DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 13 | Issue 2 Article 15

11-2014

Survival Estimation Using Bootstrap, Jackknife and K-Repeated Jackknife Methods

Johnson A. Adewara
University of Lagos, Akoka-Lagos, Nigeria, jadewara@unilag.edu.ng

Ugochukwu A. Mbata
University of Lagos, Akoka-Lagos, Nigeria, mmbataugochukwu@yahoo.com

Journal of Modern Applied Statistical Methods
November 2014, Vol. 13, No. 2, 287-306.

ISSN 1538 − 9472

Survival Estimation Using Bootstrap, Jackknife and K-Repeated Jackknife Methods

Johnson Ademola Adewara
University of Lagos
Akoka-Lagos, Nigeria

Ugochukwu Ahamefula Mbata
University of Lagos
Akoka-Lagos, Nigeria

Three re-sampling techniques are used to estimate the survival probabilities from an exponential life-time distribution. The aim is to employ a technique to obtain a parameter estimate for a two-parameter exponential distribution. The re-sampling methods considered are: Bootstrap estimation method (BE), Jackknife estimation method (JE) and the *k*-repeated Jackknife estimation method (KJE). The methods were computed to obtain the mean square error (MSE) and mean percentage error (MPE) based on simulated data. The estimates of the two-parameter exponential distribution were substituted to estimate survival probabilities. Results show that the MSE value is reduced when the *K*–repeated jackknife method is used.

Introduction

Modern statistics is anchored in the use of statistics and hypothesis tests that only have desirable and well-known properties when computed from populations that are normally distributed. While it is claimed that many such statistics and hypothesis tests are generally robust with respect to non-normality, other approaches that require an empirical investigation of the underlying population distribution or of the distribution of the statistic are possible and in some instances preferable. In instances when the distribution of a statistic, conceivably a very complicated statistic, is unknown, no recourse to a normal theory approach is available and alternative approaches are required. Statistical models and methods for estimating survival data and other time-to-event data are extensively used in many fields, including the biomedical sciences, engineering, the environmental sciences, economics, actuarial sciences, management, and the social sciences.

Johnson Ademola Adewara is a Senior Lecturer at the Distance Learning Institute at the University of Lagos. Email at jadewara@unilag.edu.ng
U.A. Mbata is a Graduate Fellow and Consulting Statistician in the Mathematics Department at the University of Lagos. Email at mmbataugochukwu@yahoo.com

Survival analysis refers to techniques for studying the occurrence and timing of events. It is concerned with studying the random variable *T*, representing the time between entry to a study and some event of interest, such as: death, the onset of disease, time until equipment failures, earthquakes, automobile accidents, time-to-promotions, time until stock market crashes, revolutions, job terminations, births, marriages, divorces, retirements or arrests. There are many different models for survival data, and what often distinguishes one model from another is the probability distribution for *T*. Resampling statistics refer to the use of the observed data or of a data generating mechanism (such as a die) to produce new hypothetical samples (resamples) that mimic an underlying population, the results of which can then be analyzed. With numerous cross-disciplinary applications especially in the sub-disciplines of the life science, resampling methods are widely used because they are options when parametric approaches are difficult to employ or otherwise do not apply.

Resampled data is derived using a manual mechanism to simulate many pseudo-trials. These approaches were difficult to utilize prior to 1980s because these methods require many repetitions. With the incorporation of computers, the trials can be simulated in a few minutes and is why these methods have become widely used. The methods that will be discussed are used to make many statistical inferences about the underlying population. The most practical use of resampling methods is to derive confidence intervals and test hypotheses. This is accomplished by drawing simulated samples from the data themselves (resamples) or from a reference distribution based on the data; afterwards, you are able to observe how the statistic of interest in these resamples behaves. Resampling approaches can be used to substitute for traditional statistical (formulaic) approaches or when a traditional approach is difficult to apply. These methods are widely used because their ease of use. They generally require minimal mathematical formulas, needing a small amount of mathematical (algebraic) knowledge. These methods are easy to understand and stray away from choosing an incorrect formula in your diagnostics.

Two general approaches considered here are: Jackknife approach and Bootstrap approach. The aim of this study is to employ a technique to obtain an estimate of the parameter of the two-parameter exponential distribution. The methods considered in this paper are: Bootstrap estimation method (BE), Jackknife estimation method (JE) and the *k*-repeated Jackknife estimation method (KJE). The estimates of the two-parameter exponential distributions are used to estimate the survival probability. Methodology under Bootstrap, Jackknife and the proposed K-repeated Jackknife is presented, followed by data analysis, results, discussion and a conclusion.

Two-Parameter Exponential Distribution

Re-sampling methods are becoming increasingly popular as statistical tools. These methods involve sampling or scrambling the original data numerous times. Two general approaches are considered here. They are: Jackknife approach and Bootstrap approach. The two-parameter exponential distribution is adopted when failure will never occur prior to some specified time, t_0. The parameter t_0 is a location parameter that shifts the distribution an amount equal to t_0 towards the right on the time line. When $t \geq t_0$, the probability density function of exponential distribution becomes:

$$f(t;\theta) = \frac{1}{\theta} exp\left(-\frac{1}{\theta}(t - t_0)\right), t \geq 0, \theta > 0, t \geq t_0 \tag{1}$$

and the survival function is given by:

$$S(t) = \int_t^\infty f(t;\theta)\,dt \tag{2}$$

where $\lambda = \frac{1}{\theta}$.

Bootstrap Estimation Method

Bootstrapping is a modern, computer-intensive, general purpose approach to statistical inference, falling within a broader class of re-sampling methods to simplify the often intricate calculations of traditional statistical theory. A parametric bootstrap method is considered in this article.

The general theory (see Rizzo, 2008) is as follows. Suppose $t_1, t_2, \ldots, t_n$ is a random sample from the distribution of *T*. An estimator $\hat{\theta}$ for a parameter θ is an n variate function $\hat{\theta} = \hat{\theta}(t_1, t_2, \ldots, t_n)$ of the sample. Functions of the estimator $\hat{\theta}$ are therefore n–variate functions of the data, also. For simplicity, let $t = (t_1, t_2, \ldots, t_n)^T \in R^n$, and $t^{(1)}, t^{(2)}, \ldots$ denote a sequence of independent random samples generated from the distribution of *T*. Random variables from the sampling distribution of $\hat{\theta}$ can be generated by repeatedly drawing independent random samples $t^{(j)}$ and computing $\hat{\theta}^{(j)} = \hat{\theta}\left(t_1^{(j)}, t_2^{(j)}, \ldots, t_n^{(j)}\right)$ for each sample. The mean of the replicates is given as

$$\hat{\theta}_B = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}^{(j)}. \tag{3}$$

The mean squared error (MSE) is defined by $MSE\left(\hat{\theta}\right) = E\left[\left(\hat{\theta}-\theta\right)^2\right]$. If m random samples $t^{(1)}, t^{(2)}, \ldots, t^{(m)}$ are generated from the distribution of T then estimate of the MSE of $\hat{\theta} = \hat{\theta}\left(t_1, t_2, \ldots, t_n\right)$ is

$$MSE\left(\hat{\theta}_B\right) = \frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{\theta}^{(j)}-\theta\right)^2 \tag{4}$$

where $\hat{\theta}^{(j)} = \hat{\theta}\left(t_1^{(j)}, t_2^{(j)}, \ldots, t_n^{(j)}\right)$.

Estimate of the standard error of the bootstrap estimate, $\hat{\theta}_B$ is given by

$$SE\left(\hat{\theta}_B\right) = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{\theta}^{(j)}-\theta\right)^2} \tag{5}$$

100(1−α)% confidence interval for θ is given by

$$\hat{\theta}_B \pm Z_{\alpha/2} SE\left(\hat{\theta}_B\right) \tag{6}$$

The mean percentage error (MPE) is

$$MPE\left(\hat{\theta}_B\right) = \frac{\left[\sum_{j=1}^{m}\frac{\left|\hat{\theta}^{(j)}-\theta\right|}{\theta}\right]}{m}. \tag{7}$$

If the bootstrap estimator $\hat{\theta}_B$ is known from (3) then the estimate of survival function is given as

$$\hat{S}_B(t) = exp\left(-\frac{t_i - t_0}{\hat{\theta}_B}\right) \tag{8}$$

Jackknife Estimation

The jackknife is a more orderly version of the bootstrap. As opposed to re-sampling randomly from the entire sample like the bootstrap does, the jackknife takes the entire sample except for 1 value, and then calculates the test statistic of interest. It repeats the process, each time leaving out a different value, and each time recalculating the test statistic. This method was introduced by Quenouille (1949) and further modification in Quenouille (1956). The theory is as follows:

Let $\hat{\theta}$ be an estimator of the parameter θ based on the complete sample of size n with g subgroups. Let $\hat{\theta}_{-i}$ be the corresponding estimator based on the sample at the i^{th} deletion. Define

$$\tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} \, (i = 1, 2, \ldots, g) \tag{9}$$

The i^{th} deletion of the total could be one individual observation or several observation. The latter case is called group- or block-based jackknife if one replication or one block observations are deleted. In equation (1) estimation $\tilde{\theta}_i$ is called the i^{th} pseudo value and the estimator in equation (2) is the jackknife estimator for the parameter θ, where θ can be a variance component, covariance component, correlation coefficient, or any other parameter of interest.

$$\tilde{\theta} = \frac{1}{g}\sum_{j=1}^{g}\tilde{\theta}_i = g\hat{\theta} - (g-1)\frac{1}{g}\sum_{j=1}^{g}\hat{\theta}_{-i} \tag{10}$$

In equation (10), $\tilde{\theta}$ is called a pseudo jackknife estimate. A t-test can then be used to test significant deviation from a given parameter value, θ_0 with degrees of freedom $g-1$ (Miller, 1974a, b). The equation (9) can be rewritten as

$$\tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} = \hat{\theta} + (g-1)\left(\hat{\theta} - \hat{\theta}_{-i}\right)(i = 1, 2, \ldots, g). \tag{11}$$

Thus, it is obvious that pseudo value $\tilde{\theta}_i$ in equation (11) is related to choices for g. When g is large, a slight difference between $\hat{\theta}$ and $\hat{\theta}_{-i}$ will cause unfavorable values. More importantly, it will potentially cause a large standard error for an estimate and thus decrease the power for the parameter being tested. If it is assumed that the estimate $\hat{\theta}_{-i}$ in equation (9) for the i^{th} deletion is unbiased, then it is easy to prove that $\bar{\hat{\theta}}$ in equation (12) is unbiased too. It is often true if $\hat{\theta}$ is an unbiased estimate of θ, then $\hat{\theta}_{-i}$ will be unbiased after a few individuals in the original data are deleted.

$$\bar{\hat{\theta}}_{jack} = \frac{\sum_{i=1}^{g} \hat{\theta}_{-i}}{g}. \tag{12}$$

In equation (12), $\bar{\hat{\theta}}$ is called a non-pseudo jackknife estimate of the parameter θ. For each non-normally distributed variable, based on the Central Limit Theorem, $\bar{\hat{\theta}}$ is approximately normally distributed when g is large. Thus, an approximate z-test can be used when g is large or t-test can be used to test significant deviation from a given parameter value, θ_0, with the degrees of freedom $g-1$. An estimate of the mean square error (MSE) of the jackknife estimate, $\bar{\hat{\theta}}_{jack}$ is given by

$$MSE\left(\bar{\hat{\theta}}_{jack}\right) = \frac{g}{g-1} \sum_{i=1}^{g} \left(\hat{\theta}_{-i} - \bar{\hat{\theta}}\right)^2 \tag{13}$$

Estimate of the standard error of the jackknife estimate, $\bar{\hat{\theta}}_{jack}$ is given by

$$SE\left(\bar{\hat{\theta}}_{jack}\right) = \sqrt{\frac{g}{g-1} \sum_{i=1}^{g} \left(\hat{\theta}_{-i} - \bar{\hat{\theta}}\right)^2} \tag{14}$$

$100(1-\alpha)\%$ confidence interval for θ is given by

$$\bar{\hat{\theta}}_{jack} \pm t_{\alpha/2, g-1} SE\left(\bar{\hat{\theta}}_{jack}\right)$$

The mean percentage error (MPE) is

$$MPE\left(\hat{\theta}_{jack}\right) = \frac{\left[\sum_{i=1}^{g} \frac{\left|\hat{\theta}_{-i} - \bar{\hat{\theta}}\right|}{\bar{\hat{\theta}}}\right]}{g}. \tag{15}$$

If the jackknife estimator $\hat{\theta}_{jack}$ is known from (12) then the estimate of survival function is given as

$$\hat{S}_{jack}(t) = exp\left(-\frac{t_i - t_0}{\bar{\hat{\theta}}_{\text{jack}}}\right). \tag{16}$$

K – Repeated Jackknife Estimation Method

The *K* – Repeated jackknife procedure is a re-sampling iterative scheme for mean square error (MSE) reduction. This involves jackknifing the observed data *k*-time, where *k* equals the sample size of the observed data. The procedure is conveniently applied when the sample size is small. The stopping rule for the repeated jackknife replications depends on the sample size of the original data. The procedure converges before or at k^{th} time, where the estimate from the jackknife replications is the same as estimator of the parameter θ based on the complete sample of size *n*. At the K^{th} time, the k^{th} – repeated jackknife estimate of bias is highly negligible.

The method involves the following steps from the usual jackknife procedure:

Step 1. Observe a random sample $T = (t_1, t_2, \ldots, t_n)$

Step 2. Compute $\hat{\theta}(t)$ a function of the data which estimates the parameter θ of the model.

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} t_i \qquad i = 1, 2, \ldots, n \tag{17}$$

Step 3. For *i* up to *n*

- generate a jackknife sample $T_{-i} = (t_1, t_{i-1}, t_{i+1}, \ldots, t_n)$ by leaving out the i^{th} observation
- calculate $\hat{\theta}_{-i}$ from each of the Jackknife sample T_{-i} by

$$\hat{\theta}_{-i} = \frac{1}{n-1}\sum_{i=1}^{n-1} T_{-i} \tag{18}$$

Step 4. Repeat step 3 using the estimates from $\hat{\theta}_{-i}$ to form pseudo samples. The new pseudo samples are used to generate another set of jackknife estimates; this is continued until the k^{th} time. This implies that the process is repeated k times, and at any given stage the preceding jackknife estimates are used as new samples in the next stage until the k^{th} time.

Step 5. At the k^{th} time the K-repeated Jackknife estimate is calculated as

$$\bar{\hat{\theta}}^{K} = \frac{1}{k}\sum_{i=1}^{n}\hat{\theta}_{i-1}^{K} \tag{19}$$

The K – repeated jackknife estimate of mean square error (MSE) is given by

$$MSE(\bar{\hat{\theta}}^{K}) = \frac{1}{k(k-1)}\sum_{i=1}^{n}\left(\hat{\theta}_{i-1}^{K} - \bar{\hat{\theta}}^{K}\right)^2 \tag{20}$$

The K – repeated jackknife estimate of standard error is given by

$$SE\left(\bar{\hat{\theta}}^{K}\right) = \sqrt{\frac{1}{k(k-1)}\sum_{i=1}^{n}\left(\hat{\theta}_{i-1}^{K} - \bar{\hat{\theta}}^{K}\right)^2} \tag{21}$$

An approximate (1–α)% confidence interval for θ is given by

$$\bar{\hat{\theta}}^{K} \pm t_{\alpha/2, K-1} SE\left(\bar{\hat{\theta}}^{K}\right) \tag{22}$$

The mean percentage error (MPE) is

$$MPE\left(\bar{\hat{\theta}}^{K}\right) = \frac{\left[\sum_{i=1}^{n}\frac{\left|\hat{\theta}_{i-1}^{K} - \bar{\hat{\theta}}^{K}\right|}{\bar{\hat{\theta}}^{K}}\right]}{k}. \tag{23}$$

If the K – repeated jackknife estimator $\bar{\hat{\theta}}^{K}$ is known from (19), then the estimate of survival function is

$$\hat{S}^{K}(t) = exp\left(-\frac{t_i - t_0}{\bar{\hat{\theta}}^{K}}\right). \tag{24}$$

The general iterative scheme is as follows: from a random sample $T = (t_1, t_2, \ldots, t_n)$

1. $\hat{\theta}^{1}_{(-1)} = \frac{1}{n-1}\sum_{i=1}^{n-1} T_{-i}$
2. $\hat{\theta}^{2}_{(-1)} = \frac{1}{n-1}\sum_{i=1}^{n-1} \hat{\theta}^{1}_{-1}$
3. $\hat{\theta}^{3}_{(-1)} = \frac{1}{n-1}\sum_{i=1}^{n-1} \hat{\theta}^{2}_{-1}$

. . . .

. . . .

K. $\hat{\theta}^{K}_{(-1)} = \frac{1}{n-1}\sum_{i=1}^{n-1} \hat{\theta}^{K-1}_{-1}$

Thus,

$$\bar{\hat{\theta}}^{K} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}^{K}_{i-1}$$

where $K = n$ (sample size) indicates the stopping rule. Other estimators such as variance, standard error and confidence interval can be estimated as in (20), (21), (22) and (23).

This study described three types of parameter estimation methods based on re-sampling technique: the bootstrap method, the jackknife method and the *k*-repeated jackknife method. However, the intention of this study is to use Monte Carlo simulated data to compare the three methods based on mean squared error (MSE) and mean percentage error (MPE), hence survival estimation.

Data and Analysis

Exploratory data analysis approach using simulated data generated by R-statistical program is adopted in this research work. This is to validate the statistical assumptions of an exponential distribution. In statistics, every statistical model has its own assumptions that have to be verified and met, to provide valid results. In the case of exponential distribution, the confidence interval for the mean life of an event requires two major assumptions: the time-to-occurrence of events of interest are independent, and the time for occurrence of event is exponentially distributed. These two statistical assumptions must be satisfied for the corresponding confidence interval to cover the true mean with the prescribed probability. The simulated data is based on random generation of values which satisfies both the assumption of independence and exponentially identical distribution. Some properties of the exponential distribution are as follows: the theoretical mean and standard deviation are equal. Hence, (1) the sample values of mean and standard deviation should be close. (2) Histogram should show that the distribution is right skewed (Median < Mean). (3) A plot of Cumulative-Failure vs. Cumulative-Time should be close to linear. (4) The regression slope of Cum-Failure vs. Cum-Time is close to the failure rate. (5) A plot of Cum-Rate vs. Cum-Failure should decrease/stabilize at the failure rate level. (6) Plots of the Exponential probability and its scores should also be close to linear. Some of these properties are explained by the exploratory data analysis displayed in Figures 1, i - xii.

Table 1. Generation Parameters

Sample 1 (Sample size = 10, λ = 0.5)	Sample 7 (Sample size = 20, λ = 1.5)
Sample 2 (Sample size = 10, λ = 1.0)	Sample 8 (Sample size = 20, λ = 2.0)
Sample 3 (Sample size = 10, λ = 1.5)	Sample 9 (Sample size = 30, λ = 0.5)
Sample 4 (Sample size = 10, λ = 2.0)	Sample 10 (Sample size = 30, λ = 1.0)
Sample 5 (Sample size = 20, λ = 0.5)	Sample 11 (Sample size = 30, λ = 1.5)
Sample 6 (Sample size = 20, λ = 1.0)	Sample 12 (Sample size = 30, λ = 2.0)

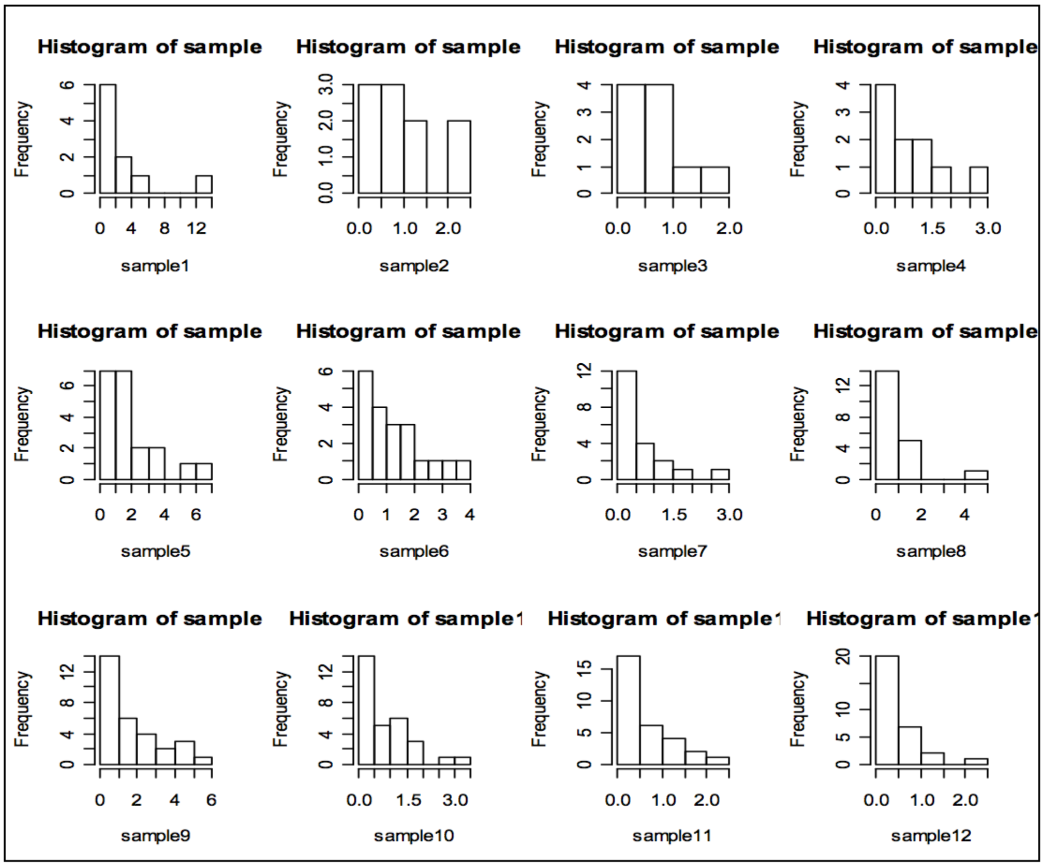

Figure 1. Histogram for each of the Randomly Generated Sample (i – xii)

Table 2. Descriptive Statistics of samples 1 to 12

Sample (i)	N	λ	Mean	Median	Std.Dev.
Sample 1	10	0.5	2.6006960	1.4973290	3.7961510
Sample 2	10	1.0	0.9211933	0.7875227	0.7792381
Sample 3	10	1.5	0.7320056	0.5990053	0.5961606
Sample 4	10	2.0	0.8510537	0.7067248	0.8785794
Sample 5	20	0.5	1.8608660	1.1843510	1.8059580
Sample 6	20	1.0	1.2379810	0.9922479	1.0967930
Sample 7	20	1.5	0.5363318	0.2470020	0.6540290
Sample 8	20	2.0	0.7281072	0.2824484	1.0112260
Sample 9	30	0.5	1.5960620	1.0194960	1.6278560
Sample 10	30	1.0	0.8559385	0.5422463	0.8296031
Sample 11	30	1.5	0.6353941	0.3864003	0.5880883
Sample 12	30	2.0	0.4639596	0.2700518	0.5150784

Results

The results of descriptive statistics show that as the sample sizes 10, 20 and 30 increase the mean and standard deviation are decreasing which satisfied one of the properties of that the theoretical mean and standard deviation are equal. The sample mean and standard deviation obtained are very close also as the value of λ increases the median values obtained get smaller. Figure 1 above shows that the observed distribution agrees with the exponential distribution property 1 and property 2 described in the data above. Figures 1, i–xii show right-skewness, which supported the attribute of an exponential distribution.

Table 3 shows the parameter estimation of the three methods. The results reveal that the estimation of the bootstrap approach is better than the other two methods that is the jackknifing and *K* repeated jackknifing. Table 4 is the result of the mean square error (MSE) of the analysis which is about the variance of the three methods. Results reveal that, as λ values increase, the results of jackknifing and *K* repeated jackknifing are better than the bootstrapped approach. Table 5 is the computation of the mean percentage error (MPE) the result shows that estimation of the bootstrap approach is better than the other two methods.

Table 3. Estimation Using the Three Methods Bootstrap, Jackknifing and K repeated jackknifing

	λ	$\hat{S}_B(t)$	$\hat{S}_{jack}(t)$	$\hat{S}^K(t)$
10	0.5	0.568858094	0.568879887	0.568879887
	1	0.476453626	0.476456925	0.476456925
	1.5	0.461343936	0.461328523	0.461328523
	2.0	0.529933691	0.529937819	0.529937819
20	0.5	0.491722891	0.491777729	0.491777729
	1	0.490229963	0.490240047	0.490240047
	1.5	0.544947075	0.544930402	0.544930402
	2.0	0.553586925	0.553580134	0.553580134
30	0.5	0.527441921	0.527445588	0.527445588
	1	0.491819455	0.491882638	0.491882638
	1.5	0.491085203	0.491099760	0.491099760
	2.0	0.528125037	0.528118624	0.528118624

Table 4. Estimation to the Bootstrap, Jackknifing and K repeated jackknifing using MSE methods

	λ	$\hat{S}_B(t)$	$\hat{S}_{jack}(t)$	$\hat{S}^K(t)$
10	0.5	0.004741437	0.004744439	0.004744439
	1	0.000554432	0.000554276	0.000554276
	1.5	0.001494291	0.001495483	0.001495483
	2.0	0.000896026	0.000896273	0.000896273
20	0.5	0.000068511	0.000067606	0.000067606
	1	0.000095454	0.000095257	0.000095257
	1.5	0.002020240	0.002018741	0.002018741
	2.0	0.002871559	0.002870831	0.002870831
30	0.5	0.000753059	0.000753260	0.000753260
	1	0.000066921	0.000065892	0.000065892
	1.5	0.000079474	0.000079214	0.000079214
	2.0	0.000791018	0.000790657	0.000790657

Table 5. Bootstrap, Jackknifing and K-repeated jackknifing using MPE methods

	λ	$\hat{S}_B(t)$	$\hat{S}_{jack}(t)$	$\hat{S}^K(t)$
10	0.5	0.137716188	0.137759774	0.137759774
	1	0.047092748	0.047086150	0.047086150
	1.5	0.077312128	0.077342954	0.077342954
	2.0	0.059867382	0.059875638	0.059875638
20	0.5	0.016554218	0.016444542	0.016444542
	1	0.019540074	0.019519906	0.019519906
	1.5	0.089894150	0.089860804	0.089860804
	2.0	0.107173850	0.107160268	0.107160268
30	0.5	0.054883842	0.054891176	0.054891176
	1	0.016361090	0.016234724	0.016234724
	1.5	0.017829594	0.017800480	0.017800480
	2.0	0.056250074	0.056237248	0.056237248

Table 6. Survival Estimation Using the Three Methods with Respect to MSE and MPE

BOOTSTRAP METHOD (1)

Size	λ	$\hat{S}_B(t)$	MSE	MPE	REMARK
10	0.5	0.568858094	0.004741437	0.137716188	1
	1	0.476453626	0.000554432	0.047092748	2,3
	1.5	0.461343936	0.001494291	0.077312128	1
	2.0	0.529933691	0.000896026	0.059867382	1
20	0.5	0.491722891	0.000068511	0.016554218	2,3
	1	0.490229963	0.000095454	0.019540074	2,3
	1.5	0.544947075	0.002020240	0.089894150	2,3
	2.0	0.553586925	0.002871559	0.107173850	2,3
30	0.5	0.527441921	0.000753059	0.054883842	1
	1	0.491819455	0.000066921	0.016361090	2,3
	1.5	0.491085203	0.000079474	0.017829594	2,3
	2.0	0.528125037	0.000791018	0.056250074	2,3

Table 6, cont'd. Survival Estimation Using the Three Methods with Respect to MSE and MPE

JACKKNIFE METHOD (2)

Size	λ	$\hat{S}_{jack}(t)$	MSE	MPE	REMARK
10	0.5	0.568858094	0.004741437	0.137716188	1
	1	0.476453626	0.000554432	0.047092748	2,3
	1.5	0.461343936	0.001494291	0.077312128	1
	2.0	0.529933691	0.000896026	0.059867382	1
20	0.5	0.491722891	0.000068511	0.016554218	2,3
	1	0.490229963	0.000095454	0.019540074	2,3
	1.5	0.544947075	0.002020240	0.089894150	2,3
	2.0	0.553586925	0.002871559	0.107173850	2,3
30	0.5	0.527441921	0.000753059	0.054883842	1
	1	0.491819455	0.000066921	0.016361090	2,3
	1.5	0.491085203	0.000079474	0.017829594	2,3
	2.0	0.528125037	0.000791018	0.056250074	2,3

K-REPEATED JACKKNIFE METHOD (3)

Size	λ	$\hat{S}^{K}(t)$	MSE	MPE	REMARK
10	0.5	0.568858094	0.004741437	0.137716188	1
	1	0.476453626	0.000554432	0.047092748	2,3
	1.5	0.461343936	0.001494291	0.077312128	1
	2.0	0.529933691	0.000896026	0.059867382	1
20	0.5	0.491722891	0.000068511	0.016554218	2,3
	1	0.490229963	0.000095454	0.019540074	2,3
	1.5	0.544947075	0.002020240	0.089894150	2,3
	2.0	0.553586925	0.002871559	0.107173850	2,3
30	0.5	0.527441921	0.000753059	0.054883842	1
	1	0.491819455	0.000066921	0.016361090	2,3
	1.5	0.491085203	0.000079474	0.017829594	2,3
	2.0	0.528125037	0.000791018	0.056250074	2,3

Discussion

The comparison of parametric estimators of exponential distribution using Bootstrap, Jackknife, and *K*-Repeated Jackknife methods indicates that the estimates of the population parameter are very close which implies that the estimators are unbiased. A comparison of the mean square error (MSE) and mean percentage error (MPE) of the estimators shows that *K*-Repeated Jackknife method has a minimum variance unbiased estimator (MVUE); irrespective of the sample size whether it is small or large at any given values of lambda (λ). The three methods are used to estimate the survival function for exponential distribution and its mean square error (MSE) and mean percentage error (MPE). The results can be deduced that the performance of the two jackknife procedures over the bootstrap procedure is 66.67% to 33.33%. This result has been able to show the effect or influence of jackknife method, especially the *k*-repeated procedure on error reduction in estimating population parameter.

Conclusion

This study demonstrates that both methods of re-sampling technique are very efficient in estimating the population parameters and their mean square errors (MSE), as viewed by Efron (1998). These methods were used to find the best minimum variance unbiased estimator, using mean square error (MSE) and mean percentage error (MPE). The estimates of the two-parameter exponential distribution are used to estimate the survival probability. The attractiveness of jackknifing and bootstrapping is that they provide investigators with an important and unattainable type of information. Jackknifing and bootstrapping have their limitations and inherent assumptions as all statistical procedures do. The three methods are computationally intensive. However, these techniques represent an important step in refining the process of data analysis more especially the *k*-repeated procedure. Hence, it can be deduced that bootstrapping is a method for evaluating the variance of an estimator while jackknife is a method for reducing the bias of an estimator, and evaluating the variance of an estimator. This is clearly shown in the MSE results. The MSE value is reduced using the *K*–repeated jackknife method.

References

Bradley, J. U. (1968). *Distribution-free statistical tests*. Englewood Cliff, NJ: Prentice-Hall.

Collett, D. (1994). *Modeling survival data in medical research*. London: Chapman and Hall.

Conover, W. J. (1980). *Practical nonparametric statistics*. New York: John Wiley and Sons.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. New York: Chapman and Hall.

Crowley, P. (1992). Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecological Systems*, *23*, 405-447.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*, 36-48.

Efron, B., & Tibshirani, R. (1998). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Fernández, A. J. (2000). Estimation and Hypothesis Testing for Exponential Lifetime Models with Double Censoring and Prior Information. *Journal of Economic and Social Research*, *2*(2), 1-17.

Good, P. L. (2005). *Resampling methods* (3rd ed.). Berlin: Birkhauser.

Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: John Wiley and Sons.

Miller, R. G. (1974a). An unbalanced jackknife. *Annals of Statistics*, *2*, 880-891.

Miller, R. G. (1974b). The jackknife-a review. *Biometrika*, *61*, 1-15.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society*, *Series B*, *11*, 68-84.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, *43*, 353-360.

Rizzo, M. L. (2008). *Statistical computing with R*. Boca Raton, FL: Chapman and Hall/CRC.

Suzuki, K. (1985a). Nonparametric estimation of lifetime distributions from a record of failures and follow-ups. *Journal of the American Statistical Association*, *80*, 68-72.

Suzuki, K. (1985b). Estimation of lifetime parameters from incomplete field data. *Technometrics*, *27*, 263-271.

Appendix

Simulated Data

Table A1. Sample Size = 10

S/N	Sample 1(λ= 0.5)	Sample 2(λ= 1.0)	Sample 3(λ= 1.5)	Sample 4(λ= 2.0)
1	0.38672500	2.27994293	0.20192160	0.01058359
2	0.11665480	0.71411938	0.48301090	2.51951031
3	0.12661390	0.75298901	0.12473380	1.94495769
4	4.42286640	2.19289233	1.99662440	0.67531853
5	2.00704600	0.04442099	0.67780240	1.05647689
6	2.34689070	0.35021824	0.75981590	0.01373747
7	0.86211020	1.01674933	0.87373650	0.08979746
8	1.03118540	0.82205630	0.21055390	0.73813115
9	12.74339800	0.02251542	1.47164790	0.07333057
10	1.96347260	1.01602885	0.52020820	1.38869338

Table A2. Sample Size = 20

S/N	Sample 5(λ= 0.5)	Sample 6(λ= 1.0)	Sample 7(λ= 1.5)	Sample 8(λ= 2.0)
1	2.62129072	0.13858938	0.61025422	0.20816883
2	3.66009415	0.23677409	1.39562321	0.10508682
3	0.28671927	0.65882621	1.64585069	0.16343073
4	0.53614812	2.03913598	1.11779841	4.47429656
5	1.67262059	3.07310425	0.64918675	0.31910786
6	0.08521250	0.85047816	0.05778008	0.21707275
7	0.84341716	0.93587816	0.04898416	0.04360368
8	1.87988871	3.97283307	0.12098188	0.11505109
9	3.13213741	1.91406287	0.55509004	0.24578897
10	6.82508233	1.04861761	0.19238185	1.05875641
11	0.02900937	0.03350444	0.26953211	0.00355591
12	1.74900498	1.30175357	0.01133644	0.05922794
13	1.24431388	1.52538027	0.28231025	0.20815442
14	1.12438751	0.06504098	2.50193518	1.48214568
15	1.02901637	1.90192968	0.52524965	0.89694421
16	2.19818977	0.71117769	0.20890800	1.53327300
17	5.75705560	2.66119925	0.12742349	1.11055824
18	0.67605113	1.38839741	0.09933566	0.81368075
19	0.79778248	0.30239189	0.08220231	0.46919296
20	1.06989471	0.00054309	0.22447181	1.03504680

Table A3. Sample Size = 30

S/N	Sample 9(λ= 0.5)	Sample 10(λ= 1.0)	Sample 11(λ= 1.5)	Sample 12(λ= 2.0)
1	1.09871091	0.17085314	0.24591320	0.03006484
2	1.30597223	0.31266839	0.78554402	0.01316790
3	0.43742079	0.21960578	0.26193004	0.08580695
4	0.57034193	1.09603105	0.07651001	0.06684230
5	1.02475744	0.54796524	0.29620540	0.75349995
6	1.53172531	0.13561858	0.58560088	0.56383185
7	0.57173448	0.46805576	0.56401687	0.10014385
8	3.18835121	0.58675535	0.41054238	0.39468972
9	1.01423546	0.41116558	1.42498655	0.26076246
10	0.05104055	3.42353290	0.03212381	0.38563026
11	2.99399940	1.09817464	0.19768464	0.96558979
12	2.95112802	0.35962685	0.24295821	0.27934118
13	4.74244122	1.98862082	0.16056365	0.01547041
14	0.04628853	1.11964839	0.24140637	0.34343548
15	5.35809191	0.48539163	0.06956623	0.80092480
16	4.26185504	1.09365222	1.49188159	1.49780414
17	0.04367701	0.73949713	0.22345808	0.90349970
18	0.05851474	0.34345758	1.56937093	0.18701462
19	0.46455153	0.29243694	0.59195204	0.00288255
20	0.09371455	0.25904558	0.42176436	0.74330912
21	2.96970220	1.79180600	1.11983745	0.25455636
22	0.54133977	0.01809066	0.28416911	0.81341579
23	1.29204462	2.94010580	0.65638599	0.10230827
24	0.02923826	0.79327459	0.12515747	2.24807427
25	2.28724168	1.13577062	2.26479793	0.08718112
26	3.62406597	0.05051664	1.73361763	1.15865739
27	0.55831489	0.53652729	0.36225826	0.02484439
28	4.24494133	1.08361205	1.41035274	0.48756216
29	0.13192788	0.25488727	0.93471466	0.25596331
30	0.39449119	1.92176066	0.27655395	0.09251206