

**CONFIDENCE BANDS FOR HAZARD RATE
UNDER RANDOM CENSORSHIP**

BY MING-YEN CHENG

Department of Mathematics, National Taiwan University, Taipei 106, Taiwan.

cheng@math.ntu.edu.tw

PETER HALL

*Centre for Mathematics and its Applications, Australian National University,
Canberra ACT 0200, Australia.*

halpstat@maths.anu.edu.au

AND DONGSHENG TU

*NCIC Clinical Trials Group, Cancer Research Institute, Queen's University,
Kingston ON K7L 3N6, Canada.*

dtu@ctg.queensu.ca

SUMMARY

We suggest a completely empirical approach to constructing confidence bands for hazard functions, based on smoothing the Nelson-Aalen estimator. In particular, we introduce a local bandwidth-choice method for the bands. Our approach uses empirical information about both the survival rate and the censoring rate, and employs undersmoothing to alleviate difficulties caused by bias. We use both Edgeworth expansion and numerical simulation, the former to develop a basic formula and the latter to modify it for general use.

Some key words: Bandwidth; Censored data; Confidence band; Coverage error; Edgeworth expansion; Hazard function; Kernel methods; Survival analysis.

1. INTRODUCTION

The shape of the hazard rate or failure rate function gives insight into the nature of the process that determines survival; see for example Aalen & Gjessing (2001). Consequently, there is often a need to estimate the hazard rate without imposing assumptions, such as shape constraints, that are not directly supported by the data. This consideration has motivated an extensive literature on nonparametric approaches to hazard rate estimation; see for example the detailed survey given by Wang (2005) and the references in §2.1 below. In the present paper we suggest nonparametric methods for constructing pointwise confidence bands for the hazard rate, based on differencing and kernel-smoothing the Nelson-Aalen cumulative estimator.

Our main challenge will be how to select the smoothing parameter. Although, in principle, a confidence band could be constructed directly from the asymptotic distributions of estimators studied by several authors, the practical performance of such an approach would leave much to be desired, since the resulting band would be heavily encumbered by bias. As a result, its coverage accuracy would be poor.

This type of difficulty also arises in nonparametric density estimation and regression, and there two approaches have been considered for alleviating it, namely explicit bias correction and undersmoothing to reduce the effects of bias. In those relatively conventional problems it is known from both theoretical and numerical studies that undersmoothing generally gives better performance than bias correction (Hall, 1992). The same may be shown to be the case here, too, but the problem remains of how to choose the extent of undersmoothing. In this paper, using a combination of theoretical and numerical arguments, we suggest a simple, practical method for locally-adaptive bandwidth choice. By deriving an asymptotic expansion for the coverage probability we determine the theoretically optimal level of undersmoothing that should be used to construct a pointwise

confidence band. However, the bandwidth that results depends on a variety of unknowns and so is not directly applicable. We therefore use the optimal bandwidth formula to suggest a simple empirical bandwidth that is appropriate in a special case, and then modify it to obtain a new formula which is suitable for general applications. The modification is based on the experience gained in extensive simulations.

Work on confidence bands has been surveyed very well in papers by Claeskens & Van Keilegom (2003) and Dumbgen (2003), and we mention here only some recent literature not mentioned there. In particular, Picard & Tribouley (2000) treated wavelet-based confidence bands for probability densities, Mao & Zhao (2003) discussed spline-based confidence bands, Sun et al. (2000) and Ojeda et al. (2004) proposed confidence bands in the context of generalised linear models, and Hall et al. (2004) introduced confidence bands for receiver operating characteristic curves.

Some of this work uses bootstrap methods, although in the present setting that technique is not attractive, for a number of reasons. First, bootstrap methods for confidence bands in nonparametric function estimation do not take account of bias, which has to be accommodated separately (Härdle & Bowman, 1988). Secondly, several derivatives of two different functions are involved in formulae for the optimal bandwidth, and so a number of different bandwidths, most of them of nonstandard sizes, need to be chosen as part of the bootstrap algorithm. Thirdly, bootstrap confidence bands are of the same asymptotic width as those produced by the simpler method suggested here, so that the bootstrap does not enjoy advantages on that score.

2. METHODOLOGY

2.1. Hazard rate estimators

A variety of methods for hazard rate estimation have been considered in

the literature, including techniques based on explicit estimation of the survival-time density and distribution, and others founded on convolution. We shall use the latter methods, of which early contributions include those of Ramlau-Hansen (1983), Tanner & Wong (1983) and Yandell (1983). Even if attention is confined to convolution there is a range of different methods, based for example on smoothing the Nelson-Aalen estimator (Müller & Wang, 1990, 1994) or on more implicit, local-polynomial ideas (Jiang & Doksum, 2003a, 2003b). The first of these approaches will form the basis for our methodology. To describe it we require a little notation, as follows.

Let C denote the censoring time, and T the true survival time, of a patient. It is assumed that C and T are independent. Put $X = \min(C, T)$ and $\delta = I(T \leq C)$, and write f and F for the density and distribution functions, respectively, of T . The hazard rate is given by

$$h(t) = \frac{f(t)}{1 - F(t)}.$$

Let $(C_j, T_j, X_j, \delta_j)$, for $1 \leq j \leq n$, denote independent values of (C, T, X, δ) , let L be the distribution function of X , and let \hat{L} be the standard empirical estimator of L , computed from the data X_1, \dots, X_n . We observe only the sample $\mathcal{S} = \{(X_1, \delta_1), \dots, (X_n, \delta_n)\}$. Write $X_{(1)} \leq \dots \leq X_{(n)}$ for the ordered values of X_1, \dots, X_n , and let $\delta_{(j)}$ be the concomitant of $X_{(j)}$ in the sequence $(X_1, \delta_1), \dots, (X_n, \delta_n)$. Then, writing b for a bandwidth and taking the kernel, K , to be a bounded, compactly supported, symmetric probability density, we define

$$\hat{h}(t) = \frac{1}{b} \sum_{j=1}^n K\left(\frac{t - X_{(j)}}{b}\right) \frac{\delta_{(j)}}{n - j + 1} = \frac{1}{bn} \sum_{j=1}^n K\left(\frac{t - X_j}{b}\right) \frac{\delta_j}{1 - \hat{L}(X_j) + n^{-1}} \quad (2.1)$$

to be our estimator. It is a smoothed version of the Nelson-Aalen estimator, and has been considered before by Ramlau-Hansen (1983), Tanner & Wong (1983), Yandell (1983) and others.

2.2. Two-sided confidence bands

It can be shown that, for $t \in (0, T]$ with $L(T) < 1$, if K is standardised so that $\int K^2 = 1$, then the estimator $\hat{h} = \hat{h}(t)$, defined at (2.1), has mean equal to $h + O(b^2)$ and variance asymptotic to $(bn)^{-1} (1-L)^{-1} h$, and that the distribution of $(\hat{h} - E\hat{h})/(\text{var } \hat{h})^{1/2}$ converges to the standard normal as n increases (Müller & Wang, 1990). Let Φ denote the standard normal distribution function, and given $0 < \alpha < \frac{1}{2}$, define x_α by $\Phi(x_\alpha) = 1 - \frac{1}{2}\alpha$. The properties discussed above suggest interpreting the interval

$$\mathcal{I}_\alpha(t) = \left[\hat{h} - (bn)^{-1/2} (1 - \hat{L})^{-1/2} \hat{h}^{1/2} x_\alpha, \hat{h} + (bn)^{-1/2} (1 - \hat{L})^{-1/2} \hat{h}^{1/2} x_\alpha \right]$$

as a two-sided confidence interval for $h(t)$, with nominal coverage $1 - \alpha$.

It may be shown by a lengthy argument, outlined in the Appendix, that the coverage error of \mathcal{I}_α is minimised by choosing the bandwidth, b , to be of size $n^{-1/3}$, and that, if b is asymptotic to a constant multiple of $n^{-1/3}$,

$$\text{pr}\{h(t) \in \mathcal{I}_\alpha(t)\} = 1 - \alpha + \text{ACE}(t, \alpha) + o(n^{-2/3}), \quad (2.2)$$

where ‘ACE’ stands for ‘asymptotic coverage error’, which refers to the first-order term in the difference between the actual and nominal coverage probabilities of a confidence interval,

$$\begin{aligned} \text{ACE}(\cdot, \alpha) = & -(bn)^{-1} \frac{x_\alpha}{h(1-L)} \left\{ \frac{1}{12} \kappa^{(4)} (x_\alpha^2 - 3) + \frac{1}{36} (\kappa^{(3)})^2 (x_\alpha^4 - 10x_\alpha^2 + 15) \right. \\ & \left. + \frac{1}{4} x_\alpha^2 (x_\alpha^2 - 1) - \frac{1}{6} \kappa^{(3)} x_\alpha^2 (x_\alpha^2 - 3) \right\} \phi(x_\alpha) \\ & + b^2 \left[\frac{h''}{h} \left\{ \frac{1}{2} \kappa_2^{(2)} + \frac{1}{6} \kappa^{(3)} \kappa_2 (3 - x_\alpha^2) \right\} \right. \\ & \left. + \kappa_2^{(2)} \frac{(2h' \ell + h \ell') (1-L) + 2h \ell^2}{2h(1-L)^2} \right] x_\alpha \phi(x_\alpha) \\ & - \frac{1}{4} nb^5 \kappa_2^2 \frac{(1-L)(h'')^2}{h} x_\alpha \phi(x_\alpha), \end{aligned} \quad (2.3)$$

$\ell = L'$ is the density corresponding to the distribution L , $\kappa_j = \int u^j K(u) du$, $\kappa^{(j)} = \int K(u)^j du$ and $\kappa_2^{(2)} = \int u^2 K(u)^2 du$. In (2.3) it is assumed that K has been standardised for scale through the condition $\kappa^{(2)} = 1$.

2.3. Bandwidth choice

We suggest choosing the bandwidth to minimise the absolute value of the asymptotic coverage error, giving $b_{\text{opt}} = B_{\text{opt}} n^{-1/3}$. However, the constant B_{opt} in this formula involves h , L and their first and second derivatives, which makes it difficult to use the formula in practice, since estimation of these unknowns leads to several other bandwidth selection problems. To avoid this impasse we borrow the ‘effective normal kernel’ idea, often used to select bandwidths for kernel density estimation. In particular, in order to produce a model which is simple and concise and at the same time allows us to include empirical information about both the survival and censoring rates, we assume that the survival-time and censoring-time distributions are exponential, with rates λ_T and λ_C , respectively. In this case, $h \equiv \lambda_T$, $h' \equiv h'' \equiv 0$ and $1 - L(t) = \exp\{-(\lambda_C + \lambda_T)t\}$. The assumption of exponential survival times is sometimes used for other purposes in survival analysis, for example for sample-size calculations.

Formula (2.3) for ACE now simplifies considerably, and it can be shown that $|\text{ACE}(t, \alpha)|$ is minimised by choosing $b = B_{\text{opt}} n^{-1/3}$, where

$$\begin{aligned} B_{\text{opt}} &= \left[B_0(\alpha) \exp\{(\lambda_C + \lambda_T)t\} / \kappa_2^{(2)} \lambda_T (\lambda_C + \lambda_T)^2 \right]^{1/3}, \\ B_0(\alpha) &= \frac{1}{6} \kappa^{(4)} (x_\alpha^2 - 3) + \frac{1}{18} (\kappa^{(3)})^2 (x_\alpha^4 - 10x_\alpha^2 + 15) \\ &\quad + \frac{1}{2} x_\alpha^2 (x_\alpha^2 - 1) - \frac{1}{3} \kappa^{(3)} x_\alpha^2 (x_\alpha^2 - 3). \end{aligned}$$

Since λ_C and λ_T are estimated root- n consistently by $\hat{\lambda}_C = (1 - \bar{\delta})/\bar{X}$ and $\hat{\lambda}_T = \bar{\delta}/\bar{X}$, where $\bar{\delta} = n^{-1} \sum_j \delta_j$ and $\bar{X} = n^{-1} \sum_j X_j$, we can compute an empirical, exponential optimum bandwidth as

$$b_{\text{opt}} = \{B_0(\alpha)/\kappa_2^{(2)} \hat{\lambda}_T (\hat{\lambda}_C + \hat{\lambda}_T)^2 n\}^{1/3} \exp\{(\hat{\lambda}_C + \hat{\lambda}_T)t/3\}.$$

Of course, this bandwidth is proportional to the simpler quantity,

$$\hat{b}_{\text{opt}} = \hat{\lambda}_T^{-1/3} (\hat{\lambda}_C + \hat{\lambda}_T)^{-2/3} n^{-1/3} \exp\{(\hat{\lambda}_C + \hat{\lambda}_T)t/3\}. \quad (2.4)$$

The theory leading to (2.4) is based on relatively high-order asymptotic arguments for estimators, in a problem where even first-order theory can be unreliable. In simpler but related settings, such as confidence-band construction in nonparametric regression and density estimation, it has proved difficult to develop effective bandwidth-choice algorithms based on asymptotic results given by, for example, Hall (1992). Problems arise because there can be as many as three quite different terms contributing to overall bias. As parameter settings are altered there is significant potential for these terms to interact, with the result that the bias contribution to coverage error can change in a way that is not easily captured in small to moderate samples. In particular, in the context of the present paper, and under the exponential model, two of the three terms on the right-hand side of (2.3) that contribute to bias, and are proportional to positive powers of b , vanish identically, but these terms do not vanish in the cases of other models.

To overcome this difficulty, we decided not to use the exact constant of proportionality that should appear in (2.4) under the exponential model, and instead chose the constant on the basis of numerical experimentation for a range of models. This leads to a smaller multiplier for (2.4) than is asymptotically optimal under the exponential model. In fact, our numerical work indicates that, in the case $\alpha = 0.05$, taking the constant to be 1, as in (2.4), is satisfactory.

3. SIMULATION STUDY

Here we report on performance of the bandwidth selector at (2.4), applied to construct the confidence band $\mathcal{I}_{0.05}(t)$. We also compare this rule with an empirical bandwidth proposed by Müller & Wang (1990, 1994) in the context of curve, rather than band, estimation. The technique of Müller and Wang produces asymptotic minimisation of asymptotic mean squared error. Hess et al. (1999) performed extensive simulations to evaluate the performance of the

hazard rate estimate based on this bandwidth. It was not proposed by Müller and Wang that their method be applied to the construction of confidence bands, but Gilbert et al. (2002) suggested, in the discussion section of their paper, that this could be done. From some viewpoints this is an attractive proposal, since in the simplest case, where there is no bias correction, it places the confidence band symmetrically around a popular and effective curve estimator; and if bias corrections are incorporated then it seems likely that the main impediment to coverage accuracy will be removed. For these reasons, a comparison between our approach and methods based on bandwidths chosen for curve estimation would be of interest to practitioners.

Therefore, both bias-ignored and bias-corrected versions of confidence intervals based on the method of Müller and Wang (1990, 1994) are considered. Given a bandwidth, b , produced by this rule, these two versions of the confidence intervals are respectively

$$\left[\hat{h} - (bn)^{-1/2}(1 - \widehat{L})^{-1/2} \hat{h}^{1/2} x_\alpha, \hat{h} + (bn)^{-1/2}(1 - \widehat{L})^{-1/2} \hat{h}^{1/2} x_\alpha \right],$$

$$\left[\hat{h} - \widehat{B} - (bn)^{-1/2}(1 - \widehat{L})^{-1/2} \hat{h}^{1/2} x_\alpha, \hat{h} - \widehat{B} + (bn)^{-1/2}(1 - \widehat{L})^{-1/2} \hat{h}^{1/2} x_\alpha \right],$$

where $\alpha = 0.05$,

$$\widehat{B} = \frac{1}{2} b^2 \tilde{h}'' \int x^2 K(x) dx, \quad \tilde{h}''(t) = \frac{1}{b_d^3} \sum_{j=1}^n K_d \left(\frac{t - X_{(j)}}{b_d} \right) \frac{\delta_{(j)}}{n - j + 1}$$

and the subscript d indicates ‘associated with estimating the second derivative of h .’ In simulations we employed the kernel

$$K_d(x) = \frac{315}{32} (-1 + 9x^2 - 15x^4 + 7x^6), \quad |x| \leq 1.$$

To calculate b_d we used the factor method proposed by Müller & Wang (1990), and there we computed an initial hazard-rate estimate employing the fourth-order kernel,

$$K_2(x) = \frac{105}{64} (1 - 5x^2 + 7x^4 - 3x^6).$$

The following scenario for a clinical-trial process was used in our simulation study. We supposed that patients were taken into the study uniformly over T_a time units. After the last patient had entered, all the patients were followed for additional T_f time units. Therefore, the total duration of the study was $T_a + T_f$ time units. It was assumed that censoring occurred only when patients had not developed the event of interest by the end of the study, and that no patient left the study early or was lost to the follow-up.

The notation C_i , T_i and X_i is as in §2.1. We generated the true survival time, X_i , from a given distribution, and then a censoring time, C_i , based on a given uniform distribution for the random study-entry time. The observed survival time and censoring indicator were then calculated for each patient, and used to compute the respective confidence intervals with nominal 95% levels. For each parameter configuration, 3,000 random samples of sizes $n = 100$ and 200 were generated. The proportion of the confidence intervals covering the true hazard rate, and the average lengths of the intervals over 3,000 samples, were used to estimate, respectively, the coverage probability and the expected length for each confidence interval.

We assumed that the true survival times were from an exponential distribution with scale parameter λ , from a Weibull or gamma distribution with shape parameter γ and scale parameter λ , or from a lognormal distribution with mean parameter μ and variance parameter γ . We treated a range of different values of λ , γ and μ in our simulations but, to save space, only the results when $\lambda = 0.05$ for exponential and Weibull distributions and $\lambda = 0.1$ for the gamma distribution, $\gamma = 2$, and $\mu = \log 10$, and for sample size 100, are summarised in Fig. 1. In the censoring step we took $T_a = 60$ and $T_f = 6$. This represented, respectively, 23%, 20%, 24% or 23% censoring rates in the cases of the exponential, Weibull, gamma and lognormal distributions.

From Fig. 1 we can see that the actual coverage of a confidence interval

based on the rule (2.4) is very close to the nominal 0.95 level, except when t is close to the right-hand boundary, in the case of the Weibull hazard rate, or to the left-hand boundary, for the lognormal hazard rate. The figure also shows that, if we use the optimal bandwidth based on mean squared error to calculate a confidence interval for the hazard rate, then the coverage probability of the resulting confidence interval is, in all cases, far below the nominal level. The explicit bias-correction approach does not rectify this problem, largely because the estimate of the second derivative is not sufficiently accurate.

To illustrate these points we mention that, in the case of gamma-distributed survival times with shape parameter 2 and scale parameter λ , the coverage probabilities of intervals based on the bandwidth at (2.4), when $n = 100$ and $t = 6, 12, 24, 36$, are 0.936, 0.928, 0.905, 0.862 for the respective values of t and for $\lambda = 0.05$; and 0.919, 0.926, 0.932, 0.911 if λ is increased to 0.075, which represents a decrease in the percentage of censoring from 50% to 34%. The respective values of coverage when the mean squared error optimal bandwidth is used are all substantially less than when employing the bandwidth at (2.4); they are 0.728, 0.686, 0.671, 0.634 when $\lambda = 0.05$ and 0.743, 0.653, 0.673, 0.627 when $\lambda = 0.075$. Coverages for the bias-corrected approach are even worse, in each case.

As expected, except at boundary points in some cases, the actual coverage of the confidence interval based on the bandwidth at (2.4) moves closer to the nominal level, 0.95, as sample size increases. The improvement is generally less marked for the other two confidence procedures. For example, in the context of gamma-distributed survival times with shape parameter 2 and scale parameter $\lambda = 0.05$, the coverage probabilities of our intervals, for $t = 6, 12, 24, 36$, are 0.936, 0.928, 0.905, 0.862, respectively, when $n = 100$, improving to 0.925, 0.937, 0.933, 0.934 when $n = 200$. On the other hand, for confidence intervals based on the mean squared error optimal bandwidth, the respective coverages are

only 0.656, 0.800, 0.656, 0.445 when $n = 100$, and 0.677, 0.795, 0.701, 0.513 when $n = 200$. In each case, one effect of increasing sample size is to reduce interval length by between 20% and 25%.

The very low coverages of confidence intervals computed using the bandwidths based on mean squared error are, of course, direct results of those intervals being too short. In almost all cases the intervals were only 50% to 70% as long as the intervals based on the bandwidth selector at (2.4). This statement is accurate quite generally; in particular, it applies to $\lambda = 0.075$ and to $n = 200$, as well as to the cases $\lambda = 0.05$ and $n = 100$ on which we reported in Fig. 1.

In survival analysis, it is known that heavy censoring in the data causes problems for many statistical procedures. There, different approaches usually are required to handle cases with heavy censoring. Therefore we also evaluated the performance of the bandwidth rule (2.4) when the survival data were heavily censored. Fig. 2 presents the results respectively for the Weibull and gamma distributions with all parameters unchanged except the scale parameter, which was 0.03, 0.02 and 0.01 for the Weibull and 0.05, 0.03 and 0.02 for the gamma. This brought the censoring rate up to respectively 39%, 59% and 86% when the survival distribution was Weibull, and 50%, 71% and 83% when the survival distribution was gamma. We can see from Fig. 2 that our rule still performs satisfactorily, except in the case of extremely heavy censorship.

We also studied the version of our bandwidth-choice method where the effect of censoring was ignored. This amounts to replacing $(\hat{\lambda}_C + \hat{\lambda}_T)$, at (2.4), by $\hat{\lambda}_T$, to which the former would be equal if our estimator of the censoring rate were zero. The impact of this change on the results in Fig. 1 is hardly detectable. However, in the case of Fig. 2, where censoring is particularly heavy, deterioration is noticeable when censoring is not taken into account.

4. DISCUSSION

Our approach has potential applications to other band-estimation problems, including those arising in nonparametric regression. There, as in this paper, an Edgeworth expansion can be used to motivate a bandwidth selector that is valid in simple cases, and numerical simulation can be employed to modify the formula so that the empirical bandwidth enjoys good performance in a much wider range of settings.

Our simulations indicated a relatively larger coverage error at the time points closer to the boundary. It is not known whether the hazard rate estimate based on local polynomial method as proposed by Jiang & Doksum (2003a, 2003b) would make any improvement since the development of the Edgeworth expansion for the coverage of the confidence interval based on this estimate is much harder. In line with other practical advice for interpretation of nonparametric survival curve estimates (Marubini & Valsecchi, 1995), we recommend avoiding drawing inference about the hazard rate at time points that are close to the boundary.

APPENDIX

Outline Proof of (2.2)

The expansions we shall give below hold under the assumption that $b \asymp n^{-1/3}$; that is, $n^{1/3}b$ is bounded away from zero and infinity as $n \rightarrow \infty$. However, expansions that have identical explicit terms, and slightly modified remainders, apply for a wider range of bandwidths, and they imply that the optimal bandwidth, in the sense of minimising coverage error, must satisfy $b \asymp n^{-1/3}$.

Define $D \equiv (bn)^{1/2} (1 - \widehat{L})^{1/2} (\widehat{h} - \mu)$, where $\mu = b^{-1} E[K\{(t - X_1)/b\} \delta_1 / \{1 - L(X_1)\}]$, and let $ct_j(D)$ denote the j th cumulant of the distribution of D . Lengthy calculations show that

$$\begin{aligned} ct_1(D) &= c (b/n)^{1/2} + o\{(bn)^{-1}\}, \\ ct_2(D) &= h (\kappa^{(2)} + 2 c_1 b^2) + o\{(bn)^{-1}\}, \end{aligned}$$

$$\text{ct}_3(D) = (bn)^{-1/2} \kappa^{(3)} h (1-L)^{-1/2} + o\{(bn)^{-1}\},$$

$$\text{ct}_4(D) = (bn)^{-1} \kappa^{(4)} h (1-L)^{-1} + o\{(bn)^{-1}\},$$

where $c_1 = \frac{1}{4} \kappa_2^{(2)} (1-L) h^{-1} \{h/(1-L)\}''$ and c denotes a generic constant, taking different values at different appearances. This leads to the Edgeworth expansion,

$$\begin{aligned} \text{pr}\left\{D/h^{1/2} (1 + c_1 b^2) \leq x\right\} &= \Phi(x) + h^{-1/2} c (b/n)^{1/2} \phi(x) \\ &\quad - (bn)^{-1/2} \frac{1}{6} \{1 + o(1)\} \kappa^{(3)} \frac{x^2 - 1}{\{h(1-L)\}^{1/2}} \phi(x) \\ &\quad - (bn)^{-1} \frac{x}{h(1-L)} \left\{ \frac{1}{24} \kappa^{(4)} (x^2 - 3) \right. \\ &\quad \left. + \frac{1}{72} (\kappa^{(3)})^2 (x^4 - 10x^2 + 15) \right\} \phi(x) + o\{(bn)^{-1}\}, \end{aligned}$$

which formula can be shown to imply that

$$\begin{aligned} &\text{pr}\left\{(bn)^{1/2} (1 - \widehat{L})^{1/2} \left| \widehat{h} - h - \frac{1}{2} b^2 \kappa_2 h'' - c n^{-1} \right| / \widehat{h}^{1/2} (1 + c_2 b^2) \leq x\right\} \\ &= 2\Phi(x) - 1 - (bn)^{-1} \frac{x}{h(1-L)} \left\{ \frac{1}{12} \kappa^{(4)} (x^2 - 3) \right. \\ &\quad \left. + \frac{1}{36} (\kappa^{(3)})^2 (x^4 - 10x^2 + 15) + \frac{1}{4} x^2 (x^2 - 1) - \frac{1}{6} \kappa^{(3)} x^2 (x^2 - 3) \right\} \phi(x) \\ &\quad + o\{(bn)^{-1}\}, \end{aligned} \tag{A.1}$$

where $c_2 = c_1 - \frac{1}{4} \kappa_2 (h''/h)$. Careful use of the delta method shows that if the ‘bias term,’ $\frac{1}{2} b^2 \kappa_2 h'' + c n^{-1}$, is omitted from inside the probability in the numerator on the left-hand side of (A.1), then the correction term,

$$b^2 \kappa_2 \frac{h''}{h} x \left\{ \frac{1}{2} + \frac{1}{6} \kappa^{(3)} (3 - x^2) \right\} \phi(x) - \frac{1}{4} n b^5 \kappa_2^2 \frac{(1-L)(h'')^2}{h} x \phi(x),$$

should be added to the right-hand side. This property, and a similar result for the change that occurs on removing the factor $1 + c_2 b^2$ inside the probability, implies that

$$\text{pr}\left\{(bn)^{1/2} (1 - \widehat{L})^{1/2} |\widehat{h} - h| / \widehat{h}^{1/2} \leq x\right\}$$

$$\begin{aligned}
&= 2\Phi(x) - 1 - (bn)^{-1} \frac{x}{h(1-L)} \left\{ \frac{1}{12} \kappa^{(4)} (x^2 - 3) \right. \\
&\quad \left. + \frac{1}{36} (\kappa^{(3)})^2 (x^4 - 10x^2 + 15) + \frac{1}{4} x^2 (x^2 - 1) - \frac{1}{6} \kappa^{(3)} x^2 (x^2 - 3) \right\} \phi(x) \\
&\quad + b^2 \left[\frac{h''}{h} \left\{ \frac{1}{2} \kappa_2^{(2)} + \frac{1}{6} \kappa^{(3)} \kappa_2 (3 - x^2) \right\} \right. \\
&\quad \left. + \kappa_2^{(2)} \frac{(2h' \ell + h \ell') (1 - L) + 2h \ell^2}{2h(1-L)^2} \right] x \phi(x) \\
&\quad - \frac{1}{4} nb^5 \kappa_2^2 \frac{(1-L)(h'')^2}{h} x \phi(x) + o\{(bn)^{-1}\}.
\end{aligned}$$

This expansion leads quickly to (2.2), with ACE there given at (2.3).

Acknowledgements. This project started while Ming-Yen Cheng and Dongsheng Tu both visited Centre for Mathematics and its Applications, Australian National University.

REFERENCES

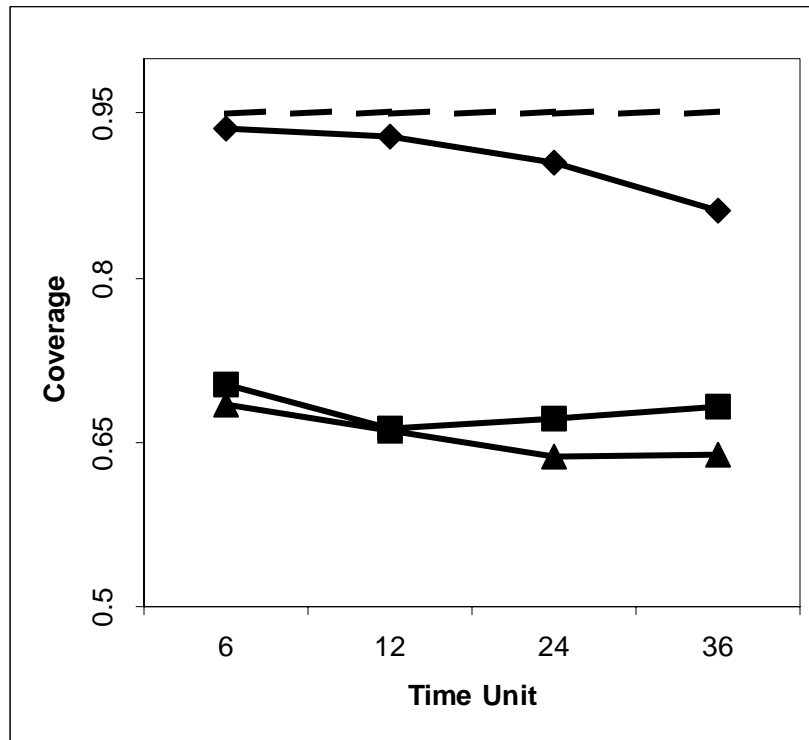
- Aalen, O.O. & Gjessing, H.K. (2001). Understanding the shape of the hazard rate: a process point of view. (with Discussion.) *Statist. Sci.* **16**, 1–22.
- Claeskens, G. & Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–84.
- Dumbgen, L. (2003). Optimal confidence bands for shaperestricted curves. *Bernoulli* **9**, 423–49.
- Gilbert, P.B., Wei, L.J., Kosorok, M.R. & Clemens, J.D. (2002). Simultaneous inferences on the contrast of two hazard functions with censored observations. *Biometrics* **58**, 773–80.
- Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20**, 675–94.
- Hall, P., Hyndman, R.J. & Fan, Y. (2004). Nonparametric confidence intervals for receiver operating characteristic curves. *Biometrika* **91**, 743–50.
- Härdle, W. & Bowman, A.W. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Am. Statist.*

Assoc. **83**, 102–10.

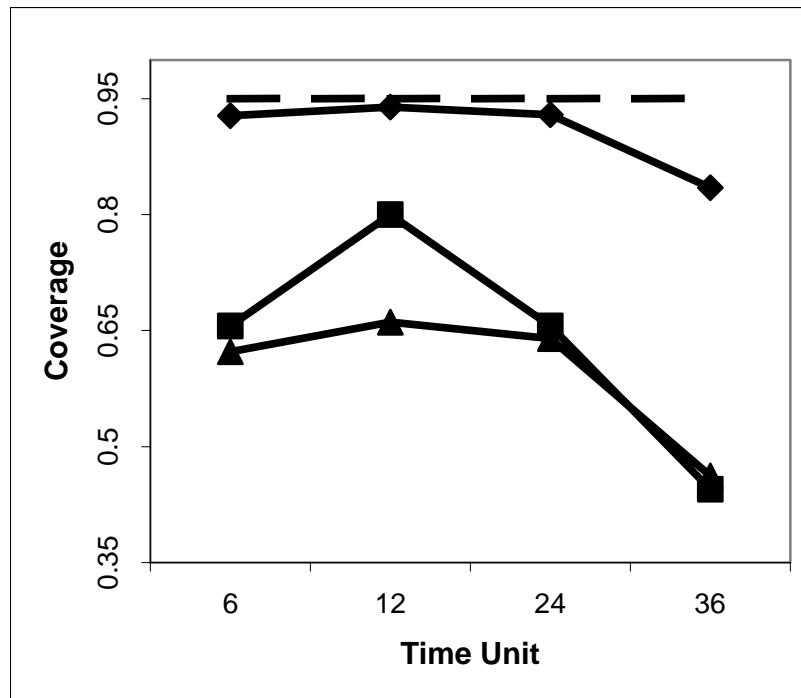
- Hess, K.R., Serachitopol, D.M. & Brown, B.W. (1999). Hazard function estimators: A simulation study. *Statist. Med.* **18**, 3075–88.
- Jiang, J. & Doksum, K.A. (2003a). Empirical plug-in curve and surface estimates. In *Mathematical and Statistical Methods in Reliability, Ser. Qual. Reliab. Eng. Statist.* **7**, Ed. B.H. Lindqvist and K.A. Doksum, pp. 433–53. River Edge, NJ: World Scientific.
- Jiang, J. & Doksum, K.A. (2003b). On local polynomial estimation of hazard functions and their derivatives under random censoring. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*, Ed. M. Moore, S. Froda and C. Léger, *IMS Lecture Notes and Monograph Series* **42**, pp. 463–82. Bethesda, MD: Institute of Mathematical Statistics.
- Mao, W.X. & Zhao, L.H. (2003). Free-knot polynomial splines with confidence intervals. *J. R. Statist. Soc. B* **65**, 901–19.
- Marubini, E. & Valsecchi, M.G. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies*. Chichester: Wiley.
- Müller, H.-G. & Wang, J.-L. (1990). Locally adaptive hazard smoothing. *Prob. Theory Rel. Fields* **85**, 523–38.
- Müller, H.-G. & Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61–76.
- Ojeda, J.L., Cristobal, J.A. & Alcala, J.T. (2004). Nonparametric confidence bands construction for GLM models with length biased data. *J. Nonparam. Statist.* **16**, 421–41.
- Picard, D. & Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28**, 298–335.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453–66.

- Sun, J.Y., Loader, C. & McCormick, W.P. (2000). Confidence bands in generalized linear models. *Ann. Statist.* **28**, 429–60.
- Tanner, M.A. & Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11**, 989–93.
- Wang, J.-L. (2005). Smoothing hazard rate. In *Encyclopedia of Biostatistics, 2nd Edition*, Ed. P. Armitage and T. Colton, **7**, pp. 4986–97. Chichester: John Wiley & Sons, Ltd.
- Yandell, B.S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Statist.* **11**, 1119–35.

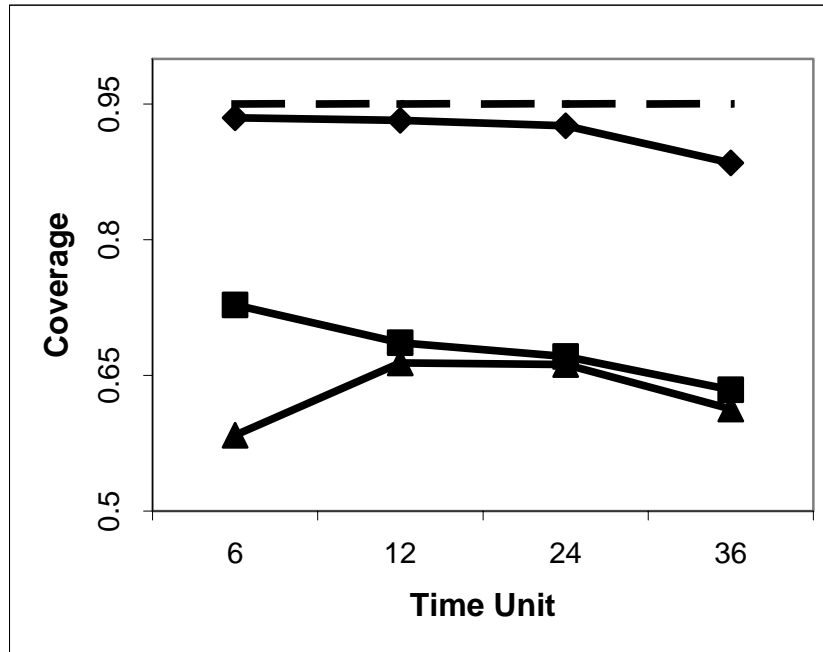
(a)



(b)



(c)



(d)

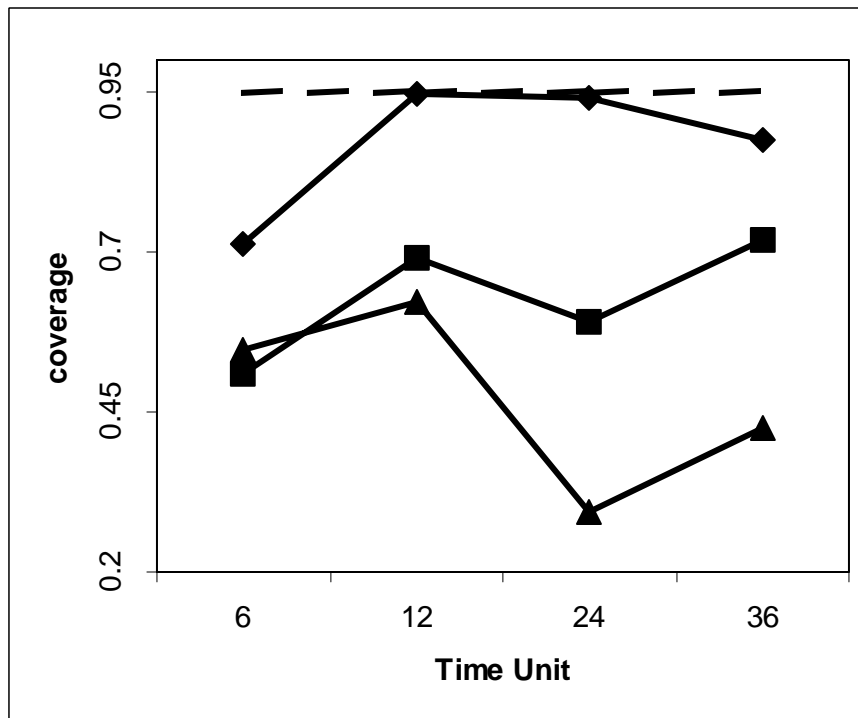
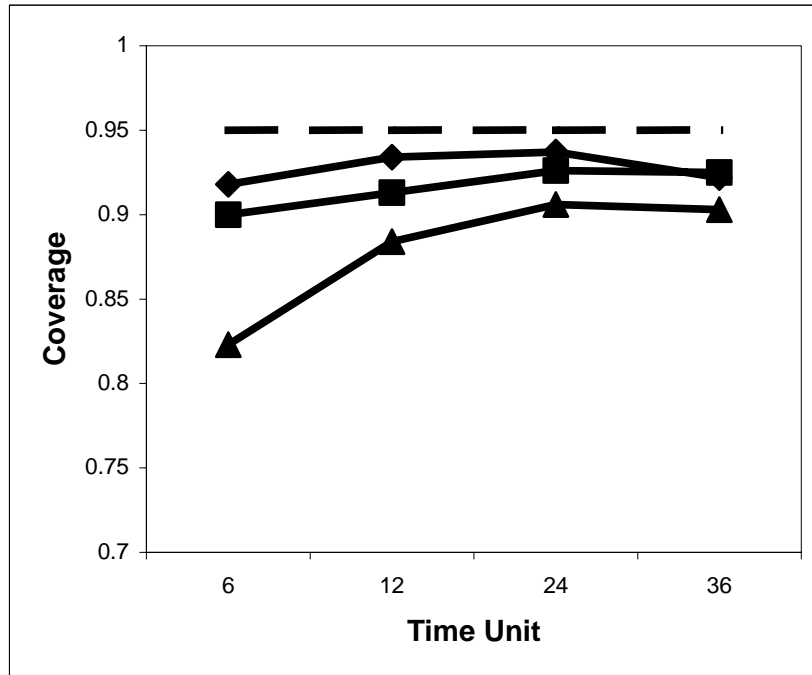


Fig. 1: Actual coverage of confidence intervals with nominal coverage 95%, for hazard rate computed from 3000 simulated sets of data with sample size 100. Dashed line: nominal level; \blacklozenge : coverage for confidence intervals with bandwidth determined by formula (2.4); \blacksquare : coverage for confidence intervals with bandwidth determined by asymptotic MSE; \blacktriangle : coverage for bias-corrected confidence intervals with bandwidth determined by asymptotic MSE. Panels correspond to data from (a) Exponential (0.05), (b) Weibull (2, 0.05), (c) Gamma (2, 0.1), (d) Lognormal (10, 2).

(a)



(b)

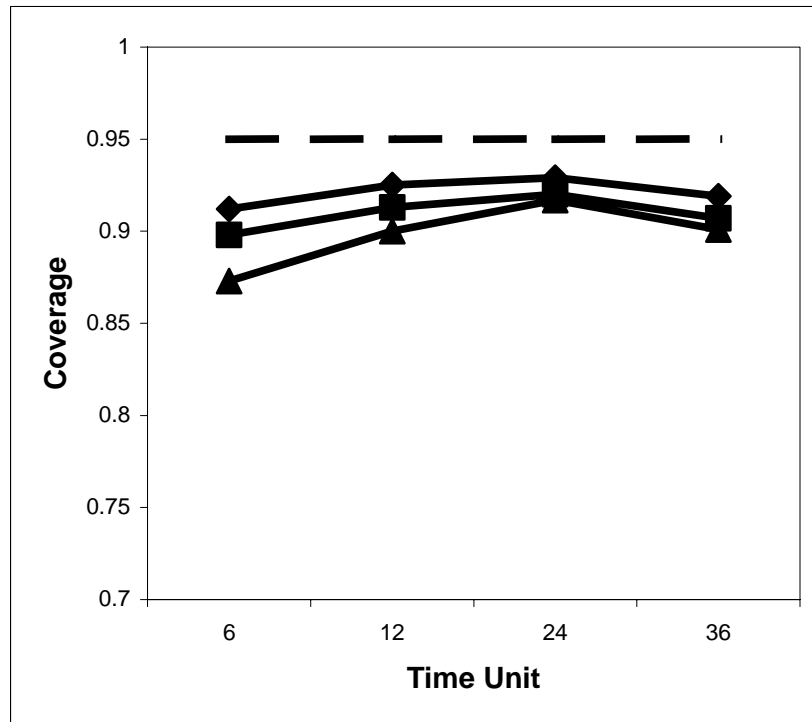


Fig. 2: Actual coverage of confidence intervals with bandwidth determined by formula (2.4) and nominal coverage 95%, for hazard rate computed from 3000 simulated sets of data with sample size 100. Dashed line: nominal level; \blacklozenge : coverage for confidence intervals with $\lambda=0.03$ for Weibull (39% censoring) and 0.05 for Gamma (50% censoring); \blacksquare : coverage for confidence intervals with $\lambda=0.02$ for Weibull (59% censoring) and 0.03 for Gamma (71% censoring); \blacktriangle : coverage for confidence intervals with $\lambda=0.01$ for Weibull (86% censoring) and 0.02 for Gamma (83% censoring). Panels correspond to data from (a) Weibull survival function and (b) Gamma survival function.