Smooth Estimation of the Survival Function for Interval Censored Data

(A99-020R)

Wei Pan
Division of Biostatistics
School of Public Health
University of Minnesota
A460 Mayo Building (Box 303)
Minneapolis, MN 55455-0378, USA
weip@biostat.umn.edu
Phone: (612)626-2705, Fax: (612)626-0660

Revised November 1999
Smooth Estimation of the Survival Function for Interval Censored Data

SUMMARY

Interval censored data arise naturally in large-scale panel studies where subjects can only be followed periodically and the event of interest can only be observed between some time intervals. To estimate the survival function the nonparametric maximum likelihood estimator (NPMLE) is commonly used, which is a step function having some large jumps. But in many applications the underlying survival function can be reasonably assumed to be smooth, then the NPMLE does not efficiently use this information. Two smooth estimators have been proposed in the literature: one is based on the kernel smoothing of the NPMLE; another is the logspline density model. However, to our knowledge there is no finite sample study yet to assess their performance in the literature. In this paper we first show by simulation that both smooth estimators improve over the NPMLE for smooth survival functions. We then apply these estimators to compare two survival curves. The test statistics based on the maximum difference between two survival functions (Kolmogorov-Smirnov test) and on the integrated weighted difference of two survival functions (IWDB test) are investigated via the bootstrap. From our simulation studies the IWDB test seems particularly promising for some stochastically ordered survival functions that do not satisfy the proportional hazards model. The methods are illustrated by reanalyzing the Breast Cosmesis Study dataset.

Key Words: Bootstrap; Cross-validation; IWDB test; NPMLE; Kernel estimation; Logspline; Kolmogorov-Smirnov test.
1. INTRODUCTION

There has been an increasing interest in survival analysis with interval censored data, where the event of interest, such as infection with a disease or other failures, is not observed exactly but only known to happen between two examination times (but may also happen before the first examination or after the last examination). Hence interval-censoring is a natural generalization of the commonly encountered right-censoring or double-censoring. Interval censored data arise often from large-scale panel studies, such as AIDS studies, where the HIV infection time can only be detected at periodic follow-ups. Some of the subjects may have missed some scheduled follow-up examinations and come back later with a changed clinical status. Turnbull introduces a self-consistency algorithm, which is a special case of the EM algorithm, to compute the NPMLE of the survival function for arbitrarily censored and truncated data. Groeneboom and Wellner give a detailed treatment of the NPMLE for interval censored data, including a faster iterated convex minorant (ICM) algorithm and some large-sample properties. The ICM will be used throughout to compute the NPMLE. For an updated review see Huang and Wellner.

In many applications, the underlying survival function may be reasonably or approximately assumed to be smooth. However, the NPMLE is a step function and does not efficiently use this information. More importantly, according to its hypothesized $n^{-1/3}$ convergence rate, the NPMLE from interval censored data tends to have a smaller number of jumps and hence larger jump sizes than the usual $n^{-1/2}$ estimators, such as the empirical distribution function for complete data or the Kaplan-Meier estimator for right censored data, where $n$ is the sample size. This can be easily visualized from the example to be discussed later (Figure 5). Hence some smooth estimators may be more desirable. The most well-known is the kernel smoothing as for right censored data, where the kernel density estimate is obtained through smoothing increments of the Kaplan-Meier estimator (or NPMLE for interval censored data) of the survival function by some kernels. But it is possible that this smoothing can not recover all the information lost in the discrete NPMLE of the survival function. Therefore a direct modeling of the survival function (or its density) as being smooth might be more efficient. The logspline model is one of these approaches. Although both the kernel smoothing and logspline model are intuitive, neither their finite sample properties nor their asymptotic theories have been well-understood for interval censored data. It is the primary goal of this paper to investigate their finite sample performances, especially compared with the NPMLE.
Although the two smooth methods indeed estimate the survival function through estimating the density, we will only concentrate on their performance on estimating the survival function. For small to medium samples (say, \( n < 100 \)) with interval censoring it seems too ambitious to estimate the density accurately. We then extend these survival estimators to two-sample comparison. A difficulty in inference with interval censored data is that even the asymptotic distribution theory of the NPMLE of the survival function has not completely been settled,\(^4\) not to mention that of the smooth estimators. Notice that the classic MLE theory\(^7\) cannot be applied directly since the number of the unknown parameters to be estimated is increasing with the sample size. One alternative is through grouping data. After grouping, the number of the unknown parameters is fixed and hence the classic MLE theory is applicable. This is the case with Petroni and Wolfe's IWD test.\(^8\) But grouping is somewhat ad hoc and may result in severe loss of information. With the current advancement of computing technology, many computing-intensive methods, such as the bootstrap and jackknife, are becoming feasible and popular. We will use the bootstrap to approximate the distribution of a test statistic.

This paper is organized as follows. We first introduce aforementioned estimators of the survival function. They are evaluated by simulations. Then we apply the two smooth estimators and the NPMLE to two sample testing, supplemented with a small-scale simulation study and a comparison with other tests, including Mantel's test and IWD test. For illustration an example is taken from the Breast Cosmesis Study, followed by a discussion.

2. ESTIMATING SURVIVAL FUNCTION

For interval-censored data we do not observe the survival times \( X_i \) but the censoring times \((U_i, V_i)\). It is known that the survival time \( X_i \) satisfies \( 0 \leq U_i < X_i \leq V_i < \infty \) for \( i = 1, \ldots, n \). Throughout this paper as usual we assume the independent censoring. We are interested in estimating the survival function \( S(x) = P(X_i > x) \) from interval censored data \( \mathcal{D} = \{(U_i, V_i)\} \).

2.1 NPMLE

With the interval censored data given above, the (conditional) log-likelihood can be written down as:\(^2\)

\[
L(S|\mathcal{D}) = \sum_{i=1}^{n} \log \left( S(U_i) - S(V_i) \right).
\]
The nonparametric maximum likelihood estimator (NPMLE) \( \hat{S} \) of \( S \) is a right-continuous survival function maximizing \( L(S|D) \). Turnbull\(^2\) shows that \( \hat{S} \) can only have jumps between the order statistics \( \tau_i \), \( i = 1, \ldots, k \), of \( U \) and \( V \), \( 1 \leq i \leq n \). Turnbull also suggests how to reduce the number of the order statistics \( \{\tau_i\} \) on which \( \hat{S} \) may have jumps. However, the probability distribution of \( \hat{S} \) within some intervals is nonidentifiable. This nonidentifiability will not influence the computation of \( \hat{S} \) by the EM algorithm. To facilitate the computation of \( \hat{S} \) by the ICM, Groeneboom and Wellner\(^4\) further restrict the NPMLE \( \hat{S} \) as a piecewise constant function with possible discontinuities only at the order statistics \( \{\tau_i\} \), and moreover \( \hat{S} \) may be less than 1 at the largest order statistic \( \tau_k \). These two different definitions seem to be non-critical. In the sequel, we adopt the second one.

### 2.2 Kernel Estimator

For right censored data the kernel density estimator based on the NPMLE (Kaplan-Meier estimator) of the survival function has been well studied (see, for example, Andersen et al.\(^9\)). Keiding\(^10\) illustrates its application to interval censored data. Specifically, suppose \( \hat{S} \) is the NPMLE of the survival function, then the kernel estimate of its density function is

\[
\hat{f}(x; h) = \frac{1}{h} \int K \left( \frac{x - t}{h} \right) d\hat{S}(t),
\]

where \( K \) is a kernel function and \( h \) a bandwidth. The kernel estimator of the survival function \( \hat{S}(x; h) \) is then obtained by integrating the density estimator \( \hat{f}(x; h) \) accordingly.

As with complete data a key and difficult step is how to choose the bandwidth. There is almost no specific discussion on this issue for interval censored data, with an exception from Groeneboom,\(^11\) where Hall’s bootstrap method\(^12\) is outlined. This approach needs a pilot bandwidth and may be sensitive to the pilot estimate. Instead we apply a \( V \)-fold likelihood cross-validation approach.\(^13,14\) It proceeds by first dividing the original data set \( D \) randomly into \( V \) nearly equally sized sets \( D^{(v)} \), \( v = 1, \ldots, V \). Suppose \( \hat{S}^{(-v)}(\cdot; h) \) is the kernel estimate of the survival function from the data \( D - D^{(v)} \). Then our bandwidth is chosen to be

\[
h_0 = \arg \max_h \sum_{v=1}^V L(\hat{S}^{(-v)}(\cdot; h)|D^{(v)}),
\]

and our kernel estimator of the survival function is \( \hat{S}(\cdot) = \hat{S}(\cdot; h_0) \).

We note that in selecting an appropriate bandwidth the usual direct methods based on the asymptotics for complete or right censored data, such as the plug-in estimators, are no longer
available, partly due to the fact that generally there is no explicit form for the NPMLE from interval censored data. For example, it is not clear how to write down the asymptotic mean integrated squared error (MISE) of a kernel estimator with bandwidth $h$. Furthermore, even though the CV has been criticized for being not efficient enough in complete data setting (e.g. Wand and Jones\textsuperscript{15}, page 86), it appears that among many of its competitors the CV is the most reliable for small samples from an extensive simulation study by Grund and Polzehl.\textsuperscript{16} Its effectiveness will be confirmed later by comparing the performance of the kernel estimator with that of the logspline estimator.

### 2.3 Logspline estimator

As mentioned before, smoothing the NPMLE may not recover some information already lost in the NPMLE. It is possible that obtaining a smooth estimate directly from the original data is more efficient. In this way the logspline approach models the log-density as a spline.\textsuperscript{6} Here we exclusively use a cubic spline, which is a twice continuously differentiable and piecewise cubic. Specifically, the timeline is partitioned into $K$ intervals $(0, t_1], ..., (t_K, \infty)$, where $\{t_i\}$ are called knots. The $K$-dimensional natural cubic spline space has a basis $\{1, B_1, ..., B_p\}$, $p = K - 1$. Then we model the density function as

$$f(x; \theta) = \exp(\sum_{i=1}^{p} \theta_i B_i(x) - C(\theta)),$$

where $\theta = (\theta_1, ..., \theta_p)'$ is the parameter vector, and $C(\theta) = \log[\int \exp(\sum_{i=1}^{p} \theta_i B_i(y))dy] < \infty$ makes $f(x; \theta)$ a proper density function. Let $S(\cdot; \theta)$ denote the corresponding survival function. The $\theta$ is estimated through maximum likelihood: $\tilde{\theta} = \arg \max_\theta L(S(\cdot; \theta)|D)$, and the logspline estimator of the survival function is $\tilde{S}(x) = S(x; \tilde{\theta})$.

An interesting feature of the logspline model discussed in Kooperberg and Stone\textsuperscript{6} is its adaptivity, where the number of knots $K$ can be adaptively chosen. The user can specify an maximum number of knots $K_0$ to be used, but a step-wise knot-deletion process based on the Wald statistic will fit a sequence of models $\tilde{S}_\nu$ indexed by $\nu$, each of which has $p_\nu$ parameters. The final model is selected to be the one minimizing the Bayesian Information Criterion $BIC_\nu = -2L(\tilde{S}_\nu|D) + (\log n)p_\nu$. Notice that other choices of the penalty term rather than $\log n$ in $BIC_\nu$ is allowable. For details see the above reference. Evidently the $K_0$ serves to control (as a prior) the maximum degree of smoothness of the logspline estimates. Kooperberg and Stone also provide an implementation of
the algorithm with user-friendly S-Plus interface, which is available through StatLib. We used their program in our simulations for the logspline estimates.

2.4 Simulation

First we need to generate some interval censored samples. The first examination time $T_i$ is taken as $\text{Uniform}(0, a)$, independent of the survival time $X_i$. To mimic many panel studies we take the length of the time interval between two examinations to be constant. Generally if we have total $k$ follow-ups after the first examination, the survival time $X_i$ is accordingly censored in one of $(0, T_i], (T_i, T_i + \text{len}_j], ..., (T_i + \text{len}_1 + ... + \text{len}_k, \infty)$, where $\text{len}_j$ for $1 \leq j \leq k$ are constants. The parameters $(a, k, \text{len}_1, ..., \text{len}_k)$ jointly determine the censoring pattern. In our simulation studies there are four configurations:

Case 1. $X_i \sim \text{Weibull}(2, 1), k = 0, a = 1.5$;
Case 2. $X_i \sim \text{Weibull}(2, 1), k = 2, a = 0.7, \text{len}_1 = 0.5, \text{len}_2 = 0.3$;
Case 3. $X_i \sim 0.6 \ast N(0.5, 0.1) + 0.4 \ast N(1, 0.1), k = 0, a = 1.5$;
Case 4. $X_i \sim 0.6 \ast N(0.5, 0.1) + 0.4 \ast N(1, 0.1), k = 2, a = 0.7, \text{len}_1 = 0.5, \text{len}_2 = 0.3$.

In Cases 1 and 3, the censoring intervals are $(0, T_i]$ and $(T_i, \infty)$ while in Cases 2 and 4 they are $(0, T_i], (T_i, T_i + 0.5], (T_i + 0.5, T_i + 0.8]$ and $(T_i + 0.8, \infty)$. The number of follow-up $k = 2$ yields less right-censoring and hence more information in resulting samples than $k = 0$ does. The density of $\text{Weibull}(2, 1)$, the Weibull distribution with shape parameter 2 and scale 1 has only one mode, while that of the normal mixture is less "smooth" with two modes. The sample size is always 100, with 200 independent replications.

All our simulation programs (including those in Section 3.2) were written in S-Plus, which are available upon request. In particular, Our kernel estimator was implemented by using Parzen's kernel with S-Plus function ksmooth. Ten-fold likelihood CV was adopted to select the bandwidth from $\{0.1, 0.2, 0.3, ..., 1.2\}$. For the four cases, the means (standard errors) of the selected bandwidths are respectively 0.50 (0.16), 0.58 (0.18), 0.33 (0.12) and 0.24 (0.08); the medians are close to the means and are respectively 0.5, 0.5, 0.3 and 0.2. The logspline fitting procedure logspline.fit may fail (with a message "no convergence") for small to medium or more heavily censored samples if we supply a larger $K_0$. For example, $K_0 = 5$ worked well for Case 4 but failed for some data sets in Case 3 and Case 1. Therefore for consistency we only used $K_0 = 4$, which works for all data sets.
but is probably too small for the normal mixture whose density has two modes. We note that the logspline software works well for right censored data. Hence conceptually this issue can be resolved by a future improved implementation with more attention paid to some numerical features with interval censoring. It is widely believed that the convergence rate of the NPMLE of the survival function is only $n^{-1/3}$ for interval censored data, in contrast to the common $n^{-1/2}$ for complete or right censored data. Accordingly we adjust the penalty term $\log n$ to $\frac{2}{3} \log n$ in the BIC model selection criterion, though it seems that the final logspline model is not very sensitive to the choice (within some ranges).

For the Weibull distribution the results are shown in Figure 1. It is evident that compared with the NPMLE the two smooth estimators improve the performance through reduced variability. All three estimators only have negligible bias, except at the right tail of the NPMLE. This bias is caused by right-censoring as for Kaplan-Meier estimator. However, the logspline estimator is apparently over-smooth without capturing the two high-curvature points in the survival function (or correspondingly the two modes in the density function) for the normal mixture with $K_0 = 4$ (Figure 2). The other two estimators are again impressive with their small bias (except the right tail of the NPMLE). In particular the kernel estimator has smaller variability than the NPMLE. But the logspline estimates do improve greatly if we use $K_0 = 5$ for Case 4 (Figure 3). Hence we are still optimistic about the performance of the logspline models.

Figures 1–3 about here

3. APPLICATION TO TWO-SAMPLE COMPARISON

Now we consider comparing two survival curves. Suppose we have two interval censored samples $D_1$ and $D_2$ with corresponding survival functions $S_1$ and $S_2$. We want to test $H_0: S_1 = S_2$ nonparametrically. This problem has been considered by many authors,$^{8,17,18}$ most of which are built on the NPMLE of the survival function. In our current framework, it is natural to consider tests based on the smooth estimators of the survival function. This latter approach is missing from the literature. Here we explore two of this kind of nonparametric tests.

The first we consider is the Kolmogorov-Smirnov test (Fleming et al.$^{19}$ for right-censored data). Suppose we have two survival function estimates $\hat{S}_1$ and $\hat{S}_2$ (which in our notation are the NPMLEs
but the same argument carries over to the kernel or logspline estimates. The test statistic is

\[ KS = \max_x |\hat{S}_1(x) - \hat{S}_2(x)|. \]

For interval censored data, it is difficult to derive the exact or asymptotic distribution of the above test statistic. Here we recourse to the bootstrap.\(^2\) Specifically, suppose that \( KS_0 \) is the test statistic computed from \( D_1 \) and \( D_2 \), and \( D = D_1 \cup D_2 \) is the pooled sample. Then with replacement from \( D \) we draw two random samples \( D_1^b \) and \( D_2^b \), each with an equal size as \( D_1 \) or \( D_2 \). From these two random samples we can calculate our test statistic \( KS_b \). This process is repeated \( B \) times. Then with a given nominal level \( \alpha \) we reject \( H_0 : S_1 = S_2 \) if \( \sum_{b=1}^B I(KS_b \geq KS_0) + 1 \leq \alpha(B + 1) \), where \( I \) is the indicator function. For a related work for complete data see Allen.\(^2\)

Another test is based on the integrated weighted difference (IWD) of two (truncated) survival functions (Pepe and Fleming\(^2\) for right censored data; Petroni and Wolfe\(^8\) for grouped interval censored data). The test statistic is

\[ IWD = \hat{\mu}_1 - \hat{\mu}_2, \text{ where } \hat{\mu}_i = \int w(t)\hat{S}_i(t)dt, \ i = 1, 2. \]

\( \hat{\mu}_i \) are the estimated weighted mean survival times (possibly truncated at right). For simplicity we only consider unit weights \( w(t) \equiv 1 \). Although the estimated mean survival time based on the NPMLE has an asymptotic normal distribution, its asymptotic variance is not easy to obtain.\(^4\) Again we use the bootstrap to estimate its distribution. To distinguish our bootstrap approach from Petroni and Wolfe's IWD test,\(^8\) where the observed information matrix from grouped data is used to approximate the variance in its asymptotic normal distribution, we denote ours as IWDB test.

### 3.2 Simulation

The random samples were generated essentially in the same way as in Section 2.4. The distribution of the first examination time is still \( \text{Uniform}(0, \alpha) \), where \( \alpha \) is chosen such that 25% observations are righted-censored. There is only one follow-up with \( \text{cen}_1 = 0.5 \) such that the censoring intervals are \((0, T_i], (T_i, T_i + 0.5] \) and \((T_i + 0.5, \infty) \). The sample size is 50 for each group.

To investigate size properties the underlying survival distribution \( \text{Weibull}(2, 1) \) is used. In order to investigate power properties, as alternatives, both the proportional hazards model (PHM) and non-PHM are considered as in Fleming \textit{et al.}\(^{19}\) In the PHM, the distribution of the first
Smooth Estimation of Survival

Weibull(2, 1), and \( S_2(x) = S_1(x)^{2.25} \). For the non-PHM, two piece-wise exponential distributions with crossing hazards are applied, which are still stochastically ordered. The configurations are shown in Figure 4 (see Fleming et al.\(^{19}\) for analytic expressions).

Figure 4 about here

We used bootstrap replication number \( B = 200 \). Since now we have a smaller sample size, only five-fold CV was used in the kernel estimator and \( K_0 = 3 \) in the logspline model. To save computing time, the CV-selected bandwidths \( h_1 \) and \( h_2 \) from the original samples \( D_1 \) and \( D_2 \) are used throughout for subsequent bootstrap samples \( D_1^b \) and \( D_2^b \) for \( 1 \leq b \leq B \). Again notice that due to numerical instability of the logspline software we only used a small \( K_0 = 3 \) for the logspline estimators. The results are shown in Table 1, with Mantel’s test\(^{22}\) and IWD tests attached for comparison. In overall the KS test is not satisfactory, due to the large variations of the survival estimates from small samples. However, the IWDB test is particularly promising for testing the crossing hazards. Notice that for the crossing hazards model the underlying survival distributions are not smooth, hence it is not surprising that the IWDB test based on the NPMLE has a slightly higher power than that from the smooth estimators. But for smooth survival distributions, such as in the PHM, the tests based on the smooth estimators evidently are favored over those on the NPMLE. The IWDB test based on the NPMLE also has a higher power than the IWD test for the crossing hazards model, hence it may suggest either that grouping data may lose a non-ignorable amount of information or that the asymptotic approximation (including deriving the variance from the observed information matrix) is not accurate enough.

Table 1 about here

4. EXAMPLE

Now we apply the methods to reanalyze the Breast Cosmesis Study dataset.\(^{17,24}\) Two medical treatments were given to some early breast cancer patients after tumorectomy: one was a mix of primary radiotherapy and adjuvant chemotherapy, and another was radiotherapy only. The interest of this clinical trial was to investigate which treatment has better long-term cosmetic effects. Interval-censoring results since the patients could only be visited every 4 to 6 months and those living farther away from the clinic had even longer followup intervals. There are respectively 46 and 48 subjects in each treatment group. We first estimate the survival function for each group.
The "survival" here is defined as being cosmetically positive. Using leave-one-out (or \( n \)-fold with \( n \) the sample size) likelihood CV the chosen bandwidths are respectively 6 and 25 months for kernel estimates. (There is no huge difference in resulting kernel smooths with bandwidth 6 or 25 for the first group when compared with the NPMLE.) With \( K_0 = 6 \) the selected final logspline model only has three knots for both groups. These two final models are fairly stable with a wide range of values of the penalty term in the BIC. The estimates are shown in Figure 5. As pointed out earlier, the NPMLE tends to have a far smaller number of jumps than the sample size \( n \), and thus much larger jumps than \( 1/n \). Notice that at the beginning two estimated survival curves cross, casting some doubts on the proportionality of the hazards. But the main difference between two survival curves remains at later times.

Now we test the null hypothesis that there is no difference between two treatments, with the alternative hypothesis that the first treatment has a better effect. With \( B = 1000 \) bootstrap replications the \( p \)-values of the IWDB test based on the NPMLE, the kernel and logspline estimators of the survival function are respectively 0.001, 0.002 and 0.004. For the KS test the \( p \)-values are 0.005, 0.030 and 0.002 respectively. (As in Figure 5 we used \( K_0 = 6 \) in the logspline models.) These results are consistent with earlier studies. Finkelstein's\(^{17} \) score test in the Cox regression model has an associated \( p \)-value 0.004. Similar results are also obtained by Fay\(^{18} \) and Pan\(^{25} \), among others. By the above evidence we believe that there is a significant long-term benefit with the mixed radiation therapy and chemotherapy.

5. DISCUSSION

We have shown that as expected the kernel and logspline estimators of the survival function may have better performances than the NPMLE for interval censored data when appropriately used. Between these two smooth estimators, the kernel smoother is not a clear loser since its pointwise mean squared errors are close to those of the logspline estimator (see Figures 1–2). Though the kernel estimator works by smoothing the NPMLE, which is discrete, we used the data again to choose the bandwidth. The logspline estimator has the potential to be more flexible in the sense that covariates can be incorporated in regression analysis.\(^{35} \) The IWDB test we proposed is particularly promising for stochastically ordered non-proportional hazards alternatives. But many theoretical
issues related with the smooth estimators still need to be tackled. Our current work suggests that such theoretical pursuits are worthwhile.

One problem with the logspline model is its current implementation from statlib does not support well small to medium sized or heavily censored samples. This issue may be largely resolved by a soon to-be-released new improved version by Clarkson and Kooperberg. It is particularly encouraging that the logspline model can be cast in a more general framework, extended linear modeling, with far-reaching applications.

An interesting question is how to estimate the standard error function (or confidence band) of the estimated survival curve. For the NPMLE with interval censored data, Huang showed that the usual bootstrap (as we used) does not work. The fundamental reason is that the convergence rate of the NPMLE is only \( n^{-1/3} \), not the usual \( n^{-1/2} \), which is often required by the usual bootstrap. However, other subsampling schemes, such as the \( m \)-out-of-\( n \) bootstrap, may work. In the \( m \)-out-of-\( n \) bootstrap, the bootstrap sample size is \( m \), which is smaller than the original sample size \( n \). In fact, it is needed that \( m/n \to 0 \) and \( m \to \infty \) as \( n \to \infty \). Then a technical difficulty is how to choose a proper \( m \) for a given sample size \( n \). A similar argument for using subsampling may also apply to our proposed KS test, though our limited simulation studies did not suggest any evidence against using the usual bootstrap.

In the NPMLE approach to the Cox proportional hazards regression with interval censored data, one needs to estimate both the regression coefficient and the baseline survival simultaneously. If the baseline survival function is of interest, one can smooth it in the same spirit as we did in Section 2.2 with a straightforward modification to the likelihood function (and using the NPMLE of the regression coefficient). It is also possible that after obtaining a smooth estimate of the baseline survival function, one can plug-in this estimate in the likelihood function and obtain an updated estimate of the regression coefficient. A more efficient (and complex) way is to model the baseline distribution as being smooth and then maximize the likelihood function to obtain the estimates of both the baseline survival function and regression coefficient (e.g. Kooperberg and Clarkson). In summary, we believe that the smooth estimation approach is an important complement to the existing NPMLE approach for interval censored data.

ACKNOWLEDGEMENT

The author wishes to thank Dr. Charles Kooperberg for kindly providing and helping with the
logspline software. The author is also grateful to the referees for detailed and helpful comments that dramatically improved the presentation.

REFERENCES


34. Politis, D. N. and Romano, J. P. ‘Large sample confidence regions based on subsamples under minimal assumptions’, *Annals of Statistics*, 22, 2031-2050.

Table 1: Sizes and powers of the tests at nominal level 0.05 (0.01) for $H_0 : S_1 = S_2$. The tests are two-sided for evaluating size but one sided for power. Sample size is 50 + 50 with 1000 replications.

<table>
<thead>
<tr>
<th>tests</th>
<th>$S_1 = S_2 = W(2)$</th>
<th>$S_1 = W(2), S_2 = S_{1.25}^2$</th>
<th>Crossing hazards</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS(NPMLE)</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>KS(Kernel)</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>KS(Logspline)</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>IWDB(NPMLE)</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>IWDB(Kernel)</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>IWDB(Logspline)</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Mantel</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>IWD</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Figure 1: Estimated survival curves for Weibull(2, 1) with interval censored data. The horizontal axes are for time $x$. Sample size is 100 with 200 replications. The solid lines are true distributions, and the dotted ones are the (pointwise) means, the lower and upper 5% quantiles of the estimates.
Figure 2: Estimated survival curves for a normal mixture with interval censored data. The horizontal axes are for time $x$. Sample size is 100 with 200 replications. The solid lines are true distributions, and the dotted ones are the (pointwise) means, the lower and upper 5% quantiles of the estimates.
Figure 3: The logspline estimates for Case 4, with $K_0 = 5$. The horizontal axis is the survival time $x$. Sample size is 100 with 200 replications. The solid line is the true distribution, and the dotted one is the (pointwise) means, the lower and upper 5% quantiles of the estimates.
Figure 4: Configurations of the two survival curves to be tested in the simulation studies. The horizontal axes are for time $x$. 
Figure 5: Estimates of the survival (defined as being cosmetically positive) probabilities for each treatment group from the Breast Cosmesis Study.