Journal of Statistical Research *ISSN 0256 - 422 X*
2012, Vol. 46, No. 1, pp. 49-63
Bangladesh

A COMPARISON OF A NEW PIECEWISE EXPONENTIAL ESTIMATOR WITH EXISTING NONPARAMETRIC ESTIMATORS OF A SURVIVAL FUNCTION: A SIMULATION STUDY

GANESH B. MALLA
Department of Mathematics and Computer Science, Xavier University, Cincinnati, OH 45207 USA
Email: mallag@xavier.edu

SUMMARY

This paper discusses a new piecewise exponential estimator (NPEE) of a survival function (SF) for censored data, which is continuous on $[0, \infty)$. For comparison purposes, we consider the Kaplan-Meier estimator (KME) and the empirical Bayes type estimator (EBE) derived by Rai et al. (1980). The EBE estimate beyond the last observation is determined solely by the prior. The NPEE retains the spirit of the KME and provides an exponential tail with a hazard rate determined by a novel nonparametric consideration. The NPEE has been compared with the KME and EBE for small sample sizes by simulation. The simulation comparisons are by the measures of bias and three norms, $(L_1, L_2$, and $L_\infty)$, for three levels of censoring, $(15\%, 50\%, 75\%)$, and two sample sizes $(10$ and $30)$. Generally speaking, the NPEE, which is asymptotically equivalent to the KME (Malla and Mukerjee (2010)), seems to be better than the KME, especially when we have heavy censoring and/or small sample sizes, and is at least as good as the EBE.

Keywords and phrases: Survival function; Piecewise exponential estimator; Nelson estimator; Kaplan-Meier estimator; Empirical Bayes type estimator; Censored data.

AMS Classification:

1 Introduction

We compare the properties of three estimators of a continuous survival function (SF), S, under random right censoring,- a new piecewise exponential estimator (NPEE), the Kaplan-Meier estimator (KME), and an empirical Bayes type estimator (EBE). The idea of the piecewise exponential estimator (PEXE) of the SF was first proposed by Kitchin, Langberg, and Proschan (1983) using the total time on test (TTT) concept. It provides a continuous estimator of the SF, but it has had limited usage possibly because the estimator is undefined beyond the last observation even in the uncensored case and for some other reasons discussed in detail in Section 2.2.1. Malla and Mukerjee (2010) introduced the NPEE for censored data by proposing a completely new nonparametric approach for estimating the hazard rates. This estimator is rather interesting. First, it is continuous on all of $[0, \infty)$ while the others are not. Second, it retains the spirit of the KME and provides an entirely different exponential tail with a hazard rate determined by a novel nonparametric consideration (see Section 2.2.2 for detail), while the other estimators are usually undefined beyond the last observation. As discussed in Malla and Mukerjee (2010), it is also interesting to see that the Nelson (1969, 1972) estimator, also called the Nelson-Aalen estimator, of the SF, may be considered as the smallest step function majorizing the NPEE with jumps at the uncensored observations. If the survivor function under consideration is known to be continuous from prior knowledge, practitioners would prefer to estimate a SF by a continuous estimator like NPEE.

Kaplan and Meier (1958) introduced the nonparametric estimator KME, also called Product-Limit Estimator, of the SF when the data are censored. It enjoys good asymptotic properties and is used extensively. Susarla and van Ryzin (1976, 1978) derived a Bayesian nonparametric estimator EBE in the same setting by using a Dirichlet process prior under squared error loss and analyzed its asymptotic properties. Although the EBE uses more of the information contained in censored data than the KME, it possesses a great deal of difficultly to the users (i) for its computation and (ii) for a subjective decision making in the choice of the prior distribution. Moreover, the EBE estimate beyond the last observation depends only on the prior chosen. This estimator has also some similarities with NPEE. Rai et al. (1980) have compared an empirical Bayes type version of this estimator with the KME using various types of norms. It should be noted that neither of these two estimators are continuous.

Malla and Mukerjee (2010) have shown that the NPEE is asymptotically equivalent to the KME. The large sample properties of the KME have been studied by Breslow and Crowley (1974), Meier (1975), and Phadia and van Ryzin (1980). Susarla and van Ryzin (1976, 1978) have studied the large sample properties of EBE and shown that the pointwise relative asymptotic efficiency of the EBE to KME is unity. Hence, from a large sample theory point of view, the estimators NPEE, KME, and EBE are equivalent.

In this paper, we examine by simulation whether or not the new piecewise exponential estimator has any advantages over these existing nonparametric estimators for the finite sample sizes. In a similar study, Rai et al. (1980) has shown that the advantages of the EBE increases dramatically as the degree of censoring increases, but the relative advantage of the EBE to the KME decreases as the sample size increases. We compare these nonparametric estimators in four cases: when the hazard rate of the true distribution is (i) constant, (ii) monotonically increasing, (iii) monotonically decreasing, and (iv) non-monotone. These cases represent the types of the survival functions that practitioners usually use in applications. The simulation comparisons are made by the measures bias and three norms (L_1 norm, L_2 norm, and L_∞norm) for small sample sizes of 15 and 30 and for three amounts of censoring (15, 50 and 75 percent). Generally speaking, the NPEE uses more of the information from the censored data, seems to be better than KME, especially when we have heavy censoring and/or small sample sizes, and is at least as good as the EBE. In general terms, the advantages of the NPEE over the KME increases considerably, but over the EBE is minimal as the amount of censoring increases.

The estimators are introduced and briefly discussed in Section 2. The details of the simulations and resulting comparisons are given in Section 3. In Section 4, we discuss the results and include some concluding remarks.

2 The Estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample from S and let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a DF, C, with no mass at 0. We observe only $\{(T_i, \delta_i) : 1 \leq i \leq n\}$, $0 < T_1 \leq T_2 \leq \cdots \leq T_n$, where $T_i = \min\{X_i, Y_i\}$ and $\delta_i = I(X_i \leq Y_i)$, which is 1 if $X_i \leq Y_i$ and 0 otherwise. We assume that the m failures or uncensored observations are at $d_1 < d_2 < \cdots < d_m$ with $d_1 > 0$, and we let $d_0 = 0$ and $d_{m+1} = \infty$. Thus, there are no ties among the uncensored observations, which occur with probability 1. Tied censored observations are ordered arbitrarily among themselves. If there are ties between censored observations and an uncensored one, the uncensored one precedes the censored ones except for the last observation where the the censored observations precede the uncensored one.

2.1 The Kaplan-Meier Estimator

Under the assumption of no ties among the uncensored observations, the Kaplan-Meier (1958) estimator of S is given by

$$\hat{S}^{KME}(t) = \prod_{T_i \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}, \quad 0 \leq t \leq T_n; \tag{2.1}$$

for $t > T_n$, $\hat{S}^{KME}(t) = 0$ if $d_m = T_n$ and it is undefined if $d_m < T_n$. In the latter case, some nonparametric suggestions have been made for this ambiguity. Efron (1967) (among others) suggested the estimator, $\hat{S}_E^{KME}$, by setting the tail estimate to 0 on $t > T_n$, which is equivalent to assuming that the last observation is uncensored whether it is censored or not, i.e., $d_m = T_n$, while Gill (1980) suggested the defective estimator, $\hat{S}_G^{KME}$, by setting $\hat{S}_G^{KME}(t) = \hat{S}^{KME}(T_n)$ for $t > T_n$.

2.2 The Piecewise Exponential Estimators

We assume that $E(X_i) \equiv \mu < \infty$ and that $\hat{S}^{KME}(T_n) = 0$, which is equivalent to the assumption that the last observation is uncensored and $d_m = T_n$. Let

$$0 \equiv a_0 < a_1 \leq a_2 \leq \cdots \leq a_m \text{ be the jumps of } \hat{S}^{KME} \text{ in magnitude}$$

at $d_0, d_1, d_2, \ldots, d_m$, respectively, and let

$$A_0 = 0 \text{ and } A_k = \textstyle\sum_{i=1}^k a_i \text{ for } 1 \leq k \leq m.$$

2.2.1 The Piecewise Exponential Estimator (PEXE)

Kitchin, Langberg and Proschan (1983) proposed an estimator, called the piecewise exponential estimator (PEXE) using the total time on test (TTT) concept. Assume that F is strictly increasing with a finite mean, μ_F, and the observations are uncensored. The TTT transform of F is defined by

$$T_F^{-1}(t) = \int_0^{F^{-1}(t)} S(u)\, du, \quad 0 \leq t \leq 1,$$

where $F^{-1}(0) = 0$ and $F^{-1}(1)$ is the right endpoint of the support of F that could be ∞. Note that $T_F^{-1}(1) = \mu_F$. It can be seen that

$$\frac{d}{dt} T_F^{-1}(t)|_{t=F(z)} = \frac{1}{\lambda_F(z)}, \quad \text{where } \lambda_F \text{ is the hazard rate of } F. \tag{2.2}$$

An empirical estimate of $T_F^{-1}(t)$ is given by

$$\hat{T}_F^{-1}(t) = \frac{1}{n}\sum_{1}^{k-1}(n-i+1)(T_i - T_{i-1}) + \left[t - \frac{k-1}{n}\right](n-k+1)(T_k - T_{k-1}), \quad \frac{k-1}{n} \le t \le \frac{k}{n},$$

for $1 \le k \le n$, where $X_0 = 0$ and an empty sum is 0. From (2), a natural estimator of the hazard rate is given by

$$\hat{\lambda}_F(t) = \frac{1}{(n-k+1)(T_k - T_{k-1})} = \frac{1}{TTT_k^F}, \quad \frac{k-1}{n} \le t \le \frac{k}{n},\ 1 \le k \le n, \tag{2.3}$$

where TTT_k^F is the total time on test by the $n-k+1$ survivors after $(k-1)th$ failure in the interval $[T_{k-1}, T_k)$, $1 \le k \le n$. Using this, one can define the PEXE of F for the uncensored case by

$$\hat{S}^{PX}(t) = e^{-\int_0^t \hat{\lambda}_F(u)du} \text{ on } [0, T_n], \text{ and undefined on } (T_n, \infty).$$

From our independence assumption, the T_i's may be considered as a random sample from the DF, $H = 1 - \bar{H}$, where $\bar{H} = (1-F)(1-C)$. Since the T_i's are uncensored, all the results above for F hold for H if the symbols H and $\bar{H}$ are used to replace F and S, respectively. Using this as an analogy, Kitchin et al. (1983) define the PEXE of S in the censored case by setting

$$\hat{\lambda}_F(t) = \left[\sum_{d_{k-1} < T_i \le d_k} TTT_i^H\right]^{-1} \equiv \frac{1}{TTT_k^F}, \quad d_{k-1} \le t < d_k, \quad 1 \le k \le m, \tag{2.4}$$

$\hat{S}^{PX}(0) = \hat{S}^{PX}(d_0) = 1$, and

$$\hat{S}^{PX}(t) = \hat{S}^{PX}(d_{k-1})e^{-\int_{d_{k-1}}^t \hat{\lambda}_F(u)du}, \quad d_{k-1} \le t < d_k, \quad 1 \le k \le m. \tag{2.5}$$

For $t > d_m$, $\hat{S}^{PX}(t)$ is left undefined even in the uncensored case.

We believe that there are two serious deficiencies in PEXE. First of all, it does not utilize the KME that has many known optimality properties. Secondly, although it is natural to estimate $\hat{\lambda}_H$ by (3) in view of (2), since $\hat{\lambda}_H = \hat{\lambda}_F + \hat{\lambda}_C$ under independence of F and C, such a justification for $\hat{\lambda}_F$ in terms of TTT_k^Fs cannot be given using (2) or any modification of it. Thus, we believe that this estimator should not be used.

2.2.2 The New Piecewise Exponential Estimator (NPEE)

Note that $\hat{S}^{KME}(t) = 1 - A_{k-1}$ for $d_{k-1} \le t < d_k,\ 1 \le k \le m$. Our new piecewise exponential estimator, $\hat{S}^{NPEE}$, of S on $[0, d_m)$ is obtained by defining the naive hazard rate, $\hat{\lambda}$, by

$$\hat{\lambda}(t) = \frac{a_k/(d_k - d_{k-1})}{1 - A_{k-1}}, \quad d_{k-1} \le t < d_k, \quad 1 \le k \le m, \tag{2.6}$$

and then defining $\hat{S}^{NPEE}$ on $[0, d_m]$ using this hazard rate:

$$\begin{aligned} \hat{S}^{NPEE}(t) &= \hat{S}^{NPEE}(d_{k-1}) e^{-\int_{d_{k-1}}^{t} \frac{a_k}{(1-A_{k-1})(d_k - d_{k-1})} du} \\ &= e^{-\hat{\Lambda}_{k-1}} e^{-\frac{a_k (t-d_{k-1})}{(1-A_{k-1})(d_k-d_{k-1})}}, \quad d_{k-1} \le t < d_k, \quad 1 \le k \le m, \end{aligned} \tag{2.7}$$

where

$$\hat{\Lambda}_k = \sum_{i=1}^{k} \frac{a_i}{1 - A_{i-1}} = \int_0^{d_k} \hat{\lambda}(u) du, \quad 1 \le k \le m, \text{ and } \Lambda_0 \equiv 0. \tag{2.8}$$

Note that $\hat{S}^{NPEE}(d_m) > 0$. Malla and Mukerjee (2010) has showed that

$$\int_0^{d_m} \hat{S}^{NPEE}(t)\, dt < \hat{\mu}_F^{KME} \equiv \int_0^{d_m} \hat{S}^{KME}(t)\, dt, \text{ the estimator of } \mu \text{ using the KME}, \tag{2.9}$$

and has extended $\hat{S}^{NPEE}$ by adding a conditionally exponential tail on $[d_m, \infty)$ so that $\hat{\mu}_F^{NPEE} \equiv \int_0^\infty \hat{S}^{NPEE}(t)\, dt = \hat{\mu}^{KME}$.
Their estimate of S for $t \in [d_m, \infty)$ is given by $\hat{S}^{NPEE}(t) = e^{-\hat{\lambda}_m} e^{-\hat{\lambda}_{tail}(t-d_m)}$
where

$$\hat{\lambda}_{tail} = e^{-\hat{\lambda}_m} / \sum_{i=1}^{m} (I_k - J_k), \tag{2.10}$$

for $1 \le k \le m$,

$$\begin{aligned} I_k &= \int_{d_{k-1}}^{d_k} \hat{S}^{KME}(t)\, dt = (1 - A_{k-1})(d_k - d_{k-1}), \text{and} \\ J_k &= \int_{d_{k-1}}^{d_k} \hat{S}^{NPEE}(t)\, dt = e^{-\hat{\lambda}_{k-1}} \frac{(1 - A_{k-1})(d_k - d_{k-1})}{a_k} \left[1 - e^{-\frac{a_k}{1-A_{k-1}}} \right]. \end{aligned}$$

Join the separate exponential 'pieces' to form the estimator.

2.3 An Empirical Bayes type Estimator

Susarla and van Ryzin (1976, 1978) derived the following nonparametric Bayes estimate of S using a Dirichlet process prior under squared error loss and analyzed its asymptotic properties:

$$\hat{S}^{EBE}(t) = \frac{\alpha(t, \infty) + N(t)}{\alpha(0, \infty) + n} \prod_{T_i \le t} \left[\frac{\alpha(T_i, \infty) + (n - i + 1)}{\alpha(T_i, \infty) + (n - i)} \right]^{1-\delta_i}, \tag{2.11}$$

where α is a finite positive measure on $(0, \infty)$ and $N(t) = \#\{T_i > t\}$. The value of $\alpha(0, \infty)$ is chosen to reflect the strength of one's belief in the prior relative to n, and the estimator becomes the KME if $\alpha(0, \infty) = 0$. The case they analyzed specifically was when $\alpha(t, \infty)/\alpha(0, \infty) = e^{-\lambda_0 t}$ for some $\lambda_0 > 0$. For the simulation study, Rai et al. (1980) has chosen two cases: $\alpha(0, \infty) = \sqrt{n}$ and $\alpha(0, \infty) = n$, in which their empirical Bayes type estimator is mean-squared consistent, and inconsistent respectively. See Rai et al. (1980) for the detailed discussions of these choices and see the pioneering paper of Ferguson (1973) for the theory of Dirichlet processes and their applications in nonparametric statistical inference. They have compared their empirical Bayes type versions of the estimator with the KME by extensive simulations and showed uniform improvement over the KME by using three types of norms: L_1 norm, L_2 norm and L_∞ norm. For our study, we choose their mean-squared consistent empirical Bayes type estimator:

$$\hat{S}^{EBE}(t) = \frac{\sqrt{n}e^{-\hat{\lambda}t} + N(t)}{\sqrt{n}+n} \prod_{T_i \le t} \left[\frac{\sqrt{n}e^{-\hat{\lambda}T_i} + (n-i+1)}{\sqrt{n}e^{-\hat{\lambda}T_i} + (n-i)} \right]^{1-\delta_i}, \tag{2.12}$$

where $\hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{m} T_i}$ is the unique maximum likelihood estimator of λ for the exponential distribution, the prior guess, under censoring (see Bartholomew (1957)).

For a pedagogical purpose, Figure 1 shows a graphical comparison of these estimators for estimating the survival function of the exponential distribution with the hazard rate, $\lambda = 5$ by using a moderate sample size n = 20 and the exponential censoring 60%. It is well known that KME overestimates the survival function. The EBE coincides with the KME at each uncensored observation, goes down as an exponential between uncensored observations and then jumps down to the KME estimate. The NPEE does not jump, but gives a continuous piecewise exponential.

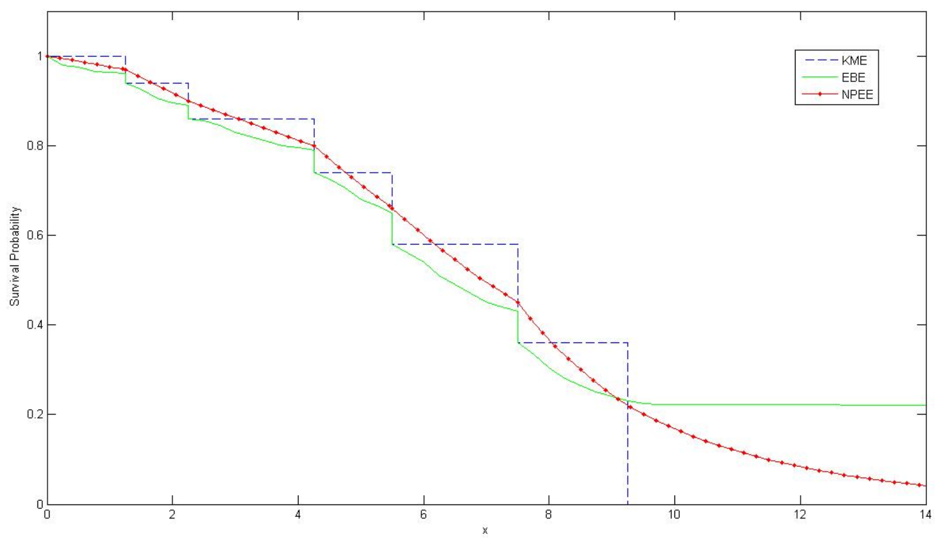
3 The Simulation Study

In this section, we present the results of a simulation study comparing three nonparametric estimators of the survival function S. The first estimator is the Kaplan-Meier estimator (KME) of (1). The second estimator is the new piecewise exponential estimator (NPEE) which is given by (6), (7) and (10). The final estimator is the empirical Bayes type estimator given by (12).

We choose the exponential distribution (density $f(t) = \lambda\, exp(-\lambda t),\ \lambda > 0, t \ge 0$) with parameter λ as a censoring distribution. By choosing the suitable value of λ, we can get different amounts of censoring in each simulation situation. The simulation model is that of random censoring: corresponding life and censoring random variables are independent and the censoring random variables are identically distributed.

In our simulation study we first use the Weibull distribution (density $f(t) = \alpha\, \lambda\, t^{\alpha-1}\, exp(-\lambda t^{\alpha})$, hazard rate $h(t) = \alpha\ \lambda\ t^{\alpha-1}$where $\alpha,\ \lambda > 0,\ t \ge 0$) with scale parameter λ and shape parameter α as our life distribution as it is flexible enough to accommodate constant ($\alpha = 1$), increasing ($\alpha > 1$), or decreasing ($\alpha < 1$) hazard rate situations of the real life distributions. We have chosen the values of the parameters as (i) $\alpha = 1$ and $\lambda = 1$, (ii) $\alpha = 2$ and $\lambda = 1$, and (iii) $\alpha = 0.5$ and $\lambda = 1$ to accommodate three possibilities of the hazard rate. The value of the scale parameter λ for the Weibull distribution is chosen 1, but can be chosen arbitrarily as monotonicity of the hazard rate is independent of the parameter λ. We note that case (i) also represents the exponential distribution

Figure 1: Graphs of the nonparametric estimators of a S.F.



and thus the simulations in Tables 1 and 2 are for the case in which Bayes prior guess of the EBE has the correct parametric form, while it has incorrect form for the simulations in Tables 3-6 and 9. The second set of simulations in Tables 7 and 8 compare the same estimators as before but for the underlying lognormal distribution (density $f(t) = exp[-\frac{1}{2}\,(\frac{\ln t-\mu}{\sigma})^2]/t(2\pi)^{1/2}\sigma$, hazard rate $h(t) = f(t)/(1 - \Phi[\frac{\ln t-\mu}{\sigma}])$ where $\sigma > 0,\ t \geq 0$) with parameters $\mu = 5$ and $\sigma = 1$. These simulation comparisons of the estimators are for the underlying distribution in which the hazard rate is non-monotone and again Bayes prior guess of the EBE has the incorrect parametric form. In fact, the hazard rate for a lognormal distribution begins at zero, rises to a maximum, then decreases very slowly to zero, irrespective of the choices of the parameters (Sweet(1990)).

These estimators are compared by simulation using bias of the estimators in Tables 2, 4, 6 and 8. The bias of the estimator $\hat{S}(t)$ of $S(t)$ is given by bias $= \hat{S}(t) - S(t)$, which is calculated at five quantiles: $q_{.1},\ q_{.2},\ q_{.5},\ q_{.8}$, and $q_{.9}$. In fact, we computed bias at 0.1(0.1)0.9 but the results were more or less the same at 0.3(0.1)0.7. So, we only included the results at $q_{.5}$ from the range in order to reduce the size of the Tables. To study the effect of the sample size and amount of censoring, comparisons are carried out for the sample sizes 15 and 30 and at three different amount of censoring: 15, 50 and 75 percent. Each simulation carried out was based on calculating the estimates of the underlying SF from 50,000 samples of the conditions of the table entry. Each entry for the bias gives the mean of these 50,000 biases calculated.

To study the effect of the sample sizes, we have presented the comparisons of the estimators for the sample sizes 10, 30, 60, 100 for three amount of censoring in Table 9. We also compare the estimators by using norms for the conditions of the table entry in tables 1, 3, 5 and 7. Rai et al. (1980) has used the norms comparison approach for the same set up as ours to compare their Bayesian type estimators of the survival function with the Kaplan-Meier estimator. The norms for comparison of the estimator $\hat{S}(t)$ of $S(t)$ are the L_1 norm, the L_2 norm, and the L_∞ norm:

$$L_1 \text{ norm} : \int_0^\infty |\hat{S}(t) - S(t)| \, dt$$
$$L_2 \text{ norm} : \sqrt{\int_0^\infty (\hat{S}(t) - S(t))^2 \, dt}$$
$$L_\infty \text{ norm:} \sup_t |\hat{S}(t) - S(t)|.$$

Columns 5 and 6 of Table 1 give the comparison of the estimators using the L_2 norm for the Weibull density above with $\alpha = 1,\ \lambda = 1$ for the sample sizes 10 and 30. The purpose of this simulation was to compare the estimation performance of the estimators as n increases and/or censoring percentage increases. The comparisons of the estimators due to all three norms in Tables 1, 3 and 5 are for the Weibull curves estimation and in Table 7 for the lognormal curve. The expected value of the L_2 norm is the integrated mean-squares error. The remaining set-up for these simulations is the same as described above for the bias calculation. Each simulation carried out in Tables 1, 3, 5 and 7 was based on calculating the mean norm of 50,000 samples with its standard error in parenthesis. The integrals for the L_1 and L_2 norms were calculated numerically using a grid of approximate length 0.01. The same grid was used for the L_∞ norm. For example, for the first entry in column three (0.280) for the estimator NPEE of Table 1, 50,000 samples each of size 10 were generated from the Weibull distribution with parameters $\alpha = 1$ and $\lambda = 1$ and an exponential distribution of a suitable parameter value was used as a censoring distribution to yield 15% censoring. For the ordered sample say $Z_{(1)}, Z_{(2)}, \ldots, Z_{(10)}$, $k \equiv 100(Z_{(10)} - Z_{(1)})$ partition points with an approximate 0.01 distance between two consecutive points between $Z_{(1)}$ and $Z_{(10)}$ were created. The L_1 norm was then approximated by $\sum_{j=1}^{k} |\hat{S}(j) - S(j)| \, \Delta_j$, where Δ_j is the distance between two consecutive partition points for the j^{th} interval and $\sum_{j=1}^{k}$ is the sum over all such intervals. This numerical integration was repeated for each of the 50,000 samples of size 10. The entry 0.280 is the mean estimate of the estimator NPEE for these 50,000 calculations and (.011) is the standard error of this mean.

The simulations have been done in MATLAB R2008b. The discussion of Tables 1-9 and conclusions are given in the Section 4.

4 Discussions and Conclusions

After a careful examination of the simulation comparisons by the bias and three norms (L_1, L_2, and L_∞) we come to the following conclusions:

(i) The continuous New piecewise exponential estimator (NPEE) which is defined on $[0, \infty)$ is at least as good as the other two estimators KME and EBE for every case simulated. When the underlying distribution is Weibull ($\alpha = 1,\ \lambda = 1$), i.e., exponential, our NPEE is better than both KME and EBE (Tables 1 and 2). We note that it is a case in which we expect the NPEE and EBE to be favored because NPEE is a piecewise exponential estimator and EBE's prior distribution is exponential. When underlying distributions are not exponential (in fact, they are Weibull ($\alpha = 2,\ \lambda = 1$), Weibull ($\alpha = 0.5,\ \lambda = 1$) or lognormal ($\mu = 5,\ \sigma = 1$)), NPEE appears at least as good as the EBE, but it is again better than the KME in terms of both the measures used. Despite the fact that EBE's prior is incorrect for these cases, EBE is the most improved estimator among the estimators considered. The dominance of the NPEE may be due to the fact that the survival functions of the underlying distributions considered are monotonically decreasing while these distributions represent various possible patterns (constant, increasing, decreasing, or increasing and decreasing) of the hazard rates of the life distributions.

(ii) All three estimators considered for this study show improvement (deterioration) with increase (decrease) in sample size (amount of censoring). The KME is the most improved (deteriorated) estimator with increase in the sample size (amount of censoring) (Tables 1, 3, 5, 7 and 9). In other words, the large sample advantages of the KME over other estimators may not show up, particularly with heavy censoring until fairly large samples are taken. The advantage of the NPEE over the EBE is minimal as the amount of censoring increases.

In Tables 2, 4, 6 and 8, a minus sign before the number indicates that the SF is underestimated by the estimator and no sign indicates that it is overestimated. As we can see the estimators NPEE and EBE are usually negatively biased while the estimator KME always shows a positive bias except for the cases in which the amount of censoring is large (75%) at the largest quantile considered. In these cases, NPEE and EBE do not show any regular pattern of bias while KME is even undefined for some cases. The undefined cases are indicated by the symbol 'NA' in the tables.

Based on these simulations, we feel that the NPEE is a predictor of the survival function that is at least as good as the EBE and better than the KME. The drawback of the NPEE is that it is more complicated to calculate than the KME, but with computer aid this should not be a problem. The complication of the choice of the priori distribution for EBE and its added complication to the calculation might make EBE a second choice over the NPEE. These simulation studies cannot prove the superiority of the NPEE over the EBE and KME as measured by bias and norms, but can make such a conclusion more plausible.

Finally, in closing, we remark that it will be interesting to see a comparison between the NPEE and EBE, EBE with a more flexible parametric family as its prior and $\alpha(0, \infty)$ as a function of the observed data. We also plan to show the weak convergence equivalence of the NPEE with the KME and use NPEE in place of the survival function in the definition of the mean residual life function to derive a continuous estimator of the MRL and study its properties.

Acknowledgements

The author gratefully acknowledges the guidance and suggestions of Dr. Hari Mukerjee and appreciates the help of Mr. Jack Butz for proofreading the manuscript. The referees' comments were valuable and have improved the paper considerably.

Table 1: Norms Comparison of the Estimators for Weibull $(\alpha = 1,\ \lambda = 1)$ Survival Curve for the Sample Sizes n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		L_1 Norm		L_2 Norm		L_∞ Norm	
Estimator	CP	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
NPEE	15	0.280(.011)	0.201(.008)	0.182(.007)	0.106(.004)	0.200(.005)	0.141(.004)
EBE		0.290(.012)	0.213(.008)	0.190(.007)	0.111(.005)	0.210(.006)	0.143(.004)
KME		0.350(.009)	0.235(.007)	0.220(.006)	0.127(.003)	0.270(.005)	0.150(.004)
NPEE	50	0.284(.012)	0.220(.008)	0.190(.008)	0.120(.005)	0.211(.007)	0.148(.005)
EBE		0.295(.013)	0.225(.009)	0.198(.008)	0.128(.004)	0.225(.008)	0.153(.005)
KME		0.361(.007)	0.271(.007)	0.275(.004)	0.206(.004)	0.342(.006)	0.247(.003)
NPEE	75	0.287(.011)	0.225(.008)	0.215(.010)	0.125(.008)	0.240(.012)	0.155(.007)
EBE		0.297(.011)	0.230(.008)	0.224(.011)	0.132(.008)	0.249(.013)	0.165(.008)
KME		0.444(.004)	0.396(.003)	0.414(.003)	0.380(.003)	0.607(.003)	0.580(.001)

Table 2: Bias Comparison of the Estimators at the Quantiles (q) for Weibull $(\alpha = 1,\ \lambda = 1)$ Survival Curve with the Sample Size n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		Bias(NPEE)		Bias(EBE)		Bias(KME)	
CP	q	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
15	0.1	-0.0036	-0.0028	-0.0045	-0.0037	0.0057	0.0041
	0.2	-0.0032	-0.0024	-0.0039	-0.0033	0.0056	0.0047
	0.5	-0.0030	-0.0023	-0.0036	-0.0028	0.0055	0.0042
	0.8	-0.0033	-0.0024	-0.0037	-0.0030	0.0054	0.0047
	0.9	-0.0036	-0.0033	-0.0042	-0.0037	0.0058	0.0041
50	0.1	-0.0038	-0.0036	-0.0047	-0.0034	0.0060	0.0050
	0.2	-0.0034	-0.0031	-0.0043	-0.0038	0.0059	0.0051
	0.5	-0.0035	-0.0028	-0.0038	-0.0035	0.0061	0.0047
	0.8	-0.0036	-0.0030	-0.0042	-0.0032	0.0056	0.0051
	0.9	-0.0039	-0.0034	-0.0044	-0.0040	0.0060	0.0054
75	0.1	-0.0043	-0.0040	-0.0047	-0.0043	0.0067	0.0052
	0.2	-0.0040	-0.0036	-0.0047	-0.0040	0.0064	0.0054
	0.5	-0.0043	-0.0035	-0.0047	-0.0042	0.0060	0.0053
	0.8	-0.0045	-0.0042	-0.0049	-0.0041	0.0060	0.0055
	0.9	-0.0040	-0.0040	-0.0045	-0.0040	0.0066	0.0053

Table 3: Norms Comparison of the Estimators for Weibull $(\alpha = 2,\ \lambda = 1)$ Survival Curve for the Sample Sizes n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		L_1 Norm		L_2 Norm		L_∞ Norm	
Estimator	CP	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
NPEE		0.411(.009)	0.390(.012)	0.342(.006)	0.338(.006)	0.335(.005)	0.329(.008)
EBE	15	0.410(.010)	0.388(.013)	0.345(.006)	0.341(.007)	0.338(.006)	0.328(.008)
KME		0.445(.007)	0.398(.010)	0.381(.004)	0.353(.004)	0.357(.005)	0.340(.007)
NPEE		0.417(.009)	0.395(.013)	0.360(.008)	0.351(.005)	0.340(.007)	0.335(.009)
EBE	50	0.416(.009)	0.397(.013)	0.366(.008)	0.348(.005)	0.342(.008)	0.340(.008)
KME		0.471(.007)	0.410(.006)	0.410(.004)	0.382(.005)	0.374(.006)	0.355(.005)
NPEE		0.421(.011)	0.395(.014)	0.371(.010)	0.355(.007)	0.345(.012)	0.337(.007)
EBE	75	0.420(.011)	0.388(.013)	0.377(.011)	0.358(.008)	0.344(.013)	0.340(.008)
KME		0.510(.005)	0.450(.006)	0.464(.003)	0.402(.004)	0.392(.003)	0.375(.003)

Table 4: Bias Comparison of the Estimators at the Quantiles (q) for Weibull $(\alpha = 2,\ \lambda = 1)$ Survival Curve with the Sample Size n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		Bias(NPEE)		Bias(EBE)		Bias(KME)	
CP	q	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
	0.1	-0.0056	-0.0055	-0.0058	-0.0056	0.0069	0.0060
	0.2	-0.0054	-0.0052	-0.0059	-0.0056	0.0067	0.0059
15	0.5	-0.0054	-0.0052	-0.0062	-0.0057	0.0067	0.0059
	0.8	-0.0055	-0.0051	-0.0061	-0.0059	0.0070	0.0064
	0.9	-0.0050	-0.0050	-0.0055	-0.0053	0.0071	0.0066
	0.1	-0.0058	-0.0055	-0.0060	-0.0057	0.0072	0.0065
	0.2	-0.0056	-0.0054	-0.0054	-0.0053	0.0070	0.0060
50	0.5	-0.0058	-0.0054	-0.0056	-0.0053	0.0069	0.0060
	0.8	-0.0057	-0.0054	-0.0058	-0.0055	0.0073	0.0066
	0.9	-0.0056	-0.0054	-0.0056	-0.0053	0.0067	0.0065
	0.1	-0.0063	-0.0058	-0.0064	-0.0062	0.0075	0.0068
	0.2	-0.0059	-0.0056	-0.0058	-0.0055	0.0075	0.0067
75	0.5	-0.0058	-0.0056	-0.0055	-0.0056	0.0074	0.0070
	0.8	-0.0058	-0.0055	-0.0059	-0.0056	0.0076	0.0070
	0.9	0.0028	0.0024	0.0030	0.0025	NA	NA

Table 5: Norms Comparison of the Estimators for Weibull $(\alpha = 0.5,\ \lambda = 1)$ Survival Curve for the Sample Sizes n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		L_1 Norm		L_2 Norm		L_∞ Norm	
Estimator	CP	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
NPEE		0.451(.009)	0.443(.012)	0.380(.007)	0.374(.009)	0.366(.005)	0.355(.008)
EBE	15	0.457(.009)	0.445(.011)	0.378(.007)	0.376(.009)	0.365(.006)	0.353(.008)
KME		0.480 (.007)	0.455(.009)	0.405(.004)	0.380(.007)	0.390(.005)	0.362(.007)
NPEE		0.458(.009)	0.450(.012)	0.383(.008)	0.377(.008)	0.372(.007)	0.366(.007)
EBE	50	0.460(.009)	0.453(.013)	0.385(.009)	0.376(.009)	0.373(.008)	0.363(.007)
KME		0.497(.007)	0.460(.008)	0.415(.004)	0.390(.005)	0.412(.006)	0.382(.005)
NPEE		0.473(.011)	0.388(.010)	0.386(.010)	0.381(.007)	0.375(.011)	0.373(.008)
EBE	75	0.474(.010)	0.391(.011)	0.387(.010)	0.378(.008)	0.374(.012)	0.376(.008)
KME		0.503(.004)	0.470(.005)	0.486(.003)	0.405(.003)	0.455(.004)	0.396(.004)

Table 6: Bias Comparison of the Estimators at the Quantiles (q) for Weibull $(\alpha = 0.5,\ \lambda = 1)$ Survival Curve with the Sample Size n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		Bias(NPEE)		Bias(EBE)		Bias(KME)	
CP	q	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
	0.1	-0.0066	-0.0065	-0.0064	-0.0063	0.0070	0.0063
	0.2	-0.0064	-0.0064	-0.0065	-0.0064	0.0072	0.0064
15	0.5	-0.0064	-0.0063	-0.0066	-0.0063	0.0072	0.0063
	0.8	-0.0065	-0.0062	-0.0065	-0.0063	0.0071	0.0062
	0.9	-0.0064	-0.0062	-0.0066	-0.0061	0.0072	0.0064
	0.1	-0.0068	-0.0066	-0.0067	-0.0064	0.0074	0.0065
	0.2	-0.0067	-0.0044	-0.0066	-0.0064	0.0075	0.0064
50	0.5	-0.0067	-0.0064	-0.0064	-0.0063	0.0075	0.0063
	0.8	-0.0067	-0.0065	-0.0065	-0.0064	0.0073	0.0065
	0.9	-0.0066	-0.0062	-0.00065	-0.0060	0.0074	0.0063
	0.1	-0.0070	-0.0068	-0.0068	-0.0068	0.0077	0.0066
	0.2	-0.0069	-0.0067	-0.0067	-0.0067	0.0077	0.0067
75	0.5	-0.0068	-0.0066	-0.0068	-0.0066	0.0076	0.0068
	0.8	-0.0067	-0.0065	-0.0069	-0.0065	0.0076	0.0070
	0.9	0.0031	0.0026	0.0032	0.0027	NA	NA

Table 7: Norms Comparison of the Estimators for Lognormal ($\mu = 5,\ \sigma = 1$) Survival Curve for the Sample Sizes n and Censoring Percentages CP (Based on $50,000$ Iterations*)

		L_1 Norm		L_2 Norm		L_∞ Norm	
Estimator	CP	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
NPEE		0.711(.015)	0.701(.014)	0.680(.017)	0.676(.013)	0.674(.012)	0.656(.014)
EBE	15	0.723(.018)	0.707(.015)	0.685(.016)	0.677(.013)	0.672(.012)	0.655(.012)
KME		0.876(.011)	0.788(.009)	0.706(.012)	0.687(.010)	0.750(.009)	0.707(.008)
NPEE		0.717(.014)	0.707(.012)	0.686(.012)	0.679(.010)	0.681(.007)	0.666(.009)
EBE	50	0.722(.014)	0.709(.013)	0.686(.012)	0.678(.011)	0.682(.008)	0.669(.009)
KME		0.910(.009)	0.845(.010)	0.755(.008)	0.699(.007)	0.790(.006)	0.730(.005)
NPEE		0.725(.011)	0.715(.010)	0.687(.010)	0.685(.007)	0.685(.011)	0.670(.007)
EBE	75	0.725(.010)	0.718(.011)	0.687(.010)	0.682(.008)	0.687(.012)	0.673(.008)
KME		0.966(.006)	0.903(.007)	0.794(.005)	0.735(.004)	0.830(.004)	0.780(.004)

Table 8: Bias Comparison of the Estimators at the Quantiles (q) for Lognormal ($\mu = 5,\ \sigma = 1$) Survival Curve with Sample Size n and Censoring Percentages CP (Based on $50,000$ Iterations*

		Bias(NPEE)		Bias(EBE)		Bias(KME)	
CP	q	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
	0.1	-0.0086	-0.0085	-0.0077	-0.0074	0.0075	0.0073
	0.2	-0.0085	-0.0084	-0.0078	-0.0074	0.0077	0.0075
15	0.5	-0.0085	-0.0084	-0.0075	-0.0073	0.0077	0.0074
	0.8	-0.0085	-0.0083	-0.0074	-0.0072	0.0078	0.0076
	0.9	-0.0082	-0.0080	-0.0076	-0.0072	0.0077	0.0076
	0.1	-0.0088	-0.0085	-0.0079	-0.0076	0.0077	0.0074
	0.2	-0.0089	-0.0086	-0.0078	-0.0076	0.0079	0.0076
50	0.5	-0.0090	-0.0087	-0.0079	-0.0075	0.0078	0.0076
	0.8	-0.0091	-0.0085	-0.0077	-0.0075	0.0077	0.0076
	0.9	-0.0090	-0.0087	-0.0076	- 0.00074	0.0079	0.0076
	0.1	-0.0090	-0.0088	-0.0080	-0.0077	0.0078	0.0076
	0.2	-0.0089	-0.0086	-0.0080	-0.0076	0.0077	0.0075
75	0.5	-0.0088	-0.0086	-0.0081	-0.0078	0.0081	0.0077
	0.8	-0.0089	-0.0085	-0.0079	-0.0075	0.0080	0.0076
	0.9	0.0034	0.0028	0.0036	0.00233	NA	NA

Table 9: L_2 Norm Comparison of the Estimators for Weibull $(\alpha = 2,\ \lambda = 1)$ Survival Curve for the Sample Sizes $10, 30, 60$ and 100 (Based on $50,000$ Iterations*)

	Sample size				
Estimator	CP	$n = 10$	$n = 30$	$n = 60$	$n = 100$
NPEE		0.342(.006)	0.338(.006)	0.320(.004)	0.308(.005)
EBE	15	0.345(.006)	0.341(.007)	0.318(.005)	0.308(.006)
KME		0.381(.004)	0.353(.004)	0.327(.004)	0.310(.004)
NPEE		0.360(.008)	0.351(.005)	0.331(.006)	0.312(.005)
EBE	50	0.366(.008)	0.348(.005)	0.330(.006)	0.311(.004)
KME		0.410(.004)	0.382(.005)	0.341(.006)	0.315(.004)
NPEE		0.371(.010)	0.355(.007)	0.335(.006)	0.315(.005)
EBE	75	0.377(.011)	0.358(.008)	0.334(.006)	0.317(.005)
KME		0.464(.003)	0.402(.004)	0.365(.003)	0.325(.003)

References

[1] Bartholomew, D. J. (1957). A problem in life testing. *Journal of the American Statistical Association*, **52**, 350-355.

[2] Breslow, N. and Crowley J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, **2**, 437-453.

[3] Efron, B. (1967). The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA*, **4**, 831-853.

[4] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209-230.

[5] Malla, G. and Mukerjee, H. (2010). A new piecewise exponential estimator of a survival function. *Statistics and Probability Letters*, **80**, 1911-1917.

[6] Gill, R. D. (1980). Censoring and Stochastic Integrals, Math. Centre Tract 124, Amsterdam: Math. Centrum.

[7] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc*, **53**, 457-481.

[8] Kitchin, J., Langberg, N., and Proschan, F. (1983). A new method for estimating life distributions from incomplete data. *Statistics and Decisions*, **1**, 241-255.

[9] Klein, J. P. and Moeschberger, M. L. (1984). The Asymptotic Bias of the Product-Limit Estimator Under Dependent Competing Risks. *Indian Journal of Productivity, Reliability and Quality Control*, **9**, 1-7.

[10] Meier, P. (1975). Estimation of a distribution function from incomplete observations. *Perspectives in Probability and Statistics, London: Academic Press*, **10**, 67-81.

[11] Nelson, W. (1969). Hazard plotting for incomplete failure data. *J. Quality Technology*, **1**, 27-52.

[12] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**, 945-966.

[13] Phadia, E. and van Ryzin, J. (1980). A note on convergence rates for the product limit estimator. *Annals of Statistics.*, **8**, 678-690.

[14] Rai, M., Susarla, V. and van Ryzin, J. (1980). Shrinkage estimation in nonparametric Bayesian survival analysis. *Communications in Statist. Simulation and Computation*, *B9*, 271-298.

[15] Susarla, V. and van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.*, **71**, 897-902.

[16] Susarla, V. and van Ryzin, J. (1978). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. *Ann. Statist.*, **6**, 755-768.

[17] Sweet, A.L. (1990). On the hazard rate of the lognormal distribution. *Reliability, IEEE Transactions*, **3**, 325-328.