Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

Predicting a Future Median Life through a Power Transformation

ZHENLIN YANG[1]

Department of Statistics and Applied Probability, National University of Singapore, 3 Science Drive 2, Singapore 117543

Abstract. A simple and unified prediction interval (PI) for the median of a future lifetime can be obtained through a power transformation. This interval usually possesses the correct coverage, at least asymptotically, when the transformation is known. However, when the transformation is unknown and is estimated from the data, a correction is required. A simple correction factor is derived based on large sample theory. Simulation shows that the unified PI after correction performs well. When compared with the existing frequentist PI′s, it shows an equivalent or a better performance in terms of coverage probability and average length of the interval. Its nonparametric aspect and the ease of usage makes it very attractive to practitioners. Real data examples are provided for illustration.

Keywords: Box-Cox transformation, lifetime distributions, median lifetime, prediction interval.

[1]email: stayzl@nus.edu.sg; phone: (65) 874-6829; fax: (65) 872-3919

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

1 Introduction

Prediction concerning future lifetimes is an important topic in reliability and medical studies. Examples include predicting the safe life of a electronic device and the recurrence times of diseases. The future median life is often called the 'typical life' in practice. It is thus useful to consider the prediction of this median life to gain information on the lasting time of a future electronic component, remaining life of a certain patient, etc. There are many distributions that can be used to model lifetimes. The most popular ones include the exponential, Weibull, gamma, lognormal, and inverse Gaussian.

Inference methods used for lifetime data, such as prediction, still faces some practical problems. For example, i) for many lifetime distributions, an exact or approximate frequentist prediction interval (PI) for a future median life is not available; ii) in practice, it is often not clear which distribution that the observations have come from and hence model selection can be a problem for practitioners; iii) the existing frequentist PI may be unsatisfactory or may be too complicated to be implemented in practice; iv) standard normal-theory linear model inference methods can not be used for the lifetime analysis, etc. Hence it is highly desirable to have a unified PI for a future median life that works for any lifetime distribution.

The Box-Cox transformation (Box and Cox, 1964) aims to transform positive continuous data to near normality so that standard normal theory inference methods can be applied. Lifetimes are positive, continuous and the median of a lifetime distribution is transformable under a one-to-one transformation. Hence, if the observations can be transformed to normality, then the usual method gives a PI for the mean/median of a transformed future observation and a simple inverse transformation gives a PI for the median of the original future observation. Clearly, this interval is correct for any sample size when exact normality can be achieved and is

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

only asymptotically correct when a certain form of symmetry (mean = median) can be achieved by the transformation. When the transformation parameter is unknown, a common practice (Hahn and Meeker, 1991, p72) is to replace it by its estimator and treat the resultant interval as the correct one, i.e., ignore the variability from estimating the transformation. No doubt, if this approach works, it gives a simple and unified PI as no specification is necessary for the exact form of the lifetime distribution. However, as we shall argue in this paper, such extra variability can not be ignored and a correction of this interval needs to be made.

Yang (1999b) applied the Box-Cox transformation technique to give a simple and unified PI for a future lifetime and showed that this unified PI often meets or outperforms the existing frequentist PI′s. Similar performance can be expected for the unified PI for a future median life as the two problems are similar in nature. However, there are major differences between the two problems: the first concerns a random variable, where the second concerns a parameter. The latter is often called a confidence interval. We call it a prediction interval since we are stressing the "future" performance. Section 2 introduces the Box-Cox PI for a future median life obtained by a simple substitution of the transformation estimator. Section 3 presents some large sample theory concerning the performance of the Box-Cox PI. Through this theory, a simple correction factor is introduced. Section 4 presents some simulation results regarding the small sample performance of the adjusted Box-Cox PI, which is also called the unified PI in this article. Section 5 presents some real life examples to illustrate the applications of this unified PI. Section 6 is a general discussion on the results obtained and their implications to reliability and medical studies.

The unified PI is shown to possess asymptotically correct coverage. Monte Carlo simulations show that it also possesses good coverage properties even when the sample sizes are not large. The new interval is also compared with the existing frequentist PI′s through simulation and real data examples and is found to be superior or equivalent. The main attractivenesses of the new interval are its simplicity and unified/nonparametric nature, which make it very appealing to practitioners.

2 The Box-Cox Prediction Interval

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ be a sample of past observations from a lifetime distribution with probability density function (pdf) represented by $g(y; \theta)$, where θ is the parameter vector. Let Y_0 be a future observation from the same distribution. We are interested in constructing a unified PI for

$$\delta_0 = \text{Median}(Y_0)$$

based on the observed value of $\mathbf{Y}$. The Box-Cox transformation technique provides a solution to this problem. We now outline the Box-Cox method for constructing prediction intervals for the median of a future lifetime observation.

For a positive random variable Y_i, Box and Cox (1964) proposed a parametric family of power transformations:

$$h(Y_i, \lambda) = \begin{cases} \frac{Y_i^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \log Y_i, & \lambda = 0, \end{cases} \tag{2.1}$$

and assumed that there exists a λ such that $h(Y_i, \lambda)$ has an $N(\mu, \sigma^2)$ distribution for some μ and σ. This assumption leads to the maximum likelihood estimators (MLE) of (λ, μ, σ) as follows:

$$\begin{aligned} \hat{\mu}(\hat{\lambda}) &= \frac{1}{n} 1_n' h(\mathbf{Y}, \hat{\lambda}) \equiv \bar{h}(\mathbf{Y}, \hat{\lambda}), \\ \hat{\sigma}(\hat{\lambda}) &= \frac{1}{\sqrt{n}} \| h(\mathbf{Y}, \hat{\lambda}) - \bar{h}(\mathbf{Y}, \hat{\lambda}) 1_n \| \\ \hat{\lambda} &= \arg\min_{\ell} \dot{Y}^{-\ell} \| h(\mathbf{Y}, \ell) - \bar{h}(\mathbf{Y}, \ell) 1_n \|, \end{aligned} \tag{2.2}$$

where 1_n is a column vector of 1′s and $\dot{Y}$ is the geometric mean of the $(Y_1, \cdots, Y_n)$. Strictly speaking, $\hat{\mu}, \hat{\sigma}$, and $\hat{\lambda}$ are not the MLE′s unless the original observations

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

came from the lognormal distribution. Thus, they are often called the **Box-Cox estimators** (Yang 1999a, b). When λ is assumed known, the restricted MLE′s of μ and σ are, respectively, $\hat{\mu}(\lambda)$ and $\hat{\sigma}(\lambda)$.

Box and Cox suggested that, once an estimated transformation $\hat{\lambda}$ is obtained, one should use $\hat{\lambda}$ for the transformation and treat the transformed observations $h(Y_1, \hat{\lambda}), \cdots, h(Y_n, \hat{\lambda})$ as normal; hence, normal-theory inference methods can be applied. This method, called the **Box-Cox transformation method**, has proven to be very useful in statistical methodology development and application promotions. Yang (1999b) investigated properties of the PI for Y_0 itself obtained by this method and found that this interval outperforms or meets the corresponding frequentist PI′s for some popular lifetime distributions such as the gamma, Weibull, lognormal, inverse Gaussian and Birnbaum-Saunders.

In predicting $E_g[h(Y_0, \lambda)]$, the mean of the transformed future observation, the Box-Cox transformation method suggests that the following pivotal quantity

$$T(\lambda) = \frac{\hat{\mu}(\lambda) - E_g[h(Y_0, \lambda)]}{\hat{\sigma}(\lambda)/\sqrt{n-1}}$$

should have a t distribution with $n - 1$ degrees of freedom. A $100(1 - \alpha)\%$ PI for $E_g[h(Y_0, \lambda)]$ can be easily obtained by standard normal theory:

$$\left\{\hat{\mu}(\lambda) \pm t_{n-1}(\alpha/2)\hat{\sigma}(\lambda)/\sqrt{n-1}\right\} \tag{2.3}$$

Now, the power transformation is monotonic. If the transformed observations are symmetrically distributed, we then have

$$\begin{aligned} E_g[h(Y_0, \lambda)] &= \text{Median}[h(Y_0, \lambda)] \\ &= h[\text{Median}(Y_0), \lambda] \\ &= h(\delta_0, \lambda). \end{aligned}$$

Based on the above, a simple inverse transformation of the interval (2.3) gives a $100(1-\alpha)\%$ PI for δ_0:

$$\left\{1+\lambda\left[\hat{\mu}(\lambda)\pm t_{n-1}(\alpha/2)\hat{\sigma}(\lambda)/\sqrt{n-1}\right]\right\}^{1/\lambda}. \quad (2.4)$$

It is obvious that the interval (2.4) is correct for any sample size if exact normality can be achieved by the transformation and is asymptotically correct as long as the mean and median of the transformed observations are the same. This indicates that, unlike the case of predicting Y_0 itself, the exact normality may not be so important to the performance of the PI for δ_0, especially when n is not small.

Finally, when λ is unknown and is estimated by $\hat{\lambda}$, a common practice is to replace λ in (2.4) by $\hat{\lambda}$ and the resulted PI for δ_0 becomes,

$$\left\{1+\hat{\lambda}\left[\hat{\mu}(\hat{\lambda})\pm t_{n-1}(\alpha/2)\hat{\sigma}(\hat{\lambda})/\sqrt{n-1}\right]\right\}^{1/\hat{\lambda}}, \quad (2.5)$$

or $\exp[\hat{\mu}(0)\pm t_{n-1}(\alpha/2)\hat{\sigma}(0)/\sqrt{n-1}]$ when $\hat{\lambda}=0$. The interval given by (2.5) will be referred to in this article as the **Box-Cox prediction interval**.

Clearly, for the PI (2.5) to have correct coverage asymptotically, it is necessary that the pivotal quantity

$$T(\hat{\lambda})=\frac{\hat{\mu}(\hat{\lambda})-h(\delta_0,\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})/\sqrt{n-1}}$$

converges to standard normal. Unfortunately, this is not true as shown in Section 3, hence a correction is necessary.

3 Large Sample Behavior of the Box-Cox PI

The PI (2.5) is simple and easy to implement. It is unified in the sense that it works with any distribution having a domain of the positive half real line. However, its analytical properties remain unknown. In particular, its coverage properties and interval size are not known. We now provide some large sample results for the limiting behavior of this PI. Let τ^2 be the asymptotic variance of $\sqrt{n}(\hat{\lambda}-\lambda)$, h_λ and $h_{\lambda\lambda}$ be the first and second partial derivatives of $h(y,\lambda)$ with respect to λ, and $\bar{h}_\lambda$ and $\bar{h}_{\lambda\lambda}$ be the corresponding averages.

Theorem 3.1 *Suppose i)* $E[h(Y_0,\lambda)] = h(\delta_0,\lambda)$; *ii)* $\hat{\lambda} \to^p \lambda$ *and* $\sqrt{n}(\hat{\lambda}-\lambda)/\tau \to^d N(0,1)$; *iii)* $E_g[h_\lambda(Y_i,\lambda)]$ *and* $E_g[h(Y_i,\lambda)h_\lambda(Y_i,\lambda)]$ *both exist; and iv)* $\bar{h}_{\lambda\lambda}(\mathbf{Y},\lambda) = O_p(1)$. *Then,*

$$T(\hat{\lambda}) \to^d N(0, 1+c^2)$$

where $c = \tau\{E_g[h_\lambda(Y_0,\lambda)] - h_\lambda(\delta_0,\lambda)\}/\sigma$. *If, further, the first six moments of* $h(Y_i,\lambda)$ *are the same as those of* $N(\mu,\sigma^2)$ *and* $\lambda\sigma/(1+\lambda\mu)$ *is small, then* $c \approx 1/\sqrt{6}$.

The proof of Theorem 3.1 is given in the Appendix. The result of Theorem 3.1 is an important one. It demonstrates that the prediction interval for the future median obtained in the regular way does not have the correct coverage. This is true even when n is large, due to the estimation of the transformation. It quantifies the effect of estimating the transformation by a simple constant and suggests that the Box-Cox PI for δ_0 given in (2.5) should be corrected to be of the form

$$\left\{1+\hat{\lambda}\left[\hat{\mu}(\hat{\lambda}) \pm k t_{n-1}(\alpha/2)\hat{\sigma}(\hat{\lambda})/\sqrt{n-1}\right]\right\}^{1/\hat{\lambda}}, \tag{3.1}$$

which becomes $\exp[\hat{\mu}(0) \pm k t_{n-1}(\alpha/2)\hat{\sigma}(0)/\sqrt{n-1}]$ when $\hat{\lambda}=0$, where

$$k = \sqrt{1+c^2} \approx \sqrt{1+1/6}.$$

Thus, $k \approx 1.0801$, which is significantly larger than one, and this certainly can not be ignored for any sample size. For example, for $n = 20$, $t_{19}(0.05) = 1.7291$, but $kt_{19}(0.05) = 1.8676$, showing a significant difference. It is easily inferred that the Box-Cox PI is too short if no correction is made. To contrast with the Box-Cox PI given in (2.5), the interval (3.1) will be referred to as the **Unified Prediction**

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

Interval or adjusted Box-Cox prediction interval. The approximate value for c is obtained under the assumption that $\lambda\sigma/(1+\lambda\mu)$ is small. This is not unrealistic since the positiveness of Y_i implies the positiveness of $1+\lambda h(Y_i,\lambda)$, which in turn implies $\lambda\sigma \ll (1+\lambda\mu)$. See Yang (1999a) for a detailed discussion on this.

Note that with the Box-Cox power transformation the condition i) of Theorem 3.1 is true for lognormal and Birnbaum-Saunders distribution (Yang, 1999b). It may not be exactly true for the other distributions such as the Weibull and gamma, but this is not a problem of practical concern. Usually, the transformed observations are nearly symmetrically distributed as symmetry is one of the goals that the Box-Cox transformation intends to achieve. Also, estimation methods specifically designed for achieving symmetry have been discussed by, among others, Taylor(1985).

4 Small Sample Behavior of the Unified PI

Theorem 3.1 shows that the unified PI given by (3.1) has the correct coverage when n is large. When n is small, its behavior is investigated by Monte Carlo simulation. Random numbers $Y_1, Y_2, \cdots Y_n$ are first generated from one of the lifetime distributions. The estimate $\hat{\lambda}$ is then obtained and the original observations are transformed according to this estimated power transformation. Finally the PI is calculated. Various popular lifetime distributions such as the Weibull, gamma, lognormal, inverse Gaussian and Birnbaum-Saunders are considered. For each parameter configuration, 10,000 PI$'$s are generated and the number of PI$'$s covering δ_0 is divided by 10,000 to give a Monte Carlo estimate of the coverage probability. The average length of these 10,000 intervals is also recorded to give a measure of the expected length of the Box-Cox prediction interval.

Exact or approximate PI$'$s are available for the lognormal (Hahn and Meeker, 1991, p56), Weibull (Nelson, 1982, p232) and gamma (Lawless, 1982, p216) distributions. Also, a distribution-free method has been reported in the literature, see for example Hahn and Meeker (1991, p82). The PI for the gamma is based on the likelihood ratio test that is too complicated to be used in practice. We thus compare the Box-Cox PI only with the Weibull, lognormal and distribution-free PI′s. The distribution-free PI is conservative in the sense that it has a coverage probability larger than the nominal level. Hence, for comparison and for checking the exact coverage probabilities of the approximate PI′s (Weibull and distribution-free), we also record the results for these three intervals. There is another reason why we record the results for the "wrong" PI′s: we wish to see what happens when none of the PI′s listed is the correct one.

Lognormal random sample. This is the case that is of greatest interest as the assumptions of the theorem are completely satisfied and an exact frequentist PI exists. The difference between the unified and exact PI′s thus reflects the pure effect of estimating the transformation. The μ and σ in the table represent the mean and standard deviation of the logged observation. When σ increases, the population skewness increases quickly. The simulated coverage probability (C.P.) and average length (A.L.) of the PI′s are summarized in Table 4.1. The results show that the Unified PI performs very well. The C.P.s and A.L.s are all very close to those of the exact PI in all cases. Simulation was also carried out for $n = 30, 50, 100$ and 200 cases. The results (not reported) show that as n increases, the C.P. approaches the nominal level. This agrees with the theory. From Table 4.1, we also see that using the Weibull interval for lognormal data will result in a PI with a very poor coverage. The distribution-free PI is indeed very conservative, especially when n is very small, with its interval length being as much as several times longer than that of the unified PI. For example, when $n = 10$ and $\sigma = 2$, the 99% distribution-free PI has an A.L. of 130.64 with a C.P. of 0.9982, whereas the corresponding values for the unified PI

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

are 28.85 and 0.9907, respectively.

Table 4.1: Simulation Results for the Lognormal Distribution

		Unified PI		Lognormal PI		Weibull PI		Dist.-Free PI	
(μ, σ)	1-δ	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.
1.0, 0.5	.90	.8898	1.60	.9018	1.58	.7950	1.41	.9364	2.60
	.95	.9476	2.01	.9493	1.95	.8636	1.67	.9778	2.92
	.99	.9923	3.00	.9895	2.90	.9439	2.24	.9983	4.84
1.0, 1.0	.90	.8836	3.50	.8957	3.42	.7926	3.27	.9312	6.72
	.95	.9472	4.48	.9441	4.36	.8579	3.98	.9774	7.13
	.99	.9901	7.29	.9907	6.86	.9436	5.42	.9981	14.45
1.0, 1.5	.90	.8937	6.07	.8983	5.89	.7856	6.01	.9288	14.55
	.95	.9477	8.05	.9517	7.82	.8618	7.47	.9787	15.03
	.99	.9914	14.61	.9900	13.87	.9376	10.83	.9985	41.92
1.0, 2.0	.90	.8889	9.90	.9044	9.52	.7874	10.27	.9317	30.89
	.95	.9511	13.74	.9487	13.26	.8576	13.13	.9789	30.79
	.99	.9907	28.85	.9888	26.73	.9406	20.12	.9982	130.64
			$n = 10 \uparrow$		$n = 20 \downarrow$				
1.0, 0.5	.90	.8884	1.09	.9000	1.05	.7504	0.99	.9236	1.50
	.95	.9506	1.33	.9511	1.28	.8490	1.19	.9622	1.65
	.99	.9892	1.85	.9905	1.76	.9357	1.57	.9923	2.47
1.0, 1.0	.90	.8936	2.29	.9008	2.20	.7559	2.26	.9196	3.33
	.95	.9472	2.79	.9496	2.68	.8373	2.69	.9573	3.58
	.99	.9921	3.95	.9897	3.76	.9416	3.59	.9920	5.82
1.0, 1.5	.90	.8967	3.67	.9007	3.48	.7628	3.85	.9221	5.75
	.95	.9477	4.57	.9500	4.34	.8354	4.69	.9603	6.05
	.99	.9923	6.74	.9901	6.35	.9381	6.43	.9945	11.11
1.0, 2.0	.90	.9007	5.39	.8974	5.06	.7555	6.03	.9244	9.21
	.95	.9483	6.85	.9462	6.43	.8385	7.44	.9603	9.48
	.99	.9903	10.75	.9900	9.95	.9337	10.59	.9930	20.32

C.P.=coverage probability, A.L.=average length.

Weibull random sample. The Weibull distribution may be one of the most frequently used lifetime distributions in practice. However, a satisfactory and easy-to-use PI still does not seem to have been obtained. We use the one reported in Nelson (1982, p232) for comparison. The cumulative distribution function takes the form $F(y) = 1 - \exp[-(y/\beta)^{\nu}]$. The three different parameter configurations results in small to large population skewness. The simulation results summarized in Table 4.2 show that the unified PI clearly outperforms the Weibull PI in terms of both coverage probability and average length of the intervals. The Weibull PI can perform

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

Table 4.2: Simulation Results for the Weibull Distribution

		Unified PI		Weibull PI		Lognormal PI		Dist.-Free PI	
(ν, β)	1-α	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.
1.0, 5.0	.90	.8886	4.56	.8492	4.36	.8860	4.42	.9362	7.95
	.95	.9494	5.76	.9061	5.36	.9539	5.74	.9818	8.55
	.99	.9909	8.85	.9616	7.53	.9904	9.59	.9981	14.12
0.5, 5.0	.90	.8912	8.20	.8463	8.16	.8866	7.29	.9345	20.43
	.95	.9449	11.25	.9026	10.94	.9494	10.83	.9771	21.25
	.99	.9914	21.50	.9601	18.09	.9895	25.51	.9981	50.10
0.1, .01	.90	.8896	5.1	.8556	10.3	.8951	3.8	.9364	1422.
	.95	.9480	16.4	.9028	40.2	.9493	28.3	.9792	1402.
	.99	.9895	392.3	.9578	1200.6	.9898	18125.	.9981	845697.
			$n = 10 \uparrow$		$n = 20 \downarrow$				
1.0, 5.0	.90	.8864	3.09	.8655	3.04	.8469	2.85	.9197	4.48
	.95	.9446	3.77	.9272	3.66	.9238	3.50	.9583	4.85
	.99	.9905	5.26	.9740	4.97	.9865	5.04	.9925	7.40
0.5, 5.0	.90	.8936	4.82	.8710	4.97	.8475	3.91	.9175	8.29
	.95	.9458	6.06	.9264	6.21	.9239	5.05	.9590	8.63
	.99	.9900	9.23	.9749	9.20	.9868	8.29	.9927	16.23
0.1, 5.0	.90	.8928	77.6	.8724	99.8	.8479	25.2	.9163	1242.6
	.95	.9446	179.8	.9255	215.2	.9316	61.6	.9585	1339.6
	.99	.9903	944.5	.9748	1539.8	.9870	658.8	.9924	27398.6

rather poorly when the sample size is small and when the data is very skewed. For example, when $n = 10$ and $\nu = 0.1$, the 99% Weibull PI has C.P. and A.L. 0.9578 and 1200, respectively, compared with 0.9895 and 392.3 for the unified PI. The lognormal PI performs in a rather unstable way as it is a wrong PI for the Weibull data. It performs well when the logarithm of the Weibull is close to normal, but otherwise poorly. Again, the distribution-free PI can give a very conservative and long interval. Notice that for the Weibull density, the symmetry condition is not satisfied with the power transformation, which may give rise to extra effects, but the simulation results do not show any significant effect due to this lack of symmetry.

Inverse Gaussian random sample. The emphasis here and in the following gamma case is to address the following questions: what will happen when none of the PI's is the right one and what is the performance of each of them? The pdf is of the form $[\beta/(2\pi y^3)] \exp[-\beta(y-\nu)^2/(2\nu^2 y)]$. The simulation results show that the unified

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

Table 4.3: Simulation Results for the Inverse Gaussian Distribution

		Unified PI		Weibull PI		Lognormal PI		Dist.-Free PI	
(ν, β)	1-α	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.
1.0, 4.0	.90	.8919	0.50	.7855	0.44	.8978	0.49	.9349	0.81
	.95	.9477	0.62	.8598	0.52	.9482	0.61	.9755	0.91
	.99	.9905	0.93	.9407	0.69	.9907	0.89	.9970	1.48
0.5, 1.0	.90	.8843	0.31	.7842	0.28	.8998	0.31	.9351	0.55
	.95	.9418	0.39	.8568	0.34	.9455	0.38	.9767	0.60
	.99	.9905	0.61	.9361	0.45	.9878	0.58	.9976	1.01
1.0, 1.0	.90	.8765	0.73	.7641	0.68	.8936	0.72	.9360	1.40
	.95	.9422	0.93	.8444	0.82	.9472	0.91	.9778	1.49
	.99	.9881	1.48	.9358	1.09	.9891	1.39	.9978	2.70
4.0, 1.0	.90	.8770	2.85	.7257	2.94	.8889	2.95	.9350	7.42
	.95	.9416	3.56	.8158	3.57	.9444	3.79	.9798	7.60
	.99	.9900	6.81	.9291	4.94	.9909	6.28	.9977	17.99
			$n = 10 \uparrow$		$n = 20 \downarrow$				
1.0, 4.0	.90	.8886	0.34	.7457	0.31	.9009	0.33	.9243	0.47
	.95	.9478	0.42	.8316	0.37	.9479	0.39	.9555	0.52
	.99	.9907	0.57	.9343	0.49	.9904	0.55	.9934	0.78
0.5, 1.0	.90	.8815	0.21	.7347	0.20	.9014	0.20	.9203	0.30
	.95	.9471	0.26	.8270	0.24	.9507	0.25	.9561	0.33
	.99	.9892	0.36	.9301	0.31	.9911	0.34	.9930	0.50
1.0, 1.0	.90	.8904	0.48	.7141	0.46	.8976	0.46	.9216	0.72
	.95	.9409	0.59	.8031	0.56	.9469	0.57	.9563	0.77
	.99	.9884	0.83	.9244	0.74	.9909	0.79	.9930	1.25
4.0, 1.0	.90	.8772	1.67	.6191	1.88	.8752	1.74	.9191	2.87
	.95	.9374	2.09	.7378	2.25	.9390	2.14	.9602	2.97
	.99	.9866	3.08	.8886	3.08	.9877	3.11	.9919	5.67

PI still performs very well. The Weibull PI performs rather poorly. It is interesting to note that the lognormal PI performs very well, and it is superior to the unified PI. This is due to the fact that the best normalizing transformation for the inverse Gaussian observations is often close to the log transformation (Yang, 1999b). The distribution-free PI performs as usual: conservatively with a long interval length. As in the Weibull case, this case also does not satisfy the symmetry condition with the power transformation, but the mean and median are very close if the population is not very skewed. See the last section for illustrations of this.

Gamma random sample. We finally use the gamma distribution to check the performance of the four PI's when none of them are exactly correct. Here β is

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

Table 4.4: Simulation Results for the Gamma Distribution

		Unified PI		Weibull PI		Lognormal PI		Dist.-Free PI	
(ν, β)	1-α	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.	C.P.	A.L.
4.0, 1.0	.90	.8946	2.17	.8197	1.91	.8956	2.13	.9328	3.3695
	.95	.9469	2.69	.8827	2.28	.9494	2.65	.9769	3.8397
	.99	.9910	3.98	.9517	3.06	.9910	3.94	.9983	6.0408
2.0, 1.0	.90	.8906	1.46	.8337	1.33	.8924	1.43	.9322	2.3698
	.95	.9466	1.82	.8953	1.59	.9507	1.79	.9811	2.6213
	.99	.9899	2.72	.9531	2.16	.9895	2.75	.9974	4.1636
1.0, 1.0	.90	.8912	0.91	.8482	0.87	.8893	0.88	.9354	1.5844
	.95	.9447	1.15	.8995	1.08	.9469	1.16	.9793	1.7290
	.99	.9895	1.78	.9591	1.51	.9915	1.93	.9983	2.8282
0.5, 1.0	.90	.8819	0.50	.8572	0.52	.8815	0.48	.9320	1.0086
	.95	.9442	0.66	.9107	0.66	.9471	0.67	.9790	1.0519
	.99	.9906	1.10	.9616	1.05	.9918	1.43	.9986	1.8889
			$n = 10 \uparrow$		$n = 20 \downarrow$				
4.0, 1.0	.90	.8844	1.50	.8349	1.36	.8811	1.43	.9187	2.0241
	.95	.9464	1.83	.8908	1.63	.9444	1.74	.9584	2.2439
	.99	.9907	2.51	.9618	2.16	.9895	2.40	.9912	3.2829
2.0, 1.0	.90	.8922	1.00	.8464	0.94	.8732	0.94	.9159	1.3845
	.95	.9460	1.22	.9080	1.13	.9360	1.15	.9565	1.5155
	.99	.9915	1.68	.9686	1.50	.9891	1.61	.9925	2.2613
1.0, 1.0	.90	.8872	0.62	.8737	0.61	.8553	0.57	.9212	0.8946
	.95	.9448	0.76	.9262	0.73	.9266	0.70	.9592	0.9687
	.99	.9915	1.06	.9740	1.00	.9886	1.01	.9929	1.4861
0.5, 1.0	.90	.8762	0.32	.8875	0.38	.8041	0.27	.9208	0.5206
	.95	.9386	0.40	.9328	0.41	.9093	0.35	.9593	0.5479
	.99	.9892	0.57	.9798	0.60	.9850	0.55	.9920	0.8987

the scale parameter and ν is the shape parameter. The results in Table 4.4 show that the unified PI is quite robust against changes of distributional shape. The Weibull PI again performs poorly although better than in the case of the inverse Gaussian random samples. The lognormal PI performs well when the data is not so skewed, but deteriorates quite significantly when data is very skewed. For example, when $n = 20$ and $\nu = 0.5$, the 90% lognormal PI has a coverage of only 80.41%.

Note that for the Weibull, inverse Gaussian and gamma distributions, simulation is also carried out for the cases of sample sizes 30 and 50. All the simulation results (not reported) are similar to the case of $n = 20$. The coverage probability does not seem to improve much for the unified PI as the sample size increases. This reflects the effect of asymmetry after transformation. This problem can be resolved by either introducing a bias correction factor or by working with a different transformation family which gives symmetric property.

5 Numerical Examples

We now present three real life examples to illustrate the Box-Cox PI and to compare it with the frequentist and distribution-free PI′s. All the three data sets have been used extensively for lifetime data analysis.

Example 5.1: Insulating Fluid Data. Nelson (1982, p228) reported data, representing the times to breakdown of an insulating fluid in an accelerated test at different test voltages for illustrating the statistical intervals for the Weibull. We take the data corresponding to voltage of 32kV: 0.270 0.400 0.690 0.790 2.750 3.910 9.880 13.949 15.930 27.799 53.239 82.847 89.282 100.575 215.099. This data is small ($n = 15$) and is moderately skewed to the right as the sample skewness $\hat{\gamma} = 1.7452$.

Example 5.2: Ball Bearing Data. The following data set, the endurance of deep groove ball bearings, is probably one of the most frequently used data set in literature for illustrating the applications of lifetime distributions. It can be found from Chhikara and Folks (1989, p73): 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40. This data is of medium sized ($n = 23$) and is slightly skewed to the right as the sample skewness $\hat{\gamma} = 0.9206$.

Example 5.3: Repair Time Data. Repair times (in hours) for an airborne communication transceiver were given in Chhikara and Folks (1989, p139) where a .5 was missed from the original paper: .2, .3, .5, .5, .5, .5, .6, .6, .7, .7, .7, .8, .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. This data set is fairly large

Table 5.1: Prediction Intervals based on Real Data Sets

	Corrected PL		Weibull PL		Lognormal PL		Dist.-Free PL	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Insulating Fluid Data: $n = 15$; $\hat{\lambda} = 0.0827$; $\hat{\gamma} = 1.7452$								
.90	3.8468	29.6616	5.3233	33.3543	3.6516	23.6134	0.7900	53.2390
.95	3.1014	35.4740	4.3593	40.7297	3.0530	28.2438	0.7900	82.8470
.99	2.0140	49.9302	2.8260	62.8286	2.1525	40.0595	0.6900	89.2820
Ball Bearing Data: $n = 23$; $\hat{\lambda} = 0.1905$; $\hat{\gamma} = 0.9206$								
.90	53.1100	79.2252	58.1365	82.5482	52.8554	76.1991	51.8400	84.1200
.95	51.0267	82.1958	56.0574	85.6097	51.0329	78.9203	48.4800	84.1200
.99	47.1511	88.2536	51.9527	92.3736	47.6548	84.5146	45.6000	98.6400
Repair Time Data: $n = 46$; $\hat{\lambda} = -0.1014$; $\hat{\gamma} = 2.8568$								
.90	1.3641	2.4424	1.8008	2.9880	1.4745	2.5307	1.0000	3.0000
.95	1.2923	2.5875	1.7122	3.1426	1.4000	2.6653	1.0000	3.3000
.99	1.1638	2.8990	1.5464	3.4795	1.2653	2.9490	1.0000	4.0000

($n = 42$) and is very skewed as the sample skewness $\hat{\gamma} = 2.8568$.

The calculated PI′s are summarized in Table 5.1. It is seen from the table that the unified PI is always one of the shorter ones and the distribution-free PI is always the longest. For the Insulating Fluid Data, the lognormal PI is the shortest one, but not much different from the unified PI. This is reasonable as the estimated transformation is very close to the log transformation. For the Ball Bearing Data, all the intervals are very close as the data is not that skewed. The repair time data is the most skewed of the three. In this case, he unified PI and the lognormal PI are almost the same. The Weibull PI can be 11-13% longer than the unified PI.

6 Discussions

A unified prediction interval for a future median life is obtained. It is shown that this interval possesses good large and small sample properties. It is simple and easy to implement and is robust against changes of distributions. This interval can be easily applied in reliability and life sciences studies to gain insights on the safe life

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

of an electronic component and the survival time of a patient. For estimating the median of the original observation though Box-Cox transformation, Carroll and Ruppert (1981) examined (in a general linear model framework) the effect of estimating the transformation in terms of the mean squared errors and found that this effect is generally not large. Our result is consistent with theirs for the special one sample model.

One may naturally think that the results in Theorem 3.1 should be directly generalizable to the general linear model, but unfortunately, unlike the case of predicting the individual Y_0, the results of Theorem 3.1 can not be generalized to the linear model case, especially the simple correction factor can not be easily obtained. Nevertheless, there is a general agreement that, no matter for a one sample model or a general linear model, the prediction interval for the median obtained by simply substituting the transformation estimator into the interval needs to be corrected.

It may be interesting to see what happens to the Box-Cox PI if no correction is applied. Our simulation results (not reported) show that this effect can not be ignored, no matter if the sample size is large or small. For example, for the lognormal case the simulated C.P. for 90% level, $n = 20$, can be as low as 0.8619 before correction, compared with the lowest 0.8884 (Table 4.1) after correction. It may also be interesting to see how much the mean and median can differ for the cases of the Weibull, inverse Gaussian and gamma distributions. This can be easily accomplished using the results of Yang (1999b, Table 2.1). We pick two inverse Gaussian cases to illustrate this point. For $(\nu, \beta) = (1, 1)$, the mean $= -0.3942$ and median $= -0.3976$ with a population skewness of 3; for $(\nu, \beta) = (4, 1)$, the mean $= 0.3492$ and median $= 0.3270$ with a population skewness of 6. This shows that the mean and median of a transformed inverse Gaussian observation will not differ much unless the population is very skewed. Similar calculations (not reported) show that the same conclusion

Appeared in: Lifetime Data Analysis, 2001, 7, 305-317.

applies to other lifetime distributions as well.

Appendix: Proof of the Theorem 3.1

We give a sketch of the proof. Detail is available from the author. By Taylor expansions, we have

$$\bar{h}(\mathbf{Y}, \hat{\lambda}) = \bar{h}(\mathbf{Y}, \lambda) + (\hat{\lambda} - \lambda)\bar{h}_\lambda(\mathbf{Y}, \lambda) + \frac{1}{2}(\hat{\lambda} - \lambda)^2[\bar{h}_{\lambda\lambda}(\mathbf{Y}, \lambda) - R_n],$$

$$h(\delta_0, \hat{\lambda}) = h(\delta_0, \lambda) + (\hat{\lambda} - \lambda)h_\lambda(\delta_0, \lambda) + \frac{1}{2}(\hat{\lambda} - \lambda)^2[h_{\lambda\lambda}(\delta_0, \lambda) - R'_n],$$

where R_n and R'_n converge to zero as $\hat{\lambda} \to \lambda$. Taking the difference, we have

$$\bar{h}(\mathbf{Y}, \hat{\lambda}) - h(\delta_0, \hat{\lambda}) = \bar{h}(\mathbf{Y}, \lambda) - h(\delta_0, \lambda) + (\hat{\lambda} - \lambda)[\bar{h}_\lambda(\mathbf{Y}, \lambda) - h_\lambda(\delta_0, \lambda)] + O_p(n^{-1}),$$

where the second term is of the same order as the first since $\bar{h}_\lambda(\mathbf{Y}, \lambda) \to^p E_g[h_\lambda(Y_0, \lambda)]$, which is different from $h_\lambda(\delta_0, \lambda)$ in general. Now, it is easy to see that

$$\hat{\sigma}^{-1}(\hat{\lambda}) = \hat{\sigma}^{-1}(\lambda) + O_p(n^{-\frac{1}{2}}) = \sigma^{-1} + O_p(n^{-\frac{1}{2}})$$

Hence, $T(\hat{\lambda})$ can be written as

$$T(\hat{\lambda}) = T(\lambda) + \sqrt{n}(\hat{\lambda} - \lambda)\{E_g[h_\lambda(Y_0, \lambda)] - h_\lambda(\delta_0, \lambda)\}/\sigma.$$

Since $T(\lambda) \to^d N(0, 1)$ and $\sqrt{n}(\hat{\lambda} - \lambda)/\tau \to^d N(0, 1)$, the result follows by showing that $T(\lambda)$ and $\hat{\lambda}$ are asymptotically independent. This can be inferred from a result of Yang (1996).

Finally, for the power transformation, one can easily show that $h_\lambda(Y_0, \lambda) = \lambda^{-1}[Y_0^\lambda \log Y_0 - h(Y_0, \lambda)]$ which converges to $\frac{1}{2}\log^2 Y_0$ as $\lambda \to 0$. Using the relation $\log Y_0 = \lambda^{-1}\log[1 + \lambda h(Y_0, \lambda)]$ and the approximation

$$\lambda \log Y \approx \log(1 + \lambda\mu) + \theta e_0 - \frac{1}{2}\theta^2 e_0^2$$

where $\theta = \lambda\sigma/(1+\lambda\mu)$ and $e_0 = h(Y_0, \lambda) - \mu$, some algebraic work leads to $c \approx \tau\theta/2\lambda$ which becomes $1/\sqrt{6}$ as $\tau \approx (\lambda/\theta)\sqrt{2/3}$ as shown by Yang (1999a, pp. 175).

Acknowledgments

The author wishes to thank the coordinating editor and the two referees for the helpful comments. A careful reading of the manuscript by my colleague, Dr. Paul Kenneth Marriott, is greatly appreciated.

References

G.E.P. Box and D.R. Cox, "An analysis of transformation (with discussion)," *J. R. Statist. Soc. B* vol. 26, pp. 211-252, 1964.

R.J. Carroll and D. Ruppert, "On prediction and the power transformation family," *Biometrika* vol.68, pp. 609-615, 1981.

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

G.J. Hahn and W.Q. Meeker, *Statistical Intervals, A Guide for Practitioners*, John Wiley & Sons, Inc., 1991.

J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc., 1982.

W. Nelson, *Applied Life Data Analysis*, John Wiley & Sons, Inc., 1982.

J.M.G. Taylor, "Power transformation to symmetry," *Biometrika* vol. 72, pp. 145-152, 1985.

Z. Yang, "Some asymptotic results for the Box-Cox transformation methodology," *Commun. Statist.-Theory Meth.* vol. 25, pp. 403-414, 1996.

Z. Yang, "Estimating a transformation and its effect on Box-Cox T-ratio," *Test* vol. 8, pp. 167-190, 1999a.

Z. Yang, "Predicting a future life time through Box-Cox transformation," *Lifetime Data Analysis* vol. 5, pp. 265-279, 1999b.