Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 134 (2005) 549–567

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Confidence sets for the maximizers of intensity functions

Andreas Futschik[a,*], Wen-Tao Huang[b]

[a] *Department of Statistics, Vienna University, Universitätsstr. 5/9, A-1010 Vienna, Austria*

[b] *Department of Management Sciences and Decision Making, Tamkang University, Tamsui, Taiwan, 251, Republic of China*

Received 14 May 2002; accepted 17 April 2004
Available online 21 July 2004

Abstract

Identifying times or time intervals when the intensity function of a Poisson process is maximal is important in a variety of practical problems, for instance in traffic control or with planning issues involving customer arrivals or accident occurrences. For this purpose, we propose confidence sets that are intuitive and easy to obtain, which makes them practicable for a quick exploratory data analysis. They may also be used in the context of mode estimation for probability densities. In the current literature, confidence sets for the mode are based on the assumption of an—at least locally—unique mode. In contrast, our approach retains the coverage probability even if the underlying intensity or density has a flat top. We even allow the intensity to be constant in the extreme.

MSC: 62G15 (primary); 62G07; 62M30 (secondary)

Keywords: Intensity functions; Mode estimation; Confidence sets; Resampling; Kernel density estimates

1. Introduction

Various practical phenomena lead to realizations of inhomogeneous Poisson processes. A frequent goal is to obtain information concerning the unknown intensity function. For this purpose nonparametric estimates for intensities have been proposed by several authors, including Diggle (1985), Diggle and Marron (1988), and Leadbetter and Wold (1983).

* Corresponding author. Tel.: +43-1-4277-38634; fax: +43-1-4277-38639.
E-mail address: andreas.futschik@univie.ac.at (A. Futschik).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.04.008

Applications have been considered, e.g., to data concerning the occurrence of explosive volcanism (Solow, 1991) and to coal mining disaster occurrences (Diggle and Marron, 1988). In these applications, a strong dependence of the estimates on the bandwidth has been observed which makes them difficult to interpret. It is therefore of interest to find out whether interesting properties of the estimated intensity are mere artifacts or actually present. For this purpose, Cowling et al. (1996) proposed methods to construct uniform confidence bands. They investigate the classical coal mining disaster data (see Barnard, 1953; Cox and Lewis, 1966), and use the fact that the lower confidence bound at the year 1870 exceeds the upper confidence limit after 1900 to confirm a decrease in disaster occurrence after 1900. Instead of this indirect approach, we focus on the direct construction of interval estimates for the point(s) where the intensity is maximal. Possible applications of such interval estimates include the identification of times where customer arrivals, accident occurrences or traffic intensities are typically maximal. An illustrating application to ship arrival data at Keelung harbor (Taiwan) will be discussed in more detail in Section 5.

As detailed below, our considered problem is related to that of mode estimation for probability densities. Mode estimation has been considered by several authors. Here, we cite only some of the relevant literature. For unimodal densities, Parzen (1962), Chernoff (1964), Romano (1988), and Grund and Hall (1995) investigate the use of mode estimates based on kernel density estimates. More recently, mode trees have been proposed by Minnotte and Scott (1993) as visual tools for identifying possible modes. Minnotte et al. (1998) considered “mode forests” as a possible robustification of mode trees. Finally, tests for unimodality or, more generally, concerning the number of modes have been investigated for instance by Hartigan and Hartigan (1985), Silverman (1981), Mammen et al. (1992) as well as by Cheng and Hall (1998). While all proposed methods assume either one or a finite number of modes, our proposed confidence sets also work with densities and intensities that have a flat top. The assumption of a finite number of extremal points is frequently reasonable in the density estimation context, but it is unclear whether the assumption is justified in the context of Poisson processes, where periods of constant intensity often seem plausible. Therefore, we do not assume unimodality and even admit the possibility of a constant intensity in the extreme case. With multiple modes, our proposed confidence sets identify those that are global extremes.

We will now state our problem more formally. Assume that we observe an inhomogeneous Poisson process $X(t)$ on a time interval $[0, m]$ whose intensity function $\lambda^*(t)$ has period 1. While periodic intensities are encountered in many situations (think e.g. of daily, monthly or yearly fluctuations), the assumption of periodicity is not essential in the derivation of asymptotic coverage probabilities. It is easy to verify that our asymptotic results can also be obtained by considering $\lambda_l^* = l\mu$ for some density μ and letting $l \to \infty$, an approach chosen e.g. in Cowling et al. (1996).

Our goal is to construct confidence sets C for the point(s) where the intensity function λ^* is maximal using the $N = X(m)$ observed jump points of $X(t)$ on $[0, m]$. Let S denote the argmax set of λ^*, i.e.

$$S = \left\{ t \in [0, 1] : \lambda^*(t) = \sup_s \lambda^*(s) \right\}.$$

To avoid trivialities, assume furthermore that λ^* is not a.e. equal to zero. Then the intensity density

$$\lambda(t) = \frac{\lambda^*(t)}{\int_0^1 \lambda^*(s)\,\mathrm{d}s}, \quad 0 \leqslant t \leqslant 1,$$

is well defined and both λ^* and λ share the same argmax-set S. Thus one may equivalently construct confidence sets for the argmax set of λ. To ensure wide applicability, we will derive our confidence sets under fairly weak smoothness conditions and assume that λ is locally (in a neighborhood of S) Lipschitz continuous, i.e. that there are β and $\varepsilon > 0$ such that

$$|\lambda(t) - \lambda(u)| \leqslant \beta |t-u| \quad \text{for all } t, u \in S_\varepsilon, \tag{1}$$

where $S_\varepsilon := \bigcup_{t \in S} U_\varepsilon(t)$ and $U_\varepsilon(t) := (t-\varepsilon, t+\varepsilon)$. This permits the underlying intensity to exhibit kinks. Assume furthermore

$$\sup_{s \in [0,1] \setminus S_\varepsilon} \lambda(s) < \lambda_m \tag{2}$$

with $\lambda_m := \max_{0 \leqslant s \leqslant 1} \lambda(s)$, a condition satisfied automatically for continuous intensities. We will denote the class of all intensity densities satisfying (1) and (2) by $L_\varepsilon(\beta)$.

Our goal is to construct confidence sets that have the correct asymptotic coverage probability, i.e. that satisfy

$$\liminf_{m \to \infty} P(t \in C) \geqslant 1 - \alpha \quad \text{for any } t \in S. \tag{3}$$

To keep notation simple, the dependence of C on m is suppressed here and subsequently.

Often each period of observation $[i, i+1]$ can be partitioned into equally spaced subintervals I_j $(1 \leqslant j \leqslant k)$ chosen to have natural interpretations. (Think of days, hours, etc.) In such a situation, it is often of interest to identify those intervals where the expected number of occurrences of events is maximal. Consider for instance our example on ship arrivals, as discussed in Section 5. If a more uniform distribution of ship arrivals is desired, a first step to achieve this would be to discourage ship arrivals during peak hours, for instance by increasing the harbor fees. Similar goals are of interest in traffic control. To achieve this, the following alternative confidence requirement is useful. Let

$$S_k^* = \left\{ \bigcup I_l : \int_{I_l} \lambda^*(s)\,\mathrm{d}s = \max_{1 \leqslant j \leqslant k} \int_{I_j} \lambda^*(s)\,\mathrm{d}s \right\}.$$

Then the requirement

$$\liminf_{m \to \infty} P(I_{l_k} \subseteq C) \geqslant 1 - \alpha \quad \text{for any nonrandom sequence } I_{l_k} \in S_k^* \tag{4}$$

provides confidence sets for those time intervals I_{l_k} where the expected cumulative intensity

$$\int_{I_{l_k}} \lambda^*(s)\,\mathrm{d}s$$

is maximal.

In the following, we propose confidence sets both for conditions (3) and (4). We consider fast and simple construction rules, based either on a partition of the observation interval or on kernel density estimates in Sections 2 and 3, as well as more elaborate confidence sets that use simulated quantiles.

The confidence sets are constructed by first considering the problem conditional on $N=n$, and then transfering the obtained results for $n \to \infty$ back to the original Poisson model by applying the strong law of large numbers on N. It can thus be immediately verified that our results carry over to the density estimation context when i.i.d. observations are available.

Section 4 contains a simulation study in which the proposed confidence sets are compared and the actual coverage probabilities are investigated both for smooth and nonsmooth intensities. We look into the effect of data-based bandwidth selection, and the issue of undersmoothing versus the use of estimates for the Lipschitz bound β. We also compare the proposed confidence sets to sets obtained by using uniform confidence bands. For this purpose, we consider bootstrap confidence bands proposed in Cowling et al. (1996), as well as the uniform bands proposed by Hall and Titterington (1988) that do not assume any smoothness besides Lipschitz continuity.

2. The partitioning approach

Consider the partition $I_1, I_2, \ldots, I_k$ of [0, 1], where $I_j=((j-1)/k,\ j/k]$ for $1\leqslant j\leqslant k$. In practical applications, it is often natural to choose k such that the resulting intervals I_j correspond to easily interpretable units, like days or hours. Since the achievable resolution depends also on the amount of available data, we assume k to be an increasing function of N.

Let

$$N_j = \sum_{l=0}^{m-1} [X(l + j/k) - X(l + (j-1)/k)]$$

denote the number of jumps falling into either I_j or one of its translates I_j+l. Our proposed confidence sets are constructed to contain all those subintervals I_l where N_l is within a certain distance of $\max_{1\leqslant j\leqslant k} N_j$. For this purpose, we use the statistics

$$D_l = \frac{\max_{1\leqslant j\leqslant k} N_j - N_l}{(\max_{1\leqslant j\leqslant k} N_j)^{1/2}}$$

and

$$D_l^* = \frac{\max_{1\leqslant j\leqslant k} N_j - N_l - b_k N/(2k^2)}{(\max_{1\leqslant j\leqslant k} N_j)^{1/2}},$$

where b_k is an estimate of the Lipschitz constant β defined in (1). Possible choices of b_k will be discussed later in this section.

More specifically, our confidence set for the regions of maximum cumulative intensity is defined as

$$C_{\mathrm{c}}^{(\mathrm{P})} = \bigcup_{j \in J_{\mathrm{c}}} I_j, \tag{5}$$

with $J_{\mathrm{c}} = \{l : D_l \leqslant q(k, 1-\alpha)\}$ and $q(k, \gamma) = (2 \log k)^{1/2} + \Phi^{-1}(\gamma)$.

A confidence set for the maximizers of the intensity itself is provided by

$$C^{(\mathrm{P})} = \bigcup_{j \in J} I_j,$$

where $J = \{l : D_l^* \leqslant q(k, 1-\alpha)\}$.

Statements (i) and (ii) of Theorem 1 below imply immediately that the confidence sets $C_{\mathrm{c}}^{(\mathrm{P})}$ and $C^{(\mathrm{P})}$ guarantee the desired coverage probabilities (4) and (3) asymptotically. Moreover, it is shown that the coverage probabilities are attained exactly in the limit for constant intensities and for intensities where S contains an interval. We need the following rather weak assumptions.

(A1) Given $N = n$ the number of cells $k = k(n)$ is chosen such that $k \to \infty$ and $nk^{-1} \to \infty$ as $n \to \infty$.
(A2) $\frac{\log k}{\log \log n} \to \infty$, and $\frac{k(\log n)^4}{n} \to 0$ as $n \to \infty$.
(A3) $\lambda \in L_\varepsilon(\beta)$.
(A4) The estimate b_k of β satisfies $(\beta - b_k)(N/k^3)^{1/2} \to 0$ in prob.

Theorem 1. *Suppose that assumptions* (A1)–(A3) *hold and that the below mentioned sequences l_k do not depend on random quantities other than k.*

(i) *If the sequence I_{l_k} satisfies $I_{l_k} \in S_k^*$ for all k, then*

$$\liminf_{m \to \infty} P(D_{l_k} - [2 \log k]^{1/2} \leqslant x) \geqslant \Phi(x) \quad \textit{for arbitrary } x.$$

(ii) *If l_k is chosen such that $I_{l_k} \cap S \neq \emptyset$ for all k, and additionally* (A4) *holds then*

$$\liminf_{m \to \infty} P(D_{l_k}^* - [2 \log k]^{1/2} \leqslant x) \geqslant \Phi(x) \quad \textit{for arbitrary } x.$$

(iii) *Finally, if the argmax set S of λ^* contains an interval, and for sequences l_k such that $I_{l_k} \cap S \neq \emptyset$ for all k, then $\lim_{m \to \infty} P(D_{l_k} - [2 \log k]^{1/2} \leqslant x) = \Phi(x)$ for arbitrary x.*

The proof of the result can be found in Section 6, an alternative proof could be given by directly approximating the Poisson by the normal distribution.

According to Theorem 1, the bias estimate should satisfy $(N/k^3)^{1/2}(b_k - \beta) \to 0$ for $C^{(\mathrm{P})}$ and $\tilde{C}^{(\mathrm{P})}$ to satisfy (3). Obviously, if k is chosen such that N/k^3 remains bounded, any consistent estimate b_k of β satisfies the required condition. Uniformly consistent kernel estimates of the derivative (see for instance Silverman, 1978) are an obvious choice to obtain an estimate b_k, but they require sufficient smoothness to work. Alternatively, a consistent estimate of β is obtained by taking the maximal differences

$$\max_{1 \leqslant i \leqslant m-1} |\hat{\lambda}_{n,h}(t_{i+1}) - \hat{\lambda}_{n,h}(t_i)|/(t_{i+1} - t_i) \tag{6}$$

of kernel density estimates $\hat{\lambda}_{n,h}(\cdot)$ on a grid $t_1, \ldots, t_{m(n)}$ of shrinking maximal width, again with the accuracy depending on the smoothness of λ.

A further alternative approach that does not require estimating β is to set $b_k = 0$ and make the bias negligible by undersmoothing. This is done by choosing k such that $Nk^{-3} \to 0$ in probability. This approach has been proposed for instance in Cowling et al. (1996).

All these methods to deal with bias carry over to the confidence sets of Section 3 that are obtained via kernel estimates.

While the above-mentioned confidence sets satisfy our asymptotic requirements, they turned out to be quite conservative in our simulations. Despite Theorem 1 (iii) this is true even with uniform intensities at least for our considered sample sizes. Similar effects occur when uniform confidence bands for density and intensity functions are based on asymptotic expansions such as those proposed by Bickel and Rosenblatt (1973).

In order to improve the actual coverage probabilities of uniform confidence bands, the use of bootstrap has been proposed by Cowling et al. (1996). Whereas the bootstrap seems to provide good results in this context, analogous resampling methods do not work when applied to D_l or D_l^*. Similar problems have been observed in several situations related to hypothesis testing. (See e.g. Beran, 1986 or Hinkley, 1987, 1988.) As a counter-example in our situation, consider constant intensities. In this case the pivotal quantity $D_{l_k} - \sqrt{\frac{N}{k}}\,\delta_s(\lambda)$, where

$$\delta_s(\lambda) := \left(\max_t \lambda(t) - \lambda(s)\right) / \left(\max_t \lambda(t)\right)^{1/2},$$

equals D_{l_k}, and $D_{l_k} - [2\log k]^{1/2} \to N(0, 1)$ in distribution according to Theorem 1 (iii). Let now $D_{l_k}^{(\mathrm{B})}$ denote the bootstrap statistic obtained by resampling from the observed jump points. This is equivalent to resampling method 3 of Cowling et al. (1996). According to the proposition below, bootstrap fails since neither the distribution of $D_{l_k}^{(\mathrm{B})} - [2\log k]^{1/2}$ nor that of $D_{l_k}^{(\mathrm{B})} - D_{l_k}$ approaches the correct limiting distribution.

Proposition 1. *Assume* (A1), *that λ is constant and that $k^{2+\alpha}/n \to 0$ for some $\alpha > 0$. Then, for arbitrary sequences l_k, conditional on $N = n$, and for $n \to \infty$*

$$D_{l_k}^{(\mathrm{B})} - 2[\log k]^{1/2} \to N(0, 2) \ \textit{in distribution}$$

and

$$D_{l_k}^{(\mathrm{B})} - D_{l_k} - (2 - \sqrt{2})[\log k]^{1/2} \to N(0, 1) \ \textit{in distribution}.$$

While further resampling methods have been proposed in Cowling et al. (1996), the essential arguments of the proof do not depend on the specific resampling method. The main reason for the bootstrap failure turns out to be that the bias of the estimate of $\max_t \lambda(t)$ is not approximated consistently. An alternative approach would be to use m out of n bootstrap that is known to work in many situations where the classical bootstrap fails. (See e.g. Bickel et al., 1997.) It turns out however, that standard consistency results that ensure consistency of subsampling like those given in Section 2.2 of Politis et al. (1999) are not applicable in our situation. A further alternative approach would be the bootstrap from a smoothed version of the empirical distribution function. Romano (1988) observed that smoothing can overcome inconsistency when bootstrapping the mode directly. A similar approach has been taken by Ziegler (2001) in the regression context.

We do not explore the possibilities of subsampling and smoothed bootstrap further here, but propose to replace the quantity $q(k,\gamma)$ used in the definition of $C_c^{(P)}$ and $C^{(P)}$ by the quantiles $\tilde{q}(k,\gamma)$ obtained by applying the below stated algorithm. The resulting confidence sets will be denoted by $\tilde{C}_c^{(P)}$, and $\tilde{C}^{(P)}$, respectively.

Algorithm.

1. Assuming that our sample includes n jump points, calculate the empirical distribution function Λ_n for the jump points $X_1,\ldots,X_n$ taken modulo one.
2. Calculate $t_i = \Lambda_n(i/k)$ for $0 \leqslant i \leqslant k$. ($t_0 = 0$) Let $s_i = t_i - t_{i-1}$ for $1 \leqslant i \leqslant k$.
3. Take k normal random variables $Y_1,\ldots,Y_k$ such that $Y_i \sim N(0,s_i)$ and let $Y^* = \sum_{1\leqslant i\leqslant k} Y_i$. Calculate

$$\tilde{D}_i = \frac{\max_{1\leqslant j\leqslant k}(Y_j - s_j Y^*) - (Y_i - s_i Y^*)}{(\max_j s_j)^{1/2}}$$

4. Repeat step 3 several (e.g. 10,000) times and calculate the empirical quantile $q^*(k,\gamma)$ of $\tilde{D}_i$. The quantile q^* can be used as replacement of $q(k)$ to decide on the inclusion of interval i into J. Notice that this quantile has to be calculated separately for each i ($1 \leqslant i \leqslant k$).

The algorithm provides a normal approximation for D_i under the null hypothesis assumption that $I_i \cap S \neq \emptyset$. More specifically, it is easily verified that for $N = n$,

$$\tilde{D}_i = \frac{\max_j B\circ\Lambda_n(I_j) - B\circ\Lambda_n(I_i)}{[\max_j \mathrm{P}_n(I_j)]^{1/2}},$$

where Λ_n denotes the empirical distribution function of the jump points of $X(t)$ taken modulo 1, and B denotes a standard Brownian bridge that is independent of Λ_n.

Theorem 2. *Under assumptions* (A1)–(A3), *the confidence sets* $\tilde{C}_c^{(P)}$ *satisfy requirements* (4), *and under* (A1)–(A4) $\tilde{C}^{(P)}$ *satisfies* (3).

3. The kernel-based approach

With the jump points $X_i \in [0,1]$ ($1 \leqslant i \leqslant N$) of the inhomogeneous Poisson process $X(t)$ taken modulo 1, the classical kernel density estimate of λ is defined as

$$\hat{\lambda}_{N,h}(x) := \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right),$$

where h depends on N. To improve estimation at the boundaries of [0, 1], the periodicity of the intensity function can obviously be exploited.

Under the assumptions of Theorem 3 below, a level $1-\alpha$ confidence set is given by

$$C^{(K)} = \left\{ x : \sqrt{\frac{Nh(N)}{\int_{-1}^{1} K^2(u)\,\mathrm{d}u}} \frac{\max_t \hat{\lambda}_{N,h}(t) - \hat{\lambda}_{N,h}(x) - b_h h c_K}{(\max_t \hat{\lambda}_{N,h}(t))^{1/2}} \leqslant q(h, 1-\alpha) \right\}, \tag{7}$$

where $c_K = \int_{-1}^{1} |u||K(u)|\,\mathrm{d}u$. Again b_h denotes an estimate of β, see Section 2 for a discussion of possible choices for b_h. Furthermore $q(h, 1-\alpha) = (2\log h^{-1})^{1/2} + \Phi^{-1}(1-\alpha)$.

In order to investigate the coverage probabilities we need the following assumptions.

(B1) The bandwidth sequence $h=h(n)$ is chosen such that $h=h(n) \to 0$ and $\frac{nh}{\log h^{-1}} \to \infty$.
(B2) The kernel K is chosen to be Lipschitz continuous and to have support $[-1, 1]$.
(B3) The estimate b_k of β satisfies $(\beta - b_k)\left(Nh^3\right)^{1/2} \to 0$ in prob.

Theorem 3. *Under assumptions* (A3) *and* (B1)–(B3) *the confidence sets* $C^{(K)}$ *satisfy* (3) *as* $m \to \infty$.

Arguments analogous to those of the proof of Theorem 2 can be used to show that the crude approximation $q(h, 1-\alpha)$ may be replaced by the following more accurate simulation-based bound. Let $Y(t) = h^{-1/2} \int K[(t-s)/h]\,\mathrm{d}B \circ F_n(s)$. Then, given $N = n$, replace $q(h, 1-\alpha)$ by the $1-\alpha$ quantile of

$$\sqrt{\frac{1}{\int_{-1}^{1} K^2(u)\,\mathrm{d}u}} \frac{\max_t Y(t) - Y(x)}{(\max_t \hat{\lambda}_{N,h}(t))^{1/2}}.$$

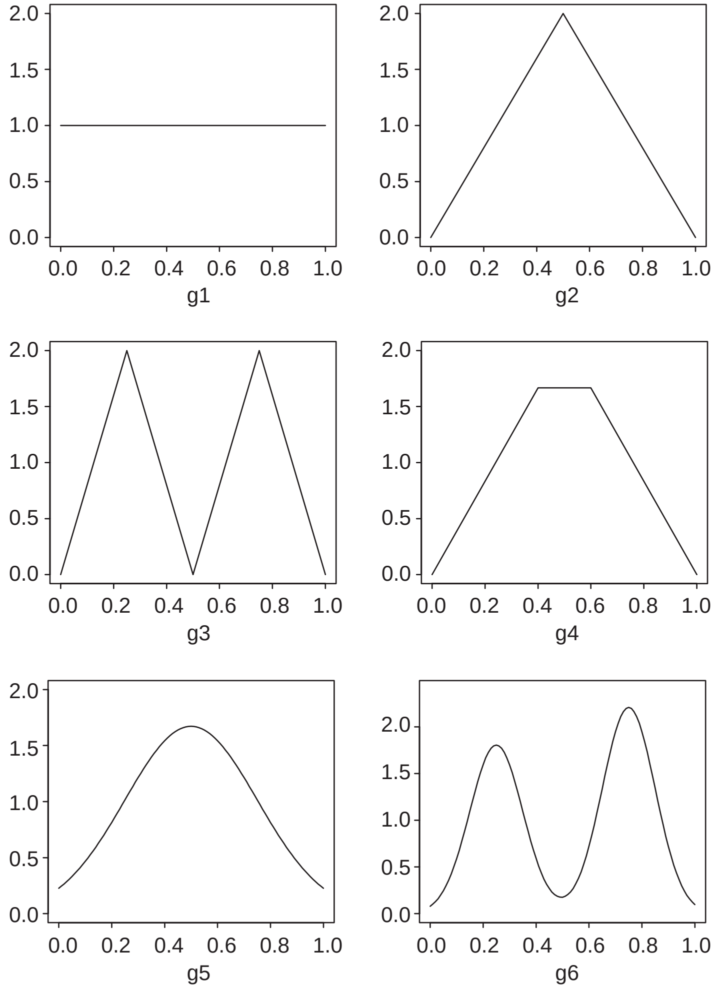
For practical convenience, the Brownian bridge B can be replaced by the uniform empirical process based on n observations.

Another practical issue is the selection of the bandwidth h. Bandwidth selection in the point process context and its analogies to density estimation is discussed in some detail in Diggle and Marron (1988). From practical point of view, it seems attractive to choose the same bandwidth for intensity estimation and confidence sets. This avoids (optical) contradictions between the curve and interval estimate, even though larger bandwidths are usually recommended for mode estimation given a sufficiently smooth underlying density.

4. Simulation results

We conducted a simulation study to investigate the performance of our proposed confidence sets and to compare them with sets constructed indirectly from uniform confidence bands by taking all points at which the upper confidence band exceeds the maximal value of the lower confidence band. As competitors, we considered the bands proposed by Hall and Titterington (1988) which require only Lipschitz continuity to be valid, as well as the resampling based bounds by Cowling et al. (1996) that have been derived under further differentiability assumptions. More specifically, resampling method 3 of Cowling et al. (1996) has been used in connection with their confidence bands $\mathscr{B}_1$ and their bias correction via undersmoothing.

We considered six different intensity functions on [0, 1] of different degree of smoothness and flatness. The corresponding densities $g_1, g_2, \ldots, g_6$ are displayed below.



We have taken $N = 200$ as well as $N = 500$ as expected numbers of Poisson jumps and 500 simulation runs have been carried out for each intensity. In all simulations, the coverage probability has been chosen equal to 0.90.

Both for the partitioning and kernel-based approach, we considered the quick and easy asymptotic critical values $q(k, 1-\alpha)$ and $q(h, 1-\alpha)$, as well as their simulation-based modifications $q^*(k, 1-\alpha)$ and $q^*(h, 1-\alpha)$. The Lipschitz bound β has been estimated according to (6). Alternatively, we considered an undersmoothing approach proposed in Cowling et al. (1996) and replaced the bandwidths h (cell widths k) by $h/2$ ($k/2$) when calculating simulation-based critical values. With the kernel estimates, we chose the Epanechnikov kernel and bandwidths by least-squares cross-validation, see Brooks and Marron (1991). A fixed number of cells ($k=10$ for $N=200$ and $k=20$ for $N=500$) has been used with the partitioning approach.

The results are summarized in Table 1. It turns out that all of our proposed confidence sets outperform those based on uniform confidence bands considerably. The replacement of $q(k,\gamma)$ by $q^*(k,\gamma)$ reduces the size of the confidence sets, and bias correction by undersmoothing provides a further significant improvement. According to our simulations, the coverage probability has been met in all considered situations, except for slight violations that turned up in some cases with the uniform intensity g_1. With g_1, the minimum empirical significance level was 0.86 and occurred with $\tilde{C}_c^{(P)}$ and for $N=200$. For all other intensities the empirical coverage probabilities have always been above the desired value.

The results were analogous both for smooth and nonsmooth intensities. Very often the partitioning approach worked better than the confidence sets $C^{(K)}$ (and modifications) based on kernel estimates. This turned out to be the case when the inverse of the number of cells $1/k$ was smaller than the average bandwidth provided by cross-validation, suggesting that small bandwidths should be desirable with respect to the size of our confidence sets. We plan to investigate this issue in future empirical work.

5. Application

One motivation of our paper has been the analysis of ship arrival data at Keelung harbor (Taiwan). In total $n=79,872$ arrivals have been recorded between July 1988 and June 1998. To provide information for personnel and harbor fee planning, our goal has been to identify time periods where the arrival intensity is maximal. For this purpose, we considered both daily and monthly fluctuations. For the monthly fluctuations, we assumed the intensity to be periodic with a period length of one year. This seems reasonable after correcting for the Chinese New Year vacation in late January or February (depending on the year), when the arrival intensity is typically lower. We considered each month separately and present results for January as an example. Based on a total of 6422 arrivals in January, Fig. 1 displays an estimate of the arrival intensity with multiple local extremes. We used the Epanechnikov kernel with bandwidth $h=1$, the wiggles are caused by the daily fluctuations.

Below we give the confidence sets $C_c^{(P)}$ and $\tilde{C}_c^{(P)}$ for the maximal average intensity based on a daily partition. The corresponding binned arrivals can be found in Fig. 2. (Divide by 10 to obtain average cumulative daily arrivals.) As coverage probability, we have chosen $1-\alpha=0.95$. As in our simulations, $\tilde{C}_c^{(P)}$ provides the smaller confidence sets.

Confidence set type	Included days	Percentage of cells included
$C_c^{(P)}$	2–7, 9–26 31	81
$\tilde{C}_c^{(P)}$	2–6, 10, 11, 13–15 17, 19–21	48

Table 1
Average sizes of confidence sets over 500 simulation runs for intensity densities g_1 to g_6

	Kernel based			Partition based			Avg. intensity		Uniform conf. bands	
n	$C^{(K)}$	$\tilde{C}^{(K)}$	$\tilde{C}_u^{(K)}$	$C^{(\mathrm{P})}$	$\tilde{C}^{(\mathrm{P})}$	$\tilde{C}_u^{(\mathrm{P})}$	$C_{\mathrm{c}}^{(\mathrm{P})}$	$\tilde{C}_{\mathrm{c}}^{(\mathrm{P})}$	$C^{(\mathrm{HT})}$	$C^{(\mathrm{CHP})}$
g_1										
200	1.00	0.99	0.88	1.00	0.98	0.90	0.99	0.86	1.00	1.00
500	1.00	0.99	0.89	1.00	0.97	0.89	0.99	0.86	1.00	1.00
g_2										
200	0.83	0.66	0.47	0.74	0.55	0.37	0.53	0.36	1.00	0.99
500	0.59	0.49	0.34	0.58	0.39	0.29	0.47	0.30	1.00	0.82
g_3										
200	0.81	0.68	0.46	0.89	0.67	0.38	0.56	0.37	1.00	1.00
500	0.66	0.55	0.35	0.66	0.48	0.32	0.49	0.33	1.00	0.99
g_4										
200	0.90	0.76	0.55	0.75	0.58	0.44	0.58	0.43	1.00	0.99
500	0.70	0.61	0.44	0.62	0.45	0.38	0.54	0.38	1.00	0.87
g_5										
200	0.91	0.77	0.57	0.79	0.59	0.43	0.60	0.42	1.00	0.99
500	0.71	0.62	0.43	0.64	0.46	0.38	0.55	0.38	1.00	0.90
g_6										
200	0.74	0.59	0.36	0.80	0.59	0.30	0.47	0.31	1.00	1.00
500	0.55	0.45	0.25	0.55	0.39	0.25	0.39	0.26	1.00	0.92

Confidence sets $C^{(\cdot)}$ use asymptotic critical values $q(\cdot)$, the sets $\tilde{C}^{\cdot}$ use simulated critical values $q^*(\cdot)$, the index $(\cdot)_u$ refers to bias correction by undersmoothing. The set $C^{(\mathrm{HT})}$ is based on uniform confidence bands according to Hall and Titterington (1988), whereas $C^{(\mathrm{CHP})}$ makes use of resampling based uniform confidence bands according to Cowling et al. (1996).

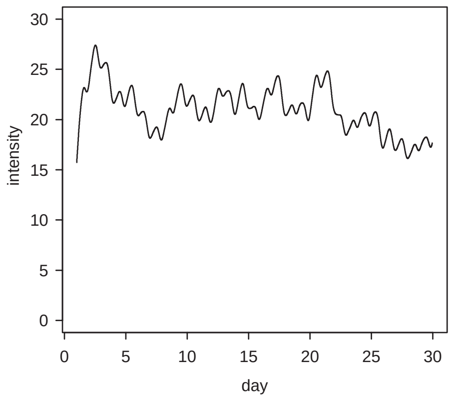


Fig. 1. Estimated intensity function for ships arriving at Keelung harbor in January.

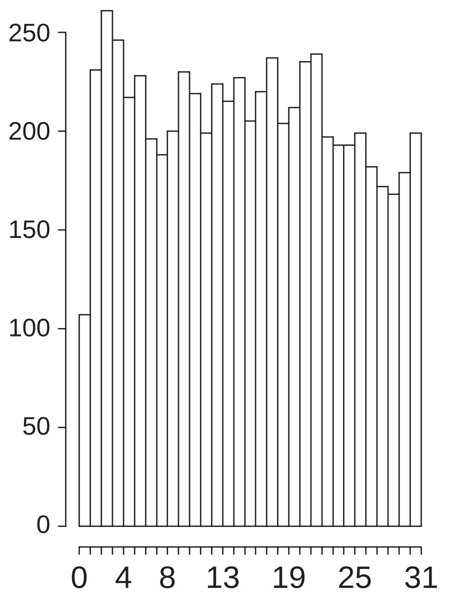


Fig. 2. Average daily number of ship arrivals in January.

With a bandwidth of 1 (day) and the undersmoothing approach, the kernel-based approach provided a still smaller confidence set, including only 21% of the month. The reason of the improvement can be found in the tendency to exclude less popular arrival times during a day when the kernel estimate is evaluated on a fine enough grid.

A look at the histogram of hourly arrivals in Fig. 3 (cumulative over the whole dataset) reveals high arrival intensities in the morning hours. Both partitioning-based confidence sets

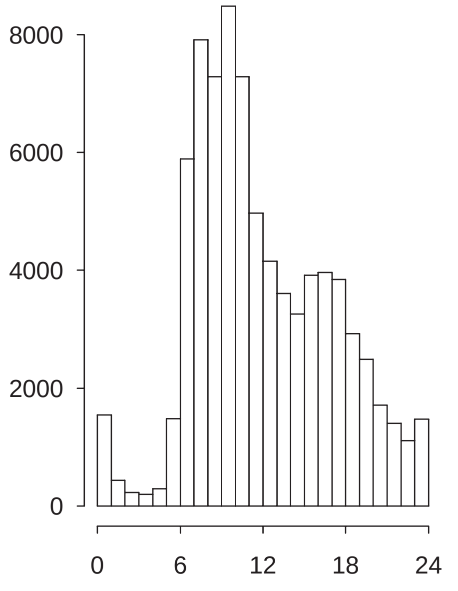


Fig. 3. Total number of arrivals for each hour in the dataset.

$C_c^{(P)}$ and $\tilde{C}_c^{(P)}$ single out the hour of 9:00–10:00 a.m. as the period of maximum intensity. This is maybe not surprising in view of the harbor fee policies that include an extra day of harbor fees for ships arriving before 9:00 a.m., causing affected ships often to wait before entering the harbor.

6. Theory

This section contains the proofs of the presented results. The proofs utilize the fact that conditionally on $N = n$, the jump points X_j have the same joint distribution as the order statistics of a sample of n independent observations $X_1, \ldots, X_n$ from density λ.

For later use, let Λ and P be the cumulative distribution function (cdf) and probability measure belonging to λ. Let furthermore Λ_n (and P_n) be the empirical distribution function (and measure) of a sample of size n from λ. Also, let $\tilde{\Lambda}_n = \sqrt{n}(\Lambda_n - \Lambda)$ and $\tilde{\mathrm{P}}_n = \sqrt{n}(\mathrm{P}_n - \mathrm{P})$ denote the corresponding empirical processes. Define $B_n(I) := B_n(b) - B_n(a)$ for Brownian bridges B_n and intervals $I = [a, b]$. For Λ_n, let as usual $\Lambda_n(I) = [\Lambda_n(a), \Lambda_n(b)]$. We first focus on the partitioning approach.

6.1. The partitioning approach

Lemma 1. *On a suitable probability space, there is a sequence of Brownian bridges B_n such that for arbitrary sequences of partitions $I_1, \ldots, I_{k(n)}$, for arbitrary sequences $l_k = l_{k(n)}$*

and as $n \to \infty$

$$\max_j P_n(I_j) - P_n(I_{l_k}) \leqslant d_{l_k} + n^{-1/2}\left[\max_j B_n \circ \Lambda(I_j) - B_n \circ \Lambda(I_{l_k})\right] + \mathrm{O}\left(\frac{(\log n)^2}{n}\right) \quad a.s. \tag{8}$$

Furthermore

$$\max_j P_n(I_j) - P_n(I_{l_k}) \leqslant d_{l_k} + n^{-1/2}\left[\max_j B_n \circ \Lambda_n(I_j) - B_n \circ \Lambda_n(I_{l_k})\right] + R_n, \tag{9}$$

where $d_{l_k} = \max_j P(I_j) - P(I_{l_k})$ *and* $R_n = \mathrm{O}\left(\left[\frac{(\log n)^2(\log \log n)}{n^3}\right]^{1/4}\right)$ *a.s. For constant intensities* $d_{l_k} = 0$, *and* "$\leqslant$" *may be replaced by* "$=$" *in* (8) *and* (9).

Proof. It is verified immediately that

$$\max_j \mathrm{P}_n(I_j) - \mathrm{P}_n(I_{l_k}) \leqslant \max_j P(I_j) - P(I_{l_k}) + n^{-1/2}\left(\max_j \tilde{\mathrm{P}}_n(I_j) - \tilde{\mathrm{P}}_n(I_{l_k})\right) \tag{10}$$

with equality for constant intensities.

According to the strong approximation due to Komlós et al. (1975), there is a sequence of Brownian bridges B_n on a suitable probability space such that $\lim\sup_n \|\tilde{\Lambda}_n - B_n \circ \Lambda\|_\infty = \mathrm{O}\left(\frac{(\log n)^2}{n^{1/2}}\right)$ a.s., and therefore

$$\max_j \tilde{\mathrm{P}}_n(I_j) - \tilde{\mathrm{P}}_n(I_{l_k}) = \max_j B_n \circ \Lambda(I_j) - B_n \circ \Lambda(I_{l_k}) + \mathrm{O}\left(\frac{(\log n)^2}{n^{1/2}}\right) \quad \text{a.s.} \tag{11}$$

This gives (8). To show (9) notice that

$$\max_j B_n \circ \Lambda(I_j) - B_n \circ \Lambda(I_{l_k}) = \max_j B_n \circ \Lambda_n(I_j) - B_n \circ \Lambda_n(I_{l_k}) + \mathrm{O}\left(\|B_n \circ \Lambda - B_n \circ \Lambda_n|_\infty\right). \tag{12}$$

But by Theorem 5.1.1 of Csörgő and Révész (1981) and by Theorem 9.25 in Karatzas and Shreve (1988) it follows that

$$\|B_n \circ \Lambda - B_n \circ \Lambda_n|_\infty = \mathrm{O}\left[\left(\frac{(\log n)^2(\log \log n)}{n}\right)^{1/4}\right] \quad \text{a.s.} \qquad \square \tag{13}$$

Proof of Theorem 1. Obviously it is sufficient to show assertion (i) conditional on $N = n$. Since, conditional on $N = n$,

$$D_{l_k} \stackrel{\mathrm{D}}{=} \sqrt{n}\,\frac{\max_{1 \leqslant j \leqslant k}[\mathrm{P}_n(I_j)] - \mathrm{P}_n(I_{l_k})}{[\max_{1 \leqslant j \leqslant k} \mathrm{P}_n(I_j)]^{1/2}}, \tag{14}$$

where $\overset{\mathrm{D}}{=}$ means equal in distribution, we may equivalently show that

$$\liminf_{n\to\infty} P\left(\sqrt{n}\frac{\max_{1\leqslant j\leqslant k}[\mathrm{P}_n(I_j)]-\mathrm{P}_n(I_{l_k})}{[\max_{1\leqslant j\leqslant k}\mathrm{P}_n(I_j)]^{1/2}}\leqslant\sqrt{2\log k}+x\right)\geqslant\Phi(x). \tag{15}$$

Now according to Lemma 1 and for $I_{l_k}\in S_k^*$,

$$\begin{aligned}&n^{1/2}\left(\max_j[\mathrm{P}_n(I_j)]-\mathrm{P}_n(I_{l_k})\right)\\&\quad\leqslant\max\ B_n\circ\Lambda(I_j)-B_n\circ\Lambda(I_{l_k})+\mathrm{O}\left(\frac{(\log n)^2}{n^{1/2}}\right)\quad\text{a.s.}\end{aligned} \tag{16}$$

Furthermore, by Theorem 9.25 in Karatzas and Shreve (1988),

$$\limsup_{k\to\infty}\max_j\ B_n\circ\Lambda(I_j)\leqslant\left(\frac{2\lambda_m\ \log(k)}{k}\right)^{1/2}+\mathrm{O}((k\ \log(k))^{-1/2})\quad\text{a.s.} \tag{17}$$

Also $B_n\circ\Lambda(I_{l_k})$ is normally $N(0,\sigma_k^2)$ distributed, where $\sigma_k^2=\Lambda(I_{l_k})(1-\Lambda(I_{l_k}))=\lambda_m/k(1+\mathrm{o}(1))$.

For the denominator we have that,

$$[\max_j\ \mathrm{P}_n(I_j)]^{-1/2}=\max_j\left([\mathrm{P}(I_j)]^{-1/2}\left[1+n^{-1/2}\frac{\tilde{\mathrm{P}}_n(I_j)}{\mathrm{P}(I_j)}\right]^{-1/2}\right), \tag{18}$$

with $\max_j\mathrm{P}(I_j)=\lambda_m/k+\mathrm{O}(k^{-2})$. Furthermore, let j^* be chosen such that $\mathrm{P}_n(I_{j^*})=\max_j\mathrm{P}_n(I_j)$. Then, $\mathrm{P}(I_{j^*})=\lambda_m/k(1+\mathrm{o_P}(1))$, and according to Shorack and Wellner (1986, p. 542), it follows that

$$\frac{\tilde{\mathrm{P}}_n(I_{j^*})}{\mathrm{P}(I_{j^*})}=\mathrm{O}\left(k\left[\frac{\log k}{k}\right]^{1/2}\right)\quad\text{a.s.}$$

By plugging in the expanded denominator, we get

$$\sqrt{n}\frac{\max_{1\leqslant j\leqslant k}[\mathrm{P}_n(I_j)]-\mathrm{P}_n(I_{l_k})}{[\max_{1\leqslant j\leqslant k}\mathrm{P}_n(I_j)]^{1/2}}\leqslant(2\ \log\ k)^{1/2}-V_{l_k}+\mathrm{o_P}(1), \tag{19}$$

where $V_{l_k}=[\max_j\mathrm{P}_n(I_j)]^{-1/2}B_n\circ\Lambda(I_{l_k})$ is asymptotically normal $N(0,1)$ distributed.

To prove assertion (ii), notice that b_k provides an asymptotically valid bias bound, if $I_{l_k}\cap S\neq\emptyset$.

For simplicity of notation, we prove assertion (iii) for constant intensities, the case where S contains an interval is analogous. By Lemma 1, equality holds in (16). Equality for the denominator follows , since $\lambda_m=1$. □

Proof of Proposition 1. Let P_n^* denote the empirical measure of a bootstrap sample. With $T_n^* = \sqrt{\max_j \mathrm{P}_n^*(I_j)}$ and $\tilde{\mathrm{P}}_n^* = \sqrt{n}(\mathrm{P}_n^* - \mathrm{P}_n)$, we may write

$$\begin{aligned} D_{l_k}^{(\mathrm{B})} &= \frac{\sqrt{n}}{T_n^*}\left(\max_j \mathrm{P}_n^*(I_j) - \mathrm{P}_n^*(I_{l_k})\right) \\ &= [T_n^*]^{-1}\left(\max_j [\tilde{\mathrm{P}}_n^*(I_j) + \tilde{\mathrm{P}}_n(I_j)] - [\tilde{\mathrm{P}}_n^*(I_{l_k}) + \tilde{\mathrm{P}}_n(I_{l_k})]\right), \end{aligned}$$

where $[T_n^*]^{-1} = \sqrt{k}(1 + \mathrm{O_P}([(k \log k)/n]^{1/2}))$. Now conditionally on the jumps $X_1, \ldots, X_n$ and as in Lemma 1, we may use a strong approximation by independent sequences of Brownian bridges B_n and B_n^* to obtain

$$\begin{aligned} D_{l_k}^{(\mathrm{B})} = [T_n]^{-1}&\left(\max_j [B_n^* \circ \Lambda_n(I_j) + B_n \circ \Lambda_n(I_j)] - [B_n^* \circ \Lambda_n(I_{l_k}) + B_n \circ \Lambda_n(I_{l_k})]\right) \\ &+ \mathrm{o_P}(1) \end{aligned} \tag{20}$$

which is equal to

$$[T_n]^{-1}\left(\max_j [\sqrt{2}\tilde{B}_n \circ \Lambda_n(I_j)] - [\sqrt{2}\tilde{B}_n \circ \Lambda_n(I_{l_k})]\right) + \mathrm{o_P}(1)$$

in distribution, for some sequence of Brownian bridges $\tilde{B}_n$. By (14) and Lemma 1 it follows that conditionally on $N = n$,

$$D_{l_k} = [T_n]^{-1}\left[\max_j B_n \circ \Lambda_n(I_j) - B_n \circ \Lambda_n(I_{l_k})\right] + \mathrm{o_P}(1). \tag{21}$$

Thus $D_{l_k}^{(\mathrm{B})}$ and $\sqrt{2}D_{l_k}$ have the same limiting distribution and the first assertion follows. The second assertion follows by subtracting (21) from (20) followed by arguments analogous to those given in the proof of Theorem 1 (iii). □

Proof of Theorem 2. Consider first $\tilde{C}_{\mathrm{c}}^{(\mathrm{P})}$ and assume that $I_{l_k} \in S_k^*$. As in the proof of Theorem 1, we may condition on $N = n$ and start from representation (14). According to (8), since $d_{l_k} = 0$ when $I_{l_k} \in S_k^*$, and by (18) it follows that

$$\begin{aligned} D_{l_k} &\leqslant \frac{\max_j B_n \circ \Lambda(I_j) - B_n \circ \Lambda(I_{l_k})}{[\max_j \mathrm{P}(I_j)]^{1/2}} + \mathrm{o_P}(1) \\ &=: \tilde{D}_{l_k}^* + \mathrm{o_P}(1) \end{aligned}$$

on a suitable probability space. Since, by Theorem 1, $D_{l_k} - \sqrt{2 \log k}$ is stochastically not larger than a normal $N(0, 1)$ random variable, the remainder term may be neglected and quantiles of $\tilde{D}_{l_k}^*$ provides critical values satisfying (4). In order to establish (4) also for $\tilde{D}_{l_k}$ it suffices to show that $\tilde{D}_{l_k}^* - \tilde{D}_{l_k} = \mathrm{o_P}(1)$ as $n \to \infty$, which follows from (12) and (13) by applying again (18).

The arguments concerning $\tilde{C}^{(\mathrm{P})}$ are analogous. □

6.2. The kernel-based approach

Lemma 2. *Assume that* $\lambda \in L_\varepsilon(\beta)$. *Then for any* $t \in S_\varepsilon$, *small enough* h *and conditionally on* $N = n$

$$|\lambda(t) - E\hat{\lambda}_{n,h}(t)| \leqslant \beta h \int_{-1}^{1} |u||K(u)|\,\mathrm{d}u.$$

Proof. The result follows by bounding

$$|\lambda(t) - E\hat{\lambda}_{n,h}(t)| \leqslant \frac{1}{h} \int |K\left(\frac{t-s}{h}\right)||\lambda(t) - \lambda(s)|\,\mathrm{d}s. \qquad \square$$

Proof of Theorem 3. Consider again the coverage probabilities conditionally on $N = n$ and let $n \to \infty$. We will first derive an upper bound for

$$a_n \frac{\max_t \hat{\lambda}_{n,h}(t)}{\sqrt{\max_t \hat{\lambda}_{n,h}(t)}} = a_n \sqrt{\max_t \hat{\lambda}_{n,h}(t)},$$

where $a_n := (nh/\int_{-1}^{1} K^2(u)\,\mathrm{d}u)^{1/2}$. By using Theorem 3.1 of Bickel and Rosenblatt (1973), we obtain that

$$\max_t \hat{\lambda}_{n,h}(t) \leqslant \lambda_m + \sqrt{\frac{2\lambda_m \log h^{-1}}{nh} \int_{-1}^{1} K^2(u)\,\mathrm{d}u} + \mathrm{o}([nh]^{-1}) \quad \text{a.s.}$$

By Taylor expansion and the above bound it follows that

$$a_n \sqrt{\max_t \hat{\lambda}_{n,h}(t)} \leqslant a_n \sqrt{\lambda_m} + \frac{1}{2}\sqrt{2 \log h^{-1}} + \mathrm{o}((nh)^{-1/2}). \tag{22}$$

Assume now that x is a point where λ attains its global maximum and consider

$$\begin{aligned}\hat{\lambda}_{n,h}(x) &= \lambda_m + (\hat{\lambda}_{n,h}(x) - E\hat{\lambda}_{n,h}(x)) + (E\hat{\lambda}_{n,h}(x) - \lambda_m) \\ &=: \lambda_m + T_{1,n} + T_{2,n}.\end{aligned}$$

According to the classical central limit theorem $T_{1,n}/(\mathrm{Var}[T_{1,n}])^{1/2} \to N(0,1)$ in distribution. Since $\max_t \hat{\lambda}_{n,h}(t) \to \lambda_m > 0$ in prob., and by a standard approximation of $\mathrm{Var}\hat{\lambda}_{n,h}(x)$ we obtain that

$$\frac{nh\,\mathrm{Var}[T_{1,n}]}{\max_t[\hat{\lambda}_{n,h}(t)] \int_{-1}^{1} K^2(u)\,\mathrm{d}u} \to 1 \quad \text{in probab.},$$

and therefore $\frac{a_n}{\sqrt{\max_t \hat{\lambda}_{n,h}(t)}} T_{1,n} \to N(0,1)$ in distribution. Furthermore, by Lemma 2 $T_{2,n} \leqslant \beta h c_k$, and according to (B3)

$$a_n \frac{\hat{\lambda}_{n,h}(x) + b_h}{\sqrt{\max_t \hat{\lambda}_{n,h}(t)}} \geqslant a_n \sqrt{\lambda_m} - \frac{1}{2}\sqrt{2 \log h^{-1}} + T^*_{1,n} + o_P(1).$$

Thus Theorem 3 follows. □

References

Barnard, G.A., 1953. Time intervals between accidents—a note on Maguire, Pearson and Wynn's paper. Biometrika 40, 212–213.
Beran, R., 1986. Simulated power functions. Ann. Statist. 14, 151–173.
Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Ann. Statist. 1, 1071–1095.
Bickel, P.J., Götze, F., Van Zwet, W.R., 1997. Resampling fewer than n observations: gains, losses and remedies for losses. Statist. Sinica 7, 1–31.
Brooks, M.M., Marron, J.S., 1991. Asymptotic optimality of the least squares cross-validation bandwidth for kernel estimates of intensity functions. Stochastic Process. Appl. 38, 157–165.
Cheng, M.-Y., Hall, P., 1998. On mode testing and empirical approximations to distributions. Statist. Prob. Lett. 39, 245–254.
Chernoff, H., 1964. Estimation of the mode. Ann. Inst. Stat. Math. 16, 31–41.
Cowling, A., Hall, P., Phillips, M.J., 1996. Bootstrap confidence regions for the intensity of a Poisson point process. J. Amer. Statist. Assoc. 91, 1516–1524.
Cox, D.R., Lewis, P.A.W., 1966. The Statistical Analysis of Series of Events. Methuen, London.
Csörgő, M., Révész, P., 1981. Strong Approximations in Probability and Statistics. Academic Press, New York.
Diggle, P., 1985. A kernel method for smoothing point process data. Appl. Statist. 34, 138–147.
Diggle, P., Marron, J.S., 1988. Equivalence of smoothing parameter selectors in density and intensity estimation. J. Amer. Statist. Assoc. 83, 793–800.
Grund, B., Hall, P., 1995. On the minimization of L_p error in mode estimation. Ann. Statist. 23, 2264–2284.
Hall, P., Titterington, D.M., 1988. On confidence bands in nonparametric density estimation and regression. J. Multivariate Anal. 27, 228–254.
Hartigan, J.A., Hartigan, P.M., 1985. The dip test of unimodality. Ann. Statist. 13, 70–84.
Hinkley, D.V., 1987. Bootstrap significance tests. in: Proceedings of the 47th Session of International Statistical Institute, Paris. pp. 65–74.
Hinkley, D.V., 1988. Bootstrap methods. J. R. Statist. Soc. B 50, 321–337.
Karatzas, I., Shreve, S.E., 1988. Brownian Motion and Stochastic Calculus. Springer, New York.
Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent R.V.'s and the sample DF.I, Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111–131.
Leadbetter, M.R., Wold, D., 1983. On estimation of point process intensities. in: Sen, P.K. (Ed.), Contributions to Statistics: Essays in Honor of Norman L. Johnson. North Holland, Amsterdam, pp. 299–312.
Mammen, E., Marron, J.S., Fisher, N.I., 1992. Some asymptotics for multimodality tests based on kernel density estimates. Probab. Theory Related Fields 91, 115–132.
Minnotte, M.C., Scott, D.W., 1993. The mode tree: a tool for visualization of nonparametric density features. J. Comput. Graph. Statist. 2, 51–68.
Minnotte, M.C., Marchette, D.J., Wegman, E.J., 1998. The bumpy road to the mode forest. J. Comput. Graph. Statist. 7, 239–251.
Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065–1076.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.
Romano, J.P., 1988. On weak convergence and optimality of kernel density estimates of the mode. Ann. Statist. 16, 629–647.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.
Silverman, B.W., 1978. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6, 177–184.
Solow, A.R., 1991. An exploratory analysis of the occurrence of explosive volcanism in the Northern Hemisphere, 1851–1985. J. Amer. Statist. Assoc. 86, 49–54.
Ziegler, K., 2001. On bootstrapping the mode in the nonparametric regression model with random design. Metrika 53, 141–170.