Exact Nonparametric Confidence Bands for the Survivor Function

Abstract: A method to produce exact simultaneous confidence bands for the empirical cumulative distribution function that was first described by Owen, and subsequently corrected by Jager and Wellner, is the starting point for deriving exact nonparametric confidence bands for the survivor function of any positive random variable. We invert a nonparametric likelihood test of uniformity, constructed from the Kaplan–Meier estimator of the survivor function, to obtain simultaneous lower and upper bands for the function of interest with specified global confidence level. The method involves calculating a null distribution and associated critical value for each observed sample configuration. However, Noe recursions and the Van Wijngaarden–Decker–Brent root-finding algorithm provide the necessary tools for efficient computation of these exact bounds. Various aspects of the effect of right censoring on these exact bands are investigated, using as illustrations two observational studies of survival experience among non-Hodgkin’s lymphoma patients and a much larger group of subjects with advanced lung cancer enrolled in trials within the North Central Cancer Treatment Group. Monte Carlo simulations confirm the merits of the proposed method of deriving simultaneous interval estimates of the survivor function across the entire range of the observed sample.

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. It was begun while the author was visiting the Department of Statistics, University of Auckland, and completed during a subsequent sojourn at the Medical Research Council Biostatistics Unit in Cambridge. The support of both institutions, in addition to that of NSERC and the University of Waterloo, is greatly appreciated.

Keywords: exact simultaneous confidence bands, survivor function, Kaplan–Meier estimator, Noe recursions, right censoring

*Corresponding author: David Matthews, Statistics & Actuarial Science, Faculty of Mathematics, University of Waterloo, Waterloo, ON N2L 3G1, Canada, E-mail: dematthews@uwaterloo.ca

1 Introduction

Owen [1] described a method of deriving exact, nonparametric, likelihood-based confidence bands for the continuous cumulative distribution function (cdf), \( F(x) \), based on the empirical cdf (ecdf) estimator. His approach involves inverting a nonparametric likelihood test of uniformity first described by Berk and Jones [2], using the recursive algorithm of Noé [3] to calculate joint probabilities from the exact null distribution of the test statistic for all sample sizes up to 1,000. An attractive feature of Owen’s method is that the Berk–Jones test statistic on which his approach relies is more efficient (in Bahadur’s sense) than any weighted Kolmogorov–Smirnov test statistic at any alternative to the uniform distribution. Jager and Wellner [4] pointed out a minor error in Owen’s derivations and extended his results to a similar family of confidence bands for \( F(x) \) based on the reversed Berk–Jones statistic. Jager and Wellner [5] subsequently introduced a new family of goodness-of-fit tests that includes both the Berk–Jones and reversed Berk–Jones statistics as special cases, as well as other familiar criteria such as the supremum version of the Anderson–Darling statistic. In Section 3 of...
their article, they outlined the null distribution theory for the complete family of statistics and remarked that calculations using the exact distributions are feasible for sample sizes up to \( n = 3,000 \) via Noé’s recursions and possibly up to \( n = 10^4 \) via the recursion of Khmaladze and Shinjikashvili [6]. More recently, Frey [7] has proposed a different approach that involves choosing the lower and upper bounds of confidence bands for \( F(x) \) to minimize a particular weighted average width criterion, and Xu et al. [8] described a computational algorithm for confidence bands that was similarly motivated.

To date, no article in the literature has addressed the related problem of deriving exact, nonparametric, likelihood-based confidence bands for the survivor function, \( F(x) \), based on the familiar Kaplan–Meier or product-limit estimator; see Kaplan and Meier [9]. Of course, in the absence of right censoring, a simple linear transformation of any of the above-mentioned confidence bands for the distribution function would suffice. However, right censoring is frequently observed in biostatistical applications, and this potential complication in the observed sample was not considered in any of these previous articles concerning confidence bands for the distribution function. Gillespie and Fisher [10], Hall and Wellner [11], Nair [12], Hollander and Peña [13] and Hollander et al. [14] all review the construction of confidence bands for \( F(x) \) that rely on the associated asymptotic theory for large sample approximations. However, an asymptotic result is not the focus of our interest.

In the balance of this article, we show how to adapt the Owen–Jager–Wellner method to calculate exact, nonparametric, likelihood-based confidence bands for \( F \), starting from the Kaplan–Meier estimator and inverting a modified Berk–Jones statistic. Of course, since these bands are likelihood-based, they inherit all the familiar, attractive properties of the empirical likelihood method – they are range preserving, transformation invariant, and their shape is determined by the observed data. To compute quantiles of the exact null distribution for finite samples of size \( n \), we use the recursions of Noé [3]. Section 2 outlines the problem and derives the cdf of the modified Berk–Jones test statistic in various simple problem settings. These solutions clearly demonstrate that the exact null distribution is sample-specific and depends on the number of complete response measurements in the sample, as well as the presence and location of incomplete observations, that is, the total sample configuration. Section 3 identifies the nonparametric likelihood confidence bands and how to compute them efficiently, in larger samples. By means of various examples considered in Section 4, we explore the effect of right censoring on the exact null distribution and the resulting confidence bands, in greater depth. A simulation study described in Section 5 provides evidence of the superiority of the proposed method of deriving exact interval estimates of \( F \) compared to competing alternative methods of inference. The final section of this article consists of some summary remarks and observations.

### 2 Exact nonparametric confidence bands for \( F(x) \) from very small samples

Without loss of generality, we assume that the observed data consist of \( n \) distinct, complete observations \( X_1, \ldots, X_n \) such that \( X_1 < \cdots < X_n \). We will denote incomplete or right-censored observations from the same sample by \( C_1, \ldots, C_k \) and assume that these are similarly ordered among themselves from smallest to largest. Thus, the total sample size is \( n + k \); the value of \( n \) is sometimes called the effective sample size. We also suppose that any censoring mechanism is such that the observed nonparametric likelihood for the survivor function is maximized by the Kaplan–Meier estimator. Of course, the presence and location of these censored observations directly affects the observed value of the Kaplan–Meier estimator derived from the sample.

We follow the development outlined in Owen [1], Owen [15] and Jager and Wellner [4]. More specifically, if \( F_0(x) \) is the true survivor function for the population from which the sample was drawn, we want to estimate exact, nonparametric confidence bands for the functional

\[
U = F_0(X);
\]
since \( F_0(x) \) is continuous, it is well-known that the distribution of \( U \) is uniform on \([0,1]\). Since \( X_1 < \cdots < X_n \), in what follows we will identify the corresponding values of \( U \) as
\[
U_n = F_0(X_1) > U_{n-1} = F_0(X_2) > \cdots > U_2 = F_0(X_{n-1}) > U_1 = F_0(X_n).
\]
The statistic identified in Berk and Jones [2] that Owen [1] inverts to derive exact, nonparametric confidence bands for the cdf is
\[
R_n = \sup_{-\infty < x < \infty} K(\hat{F}_n(x), F_0(x)),
\]
where \( \hat{F}_n(x) \) denotes the ecdf for an observed sample from \( F_0(x) \). The function
\[
K(a,b) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}
\]
is a binomial measure of discrepancy between \( a \) and \( b \); if either \( a = 0 \) or \( a = 1 \), we assume that \( 0 \log 0 = 0 \).

Following the development in Owen [1] and Jager and Wellner [4], we will replace \( \hat{F}_n(x) \) by \( \tilde{F}_n(x) \), the Kaplan–Meier estimator, based on \( n \) complete responses and any number of incomplete observations, of the true survivor function, \( F_0(X) \), which replaces \( F_0(x) \), thereby obtaining a modified Berk–Jones statistic
\[
R'_n = \sup_{-\infty < x < \infty} K(\tilde{F}_n(x), F_0(x)).
\]

We use a binomial discrepancy measure since the distribution of the \( n \) complete responses in the sample that equal or exceed any particular choice of \( x \) follows a binomial distribution with parameters \( n \) and \( F_0(x) \). As Owen observed, large values of \( K \) constitute evidence that \( \Pr(X \geq x) \neq F_0(x) \), and \( R'_n \) is based on the particular value of \( x \) in the sample at which the contradictory evidence is greatest. We also regard the \( n+1 \) distinct observed values of the Kaplan–Meier estimator as realizations of the pivotal quantity, \( U = F_0(X) \). Based on the Owen–Jager–Wellner confidence bands, we know that corresponding confidence bands for the true survivor function must consist of pairs of lower and upper limits for each observed value of the Kaplan–Meier estimator. Although the true values of \( U_1 < \cdots < U_n \) are not observed, their joint distribution is known, and from it we will derive \( n \) ordered pairs \((a_1, b_1), \ldots, (a_n, b_n)\) such that
\[
\Pr(a_1 \leq U_1 \leq b_1, \ldots, a_n \leq U_n \leq b_n) = 1 - \alpha.
\]
These \( 2n \) values, together with the trivial upper and lower limits for \( F_0(X) \) of 1 and 0, respectively, if appropriate, will form \( n + 1 \) pairs of values that constitute the exact \( 1 - \alpha \) nonparametric confidence bands for \( F_0(x) \) constructed from the Kaplan–Meier estimator.

We begin by deriving the exact null distribution of this modified Berk–Jones statistic, \( R'_n \), for samples that involve exactly one complete response. Through consideration of three different cases, we will see that this exact distribution depends crucially on the presence or absence of incomplete observations and their configuration in the sample with respect to \( X_i \).

### 2.1 The exact distribution of \( R'_1 \)

#### Case 1: \( n = 1 \) and \( k = 0 \)

Since the sample consists solely of \( X_1 \),
\[
\tilde{F}_1(t) = \begin{cases} 1 & t \leq X_1 \\ 0 & t > X_1 \end{cases},
\]
is the Kaplan–Meier estimator. It follows immediately that
Because the function when \( \lambda \)

Case 3: \( n \)

\[ K = \text{equation} \]

It follows that

This change in

\[ n \]

Case 2:

Obviously,

This value, together with

After a little effort, we can show that

It induces a corresponding change in the exact null distribution and critical value of \( R'_1 \), as well as the two pairs of lower and upper limits corresponding to the two observed values of \( \hat{F}_1(t) \), since

\[ K(1/2, U_1) = (1/2) \log(1/2U_1) + (1/2) \log(1/2(1 - U_1)) \]

\[ K(0, U_1) = \log(1/(1 - U_1)). \]

After a little effort, we can show that

\[ \Pr(R'_1 \leq \lambda) = 0.5 \left\{ 1 + \sqrt{1 - e^{-\lambda}} \right\} - e^{-\lambda} \]

is the cdf of \( R'_1 \) for this configuration of the observed data. If we call \( \lambda_{1-a} \), the \((1-a)\)-quantile of this null distribution, it follows that

\[ a_1 = \exp(-\lambda_{1-a}) \]

while \( b_1 = 0.5 \left\{ 1 + \sqrt{1 - \exp(-2\lambda_{1-a})} \right\} \) is the larger root of the equation

\[ K(1/2, U_1) = \lambda_{1-a}; \]

moreover,

\[ \Pr(a_1 \leq U_1 \leq b_1) = 1 - a. \]

Because the function \( K(1/2, U_1) \) is symmetric about \( 1/2 \), there is a second, smaller root of the equation

\[ K(1/2, U_1) = \lambda_{1-a}. \]

This value, together with \( a_1, b_1 \) and the trivial upper bound of 1 gives us two pairs of limits – \( \exp(-\lambda_{1-a}), 1 \) associated with the observed value \( \hat{F}_1(t) = 1 \), and a second pair of limits, based on the two roots of the equation \( K(1/2, U_1) = \lambda_{1-a} \), for \( \hat{F}_1(t) = 1/2 \). When \( a = 0.05, \lambda_{0.95} = 3.008014 \).

Case 3: \( n = 1 \) and \( k = 2 \)
Unless \( X_1 \) is no larger than \( C_1 \), this scenario will be no different from either Case 1 or Case 2. Therefore, we assume that \( X_1 \leq C_1 < C_2 \), and hence

\[
\hat{F}_1(t) = \begin{cases} 
1 & t \leq X_1 \\
2/3 & X_1 < t \leq C_2. 
\end{cases}
\]

Once again, the change in \( \hat{F}_1(t) \) induces changes in the exact null distribution and corresponding critical value of \( R'_1 \), as well as the pairs of limits for these two observed values of \( \hat{F}_1(t) \), since

\[
K(2/3, U_1) = (2/3) \log(2/3U_1) + (1/3) \log\{1/3(1 - U_1)\} \\
\neq K(0, U_1) = \log\{1/(1 - U_1)\}.
\]

We can show that the cdf of \( R'_1 \) corresponding to an observed sample with this configuration is

\[
\Pr(R'_1 \leq \lambda) = y_0(\lambda) - e^{-\lambda}
\]

where \( y_0(\lambda) \) is the largest root of

\[
y^2(1 - y) - \frac{4}{27} e^{-3\lambda} = 0.
\]

Obviously, the trivial upper limit of 1, together with the value \( \exp(-\lambda_{1-a}) \), are the upper and lower limits associated with the observed value \( \hat{F}_1(t) = 1 \). The two roots of the equation \( K(2/3, U_1) = \lambda_{1-a} \) become the corresponding lower and upper limits associated with \( \hat{F}_1(t) = 2/3 \). Once again, \( a_1 = \exp(-\lambda_{1-a}) \) while \( b_1 \) is the larger root of the equation

\[
K(2/3, U_1) = \lambda_{1-a},
\]

and

\[
\Pr(a_1 \leq U_1 \leq b_1) = 1 - a.
\]

When \( a = 0.05 \), the 0.95 quantile of this exact null distribution of \( R'_1 \) is \( \lambda_{0.95} = 2.995732 \).

Figure 1 displays plots of the cdf of \( R'_1 \) for each of these cases. For quite modest values of \( k \), say 3, with \( X_1 \leq C_1 < \cdots < C_6 \), numerical calculations show that the upper quantiles of the exact null distribution of \( R'_1 \) are virtually indistinguishable from those of the function \( 1 - e^{-\lambda} \). It is worth noting that convergence to this limiting cdf agrees with our intuition, since the Kaplan–Meier estimator is indistinguishable from 1 for large values of \( k \) when most right-censored observations exceed \( X_1 \). In this case, the value of the modified Berk–Jones statistic should be \( -\log U_1 \), which has the cdf

\[
\Pr(R'_1 \leq \lambda) = \Pr(U_1 \geq e^{-\lambda}) = 1 - e^{-\lambda}.
\]

These different null distributions for \( R'_1 \) demonstrate that although the number of complete responses is always one, the presence of incomplete observations and the configuration of the observed data affect both the point estimator and the corresponding exact confidence bands. Unlike the results for the cumulative distribution function obtained first by Owen [1] and subsequently corrected in Jager and Wellner [4], it seems almost certain that critical values for \( R'_1 \) will need to be evaluated on a sample-specific basis.

### 2.2 The exact distribution of \( R'_2 \)

We begin by deriving the exact null distribution for the simplest scenario, i.e. \( n = 2 \) and \( k = 0 \). Let \( U_1 \) and \( U_2 \) denote the ordered uniform pivots associated with \( X_2 \) and \( X_1 \). Since the Kaplan–Meier estimator is

\[
\hat{F}_2(t) = \begin{cases} 
1 & t \leq X_1 \\
1/2 & X_1 < t \leq X_2, \\
0 & X_2 < t.
\end{cases}
\]
it follows that
\[
R^2_0 = \max_{b_1} \{ K(0, U_1), K(1/2, U_1), K(1, U_1) \} \lor \max_{b_2} \{ K(0, U_2), K(1/2, U_2), K(1, U_2) \}
\]
\[= \max_{b_1} \{ K(0, U_1), K(1/2, U_1) \} \lor \max_{b_2} \{ K(1/2, U_2), K(1, U_2) \}, \]

where \(0 \leq U_1 < U_2 \leq 1\). The event \(R^2_0 \leq \lambda\) is equivalent to
\[
\max_{b_1} \{ K(0, U_1), K(1/2, U_1) \} \leq \lambda \quad \text{and} \quad \max_{b_2} \{ K(1/2, U_2), K(1, U_2) \} \leq \lambda.
\]

From our previous calculations in Section 2.1, we know that the pair of limits that satisfies the former event will be \(a_1 = (1/2) \{ 1 - \sqrt{1 - e^{-2\lambda}} \} \) and \(b_1 = 1 - e^{-\lambda}\), while the corresponding pair of values that satisfies the latter event will be \(a_2 = e^{-\lambda}\) and \(b_2 = (1/2) \{ 1 + \sqrt{1 - e^{-2\lambda}} \}\). This pair of events is illustrated in panels A and B of Figure 2. Panel C of the same figure shows that the joint event is an irregular pentagon in the upper half of the unit square. On this space, the joint probability density of \(U_1\) and \(U_2\) has a constant value of 2.

For \(X_1 < X_2\) and \(k = 0, \lambda = 2.02955\) is the required critical value to obtain a joint probability of 0.95. Together with the two trivial limits of 0 and 1 for \(\mathcal{F}_0(x)\), we obtain the bounds \((0, b_1)\) for \(\mathcal{F}_2(t) = 0\), \((a_1, b_2)\) for \(\mathcal{F}_2(t) = 1/2\) and \((a_2, 1)\) for \(\mathcal{F}_2(t) = 1\).
Figure 2  Various aspects of simple cases involving two complete observations with $a = 0.05$. (A) Determining lower and upper limits for $U_1$. (B) Determining lower and upper limits for $U_2$. (C) The region of positive joint probability for $U_1$ and $U_2$. (D) Cross-sections of the cdfs from the exact null distributions for Cases 1–4 discussed in Section 2.2
Although it is easy enough to evaluate the joint probability
\[ \Pr(a_1 \leq U_1 \leq b_1, a_2 \leq U_2 \leq b_2) = \Pr(R'_2 \leq \lambda) \]
directly, and hence select \( \lambda_{1-\alpha} \) such that the probability of the joint event equals \( 1-\alpha \), using the stable recursions of Noé \( \cite{3} \) to compute similar joint probabilities will be essential for \( n > 2 \). Therefore, we choose to introduce the use of this algorithm now in the solution of the problem. For a fixed choice of \( \lambda \), which determines the values \( a_1, b_1, a_2, \) and \( b_2 \), we will use Noé’s recursions to evaluate the joint probability of the corresponding event \( a_1 \leq U_1 \leq b_1, a_2 \leq U_2 \leq b_2 \). Then, finding the particular choice of \( \lambda \) such that the probability of the joint event is \( 1-\alpha \) is a straightforward, one-dimensional numerical problem. For any particular choice of \( \lambda \), finding the ordered pairs of limits \( (a_1, b_1) \) and \( (a_2, b_2) \) is simply a larger collection of one-dimensional numerical problems. As Owen \( \cite{1} \) observed, the Van Wijngaarden–Decker–Brent algorithm – implemented by Press et al. \( \cite{16} \) as the function zbrent – is an ideal choice since zbrent has the reliability of the bisection algorithm, but it often converges at a superlinear rate. Using these two numerical tools – zbrent and Noé’s recursions – we can quickly identify the exact confidence bands for any observed sample involving two complete responses and any finite number of right-censored measurements.

Panel D of Figure 2 displays a cross-section of the cdfs of \( R'_2 \) for four observed samples with different configurations involving two complete responses. These correspond to the following four cases:

- **Case 1**: \( X_1, X_2 \), with \( k = 0 \)
- **Case 2**: \( X_1 \leq C_1 < X_2 \), i.e. \( k = 1 \)
- **Case 3**: \( X_1 < X_2 \leq C_1 \), i.e. \( k = 1 \), but a different configuration from **Case 2**.
- **Case 4**: \( X_1 < X_2 \leq C_1 \leq C_2 \leq C_3 \), i.e. \( k = 3 \)

The intersection of the red horizontal line on the plot at 0.95 with each cdf identifies the four different critical values of 2.024955, 2.203956, 2.034496 and 2.502348, respectively, for these various configurations.

As in Section 2.1, these different exact null distributions for \( R'_2 \) demonstrate that although the number of complete responses is always the same, the presence of incomplete observations and the configuration of the data affect not only the point estimator but also the corresponding exact confidence bands. Clearly, a sample-specific solution for the appropriate exact null distribution will be required in every circumstance.

### 3 Evaluating the exact null distribution of \( R'_n \)

Since \( X_1 < X_2 < \cdots < X_n \), it follows immediately that
\[
1 = \hat{F}_n(X_1) > \hat{F}_n(X_2) > \cdots > \hat{F}_n(X_n) > \hat{F}_n(t) \geq 0,
\]
where \( t > X_n \) and the lower bound of zero is attained if \( X_n \) is the largest response measurement recorded in the observed sample. We will denote the corresponding distinct values of the Kaplan–Meier estimator by
\[
\hat{f}_j = \hat{F}_n(X_j),
\]
for \( j = 1, 2, \ldots, n+1 \), with \( \hat{f}_{n+1} \) representing the smallest estimated value of \( \hat{F}_n(t) \) for \( t > X_n \). Admittedly, this notation suppresses the dependence on \( n \); however, in what follows, it eases the notational burden considerably.

According to Claim 4 in Jager and Wellner \( \cite{4} \), for any fixed values \( x_1 \) and \( x_2 \) such that \( 0 < x_1 < x_2 < 1 \),
\[
\begin{align*}
(i) \quad & \max\{y | K(x_1, y) \leq \lambda, K(x_2, y) \leq \lambda \} = \max\{y | K(x_1, y) \leq \lambda \}, \\
(ii) \quad & \min\{y | K(x_1, y) \leq \lambda, K(x_2, y) \leq \lambda \} = \min\{y | K(x_2, y) \leq \lambda \}.
\end{align*}
\]
Since \( 1 > \hat{f}_2 > \cdots > \hat{f}_n > \hat{f}_{n+1} \geq 0 \), for \( j = 2, \ldots, n \), it follows at once that

(i) \( \max \{ y | K(\hat{f}_j, y) \leq \lambda, K(\hat{f}_{j+1}, y) \leq \lambda \} = \max \{ y | K(\hat{f}_{j+1}, y) \leq \lambda \} \),

(ii) \( \min \{ y | K(\hat{f}_j, y) \leq \lambda, K(\hat{f}_{j+1}, y) \leq \lambda \} = \min \{ y | K(\hat{f}_j, y) \leq \lambda \} \).

For the particular cases represented by \( \hat{f}_1 = 1 \) and \( \hat{f}_{n+1} = 0 \), we rely on the corresponding results summarized in the following lemma.

**Lemma**

(a) \[
\max \{ y | K(\hat{f}_2, y) \leq \lambda, \log 1/y \leq \lambda \} = \max \{ y | K(\hat{f}_2, y) \leq \lambda \},
\]

\[
\min \{ y | K(\hat{f}_2, y) \leq \lambda, \log 1/y \leq \lambda \} = \min \{ y | \log 1/y \leq \lambda \}.
\]

(b) \[
\max \{ y | K(\hat{f}_n, y) \leq \lambda, \log 1/(1 - y) \leq \lambda \} = \max \{ y | \log 1/(1 - y) \leq \lambda \},
\]

\[
\min \{ y | K(\hat{f}_n, y) \leq \lambda, \log 1/(1 - y) \leq \lambda \} = \min \{ y | K(\hat{f}_n, y) \leq \lambda \}.
\]

**Proof**

We outline only the proof of (a), since the argument establishing (b) is virtually the same.

For \( 0 < y < 1 \), the function \( \log 1/y \) is decreasing; it achieves its minimum value of 0 at \( y = 1 \). On the other hand, \( K(\hat{f}_2, y) \) achieves the same minimum value of 0 at \( y = \hat{f}_2 \). For \( 0 < y < \hat{f}_2 \), \( K(\hat{f}_2, y) \) is decreasing, and for \( \hat{f}_2 < y < 1 \) it increases. Both \( K(\hat{f}_2, y) \) and \( \log(1/y) \) are continuous on \( 0 < y < 1 \) and, therefore, intersect at some value of \( y > \hat{f}_2 \). It follows that

\[
\max \{ y | K(\hat{f}_2, y) \leq \lambda, \log 1/y \leq \lambda \} = \max \{ y | K(\hat{f}_2, y) \leq \lambda \}.
\]

Since \( \log 1/y > 0 \) at \( y = \hat{f}_2 \), whereas \( K(\hat{f}_2, y) = 0 \), and

\[
|d/dy K(\hat{f}_2, y)| = |(y - \hat{f}_2)/(1 - y)|/y < 1/y = |d/dy \log 1/y|,
\]

it follows that for \( 0 < y < \hat{f}_2 \), \( \log 1/y > K(\hat{f}_2, y) \). Therefore,

\[
\min \{ y | K(\hat{f}_2, y) \leq \lambda, \log 1/y \leq \lambda \} = \min \{ y | \log 1/y \leq \lambda \}.
\]

To obtain the exact nonparametric confidence bands for \( U = \mathcal{F}_0(X) \), we must evaluate the probability of the event \( R'_n \leq \lambda \), where

\[
R'_n = \sup_{X \in \mathcal{F}} K(\hat{F}_n(X), \mathcal{F}_0(X)).
\]

Since we have \( n \) complete responses, \( X_i \), there are \( n \) corresponding ordered pivots

\[
U_1 = \mathcal{F}_0(X_1) > U_{n-1} = \mathcal{F}_0(X_2) > \cdots > U_2 = \mathcal{F}_0(X_{n-1}) > U_1 = \mathcal{F}_0(X_n).
\]

Although these \( n \) pivotal quantities are not observable, their joint distribution is known and determines the distribution of \( R'_n \). In fact, because \( \hat{F}_n(X) \) has a finite set of values,

\[
R'_n = \max_{1 \leq i \leq n} \max_{1 \leq j \leq n+1} K(\hat{f}_j, U_{n+i-1})
\]

\[
= \max_{1 \leq i \leq n} \max \{ K(\hat{f}_i, U_{n+i-1}), K(\hat{f}_{i+1}, U_{n+i-1}) \}
\]

\[
= \max \{ K(1, U_n) = \log 1/U_n, K(\hat{f}_2, U_n) \} \vee \max \{ K(\hat{f}_i, U_{n+i-1}), K(\hat{f}_{i+1}, U_{n+i-1}) \}.
\]
If no incomplete observation in the sample exceeds the observed value of \(X_n\), the latter term in this expression for \( R'_n \) will subdivide into

\[
\max_{2 \leq i \leq n-1} \{ K(\hat{f}_i, U_{n-i+1}), K(\hat{f}_{i+1}, U_{n-i+1}) \} \lor \max \{ K(\hat{f}_n, U_1), K(0, U_1) = \log 1/(1 - U_1) \}.
\]

We now decompose the event \( R'_n \leq \lambda \) into separate events involving the \( n \) ordered pivotals to get

\[
\max \{ \log 1/U_n, K(\hat{f}_2, U_n) \} \leq \lambda,
\]

\[
\max \{ K(\hat{f}_i, U_{n-i+1}), K(\hat{f}_{i+1}, U_{n-i+1}) \} \leq \lambda,
\]
for \( 2 \leq i \leq n-1 \), and either

\[
\max \{ K(\hat{f}_n, U_1), K(\hat{f}_{n+1}, U_1) \} \leq \lambda
\]

if \( \hat{f}_{n+1} > 0 \), or

\[
\max \{ K(\hat{f}_n, U_1), \log 1/(1 - U_1) \} \leq \lambda
\]

if \( \hat{f}_{n+1} = 0 \). Since these \( n \) events involve the individual ordered pivotals \( U_n > U_{n-1} > \cdots > U_1 \), they correspond to lower and upper limits, \( \{ a_i \} \) and \( \{ b_i \} \), \( i = 1, \ldots, n \), respectively, for these pivotals (that also depend on both \( n \) and \( \lambda \)) such that

\[
\Pr(R'_n \leq \lambda) = \Pr(a_1 \leq U_1 \leq b_1, \ldots, a_n \leq U_n \leq b_n).
\]

Because \( K(1,y) = \log 1/y \),

\[
b_n = \max \{ y | \max \{ \log 1/y, K(\hat{f}_2, y) \} \leq \lambda \}
\]

\[
= \max \{ y | \log 1/y \leq \lambda, K(\hat{f}_2, y) \leq \lambda \}
\]

\[
= \max \{ y | K(\hat{f}_2, y) \leq \lambda \},
\]

and

\[
a_n = \min \{ y | \max \{ \log 1/y, K(\hat{f}_2, y) \} \leq \lambda \}
\]

\[
= \min \{ y | \log 1/y \leq \lambda, K(\hat{f}_2, y) \leq \lambda \}
\]

\[
= \min \{ y | \log 1/y \leq \lambda \} = e^{-\lambda}.
\]

Likewise, for \( i = 2, \ldots, n-1 \),

\[
b_{n-i+1} = \max \{ y | \max \{ K(\hat{f}_i, y), K(\hat{f}_{i+1}, y) \} \leq \lambda \}
\]

\[
= \max \{ y | K(\hat{f}_i, y) \leq \lambda, K(\hat{f}_{i+1}, y) \leq \lambda \}
\]

\[
= \max \{ y | K(\hat{f}_{i+1}, y) \leq \lambda \}
\]

and

\[
a_{n-i+1} = \min \{ y | \max \{ K(\hat{f}_i, y), K(\hat{f}_{i+1}, y) \} \leq \lambda \}
\]

\[
= \min \{ y | K(\hat{f}_i, y) \leq \lambda, K(\hat{f}_{i+1}, y) \leq \lambda \}
\]

\[
= \min \{ y | K(\hat{f}_i, y) \leq \lambda \}.
\]

If \( \hat{f}_{n+1} > 0 \), it follows immediately that
Following Owen [1, 15] and Jager and Wellner [4], we find that the exact confidence bands we seek consist of

and

However, if \( \hat{f}_{n+1} = 0 \),

and

Following Owen [1, 15] and Jager and Wellner [4], we find that the exact confidence bands we seek consist of sequences \( \{v_i\} \) and \( \{\tau_i\} \), \( i = 1, \ldots, n + 1 \) such that

Clearly, \( \tau_1 = 1 \), and if \( \hat{f}_{n+1} = 0 \), then \( v_{n+1} = 0 \). All the remaining values of \( \{v_i\} \) and \( \{\tau_i\} \) correspond to a solution of the equation

for some value of \( j = 1, \ldots, n + 1 \). In fact, we can readily deduce that

with \( v_{n+1} = \min \{y | K(\hat{f}_{n+1}, y) \leq \lambda \} \) if \( \hat{f}_{n+1} > 0 \), and \( \tau_{n+1} = 1 - e^{-\lambda} = b_1 \) if \( \hat{f}_{n+1} = 0 \).

It follows that

\[
Pr(K_n' \leq \lambda) = Pr(a_1 \leq U_1 \leq \ldots, a_n \leq U_n \leq b_n)
\]
\[
= Pr(v_n \leq U_1 \leq \tau_{n+1}, \ldots, v_1 \leq U_n \leq \tau_2)
\]
represents the probability of the required event and hence the cdf of $R_n^\lambda$. Since the joint distribution of these unobservable pivotals is known, this probability can be evaluated for any particular choice of $\lambda$ using the stable recursions of Noé.

4 Investigating the effect of right censoring

4.1 An artificial sample of 20 equi-spaced observations

To illustrate, in a very simple situation, how the critical value in the null distribution of $R_n^\lambda$ depends on the configuration of the observed data, consider Figure 3. We begin with a sample that consists of 20 equally spaced observations and suppose that exactly one measurement is incomplete. For each of the 20 possible configurations of these data, we calculated the critical value that corresponds to the 0.95-quantile of the exact null distribution of $R_{19}^\lambda$. These critical values are plotted against the location of the right-censored value in the dataset. The horizontal dashed line represents the critical value of 0.260069 found in column 5 of table 2 in Jager and Wellner [4], and which we also obtained from our algorithm using a dataset consisting of exactly 19 complete response measurements. Figure 3 clearly indicates that the critical value we require to compute exact confidence bands for the survivor function will be a sample-specific critical value, not one that can be tabulated easily, or for which a simple formula can be specified, cf. Owen [1, 15] or Jager and Wellner [4]. Moreover, as the position of the sole right-censored observation moves from an early ranking to a late value, its effect on the critical value becomes more pronounced. In this particular dataset, the largest critical value – and hence the widest confidence bands – occurs when the sole right-censored observation is the second largest value observed. In that particular case, the critical value is 0.3236684, which is almost 25% larger than the value 0.260069 indicated by the horizontal dashed line in Figure 3.

![Figure 3](image_url) Changes in the critical value of $R_{19}^\lambda$ as the location of a single incomplete observation among 19 other equi-spaced responses increases from first to last.
the graph. Nonetheless, as the next two examples clearly demonstrate, the necessary calculations to derive exact, nonparametric confidence bands for any particular sample are practicable.

### 4.2 Stage 4 non-Hodgkin's lymphoma

The data used in this example are a subset of those from 64 non-Hodgkin's lymphoma patients described in Matthews and Farewell [17]. Here, we used the subset consisting of 41 patients with Stage 4, advanced disease. The response measurement is the time, in days, from diagnosis until death, and during follow-up, only 22 (54%) of these patients died.

To illustrate the effect of early versus late censoring on the exact confidence bands, we divided the 19 right-censored observations into four groups of roughly equal size. Panel A in Figure 4 shows the resulting 95% confidence bands for the data when all 41 measurements are treated as complete. The critical value from the null distribution of $R_{0,41}^n$ used to derive these limits was 0.1353939, and the lower and upper bands for this set of 41 complete observations are clearly fairly narrow.

Figure 4B shows the 95% confidence bands derived for the same response measurements when only the five observations in Group 1 (98, 134, 917, 917 and 1,246 days) are treated as right-censored measurements. The locations of these five measurements correspond to the short vertical strokes along the lower axis of the plot. The calculated critical value from the sample-specific exact distribution of $R_{0,36}^n$ used to derive these bounds was 0.1694277, which is roughly 25% larger than the value required for the data when all 41 measurements are treated as complete; it is also 19% larger than 0.1423871, which is the 0.95-quantile for the exact null distribution corresponding to a sample consisting solely of 36 complete observations. Compare Figure 4B with its Panel A counterpart. Notice that this modest collection of early, right-censored observations appears to have little noticeable effect on the resulting confidence bands for $\hat{F}(t)$.

Figure 4C shows the effect of right censoring the ten observations at 98, 134, 917, 917, 1,246, 1,355, 1,381, 1,454, 1,487 and 1,654 days, i.e. Groups 1 and 2, indicated by short vertical marks along the lower axis of the plot. In this case, the critical value from the exact null distribution of $R_{0,31}^n$ was 0.3126413, which is almost twice the value that we used for the simultaneous 95% confidence bands in Panel B, and roughly 2.3 times the critical value when all the observations are treated as complete measurements. If we only had 31 complete observations in the sample, the 0.95-quantile for the exact null distribution would be 0.1640285, roughly half the value dictated by the presence and location of the ten censored observations in Groups 1 and 2 among the remaining 31 complete measurements. Even so, the resulting bands are still reasonably tight.

Figure 4D shows the exact 95% confidence bands that result when the ten observations in Groups 1 and 2 plus the four larger values at 1,838, 1,839, 1,943 and 1,959 days, that is, the 14 observations indicated by the short vertical lines along the lower axis, are treated as right censored. The critical value from the exact null distribution for this set of 27 complete observations and 14 incomplete measurements was 0.6260919, which represents another twofold increase in the 0.95-quantile of the exact null distribution. If the sample had consisted solely of the 27 complete observations, the appropriate critical value would have been only 0.1868825, which is roughly 30% of the critical value we were obliged to use from the exact null distribution for these 27 observations together with Groups 1–3. Of course, now the resulting confidence bands are beginning to push out towards the obvious lower and upper bounds.

Figure 5 shows the exact 95% confidence bands for the actual data – 22 complete observations and 19 incomplete values. The latter values are indicated by the short vertical strokes along the lower axis. In this case, the five observations at 2,091, 2,091, 2,159, 2,175, and 2,291 days are treated as incomplete, in addition to the 14 measurements belonging to Groups 1–3. For these data, the critical value from the null distribution of $R_{0,22}^n$ was 2.1160940, which is more than 15 times the critical value used for the confidence bands calculated from the exact null distribution.
when all 41 observations are treated as complete measurements. It is also at least nine times larger than the 0.95-quantile from the appropriate exact null distribution for a sample consisting solely of the 22 complete measurements. Notice that the upper band hardly deviates from unity, and the lower band soon declines from a maximum value of 0.12 to something hardly distinguishable from 0. When a number of the larger

**Figure 4** The changes in exact 95% confidence bounds induced by the presence of right censoring. Point estimate and bands based on (A) 41 complete observations; (B) 36 complete observations and the 5 right-censored values in Group 1; (C) 31 complete observations and the 10 right-censored values in Groups 1 and 2 and (D) 27 complete observations and the 14 right-censored values in Groups 1–3.
observations in the dataset are right censored, they affect every risk set in the calculation of the Kaplan–Meier estimator and noticeably increase the uncertainty reflected in exact confidence bands calculated from the data.

4.3 Advanced lung cancer survival

To counter the possible misapprehension that exact confidence bands may only make sense in the absence of right censoring, consider the following prospective study reported in Loprinzi et al. [18]. The data concern the survival of advanced lung cancer patients enrolled in trials run by the North Central Cancer Treatment Group. Among these 228 patients, 63 ($\approx 28\%$) have incomplete response times; the short vertical strokes along the horizontal axis in Figure 6 identify these censored observations among the rest of the data. Most occurred prior to 400 days; however, five of the six largest observations recorded were incomplete measurements. In addition, 82 of the response times, both complete and right-censored, belong to various sets of replicated responses. The critical value from the exact null distribution of $R_{139}$ corresponding to exact $95\%$ confidence bands is $\lambda_{0.95} = 0.287351$. The resulting confidence bands are displayed in Figure 6 as the two outermost survival estimates (dashed-dotted lines), along with the point estimate (solid line) and the familiar pointwise $95\%$ confidence intervals (innermost dashed lines) routinely calculated and plotted by the `survfit()` function from the `survival` library in the R software package; see R Core Team [19]. Although the 140 pointwise interval estimates derived from the distinct Kaplan–Meier estimates are usually displayed on the plot as bands that enclose the point estimate for these data, only the confidence bands that correspond to the smallest and largest estimates of the survivor function at any given time represent an exact $95\%$ confidence region for the population survivor function of advanced lung cancer patients. By joining the various endpoints of the pointwise interval estimates to mimic lower and upper estimated bounds, the typical graph produced by `survfit()` – and probably many other software packages as well – conveys a misleading overall impression that advanced lung cancer patient survival is much more precisely determined than study investigators should justifiably suppose.
5 A simulation study

We undertook a simulation study to verify the coverage properties of the proposed confidence bands and to compare that coverage with the four most competitive alternatives. Simultaneous confidence bands for the Kaplan–Meier estimator were first proposed by Gillespie and Fisher [10]. Subsequently, Hall and Wellner [11] developed bands that reduce to the well-known Kolmogorov result in the absence of censoring, and Nair [12] constructed and described equal probability (EP) confidence bands. Each set of bands, and those based on suitable transformations of the probability scale on which they are calculated, relies on the weak convergence of the stochastic process represented by the Kaplan–Meier estimator. Specific details, including suitable formulae, may be found in Borgan and Liestøl [20], who have provided the most comprehensive simulation study of pointwise and simultaneous confidence intervals based on the Kaplan–Meier estimator published in the literature.

Since Borgan and Liestøl [20] found that the most satisfactory statistical behaviour was associated with log–log or arcsine-square root transformations of the (probability) measurement scale for $\hat{F}(t)$, we chose to focus our comparison on the use of these two types of transformations of the Hall–Wellner (H–W) and EP simultaneous confidence bands, together with the exact nonparametric method outlined in Section 3. Unlike the simulation studies reported in Borgan and Liestøl [20] that relied on the use of independent samples from standard exponential distributions, we found that 10,000 replicated samples of complete or
incomplete observations from the usual uniform distribution on the unit interval could be used directly to
evaluate the coverage properties of all five methods of computing simultaneous confidence bands for \( \mathcal{F}(t) \).

In the absence of right censoring, a simulated, independent sample of size \( n \), say, from the uniform
distribution represented a known true realization from \( \mathcal{F}(t) \). Likewise, the ordered set of values \( 1, 2, \ldots, n \)
corresponded to the ranks of the ordered, simulated, observed data. The Kaplan–Meier estimate derived
from these ranks was always the same, giving rise to constant versions of each of the five possible
simultaneous confidence band alternatives. To estimate the corresponding coverage probabilities, we only
needed to compare each realization of the decreasing, ordered true values for a sample of size \( n \) with each of
the five constant versions of the simultaneous confidence bands, and determine whether or not the
estimated confidence bands simultaneously covered all \( n \) values in the true realization of \( \mathcal{F}(t) \). Those
results are summarized in each of the rows in Table 1 that correspond to a censoring proportion of 0.

Strictly speaking, the estimated coverage probabilities of the five methods are not exactly comparable.
This is because the exact method produces a simultaneous confidence band for \( \mathcal{F}(t) \) when \( t \) lies between 0
and the largest observation, regardless of its status, whereas the transformed H–W and EP alternatives
generate a simultaneous confidence band when \( t \) lies between the smallest and largest complete observa-
tions. This means that the estimated coverage probabilities for H–W- and EP-based confidence bands are
necessarily upper bounds on the values that would be directly comparable with the estimates calculated for
the exact method. Nonetheless, it seems likely that these upper bounds are probably fairly tight unless the
effective sample size is rather small, that is, when only a few observations are complete.

To investigate the effects of sample size, we considered values of \( n \), the effective sample size, equal to
10, 20, 50, and 100, and fixed the nominal overall level of confidence at 0.95. Since each of the estimated
coverage probabilities is derived from 10,000 replicated samples, the corresponding estimated standard
errors should be 0.002. Each replicate ordered sample of size \( n \) from the uniform distribution was compared
with the Kaplan–Meier-based simultaneous confidence bands generated by each of the five alternative
methods; that is, the design of the simulation study involved blocking with respect to sample-to-sample
variation so that, for a fixed choice of \( n \) and censoring proportion, the coverage probability of each method
has been estimated on the identical set of 10,000 replications. Clearly, in the absence of censoring, our
proposed method of deriving exact, simultaneous confidence bands for \( \mathcal{F}(t) \) is closer to the nominal
coverage probability than any of the four H–W- or EP-based methods previously published.

### Table 1

Estimated coverage probabilities of 95% simultaneous confidence bands for the survivor function
using five different methods of interval estimation

<table>
<thead>
<tr>
<th>Number of complete observations</th>
<th>Confidence band method</th>
<th>Censored proportion(^\dagger)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.9498</td>
<td>0.9214</td>
</tr>
<tr>
<td>20</td>
<td>0.9525</td>
<td>0.9544</td>
</tr>
<tr>
<td>50</td>
<td>0.9501</td>
<td>0.9401</td>
</tr>
<tr>
<td>100</td>
<td>0.9512</td>
<td>0.9538</td>
</tr>
</tbody>
</table>

\(^\dagger\)Proportion of the total sample that was right censored, i.e. \( p/(1 + p) \).
To obtain replicated samples of the values of \( \hat{F}(t) \) and corresponding Kaplan–Meier estimates involving \( pn \) incomplete \((p \geq 0)\) observations, as well as \( n \) complete responses, it suffices to generate the \( n \) values of \( \hat{F}(t) \) from the uniform distribution and then to associate those values, in decreasing order, with \( n \) increasing values independently sampled at random, without replacement, from the ranks 1, 2, \ldots, \( n(1+p) \). The corresponding observed status vector equals 1 for each complete observation, that is, rank with an assigned value of \( \hat{F}(t) \), and 0 otherwise. The Kaplan–Meier estimate for the resulting observed sample configuration is derived from the rank and status information. From the same Kaplan–Meier estimate, we then evaluated the five different simultaneous confidence bands at the \( n \) complete observation ranks and compared the ordered \( n \) realizations of \( \hat{F}(t) \), that is, the true values, with each of the simultaneous confidence band alternatives. Using this approach, we were able to avoid using any parametric distribution for the observation times in the simulation study. In order to investigate the effect of moderate and more pronounced right censoring, we chose to use values of \( p \) equal to 0.25 and 0.1. The various entries in Table 1 summarize the estimated coverage probabilities for each method of obtaining simultaneous confidence bands with a 0.95 nominal overall level of confidence.

Two remarks concerning these estimated coverage probabilities are warranted. First, the exact method provides simultaneous confidence bands that approximate the nominal level of coverage more consistently than any of the four H–W or EP alternatives we also investigated. Second, when the number of complete observations is substantial (either 50 or 100) and the fraction of right-censored observations is equal to half of the total sample size, i.e. \( p = 1.0 \), so that \( p/(1+p) = 0.5 \), the four H–W– and EP-based methods of simultaneous interval estimation lose a substantial fraction, if not all, of their capability to estimate \( \hat{F}(t) \) reliably whereas, by comparison, any deviations of the estimated coverage probability for the exact method from the nominal value seem negligible. Although identifying definitive reasons for the apparent collapse of the H–W– and EP-based methods of interval estimation is beyond the intended scope of this article, Figure 6 hints at a possible explanation. Notice that the EP log–log transform confidence bands displayed in the figure are much closer to the corresponding pointwise interval estimates for the data, rather than the exact nonparametric confidence bands for the survivor function of these advanced lung cancer patients. When roughly a quarter of the sample is right censored, it would appear that the EP log–log transformation-based simultaneous confidence bands are much narrower than the corresponding patient survival data warrant.

6 Discussion and conclusions

Although various researchers have described approximate confidence bands for the survivor function in the statistical literature, all of these previous methods rely on weak convergence results concerning the distribution of the estimator of \( \hat{F}(t) \). In contrast, the exact confidence bands for the cumulative distribution function that were first described by Owen [1] and subsequently corrected by Jager and Wellner [4] provide a method for investigators to obtain an exact result in samples of any size. In the absence of right censoring, a simple transformation of exact, nonparametric confidence bands based on the ecdf will yield corresponding exact confidence bands for the survivor function estimated from the same data. Similar remarks hold for the results described by Jager and Wellner [5], as well as the more recent, alternative approaches outlined in Frey [7] and Xu et al. [8]. However, the presence of right-censored observations, which is a common occurrence in biostatistical datasets, prevents investigators from using this simple transformation method, since none of these previously mentioned authors considered the possibility of censored measurements in the exact methods that they developed.

Using the framework first considered by Owen [1], we have demonstrated that the method he initially proposed can be suitably adapted to derive exact, nonparametric confidence bands for the survivor function based on the Kaplan–Meier estimator. Since this estimator is well-known and widely used to estimate the survivor function when right censoring occurs, the resulting exact confidence bands should have widespread...
applicability. The method we describe relies on inverting a modified version of the likelihood-based test statistic first proposed by Berk and Jones [2]. Although the optimality properties of the Berk–Jones statistic are no longer assured, because our method is likelihood-based, the resulting confidence bands inherit the familiar, attractive properties of the likelihood approach – they are range preserving, transformation invariant and their “shape,” that is, their symmetry or asymmetry, is determined by the data. In very small sample settings, i.e. $n = 1$ and 2, we demonstrated that the exact null distribution of this test statistic can be evaluated. We also showed that although the distribution depends primarily on the number of complete observations, it is also affected by the occurrence and configuration of incomplete measurements among the other responses. Thus, each sample configuration has its own particular exact null distribution for the modified Berk–Jones statistic. The stable recursions of Noé, combined with a good, one-dimensional root-finding algorithm such as that of Van Wijngaarden–Decker–Brent, provide a satisfactory numerical solution to the problem of evaluating the required exact confidence bands for a given set of data. For any fixed quantile of the sample-specific, exact null distribution, the results outlined in Section 3 indicate how to calculate the corresponding confidence limits from the Kaplan–Meier estimate and the associated level of confidence. Solving for the exact confidence limits that correspond to a specified overall level of confidence, such as 0.95, is then a straightforward, one-dimensional numerical problem. Various examples discussed in Section 4 explore the effects of right censoring in small-, moderate- and larger-sample problem settings. A contributed R package called \texttt{kmconfband} provides various functions to compute and display the exact bands corresponding to any single-sample \texttt{survfit} object and fixed choice of overall confidence level $0 < 1 - \alpha < 1$. A simulation study described in Section 5 provides evidence of the superiority of the proposed method of deriving exact interval estimates of $F(t)$ vis-à-vis existing alternative methods of inference.

In some study settings, investigators may only wish to obtain simultaneous confidence bands for a restricted interval on the time scale. A limited investigation confirms that the method outlined in Section 3 can be suitably adapted to address this modified inferential goal. However, due to the complex interplay of key factors – such as the effective sample size and the observed configuration of complete and incomplete responses – that determine the resulting bands, it seems quite possible that simultaneous confidence bands associated with a restricted interval could be wider than those derived from all the available information concerning patient response.

Farewell [21] shows that by simply reversing the measurement scale, that is, multiplying all the measurements in the sample by $-1$, investigators can use the Kaplan–Meier estimator to evaluate the cumulative distribution function for sample data that are either complete responses or left-censored measurements. Since nothing in the method we have described restricts the measurement scale of the observed sample in any way, it follows that the modified Berk–Jones statistic can be used, without any adaptation other than reversing the measurement scale, to derive exact, nonparametric confidence bands for the cumulative distribution function based on observed samples consisting of complete responses and left-censored measurements. The latter can often arise when the measurement system that was used has a lower limit of resolution.

In certain areas of scientific inquiry, such as lifetime and reliability studies, interval estimates of the cumulative hazard function may be preferable. The invariance of the empirical likelihood ratio statistic under one-to-one transformations guarantees that exact, nonparametric confidence bands for $F_0(t)$ can be transformed directly into exact, nonparametric confidence bands for $H_0(t) = - \log F_0(t)$, the corresponding cumulative hazard function for the population of interest. Alternatively, if using the Nelson–Aalen estimator of $H_0(t)$ is preferred – see Lawless [22], for example – it should be possible to adapt the methods outlined here to derive analogous exact, nonparametric confidence bands based on the statistical behaviour of that estimator, even when some of the sample observations are right censored. We plan to investigate this possibility in a future article.

References