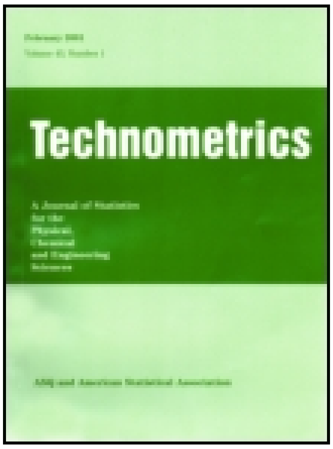

Technometrics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utch20

Confidence Bands for Survival Functions With Censored Data: A Comparative Study

Vijayan N. Nair [a]

[a] AT&T Bell Laboratories , Murray Hill , NJ , 07974

Published online: 23 Mar 2012.

TECHNOMETRICS ©, VOL. 26, NO. 3, AUGUST 1984

Confidence Bands for Survival Functions With Censored Data: A Comparative Study

Vijayan N. Nair
AT&T Bell Laboratories
Murray Hill, NJ 07974

Consider the problem of obtaining simultaneous confidence bands for the survival function $S(x)$ when the data are arbitrarily right censored. The usual pointwise confidence intervals based on Greenwood's variance formula can be adapted to yield a large-sample confidence band. This band has, in a certain sense, equal precision at each point of $S(x)$. It is compared with the censored versions of the Kolmogorov band and the Rényi band. The comparisons are made in terms of the widths and the adequacy of large-sample approximations and are carried out under various censoring models and degrees of censoring. The bands are illustrated by applying them to data from a mechanical-switch life test.

KEY WORDS: Censoring; Confidence bands; Distribution free.

1. INTRODUCTION

Arbitrarily right censored data arise commonly in industrial life testing and medical follow-up studies. It is important in these situations to obtain nonparametric point and interval estimates of the survival function S. Kaplan and Meier (1958) provided the nonparametric maximum likelihood estimate, $\hat{S}_n$, of S. Termed the product-limit estimator, $\hat{S}_n$ is a step function that reduces to the usual empirical survival function in the absence of censoring.

Simultaneous confidence bands for the survival function have only recently become available in the literature. They are based on the derivation of the limiting distribution of $\hat{S}_n$ by Efron (1967), Breslow and Crowley (1974), Aalen (1976), and Meier (1976). The most well-known confidence band thus far is that of Hall and Wellner (1980). They developed a large-sample band that reduces to the Kolmogorov band in the absence of censoring. Nair (1981) developed two classes of goodness-of-fit tests and showed that one of these classes of tests can be inverted to obtain large-sample bands for S (also see Gillespie and Fisher 1979).

The purpose of this article is twofold. First, it shows that the usual pointwise confidence intervals based on Greenwood's variance formula and the normal approximation can be easily adapted to yield a large-sample simultaneous confidence band. This band has, in a certain sense, equal precision at each point of $S(x)$.

Second, the performances of three bands are compared: the Hall–Wellner (HW) band, the equal precision (EP) band, and a censored analog of Rényi's (1953) (R) band. These three bands are chosen because they are censored versions of the three most well-known bands used with complete data. The comparisons of the bands are based on their widths and the adequacy of the large-sample approximations and are carried out under various censoring models and degrees of censoring. The bands are also illustrated and compared by applying them to data from a mechanical-switch life test.

This article is oriented toward the user; therefore, it provides only the minimum technical details necessary in developing the asymptotic theory. Furthermore, the results here are presented in terms of obtaining confidence bands for the survival function. Evidently they are also applicable to testing goodness-of-fit hypotheses. For a recent survey on general goodness-of-fit tests with censored data, see Sections 10–15 of Doksum and Yandell (1983).

2. AN EQUAL PRECISION BAND

Let $X_1^0, \ldots, X_n^0$ be independent and identically distributed (iid) with continuous distribution function F and survival function $S = 1 - F$. Let $Y_1, \ldots, Y_n$ denote the corresponding arbitrary censoring variables. The observed, arbitrarily right-censored sample is (X_i, δ_i), $i = 1, \ldots, n$, where $X_i = \min(X_i^0, Y_i)$ and $\delta_i = 0$ if $X_i^0 > Y_i$ (censored), and $\delta_i = 1$ otherwise. Let R_i be the rank of X_i in the lexicographic ordering of $(X_i, 1 - \delta_i)$, $i = 1, \ldots, n$, and let $X_{(n)} = \max(X_1, \ldots, X_n)$. The Kaplan–Meier estimator is

$$\begin{aligned}\hat{S}_n(x) &= \prod_{\{i:\, X_i \le x\}} [(n - R_i)/(n - R_i + 1)]^{\delta_i}, & x \le X_{(n)} \\ &= 0, & x > X_{(n)}.\end{aligned} \tag{2.1}$$

An estimate of the variance of $\hat{S}_n(x)$ is given by $n^{-1}\hat{S}_n^2(x)\hat{\sigma}_n^2(x)$, where

$$\hat{\sigma}_n^2(x) = n \sum_{\{i:\, X_i \le x\}} \delta_i/[(n - R_i)(n - R_i + 1)]. \quad (2.2)$$

This estimate is based on Greenwood's (1926) variance formula, originally proposed in the context of grouped data. Kaplan and Meier (1958), Breslow and Crowley (1974), and Meier (1976) showed that under appropriate conditions, it is a consistent estimate of the asymptotic variance of $\hat{S}_n(x)$.

The usual large-sample pointwise confidence intervals for $S(x)$, based on (2.2) and the normal approximation, are

$$\hat{S}_n(x) \pm z_{\alpha/2} n^{-1/2} \hat{S}_n(x)\hat{\sigma}_n(x). \quad (2.3)$$

Here $z_{\alpha/2}$ is the $(1 - \alpha/2)$th standard normal quantile. Thomas and Grunkemeier (1975) examined how well the large-sample approximation in (2.3) works and compared it with other pointwise confidence intervals.

A simultaneous large-sample confidence band for S can be obtained essentially by replacing $z_{\alpha/2}$ in (2.3) by an appropriately larger critical value. But first some preliminaries are needed. The asymptotic distribution of the Kaplan–Meier estimator was studied by Breslow and Crowley (1974) under a random censoring model, where the censoring variables Y_i are iid with (right continuous) survival function H and are independent of the X_i^0's. Meier (1976) showed that the asymptotic theory developed under the random censoring model continues to hold, under an appropriate asymptotic framework, even when the Y_i's are fixed, arbitrary values. This article will consider only the random censoring model. Then the X_i's are also iid with survival function $L(x) = S(x)H(x)$.

Fix $0 < T < \infty$ so that $L(T) > 0$. Because of a technical difficulty with the limiting distribution of $\hat{S}_n$ (see the Appendix for details), simultaneous confidence bands can be obtained only for $0 \le x \le T$. Although I will not state it explicitly, it should be noted that all of the bands in this article are valid only for $0 \le x \le T$ for some T with $L(T) > 0$. In practice one can choose $T < T_n$, the largest uncensored observation, to ensure that $L(T) > 0$.

Let

$$\hat{K}_n(x) = \hat{\sigma}_n^2(x)/1 + \hat{\sigma}_n^2(x).$$

As will be seen, $\hat{K}_n(x)$ displays an important role in developing asymptotically distribution-free confidence bands with censored data. It behaves like an empirical distribution function: $0 \le \hat{K}_n(x) \le 1$ and $\hat{K}_n(x)$ is nondecreasing in x. When there is no censoring,

$$\hat{K}_n(x) \equiv \hat{F}_n(x) = 1 - \hat{S}_n(x),$$

the usual empirical distribution function (Hall and Wellner 1980). Let

$$K(x) = \sigma^2(x)/1 + \sigma^2(x),$$

where

$$\sigma^2(x) = \int_0^x [L(z)S(z)]^{-1}\, dF(z). \quad (2.4)$$

Then, as $n \to \infty$, $\hat{K}_n(x) \to K(x)$ a.s. uniformly in $x \in (0, T)$ (Hall and Wellner 1980).

To define the large-sample band, let $0 < a < b < 1$ be fixed. The Appendix shows that under a random censoring model,

$$P\{S(x) \in \hat{S}_n(x) \pm e_\alpha n^{-1/2} \hat{S}_n(x)\hat{\sigma}_n(x) \quad \forall x : a \le \hat{K}_n(x) \le b\} \to 1 - \alpha \quad (2.5)$$

as $n \to \infty$. Here $e_\alpha = e_\alpha(a, b)$ satisfies

$$P\left\{\sup_{a \le u \le b} \frac{|W^0(u)|}{[u(1-u)]^{1/2}} \le e_\alpha\right\} = 1 - \alpha, \quad (2.6)$$

where $W^0(u)$ is a tied-down Wiener process on (0, 1). Thus (2.5) provides a large-sample distribution-free band for S. To construct this band, one must choose values of a and b. It can be shown that the critical value $e_\alpha \to \infty$ as $a \to 0$ or $b \to 1$. So there is a trade-off between the width and the coverage of the band. In practice, the choices $a = .05$ or $.1$ and $b = \min(.9, \hat{K}_n(T_n))$ or $\min(.95, \hat{K}_n(T_n))$, where T_n is the largest uncensored observation, appear to be reasonable.

Borokov and Sycheva (1968) studied the distribution of the statistic in (2.6) and showed that for the one-sided version

$$P\left\{\sup_{a \le u \le b} \frac{W^0(u)}{[u(1-u)]^{1/2}} > x\right\}$$

can be well approximated for x in the upper tail by

$$A(x) = x \exp(-x^2/2) \log\left[\frac{(1-a)b}{a(1-b)}\right] \Big/ \sqrt{8\pi}.$$

One can obtain an excellent approximation for e_α from this by setting $A(e_\alpha) = \alpha/2$. Some critical values for different choices of α and $a = 1 - b$ are given in Table 2 (Section 5).

The band (2.5) differs from the pointwise intervals (2.3) only through the critical values. It is valid simultaneously for all $x \le T$ for which $a \le \hat{K}_n(x) \le b$. Like the pointwise confidence intervals, its width is proportional to its estimated standard deviation. In this sense, the band has equal precision at all of the x values for which it is valid. When there is no censoring, $\hat{K}_n(x) \equiv \hat{F}_n(x)$ so that the band reduces to

$$\hat{S}_n(x) \pm e_\alpha \sqrt{[\hat{S}_n(x)(1 - \hat{S}_n(x))]/n}.$$

3. HW AND R BANDS

The HW band—the censored version of the Kolmogorov band—and the R band—the censored version of the Rényi band—will now be defined. The HW band of Hall and Wellner (1980) is

$$\hat{S}_n(x) \pm h_\alpha n^{-1/2}\hat{S}_n(x)/\bar{K}_n(x), \tag{3.1}$$

where $\bar{K}_n = 1 - \hat{K}_n$. As mentioned before, the band is only valid for $0 \le x \le T$ with $L(T) > 0$. For complete samples this band reduces to the Kolmogorov band restricted to a finite interval. Hall and Wellner (1980) showed that the asymptotic critical value for the Kolmogorov band in complete samples provides an upper bound for the asymptotic critical value h_α in (3.1). Thus h_α can be obtained by setting $B(h_\alpha) = \alpha$, where

$$B(x) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2x^2),$$

which can be well approximated here by $2\exp(-2x^2)$.

The R band considered in this article is

$$\hat{S}_n(x) \pm r_\alpha n^{-1/2}\hat{S}_n(x). \tag{3.2}$$

It is valid for all x for which $0 \le \hat{K}_n(x) \le b$, where $0 < b < 1$ is fixed. An advantage of this band is that it is simple to compute, since it depends only on $\hat{S}_n(x)$. As will be seen later, this band is geared toward estimating the upper tail of the distribution and tends to be very wide in the other regions. As with the EP band, one must bound b away from 1. In fact, the critical value r_α is equal to $(b/(1-b))^{1/2}w_\alpha$, where w_α is the $(1-\alpha)$th quantile of the supremum of the absolute value of a Wiener process on (0, 1). So $r_\alpha \to \infty$ as $b \to 1$. For practical purposes the choices $b = \min(.8, \hat{K}_n(T_n))$ or $\min(.9, \hat{K}_n(T_n))$, where T_n is the largest uncensored observation, appear reasonable. From Feller (1971, p. 343), one sees that w_α can be obtained by taking $B(w_\alpha) = 1 - \alpha$, where

$$B(x) = \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k \exp(-\pi^2(2k+1)^2/8x^2)/(2k+1)$$

$$\simeq 4\Phi(x) - 2$$

for x in the upper tail and $\Phi(x)$ is the standard normal distribution function.

All three bands—EP, HW, and R—are obtained by inverting a test statistic of the form

$$\sup_{0<x<T} n^{1/2} \frac{|\hat{S}_n(x) - S_0(x)|\,\psi(\hat{K}_n(x))}{\hat{S}_n(x)/\bar{K}_n(x)}, \tag{3.3}$$

where $S(x) = S_0(x)$ is the hypothesis being tested. The EP band corresponds to $\psi(u) = [u(1-u)]^{-1/2}$ for $a \le u \le b$, and $= 0$ otherwise. The HW band is obtained by taking $\psi(u) \equiv 1$. For the R band, take $\psi(u) = (1-u)^{-1}$ for $0 \le u \le b$, and $= 0$ otherwise. This band is a censored analog Rényi's band that gives more weight to the upper tail.

When there is no censoring, (3.3) reduces to

$$\sup_{0<x<T} n^{1/2}|\hat{F}_n(x) - F_0(x)|\,\psi(\hat{F}_n(x)), \tag{3.4}$$

where $\hat{F}_n = 1 - \hat{S}_n$ and $F_0 = 1 - S_0$. The usual bands in complete samples are based on a statistic similar to (3.4) but with $\psi(F_0(x))$ in place of $\psi(\hat{F}_n(x))$. It is also possible to develop bands from the statistic (3.3) with $\hat{S}(x)$ in the denominator replaced by $S_0(x)$. These bands will not be symmetric about $\hat{S}_n(x)$. Furthermore, they do not reduce in the absence of censoring to the nice form in (3.4). But the two forms are asymptotically equivalent. The class of procedures discussed in Nair (1981) includes these asymptotically equivalent variants of the EP and R bands, but their properties were not studied in detail.

4. AN ILLUSTRATIVE EXAMPLE

Here is an illustration of the use of these bands on data from a mechanical-switch life test. The data in Table 1 are the failure times (measured in millions of operations) of 40 randomly selected mechanical switches. They were tested in a facility with 40 test positions. Three of the test positions became available much later than the others, so the three switches tested at these positions were still operating at the termination of the test. The corresponding censored observations are indicated by the code 0 in Table 1. There were also two possible modes of failure—essentially two different springs—for these switches. When a switch failed, its mode of failure was noted. These modes are indicated by the codes A and B in Table 1. The two springs were identical in construction but subjected to different stress levels, so the life distributions of the two failure modes are likely to be different.

There are many quantities of interest here, so there are many ways in which this data set can be analyzed. For example, one may study the life distribution of the switch only, in which case the failure modes can be

Table 1. *Failure Times (in millions of operations) for a Mechanical-Switch Life Test*

Time	*Mode*	*Time*	*Mode*
1.151	B	2.119	B
1.170	B	2.135	A
1.248	B	2.197	A
1.331	B	2.199	B
1.381	B	2.227	A
1.499	A	2.250	B
1.508	B	2.254	A
1.534	B	2.261	B
1.577	B	2.349	B
1.584	B	2.369	A
1.667	A	2.547	A
1.695	A	2.548	A
1.710	A	2.738	B
1.955	B	2.794	A
1.965	A	2.883	0
2.012	B	2.883	0
2.051	B	2.910	A
2.076	B	3.015	A
2.109	A	3.017	A
2.116	B	3.793	0

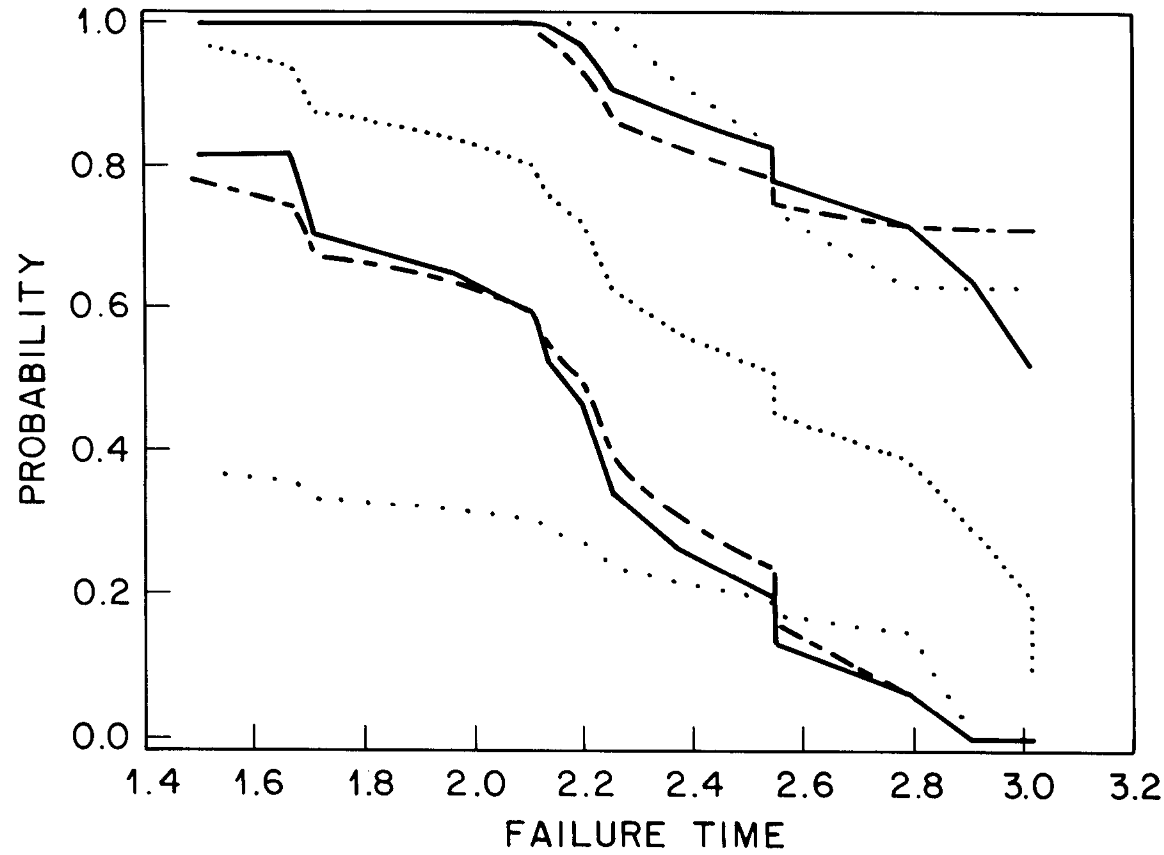

Figure 1. *$\hat{S}_n$ and 90% confidence bands for mechanical switch data: $\hat{S}_n$ ⋯⋯; Hall–Wellner (HW) band – – – –; equal precision (EP) band ————; and censored analog of Rényi's (R) band · · · ·.*

ignored. One may also be interested in the life distribution of each failure mode separately or in comparing the two distributions. Here the occurrence of a failure from mode A prevents us from observing a failure from mode B. So every occurrence time for failure mode A is a right censoring time for failure mode B and vice versa. This is an example of a situation for which the random censoring model described in Section 2 applies, although the independence of the two failure modes may be questionable.

Here attention is restricted to nonparametric estimation of the distribution for failure mode A. This model is chosen because for mode B there is a lot of censoring in the upper tail. The Kaplan–Meier (1958) estimate $\hat{S}_n$ for mode B goes down to only about .38, so the relative performances of the bands in the upper tail will not be seen.

Figure 1 shows $\hat{S}_n$ and the three 90% simultaneous confidence bands for the survival function of failure mode A. The last observed failure time is 3.017 million operations, for which $\hat{S}_n = .097$. In computing these bands, T was chosen to be 3.015 million operations, the second largest uncensored observation. The EP band shown corresponds to $a = 1 - b = .05$ with critical value $e_\alpha = 2.91$. This choice of a and b excluded the smallest and largest observations, but the band was extended to cover these values by using the monotonicity of S and by using the obvious bounds of 0 and 1. The HW band shown was computed with critical value $h_\alpha = 1.22$, and the band was extended, again, to cover the last observation. The upper band was not monotonic; it was increasing for the last three values. The band at these values was replaced by the bound from the fourth largest observation. This explains the flatness of the upper band on the right side. This lack of monotonicity can also occur for the EP band. The R band shown was computed with $b = .2$ and critical value $r_\alpha = 3.92$. This excluded the last three observations, but the band was extended, again, to cover these values.

Figure 1 shows that the R band performs poorly. It is narrower than the other two bands in a small region near the upper tail at the expense of being very wide in the other parts of the distribution. The HW and EP bands are competitive. The EP band is narrower in the tails, whereas the HW band performs better in the middle. This is qualitatively similar to their relative performances in complete samples.

5. ADEQUACY OF LARGE-SAMPLE APPROXIMATIONS

This section examines how well the large-sample distribution-free bands do in finite sample situations. A Monte-Carlo simulation is used to determine the

Table 2. *Asymptotic Critical Values and the Achieved Confidence Levels in the Uncensored Case for n = 200*

Confidence Band	$\alpha = .01$		$\alpha = .05$		$\alpha = .10$		$\alpha = .20$	
	Critical Value	Achieved Level	Critical Value	Achieved Level	Critical Value	Achieved Level	Critical Value	Achieved Level
HW	1.63	.009	1.36	.05	1.22	.10	1.07	.19
EP: $a = 1 - b = .01$	3.81	.07	3.31	.14	3.08	.19	2.81	.26
EP: $a = 1 - b = .05$	3.68	.023	3.16	.07	2.91	.11	2.62	.19
EP: $a = 1 - b = .1$	3.59	.014	3.06	.05	2.79	.10	2.48	.19
R: $b = .8$	5.62	.015	4.48	.06	3.92	.11	3.70	.21

NOTE: HW = Hall-Wellner. EP = equal precision. R = censored analog of Rényi's.

achieved levels of the bands under several censoring situations. To isolate the effects of censoring, one must first examine how the large-sample approximations work in complete samples. All of the simulation results reported in this article were computed on the CRAY-1 computer at the Bell Laboratories Murray Hill Computer Center, using the uniform random number generator in the PORT Mathematical Subroutine Library.

Table 2 gives the asymptotic critical values for the HW band; the EP bands with $a = 1 - b = .01$, .05, and .1; and the R band with $b = .8$ for four α-values—.01, .05, .1, and .2. Table 2 also gives the confidence levels achieved with these critical values in complete samples of size $n = 200$. These results, as well as all others in this section, are based on 10,000 simulations, so the standard errors involved are less than .005. In fact if the true error is $\simeq .01$, the standard errors are $\simeq .001$. So some of the smaller values in Table 2 are extended to three decimal places.

As noted earlier, the HW band reduces to the Kolmogorov band in complete samples, where its performance has been studied extensively. It is known that the asymptotic critical values provide a good approximation for $n = 200$, and this is confirmed by the values in Table 2. For the EP band with $a = 1 - b = .01$, the asymptotic critical values are too small. The approximation improves as $a = 1 - b$ increases, and it is in fact good for $a = 1 - b = .1$. In general the approximation also gets better as α increases. For the R band with $b = .8$, the asymptotic critical values are slightly small, although the approximation is reasonable overall.

Based on their performances in complete samples, I selected the HW band, the EP bands with $a = 1 - b = .05$ and .1, and the R band with $b = .8$ and examined how they do in the presence of censoring. This was done under the random censoring model described in Section 2. In fact I considered three different random censoring models: (a) the proportional hazards or Koziol–Green (KG) model, where $H = S^{\theta}$ for some θ; (b) one where S = standard exponential and H = uniform $(0, b)$; and (c) one where $S = \exp(-x^2)$ (Weibull) and H = uniform $(0, b)$. The parameters θ and b determine the degree of censoring. Three different censoring probabilities—.25, .5, and .75—and three different sample sizes—25, 50, and 100—are considered.

Under the KG model, the distribution of the test statistic (3.3) does not depend on the underlying distribution S if the hypothesis $S(x) = S_0(x)$ holds. This is easily seen by transforming both the X_i^0's and the Y_i's in the random censoring model by $F_0(x) = 1 - S_0(x)$. Hence the test statistic (3.3) is distribution free even in finite samples. The distribution depends only on the (censoring) parameter θ: for $\theta = 0$ there is no censoring, for $\theta = 1$ there is 50% censoring, and for $\theta = \infty$ there is complete censoring. Under the other two censoring models considered, the test statistic (3.3) is not distribution free. The KG model may be reasonable in the mechanical-switch life test example described in Section 4.

Tables 3–5 give the confidence levels obtained from the asymptotic critical values in finite samples under three different degrees of censoring. Overall, the asymptotic critical values provide a reasonable approximation for the HW and EP bands, even with sample size $n = 25$. The approximations are in fact good for the HW band and the EP band with $a = 1 - b = .1$, especially when there is only 25% censoring. Furthermore, there is essentially no difference in the levels among the three censoring models with 25% censoring. For the R band, however, the asymptotic critical values provide a poor approximation.

The quality of the approximations worsens as the censoring proportion increases. With increased censoring, the exact critical values in finite samples get larger. Furthermore, differences begin to emerge among the censoring models: The finite-sample critical values in the KG model are much larger than those from the exponential model with uniform censoring, and the Weibull model with uniform censoring is between the two. One reason for this is that with the uniform censoring distribution on $(0, b)$, the survival function $S(x)$ cannot be estimated for $x > b$; for the exponential distribution, $S(b) = .203$ with 50% censoring and $S(b) = .546$ with 75% censoring, so only a portion of the distribution can be estimated. For the

Table 3. *Achieved Confidence Levels With 25% Censoring*

Confidence Band	Asymptotic Level	KG n = 25	KG 50	KG 100	Exponential/Uniform 25	Exponential/Uniform 50	Exponential/Uniform 100	Weibull/Uniform 25	Weibull/Uniform 50	Weibull/Uniform 100
HW	.01	.009	.010	.008	.009	.008	.008	.009	.010	.007
	.05	.04	.05	.05	.04	.05	.04	.04	.04	.05
	.10	.09	.09	.09	.09	.09	.09	.09	.09	.09
	.20	.18	.19	.19	.18	.19	.19	.18	.18	.19
EP: $a = 1 - b = .05$	.01	.04	.03	.03	.04	.03	.03	.04	.04	.03
	.05	.08	.08	.07	.08	.08	.07	.09	.08	.08
	.10	.12	.12	.12	.12	.12	.12	.14	.12	.12
	.20	.19	.19	.20	.19	.19	.19	.20	.20	.19
EP: $a = 1 - b = .10$	.01	.022	.021	.014	.023	.021	.014	.024	.023	.015
	.05	.06	.06	.05	.07	.06	.05	.06	.06	.08
	.10	.10	.11	.10	.10	.11	.10	.11	.11	.12
	.20	.18	.19	.19	.17	.19	.19	.18	.19	.19
R: $b = .80$	.01	.08	.05	.03	.08	.05	.03	.08	.05	.03
	.05	.15	.11	.08	.15	.10	.08	.15	.11	.08
	.10	.21	.17	.14	.20	.16	.13	.21	.17	.13
	.20	.31	.27	.24	.29	.25	.23	.31	.28	.23

NOTE: KG = Koziol–Green model. See Note to Table 2 for other abbreviations.

KG model, however, the entire distribution can be estimated even with extremely heavy censoring.

The relative performances of the two EP bands suggest that the asymptotic approximations work better as $a = 1 - b$ gets larger. This is similar to the situation in complete samples. As in complete samples, for $a = .05$ the approximation is poorer for smaller α-values. Comparing the performance of the EP bands in complete samples (Table 2) reveals that a large part of the error in the approximations in Tables 3–5 can be ascribed to the difference between the asymptotic and finite-sample critical values in complete samples.

I repeated the simulations for the EP bands in Tables 3–5 with the critical values from complete samples for $n = 200$ (obtained by simulation) and found that the new confidence levels were comparable to those of the HW band.

Overall, the approximations improve with sample size, and this is particularly evident for the R band, where in small samples the approximations do not work well.

These results suggest the following conclusions in general situations: When there is light censoring (25% or less), the asymptotic critical values provide a good approximation for the HW and EP bands even in small samples, irrespective of the censoring mechanism. When there is heavier censoring, the approximations are still reasonable although not as good. In these situations, the approximations are better if $\hat{S}_n(T_n)$ (where T_n is the largest uncensored observation) is not close to zero than if $\hat{S}_n(T_n)$ is close to zero. In a particu-

Table 4. *Achieved Confidence Levels With 50% Censoring*

Confidence Band	Asymptotic Level	KG n = 25	KG 50	KG 100	Exponential/Uniform 25	Exponential/Uniform 50	Exponential/Uniform 100	Weibull/Uniform 25	Weibull/Uniform 50	Weibull/Uniform 100
HW	.01	.023	.023	.016	.012	.017	.012	.019	.018	.012
	.05	.07	.07	.06	.05	.05	.06	.06	.06	.05
	.10	.12	.12	.11	.09	.10	.10	.11	.11	.10
	.20	.20	.21	.20	.16	.19	.20	.19	.20	.19
EP: $a = 1 - b = .05$	.01	.05	.04	.03	.04	.03	.022	.05	.04	.03
	.05	.10	.09	.07	.08	.07	.07	.11	.09	.07
	.10	.14	.13	.12	.12	.11	.11	.14	.13	.12
	.20	.21	.20	.20	.17	.17	.17	.21	.21	.19
EP: $a = 1 - b = .10$	.01	.03	.029	.021	.024	.023	.017	.03	.03	.018
	.05	.08	.08	.06	.06	.06	.06	.08	.07	.04
	.10	.12	.12	.11	.10	.10	.10	.11	.12	.11
	.20	.19	.20	.20	.16	.18	.18	.18	.20	.19
R: $b = .80$	.01	.10	.07	.05	.07	.06	.04	.09	.07	.04
	.05	.19	.15	.12	.14	.12	.10	.18	.14	.10
	.10	.25	.22	.18	.20	.18	.16	.25	.20	.16
	.20	.36	.33	.29	.30	.28	.25	.35	.30	.25

NOTE: See Notes to Tables 2 and 3.

Table 5. *Achieved Confidence Levels for n = 100 With 75% Censoring*

Confidence Band	Asymptotic Level	KG	Exponential/ Uniform	Weibull/ Uniform
HW	.01	.05	.009	.022
	.05	.10	.03	.06
	.10	.15	.07	.10
	.20	.24	.13	.19
EP: $a = 1 - b = .05$	.01	.05	.026	.03
	.05	.10	.06	.08
	.10	.14	.08	.11
	.20	.21	.13	.17
EP: $a = 1 - b = .10$	.01	.04	.018	.027
	.05	.09	.05	.07
	.10	.14	.08	.10
	.20	.21	.13	.17
R: $b = .80$	.01	.08	.022	.03
	.05	.16	.07	.09
	.10	.22	.12	.15
	.20	.32	.21	.24

NOTE: See Notes to Tables 2 and 3.

lar application, $\hat{K}_n(T_n)$ may be much smaller than one, so the EP or R bands can be constructed with $b < \hat{K}_n(T_n)$. For these bands the accuracy of the approximations gets better as b decreases. For the HW band, the accuracy is not likely to get better unless $\varepsilon = 1 - \hat{K}_n(T_n)$ is larger than .25. For such large values of ε, the asymptotic critical values from the Kolmogorov band will be too conservative, and the exact asymptotic critical values given in Hall and Wellner (1980) should be used.

Koziol (1980) observed that for the goodness-of-fit test corresponding to the HW band, the asymptotic critical values do not provide a good approximation in small samples. His test statistic, however, is slightly different than that proposed by Hall and Wellner (1980)—it has S_0 rather than $\hat{S}_n$ in the denominator of (3.3). His conclusions are also based on only 1,000 simulations. My results indicate that the asymptotic critical values for the HW band provide a good approximation with sample sizes as small as $n = 25$.

6. COMPARISON OF THE WIDTHS

In this section, the EP band and the R band are compared with the HW band in terms of the ratio of their limiting squared widths. To get the limiting widths, let $K(x)$ be defined by (2.4). Then it can be shown that the ratio of the limiting squared widths of

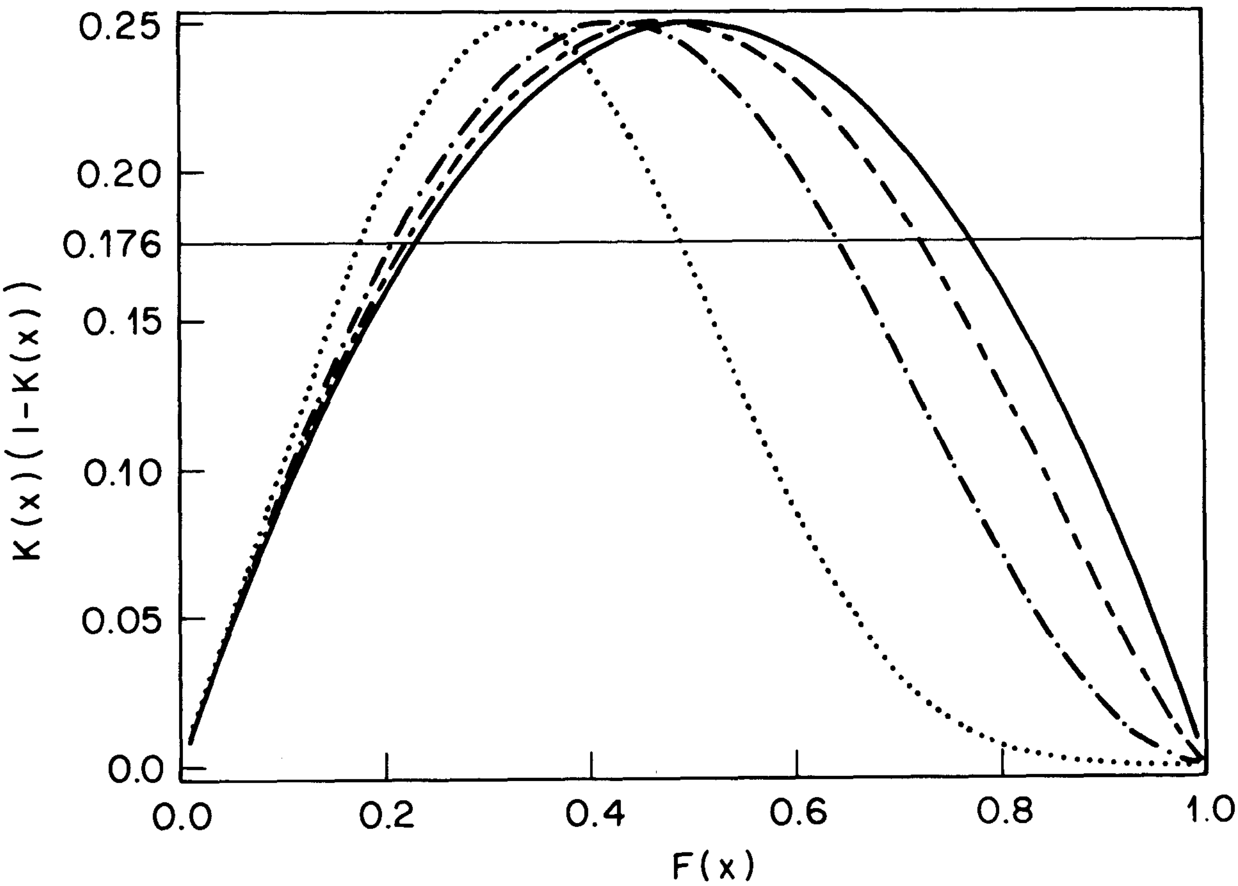

Figure 2. *Comparison of the widths of the EP band and the HW band under the KG model: no censoring* ———; *25% censoring* — — — –; *50% censoring* —·—·—; *and 75% censoring* ······. *The horizontal line corresponds to* $y = (h_\alpha/e_\alpha)^2 = .176$ *for* $\alpha = .1$, $a = 1 - b = .05$. (*See Figure 1 caption for explanation of abbreviations.*)

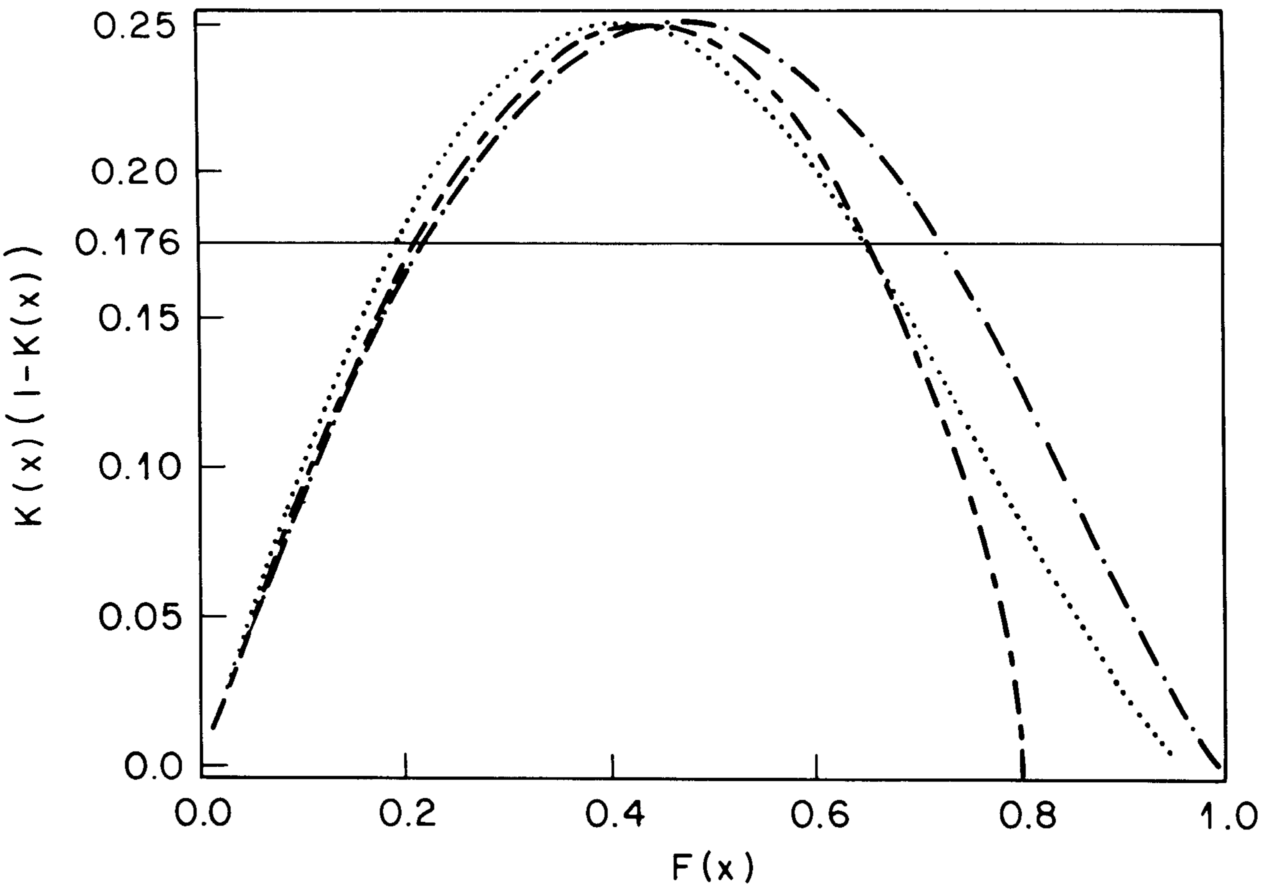

Figure 3. *Comparison of the widths of the EP band and the HW band under different censoring models with 50% censoring : KG —·—·—; exponential/uniform — — — –; and Weibull/uniform ······. The horizontal line corresponds to* $y = (h_\alpha/e_\alpha)^2 = .176$ *for* $\alpha = .1, a = 1 - b = .05$. *(See Figure 1 caption for explanation of abbreviations.)*

the EP band to the HW band is

$$w_1(x) = (e_\alpha/h_\alpha)^2 K(x)(1 - K(x)).$$

Since $K(x)$ reduces to $F(x)$ in the absence of censoring, $w_1(x)$ provides the usual comparison between the EP band and the Kolmogorov band in complete samples. Note that $w_1(x)$ (which measures the asymptotic relative efficiency of the HW band to the EP band at $x) \to 0$ as $x \to 0$ or ∞.

Figure 2 provides plots of $K(x)(1 - K(x))$ under the KG model for four different censoring proportions—0, .25, .5, and .75. The x axis is scaled in terms of $t = F(x)$. For the uncensored case, therefore, the plotted curve is just $t(1 - t)$. To obtain a comparison between the HW band and the EP band for a particular α and choice of a and b, one can draw a horizontal line at $y = (h_\alpha/e_\alpha)^2$; for x values where the curve is above the line, the HW band is narrower, and for the other values the EP band is narrower.

The horizontal line shown in Figure 2 is at $y = (h_\alpha/e_\alpha)^2 = .176$, and it corresponds to $\alpha = .1$ and $a = 1 - b = .05$. In this case, the HW band is narrower for the x values for which $F(x) \in (.23, .77)$. As the censoring proportion increases, the relative performance of the EP band to the HW band gets better. For example, with 25% censoring the HW band is better only in the region $F(x) \in (.22, .72)$. With 50% censoring, this region changes to $F(x) \in (.20, .64)$, and with 75% censoring it changes to $F(x) \in (.18, .48)$. Similar comparisons can be made from Figure 2 for other values of α and choices of a and b for the EP band. Of course the comparisons are valid only in the region $K(x) \in (a, b)$.

Figure 3 shows how the relative performance of the EP and HW bands change as the censoring models change. The curves plotted in Figure 3 are $K(x)(1 - K(x))$ with 50% censoring under three different models: the KG model, the exponential distribution with uniform censoring, and the Weibull distribution with uniform censoring. The horizontal line shown in Figure 3 is again $y = .176$. Compared to the KG model, the relative performance of the EP band is much better in the exponential case and slightly better in the Weibull case. This is because for the exponential distribution with uniform censoring, $K(x) \to 1$ rapidly and $= 1$ when $F(x) = .8$. On the other hand, for the KG model $K(x) = 1$ only when $F(x) = 1$.

To summarize the comparisons between the

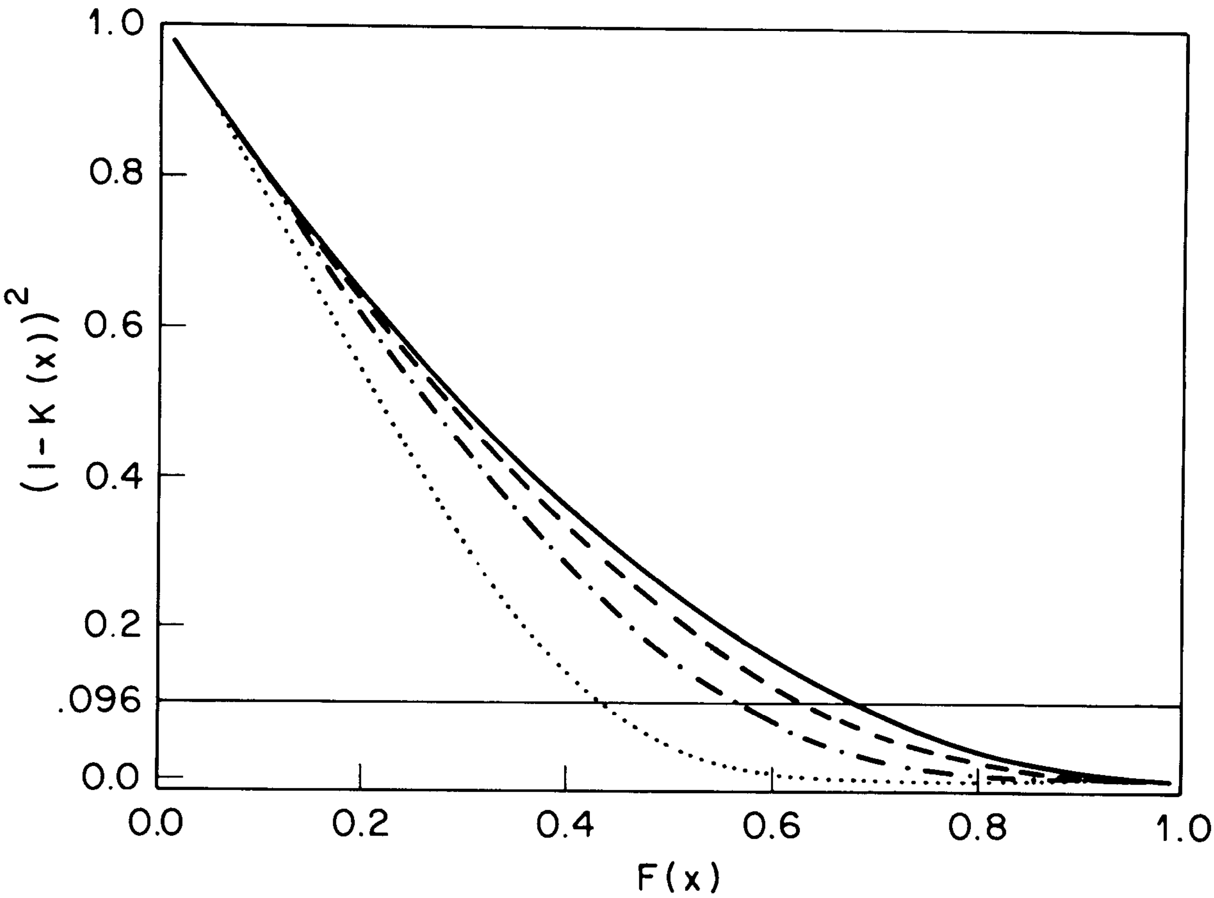

Figure 4. *Comparison of the widths of the R band to the HW band under the KG model: no censoring* ————; *25% censoring* — — — —; *50% censoring* —·—·—; *and 75% censoring* ······. *The horizontal line corresponds to* $y = (h_\alpha/r_\alpha)^2 = .176$ *for* $\alpha = .1, b = .8$. (*See Figure 1 caption for explanation of abbreviations.*)

two bands by a single number, a good measure is $\int w_1(x)\, d\tilde{F}(x)$, where $\tilde{F}$ is the probability measure obtained by restricting F to the region of comparison $K(x) \in (a, b)$. For the KG model and for $\alpha = .05$ and $a = 1 - b = .05$, the value of this measure turns out to be 1.15 with no censoring, 1.09 with 25% censoring, .98 with 50% censoring, and only .74 with 75% censoring. As was seen earlier, the EP band performs better as the censoring proportion increases.

Now turn to the comparison of the R band with the HW band. It is easily shown that the ratio of the squared limiting widths of the two bands is

$$w_2(x) = (r_\alpha/h_\alpha)^2(1 - K(x))^2.$$

Figure 4 gives plots of $(1 - K(x))^2$ for four censoring proportions—0, .25, .5, and .75—under the KG model. The horizontal line shown is $y = (h_\alpha/r_\alpha)^2 = .096$, and it corresponds to $\alpha = .1$ and $b = .8$ for the R band. The R band is better in a small region in the upper tail, but it performs poorly in the other parts of the distribution. Like the EP band, however, the relative performance of the R band to the HW band improves as the censoring proportion increases.

A note of caution is necessary in interpreting the asymptotic results in this section. It is possible that in small samples all three of the upper bands are larger than one in the lower tail of the distribution, and all three of the lower bands are smaller than zero in the upper tail of the distribution. So even though the EP band is narrower than the HW band in the tails, the practical implication of this may be more limited. See Figure 1 on the mechanical-switch life test, for example.

7. CONCLUDING REMARKS

The HW and EP bands are competitive, with the former being narrower in the middle and the latter being narrower in the tails. The relative performance of the EP band to the HW band gets better with censoring. The asymptotic critical values provide reasonable approximations in finite samples for both bands. The R band, however, seems to perform poorly.

Like the pointwise confidence intervals, the EP band has the same (estimated) precision at each point of x. This is an intuitively appealing property for a simultaneous confidence band. If the band fails to include the true survival function, then it is as likely to fail at one point as at any other point. It is also easier to introduce and explain the EP band to nonstatisticians because it is just the simultaneous version of the commonly used pointwise confidence intervals. These remarks, of course, apply equally well to complete data situations. Despite these reasons, however, the EP band does not seem to be as popular as the Kolmogorov band. There are several possible reasons for this. Perhaps the most important reason is that the EP band is valid only in a restricted region $a \le \hat{K}_n(x) \le b$, and the user has to make subjective choices for the values of a and b. From practical experience, however, the choices $a = 1 - b = .05$ or $.1$ seem to provide a reasonable compromise between the width and the coverage of the band. Although it is true that the Kolmogorov band is unrestricted and is valid even in the tails, it turns out to be rather wide in the tails of the distribution and, hence, not very useful in the tail region. A second criticism of the EP band is that the distribution of the test statistic has not been as extensively studied and tabulated, so it is harder to obtain critical values. There have been some recent papers (Noe 1972, Niederhausen 1981, and Kotel'nikova and Chmaladze 1983), however, that have provided extensive critical values in the uncensored case. In any event, with the present day computing resources, it would be relatively easy to obtain excellent approximations for any required distribution.

Finally note that the bands in this paper are all based on the test statistic (3.3), which measures a maximum weighted distance between $\hat{S}_n$ and S. An alternative approach is to transform $\hat{S}_n$ and S and compute the maximal difference between these transformed values. In the uncensored case, Michael (1983) showed that the arc sine–square root transformation does well. Although asymptotically equivalent to the EP band (which follows from a Taylor series expansion), the procedure based on the arc sine–square root transformation seems to perform better in small samples. It would be interesting to investigate a similar use of transformations with multiply right-censored data.

APPENDIX

I now discuss the limiting distribution of the Kaplan–Meier estimator and show that all of the bands discussed in the article are asymptotically distribution free. This is done by showing that under the hypothesis $S(x) = S_0(x)$, the limiting distribution of the test statistic (3.3) does not depend on $S(x)$. The result is shown under the random censoring model described in Section 2. But Meier's (1976) results imply that it also holds under a fixed censorship model.

Let

$$T_L = \inf \{t : L(T) = 0\},$$

and consider $T < T_L$. Let $K(x)$ be defined as in Section 2 and $\bar{K} = 1 - K$. The following theorem is from Breslow and Crowley (1974), and the form in which it is stated is from Hall and Wellner (1980). Breslow and Crowley (1974) proved the result when the censoring distribution H is continuous, but as indicated by Gill (1983), this assumption is not necessary.

Theorem: Under the random censoring model, as $n \to \infty$,

$$n^{1/2}\left(\frac{\hat{S}_n(x) - S(x)}{S(x)/\bar{K}(x)}\right) \xrightarrow{\mathscr{D}} W^0(K(x)), \qquad 0 \le x \le T,$$

where $W^0(u)$ is a tied-down Wiener process on (0, 1).

Let $\psi(u)$ be a nonnegative function on (0, 1) that is continuous a.s. with respect to the Lesbesque measure. Then by Theorem 5.1 in Billingsley (1968),

$$\sup_{0 \le x \le T} n^{1/2}\left|\frac{\hat{S}_n(x) - S(x)}{S(x)/\bar{K}(x)}\right| \psi(K(x))$$

$$\xrightarrow{\mathscr{D}} \sup_{0 \le x \le T} |W^0(K(x))| \, \psi(K(x))$$

$$= \sup_{0 \le u \le K(T)} |W^0(u)| \psi(u).$$

The last equality follows from the fact that $0 \le K(x) \le 1$ and $K(x)$ is continuous on $[0, T]$. By Theorem 5.5 in Billingsley (1968) and the fact that $\hat{S}_n(x) \to S(x)$ a.s. and $\hat{K}_n(x) \to K(x)$ a.s. uniformly in $x \in [0, T]$, this result continues to hold if $S(x)$ and $\bar{K}(x)$ are replaced in the denominator of the test statistic by $\hat{S}_n(x)$ and $\bar{K}_n(x)$, and $\psi(K(x))$, by $\psi(\hat{K}_n(x))$.

So except for $K(T)$, the limiting distribution of the test statistic (3.3) does not depend on $S(x)$. For the EP band, the limiting distribution is that of

$$\sup_{a \le u \le b} |W^0(u)| \, [u(1 - u)]^{-1/2},$$

if b is chosen to be less than $K(T)$. Similarly for the R band, the limiting statistic is

$$\sup_{0 \le u \le b} |W^0(u)| \, (1 - u)^{-1}.$$

For the HW band with $\psi(u) \equiv 1$, the critical values from the statistic

$$\sup_{0 \le u \le 1} |W^0(u)|$$

provide an upper bound. This is reasonable unless $K(T)$ is much smaller than one. In this case the exact critical values provided in Hall and Wellner (1980) should be used.

These bands are valid only for $x \le T < T_L$, and typically in applications one will take $T < T_n$, the largest uncensored observation. Gill (1983) showed that the HW band is valid for

$$0 \leq x \leq T_n \quad \text{if} \quad \int_0^{T_L} [H(t)]^{-1} \, dF(t) < \infty.$$

Of course there is no easy way of checking this condition in nonparametric situations. Whether the EP and R bands can be similarly extended when $K(T_L^-) < 1$ is unknown. Such an extension would allow the bands used with the truncation case (censoring distribution H degenerate) to be treated as a special case of the results in this article.

[*Received October 1983. Revised February 1984.*]

REFERENCES

AALEN, O. (1976), "Nonparametric Inference in Connection With Multiple Decrement Models," *Scandinavian Journal of Statistics*, 3, 15–27.

BILLINGSLEY, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

BOROKOV, A. A., and SYCHEVA, N. M. (1968), "On Asymptotically Optimal Nonparametric Criteria," *Theory of Probability and Its Applications*, 13, 359–363.

BRESLOW, N., and CROWLEY, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimate Under Random Censorship," *Annals of Statistics*, 2, 437–453.

DOKSUM, K. A., and YANDELL, B. S. (1983), "Tests for Exponentiality," Technical Report 24, University of California, Berkeley, Dept. of Statistics.

EFRON, B. (1967), "The Two-Sample Problem With Censored Data," *Proceedings of the Fifth Berkeley Symposium*, 4, 831–853.

FELLER, W. (1971), *An Introduction to Probability Theory and Its Applications* (Vol. 2, 2nd ed.), New York: John Wiley.

GILL, R. (1983), "Large Sample Behavior of the Product-Limit Estimator on the Whole Line," *Annals of Statistics*, 11, 49–58.

GILLESPIE, M. J., and FISHER, L. (1979), "Confidence Bands for the Kaplan–Meier Survival Estimate," *Annals of Statistics*, 7, 920–924.

GREENWOOD, M. (1926), "The Natural Duration of Cancer," *Reports on Public Health and Medical Subjects*, Vol. 33, London: H. M. Stationery Office.

HALL, W. J., and WELLNER, J. A. (1980), "Confidence Bands for a Survival Curve," *Biometrika*, 67, 133–143.

KAPLAN, E. L., and MEIER, P. (1958), "Nonparametric Estimation From Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.

KOTEL'NIKOVA, V. F., and CHMALADZE, E. V. (1983), "On Computing the Probability of an Empirical Process Not Crossing a Curvilinear Boundary," *Theory of Probability and Its Applications*, 27, 640–648.

KOZIOL, J. A. (1980), "Goodness-of-Fit Tests for Randomly Censored Data," *Biometrika*, 67, 693–696.

MEIER, P. (1976), "Estimation of the Distribution From Incomplete Observations," in *Perspectives in Probability and Statistics*, London: Applied Probability Trust, 67–88.

MICHAEL, J. A. (1983), "The Stabilized Probability Plot," *Biometrika*, 70, 11–17.

NAIR, V. N. (1981), "Plots and Tests for Goodness of Fit With Randomly Censored Data," *Biometrika*, 68, 99–103.

NIEDERHAUSEN, H. (1981), "Tables of Significance Points for the Variance Weighted Kolmogorov Statistic," Technical Report 298, Stanford University, Dept. of Statistics.

NOE, M. (1972), "The Calculation of Distributions of Two-Sided Kolmogorov–Smirnov Type Statistics," *Annals of Mathematical Statistics*, 43, 58–64.

RÉNYI, A. (1953), "On the Theory of Order Statistics," *Acta Mathematicarum Academy Scientiarum, Hungary*, 4, 191–231.

THOMAS, D. R., and GRUNKEMEIER, G. L. (1975), "Confidence Interval Estimation of Survival Probabilities for Censored Data," *Journal of the American Statistical Association*, 70, 865–871.