the Weibull observations to near-normality, construct an interval for the transformed future median and then invert. The best power transformation is obtained through minimizing the Kullback-Leibler information. Both the large sample and small sample properties of the transformation-based interval are studied and simple correction factors are introduced. It is shown that the transformation-based prediction interval with simple corrections outperforms the existing one in terms of coverage probability.

Considering both simplicity and accuracy, the results obtained in this paper favour the transformation approach. Similar results can be expected for the general Weibull prediction problems concerning the percentiles or reliabilities. The transformation approach for prediction interval construction may be also applicable to the failure time regression case where the failure times are Weibull distributed, but dependent on certain concomitant variables. It might be interesting to give theoretical investigations for these situations.

The Box-Cox transformation is usually applied to the complete data. The effect of censoring requires further investigation that is beyond the scope of this work and the theoretical results as those in Theorem 3.1 are not readily available. However, if one is interested in using the transformation approach to do prediction with censored data, one possible way is to adapt the general method described by Meeker and Escobar (1998, p296) for a (transformed) location and scale family. Clearly, the issue of the effect of estimating transformation becomes more complicated. We will pursue this study in a future paper.

Acknowledgements

We are grateful to the referee for the comments that led to improvements of this article.

Appendix : Proof of Theorem 3.1

The proof of Theorem 3.1 requires two lemmas that are given below.

Lemma A. Let X be a Weibull random variable with scale and shape parameters α and β, respectively. Then $E\left(X^{\lambda}\log X\right)$ is finite for some real $\lambda > -\beta$.

Proof: First, we have that

$$E\left(X^{\lambda}\log X\right) = \int_0^{\infty}(x^{\lambda}\log x)\beta\alpha^{-\beta}x^{\beta-1}\exp\left[-(x/\alpha)^{\beta}\right]dx\,.$$

Let $y = (x/\alpha)^{\beta}$, the above integral becomes,

$$\alpha^{\lambda}\int_0^{\infty}(\log\alpha+\beta^{-1}\log y)y^{\lambda/\beta}\exp(-y)dy$$

$$= \alpha^{\lambda}\left[\log\alpha\Gamma(1+\lambda/\beta)+\beta^{-1}\Gamma(1+\lambda/\beta)\Psi(1+\lambda/\beta)\right],$$

where $\Gamma(\cdot)$ is the gamma function and $\Psi(\cdot)$ is the digamma function that is the derivative of $\log\Gamma(\cdot)$. Clearly the last integral is defined if $\lambda > -\beta$.

Lemma B. Let $\hat{\lambda}$ be the MLE of λ_0, i.e., $\hat{\lambda} = \ell_0 \hat{\beta}$, with ℓ_0 being the solution of (2.1) and $\hat{\beta}$ being the solution of (2.2), then $s^2(\hat{\lambda}) \xrightarrow{p} \sigma_0^2$.

Proof: A first-order Taylor expansion gives

$$s^2(\hat{\lambda}) = s^2(\lambda_0) + (\hat{\lambda}-\lambda_0)\left\{\frac{2}{n-1}\sum_{i=1}^{n}\left[X_i(\lambda_0)X_{i\lambda_0}(\lambda_0)\right]-\frac{n}{n-1}\bar{X}(\lambda)\bar{X}_{\lambda_0}(\lambda_0)+R_n\right\},$$

where $X_{i\lambda_0}(\lambda_0) = dX_i(\lambda_0)/d\lambda_0$ and $\bar{X}_{\lambda_0}(\lambda_0) = \frac{1}{n}\sum_{i=1}^{n} X_{i\lambda_0}(\lambda_0)$. Lemma A and the weak laws of large numbers (WLLN) ensure that the first two quantities in the curling brackets converge in probability. Hence $R_n \to 0$ as $\hat{\lambda} \to \lambda_0$, which means the whole second term converges in probability to zero. It is easy to see by WLLN that $s^2(\lambda_0) \xrightarrow{p} \sigma_0^2$, hence the result follows.

Proof of Theorem 3.1: We argue that $T^*(\hat{\lambda})$ is asymptotically normal and then find the asymptotic mean and variance of it. First, the joint asymptotic normality of $\sqrt{n}(\hat{\beta}-\beta)$ and $\sqrt{n}(\hat{\mu}-\mu_0)$, where $\hat{\mu} = \bar{X}(\ell_0\hat{\beta})$, can be easily established by the standard theory of M-estimation based on the estimating equation

$$n^{-1}\sum_1^n \psi_i(X_i;\hat{\beta},\hat{\mu}) = 0,$$

where $\psi_i' = \{\, X_i^{\hat{\beta}}(\hat{\beta}\log X_i - \hat{\beta}\bar{g} - 1)\ ,\ X_i^{\ell_0\hat{\beta}} - \hat{\mu}\,\}$ and $\bar{g} = n^{-1}\sum_i^n \log X_i$. Now a first-order Taylor expansion gives $c_b\theta_0^{\hat{\lambda}} = \mu_0 + (\hat{\lambda}-\lambda_0)\,c_b\theta_0^{\lambda_0}\log\theta_0 + O_p(n^{-1})$, hence

$$\sqrt{n}\left[\bar{X}(\hat{\lambda}) - c_b\theta_0^{\hat{\lambda}}\right]/\sigma_0 = \sqrt{n}(\hat{\mu}-\mu_0)/\sigma_0 + c_b\theta_0^{\lambda_0}\log\theta_0\,\sqrt{n}(\hat{\lambda}-\lambda_0)/\sigma_0 + O_p(n^{-1/2}).$$

It follows that $\sqrt{n}\left[\bar{X}(\hat{\lambda})-c_b\theta_0^{\hat{\lambda}}\right]/\sigma_0$ is also asymptotically normal. Hence Lemma B and Slutsky's theorem can be applied to give the asymptotic normality of $T^*(\hat{\lambda})$ in Theorem 3.1. What is left is to find the asymptotic variance of $T^*(\hat{\lambda})$. A first-order Taylor expansion of $U(\hat{\lambda})=\sqrt{n}\left[\bar{X}(\hat{\lambda})-c_b\theta_0^{\hat{\lambda}}\right]$ around λ_0 gives

$$U(\hat{\lambda}) = U(\lambda_0) + \sqrt{n}(\hat{\lambda}-\lambda_0)[\,\bar{X}_{\lambda_0}(\lambda_0)-c_b\theta_0^{\lambda_0}\log\theta_0\,] + O_p(n^{-1/2}).$$

Lemma A and WLLN show that $\bar{X}_{\lambda_0}(\lambda_0) \xrightarrow{p} E(X_1^{\lambda_0}\log X_1)$, hence

$$U(\hat{\lambda}) = U(\lambda_0) + \sqrt{n}(\hat{\lambda}-\lambda_0)[\,E(X_1^{\lambda_0}\log X_1)-c_b\theta_0^{\lambda_0}\log\theta_0\,] + O_p(n^{-1/2}).$$

Using the results in Lemma A, some simple calculations give the asymptotic variance of $U(\hat{\lambda})$ and hence the asymptotic variance of $T^*(\hat{\lambda})$.

The asymptotic mean zero is obvious since c_b is defined as the ratio of the mean and median of X^{λ_0}.

References

Bain, L.J. and Engelhardt, M. (1991), *Statistical Analysis of Reliability and Life-Testing Models*, New York: Marcel Dekker.

Bilikan, J.E., Moore A.H. and Petrick, G.L. (1979), "*K* Sample ML Ratio Test for Change of Shape Parameter," *IEEE Transactions on Reliability*, **28**, 47-50.

Box, G.E.P. and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society - Series B*, **26**, 211-252.

Carroll, R. J. and Rupper, D. (1981), "On Prediction and the Power Transformation Family," *Biometrika*, **68**, 609-615.

——— (1991), "Prediction and Tolerance Intervals With Transformation and/or Weighing," *Technometrics*, **33**, 197-210.

Cheng, R.C.H. and Iles, T.C. (1990), "Embedded Models in Three Parameter Distributions and Their Estimation," *Journal of the Royal Statistical Society - Series B*, **52**, 135-149.

Chhikara, R. S. and Folks, J. L. (1989), *The Inverse Gaussian Distribution: Theory, Methodology and Applications*, New York: Marcel Dekker.

Collins, S. (1991), "Prediction Techniques for Box-Cox Regression Model," *Journal of Business and Economic Statistics*, **9**, 267-277.

Duan, N. (1993), "Smearing Estimate: A Nonparametric Retransformation Method *Journal of the American Statistical Association*, **78**, 605-610.

Hamilton, S. A. and Taylor, J. M. G. (1993), "A Comparison of the Box-Cox Transformation Method and Nonparametric Methods for Estimating Quantiles in Clinical Data with Repeated Measures," *Journal of the Statistical Computation and Simulation*, **45**, 185-201.

Hahn, G. J. and Meeker, W. Q. (1991), *Statistical Intervals: A Guide for Practitioners*, New York: Wiley.

Hernandez, F. and Johnson, R.A. (1980), "The Large-Sample Behavior of Transformations to Normality," *Journal of the American Statistical Association*, **75**, 855-861.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994), *Continuous Univariate Distributions*, New York: Wiley.

Kullback, S. (1968), *Information Theory and Statistics*, New York: Dover Publication.

Lawless, J. F. (1974), "On the Prediction of Survival Time for Individual Systems," *IEEE Transactions on Reliability*, **23**, 235-241

——— (1978), "Confidence Interval Estimation for the Weibull and Extreme Value Distributions," *Technometrics,* **20**, 355-368.

Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data,* John Wiley & Sons, Inc.

Nelson, W. (1982), *Applied Life Data Analysis,* New York: Wiley.

Sakia, R.M. (1992), "The Box-Cox Transformation Technique - a Review," *Statistician*, **41**, 169-178.

Taylor, J. M. G. (1985), "Measures of Location of Skewed Distributions Obtained Through Box-Cox Transformation," *Journal of the American Statistical Association*, **80**, 427-432.

Weibull, W. (1951), "A Statistical Distribution of Wide Applicability," *Journal of Applied Mechanics*, **18**, 293-297.

Yang, Z.L. (1999a). "Estimating a Transformation and its Effect on Box-Cox *T*-ratio," *Test*, **8**, 167-190.

——— (1999b). "Predicting a Future Lifetime Through Box-Cox Transformation," *Lifetime Data Analysis,* **5**, 265-279.