Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Bootstrap Confidence Regions for the Intensity of a Poisson Point Process

Ann Cowling , Peter Hall & Michael J. Phillips

To cite this article: Ann Cowling , Peter Hall & Michael J. Phillips (1996) Bootstrap Confidence Regions for the Intensity of a Poisson Point Process, Journal of the American Statistical Association, 91:436, 1516-1524

To link to this article: https://doi.org/10.1080/01621459.1996.10476719

Bootstrap Confidence Regions for the Intensity of a Poisson Point Process

Ann COWLING, Peter HALL, and Michael J. PHILLIPS

Bootstrap methods are developed for constructing confidence regions for the intensity function of a nonstationary Poisson process. Several different resampling algorithms are suggested, ranging from resampling a Poisson process with intensity equal to that estimated nonparametrically from the data to resampling the data points themselves in the same manner that the bootstrap is used in problems involving independent and identically distributed observations. For each different bootstrap method, various percentile-t ways of constructing confidence bands are described. The effectiveness of these different approaches is demonstrated both theoretically and numerically, for real and simulated data. Issues such as bias correction are addressed.

KEY WORDS: Confidence band; Inhomogeneous Poisson process; Intensity function; Kernel estimation.

1. INTRODUCTION

Nonparametric methods for estimating a varying point process intensity have been developed by a number of authors, including Diggle (1985), Diggle and Marron (1988), Ellis (1986), and Leadbetter and Wold (1983). Point estimates may be difficult to interpret without some idea of their accuracy. That typically requires confidence intervals or bands, and the problem of constructing such regions can be rather awkward in the context of dependent data, such as that derived from a Poisson process. Here we develop several simple and, we argue, attractive bootstrap methods for computing confidence regions for Poisson intensity functions. They use a variety of different resampling algorithms and different variants of the percentile-t bootstrap.

Thus the context of our article is that of the bootstrap for dependent processes. Although we treat point processes rather than time series, the setting is in principle not unlike that of bootstrapping a time series when a structural model is available for the type of dependence (e.g., an autoregression), yet the distribution of perturbations is virtually arbitrary. This type of problem has been treated by, for example, Bose (1988), but there are fundamental differences between that work and our own. In particular, in the point process context, with no structural assumptions made about the intensity function, the quantity of interest can be accessed only via statistical smoothing, and so smoothing is an essential feature of all our algorithms. This is not the case with more traditional inferential problems involving dependent data. There are of course other bootstrap approaches to statistical inference under dependence, such as the block bootstrap (see, e.g., Carlstein 1986, Hall 1985, or Künsch 1989), but they are not closely related to the methods that we develop.

There is a parallel between the technology developed here and that appropriate in the setting of nonparametric curve estimation from noisy data. This analogy has been discussed extensively by Diggle and Marron (1988), and so we shall not dwell on it here, except to note that literature in the curve estimation context distantly related to our own includes work of Härdle and Bowman (1988) and Härdle and Marron (1991) on bootstrap confidence bands for regression curves. The Poisson process context treated in this article is distinguished by, among other things, the need to develop a different approach to bootstrapping, with consequently different technology.

A major way in which our work differs from that of earlier authors is that we do not take a "Cox process" view of the intensity estimation problem. Earlier contributions (Diggle 1985; Diggle and Marron 1988) considered the intensity of the observed Poisson process to be a realization of a stationary stochastic process, and assessed performance of their estimation procedures by taking averages over that process as well as over the observed data conditional on intensity. This simplified their theoretical development considerably, because the estimation problem was transformed to one for stationary rather than "time-varying" processes. However, we feel that the Cox process view is much less appropriate in the confidence region problem than it was for point estimation, because it makes little sense to construct a confidence region for the average intensity. Therefore, we treat the Poisson process as genuinely nonstationary. This requires us to develop new theory for point estimation in this more complex case, complementing work of Diggle and Marron on point estimation, as well as contributing new results on confidence regions.

All of our work has a straightforward generalization to the case of multivariate Poisson processes, where intensity is a function from a multivariate Euclidean space to the set of positive numbers. The manner in which our univariate algorithms should be modified is so straightforward that it seems unnecessary to comment further. Generalizations to processes other than Poisson are possible, but unless they are stationary, or replicated data are available, it would seem to be essential to model the structural nature of dependence. Confidence regions for the (scalar) intensity of a stationary non-Poisson process, without structural assumptions, may be constructed using the block bootstrap; Hall (1985) has

Ann Cowling is Statistician/Biometrician, Commonwealth Scientific and Industrial Research Organisation Division of Fisheries, Hobart, TAS 7001, Australia; Peter Hall is Professor, Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia; and Michael J. Phillips is Lecturer, Department of Mathematics and Computer Science, University of Leicester, LE1 7RH, U.K. The authors thank the referees and Anthony Davison for helpful comments that have improved the article.

discussed the concept of the block bootstrap and its application in the context of multivariate coverage processes. However, because this approach does not require smoothing techniques, then it is really beyond the scope of this article, and so will not be treated here.

Section 2 defines basic kernel-type estimators, describes our bootstrap algorithms and confidence intervals and bands, and discusses bias corrections. Section 3 applies these to both real and synthetic data, addressing the problem of finite-sample coverage accuracy. Section 4 develops theory for intervals and bands. Proofs of the results there are available in a longer version of the manuscript, obtainable from the authors.

2. METHODOLOGY

2.1 Summary

In Section 2.2 the basic kernel estimator of a Poisson intensity function is presented and formulas for its asymptotic bias and variance are given. Boundary correction of this estimator is outlined in Section 2.6. Our three basic methods for resampling are discussed in Section 2.3. These are applied to the development of confidence bands in Section 2.4, where four different percentile-t constructions and one percentile method are detailed. Each is suitable for use with each of the resampling methods. The confidence regions introduced in Section 2.4 are perhaps more accurately described as regions for $E(\hat{\lambda})$ rather than for λ. However, they may be bias corrected to produce regions for λ, as shown in Section 2.5.

2.2 Estimate of Intensity

Suppose that a Poisson point process $\{X_i\}$, conceptually on the infinite interval $(-\infty, \infty)$, is observed on $\mathcal{I} = (0, 1)$, with points $0 < X_1 \le X_2 \le \cdots \le X_N < 1$. A kernel estimator of the varying intensity $\lambda(x)$ is given by

$$\hat{\lambda}(x) = h^{-1} \sum_{i=1}^{N} K\{(x - X_i)/h\}, \tag{1}$$

where K is a kernel function and h denotes bandwidth (see, e.g., Diggle and Marron 1988). We address the matter of edge effects in Section 2.6. For now, take $\hat{\lambda}$ to be the estimator defined by (1), perhaps incorporating a modification toward the ends of $\mathcal{I}$.

Statistical consistency of $\hat{\lambda}$ for λ under smoothness conditions alone can occur only in a relative sense, and it requires that λ increase. In particular, it may be shown that if $\lambda = l\mu$ where μ is a fixed function and l is a scalar, then $\hat{\lambda}/l$ converges to μ in mean square if and only if $h \to 0$ and $hl \to \infty$ as $l \to \infty$.

Under these consistency assumptions and using a second-order kernel, we show in Section 4 that the bias and variance of $\hat{\lambda}$ are given approximately by $\frac{1}{2}\kappa h^2 \lambda''$ and $h^{-1}\lambda \int K^2$, where $\kappa = \int y^2 K(y)\, dy$. Thus both the mean and variance of $\hat{\lambda}$ are proportional to λ, a fact used to motivate various percentile-t bootstrap approaches to constructing confidence regions.

We assume throughout the article that the kernel K is a compactly supported, bounded, symmetric probability density function such as the quartic kernel,

$$K(x) = .9375(1 - x^2)^2 I_{[-1,1]}(x).$$

This guarantees that $\hat{\lambda}$ is nonnegative, an essential property for an estimate of the intensity of a point process.

Using the model $\lambda(x) = l\mu(x)$ considered earlier, λ is of size l. In view of the foregoing bias and variance formulas, the bias is of order $h^2 l$ and the error of $\hat{\lambda}$ about its mean is of order $(l/h)^{1/2}$. These two quantities are of equal size when $h \sim l^{-1/5}$, which is the optimal order of bandwidth in the sense of minimizing error in any L_p metric for finite p. More concisely, mean squared error (MSE) at a point is given by

$$\text{MSE} = E(\hat{\lambda} - \lambda)^2 \sim \frac{1}{4}\kappa^2 h^4 \lambda''^2 + h^{-1}\lambda \int K^2, \tag{2}$$

and the right side is minimized by taking

$$\begin{aligned} h &= \left(\lambda \int K^2 \Big/ \lambda''^2 \kappa^2\right)^{1/5} \\ &= l^{-1/5} \left(\mu \int K^2 \Big/ \mu''^2 \kappa^2\right)^{1/5}. \end{aligned} \tag{3}$$

This is of course the analog of the more familiar bandwidth choice formula for kernel density estimators. Such analogies are discussed in considerably greater detail by Diggle and Marron (1988), albeit for Cox processes rather than in the present context.

The foregoing statements about variance, bias, and optimal bandwidth choice apply of course to estimation of λ at a point, and to simultaneous estimation at any finite number of points. Mean integrated squared error (MISE) over an interval $\mathcal{I}$ is readily derived by integrating formula (2):

$$\begin{aligned} \text{MISE} &= \int \{E(\hat{\lambda} - \lambda)^2\} \\ &\sim h^{-1} \left(\int \lambda\right)\left(\int K^2\right) + \frac{1}{4}\kappa^2 h^4 \int \lambda''^2, \end{aligned}$$

and the minimizer of the right side is given by

$$h_{\text{opt}} = l^{-1/5} \left\{ \left(\int \mu\right)\left(\int K^2\right) \Big/ \left(\int \mu''^2\right)\kappa^2 \right\}^{1/5}; \tag{4}$$

compare (3). In particular, the optimal bandwidth is of size $l^{-1/5}$.

Asymptotic theory for a simultaneous confidence band is best described in the L_∞ metric, rather than an L_p metric for finite p. We show in Section 4 that if $\mathcal{I}_\varepsilon$ denotes the interval $(\varepsilon, 1 - \varepsilon)$, then the supremum of error about the mean over $\mathcal{I}_\varepsilon$ is of size $(l/h)^{1/2}(\log l)^{1/2}$ (compared with $(l/h)^{1/2}$ in the case of an L_p metric), and the supremum of bias is of size $h^2 l$ (the same as in an L_p metric).

2.3 Methods for Bootstrapping

There are several methods of bootstrapping the process $\{X_i\}$ to obtain a resample $\{X_i^*\}$. We give three of them here. The second and third methods are motivated by the fact that conditional on $N = n, X_1, \ldots, X_N$ denote the order statistics of a sample of size n drawn from the distribution with density $f(x) = \lambda(x)/\Lambda(1)$, where

$$\Lambda(x) = \int_0^x \lambda(y)\, dy.$$

Method 1. Generate the X_i^*'s as points of a Poisson process of intensity $\hat{\lambda}$. Thus the resampled Poisson process on the interval $\mathcal{I}$ may be written as $X_1^*, \ldots, X_{N^*}^*$, where N^* has a Poisson distribution with intensity $\hat{\Lambda}(1)$ and

$$\hat{\Lambda}(x) = \int_0^x \hat{\lambda}(y)\, dy.$$

Method 2. Conditional on $\mathcal{X} = \{X_1, \ldots, X_N\}$, let N^* have the Poisson distribution with parameter $\hat{\Lambda}(1)$. Draw $X_1^*, \ldots, X_{N^*}^*$ by sampling randomly with replacement, N^* times, from $\mathcal{X}$.

Method 3. Here we do the same as in method 2, except that we replace N^* by N. Thus, conditional on $\mathcal{X}$, we draw $X_1^*, \ldots, X_N^*$ by sampling uniformly, with replacement, from $\mathcal{X}$, and take this collection to be our resample on $\mathcal{I}$.

Methods 2 and 3 produce resamples that include ties, whereas method 1 does not. Methods 1 and 2 give point processes on $\mathcal{I}$ for which the number of points varies from one resample to the next, whereas that number is fixed in the case of method 3.

The bootstrap version of $\hat{\lambda}$ is

$$\hat{\lambda}^*(x) = h^{-1} \sum_i K\{(x - X_i^*)/h\},$$

where the sum is over $1 \le i \le N^*$ in the case of methods 1 and 2 and over $1 \le i \le N$ for method 3.

Note particularly that in the case of methods 2 and 3, $E(\hat{\lambda}^* | \mathcal{X}) = \hat{\lambda}$. This equality is false for method 1, where, to a good approximation,

$$E(\hat{\lambda}^* | \mathcal{X}) \sim \int \hat{\lambda}(x - hy) K(y)\, dy.$$

(A rigorous proof of this result is given in Sec. 4; see Thm. 4.3.) The relative simplicity of $E(\hat{\lambda}^* | \mathcal{X})$ under methods 2 and 3 is perhaps one reason for preferring these methods over method 1.

Resampling under method 1 may be accomplished by first resampling N^* as in method 2 and then drawing $X_1^*, \ldots, X_{N^*}^*$ by sampling randomly with replacement, N^* times, from the distribution with density $\hat{\lambda}/\hat{\Lambda}(1)$. This in turn is equivalent to defining $X_i^* = Y_i^* + hZ_i^*$, where the Y_i^*'s are drawn randomly with replacement from $\mathcal{X}$ and, conditional on $\mathcal{X}$, the Z_i^*'s are independent and identically distributed with density K. Comparing this view of method 1 with method 2, we see that method 1 might be termed a smoothed version of method 2, with the role of the empirical distribution in method 2 replaced in method 1 by that of a kernel-smoothed empirical distribution. The smoothing step does not have to be conducted using the same bandwidth or kernel function as those used to estimate λ in the final analysis; indeed, a different kernel function could be used as well.

Next we discuss techniques for constructing confidence regions for λ, using any of the three methods of resampling. There are basically two approaches—percentile and percentile-t—either of which might need to be bias corrected in some way and both of which are applicable to various types of regions (e.g., symmetric or equal-tailed). We prefer the percentile-t approach and focus our attention there.

2.4 Confidence Bands Before Bias Correction

Let $\mathcal{B}$ denote a connected, voidless random subset of the rectangle $(0,1) \times (0,\infty)$, such that $\mathcal{B} \cap \{(x,y)\colon 0 < y < \infty\}$ is nonempty for each $x \in (0,1)$. We call $\mathcal{B}$ a confidence region for λ over the set $\mathcal{S} \in (0,1)$ with coverage α if

$$P\{(x, \lambda(x)) \in \mathcal{B} \quad \forall\, x \in \mathcal{S}\} = \alpha.$$

The coverage level α is said to be nominal if this identity holds true asymptotically, as λ increases. Typically, $\mathcal{S}$ will be either a point or a finite set of points in the interval $(0,1)$, in which case confidence intervals result, or an interval such as $\mathcal{I}_\varepsilon = (\varepsilon, 1-\varepsilon)$, where $0 < \varepsilon < \frac{1}{2}$, when our construction produces a confidence band. It is often informative to display pointwise confidence intervals in the form of a band (see, e.g., Silverman 1986 or Wahba 1983), even though a simultaneous confidence band will result only by taking $\mathcal{S}$ to be an interval.

Define $\hat{\lambda}_1 = E(\hat{\lambda}^* | \mathcal{X})$. Percentile-$t$ bootstrap confidence regions, such as those given by $\mathcal{B}_1$ to $\mathcal{B}_3$ later in this section, for λ are based on the assertion that the bootstrap distribution of the stochastic process

$$T^*(x) = \{\hat{\lambda}^*(x) - \hat{\lambda}_1(x)\}/\hat{\lambda}^*(x)^{1/2}, \qquad x \in \mathcal{I},$$

is an approximation to the unconditional distribution of

$$T(x) = [\hat{\lambda}(x) - E\{\hat{\lambda}(x)\}]/\hat{\lambda}(x)^{1/2}, \qquad x \in \mathcal{I}.$$

The validity of this claim is established in an asymptotic sense in Section 4. (Standardization by the square root of intensity is motivated by the fact that $\operatorname{var}(\hat{\lambda})$ is approximately proportional to λ, so that T and T^* are approximately pivotal.) For example, one may define t_1 by

$$P\{|T^*(x)| \le t_1, \text{ all } x \in \mathcal{S} | \mathcal{X}\} = \alpha$$

and take the band to be

$$\mathcal{B}_1 = \{(x,y)\colon x \in \mathcal{S}, \max[0, \hat{\lambda}_1(x) - t_1 \hat{\lambda}(x)^{1/2}] < y < \hat{\lambda}_1(x) + t_1 \hat{\lambda}(x)^{1/2}\}. \quad (5)$$

This produces a confidence region whose width at a point x, being proportional to $\hat{\lambda}(x)^{1/2}$, reflects the variability of the point estimate $\hat{\lambda}(x)$.

A confidence region similar to $\mathcal{B}_1$ may be derived by noting that because the variance of $\hat{\lambda}$ is very nearly proportional to its mean, the square-root transformation is approximately variance-stabilizing. Thus we might base variance-stabilized (or implicitly percentile-t) confidence regions on the assertion that the bootstrap distribution of $U^*(x) = \hat{\lambda}^*(x)^{1/2} - \hat{\lambda}_1(x)^{1/2}$ is an approximation to the distribution of $U(x) = \hat{\lambda}(x)^{1/2} - E\{\hat{\lambda}(x)^{1/2}\}$. This argument suggests defining t_2 by

$$P\{|U^*(x)| \le t_2, \text{ all } x \in \mathcal{S}|\mathcal{X}\} = \alpha,$$

finding the solution for the random function B of

$$|\hat{\lambda}_1^{1/2} - B^{1/2}| = t_2,$$

and taking

$$\mathcal{B}_2 = \{(x, y): x \in \mathcal{S}, (\hat{\lambda}_1^{1/2} - t_2)^2 < y < (\hat{\lambda}_1^{1/2} + t_2)^2\}.$$

A third band, of the equal-tailed type, may be defined by selecting t_3 and t_4 such that

$$P\{t_3 \le T^*(x) \le t_4, \text{ all } x \in \mathcal{S}|\mathcal{X}\} = \alpha,$$

and

$$P\{(T^*(x) \le t_3, \text{ all } x \in \mathcal{S}|\mathcal{X}\} = P\{T^*(x) \ge t_4, \text{ all } x \in \mathcal{S}|\mathcal{X}\},$$

and taking

$$\mathcal{B}_3 = \{(x, y): x \in \mathcal{S}, \max[0, \hat{\lambda}_1(x) - t_4\hat{\lambda}(x)^{1/2}] < y < \hat{\lambda}_1(x) - t_3\hat{\lambda}(x)^{1/2}\}.$$

The first and last of these techniques are of the percentile-t type, in that the studentized ratio $\{\hat{\lambda}^*(x) - \hat{\lambda}_1(x)\}/\hat{\lambda}^*(x)^{1/2}$, rather than just its numerator $V^*(x) = \hat{\lambda}^*(x) - \hat{\lambda}_1(x)$, is bootstrapped and the second technique is implicitly percentile-t. The advantages of percentile-t in other contexts involving curve estimation have been described by Hall (1992, chap. 4). Though Hall's results do not apply directly to the present case, the results may be modified so that they do so. In particular, percentile-t produces better approximations to second-order features of confidence intervals than the percentile method.

Nonbootstrap bands, based on extreme-value approximations, may also be considered. A band of that type, $\mathcal{B}_4$, is introduced in Section 4 in a discussion of the asymptotic distribution of stochastic processes such as T.

2.5 Bias Correction

The methods described earlier for constructing confidence regions produce regions for the expected value $E(\hat{\lambda})$ rather than for λ itself. Depending on the size of bandwidth, the regions may require correction for bias. This problem is simplest in the case of estimation by an interval, or simultaneously by a finite number of intervals; there, the bias correction is unnecessary if and only if bias is negligible or, equivalently, if and only if $h = o(l^{-1/5})$. If $\mathcal{S} = \mathcal{I}_\varepsilon$, so that the confidence region is a band, and if $\mathcal{B}_1$ is defined as in (5), then its coverage probability equals $\alpha + o(1)$ only if $h = o\{(l \log l)^{-1/5}\}$. This follows from the discussion in the last paragraph of Section 2.2. The condition demands that h be significantly smaller (albeit by only a logarithmic factor) than would be required for optimal pointwise performance. An alternative approach is to note that bias is given approximately by $\frac{1}{2}\kappa h^2 \lambda''(x)$ and to explicitly correct a confidence region $\mathcal{B}$ for bias; for example, replacing $\mathcal{B}$ by

$$\mathcal{B}' = \left\{(x, y): \left(x, y + \frac{1}{2}\kappa h^2 \tilde{\lambda}''(x)\right) \in \mathcal{B}\right\},$$

where $\tilde{\lambda}''$ denotes an estimator of λ'', perhaps equal to the second derivative of a version of $\hat{\lambda}$ but using a different, larger value of h in its construction. Theoretical aspects of bias correction are addressed in Section 4.

2.6 Corrections for Edge Effects

Although the bias of the estimator in (1) is of size $O(h^r)$ in the interior of the estimation interval, it is of size $O(1)$ near the boundaries. This increase in bias toward the ends of the interval is a particular problem in intensity estimation, as the process often continues but is not observed outside the observation interval. Ideally, the bias would be of the same size in the boundaries as in the interior.

Diggle (1985) and Diggle and Marron (1988) explicitly corrected the estimator for edge effects, using rescale-the-kernel and reflection methods, both of which reduce the bias to $O(h)$ in the boundary regions. However, we found this level of bias correction inadequate, especially for constructing confidence bands. In another article (Cowling and Hall 1996), we introduce a method of bias correction based on generating pseudodata outside the estimation interval using the order statistics of the sample, then applying the usual estimator to the pseudodata as well as to the actual data. This method has bias of $O(h^r)$ in the boundary regions, the same size as in the interior.

In an extensive simulation study in the density estimation context (which is detailed in Cowling 1995), we show that the pseudodata method outperforms the reflection method in terms of both reducing bias and reproducing the slope of the true curve at the boundary. Our estimator even outperforms the traditional boundary kernel method of bias correction (see, e.g., Gasser, Müller, and Mammitzsch 1985) for moderate and small sample sizes when the slope increases away from the boundary.

In Section 3 we use our pseudodata method of edge correction, generating pseudodata using $X_{(-i)} = -5X_{(i/3)} - 4X_{(2i/3)} + 10/3X_{(i)}$. (Here $X_{(t)}$ is defined by linear interpolation when t is not an integer.)

3. NUMERICAL RESULTS

Our motivation for treating this problem was provided by the classic coal mining disaster dataset given by Jarrett (1979) and previously studied by (among others) Maguire, Pearson, and Wynn (1952), Barnard (1953), Cox and Lewis (1966), Raftery and Akman (1986), Siegmund (1988), and Carlin, Gelfand, and Smith (1992). The dataset consists of

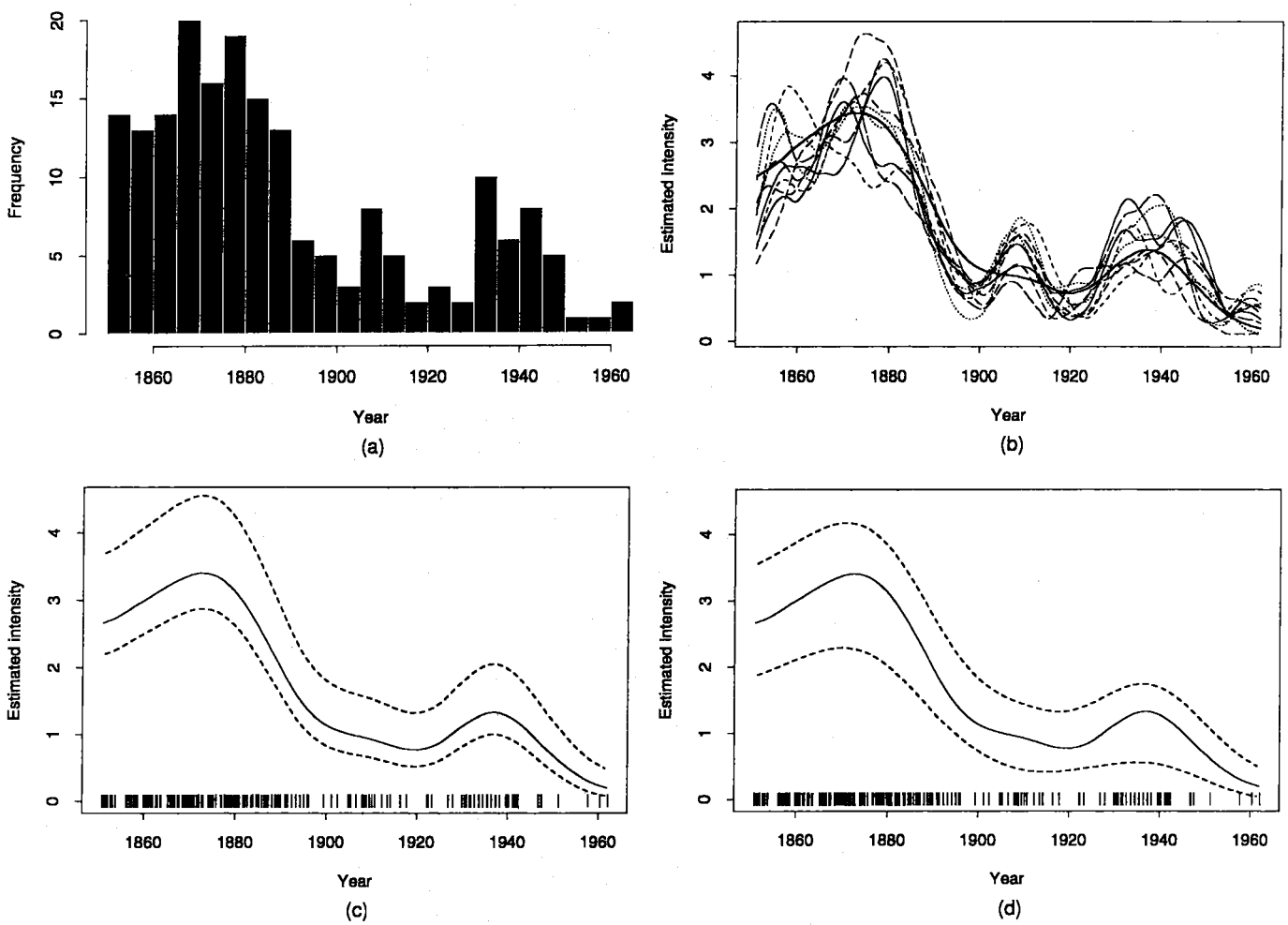

Figure 1. Estimated Intensity of Coal Mining Disasters. (a) Histogram of coal mining data; (b) $\hat{\lambda}^*$ for 10 bootstrap resamples from $\hat{\lambda}$ (resampling method 3, $\hat{\lambda}^*$ calculated using $h_{opt}/2$); (c) bootstrap confidence band for $\hat{\lambda}$ ($\mathcal{B}_1$ and resampling method 1); (d) bootstrap confidence band for $\hat{\lambda}$ ($\mathcal{B}_3$ and resampling method 3).

the time intervals between accidents involving more than 10 deaths in British coal mines between March 15, 1851 and March 22, 1962. A histogram of the data is shown in Figure 1a.

The aim of a statistical analysis of industrial accident data such as this is typically to detect a change in the rate of occurrence of these events. In their analyses of the coal mining data, Barnard (1953) and Cox and Lewis (1966) fitted an exponential model to the intensity function, thereby showing a general decreasing trend in the accident rate. Raftery and Akman (1986), Siegmund (1988), and Carlin et al. (1992) fitted models with two constant intensities and a changepoint. These changepoint analyses all found that the postulated change most probably occurred around 1890.

The decrease in the intensity at this time has been attributed to a severe decline in labor productivity starting at the end of the 1880s and the emergence of the Miners' Federation at the end of 1889 (see Raftery and Akman 1986). However, it is plausible that the alterations in accident rates came about gradually, rather than suddenly. Our approach enables gradual changes to be treated without imposing structural assumptions. In our view, the kernel intensity estimator proposed by Diggle (1985) and Diggle and Marron (1988) and shown in Figure 1, b–d, is a more flexible tool for preliminary analysis than those discussed earlier. Figure 1 demonstrates that although both the exponential and changepoint models capture some general features of the histogram of the data, the kernel estimate reveals not only the same decreasing trend, but also additional features—the peaks at about 1870 and 1940 that are also shown in the histogram—not accounted for by the parametric models. By adding confidence bands to the kernel estimator of intensity, we can answer the same question of global trend that is addressed in the parametric analyses, and moreover can determine whether the additional information about the data—the apparent peaks—is likely to be real or merely a chimera of the nonparametric methodology used.

In our analysis of the coal mining data, we used least squares cross-validation for bandwidth selection; this was shown by Brooks and Marron (1991) to be asymptotically optimal for kernel intensity estimation.

Figure 1, c–d, illustrates two of the confidence bands and two of the resampling methods described in Section 2. The band based on resampling method 1 is smoother than the band based on resampling method 3; both are centered around $E(\hat{\lambda}^*|\mathcal{X})$, but in the former case $E(\hat{\lambda}^*|\mathcal{X}) =$

Table 1. Coverage Error of Confidence Bands for λ With Nominal Level .95

		μ_1			μ_2			μ_3		
l		M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
100	$\mathcal{B}_1$	30 (10)	10 (14)	5 (21)	20 (12)	25 (11)	10 (14)	5 (15)	−5 (16)	−40 (20)
	$\mathcal{B}_2$	35 (9)	5 (15)	−5 (16)	15 (13)	−5 (16)	−5 (16)	5 (15)	−20 (18)	−35 (20)
	$\mathcal{B}_3$	20 (12)	30 (10)	10 (14)	0 (15)	−10 (17)	15 (13)	−20 (18)	−5 (16)	−15 (17)
	$\mathcal{B}_4$	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)
300	$\mathcal{B}_1$	20 (12)	20 (12)	10 (14)	25 (11)	5 (15)	−5 (16)	10 (14)	25 (11)	0 (15)
	$\mathcal{B}_2$	35 (9)	30 (10)	10 (14)	10 (14)	−10 (17)	−5 (16)	10 (14)	20 (12)	0 (15)
	$\mathcal{B}_3$	25 (11)	30 (10)	10 (14)	10 (14)	5 (15)	0 (15)	10 (14)	20 (12)	0 (15)
	$\mathcal{B}_4$	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)
500	$\mathcal{B}_1$	15 (13)	−5 (16)	5 (21)	10 (14)	−5 (16)	−20 (12)	10 (14)	−25 (19)	−5 (16)
	$\mathcal{B}_2$	10 (14)	10 (14)	0 (15)	5 (15)	−15 (17)	−15 (17)	−10 (17)	−30 (19)	5 (15)
	$\mathcal{B}_3$	20 (12)	20 (12)	−5 (16)	10 (14)	−25 (19)	0 (16)	−10 (17)	−30 (10)	5 (15)
	$\mathcal{B}_4$	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)	50 (0)

NOTE: Coverage errors shown are given by (coverage − .95) × 1,000 for the four confidence bands ($\mathcal{B}_1, \ldots, \mathcal{B}_4$) and three resampling methods ($M_1, \ldots, M_3$) discussed in Section 2, for the values of l given, and for the functions μ_1, μ_2, and μ_3 described in Section 3. Standard errors are shown in parentheses.

$\int \hat{\lambda}(x - hy)K(y)\,dy$, whereas in the latter $E(\hat{\lambda}^*|\mathcal{X}) = \hat{\lambda}$. The confidence band limits for $\mathcal{B}_1$ are equally spaced above and below $E(\hat{\lambda}^*|\mathcal{X})$, whereas the band for $\mathcal{B}_3$ is unequally placed because it is an equal-tailed band.

In both bands the lower confidence limit at $x = 1851$ is above the upper confidence limit for $x > 1900$, providing strong evidence of the general decreasing trend revealed by the parametric analyses. Furthermore, both bands suggest that constant-with-changepoint intensities do not enjoy a high likelihood. Figure 1b shows $\hat{\lambda}^*$ for 10 randomly chosen bootstrap resamples (with $\hat{\lambda}^*$ calculated using $h/2$; see later). The peaks at $x = 1870$ and 1940 (also at 1910) tend to reappear in the resamples. In Figure 1d, smoothing the peaks at 1870 and 1940 out of the curve by hand requires crossing the lower confidence band. These peaks may well be real features of the data.

We also conducted a large simulation study to obtain empirical confirmation that our proposed methodology produces confidence bands with coverage probabilities close to the nominal levels. We used the three methods of bootstrap resampling described in Section 2.3 and the methods of confidence band construction resulting in $\mathcal{B}_1, \ldots, \mathcal{B}_3$ in Section 2.4 and $\mathcal{B}_4$ in Section 4. The simulations were based on the asymptotic model $\lambda(x) = l\mu(x)$. We used $l = 100, 300$, and 500, and three functions $\mu(x)$ having the same underlying shape but different degrees of roughness:

$$\mu_1(x) = \frac{3}{2} - x + \frac{1}{4}\sin(4\pi x),$$

$$\mu_2(x) = \frac{3}{2} - x + \frac{1}{4}\sin(3\pi x),$$

and

$$\mu_3(x) = \frac{3}{2} - x + \frac{1}{6}\sin(3\pi x).$$

(The roughness ratios, measured in terms of $\int \mu''^2$, are 4:2.25:1.5.) We used $B = 200$ bootstrap resamples.

To reduce computing time in the simulation study, we used the asymptotically optimal choice, h_{opt}, given in (4). We constructed confidence intervals on $\mathcal{I}_\varepsilon = [.2, .8] \subseteq [h_{\text{opt}}, 1 - h_{\text{opt}}]$, so as not to complicate issues of confidence band construction with issues related to boundary correction.

To correct the confidence bands for bias, we first tried undersmoothing as outlined in Section 2.5. However, this produced oscillatory confidence bands that covered the true curve only about half the time. The approach we took in this article was to use h_{opt} in calculating $E(\hat{\lambda})$ and use $h_{\text{opt}}/2$ for $\hat{\lambda}^*$. Thus the bands may be incorrectly centered, especially near peaks and troughs in the curve. However, the additional width of the bands induced by the smaller bandwidths used in the bootstrap resamples adjusts for this and gives acceptable coverage probabilities. We recommend using an even smaller bandwidth when resampling from the smoothed empirical distribution; we used $h_{\text{opt}}/4$, as we found that the coverage probabilities for method (1) were a little low using $h_{\text{opt}}/2$ in the resampling stage.

Table 1 shows the coverage errors (×1,000) obtained using 200 samples. The nominal coverage probability of these intervals is .95. To indicate the Monte Carlo variability in these estimates, the table also shows the standard error of each coverage error calculated by $1{,}000 \times \{\hat{p}(1 - \hat{p})/200\}^{1/2}$, where $\hat{p}$ is the observed coverage probability. The table shows that all the confidence bands except $\mathcal{B}_4$ had actual coverage probabilities close to the nominal level. The extreme-value confidence bands $\mathcal{B}_4$ are consistently too wide, having actual coverage probability of 1.00. Further, there was little apparent difference between the resampling methods. (More details of the simulation can be found in Cowling 1995.)

4. THEORY

We cannot expect to estimate λ well using only local information, unless the points X_i are closely spaced. Thus asymptotic theory describing consistency must allow λ to increase without bound. We adopt the asymptotic model

$$\lambda(x) = l\mu(x),$$

where the smooth function μ is held fixed and the scalar parameter $l \geq 1$ diverges. In this section we develop asymptotic theory describing the distribution of the stochastic pro-

cess $\hat\lambda$ and its bootstrap counterpart $\hat\lambda^*$ under the model just described.

Our first result describes a Gaussian approximation to the distribution of $\hat\lambda$ that is sufficiently accurate for the development of first-order limit theory. Let $\hat\lambda^*$ be derived by any one of the three different bootstrap algorithms and write E' for expectation conditional on $\mathcal{X}$.

Theorem 4.1. Assume that K is compactly supported with $\text{supp}(K) \subseteq (-c, c)$; that K'' exists and satisfies a Lipschitz condition of order 1; that $0 < \inf_{\mathcal{I}} \mu \le \sup_{\mathcal{I}} \mu < \infty, \sup_{\mathcal{I}} |\mu'| < \infty, \mu'$ is Hölder continuous; and that for some $\xi > 0, l^{-\xi} \ge h \ge l^{-(1/4)+\xi}$. Then we may write

$$\hat\lambda - E(\hat\lambda) = (h^{-1}l)^{1/2}(\mu^{1/2}U_n + R_{1n})$$

and

$$\hat\lambda^* - E'(\hat\lambda^*) = (h^{-1}l)^{1/2}(\mu^{1/2}U_n^* + R_{1n}^*),$$

where U_n, and U_n^* conditional on $\mathcal{X}$, are stationary Gaussian processes on (0, 1) with identical distributions, zero means, and covariances

$$\begin{aligned}\text{cov}\{U_n(x_1), U_n(x_2)\} &= \gamma(x_1 - x_2)\\ &\equiv \int K(y)K\{y + (x_1 - x_2)h^{-1}\}\,dy;\end{aligned}$$

and for some $\eta > 0$, and all $k \ge 1$ and $\varepsilon \in [ch, \frac{1}{2})$,

$$E\left\{\sup_{\varepsilon \le x \le 1-\varepsilon} |R_{1n}(x)|^k + \sup_{\varepsilon \le x \le 1-\varepsilon} |R_{1n}^*(x)|^k\right\} = O(l^{-k\eta})$$

as $l \to \infty$.

An immediate consequence of Theorem 4.1 is that $\hat\lambda - E(\hat\lambda)$ is asymptotically normally distributed with zero mean and variance $h^{-1}\lambda \int K^2$; and that this asymptotic distribution is estimated consistently by the bootstrap. The theorem may also be used to develop an extreme-value approximation to the distribution of $\sup |\{\hat\lambda - E(\hat\lambda)\}/\hat\lambda^{1/2}|$. Indeed, letting $\varepsilon \in [ch, \frac{1}{2})$ and defining $\mathcal{I}_\varepsilon = (\varepsilon, 1 - \varepsilon)$ and

$$a_l = \left(2\log\left[\left\{\int (K')^2 \Big/ \int K^2\right\}^{1/2} (1 - 2\varepsilon)(2\pi h)^{-1}\right]\right)^{1/2},$$

we claim that

$$T \equiv a_l\left[\left(h \Big/ \int K^2\right)^{1/2} \sup_{\mathcal{I}_\varepsilon} |\{\hat\lambda - E(\hat\lambda)\}/\hat\lambda^{1/2}| - a_l\right] \quad (6)$$

has limiting distribution function $G(z) = \exp(-2e^{-z})$. In principle, this result could be used to develop simultaneous confidence bands for $E(\hat\lambda)$ over $\mathcal{I}_\varepsilon$, circumventing the need for bootstrap methods. However, we show shortly that this proposal is somewhat impractical, not in the least because the rate of convergence in the limit theorem $P(T \le z) \to G(z)$ is exceedingly slow, only $O\{(\log l)^{-1}\}$ as $l \to \infty$.

To describe the asymptotic properties of T in more detail, let $\mathcal{S}$ denote any subset of $\mathbb{R}$ and define

$$\Delta_1(z, a, b) = P\left(b\left[\left(h \Big/ \int K^2\right)^{1/2} \sup_{\mathcal{I}_\varepsilon}\{\hat\lambda - E(\hat\lambda)\} \div \hat\lambda^{1/2} - a\right] \le z\right) - \exp(-e^{-z}),$$

$$\Delta_2(z, a, b) = P\left(b\left[\left(h \Big/ \int K^2\right)^{1/2} \sup_{\mathcal{I}_\varepsilon} |\{\hat\lambda - E(\hat\lambda)\} \div \hat\lambda^{1/2}| - a\right] \le z\right) - \exp(-2e^{-z}),$$

and

$$\delta_j(\mathcal{S}, a, b) = \sup_{z \in \mathcal{S}} |\Delta_j(z, a, b)|.$$

If $\mathcal{S}$ contains three or more distinct points, then, no matter how a and b are chosen as functions of h or l, $(\log l)^{-1} = O\{\delta_j(\mathcal{S}, a, b)\}$, and for $a = b = a_l$ and $\mathcal{S} = \mathbb{R}, \delta_j(\mathbb{R}, a_l, b_l) = O\{(\log l)^{-1}\}$. The case $j = 1$ follows from theorem 2.1 of Hall (1991), our Theorem 4.1, and the fact that $E(\hat\lambda) = \lambda + O(h^2 l)$. The case $j = 2$ may be treated similarly. (See also Bickel and Rosenblatt 1973 and Konakov and Piterbarg 1982, 1983, 1984.)

Next we introduce confidence bands based on these extreme-value approximations. The theory just recounted may be used to describe the coverage accuracy of those bands. Assume the conditions of Theorem 4.1, let z_α denote the solution of the equation $\exp(-2e^{-z_\alpha}) = \alpha$, put $t = (\int K^2/h)^{1/2}(a_l + a_l^{-1}z_\alpha)$, and define

$$\mathcal{B}_4 = \{(x, y): x \in \mathcal{S}, \hat\lambda(x) - t\hat\lambda(x)^{1/2} \le y \le \hat\lambda(x) + t\hat\lambda(x)^{1/2}\}.$$

We may straightaway deduce from the results in the previous paragraph that $\mathcal{B}_4$ is a confidence band for $E(\hat\lambda)$ on $\mathcal{I}_\varepsilon$, with asymptotic coverage α, but that the coverage accuracy uniformly over α cannot be better than order $(\log l)^{-1}$. In fact,

$$P\{(x, E\{\hat\lambda(x)\}) \in \mathcal{B}_4, \text{ all } x \in \mathcal{I}_\varepsilon\} = \alpha + O\{(\log l)^{-1}\} \quad (7)$$

uniformly in α,

$$(\log l)^{-1} = O\left[\sup_\alpha |P\{(x, E\{\hat\lambda(x)\}) \in \mathcal{B}_4, \text{ all } x \in \mathcal{I}_\varepsilon\} - \alpha|\right]. \quad (8)$$

However, like the bootstrap bands introduced in Section 2.4, $\mathcal{B}_4$ is a confidence band for $E\hat\lambda$ rather than for λ itself. Therefore, it will not necessarily cover λ with even the desired asymptotic probability α, let alone with accuracy $O\{(\log l)^{-1}\}$, unless h is small enough to make the bias negligible.

Under more stringent conditions, including the requirement that μ'' exist and be continuous, it may be shown that

$$E(\hat{\lambda}) = \lambda + \frac{1}{2}\kappa h^2 \lambda'' + o(h^2 l) \tag{9}$$

uniformly in $x \in \mathcal{I}_\varepsilon$, where $\kappa = \int z^2 K(z)\,dz$ (see Theorem 4.3). Therefore, if we ask that $h = o\{(l^{-1}\log l)^{1/5}\}$, then bias, which is of order $h^2 l$, will be negligibly small in the context of the confidence band, so that

$$P\{(x, \lambda(x)) \in \mathcal{B}_4, \text{ all } x \in \mathcal{I}_\varepsilon\} = \alpha + o(1); \tag{10}$$

compare (7). An alternative approach is to explicitly bias correct the band $\mathcal{B}_4$, as noted in Section 2.6.

Our next result describes the Gaussian approximation of Theorem 4.1 in more detail. Theorem 4.2 is sufficiently detailed to imply that the percentile-t bootstrap approximation to the distribution of T is substantially more accurate than the extreme-value approximation described earlier, as we shortly relate.

Define $M(x) = \int_{0<y<x} \mu(y)\,dy$.

Theorem 4.2. Assume the conditions of Theorem 4.1. Then we may write

$$\hat{\lambda} - E(\hat{\lambda}) = h^{-1} l^{1/2} \int W\{M(x - hy)\} K'(y)\,dy + (h^{-1}l)^{1/2} R_{2n}, \tag{11}$$

and

$$\hat{\lambda}^* - E'(\hat{\lambda}^*) = h^{-1} l^{1/2} \int W^*\{M(x - hy)\} K'(y)\,dy + (h^{-1}l)^{1/2} R^*_{2n}, \tag{12}$$

where W and W^* conditional on $\mathcal{X}$, are standard Wiener processes on $(0, \infty)$, and for all $k \geq 1$ and $\varepsilon \in [ch, \frac{1}{2})$ and some $\beta > 0$,

$$E\left\{\sup_{\varepsilon \leq x \leq 1-\varepsilon} |R_{2n}(x)|^k + (h^2 l)^{-k/2} \sup_{\varepsilon \leq x \leq 1-\varepsilon} |R^*_{2n}(x)|^k\right\} = O[\{(hl)^{-1/2}\log l\}^k l^\beta].$$

Construction of the processes W and W^* depends on l, and in the case of W^*, also on $\mathcal{X}$, although of course they have the same distribution on all occasions. The connection between Theorems 4.1 and 4.2 may be drawn by noting that under the conditions of Theorem 4.1,

$$h^{-1/2}\mu(x)^{-1/2} \int W\{M(x - hy)\} K'(y)\,dy = Z(x) + R(x),$$

where Z is a stationary Gaussian process with covariance γ, and

$$E\left\{\sup_{\varepsilon \leq x \leq 1-\varepsilon} |R(x)|^k\right\} = O(l^{-k\eta})$$

for some $\eta > 0$ and all $k \geq 1$.

The main practical significance of Theorem 4.2 derives from the fact that the first terms on the right sides of (11) and (12) have identical distributions, thereby ensuring the accuracy of the bootstrap approximation to the distribution of the stochastic process $\hat{\lambda} - E(\hat{\lambda})$. Furthermore, the remainder terms R_{2n} and R^*_{2n} are of much smaller order than $(\log l)^{-1}$, guaranteeing that the bootstrap approximation is more accurate than the extreme-value one. Indeed, it may be proved from Theorem 4.2 that with T defined by (6) and

$$T^* = a_l \left\{\left(h \Big/ \int K^2\right)^{1/2} \sup_{\mathcal{I}_\varepsilon} |(\hat{\lambda}^* - E'\hat{\lambda}^*)/\hat{\lambda}^{*1/2}| - a_l\right\},$$

we have

$$\sup_{-\infty < z < \infty} |P(T \leq z) - P(T^* \leq z|\mathcal{X})| = O\{(hl)^{-\eta}\}$$

for some $\eta > 0$, with probability 1. This level of accuracy is exponentially greater than that provided by the extreme-value approximation. The bootstrap does not appear to have any classical competitors in this particular problem.

Take $\mathcal{S} = \mathcal{I}_\varepsilon = (\varepsilon, 1 - \varepsilon)$, so that the region $\mathcal{B}_1$ is a confidence band. The foregoing argument may be used to show that

$$P\{(x, E\hat{\lambda}(x)) \in \mathcal{B}_1, \text{ all } x \in \mathcal{I}_\varepsilon\} = \alpha + O\{(hl)^{-\eta}\} \tag{13}$$

for all $\eta > 0$, uniformly in α. Compare this level of accuracy to that provided by the extreme-value approximation, indicated by (7) and (8).

Of course, the level of coverage accuracy evidenced by (13) is for the expected value of $\hat{\lambda}$, not for λ. Undersmoothing (e.g., choosing h to be of order $l^{-(1/5)-\eta}$ for some $\eta > 0$) or explicit bias correction is required if the coverage error of the band $\mathcal{B}_1$ is to be of order $l^{-\eta'}$ for some $\eta' > 0$, compared to $(\log l)^{-1}$ for the extreme-value band $\mathcal{B}_4$.

We conclude this section by describing details of bias and variance formulas, such as (9). These results lead directly to the bias corrections discussed in Section 2.6. There it was claimed that if K is a second-order kernel, then $E(\hat{\lambda})$ is given approximately by

$$\lambda_1(x) = \int \lambda(x - hy) K(y)\,dy = \lambda(x) + \frac{1}{2}\kappa h^2 \lambda''(x) + o(h^2 l), \tag{14}$$

provided that μ'' exists and is continuous. The order of this approximation to $E(\hat{\lambda})$ is described by the following theorem, which also treats the first of the three different bootstrap methods.

Define $\hat{\lambda}_2(x) = \int \hat{\lambda}(x - hy) K(y)\,dy$, and let $\varepsilon \in [ch, \frac{1}{2})$.

Theorem 4.3. Assume the conditions of Theorem 4.1. Then

$$E(\hat{\lambda}) = \lambda_1 + O(h^{-3} l^{-1/2} + h^{-1+\zeta}) \tag{15}$$

uniformly on $\mathcal{I}_\varepsilon$, and in the case of the first of the three bootstrap methods,

$$E'(\hat{\lambda}^*) = \hat{\lambda}_2 + O(h^{-3} l^{-1/2} + h^{-1+\zeta}) \tag{16}$$

uniformly on $\mathcal{I}_\varepsilon$, with probability 1.

To appreciate the implications of the order of error in (15), observe that under the conditions of Theorem 4.1, $h \geq l^{-(1/4)+\xi}$ for some $\xi > 0$. It follows that $h^{-3}l^{-1/2} + h^{-1+\zeta} = o(h^2 l)$, and so (10) and (11) together imply that if K is a second-order kernel with $\int y^2 K(y)\, dy = \kappa$, say, then (9) holds.

Formula (9) is not always extendable to the bootstrap case in the form

$$E'(\hat{\lambda}^*) - \hat{\lambda} = \frac{1}{2}\kappa h^2 l\hat{\mu}'' + o(h^2 l) = \frac{1}{2}\kappa h^2 l\mu'' + o(h^2 l) \tag{17}$$

with probability 1. The reason is that depending on choice of $h, \hat{\mu}''$ may not be consistent for μ''. It may be shown that $\hat{\mu}'' = \hat{\lambda}''/l$ has variance of order $(h^5 l)^{-1}$, which does not converge to zero unless h is of larger order than $l^{-1/5}$. However, if $l^{-\xi} \geq h \geq l^{-(1/5)+\xi}$ for some $\xi > 0$, and if μ'' exists and satisfies a Lipschitz condition, then (17) may be shown to hold with probability 1. This situation is analogous to that which arises in density estimation, where the second derivative of a kernel estimator is consistent for the second derivative of the true density only if the bandwidth is of sufficiently large an order of magnitude.

Finally, to complement the bias formulas discussed earlier, we note that expressions for the variance of $\hat{\lambda}(x)$ and for the conditional variance of $\hat{\lambda}^*$ are readily deduced from Theorem 4.1: $\text{var}\{\hat{\lambda}(x)\} \sim h^{-1}l\mu(x)\int K^2$ and $\text{var}\{\hat{\lambda}^*(x)|\mathcal{X}\} \sim h^{-1}l\mu(x)\int K^2$, with probability 1.

5. CONCLUSIONS

The simulations showed that all three resampling methods produced confidence bands with low coverage errors. Thus in practice, there is no need to do a computationally intensive parametric simulation from a Poisson process, as the simpler resampling method (3)—the method used in the independent and identically distributed case—seems to work as well.

The three bootstrap confidence bands all had coverage probabilities close to their nominal levels, whereas the extreme value confidence band was consistently too wide, having a coverage probability of 1.00, giving empirical confirmation that the bootstrap is the only viable method of determining confidence bands for kernel intensity estimates.

We used undersmoothing for bias correction as it is much simpler than explicit bias correction. The simulations also confirmed that this method of bias correction produced acceptable results.

[Received December 1994. Revised February 1996.]

REFERENCES

Barnard, G. A. (1953), "Time Intervals Between Accidents—A Note on Maguire, Pearson and Wynn's Paper," *Biometrika*, 40, 212–213.

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1, 1071–1093.

Bose, A. (1988), "Edgeworth Correction by Bootstrap in Autoregressions," *The Annals of Statistics*, 16, 1709–1722.

Brooks, M. M., and Marron, J. S. (1991), "Asymptotic Optimality of the Least-Squares Cross-Validation Bandwidth for Kernel Estimates of Intensity Functions," *Stochastic Processes and Their Applications*, 38, 157–165.

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian Analysis of Changepoint Problems," *Applied Statistics*, 41, 389–405.

Carlstein, E. (1986), "The Use of Subseries Methods for Estimating the Variance of a General Statistic From a Stationary Time Series," *The Annals of Statistics*, 14, 1171–1179.

Cowling, A. (1995), "Some Problems in Curve Estimation," unpublished Ph.D. thesis, Australian National University, Centre for Mathematics and Its Applications.

Cowling, A., and Hall, P. (1996), "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation," *Journal of the Royal Statistical Society*, Ser. B, 58, 551–563.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, London: Methuen.

Diggle, P. (1985), "A Kernel Method for Smoothing Point Process Data," *Applied Statistics*, 34, 138–147.

Diggle, P., and Marron, J. S. (1988), "Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation," *Journal of the American Statistical Association*, 83, 793–800.

Ellis, S. P. (1986), "A Limit Theorem for Spatial Point Processes," *Advances in Applied Probability*, 18, 649–659.

Gasser, T., Müller, H. G., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society*, Ser. B, 47, 238–252.

Hall, P. (1985), "Resampling a Coverage Process," *Stochastic Processes and Their Applications*, 20, 231–246.

——— (1991), "On Convergence Rates of Suprema," *Probability Theory and Related Fields*, 89, 447–455.

——— (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Härdle, W., and Bowman, A. W. (1988), "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands," *Journal of the American Statistical Association*, 83, 102–110.

Härdle, W., and Marron, J. S. (1991), "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *The Annals of Statistics*, 19, 778–796.

Jarrett, R. G. (1979), "A Note on the Intervals Between Coal-Mining Disasters," *Biometrika*, 66, 191–193.

Konakov, V. D., and Piterbarg, V. I. (1982), "Convergence Rate of Maximum Deviation Distributions of Gaussian Processes and Empirical Densities I," *Theory of Probability and Its Applications*, 27, 759–779.

——— (1983), "Rate of Convergence of Maximal Deviations of Gaussian Processes and Empirical Density Functions II," *Theory of Probability and Its Applications*, 28, 172–178.

——— (1984), "On the Convergence Rate of Maximal Deviation Distribution for Kernel Regression Estimates," *Journal of Multivariate Analysis*, 15, 279–294.

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 17, 1217–1241.

Leadbetter, M. R., and Wold, D. (1983), "On Estimation of Point Process Intensities," in *Contributions to Statistics: Essays in Honour of Norman L. Johnson*, ed. P. K. Sen, Amsterdam: North-Holland.

Maguire, B. A., Pearson, E. S., and Wynn, A. H. A. (1952), "The Time Intervals Between Industrial Accidents," *Biometrika*, 39, 168–180.

Raftery, A. E., and Akman, V. E. (1986), "Bayesian Analysis of a Poisson Process With a Change-Point," *Biometrika*, 73, 85–89.

Siegmund, D. (1988), "Confidence Sets in Change-Point Problems," *International Statistical Review*, 56, 31–48.

Silverman, B. W. (1986), "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 47, 1–52.

Wahba, G. (1983), "Bayesian Confidence Intervals for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society*, Ser. B, 45, 133–150.