

11-2014

Double Bootstrap Confidence Interval Estimates with Censored and Truncated Data

Jayanthi Arasan

University Putra Malaysia, Serdang, Selangor, Malaysia, jayanthi@upm.edu.my

Mohd B. Adam

University Putra Malaysia, Serdang, Selangor, Malaysia, bakri@upm.edu.my

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Arasan, Jayanthi and Adam, Mohd B. (2014) "Double Bootstrap Confidence Interval Estimates with Censored and Truncated Data," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 2 , Article 22.
DOI: 10.22237/jmasm/1414815660

Double Bootstrap Confidence Interval Estimates with Censored and Truncated Data

Jayanthi Arasan
University Putra Malaysia
Serdang, Selangor, Malaysia

Mohd Bakri Adam
University Putra Malaysia
Serdang, Selangor, Malaysia

Traditional inferential procedures often fail with censored and truncated data, especially when sample sizes are small. In this paper we evaluate the performances of the double and single bootstrap interval estimates by comparing the double percentile (DB-p), double percentile-t (DB-t), single percentile (B-p), and percentile-t (B-t) bootstrap interval estimation methods via a coverage probability study when the data is censored using the log logistic model. We then apply the double bootstrap intervals to real right censored lifetime data on 32 women with breast cancer and failure data on 98 brake pads where all the observations were left truncated.

Keywords: Double bootstrap, censored, simulation, truncated, survival

Introduction

Modeling and data analysis is never complete without reliable statistical inferential procedures such as the confidence intervals or hypothesis testing. These are powerful tools that help us make certain conclusions regarding the population and its parameters based on sample data. The confidence interval can also be used to indicate the reliability of our estimates. However, it is not easy to obtain the exact solutions for some of these inferential procedures especially in cases involving more complex data structures such as incomplete, censored or truncated data. Thus, many have resorted to the much simplified techniques based on the asymptotic normality of the maximum likelihood estimates. In recent years, the bootstrapping techniques have taken over some of these existing methods because they relieve us from certain assumptions based on the asymptotic statistical theory. Thus, the

*Dr. Arasan is a lecturer in the Computational Statistics and Operations Research Laboratory, Institute for Mathematical Research. Email her at jayanthi@upm.edu.my.
Dr. Adam is a lecturer in the Computational Statistics and Operations Research Laboratory, Institute for Mathematical Research. Email him at bakri@upm.edu.my.*

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

bootstrap is widely preferred when sample sizes are low or data is censored or truncated since the Wald is known to be highly asymmetrical in these cases with actual coverage probability significantly different from the nominal specification as discussed by Jeng and Meeker (2000), Doganaksoy and Schmee (1993a; 1993b) and Arasan and Lunn (2009).

The bootstrap method as described by Efron and Tibshirani (1993) is a direct application of the plug-in principle which is a way of understanding the population based on estimates from random samples drawn from the population. The standard bootstrap confidence interval is constructed from information obtained directly from the data without any theoretical assumptions. In recent years the double bootstrap has gained a lot of popularity because it typically has a higher order of accuracy. According to McCullough and Vinod (1998), and Letson and McCullough (1998), the double bootstrap enhances the accuracy of the ordinary bootstrap by estimating an error and then using this error to adjust the ordinary bootstrap in order to reduce its error.

Efron (1993) introduced the bootstrap percentile (B-p), percentile-t (B-t), and the bias-corrected percentile (BCa) intervals, see also Efron (1981b; 1981a). Efron (1985) claims that the bootstrap confidence interval reduces most of the error in standard approximation. He also describes some of these intervals as invariant under transformation, thus producing accurate results without involving knowledge of the normalizing transformations. An extensive survey of different bootstrap methods for producing good confidence interval estimates is given in DiCiccio and Efron (1996). Singh (1981) established the second order accuracy of the bootstrap confidence interval by applying Edgeworth theory to the B-t interval. Hall (1986; 1988a; 1988b) examined several different bootstrap interval estimation methods that can be used in both parametric and nonparametric settings and concluded that B-t and BCa methods were superior to other methods. More applied works on bootstrap confidence intervals were done by Arasan and Lunn (2008), Robinson (1983), Schenker (1985), and Jeng and Meeker (2000).

The Model

The log-logistic distribution is very popular in survival studies because it has a hazard rate that increases in the beginning and slowly starts to decrease after a finite time. These types of non-monotonic hazard rates are very popular in medical studies especially those involving lung cancer, breast cancer and kidney or heart transplant patients. This distribution has been studied by various authors such as Bennet (1983) who explored and provided the linear model for the log odds on

survival. Gupta et al. (1999) proved analytically that unique maximum likelihood estimates exist for the parameters of this model and analyzed a lung cancer data. Mazucheli et al. (2005) compared the accuracy of Wald confidence interval with the B-p and B-t intervals for the mode of the hazard function of the log logistic distribution. Other authors who have done significant work using this model are [Cox and Lewis (1966)], [Cox, Oakes, O'Quigley and Struthers (1982)]. The model can also easily be extended to accommodate covariates, truncated data and all types of censored observations such as left, right and interval. More discussions on truncated data can be found in Lawless (1982).

Lifetime data are sometimes truncated due to some of the conditions in the study design. When the lifetime t_i for the i^{th} subject is forced to lie between the interval $[u_i, v_i]$, where u_i and v_i are left and right truncation times respectively, then t_i is said to be either left or right truncated. Subjects who do not experience the event within this window will not be included in the study. So, subjects are left truncated only if they were already at risk before entering the study. So the current lifetime of subject i at selection is $t_i \geq u_i$ where $u_i > 0$ and. Similarly right truncated data are data where $t_i \leq v_i$. Thus, right truncation occurs when all the subjects have already experienced the event of interest when they enter the study. On the other hand right censoring occurs when a subject's event time is unknown due to reasons such as study has ended or subject has left the study. Left censoring occurs when the event of interest has already occurred before the study started. In both cases subjects are still included in the study but it is acknowledged that their event time is above or below a certain point.

There are several different parameterizations for the log logistic distribution. If $-\infty < \delta < \infty$ is the scale parameter and $\beta > 0$ is the shape parameter the density (pdf) and survivor function of the log logistic are

$$f(t, \delta, \beta) = \frac{e^{\delta} \beta t^{\beta-1}}{[1 + e^{\delta} t^{\beta}]^2}$$

$$S(t, \delta, \beta) = \frac{1}{1 + e^{\delta} t^{\beta}}.$$

Suppose we have both right censored and uncensored lifetimes for $i = 1, 2, \dots, n$ observations. Let c_i be the indicator variable assuming the value of 1 if data is uncensored or 0 otherwise. The log-likelihood function for the full sample is given by

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

$$\begin{aligned}
 l(\delta, \beta) &= \ln \left\{ \prod_{i=1}^n [f(t_i; \delta; \beta)]^{c_i} [S(t_i; \delta; \beta)]^{(1-c_i)} \right\} \\
 &= \sum_{i=1}^n c_i \left((\beta - 1) \ln(t_i) + \ln(\beta) + \delta - 2 \ln(1 + t_i^\beta e^\delta) \right) \\
 &\quad - (1 - c_i) \ln(1 + t_i^\beta e^\delta)
 \end{aligned} \tag{1}$$

Suppose there is left truncated data for $i = 1, 2, \dots, n$ observations. Let u_i be the left truncation time for the i^{th} subject. The log-likelihood function for the full sample is given by

$$\begin{aligned}
 l(\delta, \beta) &= \ln \left\{ \prod_{i=1}^n \frac{f(t_i; \delta; \beta)}{S(u_i; \delta; \beta)} \right\} \\
 &= \sum_{i=1}^n \left((\beta - 1) \ln(t_i) + \ln(\beta) + \delta - 2 \ln(1 + t_i^\beta e^\delta) + \ln(1 + u_i^\beta e^\delta) \right)
 \end{aligned} \tag{2}$$

Confidence interval estimates

A bootstrap procedure depends on how the bootstrap sampling is done, namely parametric or nonparametric sampling procedures. In the parametric bootstrap sampling procedure, B bootstrap samples each of size n are generated from the assumed parametric distribution. Then, the bootstrap estimates, $\hat{\theta}_b^*$, $b = 1, 2, \dots, B$ are estimates calculated from each of these bootstrap samples of size n . The nonparametric procedure requires the sampling of a large number of B bootstrap samples with replacement from the original data set with each observation having equal probability of being chosen. This technique of resampling clearly requires the assumption that the data are independent. Following that, the bootstrap estimates are calculated in the same way as described before. In this research we employ the nonparametric bootstrap sampling procedure since we wish to incorporate censored and truncated observations in our dataset.

Single bootstrap without pivot (Percentile interval or B-p)

A clear and thorough understanding of the single bootstrap interval estimation procedures is essential before moving on to any of the double bootstrap methods. If θ is our parameter of interest and $\hat{\theta}$ its estimate using sample data, then $\hat{\theta}_b^*$ is the estimate of θ using the b^{th} bootstrap sample. The B-p method is rather simple and

constructs confidence intervals directly from the percentiles of the bootstrap distribution of the estimated parameters. In this procedure B bootstrap samples, y^b , $b = 1, 2, \dots, B$ will be generated using the nonparametric bootstrap sampling method. Following that, estimate the bootstrap version of the parameter estimates, $\hat{\theta}_b^*$ for each of the bootstrap sample, y^b . The $100(1-\alpha)\%$ percentile interval for θ is $\{\hat{\theta}_{[l]}^*, \hat{\theta}_{[u]}^*\}$, where, $l = B \times \frac{\alpha}{2}$, $u = B \times \left(1 - \frac{\alpha}{2}\right)$ and $\hat{\theta}_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ is the ordered list of the B values of $\hat{\theta}^*$. For ease of computation and accuracy large values of B that give integer values of l and u should preferably be chosen.

The B-p method is said to be transformation-respecting and has the ability to automatically produce accurate results without any normalizing transformations as described in Efron and Tibshirani (1993). Thus it becomes especially useful when the distribution of $\hat{\theta}$ is not approximately normally distributed, since in this case the Wald interval would not perform well unless an appropriate transformation is used.

Single bootstrap with pivot (Percentile-t interval or B-t)

The B-t method involves a bit more work than the B-p interval since it requires the standard error of an estimate. In this method, a bootstrap table consisting the percentiles of the bootstrap version of the standardized values of the parameter estimates (approximate pivot) is constructed using the available data. The property of the approximate pivot whose distribution is approximately the same for all parameter allows the formation of this bootstrap distribution. Following that, this bootstrap table is used to construct the B-t confidence intervals. The main highlight of this method is that it is only dependent on the data in hand and does not require any normal theory assumptions. However, depending on the data available, the bootstrap distribution produced (B-t percentiles) can be asymmetric about 0, which may produce more asymmetrical intervals although at a much better coverage probability.

In this procedure, compute $\hat{\theta}_b^*$ for $b = 1, 2, \dots, B$ bootstrap samples and obtain

$$R_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{se(\hat{\theta}_b^*)}. \tag{3}$$

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Here, $se(\hat{\theta}_b^*)$ is the estimated standard error of $\hat{\theta}_b^*$ for the bootstrap sample, y^b . In B-t confidence interval the $\frac{\alpha}{2}$ quantile are based on B-t percentiles which can be obtained using the data. In order to do this obtain $R_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ which is the ordered list of the B values of R^* . The $\frac{\alpha}{2}$ percentile of R^* is then the value $R_{[B-\frac{\alpha}{2}]}^*$. Then, the $100(1-\alpha)\%$ confidence interval for θ is $\{\hat{\theta} - R_{[l]}^*.se(\hat{\theta}), \hat{\theta} - R_{[u]}^*.se(\hat{\theta})\}$ where, $l = B \times (1 - \frac{\alpha}{2})$ and $u = B \times \frac{\alpha}{2}$.

Double bootstrap procedures

The double bootstrap procedure requires resampling from bootstrap samples in order to further reduce the bias and correct the errors in the bootstrap procedures, see Martin (1992). Similarly a double bootstrap confidence interval procedure is a further iteration to the ordinary bootstrap confidence interval procedure which would further reduce the order of magnitude of coverage error. Both B-p and B-t, has under mild regularity conditions, a 2 sided coverage error equals $O(n^{-1})$, at nominal level α . It follows that a further iteration of the ordinary bootstrap confidence interval would further reduce the order of magnitude of coverage error to $O(n^{-2})$.

Double bootstrap without pivot (Double B-p)

The double bootstrap without a pivot or double B-p is given by Shi (1992) and also discussed by Letson and Mccullough (1998) and is constructed as follows. First, draw B single bootstrap samples, denoted $y^1, y^2, \dots, y^b, b = 1, 2, \dots, B$. Then, for each b draw another $c = 1, 2, \dots, C$ bootstrap resamples. Following that calculate $\hat{\theta}_{bc}^{**}$ for each double bootstrap samples. In the next step we have to calculate the number of $\hat{\theta}_{bc}^{**}$ that is lesser or equal to $\hat{\theta}$ for each c and divide this number by C

$$Q_b = \frac{\#(\hat{\theta}_{bc}^{**} \leq \hat{\theta})}{C} \quad (4)$$

Following that if the ordered values of Q_b are $Q_{[1]}, Q_{[2]}, \dots, Q_{[B]}$ then the $(1-\alpha)\%$ double percentile bootstrap confidence interval for θ is $\{\hat{\theta}_{[l]}^*, \hat{\theta}_{[u]}^*\}$ where

$l = B.Q_{\left[\frac{\alpha}{2} \times B\right]}$, $u = B.Q_{\left[\left(1-\frac{\alpha}{2}\right) \times B\right]}$ and $\hat{\theta}_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ is the ordered list of the B values of $\hat{\theta}^*$. In cases where l and u are not integers, they should be rounded to the nearest integer lesser than or equal to their values.

Double bootstrap with pivot (Double B-t)

The double bootstrap with pivot or double B-t was discussed Mccullough and Vinod (1998) and also Letson and Mccullough (1998). In order to construct the double B-t confidence interval for the parameter, θ , recall that for the B-t interval we need to compute R_b^* as given by (3). Following that we now have to resample C double bootstrap samples from each of the single bootstrap samples and obtain

$$R_{bc}^{**} = \frac{\hat{\theta}_{bc}^{**} - \hat{\theta}_b^*}{se(\hat{\theta}_{bc}^{**})}.$$

In the next step we have to calculate the number of times the second stage root R_{bc}^{**} is lesser or equal to the first stage root R_b^* for each c and divide this number by C

$$Z_b = \frac{\#(R_{bc}^{**} \leq R_b^*)}{C}. \tag{5}$$

Then, if the ordered values of Z_b are $Z_{[1]}, Z_{[2]}, \dots, Z_{[B]}$, the $(1-\alpha)\%$ double B-t confidence interval for θ is $\left\{ \hat{\theta} - R_{[l]}^* . se(\hat{\theta}), \hat{\theta} - R_{[u]}^* . se(\hat{\theta}) \right\}$ where $l = B.Z_{\left[\left(1-\frac{\alpha}{2}\right) \times B\right]}$,

$u = B.Z_{\left[\frac{\alpha}{2} \times B\right]}$ and $R_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ is the ordered list of the B values of R^* .

Simulation study

A simulation study was conducted using $N = 1000$ samples of size $n = 25, 30, 40$ and 50 to compare the performance of the confidence interval estimates discussed in the previous section for the parameters of the log logistic model with censored data. We used $\alpha = 0.05$ and $\alpha = 0.10$ where α is the nominal error probability and censoring proportion (cp) of 10% and 15%. We compared a total of 4 methods

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

namely the B-p, B-t, DB-p and DB-t. Following that, we calculated the estimated coverage probability error by adding the number of times in which an interval did not contain the true parameter value divided by the total number of samples.

The estimated left (right) error probability was calculated by adding the number of times the left (right) endpoint was more (less) than the true parameter value divided by the total number of samples. Following Doganaksoy and Schmee (1993a) if the total error probability is greater than $\alpha + 2.58s.e(\hat{\alpha})$, then the method is termed anticonservative, and if it is lower than $\alpha - 2.58s.e(\hat{\alpha})$, the method is termed conservative. The estimated error probabilities are called symmetric when the larger error probability is less than 1.5 times the smaller one.

The value of $\delta = -18$ and $\beta = 4$ were chosen as the parameters of the model to simulate failure times that mimic those seen in automobile brake pad failures. Suppose there are $i = 1, 2, \dots, n$ observations. Random numbers from the uniform distribution on the interval (0,1), u_i , was generated to produce t_i for the i^{th} observation

$$t_i = \left[\frac{1}{e^\delta} \left(\frac{1}{u_i} - 1 \right) \right]^{(1/\beta)} .$$

To obtain censored observation in our data, the censoring time for the i^{th} observation, c_i were simulated from the exponential distribution with parameter μ where the value of μ could be adjusted to obtain the desired approximate censoring proportion in our data. Following that t_i will be censored at c_i if $t_i > c_i$ and uncensored otherwise. The simulation study was carried out via the FORTRAN programming language.

Simulation results

Depicted in Table 1 are the summary of the estimated left, right and total error probabilities for the different methods discussed in the previous section. The results using the B-p method were omitted from the discussion due to the method's poor performance when compared to the other methods. The B-p interval uses the empirical distribution and tends to fail when the distribution of $\hat{\theta}^*$ is highly skewed which is rather common when bootstrapping censored and truncated observations. Inclusion of the B-p results would require a substantial increase in the graphical scale. Thus, we decided only to compare and display the results for the remaining 3 methods in all discussions that follow. More comprehensive results are given in

Tables 2–5. The left and right estimated error probabilities should preferably be equal or close to $\alpha/2$. The overall performances of the different methods were judged based on the total number of anticonservative (AC), conservative (C) and asymmetrical (AS) intervals. We are also interested in methods that behave well at different nominal levels and censoring proportions. Figures 1–8 compare the results of the coverage probability study using different methods graphically.

The alternative computer intensive methods are usually employed to relieve us from tedious calculations and asymptotic normality assumptions. Thus, we wish to see them perform well especially at smaller sample sizes where the intervals based on asymptotic normality usually fail. Based on the results of the simulation study, we see that the DB-t intervals are more reliable than the DB-p and B-t methods. The DB-t method does not produce any conservative or anticonservative intervals for both parameters δ and β , even when censoring proportion in the data is high (cp = 15%).

The DB-t method produced very few asymmetrical intervals, especially for the parameter β at $\alpha = 0.05$ (see Table 1). The estimated error probabilities for the DB-t is also always closer to the nominal compared to the other methods even when the censoring proportion is high and sample size is low ($n = 25$, see Figures 1–8 and Tables 2–3). All methods seem to produce fewer conservative, anticonservative and asymmetrical intervals when $\alpha = 0.10$. So overall we can conclude that the DB-t interval is the best method to employ when dealing with censored data especially at very low sample sizes ($n = 25$). The DB-p tends to work slightly better than the B-t method but both these methods do not perform as well as the DB-t when samples sizes are low ($n = 25$).

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Table 1. Summary of the interval estimates at $\alpha = 0.05$ and 0.10 , $cp = 10\%$ and 15%

		Parameters						
		δ			β			
Methods		C	AC	AS	C	AC	AS	
$\alpha = 0.05$	$cp = 10\%$	Double B-p	0	1	2	0	0	4
		Single B-t	0	1	3	0	2	3
		Double B-t	0	0	1	0	0	3
	$cp = 15\%$	Double B-p	0	1	4	0	0	4
		Single B-t	0	2	4	0	2	4
		Double B-t	0	0	1	0	0	1
$\alpha = 0.10$	$cp = 10\%$	Double B-p	0	0	1	0	0	2
		Single B-t	0	0	2	0	0	2
		Double B-t	0	0	0	0	0	1
	$cp = 15\%$	Double B-p	0	0	2	0	0	3
		Single B-t	0	1	3	0	1	2
		Double B-t	0	0	1	0	0	1

Table 2. Coverage probability of interval estimates for δ at $\alpha = 0.05$

Methods	n	$cp = 10\%$			$cp = 15\%$		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.045	0.028	0.073	0.049	0.024	0.073
	30	0.036	0.026	0.062	0.043	0.020	0.063
	40	0.030	0.022	0.052	0.033	0.020	0.053
	50	0.029	0.017	0.046	0.032	0.015	0.047
Single B-t	25	0.056	0.020	0.076	0.060	0.021	0.081
	30	0.043	0.022	0.065	0.049	0.022	0.071
	40	0.038	0.024	0.062	0.040	0.018	0.058
	50	0.037	0.025	0.062	0.040	0.022	0.062
Double B-t	25	0.029	0.028	0.057	0.033	0.024	0.057
	30	0.022	0.030	0.052	0.019	0.022	0.041
	40	0.016	0.023	0.039	0.014	0.023	0.037
	50	0.014	0.022	0.036	0.018	0.022	0.040

ARASAN & ADAM

Table 3. Coverage probability of interval estimates for β at $\alpha = 0.05$

Methods	n	cp = 10%			cp = 15%		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.006	0.039	0.045	0.002	0.048	0.050
	30	0.008	0.033	0.041	0.004	0.038	0.042
	40	0.010	0.030	0.040	0.013	0.032	0.045
	50	0.014	0.031	0.045	0.012	0.035	0.047
Single B-t	25	0.021	0.055	0.076	0.022	0.058	0.080
	30	0.025	0.045	0.070	0.025	0.045	0.070
	40	0.022	0.035	0.057	0.023	0.043	0.066
	50	0.025	0.033	0.058	0.019	0.037	0.056
Double B-t	25	0.028	0.024	0.052	0.023	0.025	0.048
	30	0.031	0.021	0.055	0.021	0.021	0.042
	40	0.028	0.013	0.041	0.027	0.010	0.037
	50	0.027	0.015	0.042	0.022	0.015	0.037

Table 4. Coverage probability of interval estimates for δ at $\alpha = 0.10$

Methods	n	cp = 10%			cp = 15%		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.068	0.043	0.111	0.070	0.042	0.112
	30	0.057	0.041	0.098	0.064	0.041	0.105
	40	0.052	0.045	0.097	0.054	0.043	0.097
	50	0.059	0.047	0.106	0.060	0.040	0.100
Single B-t	25	0.071	0.043	0.114	0.085	0.045	0.130
	30	0.062	0.045	0.107	0.077	0.038	0.115
	40	0.063	0.044	0.107	0.070	0.045	0.115
	50	0.068	0.043	0.111	0.069	0.047	0.116
Double B-t	25	0.043	0.053	0.096	0.049	0.043	0.092
	30	0.037	0.053	0.090	0.033	0.045	0.078
	40	0.040	0.047	0.087	0.030	0.050	0.080
	50	0.041	0.046	0.087	0.050	0.046	0.096

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Table 5. Coverage probability of interval estimates for β at $\alpha = 0.10$

Methods	n	cp = 10%			cp = 15%		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.036	0.064	0.100	0.029	0.071	0.100
	30	0.037	0.059	0.096	0.035	0.068	0.103
	40	0.040	0.052	0.092	0.038	0.054	0.092
	50	0.040	0.056	0.096	0.036	0.055	0.091
Single B-t	25	0.046	0.077	0.123	0.046	0.086	0.132
	30	0.046	0.065	0.111	0.040	0.078	0.118
	40	0.040	0.069	0.109	0.047	0.066	0.113
	50	0.046	0.065	0.111	0.049	0.066	0.115
Double B-t	25	0.054	0.040	0.094	0.051	0.042	0.093
	30	0.055	0.035	0.090	0.047	0.032	0.079
	40	0.048	0.032	0.080	0.050	0.027	0.077
	50	0.048	0.039	0.087	0.046	0.039	0.085

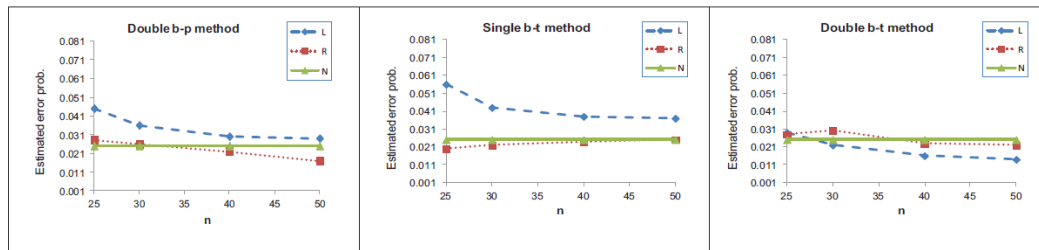


Figure 1. Interval estimates at $\alpha = 0.05$, cp = 10% for parameter δ

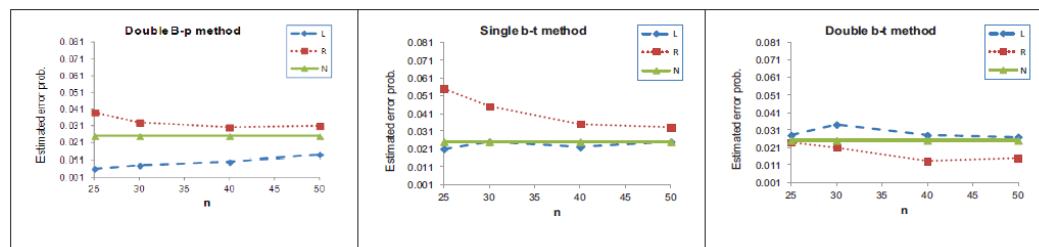


Figure 2. Interval estimates at $\alpha = 0.05$, cp = 10% for parameter β

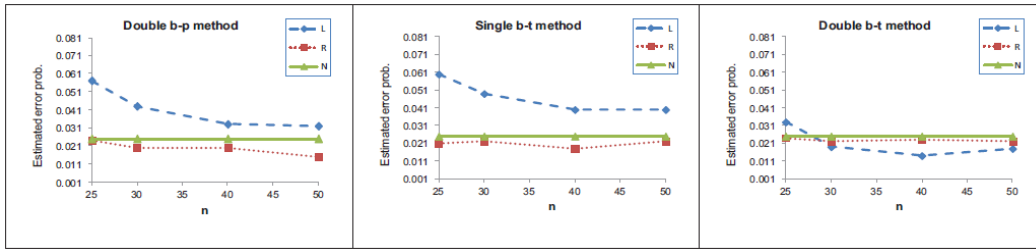


Figure 3. Interval estimates at $\alpha = 0.05$, $cp = 15\%$ for parameter δ

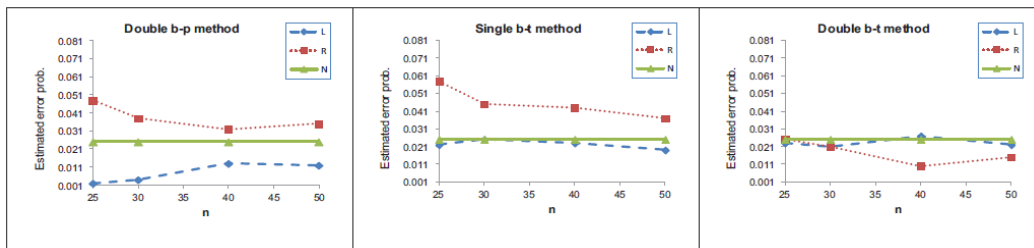


Figure 4. Interval estimates at $\alpha = 0.05$, $cp = 15\%$ for parameter β

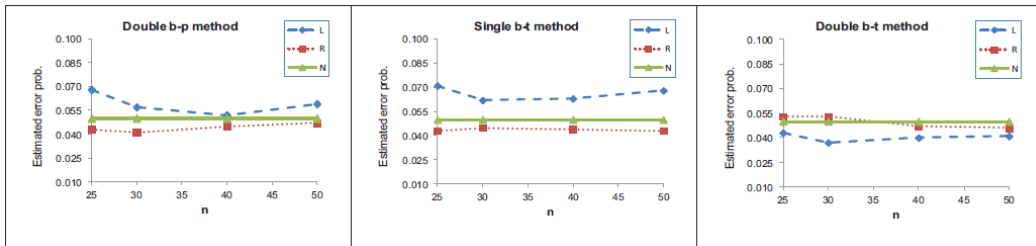


Figure 5. Interval estimates at $\alpha = 0.10$, $cp = 10\%$ for parameter δ

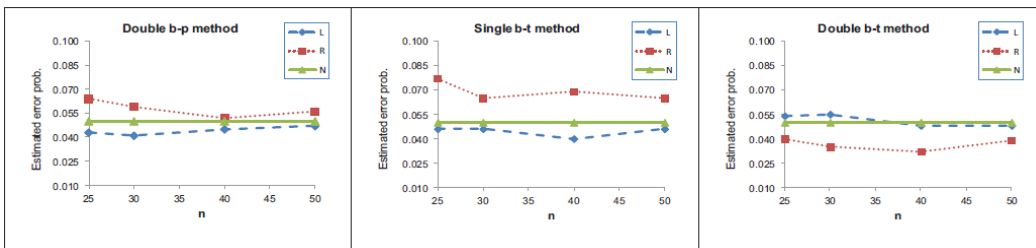


Figure 6. Interval estimates at $\alpha = 0.10$, $cp = 10\%$ for parameter β

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

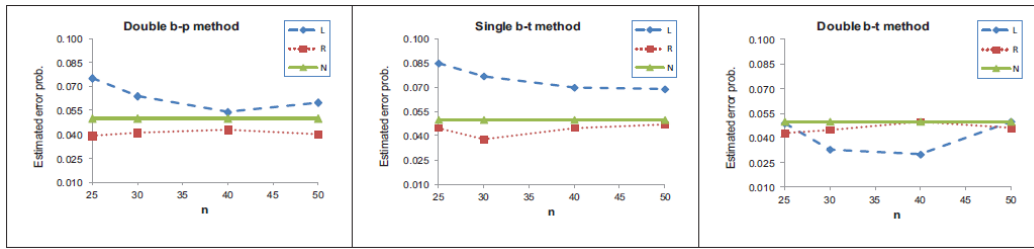


Figure 7. Interval estimates at $\alpha = 0.10$, $cp = 15\%$ for parameter δ

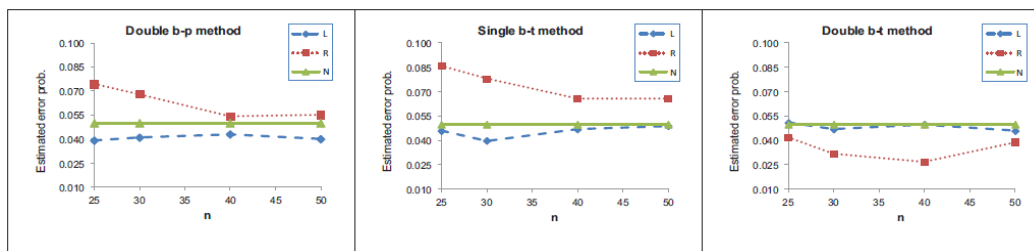


Figure 8. Interval estimates at $\alpha = 0.10$, $cp = 15\%$ for parameter β

Real data analysis

To illustrate the application of the double bootstrap confidence interval procedures, we will consider two data sets dealing with censored and truncated observations respectively. First is the data on breast cancer by Leathem and Brooks (1987) on the lifetimes of 32 women whose tumor has potential to metastasize thus classified as positive staining. 11 of the observations were censored which make the censoring proportion almost 34%.

Considering the sample size and the censoring proportion in the data, any use of inferential procedures based on the asymptotic normality of the maximum likelihood estimates is not advisable. The second is 98 left truncated data on the lifetimes of the brake pads of automobiles. The left truncated lifetimes u_i is the current odometer reading for each car. Only cars that had initial pads were selected and the remaining, and initial pad thickness, were used to estimate t_i . Although the simulation study did not extend to include any truncated data, we believe we may generalize the results obtained using censored data to truncated data since in both cases the distribution of the bootstrap estimates will be skewed and far from normal, see Figures 10 and 12.

Right censored data

The log logistic distribution fits the breast cancer data well as shown by the probability plot in Figure 9. Table 6 shows the parameter estimates when the log logistic distribution is fit to the data and the 95% confidence intervals using the Wald, DB-t, DB-p and B-t interval estimation procedures.

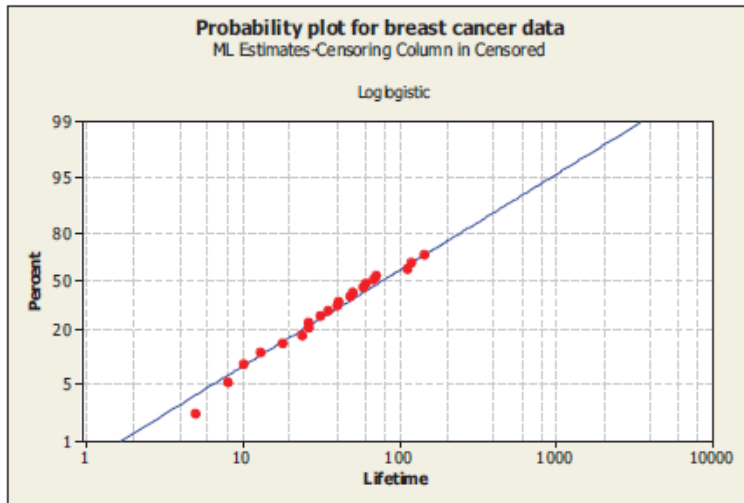


Figure 9. Log logistic probability plot for the breast cancer data

Table 6. MLE of breast cancer data and 95% confidence intervals

θ	$\hat{\theta}$	SE	Wald	B-t	DB-p	DB-t
δ	-5.187	0.953	(-7.054,-3.3189)	(-7.154,-3.549)	(-6.824,-3.744)	(-8.231,-3.824)
β	1.200	0.220	(0.769,1.633)	(0.811,1.666)	(0.826,1.611)	(0.824,1.677)

Figure 10 shows the histogram of 1000 bootstrap replications of $\hat{\delta}$ and $\hat{\beta}$. We can clearly see that both the histogram are not very close to normal shape and appear to be skewed especially the distribution of $\hat{\delta}^*$. Goodness of fit test based on the Anderson-Darling and Kolmogorov-Smirnov had also rejected the assumption of normality at $\alpha = 0.05$. In this case we can expect a disagreement between the standard normal interval and intervals based on the bootstrap methods as highlighted by Efron and Tibshirani (1993).

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

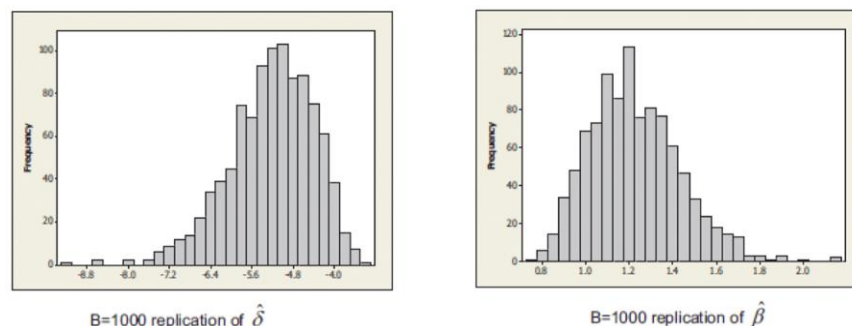


Figure 10. Bootstrap replication of $\hat{\delta}$ and $\hat{\beta}$ for breast cancer data

Based on the coverage probability study, we could see the DB-t intervals produces the most reliable intervals when sample sizes are small and data is highly censored. From Figure 10, we can see that the distribution of $\hat{\delta}^*$ is more skewed than $\hat{\beta}^*$. Thus, as expected the DB-t interval for the parameter δ tends to disagree more with the Wald interval, than the parameter β . The DB-t interval for the parameter δ is also wider than other intervals because it tries to accommodate the skewness in the distribution of the bootstrap estimates, which eventually increases the probability of the true parameter value lying within this interval as verified by the results of the coverage probability study. Other intervals, though narrower may fail to include the true parameter value. Observe that for the parameter β the Wald interval is wider than the DB-t intervals. So the DB-t interval has the ability to adjust itself according to the distribution of the bootstrap estimates which is directly linked to the data in hand and not dependent on any theoretical assumptions, which may fail when the normal approximation is simply not true. So it would actually be more practical to employ the DB-t interval in this case.

Based on the results of the coverage probability study we can see that the DB-p and B-t intervals do not perform as well as the DB-t when sample sizes are small. Their performance tends to improve gradually when $n > 30$. However we have included the interval estimates based on these methods merely to do some comparison study. As we can see the B-t interval is the narrowest among the 4 intervals and there is much doubt as to whether the true parameter value will actually be included within this interval. The DB-p interval is much closer to the Wald interval than the DB-t, especially for parameter δ . This only makes it clearer that the DB-t interval will be more reliable since DB-p did not perform as well as DB-t in the coverage probability study.

Left truncated data

Table 7: MLE of brake pad data and 95% confidence intervals

θ	$\hat{\theta}$	SE	Wald	B-t	DB-p	DB-t
δ	-17.062	1.619	(-20.235,-13.889)	(-20.381,-13.810)	(-20.246,-13.765)	(-20.686,-14.005)
β	4.137	0.383	(3.387,4.888)	(3.397,4.892)	(3.379,4.831)	(3.427,4.974)

Table 7 shows the parameter estimates when the log logistic distribution is fit to the brake pad data and the 95% confidence intervals using the Wald, DB-t, DB-p and

B-t interval estimation procedures. It is known that $e_i = \frac{S(t_i; \hat{\theta})}{S(u_i; \hat{\theta})}$ is $U(0,1)$ given u_i .

Thus, can plot the uniform residual, e_i against the uniform quantile to see if the log logistic distribution fits data as given in Figure 11. As we can see the model fits the data quite well.

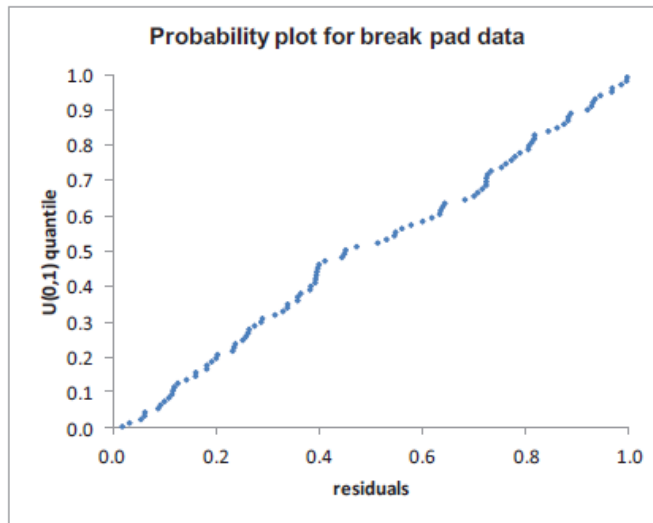


Figure 11. Log logistic probability plot for break pad data

Figure 12 shows the histogram of 1000 bootstrap replications of $\hat{\delta}$ and $\hat{\beta}$. We can also observe that for the left truncated data both the histogram appear to be skewed. The goodness of fit test based on the Anderson-Darling and Kolmogorov-

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Smirnov had again rejected the assumption of normality at $\alpha = 0.05$. As with the breast cancer data we can expect a disagreement between the standard normal interval and intervals based on the bootstrap methods.

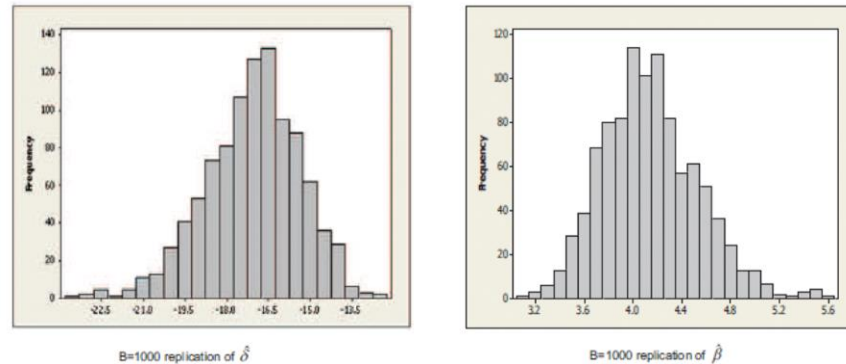


Figure 12. Bootstrap replication of $\hat{\delta}$ and $\hat{\beta}$ for break pad data

The DB-t interval for both the parameters tends to disagree with the Wald interval due to the skewness in the distribution of the bootstrap estimates. The DB-p gives the narrowest interval followed by the Wald. However as we discussed earlier we should be cautious since these intervals may fail to include the true parameter value. Similar pattern is displayed for parameter β where the DB-t interval is the widest but now the DB-p interval is the narrowest. For both parameters the B-t interval and DB-p interval seem to be rather close to the Wald interval which again makes it clearer that the DB-t interval will be more reliable as the B-t and DB-p interval did not perform as well as the DB-t in the coverage probability study.

Discussion

It may appear as if the Wald would suffice as a confidence interval estimate due to its simplicity but this may not be true with smaller data sets that are censored or truncated. So while the Wald can still be employed especially when samples sizes are large and censoring proportions in the data is low, alternative bootstrap methods such as the DB-t should be employed otherwise. Although the DB-t method is slightly more computational compared to the Wald, with the existence of fast computers and parallel computing techniques, these results can be obtained very

quickly especially for small data sets. The performances of the B-t and DB-p method do not seem to be significantly better than the DB-t methods. Thus, we would not recommend employing them for the construction of confidence intervals for the parameters of this model.

Keep in mind that the bootstrap confidence interval methods was initially introduced by Efron and Tibshirani (1993) mostly for use with location statistics such as the mean, median and trimmed mean. However, recently it is being used for more complicated statistics, especially those with standard errors that are attainable. Through simulation studies, we are able to assess if the bootstrap methods can be extended to other general problems, such as constructing confidence interval estimates for model parameters with censored and truncated data. They offer an alternative rather than depending solely on interval estimates that are based on asymptotic normality theory.

Although many are still skeptical about these methods, we can't deny that they provide us with an opportunity to perform comparison study which in some cases may lead to estimates that are better than those produced by traditional methods. The computation time for the double bootstrap is 2 times longer than the ordinary or single bootstrap procedures and negligible for the Wald procedure. For analyzing the breast cancer data, the computation time using the FORTRAN programming language only took 0.015 seconds for the single bootstrap procedures and 0.031 seconds for double bootstrap procedures. The single bootstrap procedure for the brake pad data took 0.03 seconds and the double bootstrap procedures took 0.06 seconds. So, the results can be obtained extremely quickly in only matter of seconds. Thus, the argument of bootstrap methods being heavily computational is not applicable anymore due to the availability of very fast computers.

The methods discussed here can be applied to the parameters of other survival models involving censored or truncated data. The log logistic model discussed here was chosen mainly due to its popularity in most cancer studies and its ability to accommodate both fixed and time dependent covariates easily. When the data in hand has more complicated structures such as double or interval censored the construction of the confidence intervals using solely methods based on asymptotic normality becomes unreliable. Thus, it's important and also interesting to see if these alternative methods provide us with more appealing solutions.

References

- Arasan, J., & Lunn, M. (2008). Alternative interval estimation for parameters of bivariate exponential model with time varying covariate. *Computational Statistics*, 23, 605–622.
- Arasan, J., & Lunn, M. (2009). Survival model of a parallel system with dependent failures and time varying covariates. *Journal of Statistical Planning and Inference*, 139(3), 944–951.
- Bennet, D. S. (1983). Log-Logistic Regression Models for Survival Data. *Applied Statistics*, 32(2), 165-171.
- Cox, D. R., & Lewis, P. A. (1966). *Statistical Analysis of Series of Events*. London: Methuen.
- Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman and Hall.
- Diciccio, T. J., & Efron, B. (1996). Bootstrap confidence interval. *Statistical Science*, 11(3), 189–228.
- Doganaksoy, N. & Schmee, J. (1993a). Comparison of approximate confidence intervals for distributions used in life-data analysis. *Technometrics*, 35(2), 175–184.
- Doganaksoy, N., & Schmee, J. (1993b). Comparison of approximate confidence intervals for smallest extreme value distribution simple linear regression model under time censoring. *Communications in Statistics – Simulation and Computation*, 2, 175–184.
- Efron, B. (1981a). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9, 139–172.
- Efron, B. (1981b). Nonparametric standard errors: the jackknife, the bootstrap and other methods. *Biometrika*, 68, 589–599.
- Efron, B. (1985). Bootstrap confidence interval for a class of parametric problem. *Biometrika*, 72, 45–58.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Gupta, R. C., Akman, O., & Lvin, S. (1999). A study of log-logistic model in survival analysis. *Biometrical Journal*, 41(4), 431–443.
- Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics*, 14, 1431–1452.

- Hall, P. (1988a). Theoretical comparison of bootstrap confidence intervals (with discussion). *Annals of Statistics*, 16, 927–953.
- Hall, P. (1988b). On symmetric bootstrap confidence intervals. *Journal of Royal Statistical Society, B(50)*, 35–45.
- Jeng, S. L., & Meeker, W. Q. (2000). Comparison for approximate confidence interval procedure for type I censored data. *Technometrics*, 42(2), 135–148.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley and Sons.
- Leathem, A. J., & Brooks, S. (1987). Predictive value of lectin binding on breast-cancer recurrence and survival. *The Lancet*, 329(8541), 1054-1056.
- Letson, D. & McCullough, B. D. (1998). Better confidence intervals: the double bootstrap with no pivot. *American journal of agricultural economics*, 80(3), 552–559.
- Martin, M. A. (1992). On the double bootstrap. New York: Springer.
- Mazucheli, J., Barros, E. A. C., & Achcar, J. A. (2005). Bootstrap confidence intervals for the mode of the hazard function. *Computer Methods and Programs in Biomedicine*, 79(1), 39–47.
- McCullough, B. D., & Vinod, H. D. (1998). Implementing the double bootstrap. *Computational Economics*, 12(1), 79–95.
- O’Quigley, J. & Struthers, L. (1982). Survival models based upon the logistic and log-logistic distributions. *Computer Programs in Biomedicine*, 15(1), 3–11.
- Robinson, J. A. (1983). Bootstrap confidence intervals in location-scale models with progressive censoring. *Technometrics*, 25(2), 179–187.
- Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of American Statistical Association*, 80, 360–361.
- Shi, S. G. (1992). Accurate and efficient double-bootstrap confidence limit method. *Computational Statistics & Data Analysis*, 13(1), 21–32.
- Singh, K. (1981). On the asymptotic accuracy of Efron’s bootstrap. *Annals of Statistics*, 9, 1187–1195.