Confidence Bands for the Survival Curve Based on the Kaplan-Meier Estimate

Ann Walker
Stat 570
June 14, 1989

Abstract: It is important in the practice of statistics to accompany an estimate with a confidence band to indicate the uncertainty of the estimate. There are several alternative methods for calculating confidence bands for the survival curve based on the Kaplan-Meier estimate. These methods include equal precision bands, Hall-Wellner bands and bootstrapped confidence bands. The goals of this paper are to review these methods and to implement the computations in the S programming language.

Introduction

The Kaplan-Meier (product limit) estimator of the survival function for right censored data is useful in the analysis of medical clinical trials and industrial life testing. It is the nonparametric maximum likelihood estimate for the survival curve, has an asymptotic normal distribution, and is relatively easy to compute (Efron, 1981; Miller, 1983).

The Kaplan-Meier estimate of the survivor function is defined as follows. Let $X_1^\circ, \ldots, X_n^\circ$ denote a sample of independent, identically distributed (iid) survival times with distribution function F and survivor function $S = 1 - F$. Let $Y_1^\circ, \ldots, Y_n^\circ$ denote the corresponding iid censoring variables (times) with distribution function G, where F and G are independent. The values (X_i, δ_i), $i = 1, \ldots, n$ (define indicator) where $X_i = \min(X_i^\circ, Y_i^\circ)$ and $\delta_i = I[X_i = X_i^\circ]$ (what is I(·)?), $i = 1, \ldots, n$ are observed. Therefore, the observed survival time is X_i; δ_i indicates whether the observation was censored or uncensored. Let R_i denote the rank of X_i in the lexicographic ordering of the observed survival times. The δ_i are the corresponding indicator values. Then the Kaplan-Meier estimator of $S = 1 - F$ is

$$
\begin{aligned}
\hat{S}_n(x) &= \Pi_{\{i:X_i \le x\}}[(n - R_i)/(n - R_i - 1)]^{\delta i} \quad , x \le \max(X_i) \\
&= 0 \quad , \text{otherwise.}
\end{aligned}
$$

If a tie occurs between a censored and an uncensored observation, say at time t, the failure is considered to have occurred a short time before t, $t - \Delta$, and the censored observation a short time after t, $t + \Delta$. This is consistent with Kaplan and Meier's procedure for handling ties (Gross) + Clark (1975).

The variance of the product limit estimator can be estimated using Greenwood's formula (Miller, 1983). The estimated variance is $\hat{S}_n^2(x)\hat{\sigma}_n^2(x)/n$, where $\hat{\sigma}_n^2(x) = n\Sigma_{\{i:X_i \le x\}}\ \delta_i/[(n - R_i)(n - R_i + 1)]$. Efron (1981) proves that the bootstrapped estimate for the standard deviation of the product limit estimate is approximately the same as that computed using Greenwood's formula.

A number of alternatives are available for calculating simultaneous confidence bands for the survival curve based on the Kaplan-Meier estimate. The "equal precision" (Nair, 1984) and Hall-Wellner bands (Hall, 1980) are both based on the asymptotic distribution of the product limit estimator. However, confidence bands based on asymptotic results have been shown to be less accurate than confidence bands constructed by bootstrapping (Akritas, 1986).

The goals of this paper are to review the alternative methods for calculating simultaneous confidence bands for the survivor function based on the Kaplan-Meier estimate and to illustrate the confidence band calculations as implemented in the S programming language. The S language is described in Becker et al. Equal precision and Hall-Wellner bands based on the asymptotic distribution of the product limit estimator are discussed as well as bootstrapped confidence bands using the Hall-Wellner method.

Asymptotic Confidence Bands

Confidence bands based on the asymptotic distribution of the product limit estimate are valid only for a finite interval $0 \leq x \leq T < \infty$. In practice, $T < T_n$, where T_n is the largest uncensored observation. The restriction is due to problems with the limiting distribution of the Kaplan-Meier estimator (Nair, 1984). Nair defines an "equal precision" band under a random censoring model to be

$\hat{S}_n(x) \pm e_\alpha \sqrt{n} \hat{S}_n(x) \hat{\sigma}_n(x)$, $\forall$ x: $a \leq \hat{K}_n(x) \leq b$, where a and b are fixed such that $0 < a < b < 1$, and $\hat{K}_n(x) = \hat{\sigma}_n(x)/(1 + \hat{\sigma}_n(x))$. The critical value e_α can be approximated by solving $A(e_\alpha) = \alpha/2$, where

$$A(x) = x\exp(-x^2/2)\log[(1-a)b/a(1-b)]/8\pi.$$

Nair states that values of $a = 0.05$ or 0.1 and $b = \min(.95, \hat{K}_n(T_n))$ or $\min(0.9, \hat{K}_n(T_n))$ offer a "reasonable balance between the width and coverage of the band". The band is "equal precision" in the sense that the width of the band is proportional to its estimated standard deviation. The problem with the equal precision band is that it is not valid in the tails where $a \le \hat{K}_n(x) \le b$ is not satisfied.

The Hall-Wellner band is also a simultaneous confidence band valid only for

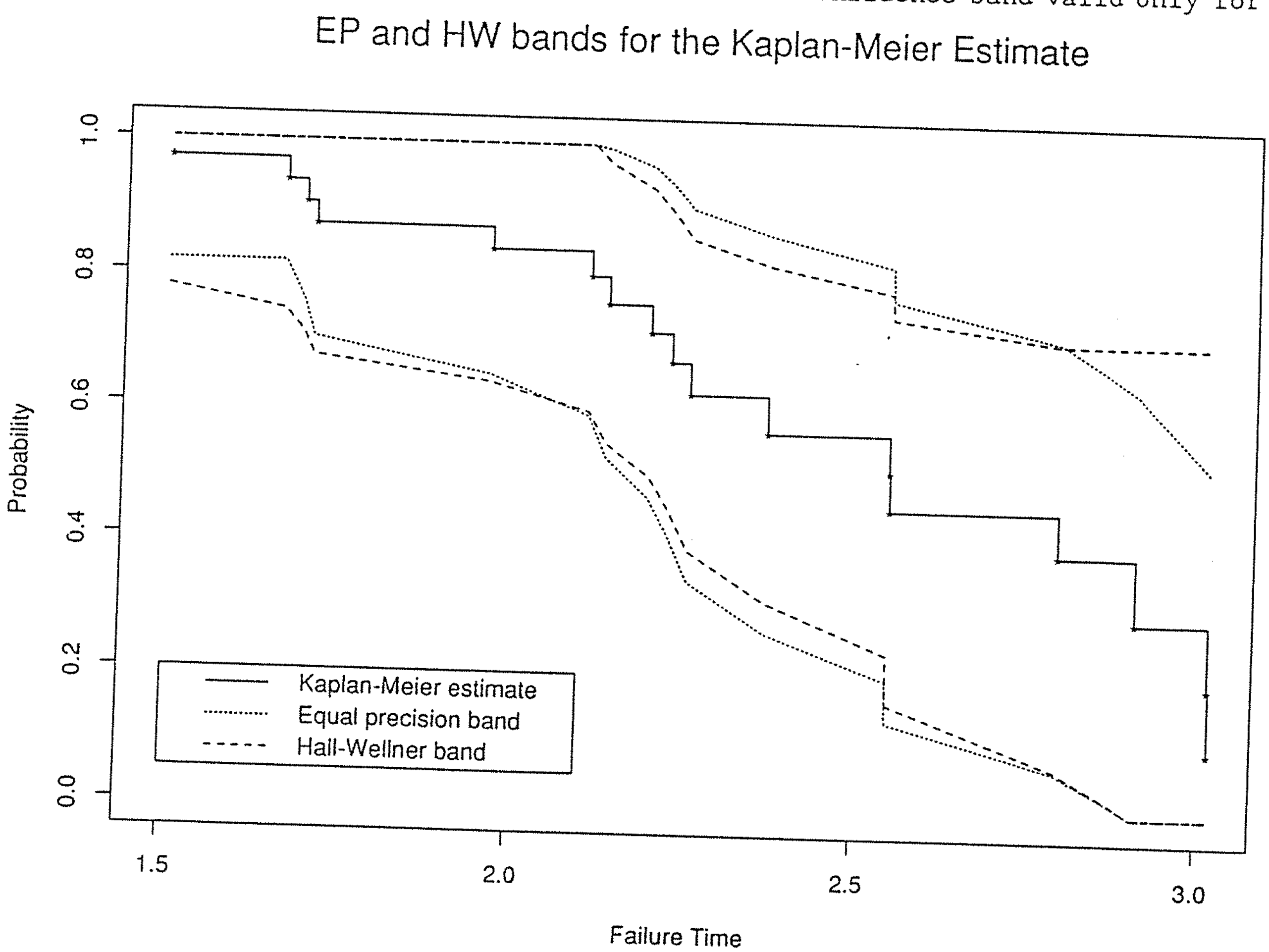

FIGURE 1: 90% Hall-Wellner and Equal Precision bands for the Kaplan-Meier estimate for mechanical switch life test data. The Hall-Wellner band is narrower in the center. The Equal Precision band is narrower in the tails.

$0 \le x \le T$. The Hall-Wellner band is defined by

$$\hat{S}_n(x) \pm h_\alpha \hat{S}_n(x)/\sqrt{n}(1 - \hat{K}_n(x))$$

where h_α can be approximated by solving $2\exp(-2h_\alpha^2) = \alpha$, and $\hat{K}_n(x)$ is defined as above. One deficiency of the Hall-Wellner and equal precision bands is that they tend to be increasing in the tails (not monotone), violating a basic property of survival curves.

Figure 1 is the Kaplan-Meier estimate for the survival curve with 90% equal precision and Hall-Wellner confidence bands. The graph was generated using the local 'S' function 'kap' which duplicates the graph done by Nair (1984). The data, shown in table 1 are the failure times for a mechanical switch life test (Nair, 1984). The critical value for the equal precision band was $e_\alpha = 2.91$ which corresponds to $a = 0.05$ and $b = 0.95$. See Nair (1984) for a table containing critical values for other choices of a and b and for various α levels. The first and last observations were not contained in the coverage of the band due to the choices of a and b. The band was extended to cover these values by assigning them the same value as the nearest valid bound. The critical value for the Hall-Wellner band was $h_\alpha = 1.22$. The last three observations for the upper Hall-Wellner band were not monotone but were increasing. The bound of the fourth largest observation was substituted for these values. The adjustments used here are the conventions suggested by Nair (1984). Both bands were computed for $0 \le x \le T$, where T was the second largest uncensored observation. As can be seen in Figure 1, the Hall-Wellner band is narrower in the center while the equal precision band is narrower in the tails. This is typical behavior for the two types of bands (Nair, 1984).

The function 'kap' has several options. It can be used to generate a plot of the Kaplan-Meier survival curve estimate for right censored data. For example, the command kap(X), where X is a matrix whose first column contains survival times and whose second column contains the corresponding indicator values, will result in a graph of the Kaplan-Meier survival curve. Note that ties between censored and uncensored observations are handled by considering the uncensored observation to have occurred before the censored observation.

The function 'kap' can also be used to add either "equal precision" bands, Hall-Wellner bands or both to the graph of the Kaplan-Meier survival curve. In order to generate equal precision bands, the critical value for the equal precision band along with the corresponding values for a and b must be passed into the function. For example, kap(X,a=0.05,b=0.95,e=2.91) will result in the addition of a 90% equal precision band to the Kaplan-Meier survival curve as in Figure 1. The function computes the equal precision band for $T < T_n$, the largest uncensored observation. The standard deviation of the Kaplan-Meier survival curve is estimated using Greenwood's formula. If the bounds for any of the observations are greater than one or less than zero the bound at that point is replaced with one or zero, respectively. Observations for which $\hat{K}_n(x) < a$ or $\hat{K}_n(x) > b$ are assigned the value of the nearest valid bound.

In order to generate the Hall-Wellner bands, X and a critical value must be passed into the function. For example, kap(X,h=1.22) will result in the addition of a 90% Hall-Wellner confidence band to the Kaplan-Meier survival curve as in Figure 1. The standard deviation of the Kaplan-Meier survival curve is estimated using Greenwood's formula. If any of the calculated bounds are greater than one or less than zero, the bound at that observation is replaced with one or zero, respectively. If the band is increasing in the tails these values are replaced by the nearest nonincreasing bound. Both the equal precision and Hall-Wellner bands can be added to the Kaplan-Meier graph by using the command kap(X,a=0.05,b=0.95,e=2.91,h=1.22).

The approximations of the critical values for both the equal precision and Hall-Wellner bands become less accurate as the proportion of censoring increases. However, in the situation where there is light censoring (25% or less) the asymptotic critical values give a good approximation for the Hall-Wellner and equal precision bands. This holds for small samples, regardless of the censoring mechanism (Nair, 1984). In general, as the censoring proportion increases, the relative performance of the equal precision to the Hall-Wellner band improves (Nair, 1984).

Bootstrapped Confidence Bands

Bootstrapped nonparametric confidence bands have been shown to be more accurate than confidence bands based on the asymptotic distribution of the estimators (Akritas, 1986). The following is a nontechnical description of the bootstrap procedure for computing critical values for confidence bands using the Hall-Wellner method. For a rigorous mathematical discussion, see Akritas (1986).

The bootstrap sample must first be determined by sampling n times, at random, with replacement from $X_1^\circ, \ldots, X_n^\circ$, and independently sampling n times at random, with replacement from $Y_1, \ldots, Y_n$. Y_i° ? The resulting bootstrapped samples are denoted $X_1^{\circ *}, \ldots, X_n^{\circ *}$ and $Y_1^*, \ldots, Y_n^*$. The "observed" data would then be $X_i^* = \min(X_i^{\circ *}, Y_i^*)$, $\delta_i^* = I[X_i^* = X_i^{\circ *}]$. This method has been shown by Efron to be equivalent to taking a sample (X_i^*, δ_i^*),

what if X_i° is not observed?

$i = 1, \ldots, n$ at random, with replacement from (X_i, δ_i), $i = 1, \ldots, n$ (Akritas, 1986).

In order to compute a $(1 - \alpha)100\%$ confidence band, calculate

$\sqrt{n}\sup\{|(\hat{S}_n^*(x) - \hat{S}_n(x))(1 - \hat{K}_n(x))/\hat{S}_n(x)|\}$, $0 \le x \le T$, where T is the second largest uncensored observation, $\hat{S}_n(x)$ and $\hat{K}_n(x)$ are defined as above, and $\hat{S}_n^*(x)$ is the Kaplan-Meier estimate for the bootstrapped sample, for each of 200 to 300 bootstrap samples. The bootstrap critical value is then approximated by the $(1 - \alpha)100^{th}$ percentile of these numbers (c_n). The confidence band based on the Hall-Wellner method is

$$\hat{S}_n(x) \pm (c_n/\sqrt{n})\hat{S}_n(x)/(1 - \hat{K}_n(x)) \quad \text{(Akritas, 1986).}$$

The bootstrapped band, like the Hall-Wellner band is not monotone in the tails. The increasing bounds are assigned the value of the first valid bound immediately preceding them. The bootstrapped confidence band is narrower than the Hall-Wellner band. Hence, it is less conservative. The bootstrapped band also gives accurate coverage probabilities for discrete data which is not true of the Hall-Wellner band. Akritas (1986) recommends using the bootstrapped confidence bands rather than the Hall-Wellner bands in the case of discrete or grouped data.

The local 'S' function 'bboot' computes the bootstrapped critical value c_n. The function 'bboot' is called by bboot(X,α,N), where N is the number of bootstrapped samples to be computed. The function 'bboot' first computes the Kaplan-Meier survival curve for the original data. Bootstrapped samples are then obtained by sampling at random, with replacement, n times from the survival times and independently from the indicator values. The Kaplan-Meier survival curve is then calculated for the bootstrapped sample. The expression

Is this the procedure described on the previous page?

$$\sqrt{n}\max\{|(\hat{S}_n^*(x) - \hat{S}_n(x))(1 - \hat{K}_n(x))/\hat{S}_n(x)|\} \qquad (1)$$

Is this really needed?

is evaluated for each bootstrapped sample for $0 \le x \le T$, where T is the second largest uncensored observation. (Note that the maximum is evaluated rather than the supremum.) The expression (1) is evaluated at each observed survival time, whether the observation is censored or not. The $(1 - \alpha)100^{th}$ percentile of the numbers generated by (1) is then computed to yield c_n. The disadvantage of the bootstrap method for computing confidence bands is that it is computationally

for n = ?

intensive, requiring approximately seven hours to run on a MicroVAX II. In order to obtain bootstrapped confidence bands simply call kap(X,h=c_n).

What recommendations on n = # of ~~bootstrap~~ samples?

Summary

It is important in the practice of statistics to accompany an estimate with a confidence band to indicate the uncertainty of the estimate. There are several alternative methods for calculating confidence bands for the survival curve based on the Kaplan-Meier estimate. These methods include equal precision bands, Hall-Wellner bands and bootstrapped confidence bands. If practical, bootstrapped confidence bands should be used rather than those based on the asymptotic distribution of the Kaplan-Meier estimate. The bootstrapped confidence bands give accurate coverage probabilities which is not necessarily true of the asymptotic confidence bands.

REFERENCES

Akritas, M.G. (1986), "Bootstrapping the Kaplan-Meier Estimator," Journal of the American Statistical Association, 81, 1032-1038.

Becker, B.A., Chambers, J.M. and Wilks, A.R. (1988), The New S Language: A Programming Environment for Data Analysis and Graphics, Wadsworth Brooks/Cole, California.

Efron, B. (1981), "Censored Data and the Bootstrap," Journal of the American Statistical Association, 76, 312-319.

Gross, A.J. and Clark, V.A. (1975), Survival Distributions: Reliability Applications in the Biomedical Sciences, John Wiley & Sons, New York.

Hall, W.J. and Wellner, J.A. (1980), "Confidence Bands for a Survival Curve," Biometrika, 67, 133-143.

Miller, R. G. (1983), "What Price Kaplan-Meier?," Biometrics, 39, 1077-1081.

Nair, V. N. (1984), "Confidence Bands for Survival Functions with Censored Data: A Comparative Study," Technometrics, 26, 265-275.