SMALL AND LARGE SAMPLE PERFORMANCE OF KAPLAN-MEIER AND SHRUNKEN KAPLAN-MEIER SURVIVAL FUNCTIONS

QAMRUZ ZAMAN*, YOUSAF HAYAT**, S.M. SUHAIL***,
MUNIR KHAN**** and SYED WAQAR SHAH*****

* *Department of Statistics, University of Peshawar, Peshawar – Pakistan.*
** *Department of Mathematics, Statistics and Computer Science, NWFP Agricultural University, Peshawar – Pakistan.*
*** *Department of Animal Breeding and Genetics, NWFP Agricultural University, Peshawar – Pakistan.*
**** *Department of Agricultural Economics, NWFP Agricultural University, Peshawar – Pakistan.*
***** *NWFP University of Engineering and Technology, Peshawar – Pakistan.*

ABSTRACT

Simulations were conducted to compare the Kaplan-Meier survival function with the Shrunken Kaplan-Meier survival function. Using the mean square error and pitman closeness criterion, the Shrunken Kaplan-Meier estimator is seem to perform better as compared to the Kaplan-Meier estimator especially for Weibull distribution and for a variety of censoring percentages. In case of exponential and log-logistic survival distributions, the results are less clear and not up to the mark. In addition, two variance estimators of Shrunken survival function are proposed. Simulation results show that Shrunken Kaplan-Meier survival function considerably perform better for the small samples.

***Key Words:** Kaplan-Meier survival function, Shrunken Kaplan-Meier survival function, Weibull distribution, Exponential distribution, Log-logistic distribution, Pitman closeness criterion*

Citation: Zaman, Q., Y. Hayat, S.M. Suhail, M. Khan and S.W. Shah. 2009. Small and large sample performance of Kaplan-Meier and Shrunken Kaplan-Meier survival functions. Sarhad J. Agric. 25(4): 671-680.

INTRODUCTION

Survival analysis deals with deaths in biological organism and failure in mechanical systems. It is also called the reliability analysis in engineering and duration modeling in economics or sociology (http://en.wikipedia.org/wiki/Survival_analysis). Generally, it involves the modeling of time to event data and is applicable in almost all the research disciplines like agriculture, plant and animal breeding experiments, medical sciences, statistical genetics, and crime analysis etc. for studying the time to failure of different organism and/or the reliability of different systems. In the present study, small and large sample performance of Kaplan-Meier and Shrunken Kaplan-Meier survival functions has been investigated; in addition, two new survival functions in the framework of Kaplan-Meier and Shrunken Kaplan-Meier survival functions are proposed. All these methods are defined in the following sections.

MATERIALS AND METHODS

Kaplan-Meier and Shrunken Kaplan-Meier Survival Functions

Let $x_1, x_2, \ldots, x_n$ be a sample of independent survival times with distribution function F(x). Let the censoring times $c_1, c_2, \ldots, c_n$ be independently distributed according to G(c). The survival time's x_i and censoring time's c_i are assumed to be independent.
Let $T_i = \min(x_i, c_i)$ and $\Delta_i = I(x_i \le c_i) = 1$ if $x_i \le c_i$, and 0 otherwise.
The Kaplan-Meier (1958) product limit estimator of S(x) = 1- F(x) is defined by

$$\hat{S}_{KM}(x) = \prod_{t_i \le x} (n_i - e_i) / n_i \qquad (1)$$

where n_i is the number of individuals who are alive just before time t_i and e_i denotes the number of events at that time.
The methods of Greenwood (1926) and Peto *et al* (1977) are most commonly used for estimating the variance of the Kaplan-Meier survival function, which are reproduced in equation (2) and (3), respectively;

$$V_G(\hat{S}_{KM}(x)) = \hat{S}_{KM}^{2}(x)\sum_{t_i \le x}(e_i / n_i(n_i - e_i)) \quad (2)$$

and

$$V_P(\hat{S}_{KM}(x)) = \hat{S}_{KM}^{2}(x)(1-\hat{S}_{KM}(x)) / n_i \quad (3)$$

In case of moderate and heavy censoring, both of these estimators (as mentioned in equation (2 and 3) under estimate the true variance (Borkowf, 2005). To overcome this, Borkowf (2005) proposed a survival function under the framework of the Kaplan-Meier survival function and is called Shrunken Kaplan-Meier survival function. The Shrunken Kaplan-Meier survival function having *n* number of cases in the study is defined by the expression (4).

$$\hat{S}_{KM}^{*}(x) = (n-1)/n\ \hat{S}_{KM}(x) + 1/2n \quad (4)$$

For the problem of underestimation of variance, Borkowf introduced two new estimators; one is based on the Kaplan-Meier survival function and is defined as

$$V(\hat{S}_{KMH}(x)) = \hat{S}_{KM}(x)(1-\hat{S}_{KM}(x)) / (n - cum_c) \quad (5)$$

where, cum_c is cumulative censoring.

The other estimator, which he named an adjusted hybrid variance estimator, is defined as;

$$V(\hat{S}_{KMH}^{*}(x)) = \hat{S}_{KM}^{*}(x)(1-\hat{S}_{KMH}^{*}(x)) / (n - cum_c) \quad (6)$$

Borkowf (2005) proved that these estimators performed better as compared to the Greenwood and Peto's estimators. Borkowf in his study analysed only the variance estimators, while in this study, first we compare the two survival functions theoretically and then a simulation analysis is performed to make the decision on the basis of familiar mean square error and Pitman Closeness Criterion (Keating *et al.* 1993). We also propose two new variance estimators utilizing the method of Shrunken Kaplan-Meier survival functions. These estimators are compared with other variance estimators through extensive simulations by different sample scenarios.

Comparison of Kaplan-Meier and Shrunken Kaplan-Meier Survival Functions

By definition, Kaplan-Meier survival function is

$$\hat{S}_{KM}(x) = \prod_{t_i \le x}(n_i - e_i)/n_i$$

and Shrunken Kaplan-Meier function can be expressed as

$$\hat{S}_{KM}^{*}(x) = (n-1)\hat{S}_{KM}(x)/n + 1/2n$$

We mainly use mostly the same notations as mentioned by Borkowf.

We start from the zero time, since at this point no event has occurred, so $e_i = 0$ and

$$\hat{S}_{KM}(0) = p_0 = n/n = n_i/n_i = 1,$$

and

$$\hat{S}_{KM}^{*}(0) = (n-1)/n + 1/2n = (2n-1)/2n, \text{ where } (2n-1)/2n < 1$$

therefore, $\hat{S}_{KM}^{*}(0) < \hat{S}_{KM}(0)$ (7)

So, the starting value of Shrunken survival function is always less than 1.

At the first observed time, there will be two possibilities, event or censoring. If the first observed time is censored then

$$\hat{S}_{KM}(1) = \hat{S}_{KM}(0) * (n_1 - 0)/n_1 = p_0 * p_1 = 1 * n_1/n_1 = 1 \quad (8)$$

In case of no event at the first observed time

$$\hat{S}_{KM}(1) = \hat{S}_{KM}(0) = 1 \tag{9}$$

Similarly,

$\hat{S}_{KM}^{*}(1) = (n-1)/n + 1/2n = \hat{S}_{KM}^{*}(0)$, so using equation (7) it is concluded that

$$\hat{S}_{KM}^{*}(1) < \hat{S}_{KM}(1) \tag{10}$$

If the first observed time is event, then

$$\hat{S}_{KM}(1) = \hat{S}_{KM}(0) * (n_1 - e_1)/n_1 = p_0 * p_1 = 1 * p_1 < 1$$

As $p_1 = (n_1 - e_1)/n_1 < 1$, therefore, $\hat{S}_{KM}(1) < \hat{S}_{KM}(0)$ (11)

In the Shrunken method, $1/2n$ is a constant factor, so the probability from one time to another time depends on $(n-1)/n\hat{S}_{KM}(x)$. From equation (11), it can be written as;

$(n-1)/n\hat{S}_{KM}(1) < (n-1)/n\hat{S}_{KM}(0)$, therefore, $(n-1)/n\hat{S}_{KM}(1) + 1/2n < (n-1)/n\hat{S}_{KM}(0) + 1/2n$

$$\Rightarrow \hat{S}_{KM}^{*}(1) < \hat{S}_{KM}^{*}(0) \tag{12}$$

The Kaplan-Meier probability at time t_1 is greater than the Shrunken Kaplan-Meier probability, if

$$\hat{S}_{KM}^{*}(1) < \hat{S}_{KM}(1);\ (n-1)/n\,\hat{S}_{KM}(1) + 1/2n < \hat{S}_{KM}(1),\ \text{if}\ (n-1)/n + n_1/2n(n_1 - e_1) < 1$$

Similarly, $\hat{S}_{KM}^{*}(2) < \hat{S}_{KM}(2)$, if $(n-1)/n + (n_1 n_2)/2n(n_1 - e_1)(n_2 - e_2) < 1$

In general, $\hat{S}_{KM}^{*}(x) < \hat{S}_{KM}(x)$, if $(n-1)/n + \prod_{i=1}^{x} n_i / 2n(n_i - e_i) < 1$

Variance Estimators

We develop a new variance estimator of Shrunken Kaplan-Meier survival function by using the variance property on equation (4), we get

$$V\left(\hat{S}_{KM}^{*}(x)\right) = \left((n-1)/n\right)^2 V\left(\hat{S}_{KM}(x)\right) \tag{13}$$

If we use the Greenwood variance formula, then the new estimator is

$$V\left(\hat{S}_{KM}^{*}(x)\right) = \left((n-1)/n\right)^2 \hat{S}_{KM}^{2}(x)\left(\sum_{t_i \le x} e_i / n_i (n_i - e_i)\right) \tag{14}$$

In case of Peto's variance *i.e*.

$$V\left(\hat{S}_{KM}^{*}(x)\right) = \left((n-1)/n\right)^2 \hat{S}_{KM}^{2}(x)\left(1 - \hat{S}_{KM}(x)\right)/n_i \tag{15}$$

For large "n" these two variances reduce to the Greenwood's and Peto's variance respectively.

Comparison of the two New Variances with Greenwood's and Peto's Variance

$$V\left(\hat{S}_{KM}^{*}(x)\right) = \left((n-1)/n\right)^2 \hat{S}_{KM}^{2}(x)\left(\sum_{t_i \le x} e_i / n_i (n_i - e_i)\right)$$

$$V_G\left(\hat{S}_{KM}^{\;*}(x)\right) = \left((n-1)/n\right)^2 V_G\left(\hat{S}_{KM}(x)\right) \quad (16)$$

Since, $(n-1)/n < 1$, so $V_G\left(\hat{S}_{KM}^{\;*}(x)\right) < V_G\left(\hat{S}_{KM}(x)\right)$ (17)

Variance is smaller than the Greenwood's variance.
The new variance based on Peto's formula,

$$V_P\left(\hat{S}_{KM}^{\;*}(x)\right) = ((n-1)/n)^2 \hat{S}_{KM}^2(x)\left(1-\hat{S}_{KM}(x)\right)/n_i = ((n-1)/n)^2 V_P\left(\hat{S}_{KM}(x)\right) \quad (18)$$

This shows that

$$V_P\left(\hat{S}_{KM}^{\;*}(x)\right) < V_P\left(\hat{S}_{KM}(x)\right) \quad (19)$$

If the data is free from censoring, the Greenwood's variance is reduced to a binomial variance, i.e. $V_G\left(\hat{S}_{KM}(x)\right) = \hat{S}_{KM}(x)\left(1-\hat{S}_{KM}(x)\right)/n_1 = V_B\left(\hat{S}_{KM}(x)\right)$

In that case, $V_G\left(\hat{S}_{KM}^{\;*}(x)\right) = ((n-1)/n)^2 V_B\left(\hat{S}_{KM}(x)\right)$ which shows that the proposed variance is less than the binomial variance i.e.

$$V_G\left(\hat{S}_{KM}^{\;*}(x)\right) < V_B\left(\hat{S}_{KM}(x)\right) \quad (20)$$

Comparison of Proposed Variances with the Variances Proposed by Borkowf

As Borkowf showed that

$$V\left(\hat{S}_{KMH}(x)\right) > V_G\left(\hat{S}_{KM}(x)\right) \quad (21)$$

and $V\left(\hat{S}_{KMH}(x)\right) > V_P\left(\hat{S}_{KM}(x)\right)$ (22)

Similarly, $V\left(\hat{S}_{KMH}^{\;*}(x)\right) > V_G\left(\hat{S}_{KM}(x)\right)$ (23)

$$\Rightarrow V\left(\hat{S}_{KMH}^{\;*}(x)\right) > V_P\left(\hat{S}_{KM}(x)\right) \quad (24)$$

By comparing this, it is concluded that the proposed variances may always give the smaller values as compared to the aforementioned four existing methods. So, we do not consider it for the analysis. It is proved both by the simulation and by the practical application that the adjusted hybrid variance estimator is the best choice in case of moderate to heavy censoring. As the new variance estimator is less than in every situation, so we do not apply it on the real data set.

RESULTS AND DISCUSSION

Monte Carlo simulations were conducted to compare the efficiency of the aforementioned methodologies. A uniform density U(0, b) for the censoring distribution while adjusting parameter "b" to provide 15%, 30%, 45%, 60% and 75% censoring was used to achieve this objective. For survival distributions, the exponential, weibull and log-logistic distributions were selected. For comparison, we decided to choose points of the form $F_x^{-1}(d_j)$ with fixed d_i's (d_1= 0.1, d_3 = 0.3, d_5 = 0.5, d_7 = 0.7, d_9 = 0.9). We generated 500 data sets of survival times with various sample sizes (n = 35, 70, 140, 280). To make the phenomena random, instead of using the subjective approach for selection of sample size, a different approach following three different steps was adopted. In the first step, a sample from 30 to 50 units (in order to select a small sample) was selected, replicated twice and then replicated the units obtained from the 2nd step to enter in the third step. At last, the units of 3rd steps were further replicated two times to achieve a larger sample size. By doing this, a sample of size 35 was randomly generated and following the steps a sample of size 280 was obtained. In addition, to compare the aforementioned methods the censoring times from the uniform distribution were generated for each of the survival time data sets. These data sets will built our confidence to draw conclusions about the choice of selecting better estimator for both the small and large data sets at uniform conditions.

With the usual logic of survival time and censoring time, we got the required times. For each data set, the mean square errors of the Kaplan-Meier as well as of the Shrunken Kaplan-Meier survival functions were computed. Similarly, probabilities were computed for the pitman closeness criterion. To check whether these two methods reach the same conclusion, we combined all these results (Table I). Here, the results for n = 140 are not mentioned as it almost yield the same results as n = 280, although these results are provided Figure 1 and 2.

Table I shows that the decision on basis of the ratio of mean square error and that of Pitman Closeness Criterion do agree at most of the points. Pitman closeness criterion is more useful in a situation, when the mean square errors of both estimators are equal to zero. In this case, if it is desired to choose the best estimator at that point, then the Pitman closeness criterion is helpful. In case of heavy censoring, which is usually the case in reality, and yield zeros mean square errors for the lower deciles. Table I indicates that Weibull survival distribution gives more satisfactory results for all samples sizes, while the other two distributions give mixed results for large sample sizes. In our simulation study, in all situations, Pitman's criteria give the decision in favour of new estimator.

Table I ***Ratio of the mean square errors of Kaplan-Meier and Shrunken Kaplan-Meier estimators, Pitman closeness criterion based on 500 samples of size n***

Dist. of survivor	Dist of Censoring	%	n	$d_{.1}$		$d_{.3}$		$d_{.5}$		$d_{.7}$		$d_{.9}$	
				R.MSE	Pcc	R.MSE	Pcc	R.MSE	Pcc	R.MSE	Pcc	R.MSE	Pcc
E(1)	U(0,6.5)	15	35	1.016	0.516	0.957	0.480	0.951	0.890	0.953	0.486	0.975	0.486
			70	1.006	0.480	0.981	0.494	0.972	0.912	0.973	0.498	1.000	0.490
			280	0.998	0.512	0.993	0.486	0.993	0.956	0.994	0.508	0.993	0.514
	U(0,3.2) \	30	35	1.111	0.584	0.951	0.504	0.944	0.882	0.954	0.510	1.027	0.440
			70	1.047	0.496	0.977	0.480	0.972	0.912	0.976	0.476	0.997	0.488
			280	1.005	0.456	0.997	0.470	0.997	0.968	0.995	0.492	1.000	0.492
	U(0, 1.8)	45	35	0/0	1.000	0.962	0.498	0.944	0.896	0.952	0.498	1.021	0.462
			70	0/0	1.000	0.974	0.506	0.971	0.942	0.976	0.478	0.995	0.490
			280	0/0	1.000	0.994	0.498	0.993	0.968	0.993	0.478	1.003	0.460
	U(0,1.1)	60	35	0/0	1.000	0/0	1.000	0.943	0.942	0.951	0.534	1.024	0.454
			70	0/0	1.000	0/0	1.000	0.972	0.942	0.976	0.498	0.999	0.496
			280	0/0	1.000	0/0	1.000	0.993	0.972	0.995	0.518	0.999	0.490
	U(0, 0.6)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.952	0.542	1.020	0.472
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.976	0.470	0.995	0.512
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.995	0.504	0.997	0.502
E(0.5)	U(0,12)	15	35	1.021	0.528	0.953	0.500	0.951	0.896	0.955	0.518	0.984	0.468
			70	1.007	0.498	0.976	0.496	0.971	0.912	0.978	0.494	0.999	0.454
			280	1.003	0.490	0.998	0.462	0.993	0.948	0.995	0.522	0.996	0.526
	U(0, 6)	30	35	1.096	0.610	0.951	0.504	0.944	0.886	0.952	0.510	1.027	0.432
			70	1.051	0.506	0.978	0.496	0.972	0.912	0.976	0.496	0.996	0.486
			280	1.001	0.472	0.998	0.478	0.993	0.964	0.995	0.502	0.998	0.504
	U(0, 3.5)	45	35	0/0	1.000	0.966	0.482	0.944	0.918	0.944	0.520	0.972	0.502
			70	0/0	1.000	0.972	0.498	0.972	0.924	0.975	0.514	0.984	0.502
			280	0/0	1.000	0.991	0.522	0.993	0.970	0.996	0.428	1.005	0.474
	U(0, 2.2)	60	35	0/0	1.000	0/0	1.000	0.944	0.936	0.944	0.534	0.996	0.486
			70	0/0	1.000	0/0	1.000	0.972	0.936	0.970	0.538	0.970	0.528
			280	0/0	1.000	0/0	1.000	0.993	0.978	0.995	0.508	0.993	0.520
	U(0, 1.3)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.955	0.530	1.018	0.462
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.976	0.478	0.997	0.500
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.994	0.512	0.997	0.502

E(1.5)	U(0, 4)	15	35	1.021	0.528	0.953	0.500	0.951	0.896	0.955	0.518	0.984	0.468
			70	1.007	0.498	0.976	0.496	0.971	0.912	0.978	0.494	0.999	0.454
			280	1.003	0.490	0.998	0.462	0.993	0.948	0.995	0.522	0.996	0.526
	U(0, 2.1)	30	35	1.109	0.594	0.950	0.500	0.945	0.890	0.953	0.516	1.026	0.438
			70	1.050	0.486	0.977	0.484	0.972	0.930	0.976	0.476	0.995	0.486
			280	1.004	0.466	0.998	0.472	0.993	0.978	0.995	0.498	0.999	0.500
	U(0,1.2)	45	35	0/0	1.000	0.966	0.484	0.944	0.912	0.945	0.510	0.971	0.512
			70	0/0	1.000	0.972	0.492	0.972	0.924	0.975	0.516	0.982	0.506
			280	0/0	1.000	0.990	0.534	0.993	0.966	0.996	0.438	1.006	0.476
	U(0, 0.8)	60	35	0/0	1.000	0/0	1.000	0.944	0.920	0.944	0.502	0.989	0.486
			70	0/0	1.000	0/0	1.000	0.972	0.952	0.969	0.540	0.973	0.528
			280	0/0	1.000	0/0	1.000	0.993	0.962	0.995	0.514	0.993	0.510
	U(0, 0.4)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.946	0.540	0.966	0.504
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.976	0.490	1.002	0.460
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.993	0.502	1.007	0.448
W(1, 0.5)	U(0, 3.2)	15	35	0/0	1.000	0.914	1.000	0.948	0.996	0.886	0.822	0.779	0.988
			70	0/0	1.000	0.947	1.000	0.970	0.996	0.921	0.910	0.862	1.000
			280	0/0	1.000	0.987	1.000	0.993	1.000	0.976	0.998	0.959	1.000
	U(0, 1.6)	30	35	0/0	1.000	1.000	0.996	0.940	0.982	0.878	0.836	0.761	0.986
			70	0/0	1.000	0.996	1.000	0.970	0.998	0.923	0.900	0.855	1.000
			280	0/0	1.000	0.988	1.000	0.993	1.000	0.976	0.998	0.959	1.000
	U(0, 0.9)	45	35	0/0	1.000	0/0	1.000	0.940	0.980	0.876	0.802	0.747	0.988
			70	0/0	1.000	0/0	1.000	0.971	0.996	0.923	0.904	0.855	1.000
			280	0/0	1.000	0/0	1.000	0.993	0.998	0.976	0.996	0.959	1.000
	U(0, 0.55)	60	35	0/0	1.000	0/0	1.000	0.978	0.992	0.876	0.824	0.747	0.992
			70	0/0	1.000	0/0	1.000	0.977	0.992	0.923	0.910	0.853	0.998
			280	0/0	1.000	0/0	1.000	0.993	1.000	0.976	0.994	0.958	1.000
	U(0, 0.33)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.883	0.788	0.743	0.988
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.927	0.882	0.850	1.000
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.976	0.994	0.958	1.000
W(1, 1.5)	U(0,10)	15	35	1.038	0.014	1.004	0.270	0.962	0.902	0.914	0.648	1.133	0.230
			70	1.023	0.000	1.004	0.222	0.976	0.960	0.938	0.674	1.118	0.206
			280	1.006	0.000	1.001	0.158	0.993	1.000	0.970	0.732	1.052	0.022
	U(0, 4.7)	30	35	1.038	0.024	1.001	0.308	0.956	0.900	0.910	0.624	1.114	0.254
			70	1.024	0.000	1.004	0.234	0.976	0.952	0.945	0.628	1.117	0.148
			280	1.006	0.000	1.001	0.150	0.993	1.000	0.969	0.730	1.051	0.012
	U(0, 2.8)	45	35	1.030	0.072	0.999	0.358	0.955	0.888	0.905	0.650	1.109	0.306
			70	1.020	0.014	1.002	0.270	0.975	0.958	0.945	0.650	1.114	0.166
			280	1.006	0.000	1.001	0.150	0.993	1.000	0.973	0.686	1.050	0.032
	U(0, 1.7)	60	35	0/0	1.000	0.994	0.432	0.953	0.928	0.912	0.608	1.130	0.286
			70	0/0	1.000	1.001	0.308	0.975	0.962	0.951	0.586	1.114	0.158
			280	0/0	1.000	1.001	0.186	0.993	0.998	0.972	0.756	1.051	0.038
	U(0, 0.9)	75	35	0/0	1.000	0/0	1.000	0.966	0.960	0.919	0.602	1.102	0.294
			70	0/0	1.000	0/0	1.000	0.976	0.968	0.957	0.576	1.093	0.224
			280	0/0	1.000	0/0	1.000	0.993	0.996	0.977	0.688	1.047	0.032
	U(0, 18)	15	35	1.087	0.542	0.940	0.524	0.944	0.900	0.950	0.498	0.979	0.496
			70	1.031	0.488	0.980	0.478	0.972	0.918	0.979	0.464	0.996	0.494
			280	1.010	0.462	0.994	0.488	0.993	0.956	0.995	0.508	0.995	0.512

Logl (1, 1)	U(0, 6.3)	30	35	0/0	1.000	0.956	0.494	0.944	0.888	0.952	0.504	1.011	0.448
			70	0/0	1.000	0.971	0.518	0.972	0.920	0.976	0.492	1.005	0.484
			280	0/0	1.000	0.994	0.480	0.993	0.970	0.994	0.514	0.993	0.512
	U(0,3)	45	35	0/0	1.000	0.982	0.510	0.943	0.910	0.948	0.520	0.995	0.482
			70	0/0	1.000	0.977	0.492	0.972	0.936	0.975	0.488	1.000	0.478
			280	0/0	1.000	0.993	0.504	0.993	0.966	0.994	0.486	0.996	0.528
	U(0, 1.5)	60	35	0/0	1.000	0/0	1.000	0.946	0.954	0.950	0.524	0.982	0.472
			70	0/0	1.000	0/0	1.000	0.972	0.976	0.976	0.494	0.992	0.494
			280	0/0	1.000	0/0	1.000	0.993	0.974	0.994	0.492	0.996	0.504
	U(0, 0.8)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.954	0.518	1.016	0.492
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.976	0.474	1.007	0.462
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.993	0.524	0.994	0.508
	U(0. 70)	15	35	0/0	1.000	0.948	0.518	0.950	0.924	0.958	0.492	0.997	0.474
			70	0/0	1.000	0.972	0.502	0.972	0.908	0.971	0.498	1.005	0.458
			280	0/0	1.000	0.997	0.460	0.993	0.952	0.995	0.484	1.005	0.470
Logl(1, 0.5)	U(0,16)	30	35	0/0	1.000	0.952	0.500	0.945	0.900	0.939	0.524	0.937	0.498
			70	0/0	1.000	0.981	0.474	0.972	0.932	0.967	0.522	0.979	0.518
			280	0/0	1.000	0.993	0.514	0.993	0.952	0.993	0.492	1.000	0.506
	U(0,4)	45	35	0/0	1.000	0/0	1.000	0.944	0.916	0.948	0.518	0.998	0.476
			70	0/0	1.000	0/0	1.000	0.972	0.946	0.973	0.474	0.990	0.514
			280	0/0	1.000	0/0	1.000	0.993	0.966	0.992	0.532	0.990	0.498
	U(0, 1.2)	60	35	0/0	1.000	0/0	1.000	0.966	0.982	0.960	0.482	1.011	0.434
			70	0/0	1.000	0/0	1.000	0.976	0.976	0.973	0.490	0.985	0.470
			280	0/0	1.000	0/0	1.000	0.993	0.990	0.997	0.444	1.002	0.486
	U(0, 0.3)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.946	0.532	0.987	0.472
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.967	0.550	0.983	0.494
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.992	0.502	0.991	0.518
	U(0, 12.5)	15	35	1.056	0.484	0.943	0.522	0.943	0.900	0.950	0.500	0.997	0.446
			70	0.977	0.524	0.973	0.508	0.972	0.918	0.969	0.544	0.992	0.494
			280	1.003	0.480	1.000	0.472	0.993	0.954	0.993	0.510	0.993	0.496
Logl(1, 1.5)	U(0, 5)	30	35	1.104	0.724	0.955	0.502	0.942	0.886	0.940	0.486	0.988	0.468
			70	1.070	0.628	0.968	0.500	0.971	0.926	0.968	0.492	0.988	0.456
			280	1.005	0.502	0.995	0.468	0.993	0.964	0.994	0.532	0.997	0.508
	U(0, 2.9)	45	35	0/0	1.000	0.964	0.496	0.944	0.900	0.949	0.548	0.981	0.490
			70	0/0	1.000	0.979	0.452	0.972	0.920	0.980	0.510	0.965	0.530
			280	0/0	1.000	0.991	0.512	0.993	0.950	0.995	0.502	1.000	0.490
	U(0, 1.7)	60	35	0/0	1.000	0/0	1.000	0.943	0.928	0.949	0.532	0.967	0.514
			70	0/0	1.000	0/0	1.000	0.971	0.930	0.980	0.486	1.009	0.466
			280	0/0	1.000	0/0	1.000	0.993	0.958	0.995	0.496	0.999	0.470
	U(0, 1)	75	35	0/0	1.000	0/0	1.000	0/0	1.000	0.946	0.570	0.987	0.482
			70	0/0	1.000	0/0	1.000	0/0	1.000	0.972	0.512	1.001	0.472
			280	0/0	1.000	0/0	1.000	0/0	1.000	0.993	0.482	1.002	0.476

R.MSE = Ratio of mean square errors; Pcc= Pitman Closeness Criterion; d_1, d_3, d_5, d_7 and d_9 are the deciles

Figure 1 and 2 represent the Pitman closeness criterion and mean square error curves of the two estimators, at d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8 and d_9, for the case where the samples of size n= 35, 70, 140 and 280 respectively, come from W(1, 0.5) and censoring is uniform. By comparing these results (Fig 1 and 2) , it is evident that the mean square error gives a clearer picture and reveals that the proposed method is better than the traditional method for small sample size couple with moderate to heavy censoring. While in case of large sample size, it behaves like the Kaplan-Meier survival function.

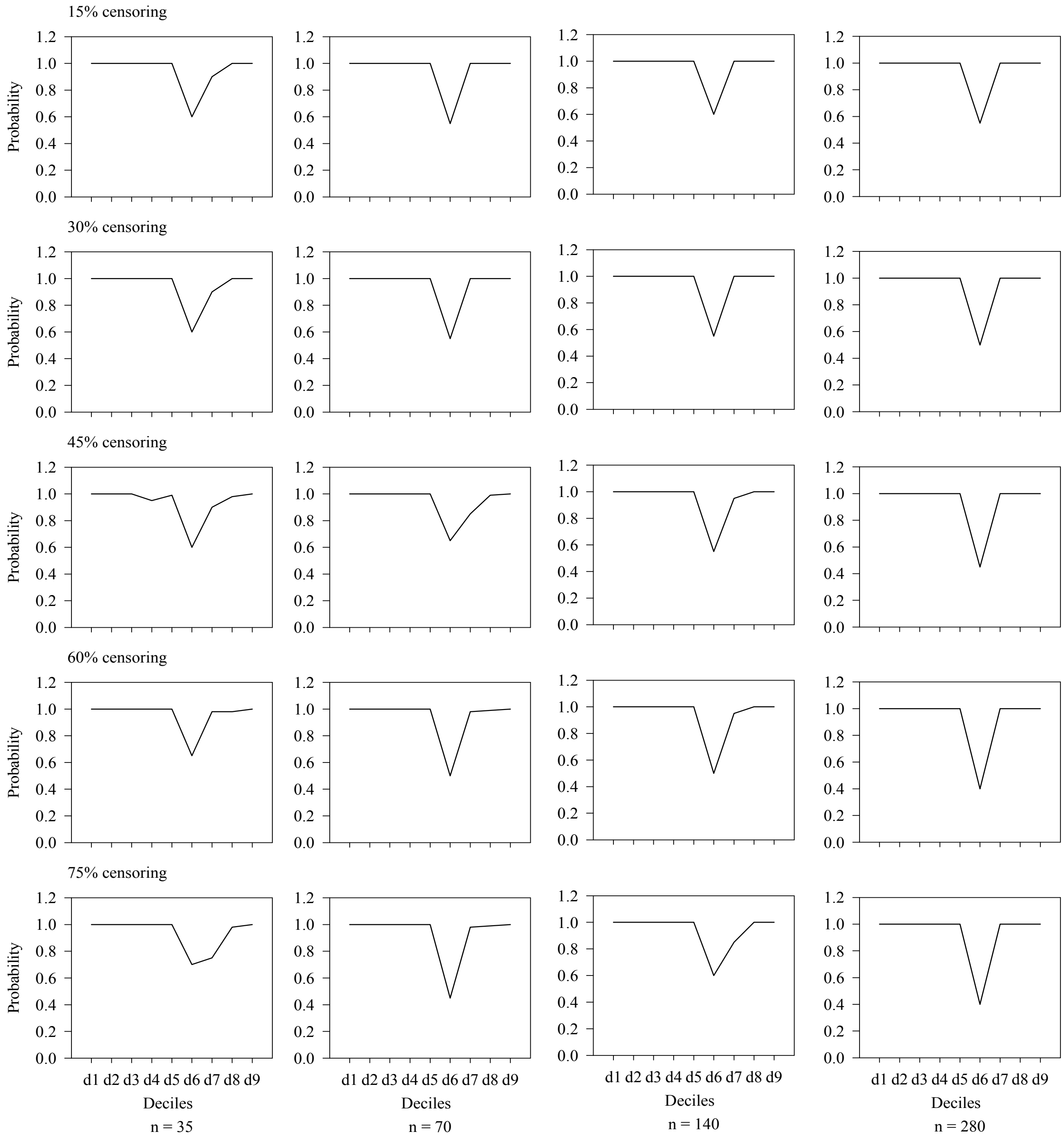

Fig. 1 Curves of Pitman closeness criterion from Weibull survival distribution and uniform censoring distribution, with censoring percentages and sample size n

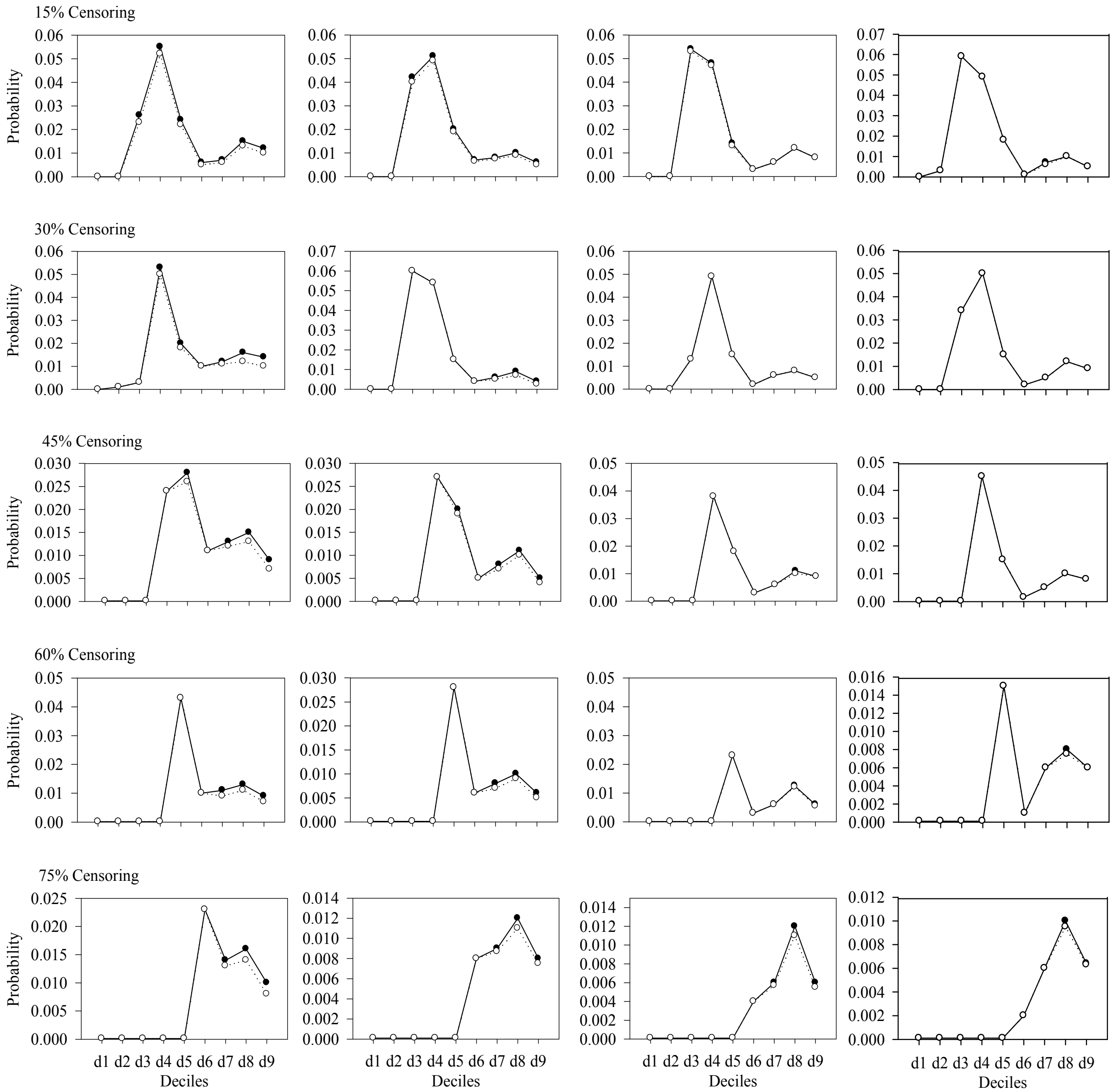

Fig. 2. Solid curve represents the Mean Square errors of Kaplan-Meier survival function and dotted curve represents the Mean Square errors of the Shrunken Kaplan-Meier survival function

CONCLUSION AND RECOMMENDATION

Our results extend the results of earlier study. Our simulations were considerably more ambitious. We used four sample sizes, three survival distributions and four levels of censoring. Our results demonstrated that Shrunken Kaplan-Meier estimators generally performed better for small samples. If the data follow the Weibull distribution, then the results are more in favour of the Shrunken Kaplan-Meier survival function, but it gives mixed results if we use the exponential or log-logistic distribution, especially for large sample size, which may be due to the fact that the Shrunken Kaplan-Meier survival function is more suitable for the small samples.

REFERENCES

Borkowf, C. B. 2005. A simple hybrid variance estimator for the Kaplan-Meier survival function. Statist. in Medicine. 24: 827-851.

Greenwood, M. 1926. A report on the natural duration of cancer. In Reports on public Health and Medical Subjects. His Majesty's Stationery Office, London. 33, 1-26.

Kaplan, E.L. and P. Meier. 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53: 457-481.

Keating, J.P., R.L. Mason and P.K. Sen. 1993. Pitman's Measure of Closeness: A comparison of statistical estimators, SIAM, Philadelphia. In Chaturvedi, A. and Shalabh. 2004. Risk and Pitman closeness properties of feasible generalized double k-class estimators in linear regression models with non-spherical disturbances under balanced loss function. J. Multivariate Analysis. 90(2): 229-256.

Peto, R., M.C. Pike, P. Armitage, N.E. Breslow, D.R. Cox, S.V. Howard, N. Mantel, K. McPherson, J. Peto and P.G. Smith. 1977. Design and analysis of randomized clinical trials requiring prolonged observation of each patient - II. Analysis and examples. British J. Cancer. 35(1): 1-39.