The Annals of Statistics
2007, Vol. 35, No. 6, 2691–2722
DOI: 10.1214/009053607000000280

SOME THEORETICAL RESULTS ON NEURAL SPIKE TRAIN PROBABILITY MODELS

BY HOCK PENG CHAN AND WEI-LIEM LOH

National University of Singapore

This article contains two main theoretical results on neural spike train models, using the counting or point process on the real line as a model for the spike train. The first part of this article considers template matching of multiple spike trains. P-values for the occurrences of a given template or pattern in a set of spike trains are computed using a general scoring system. By identifying the pattern with an experimental stimulus, multiple spike trains can be deciphered to provide useful information.

The second part of the article assumes that the counting process has a conditional intensity function that is a product of a free firing rate function s, which depends only on the stimulus, and a recovery function r, which depends only on the time since the last spike. If s and r belong to a q-smooth class of functions, it is proved that sieve maximum likelihood estimators for s and r achieve the optimal convergence rate (except for a logarithmic factor) under L_1 loss.

1. Introduction. In the field of neuroscience, it is generally acknowledged that neurons are the basic units of information processing in the brain. They play this role by generating characteristic and highly peaked electric action potentials of very short duration, or more simply, *spikes* (cf. Dayan and Abbott [11]). These spikes can travel along nerve fibers that extend over relatively long distances to other cells. The temporal pattern of these spikes depends dynamically on the stimuli of the neuron or the biochemicals induced by the spikes of other neurons. The collection of such spikes generated by a neuron over a time period is called a *spike train*. In this way, information is transmitted via spike trains. Because the spikes are of very short duration and are highly peaked, point processes or counting processes are the most commonly used probability models for neural spike trains, with points on the time axis representing the temporal locations of the spikes (cf. Brillinger [3]).

Sections 2 and 3 deal with the detection of multiple spike train patterns. Let $N_i(T)$ be the number of spikes of the ith template neuron in the time interval $[0, T)$ and $\mathbf{w}^{(i)} = \{w_1^{(i)}, \dots, w_{N_i(T)}^{(i)}\}$ the corresponding spike times. We are

Received June 2005; revised May 2006.

AMS 2000 subject classifications. Primary 62E20; secondary 62G20, 62M20.

Key words and phrases. Boundary crossing probability, conditional intensity, counting process, importance sampling, neural spike train, Poisson process, scan statistics, sieve maximum likelihood estimation, template matching.

interested in establishing the presence of the template $\mathbf{w} := (\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(d)})$ inside a longer spike train pattern $\mathbf{y} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(d)})$. Loosely speaking, the template $\mathbf{w}$ is said to have occurred at time t in the spike trains $\mathbf{y}$ if for most $y \in \mathbf{y}^{(i)} \cap [t, t+T)$, $1 \le i \le d$, there exists $w \in \mathbf{w}^{(i)}$ such that $y - t$ is close to w. A more precise definition of a match, via a user-chosen score function, is given in Section 2. When the number of matches is significantly large, we can identify the onset of the patterns $\mathbf{w}$ in $\mathbf{y}$ with the stimulus provided when $\mathbf{w}$ are recorded. For example, $\mathbf{w}$ can be the spike times of an assembly of neurons of a zebra finch when its own song is played while it is awake and $\mathbf{y}$ the spike trains of the same assembly when it is sleeping. The replaying of these patterns during sleep has been observed and hypothesized to play an important role in bird song learning (cf. Dave and Margoliasch [10] and Mooney [22]).

It was observed in Brown, Kass and Mitra [4] that "research in statistics and signal processing on multivariate point process models has not been nearly as extensive as research on models of multivariate continuous-valued processes" in a section titled "Future challenges for multiple spike train data analysis." In Sections 2 and 3, an asymptotic theory of scan statistics in multivariate point processes is developed and applied to the template matching problem. The finite-sample accuracy of these results is then checked via computer experiments.

The second part of the article assumes that the spike train is modeled as a counting process with a conditional intensity function that is a product of a free firing rate function, which depends only on the stimulus, and a recovery function, which depends only on the time since the last spike. More specifically, let $N(t)$ denote the number of spikes on the interval $[0, t)$ and $w_1 < \cdots < w_{N(t)}$ be the spike times occurring in $[0, t)$. Suppose the following conditional intensity function exists: $\lambda(t|w_1, \ldots, w_{N(t)}) = \lim_{\delta \downarrow 0} \delta^{-1} E[N(t+\delta) - N(t)|w_1, \ldots, w_{N(t)}]$, a.s.

In the neuroscience literature, a number of probability models for λ have been proposed. One of the simplest is when λ depends only on t. This leads to a nonhomogeneous Poisson process (cf. Ventura et al. [29]). It is well known that for a short period of time after a spike has been discharged, it is more difficult, or even impossible, for a neuron to fire another spike (cf. Dayan and Abbott [11]). Such a time interval is called the *refractory period*. The main drawback of the nonhomogeneous Poisson process model is that it does not incorporate the refractory period of the neuron. To account for this, a number of researchers (cf. Johnson and Swami [17], Miller [21], Berry and Meister [1] and Kass and Ventura [18]) have proposed modeling the conditional intensity function λ by

$$\lambda_1\bigl(t|w_1, \ldots, w_{N(t)}\bigr) = \begin{cases} s(t), & \text{if } N(t) = 0, \\ s(t) r\bigl(t - w_{N(t)}\bigr), & \text{if } N(t) \ge 1, \end{cases} \tag{1}$$

where s, r are nonnegative functions. s and r are known as the *free firing rate function* and the *recovery function*, respectively. This model is Markovian in that it depends only on the present time t and the duration $t - w_{N(t)}$ since the last spike.

Section 4 considers sieve maximum likelihood estimation of s and r in (1) based on n independent realizations of $N(t)$, $t \in [0, T)$, where $0 < T < \infty$. Here, we assume that the true free firing rate function s and recovery function r both lie in the class of q-smooth functions $\Theta_{\boldsymbol{\kappa},q}$ [defined as in (56)]. Assuming that there exists an absolute refractory period, it is proved in Theorems 4 and 5 that the sieve MLE's for s and r both achieve essentially the optimal convergence rate (except for a logarithmic factor) under L_1 loss.

2. Template matching with continuous kernels. Let $\mathbf{w} = (\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(d)})$ be the spike train pattern of an assembly of d neurons recorded when an experimental stimulus is provided to a subject, where $\mathbf{w}^{(i)} = \{w_1^{(i)}, \dots, w_{N_i(T)}^{(i)}\}$ are the spike times of the ith neuron over the period $[0, T)$. The same neurons are subsequently observed for a longer time period when the subject is engaged in other activities and the corresponding spike trains $\mathbf{y} = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(d)})$ are checked for occurrences of the template $\mathbf{w}$.

For $t \geq 0$, let $\mathbf{y}_t = (\mathbf{y}_t^{(1)}, \dots, \mathbf{y}_t^{(d)})$, where $\mathbf{y}_t^{(i)} = \{y - t : y \in \mathbf{y}^{(i)} \cap [t, t+T)\}$. There are various algorithms in the neuroscience literature that have been used to determine if there is a close match between $\mathbf{y}_t$ and $\mathbf{w}$. In Grün, Diesmann and Aertsen [15], T is chosen small and a match is declared if $\{1 \leq i \leq d : \mathbf{w}^{(i)} = \varnothing\} = \{1 \leq i \leq d : \mathbf{y}_t^{(i)} = \varnothing\}$. In the sliding sweeps algorithm (cf. Dayhoff and Gerstein [12] and Nádasdy et al. [23]), a match is declared if

$$\sup_{1 \leq i \leq d} \sup_{w \in \mathbf{w}^{(i)}} \inf_{y - t \in \mathbf{y}_t^{(i)}} |y - t - w| \leq \Delta,$$

where $\Delta > 0$ is a predetermined constant. In this section, we shall study the pattern-filtering algorithm (cf. Chi, Rauske and Margoliasch [8]), which uses a scoring system to measure the proximity between $\mathbf{w}$ and $\mathbf{y}_t$.

Let f be a nonincreasing and nonconstant function on $[0, \infty)$ with $f(0) > 0$. The score between $\mathbf{w}$ and $\mathbf{y}_t$ is given by

$$S_t = \sum_{i=1}^{d} S_t^{(i)}, \qquad \text{where } S_t^{(i)} = T^{-1} \sum_{y - t \in \mathbf{y}_t^{(i)}} \max_{w \in \mathbf{w}^{(i)}} f(|y - t - w|). \tag{2}$$

For a given template $\mathbf{w}$, define the kernel functions

$$g_{\mathbf{w}}^{(i)}(u) = \Big[\max_{w \in \mathbf{w}^{(i)}} f(|u - w|)\Big] \mathbf{1}_{\{0 \leq u < T\}} \qquad \forall i = 1, \dots, d. \tag{3}$$

We can then also express $S_t^{(i)} = T^{-1} \sum_{y \in \mathbf{y}^{(i)}} g_{\mathbf{w}}^{(i)}(y - t)$. The graph of $S_t^{(i)}$ against t is thus a normalized sum of the kernels $g_{\mathbf{w}}^{(i)}(y - \cdot)$ over all $y \in \mathbf{y}^{(i)}$. We declare a match between $\mathbf{y}_t$ and $\mathbf{w}$ to be present when the proximity score S_t exceeds a predetermined threshold level c. To prevent overcounting, a match at time t is declared to be new only if the overlap between $[t, t + T)$ and the time interval of the previous new match is less than αT for some $0 < \alpha < 1$. More specifically, let $\sigma_1 = \inf\{t : S_t \geq c\}$ and $\sigma_{j+1} = \inf\{t > \sigma_j + (1-\alpha)T : S_t \geq c\}$ for $j \geq 1$. The number of new matches between the spike trains $\mathbf{y}$ over the time interval $[0, a+T)$ and the template $\mathbf{w}$ is then $U_a := \sup\{j : \sigma_j \leq a\}$, with the convention that $U_a = 0$ if $\sigma_1 > a$.

To prevent the occurrence of too many (false) matches when $\mathbf{y}$ is pure noise, the threshold level c must be chosen reasonably large. For a large, there can be, on average, more than one new (false) match between $\mathbf{y}$ and $\mathbf{w}$. The Poisson distribution is often used for modeling U_a to compute the p-value under such circumstances. For small a, the occurrence of a single match would itself be rare and we can use the probability of having at least one match as the p-value. For this purpose, we study

$$M_a := \sup_{0 \leq t \leq a} S_t \quad \text{and} \quad V_c := \inf\{t : S_t \geq c\},$$

which are the scan statistic and its dual, the time to detection, respectively. In this section, we obtain their asymptotic distributions when f is continuous on $[0, \infty)$ and deal with discontinuous score functions in Section 3.

2.1. *Main results.* Let $\mathbf{y}^{(i)}$, $i = 1, \ldots, d$, be independent Poisson processes with constant intensity $\lambda_i > 0$. Consider the following regularity conditions on $\mathbf{w}$ and f.

(A1) Let $\mathbf{w}^{(i)} = \mathbf{w}_*^{(i)} \cap [0, T)$, where $\mathbf{w}_*^{(1)}, \ldots, \mathbf{w}_*^{(d)}$ are point processes on $[0, \infty)$ with each $\mathbf{w}_*^{(i)}$ ergodic, stationary and having nonconstant interarrival times.

(A2) Let f be continuous and let there be a possibly empty finite set H such that the second derivative of f is uniformly continuous and bounded over any interval inside $\mathbb{R}^+ \setminus H$. Moreover,

$$0 < \sup_{x \in \mathbb{R}^+ \setminus H} \left| \frac{d}{dx} f(x) \right| < \infty \quad \text{and} \quad \lim_{x \to \infty} f(x) > -\infty. \tag{4}$$

Let $\mu_{\mathbf{w}} = T^{-1} \sum_{i=1}^d \lambda_i \int_0^T g_{\mathbf{w}}^{(i)}(u)\, du$ be the expected value of S_t conditioned on known $\mathbf{w}$ and let the large deviation rate function of S_t be

$$\phi_{\mathbf{w}}(c) = \sup_{\theta > 0} \left[\theta c - T^{-1} \sum_{i=1}^d \lambda_i \int_0^T \left(e^{\theta g_{\mathbf{w}}^{(i)}(u)} - 1 \right) du \right] \qquad \text{for } c > \mu_{\mathbf{w}}. \tag{5}$$

We shall denote by $\theta_{\mathbf{w}}$ $(= \theta_{\mathbf{w},c})$ the unique value of $\theta > 0$ that attains the supremum on the right-hand side of (5). By the stationarity of $\mathbf{w}_*^{(i)}$ in (A1), for all $y \in \mathbb{R}$, the distribution of $\max_{w \in \mathbf{w}_*^{(i)}} f(|y - w|)$ is equal to the distribution of $Z_i := \max_{w \in \mathbf{w}_*^{(i)}} f(|w|)$. Hence, by (A1) and (A2), for all $\theta > 0$,

$$T^{-1} \int_0^T e^{\theta g_{\mathbf{w}}^{(i)}(u)}\, du \to E e^{\theta Z_i} \quad \text{and} \quad T^{-1} \int_0^T g_{\mathbf{w}}^{(i)}(u)\, du \to E Z_i \qquad \text{a.s.} \tag{6}$$

as $T \to \infty$. Let $\mu = \sum_{i=1}^{d} \lambda_i E Z_i$ and

$$\phi(c) = \sup_{\theta > 0} \left[\theta c - \sum_{i=1}^{d} \lambda_i (E e^{\theta Z_i} - 1) \right] \qquad \text{for } c > \mu. \tag{7}$$

Let θ_* $(= \theta_{*,c})$ be the unique value of $\theta > 0$ attaining the supremum on the right-hand side of (7). Then by (5)–(7), $\mu_{\mathbf{w}} \to \mu$, $\phi_{\mathbf{w}} \to \phi$ [pointwise on (μ, ∞)] and $\theta_{\mathbf{w}} \to \theta_*$ a.s. as $T \to \infty$. Similarly,

$$\begin{aligned} v_{\mathbf{w}} &:= T^{-1} \sum_{i=1}^{d} \lambda_i \int_0^T [g_{\mathbf{w}}^{(i)}(u)]^2 e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\, du, \\ \tau_{\mathbf{w}} &:= T^{-1} \sum_{i=1}^{d} \lambda_i \int_0^T \left[\frac{d}{du} g_{\mathbf{w}}^{(i)}(u) \right]^2 e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\, du \end{aligned} \tag{8}$$

both converge almost surely to positive constants as $T \to \infty$. Let $P_{\mathbf{w}}$ denote the probability measure conditioned on a known $\mathbf{w}$ and $\zeta_{\mathbf{w}} = (2\pi)^{-1} (\tau_{\mathbf{w}} / v_{\mathbf{w}})^{1/2}$.

PROPOSITION 1. *Assume* (A1)–(A2). *Then for any* $t \ge 0$, $\Delta > 0$ *and* $c > \mu$,

$$P_{\mathbf{w}} \left\{ \sup_{t < u \le t + \Delta} S_u \ge c \right\} \sim \Delta \zeta_{\mathbf{w}} e^{-T \phi_{\mathbf{w}}(c)} \qquad \textit{a.s. as } T \to \infty. \tag{9}$$

By piecing together the boundary crossing probabilities in (9), we obtain the following.

THEOREM 1. *Assume* (A1)–(A2).

(a) *Let* $c > \mu$. *The distribution* (*conditional on* $\mathbf{w}$) *of* $\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} V_c$ *converges to the exponential distribution with mean* 1 *almost surely as* $T \to \infty$.

(b) *Let* $a \to \infty$ *as* $T \to \infty$ *in such a way that* $(\log a)/T$ *converges to a positive constant. Let* $c_{\mathbf{w}} > \mu_{\mathbf{w}}$ *satisfy* $\phi_{\mathbf{w}}(c_{\mathbf{w}}) = (\log a)/T$. *Then for any* $z \in \mathbb{R}$,

$$P_{\mathbf{w}}\{\theta_{\mathbf{w}} T (M_a - c_{\mathbf{w}}) - \log \zeta_{\mathbf{w}} \ge z\} \to 1 - \exp(-e^{-z}) \qquad \textit{a.s. as } T \to \infty.$$

(c) *Let* $a \to \infty$ *as* $T \to \infty$ *in such a way that* $(\log a)/T$ *converges to a positive constant. Let* c $(= c_T)$ *be such that* $\eta_{\mathbf{w}} := a \zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} \to \eta > 0$ *almost surely. Then*

$$P_{\mathbf{w}}\{U_a = k\} - e^{-\eta_{\mathbf{w}}} \frac{\eta_{\mathbf{w}}^k}{k!} \to 0 \qquad \text{a.s. } \forall k = 0, 1, \ldots. \tag{10}$$

REMARK 1. In Theorem 1 of Chi [7], it was shown [without the regularity condition (A2)] that $\lim_{T \to \infty} T^{-1} \log V_c \to \phi(c)$ a.s. for all $c > \mu$. The question of whether $\log V_c = T\phi_{\mathbf{w}}(c) + o(T^{1/2})$ was also raised in a remark on page 157. It follows from Theorem 1(a) that the more precise $\log V_c = T\phi_{\mathbf{w}}(c) + O_P(1)$ holds.

REMARK 2. We can extend Theorem 1(c) to deal with piecewise constant rate functions $\lambda_i(t)$. Let $(a_0 =)0 < a_1 < \cdots < a_{n-1} < a(= a_n)$ be such that $(a_j - a_{j-1})$ are large compared to T for all j. Let $\lambda_i(t) = \lambda_{ij}$ for all $a_{j-1} < t \le a_j$. Define $\phi_{\mathbf{w},j}$, $\tau_{\mathbf{w},j}$ and $\upsilon_{\mathbf{w},j}$ as in (5) and (8), with λ_i replaced by λ_{ij}. Then Poisson approximation can be used with $\eta_{\mathbf{w}} = \sum_{j=1}^n (a_j - a_{j-1})\zeta_{\mathbf{w},j} e^{-T\phi_{\mathbf{w},j}(c)}$, where $\zeta_{\mathbf{w},j} = (2\pi)^{-1}(\tau_{\mathbf{w},j}/\upsilon_{\mathbf{w},j})^{1/2}$.

2.2. *Implementation.* We conduct a small-scale simulation study here to test the finite-sample accuracy of the analytic approximations in Theorem 1. An alternative to analytic approximations is to compute the p-values $p_{\mathbf{w}} := P_{\mathbf{w}}\{M_a \ge c\}$ via direct Monte Carlo simulations. However, as p-values of interest are often small, a large number of simulation runs is required for these estimations to be accurate. The computational cost is compounded when the time period $[0, a + T)$ of $\mathbf{y}^{(i)}$ is large.

We shall now introduce an importance sampling alternative for the simulation of p-values. Even though the probability of interest is with respect to a homogeneous Poisson process $\mathbf{y}^{(i)}$, a nonhomogeneous Poisson process is used to generate $\mathbf{y}^{(i)}$ so that $\{M_a \ge c\}$ is encountered more frequently. Likelihood ratio weights reflecting the change of measure are then introduced to ensure that estimates are unbiased with respect to the underlying homogeneous Poisson process. The same change of measure is also used to prove Proposition 1. Analogous techniques for computing p-values have been used in sequential analysis (cf. Siegmund [26]), change-point detection (cf. Lai and Shan [20]) and DNA sequence alignments (cf. Chan [5]).

Let $P_{\theta,t}$ denote the probability measure under which $\mathbf{y}^{(i)}$ is generated as a Poisson point process with intensity $\eta_i(v) = \lambda_i e^{\theta g_{\mathbf{w}}^{(i)}(v-t)}$ for each i. Note that $g_{\mathbf{w}}^{(i)}(v - t) = 0$ for $v \notin [t, t + T)$ and hence the change of measure occurs only for the generation of spikes in the interval $[t, t + T)$. The likelihood of $\mathbf{y}_t^{(i)}$ under $P_{\theta,t}$ is given by

$$L_{\theta,t}(\mathbf{y}_t^{(i)}) = \exp\left(-\lambda_i \int_0^T e^{\theta g_{\mathbf{w}}^{(i)}(u)}\, du\right) \prod_{y \in \mathbf{y}_t^{(i)}} \lambda_i e^{\theta g_{\mathbf{w}}^{(i)}(y-t)}.$$

Hence, by (5), the likelihood ratio

$$\begin{aligned}\frac{dP_{\theta_{\mathbf{w}},t}}{dP_{\mathbf{w}}}(\mathbf{y}) &= \prod_{i=1}^d \frac{L_{\theta_{\mathbf{w}},t}(\mathbf{y}_t^{(i)})}{L_{0,t}(\mathbf{y}_t^{(i)})} \\ &= \prod_{i=1}^d \exp\left[\theta_{\mathbf{w}} T S_t^{(i)} - \lambda_i \int_0^T \left(e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)} - 1\right) du\right] \qquad (11)\end{aligned}$$

$$
\begin{aligned}
&= \exp\left[\theta_{\mathbf{w}} T S_t - \sum_{i=1}^{d} \lambda_i \int_0^T \big(e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)} - 1\big)\, du\right] \\
&= \exp[T\phi_{\mathbf{w}}(c) + T\theta_{\mathbf{w}}(S_t - c)].
\end{aligned}
$$

In our importance sampling algorithm, we first select a small $\Delta > 0$ such that $J := a/\Delta$ is a positive integer. For each simulation run, we generate j randomly from $\{0, \dots, J\}$ followed by $\mathbf{y}$ from $P_{\theta_{\mathbf{w}}, j\Delta}$. The estimate

$$
\begin{aligned}
\widehat{p} &= (J+1)\left[\sum_{j=0}^{J} \frac{dP_{\theta_{\mathbf{w}}, j\Delta}}{dP_{\mathbf{w}}}(\mathbf{y})\right]^{-1} \mathbf{1}_{\{M_a \ge c\}} \\
&= (J+1)\exp\left[\sum_{i=1}^{d} \lambda_i \int_0^T \big(e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)} - 1\big)\, du\right]\left(\sum_{j=0}^{J} e^{\theta_{\mathbf{w}} T S_{j\Delta}}\right)^{-1} \mathbf{1}_{\{M_a \ge c\}}
\end{aligned} \tag{12}
$$

is then unbiased for $p_{\mathbf{w}}$. The averages of (12) over all the simulation runs is then the importance sampling estimate of $p_{\mathbf{w}}$.

EXAMPLE 1. Consider the Hamming window function

$$
f(t) = \begin{cases} \dfrac{1}{2}(1-\beta) + \dfrac{1}{2}(1+\beta)\cos\left(\dfrac{\pi t}{\epsilon}\right), & \text{if } 0 \le t < \epsilon, \\ -\beta, & \text{if } t \ge \epsilon, \end{cases} \tag{13}
$$

with $\epsilon = 5$ ms and $\beta = 0.4$ (see, e.g., Chi, Rauske and Margoliash [8]).

We generate a template $\mathbf{w}$ over the time interval from 0 to $T = 500$ ms on $d = 4$ spike trains, with interarrival distance X ms between two spikes satisfying

$$
P\{X \le x\} = 1 - e^{-(x-1)^+/24}. \tag{14}
$$

This corresponds to an absolute refractory period or "dead time" of 1 ms after each spike in $\mathbf{w}^{(i)}$ before the next spike can be generated. Further discussion and results on the implications of a refractory period in estimating the spike train intensity will be provided in Section 4. In our computer experiment, a total of 80 spikes were first generated on the four spike trains using (14).

To compute p-values using direct Monte Carlo simulations, we generated 2000 realizations of $\mathbf{y}$ by using Poisson point processes with constant intensity $\lambda_i = 0.04$ ms^{-1} on the interval from 0 to $a + T = 20$ s. The proportion of times that $\{M_a \ge c\}$ occurs is taken as the estimate of $p_{\mathbf{w}}$. For importance sampling, 2000 simulation runs were also executed using the algorithm described earlier by choosing $\Delta = 0.2$ ms.

For the analytic approximation, we apply Theorem 1(a), which gives us

$$
\begin{aligned}
P_{\mathbf{w}}\{M_a \ge c\} &= P_{\mathbf{w}}\{V_c \le a\} \\
&\doteq 1 - \exp\big(-a\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)}\big).
\end{aligned} \tag{15}
$$

We see from the results summarized in Table 1 that there is substantial variance reduction when importance sampling is used. The analytic approximations have also

TABLE 1
Estimates of $P_{\mathbf{w}}\{M_a \geq c\} \pm$ standard error

c	**Direct MC**	**Imp. sampling**	**Anal. approx. (15)**
0.017	0.037 ± 0.004	0.0387 ± 0.0019	0.0383
0.018	0.024 ± 0.003	0.0237 ± 0.0012	0.0241
0.019	0.016 ± 0.003	0.0158 ± 0.0008	0.0149
0.020	0.009 ± 0.002	0.0095 ± 0.0005	0.0091
0.021	0.005 ± 0.002	0.0054 ± 0.0003	0.0055
0.022	0.003 ± 0.001	0.0033 ± 0.0002	0.0033

been shown to be quite accurate, lying within two standard errors of the importance sampling estimate in all cases considered.

2.3. *Proofs.* We preface the proofs of Proposition 1 and Theorem 1 with the following preliminary lemmas. We shall let $\lfloor\cdot\rfloor$ denote the greatest integer function. Let $P_{\theta_{\mathbf{w}},t}$ be the change of measure defined at the beginning of Section 2.2 and let $P_{\theta_{\mathbf{w}}} = P_{\theta_{\mathbf{w}},0}$.

LEMMA 1. *Let $t \geq 0$ and $c > \mu_{\mathbf{w}}$. Then*

$$P_{\mathbf{w}}\{S_t \geq c\} \sim (2\pi v_{\mathbf{w}})^{-1/2}\theta_{\mathbf{w}}^{-1}T^{-1/2}e^{-T\phi_{\mathbf{w}}(c)} \qquad \text{a.s. as } T \to \infty.$$

PROOF. Let $E_{\theta,t}$ denote expectation with respect to the probability measure $P_{\theta,t}$. Let $I_T = [z_T, z_T + \varepsilon_T)$ with $\varepsilon_T = o(T^{-1/2})$. Then, by (11),

$$P_{\mathbf{w}}\{T^{1/2}(S_t - c) \in I_T\} = e^{-T\phi_{\mathbf{w}}(c)} E_{\theta_{\mathbf{w}},t}\big[e^{T\theta_{\mathbf{w}}(c-S_t)}\mathbf{1}_{\{T^{1/2}(S_t-c)\in I_T\}}\big]. \tag{16}$$

By similar computations, for any $y \in \mathbb{R}$,

$$\begin{aligned} P_{\mathbf{w}}\{S_t \geq c + y\} &= e^{-T\phi_{\mathbf{w}}(c)} E_{\theta_{\mathbf{w}},t}\big[e^{T\theta_{\mathbf{w}}(c-S_t)}\mathbf{1}_{\{S_t \geq c+y\}}\big] \\ &\leq e^{-T\phi_{\mathbf{w}}(c) - T\theta_{\mathbf{w}}y}. \end{aligned} \tag{17}$$

Under $P_{\theta_{\mathbf{w}},t}$, $TS_t^{(i)} = \sum_{y\in\mathbf{y}^{(i)}} g_{\mathbf{w}}^{(i)}(y-t)$ is compound Poisson with expected number of summands $\eta_i = \lambda_i \int_0^T e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\,du$, each summand identically distributed as $g_{\mathbf{w}}^{(i)}(U_i)$, where U_i has positive density $(\lambda_i/\eta_i)e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}$ for $u \in [0, T)$. We note that

$$\begin{aligned} E_{\theta_{\mathbf{w}},t}\big[g_{\mathbf{w}}^{(i)}(U_i)\big] &= (\lambda_i/\eta_i)\int_0^T g_{\mathbf{w}}^{(i)}(u) e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\,du \\ &= \eta_i^{-1}\frac{d}{d\theta}\int_0^T \lambda_i e^{\theta g_{\mathbf{w}}^{(i)}(u)}\,du\bigg|_{\theta=\theta_{\mathbf{w}}}. \end{aligned}$$

Since $\theta_{\mathbf{w}}$ maximizes the right-hand side of (5), it follows that

$$
\begin{aligned}
E_{\theta_{\mathbf{w}},t}[TS_t] &= \sum_{i=1}^{d} \eta_i E_{\theta_{\mathbf{w}}}\big[g_{\mathbf{w}}^{(i)}(U_i)\big] \\
&= \frac{d}{d\theta} \sum_{i=1}^{d} \lambda_i \int_0^T \big(e^{\theta g_{\mathbf{w}}^{(i)}(u)} - 1\big)\, du\Big|_{\theta=\theta_{\mathbf{w}}} = Tc.
\end{aligned} \tag{18}
$$

A compound Poisson $Y = \sum_{j=1}^{N} Y_j$ satisfies $\mathrm{Var}(Y) = (EN)(EY_1^2)$. Hence, by (8),

$$
\mathrm{Var}_{\theta_{\mathbf{w}},t}(S_t) = T^{-2} \sum_{i=1}^{d} \eta_i \int_0^T \big[g_{\mathbf{w}}^{(i)}(u)\big]^2 (\lambda_i/\eta_i) e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\, du = T^{-1} v_{\mathbf{w}}. \tag{19}
$$

By (18) and (19), $T^{1/2}(S_t - c)$ is asymptotically normal with mean 0 and variance $v_{\mathbf{w}}$. Hence, by equation (5) of Stone [28],

$$
\begin{aligned}
&P_{\theta_{\mathbf{w}},t}\{T^{1/2}(S_t - c) \in I_T\} \\
&\quad = (2\pi v_{\mathbf{w}})^{-1/2} \int_{I_T} e^{-z^2/(2v_{\mathbf{w}})}\, dz + o_T(1)(\varepsilon_T + T^{-1/2})
\end{aligned} \tag{20}
$$

almost surely as $T \to \infty$, where $o_T(1)$ does not depend on ε_T and z_T. Let $\varepsilon_T T^{1/2} \to 0$ sufficiently slowly that $o_T(1)/(\varepsilon_T T^{1/2}) \to 0$. Then, by (16), (17) and (20),

$$
\begin{aligned}
P_{\mathbf{w}}\{S_t \ge c\} &= \sum_{k=0}^{\lfloor \varepsilon_T^{-1} \rfloor} P_{\mathbf{w}}\{k\varepsilon_T \le T^{1/2}(S_t - c) < (k+1)\varepsilon_T\} + P_{\mathbf{w}}\{S_T \ge c + T^{-1/2}\} \\
&\sim (2\pi v_{\mathbf{w}})^{-1/2} e^{-T\phi_{\mathbf{w}}(c)} \int_0^\infty e^{-T^{1/2}\theta_{\mathbf{w}} z - z^2/(2v_{\mathbf{w}})}\, dz \qquad \text{a.s. as } T \to \infty,
\end{aligned}
$$

and Lemma 1 holds. □

LEMMA 2. *Assume* (A1)–(A2). *There exists $\epsilon_T = o(T^{-1/2})$ such that for all uniformly bounded intervals $I_{1,T}$, $I_{2,T}$ of length ϵ_T, as $T \to \infty$,*

$$
\begin{aligned}
&P_{\theta_{\mathbf{w}},t}\Big\{T^{1/2}\Big(S_t - c, \frac{d}{dx} S_x\Big|_{x=t}\Big) \in I_{1,T} \times I_{2,T}\Big\} \\
&\quad \sim (2\pi)^{-1} (v_{\mathbf{w}} \tau_{\mathbf{w}})^{-1/2} \Big(\int_{z_1 \in I_{1,T}} e^{-z_1^2/(2v_{\mathbf{w}})}\, dz_1\Big) \\
&\qquad \times \Big(\int_{z_2 \in I_{2,T}} e^{-z_2^2/(2\tau_{\mathbf{w}})}\, dz_2\Big) \qquad \textit{a.s.}
\end{aligned} \tag{21}
$$

PROOF. By stationarity, we may assume without loss of generality that $t = 0$. Under $P_{\theta_{\mathbf{w}}}$, $(TS_0^{(i)}, T\frac{d}{dx} S_x^{(i)}|_{x=0})'$ is bivariate compound Poisson with Poisson mean $\eta_i = \lambda_i \int_0^T e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\, du$ and each summand distributed as $(g_{\mathbf{w}}^{(i)}(U_i),$

$-\frac{d}{du}g_{\mathbf{w}}^{(i)}(u)|_{u=U_i})'$, where U_i is a random variable on $[0, T)$ with density $(\lambda_i/\eta_i)\times e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}$.

We shall now compute the means and covariances of $(S_0, \frac{d}{dx}S_x|_{x=0})' = \sum_{i=1}^d (S_0^{(i)}, \frac{d}{dx}S_x^{(i)}|_{x=0})'$ under $P_{\theta_{\mathbf{w}}}$. Since $g_{\mathbf{w}}^{(i)}$ is bounded, it follows that

$$
\begin{aligned}
E_{\theta_{\mathbf{w}}}\left[\frac{d}{dx}S_x\Big|_{x=0}\right] &= T^{-1}\sum_{i=1}^d \eta_i E_{\theta_{\mathbf{w}}}\left[-\frac{d}{du}g_{\mathbf{w}}^{(i)}(u)\Big|_{u=U^{(i)}}\right] \\
&= -T^{-1}\sum_{i=1}^d \lambda_i \int_0^T \left[\frac{d}{du}g_{\mathbf{w}}^{(i)}(u)\right] e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\,du \\
&= -T^{-1}\sum_{i=1}^d \lambda_i \theta_{\mathbf{w}}^{-1}\big(e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(T)} - e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(0)}\big) \qquad (22) \\
&= O(T^{-1}) \qquad \text{a.s.}
\end{aligned}
$$

The bivariate compound Poisson $(Y, Z)' = \sum_{j=1}^N (Y_j, Z_j)'$ satisfies $\mathrm{Cov}(Y, Z) = E(N)E(Y_1 Z_1)$. Hence,

$$
\begin{aligned}
\mathrm{Cov}_{\theta_{\mathbf{w}}}\left(S_0, \frac{d}{dx}S_x\Big|_{x=0}\right) &= -T^{-2}\sum_{i=1}^d \eta_i E_{\theta_{\mathbf{w}}}\left[g_{\mathbf{w}}^{(i)}(U_i)\left(\frac{d}{du}g_{\mathbf{w}}^{(i)}(u)\Big|_{u=U_i}\right)\right] \\
&= -T^{-2}\sum_{i=1}^d \lambda_i \int_0^T \left[\frac{d}{du}g_{\mathbf{w}}^{(i)}(u)\right] g_{\mathbf{w}}^{(i)}(u) e^{\theta g_{\mathbf{w}}^{(i)}(u)}\,du \\
&= -T^{-2}\sum_{i=1}^d \lambda_i \big[\theta_{\mathbf{w}}^{-1} g_{\mathbf{w}}^{(i)}(u) - \theta_{\mathbf{w}}^{-2}\big] e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u)}\Big|_{u=0}^{u=T} \qquad (23) \\
&= O(T^{-2}) \qquad \text{a.s.}
\end{aligned}
$$

Since $\mathrm{Var}_{\theta_{\mathbf{w}}}(S_0) \sim T^{-1}v_{\mathbf{w}}$ and $\mathrm{Var}_{\theta_{\mathbf{w}}}(\frac{d}{dx}S_x|_{x=0}) \sim T^{-1}\tau_{\mathbf{w}}$, it follows from equation (5) of Stone [28], (18), (22) and (23) that

$$
\begin{aligned}
&P_{\theta_{\mathbf{w}},t}\left\{T^{1/2}\left(S_t - c, \frac{d}{dx}S_x\Big|_{x=t}\right) \in I_{1,T} \times I_{2,T}\right\} \\
&\quad = (2\pi)^{-1}(v_{\mathbf{w}}\tau_{\mathbf{w}})^{-1/2} \\
&\qquad \times \left(\int_{z_1 \in I_{1,T}} e^{-z_1^2/(2v_{\mathbf{w}})}\,dz_1\right)\left(\int_{z_2 \in I_{2,T}} e^{-z_2^2/(2\tau_{\mathbf{w}})}\,dz_2\right) + o_T(1)(\varepsilon_T^2 + T^{-1}),
\end{aligned}
$$

where $o_T(1)$ does not depend on $I_{j,T}$, $j = 1, 2$. Lemma 2 follows by selecting ε_T such that $\varepsilon_T T^{1/2} \to 0$ and $o_T(1)/\varepsilon_T^2 T \to 0$. $\square$

LEMMA 3. *Let κ, T, K and c be positive constants. Let*

$$s(u) = c + z_1 T^{-1/2} + u z_2 T^{-1/2} - u^2 K/2.$$

Then $\sup\{s(u):0<u<\kappa T^{-1/2}\}\geq\max\{c,s(0),s(\kappa T^{-1/2})\}$ *if and only if* $z_2/K\in[0,\kappa]$ *and* $z_1\geq -z_2^2/(2KT^{1/2})$.

PROOF. The quadratic s is maximized at $u=z_2/(KT^{1/2})$. If $z_2/K\in[0,\kappa]$, then

$$\sup\{s(u):0<u<\kappa T^{-1/2}\}=s(z_2/KT^{1/2})=c+z_1/T^{1/2}+z_2^2/2KT$$

and Lemma 3 easily follows. □

Let $A_t(=A_{t,\kappa,c,T})=\{\sup_{t<u<t+\kappa T^{-1/2}}S_u\geq\max(c,S_t,S_{t+\kappa T^{-1/2}})\}$ and $D_t(=D_{t,\kappa,c,T})=A_t\cap\{S_t\geq c-\kappa^2\theta_{\mathbf{w}}\tau_{\mathbf{w}}(2T)^{-1}\}$.

LEMMA 4. *Assume* (A1)–(A2). *Then for any* $\kappa>0$, $t\geq 0$ *and* $c>\mu$,

$$P_{\mathbf{w}}(A_t)\sim P_{\mathbf{w}}(D_t)\sim\kappa T^{-1/2}\zeta_{\mathbf{w}}e^{-T\phi_{\mathbf{w}}(c)}\qquad a.s., \tag{24}$$

where $\zeta_{\mathbf{w}}=(2\pi)^{-1}(\tau_{\mathbf{w}}/v_{\mathbf{w}})^{1/2}$.

PROOF. Assume without loss of generality that $t=0$. Let H_i $(=H_{i,\mathbf{w}})$ be the set of all v such that a second derivative does not exist at $g_{\mathbf{w}}^{(i)}(v)$. By (A1)–(A2), the number of elements in H_i is $O(T)$ a.s. for all i. Let $0<u<\kappa T^{-1/2}$ and let $y\in\mathbf{y}^{(i)}$ be such that $y-h\notin(0,u)$ for all $h\in H_i$. Then by (A2) and the mean value theorem,

$$g_{\mathbf{w}}^{(i)}(y-u)-g_{\mathbf{w}}^{(i)}(y)+u\frac{d}{dv}g_{\mathbf{w}}^{(i)}(v)\Big|_{v=y}=\frac{u^2}{2}\left(\frac{d^2}{dv^2}g_{\mathbf{w}}^{(i)}(v)\Big|_{v=\xi}\right) \tag{25}$$

for some $y-u\leq\xi\leq y$. If $y\in\mathbf{y}^{(i)}\cap(h,h+u)$ for some $h\in H_i\setminus\{0,T\}$, then

$$\begin{aligned}
&g_{\mathbf{w}}^{(i)}(y-u)-g_{\mathbf{w}}^{(i)}(y)+u\frac{d}{dv}g_{\mathbf{w}}^{(i)}(v)\Big|_{v=y}\\
&\quad=\int_{y-u}^{y}\left(\frac{d}{dv}g_{\mathbf{w}}^{(i)}(v)\Big|_{v=y}-\frac{d}{d\xi}g_{\mathbf{w}}^{(i)}(\xi)\right)d\xi \qquad (26)\\
&\quad=(h+u-y)\left(\frac{d}{dv}g_{\mathbf{w}}^{(i)}(v)\Big|_{v\downarrow h}-\frac{d}{dv}g_{\mathbf{w}}^{(i)}(v)\Big|_{v\uparrow h}\right)+o(u^2).
\end{aligned}$$

The case $y-T\in(0,u)$ and $y\in(0,u)$ has negligible contribution and hence by adding up (25) and (26) over $y\in\mathbf{y}^{(i)}$ for $i=1,\dots,d$ and dividing by T, we obtain

$$S_u-S_0-u\frac{d}{dv}S_v\Big|_{v=0}=-\frac{C_{\mathbf{w},u}u^2}{2}, \tag{27}$$

where $C_{\mathbf{w},u}$ is an expression derived from the right-hand sides of (25) and (26). It is shown in Chan and Loh [6] that

$$\lim_{T\to\infty}\sup_{0<u<\kappa T^{-1/2}}u^2T|C_{\mathbf{w},u}-\theta_{\mathbf{w}}\tau_{\mathbf{w}}|\to 0\qquad\text{a.s. under } P_{\theta_{\mathbf{w}}}. \tag{28}$$

Then, by Lemma 2, the change of variables

$$z_1 = T^{1/2}(S_0 - c) \quad \text{and} \quad z_2 = T^{1/2}\frac{d}{dx}S_x\Big|_{x=0}, \tag{29}$$

substituting $K = \theta_{\mathbf{w}}\tau_{\mathbf{w}}$ into Lemma 3, (11) and (27), we have

$$\begin{aligned}
P_{\mathbf{w}}(A_0) &= e^{-T\phi_{\mathbf{w}}(c)} E_{\theta_{\mathbf{w}}}\big[e^{T\theta_{\mathbf{w}}(c-S_0)}\mathbf{1}_{\{\sup_{0<u<\kappa T^{-1/2}} S_u \ge \max(c, S_0, S_{\kappa T^{-1/2}})\}}\big] \\
&\sim \frac{e^{-T\phi_{\mathbf{w}}(c)}}{2\pi(v_{\mathbf{w}}\tau_{\mathbf{w}})^{1/2}} \int_0^{\kappa\theta_{\mathbf{w}}\tau_{\mathbf{w}}} \int_{-z_2^2/(2\theta_{\mathbf{w}}\tau_{\mathbf{w}}T^{1/2})}^{\infty} e^{-T^{1/2}\theta_{\mathbf{w}}z_1 - z_1^2/2v_{\mathbf{w}} - z_2^2/2\tau_{\mathbf{w}}}\,dz_1\,dz_2 \\
&\sim \frac{e^{-T\phi_{\mathbf{w}}(c)}}{(2\pi)(v_{\mathbf{w}}\tau_{\mathbf{w}})^{1/2}} \\
&\quad \times \int_0^{\kappa\theta_{\mathbf{w}}\tau_{\mathbf{w}}} e^{-z_2^2/2\tau_{\mathbf{w}}}\big(-T^{-1/2}\theta_{\mathbf{w}}^{-1}e^{-T^{1/2}\theta_{\mathbf{w}}z_1}\big)\Big|_{z_1=-z_2^2/(2\theta_{\mathbf{w}}\tau_{\mathbf{w}}T^{1/2})}^{z_1=\infty}\,dz_2,
\end{aligned}$$

which gives us the right-hand side of (24). Since the constraint $z_1 \ge -z_2^2/(2\theta_{\mathbf{w}}\tau_{\mathbf{w}}T^{1/2}) \ge -(\kappa\theta_{\mathbf{w}}\tau_{\mathbf{w}})^2/(2\theta_{\mathbf{w}}\tau_{\mathbf{w}}T^{1/2})$ is satisfied in the integrals above, it follows from (29) that $P_{\mathbf{w}}(D_0)$ is also asymptotically equal to the right-hand side of (24). $\square$

The proof of the next lemma is shown in Chan and Loh [6].

LEMMA 5. *Assume* (A1)–(A2). *There exists $r_\kappa = o(\kappa)$ as $\kappa \to \infty$ such that for all $t \ge 0$, with probability* 1,

$$\begin{aligned}
&\sum_{\ell=2}^{\lfloor T^{3/2}/\kappa\rfloor - 1} P_{\mathbf{w}}\{S_t \ge c - \kappa^2\theta_{\mathbf{w}}\tau_{\mathbf{w}}(2T)^{-1}, S_{t+\ell\kappa T^{-1/2}} \ge c - \kappa^2\theta_{\mathbf{w}}\tau_{\mathbf{w}}(2T)^{-1}\} \\
&\qquad + P_{\mathbf{w}}(A_t \cap A_{t+\kappa T^{-1/2}}) \le r_\kappa T^{-1/2}e^{-T\phi_{\mathbf{w}}(c)},
\end{aligned} \tag{30}$$

for all large T.

PROOF OF PROPOSITION 1. By stationarity, we may assume without loss of generality that $t = 0$. By Lemmas 1, 4, 5 and the inequalities

$$\begin{aligned}
\sum_{q=0}^{\lfloor \Delta/(\kappa T^{-1/2})\rfloor - 1}\Bigg[& P_{\mathbf{w}}(D_{q\kappa T^{-1/2}}) - P_{\mathbf{w}}\big(A_{q\kappa T^{-1/2}} \cap A_{(q+1)\kappa T^{-1/2}}\big) \\
& - \sum_{\ell=2}^{\lfloor \Delta/(\kappa T^{-1/2})\rfloor - 1 - q} P_{\mathbf{w}}\bigg\{S_{q\kappa T^{-1/2}} \ge c - \frac{\kappa^2\theta_{\mathbf{w}}\tau_{\mathbf{w}}}{2T}, \\
& \qquad\qquad S_{(q+\ell)\kappa T^{-1/2}} \ge c - \frac{\kappa^2\theta_{\mathbf{w}}\tau_{\mathbf{w}}}{2T}\bigg\}\Bigg]
\end{aligned}$$

$$\leq P_{\mathbf{w}}\left\{\sup_{0<u\leq\Delta} S_u \geq c\right\} \leq \sum_{q=0}^{\lfloor\Delta/(\kappa T^{-1/2})\rfloor} [P_{\mathbf{w}}(A_{q\kappa T^{-1/2}}) + P_{\mathbf{w}}\{S_{q\kappa T^{-1/2}} \geq c\}],$$

it follows that for any $\varepsilon > 0$, there exists κ arbitrarily large such that $r_\kappa \leq \varepsilon\kappa\zeta_{\mathbf{w}}$, $(2\pi v_{\mathbf{w}})^{-1/2}\theta_{\mathbf{w}}^{-1} \leq \varepsilon\kappa\zeta_{\mathbf{w}}$ and the inequality

$$\left| P_{\mathbf{w}}\left\{\sup_{0<u\leq\Delta} S_u \geq c\right\} \Big/ (\Delta\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)}) - 1 \right| \leq 2\varepsilon$$

holds for all large T with probability 1. It remains to choose ε arbitrarily small. □

PROOF OF THEOREM 1. Let $z \in \mathbb{R}$ and let ξ $(= \xi_{\mathbf{w}})$ be such that $\xi/T \to \infty$ and k $(= k_{\mathbf{w}}) := ze^{T\phi_{\mathbf{w}}(c)}/\zeta_{\mathbf{w}}\xi$ is a positive integer tending to infinity almost surely. Define $B_j = \{\sup_{(j-1)\xi\leq t<j\xi-T} S_t \geq c\}$ and $C_j = \{\sup_{j\xi-T\leq t\leq j\xi} S_t \geq c\}$. Then

$$P_{\mathbf{w}}\left(\bigcup_{j=1}^{k} B_j\right) \leq P_{\mathbf{w}}\{\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} V_c \leq z\} \leq P_{\mathbf{w}}\left(\bigcup_{j=1}^{k} B_j\right) + \sum_{j=1}^{k} P_{\mathbf{w}}(C_j). \tag{31}$$

Conditioned on known $\mathbf{w}$, the event B_j depends only on the spike train times of $\mathbf{y}^{(i)}$ lying inside $[(j-1)\xi, j\xi)$. Since these intervals are disjoint for different j, it follows that $B_1, \ldots, B_k$ are independent, conditioned on $\mathbf{w}$. By Lemmas 4 and 5, it follows that with probability 1,

$$\begin{aligned} &P_{\mathbf{w}}(B_j) \sim (\xi - T)\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} \sim z/k, \\ &P_{\mathbf{w}}(C_j) \sim T\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} \qquad \forall 1 \leq j \leq k. \end{aligned} \tag{32}$$

Since $k \to \infty$ a.s. as $T \to \infty$, with probability 1, we have

$$P_{\mathbf{w}}\left(\bigcup_{j=1}^{k} B_j\right) = 1 - \prod_{j=1}^{k} P_{\mathbf{w}}(B_j^c) = 1 - (1 - z/k)^k + o(1) \to 1 - e^{-z}. \tag{33}$$

Moreover, because $\xi/T \to \infty$, it follows from (32) that with probability 1,

$$\sum_{j=1}^{k} P_{\mathbf{w}}(C_j) \sim Tk\zeta_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} = o(1). \tag{34}$$

Hence, (a) follows from (31), (33) and (34). The proofs of (b) and (c) use similar techniques and shall be omitted here. □

3. Template matching with discontinuous kernels. In this section, we obtain analogues of Proposition 1 and Theorem 1 when the score function f contains discontinuities. A typical example is the box kernel

$$f(x) = \begin{cases} 1, & \text{if } x < \varepsilon, \\ -\beta, & \text{if } x \geq \varepsilon, \end{cases} \tag{35}$$

where β, ε are positive real numbers. Instead of (A2), we assume the following.

(A2)$'$ Let f be a discontinuous function and let there be a finite set H such that the first derivative of f exists and is uniformly continuous over any interval within $\mathbb{R}^+ \setminus H$. Moreover, (4) holds.

Under (A2)$'$, the values of f may be concentrated on $0, \pm q, \pm 2q, \ldots$ for some $q > 0$.

DEFINITION. Let $L(f) = \{f(x) : x \geq 0\}$. We say that f is *arithmetic* if

$$L(f) \subseteq q\mathbb{Z} \qquad \text{for some } q > 0. \tag{36}$$

Moreover, if q is the largest number satisfying (36), then we say that f is *arithmetic with span* q. If (36) is not satisfied for all $q > 0$, we say that f is *nonarithmetic*.

For example, if β in (35) is irrational, then f is nonarithmetic, while if $\beta = s/r$ for coprimes r and s, then f is arithmetic with span $q = r^{-1}$. We write, for $i = 1, \ldots, d$,

$$g_{\mathbf{w}}^{(i)}(u+) = \lim_{v \downarrow u} g_{\mathbf{w}}^{(i)}(v), \qquad g_{\mathbf{w}}^{(i)}(u-) = \lim_{v \uparrow u} g_{\mathbf{w}}^{(i)}(v),$$

$$\delta_i(u) = g_{\mathbf{w}}^{(i)}(u-) - g_{\mathbf{w}}^{(i)}(u+), \qquad D_i = \{u \in (0, T) : \delta_i(u) \neq 0\},$$

where $g_{\mathbf{w}}^{(i)}$ is defined in (3).

Let $\phi_{\mathbf{w}}$, $\theta_{\mathbf{w}}$, $v_{\mathbf{w}}$ and μ be defined as in Section 2.1. If $D_i \neq \varnothing$ for some i, we can define a probability mass function $h_{\mathbf{w}}^*$ taking values in $\{\delta_i(u)\}_{u \in D_i, 1 \leq i \leq d}$ such that

$$h_{\mathbf{w}}^*(x) = \sum_{i=1}^d \lambda_i \sum_{u \in D_i} e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u-)} \mathbf{1}_{\{\delta_i(u) = x\}} \Big/ \sum_{i=1}^d \lambda_i \sum_{u \in D_i} e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u-)}. \tag{37}$$

Let E_* denote expectation when $X_1, X_2, \ldots$ are independent identically distributed random variables with probability mass function $h_{\mathbf{w}}^*$. Define

$$\omega_b = \inf\{n : X_1 + \cdots + X_n \geq b\} \quad \text{and} \quad R_b = X_1 + \cdots + X_{\omega_b}. \tag{38}$$

The overshoot constant is then defined as

$$\nu_{\mathbf{w}} = \lim_{b \to \infty} E_* e^{-\theta_{\mathbf{w}}(R_b - b)}, \tag{39}$$

where b is taken to be a multiple of χ if $h_{\mathbf{w}}^*$ is arithmetic with span χ. Note that the statement "$h_{\mathbf{w}}^*$ is arithmetic with span χ" implies that $\{\delta_i(u)\}_{u \in D_i, 1 \leq i \leq d} \subset \chi\mathbb{Z}$. The constants $\nu_{\mathbf{w}}$ have been well studied in sequential analysis; see, for example, Siegmund [27] for the existence of the limits in (39). Let us define

$$\zeta'_{\mathbf{w}} = (2\pi T v_{\mathbf{w}})^{-1/2} \nu_{\mathbf{w}} K_{\mathbf{w}} \sum_{i=1}^d \lambda_i \sum_{u \in D_i} \delta_i(u) e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u-)}, \tag{40}$$

where

$$K_{\mathbf{w}} = \begin{cases} 1, & \text{if } h^*_{\mathbf{w}} \text{ is nonarithmetic,} \\ \dfrac{1}{\theta_{\mathbf{w}}\chi}(1 - e^{-\theta_{\mathbf{w}}\chi}), & \text{if } h^*_{\mathbf{w}} \text{ is arithmetic with span } \chi \\ & \text{and } f \text{ is nonarithmetic,} \\ \dfrac{q}{\chi}\left(\dfrac{1 - e^{-\theta_{\mathbf{w}}\chi}}{1 - e^{-\theta_{\mathbf{w}}q}}\right), & \text{if } h^*_{\mathbf{w}} \text{ arithmetic with span } \chi \\ & \text{and } f \text{ arithmetic with span } q. \end{cases}$$

Since we can express each $\delta_i(u)$, $u \in D_i$, in the form $g_1 - g_2$ for $g_1, g_2 \in L(f)$, it follows that if f is arithmetic, then $h^*_{\mathbf{w}}$ is arithmetic and χ/q is a positive integer. Analogously to Proposition 1 and Theorem 1, we can obtain the asymptotic boundary crossing probabilities of S_t, the asymptotic distribution of the scan statistic M_a and the time to detection V_c for kernels with discontinuities.

PROPOSITION 2. *Assume* (A1), (A2)′ *and let* $\Delta > 0$, $t \geq 0$. *If f is nonarithmetic and $c > \mu$, then*

$$P_{\mathbf{w}}\left\{\sup_{t<u\leq t+\Delta} S_u \geq c\right\} \sim \Delta\zeta'_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} \qquad \textit{a.s. as } T \to \infty. \tag{41}$$

If f is arithmetic with span q, then (41) *also holds under the convention that*

$$Tc(= Tc_T) \in q\mathbb{Z} \qquad \textit{with } c \to c' \textit{ as } T \to \infty \textit{ for some } c' > \mu. \tag{42}$$

THEOREM 2. *Assume* (A1) *and* (A2)′ *and let f be nonarithmetic.*

(a) *Let $c > \mu$. Then the distribution* (*conditional on* $\mathbf{w}$) *of* $\zeta'_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} V_c$ *converges to the exponential distribution with mean* 1 *almost surely as* $T \to \infty$.

(b) *Let $a \to \infty$ as $T \to \infty$ in such a way that* $(\log a)/T$ *converges to a positive constant. Let $c_{\mathbf{w}} > \mu_{\mathbf{w}}$ satisfy $\phi_{\mathbf{w}}(c_{\mathbf{w}}) = (\log a)/T$. Then for any $z \in \mathbb{R}$,*

$$P_{\mathbf{w}}\{\theta_{\mathbf{w}} T(M_a - c_{\mathbf{w}}) - \log \zeta'_{\mathbf{w}} \geq z\} \to 1 - \exp(-e^{-z}) \qquad \textit{a.s. as } T \to \infty.$$

(c) *Let $a \to \infty$ as $T \to \infty$ in such a way that* $(\log a)/T$ *converges to a positive constant. Let c $(= c_T)$ be such that $\eta_{\mathbf{w}} := a\zeta'_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)} \to \eta > 0$ almost surely. Then*

$$P_{\mathbf{w}}\{U_a = k\} - e^{-\eta_{\mathbf{w}}} \frac{\eta^k_{\mathbf{w}}}{k!} \to 0 \qquad \textit{a.s. } \forall k = 0, 1, \ldots. \tag{43}$$

If f is arithmetic with span q, then (a) *and* (c) *also hold under the convention* (42).

EXAMPLE 2. We conduct here a simulation study similar to Example 1. The generation of $\mathbf{w}$ and $\mathbf{y}$ are as in Example 1, but the box kernel (35) is used instead of the Hamming window function (13) when computing $g^{(i)}_{\mathbf{w}}$. We choose parame-

TABLE 2
Estimates of $P_{\mathbf{w}}\{M_a \geq c\} \pm$ *standard error with* $a + T = 20$ *s*

c	**Direct MC**	**Imp. sampling**	**Anal. approx. (44)**
0.065	0.029 ± 0.004	0.0300 ± 0.0016	0.0289
0.066	0.019 ± 0.003	0.0218 ± 0.0012	0.0207
0.067	0.012 ± 0.002	0.0140 ± 0.0008	0.0144
0.068	0.008 ± 0.002	0.0103 ± 0.0006	0.0101
0.069	0.005 ± 0.002	0.0067 ± 0.0004	0.0070
0.070	0.003 ± 0.001	0.0051 ± 0.0003	0.0047

ters $\varepsilon = 4$ ms and $\beta = 0.3$. Hence, f is arithmetic with span $q = 0.1$ and $h^*_{\mathbf{w}}$ arithmetic with span $\chi = 1.3$. In fact, $h^*_{\mathbf{w}}$ is positive only on the values -1.3 and 1.3 and hence $\nu_{\mathbf{w}} = 1$. In the template $\mathbf{w}$, there were a total of 2×59 elements in $\bigcup_i D_i$ with half of all $u \in D_i$ satisfying $\delta_i(u) = 1.3$ and the other half satisfying $\delta_i(u) = -1.3$. Hence, $h^*_{\mathbf{w}}(-1.3) = e^{-0.3\theta_{\mathbf{w}}}/(e^{\theta_{\mathbf{w}}} + e^{-0.3\theta_{\mathbf{w}}})$ and $h^*_{\mathbf{w}}(1.3) = e^{\theta_{\mathbf{w}}}/(e^{\theta_{\mathbf{w}}} + e^{-0.3\theta_{\mathbf{w}}})$. This information is used in the computation of $\zeta'_{\mathbf{w}}$ in the approximation

$$P_{\mathbf{w}}\{M_a \geq c\} = P_{\mathbf{w}}\{V_c \leq a\} \doteq 1 - \exp\bigl(-a\zeta'_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)}\bigr), \tag{44}$$

an analogue of (15) that follows from Theorem 2(a).

In Table 2, we compare the analytical approximation (44) with both direct Monte Carlo simulations and importance sampling, with 2000 simulation runs used to obtain each entry. The variance reduction when using importance sampling is similar to that seen in Example 1 and the technique is indeed an effective time-saving device for computing p-values, especially when they are small. The analytic approximations are also accurate and agree with the simulation results that were obtained.

In addition to the above simulation study, we also conducted a similar exercise to check the accuracy of the Poisson approximation of U_a in (43), this time with $a + T = 200$ s and threshold level $c = 0.0614$. The maximal proportion of overlap between two matches is chosen to be $\alpha = 0.8$. The analytical approximations are compared with 2000 direct Monte Carlo simulation runs and the results are recorded in Table 3 (with standard errors in parentheses). Again, we see that the analytical approximations are quite accurate and this indicates the usefulness of the asymptotic results in Theorem 2 for estimating p-values.

We shall now prove Proposition 2 and Theorem 2. The next result shows that

$$h_{\mathbf{w}}(x) := \sum_{i=1}^{d} \lambda_i \sum_{u \in D_i} e^{\theta_{\mathbf{w}} g^{(i)}_{\mathbf{w}}(u+)} \mathbf{1}_{\{\delta_i(u)=x\}} \Big/ \sum_{i=1}^{d} \lambda_i \sum_{u \in D_i} e^{\theta_{\mathbf{w}} g^{(i)}_{\mathbf{w}}(u+)} \tag{45}$$

and $h^*_{\mathbf{w}}$ [see (37)] are asymptotically conjugate probability mass functions.

TABLE 3
Estimates of $P_{\mathbf{w}}\{U_a = k\}$ and $\eta_{\mathbf{w}} = E_{\mathbf{w}}(U_a)$

k	**0**	**1**	**2**	**3**	**4**	**5**	**≥6**	$\eta_{\mathbf{w}}$
Poisson	0.336	0.366	0.200	0.073	0.020	0.004	0.001	1.09
Direct	0.328	0.363	0.195	0.084	0.024	0.005	0.001	1.13
MC	(0.011)	(0.011)	(0.009)	(0.006)	(0.003)	(0.002)	(0.001)	(0.02)

LEMMA 6. *There exists $\gamma_{\mathbf{w}} = 1 + O(T^{-1})$ a.s. such that $h^*_{\mathbf{w}}(x) = \gamma_{\mathbf{w}} e^{\theta_{\mathbf{w}} x} h_{\mathbf{w}}(x)$ for all x.*

PROOF. Let $u \in D_i$ with $w_j^{(i)} < u < w_{j+1}^{(i)}$ for adjacent spikes $w_j^{(i)}$, $w_{j+1}^{(i)} \in \mathbf{w}^{(i)}$. Then, by the symmetry of $g_{\mathbf{w}}^{(i)}$ in the interval $(w_j^{(i)}, w_{j+1}^{(i)})$ about its midpoint $(w_j^{(i)} + w_{j+1}^{(i)})/2$, it follows that $v := w_{j+1}^{(i)} - (y - w_j^{(i)})$ lies inside D_i and $g_{\mathbf{w}}^{(i)}(v-) = g_{\mathbf{w}}^{(i)}(u+)$. Hence, $\gamma_{\mathbf{w}}$, which we define here to be the ratio of the denominators on the right-hand sides of (37) and (45), is $1 + O(T^{-1})$ a.s. with the $O(T^{-1})$ coming from $u \in D_i$ occurring before the first spike or after the last spike in $\mathbf{w}^{(i)}$. Lemma 6 holds since $e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u-)} \mathbf{1}_{\{\delta_i(u) = x\}} = e^{\theta_{\mathbf{w}}[x + g_{\mathbf{w}}^{(i)}(u+)]} \mathbf{1}_{\{\delta_i(u) = x\}}$. $\square$

LEMMA 7. *Assume* (A1) *and* (A2)$'$. *For all $\varepsilon > 0$, there exists κ sufficiently large enough that for any $t \geq 0$, with probability* 1,

$$\left| \frac{P_{\mathbf{w}}\{S_t < c, \sup_{t < u \leq t + \kappa T^{-1}} S_u \geq c\}}{\kappa T^{-1} \zeta'_{\mathbf{w}} e^{-T\phi_{\mathbf{w}}(c)}} - 1 \right| \leq \varepsilon \qquad \textit{for all large } T,$$

where $\zeta'_{\mathbf{w}}$ is defined in (40) *and $c > \mu$ if f is nonarithmetic*, *while c satisfies* (42) *if f is arithmetic with span q.*

PROOF. Assume without loss of generality that $t = 0$ and let $G_i = \bigcup_{v \in D_i} (v, v + \kappa T^{-1}]$. We can write $T S_0 = T S'_0 + J_0$, where

$$S'_0 = T^{-1} \sum_{i=1}^{d} \sum_{y \in \mathbf{y}^{(i)}, y \notin G_i} g_{\mathbf{w}}^{(i)}(y) \quad \text{and} \quad J_0 = \sum_{i=1}^{d} \sum_{y \in \mathbf{y}^{(i)} \cap G_i} g_{\mathbf{w}}^{(i)}(y).$$

The random variables S'_0 and J_0 are independent because they are functions of the Poisson processes $\mathbf{y}^{(i)}$ over disjoint subsets of the real line. Let us first consider f arithmetic with span q. Then $T S'_0$ and J_0 are both integral multiples of q. Since f is constant between jumps, we can express $T S_u = T S'_0 + J_u$ [see (2)], where

$$J_u = \sum_{i=1}^{d} \sum_{y \in \mathbf{y}^{(i)} \cap G_i} g_{\mathbf{w}}^{(i)}(y - u) \qquad \forall u \in (0, \kappa T^{-1}). \tag{46}$$

Hence, both $S_0 < c$ and $\sup_{0<u\le\kappa T^{-1}} S_u \ge c$ occur if and only if

$$\sup_{0<u\le\kappa T^{-1}} (J_u - J_0) \ge \ell q \quad \text{and} \quad S_0' = c - T^{-1}kq \tag{47}$$
$$\text{for some } \ell \ge 1 \text{ and } k = J_0/q + \ell.$$

We shall now consider the probability measure $P_{\theta_{\mathbf{w}}}$ and its associated expectation $E_{\theta_{\mathbf{w}}}$ as defined at the beginning of Sections 2.2 and 2.3. By (18), $E_{\theta_{\mathbf{w}}}[S_0'] = c + O(T^{-1})$ a.s. and hence, by the local limit theorem for lattice random variables (see, e.g., Theorem 15.5.3 of Feller [14]),

$$P_{\theta_{\mathbf{w}}}\{S_0' = c - T^{-1}qk\} \sim q/(2\pi T v_{\mathbf{w}})^{1/2} \qquad \text{a.s.} \tag{48}$$

for any integer k. Since S_0' and $(J_u)_{0<u\le\kappa T^{-1}}$ are independent, it follows from (5), the change of measure (11), (47) and (48) that

$$\begin{aligned} &P_{\mathbf{w}}\Big\{S_0 < c, \sup_{0<u\le\kappa T^{-1}} S_u \ge c\Big\} \\ &\quad = E_{\theta_{\mathbf{w}}}\bigg[\frac{dP_{\mathbf{w}}}{dP_{\theta_{\mathbf{w}}}}(\mathbf{y})\mathbf{1}_{\{S_0<c,\sup_{0<u\le\kappa T^{-1}} S_u\ge c\}}\bigg] \\ &\quad \sim \frac{q}{(2\pi T v_{\mathbf{w}})^{1/2}} e^{-T\phi_{\mathbf{w}}(c)} \sum_{\ell=1}^{\infty} e^{\theta_{\mathbf{w}}\ell q} P_{\theta_{\mathbf{w}}}\Big\{\sup_{0<u\le\kappa T^{-1}} (J_u - J_0) \ge \ell q\Big\}. \end{aligned} \tag{49}$$

Since $g_{\mathbf{w}}^{(i)}$ is piecewise constant, the graph of $(J_u - J_0)$ against u is also piecewise constant, with jumps of $\delta_i(y-u)$ whenever $y - u \in D_i$ for some $y \in \mathbf{y}^{(i)}$, $1 \le i \le d$; see (46). Let N_* be the total number of spikes in $\bigcup_{1\le i\le d}(\mathbf{y}^{(i)} \cap G_i)$. Then there are N_* such jumps and

$$\sup_{0<u\le\kappa T^{-1}} (J_u - J_0) = \sup_{1\le j\le N_*} (X_1 + \cdots + X_j), \tag{50}$$

where X_j is the jth jump and has probability mass function $h_{\mathbf{w}}$. Moreover, $X_1, X_2, \ldots$ are independent, conditioned on N_*, which is Poisson with mean

$$EN_* = \kappa T^{-1} \sum_{i=1}^{d} \lambda_i \sum_{u\in D_i} e^{\theta_{\mathbf{w}} g_{\mathbf{w}}^{(i)}(u+)}. \tag{51}$$

If $r \in \{0, \ldots, \chi/q - 1\}$ and $s \in \mathbb{Z}^+$, then $R_{s\chi - rq} = R_{s\chi}$; see (38). Let E_* and P_* denote the expectation and probability measure, respectively, when $X_1, X_2, \ldots$ are independent identically distributed with probability mass function $h_{\mathbf{w}}^*$. It then follows from a change of measure to P_*, (50) and Lemma 6 that

$$
\begin{aligned}
&\sum_{\ell=1}^{\infty} e^{\theta_{\mathbf{w}}\ell q} P_{\theta_{\mathbf{w}}}\Big\{\sup_{0<u\le\kappa T^{-1}}(J_u - J_0)\ge \ell q\Big\}\\
&\qquad = \gamma_{\mathbf{w}}^{-1}\sum_{\ell=1}^{\infty} E_*\big[e^{-\theta_{\mathbf{w}}(R_{\ell q}-\ell q)}\mathbf{1}_{\{\sup_{1\le j\le N_*}(X_1+\cdots+X_j)\ge \ell q\}}\big]\\
&\qquad = \gamma_{\mathbf{w}}^{-1}E_*\Bigg[\sum_{r=0}^{\chi/q-1}\sum_{s=1}^{\infty} e^{-\theta_{\mathbf{w}}[R_{s\chi}-(s\chi-rq)]}\mathbf{1}_{\{\sup_{1\le j\le N_*}(X_1+\cdots+X_j)\ge s\chi\}}\Bigg]\\
&\qquad \sim \chi^{-1}\Bigg(\sum_{r=0}^{\chi/q-1} e^{-\theta_{\mathbf{w}} rq}\Bigg)\nu_{\mathbf{w}} E_*\Big[\sup_{1\le j\le N_*}(X_1+\cdots+X_j)\Big].
\end{aligned}
\tag{52}
$$

Since $E_*X_i > 0$ for all large T and the almost sure limit of EN_* [see (51)] is proportional to κ, it follows that there exists κ sufficiently large that

$$
\left|\frac{E_*[\sup_{1\le j\le N_*}(X_1+\cdots+X_j)]}{(EN_*)(E_*X_1)} - 1\right| < \frac{\varepsilon}{2}
\tag{53}
$$

for all large T. Since $(q/\chi)\sum_{r=0}^{\chi/q-1} e^{-\theta_{\mathbf{w}} rq} = K_{\mathbf{w}}$, Lemma 7 follows from (37), (49) and (51)–(53). When f is nonarithmetic, the local limit result (21) with $I_{2,T}=\mathbb{R}$ and $t=0$ is used in place of (48). □

The next lemma is proved in Chan and Loh [6].

LEMMA 8. *Assume* (A1) *and* (A2)′. *Let* $A_{t,v} = \{S_t < c, \sup_{t<u\le t+v} S_u \ge c\}$. *There exist* $r_\kappa = o(\kappa)$ *as* $\kappa\to\infty$, $\ell_0 > 0$ *and* $\gamma > 0$ *such that with probability* 1,

$$
\begin{aligned}
&P_{\mathbf{w}}\Big\{S_t < c-\kappa\gamma T^{-1}, \sup_{t<u\le t+\kappa T^{-1}} S_u \ge c\Big\}\\
&\quad + P_{\mathbf{w}}\big(A_{t,\kappa T^{-1}}\cap A_{t+\kappa T^{-1},(\ell_0-1)\kappa T^{-1}}\big)\\
&\quad + \sum_{\ell=\ell_0}^{\lfloor T^2/\kappa\rfloor} P_{\mathbf{w}}\{S_t \ge c-\kappa\gamma T^{-1}, S_{t+\ell\kappa T^{-1}} \ge c-\kappa\gamma T^{-1}\}\\
&\quad \le r_\kappa T^{-1/2} e^{-T\phi_{\mathbf{w}}(c)},
\end{aligned}
\tag{54}
$$

for all $t\ge 0$ *and large* T.

PROOF OF PROPOSITION 2 AND THEOREM 2. By stationarity, we may assume without loss of generality that $t=0$. Select $\ell_0 > 0$ and $\gamma > 0$ such that (54) is satisfied. Let $D_{y,v} = A_{y,v} \cap \{S_y \geq c - \kappa\gamma T^{-1}\}$. Then (41) follows from Lemmas 1, 7, 8 and the inequalities

$$\sum_{q=0}^{\lfloor \Delta/(\kappa T^{-1})\rfloor - 1} \Bigg[P_{\mathbf{w}}(D_{q\kappa T^{-1},\kappa T^{-1}}) - P_{\mathbf{w}}\big(A_{q\kappa T^{-1},\kappa T^{-1}} \cap A_{(q+1)\kappa T^{-1},(\ell_0-1)\kappa T^{-1}}\big)$$

$$- \sum_{\ell=\ell_0}^{\lfloor \Delta/(\kappa T^{-1})\rfloor - 1 - q} P_{\mathbf{w}}\{S_{q\kappa T^{-1}} \geq c - \kappa\gamma T^{-1}, S_{(q+\ell)\kappa T^{-1}} \geq c - \kappa\gamma T^{-1}\}\Bigg]$$

$$\leq P_{\mathbf{w}}\Big\{\sup_{0<u\leq\Delta} S_u \geq c\Big\} \leq P_{\mathbf{w}}\{S_0 \geq c\} + \sum_{q=0}^{\lfloor \Delta/(\kappa T^{-1})\rfloor} P_{\mathbf{w}}(A_{q\kappa T^{-1},\kappa T^{-1}}),$$

with κ arbitrarily large; see, for example, the proof of Proposition 1. The proof of Theorem 2 proceeds as in the proof of Theorem 1, replacing $\zeta_{\mathbf{w}}$ by $\zeta'_{\mathbf{w}}$. $\square$

4. Sieve maximum likelihood estimation. In the second part of this article, we assume that the spike train N is modeled by a counting process with conditional intensity λ_1, as given by (1). Suppose that a realization of N is observed on the interval $[0, T)$, $0 < T < \infty$, and that the spike times are $0 < w_1 < \cdots < w_{N(T)} < T$. Let $\mathbf{w}_T = \{w_1, \ldots, w_{N(T)}\}$ denote the point process corresponding to $N(t)$, $t \in [0, T)$, and $\mathcal{N}$ be the set of all possible realizations of $\mathbf{w}_T$. It follows from Daley and Vere-Jones [9] that the likelihood is the local Janossy density, given by

$$p_{s,r}(\mathbf{w}_T) = e^{-\int_0^T s(t) r(t - w_{N(t)})\,dt} \prod_{j=1}^{N(T)} s(w_j) r(w_j - w_{j-1}), \tag{55}$$

where $r(t - w_0) = 1$ for all $t \in [0, T)$. Next, let q, q_0, q_1 be constants such that $q = q_0 + q_1$, q_0 is a nonnegative integer and $0 < q_1 \leq 1$, and let $\boldsymbol{\kappa} = (\kappa_0, \ldots, \kappa_{q_0+1})$ be a vector of strictly positive constants. Here, we assume that the true free firing rate function s and the recovery function r lie in the q-smooth function class $\Theta_{\boldsymbol{\kappa},q}$ where

$$\Theta_{\boldsymbol{\kappa},q} = \Bigg\{ f = g^2 : g \in \mathcal{C}^{q_0}[0, T), \min_{t\in[0,T)} g(t) \geq 0, \max_{t\in[0,T)} \left| \frac{d^j}{dt^j} g(t) \right| < \kappa_j,$$

$$j = 0, \ldots, q_0, \left| \frac{d^{q_0}}{dt^{q_0}} g(t_1) - \frac{d^{q_0}}{dt^{q_0}} g(t_2) \right| < \kappa_{q_0+1} |t_1 - t_2|^{q_1}, \tag{56}$$

$$\forall t_1, t_2 \in [0, T) \Bigg\}.$$

Let $\{0 < \delta_n \leq 1 : n = 1, 2, \ldots\}$ be a sequence of constants (to be suitably chosen later and where δ_n depends only on n) such that $\delta_n \to 0$ as $n \to \infty$. Define

$$\Theta_{\boldsymbol{\kappa},q,n} = \Theta_{\boldsymbol{\kappa},q} \cap \Big\{ f = g^2 : g \in \mathcal{C}^{q_0}[0, T), \min_{t\in[0,T)} g(t) \geq \delta_n \Big\}.$$

Let $\Theta_{\boldsymbol{\kappa},q}$ and $\Theta_{\boldsymbol{\kappa},q,n}$ be endowed with the metrics $\rho_{\Theta_{\boldsymbol{\kappa},q}}$ and $\rho_{\Theta_{\boldsymbol{\kappa},q,n}}$, respectively, where $\rho_{\Theta}(f_1, f_2) = \sup_{t\in[0,T)} |f_1^{1/2}(t) - f_2^{1/2}(t)|, \forall f_1, f_2 \in \Theta$, $\Theta \in \{\Theta_{\boldsymbol{\kappa},q}, \Theta_{\boldsymbol{\kappa},q,n}\}$. We observe that any $f \in \Theta_{\boldsymbol{\kappa},q}$ can be approximated arbitrarily closely by $(f^{1/2} + \delta_n)^2 \in \Theta_{\boldsymbol{\kappa},q,n}$ by choosing n sufficiently large. Consequently, a sieve for the parameter space of (s, r) can now be expressed as $\Theta^2_{\boldsymbol{\kappa},q,n}$ with metric $\rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}$, where

$$\begin{aligned}
&\rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}((f_1, g_1), (f_2, g_2)) \\
&\qquad = \rho_{\Theta_{\boldsymbol{\kappa},q,n}}(f_1, f_2) + \rho_{\Theta_{\boldsymbol{\kappa},q,n}}(g_1, g_2) \qquad \forall (f_1, g_1), (f_2, g_2) \in \Theta^2_{\boldsymbol{\kappa},q,n}.
\end{aligned}$$

Next, let $\mathcal{F}_{\boldsymbol{\kappa},q,n} = \{p_{s_1,r_1}$ is as in (55): $(s_1, r_1) \in \Theta_{\boldsymbol{\kappa},q,n^2}\}$ be endowed with the Hellinger metric $\rho_{\mathcal{F}_{\boldsymbol{\kappa},q,n}}$. More precisely, writing $\mathbf{w}_j = \{w_1, \ldots, w_j\}$,

$$\begin{aligned}
\rho_{\mathcal{F}_{\boldsymbol{\kappa},q,n}}(p_{s_1,r_1}, p_{s_2,r_2}) &= \| p^{1/2}_{s_1,r_1} - p^{1/2}_{s_2,r_2} \|_2 \\
&= \left\{ \sum_{j=0}^{\infty} \int_{0<w_1<\cdots<w_j<T} [p^{1/2}_{s_1,r_1}(\mathbf{w}_j) - p^{1/2}_{s_2,r_2}(\mathbf{w}_j)]^2 \, d\mathbf{w}_j \right\}^{1/2}.
\end{aligned}$$

For $\varepsilon > 0$, let $\Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon) \subseteq \Theta^2_{\boldsymbol{\kappa},q,n}$ denote a finite ε-net for $\Theta^2_{\boldsymbol{\kappa},q,n}$ with respect to the metric $\rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}$. This implies that $\mathrm{card}(\Theta^2{}_{\boldsymbol{\kappa},q,n}(\varepsilon)) < \infty$ and that for each $(s_1, r_1) \in \Theta^2_{\boldsymbol{\kappa},q,n}$, there exists an $(s_2, r_2) \in \Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon)$ such that $\rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}((s_1, r_1), (s_2, r_2)) \le \varepsilon$.

Now, suppose that for each $\varepsilon > 0$, there exist measurable nonnegative functions $f_{l,\varepsilon}$ and $f_{u,\varepsilon}$ on $\Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon) \times \mathcal{N}$ such that for each $(s_1, r_1) \in \Theta^2_{\boldsymbol{\kappa},q,n}$, there exists some $(s_2, r_2) \in \Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon)$ satisfying:

(C1) $\rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}((s_1, r_1), (s_2, r_2)) \le \varepsilon$;
(C2) $f_{l,\varepsilon}((s_2, r_2), \mathbf{w}_T) \le p_{s_1,r_1}(\mathbf{w}_T) \le f_{u,\varepsilon}((s_2, r_2), \mathbf{w}_T)$, a.s.;
(C3) $\{\sum_{j=0}^{\infty} \int_{0<w_1<\cdots<w_j<T} [f^{1/2}_{u,\varepsilon}((s_2, r_2), \mathbf{w}_j) - f^{1/2}_{l,\varepsilon}((s_2, r_2), \mathbf{w}_j)]^2 \, d\mathbf{w}_j\}^{1/2} \le \varepsilon$.

For $\varepsilon > 0$, the ε-entropy of $\Theta^2_{\boldsymbol{\kappa},q,n}$ with respect to $\rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}$ is defined to be

$$\begin{aligned}
&H(\varepsilon, \Theta^2_{\boldsymbol{\kappa},q,n}, \rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}) \\
&\qquad = \log[\min\{\mathrm{card}(\Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon)) : \Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon) \text{ is a } \varepsilon\text{-net for } \Theta^2_{\boldsymbol{\kappa},q,n} \\
&\qquad\qquad\qquad \text{with respect to the metric } \rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}\}].
\end{aligned}$$

The ε-entropies of $\Theta_{\boldsymbol{\kappa},q}$ and $\Theta_{\boldsymbol{\kappa},q,n}$ are defined in a similar manner. The ε-entropy of $\mathcal{F}_{\boldsymbol{\kappa},q,n}$ with bracketing with respect to the metric $\rho_{\mathcal{F}_{\boldsymbol{\kappa},q,n}}$ is defined to be

$$\begin{aligned}
&H^B(\varepsilon, \mathcal{F}_{\boldsymbol{\kappa},q,n}, \rho_{\mathcal{F}_{\boldsymbol{\kappa},q,n}}) \\
&\qquad = \log[\min\{\mathrm{card}(\Theta^2_{\boldsymbol{\kappa},q,n}(\varepsilon)) : \text{(C1), (C2) and (C3) are satisfied}\}].
\end{aligned}$$

We observe from Kolmogorov and Tihomirov [19], page 308, and Dudley [13], page 11, that the ε-entropy of $\Theta^2_{\boldsymbol{\kappa},q,n}$ satisfies

$$H(\varepsilon, \Theta^2_{\boldsymbol{\kappa},q,n}, \rho_{\Theta^2_{\boldsymbol{\kappa},q,n}}) \le 2H(\varepsilon/2, \Theta_{\boldsymbol{\kappa},q}, \rho_{\Theta_{\boldsymbol{\kappa},q}}) \le 2^{(q+1)/q}\varepsilon^{-1/q}C_{\boldsymbol{\kappa},q}, \tag{57}$$

where $C_{\boldsymbol{\kappa},q}$ is a constant depending only on $\boldsymbol{\kappa}$ and q. Next, let $f : \mathcal{N} \to \mathbb{R}$ be a nonnegative function such that $\sum_{j=0}^{\infty}\int_{0\le w_1<\cdots<w_j<T} f(\mathbf{w}_j)\,d\mathbf{w}_j < \infty$. We follow Wong and Shen [30] in defining $Z_f(\mathbf{w}_T) = \log[f(\mathbf{w}_T)/p_{s,r}(\mathbf{w}_T)]$, where s is the true free firing rate function and r the true recovery function. For $\tau > 0$, we write

$$\tilde{Z}_f(\mathbf{w}_T) = \begin{cases} Z_f(\mathbf{w}_T), & \text{if } Z_f(\mathbf{w}_T) \ge -\tau, \\ -\tau, & \text{if } Z_f(\mathbf{w}_T) < -\tau. \end{cases}$$

Let $\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n} = \{\tilde{Z}_{p_{s_1,r_1}} : p_{s_1,r_1} \in \mathcal{F}_{\boldsymbol{\kappa},q,n}\}$ be the space of truncated log-likelihood ratios [based on one (spike train) observation]. Define $H^B(\varepsilon, \tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}, \rho_{\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}})$ to be the ε-entropy of $\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}$ with bracketing with respect to the metric

$$\rho_{\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}}(\tilde{Z}_{p_{s_1,r_1}}, \tilde{Z}_{p_{s_2,r_2}}) = \{E_{s,r}\{[\tilde{Z}_{p_{s_1,r_1}}(\mathbf{w}_T) - \tilde{Z}_{p_{s_2,r_2}}(\mathbf{w}_T)]^2\}\}^{1/2}. \tag{58}$$

In this section, $E_{s,r}$ and $P_{s,r}$ denote expectation and probability, respectively, when the true free firing rate function is s and the recovery function is r. We observe from Lemma 9 below that

$$H^B(\varepsilon, \tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}, \rho_{\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}}) \le C^*_{\boldsymbol{\kappa},q}\left(\frac{2e^{\tau/2}}{\varepsilon}\right)^{1/q}. \tag{59}$$

4.1. *Main results.* Suppose we have n i.i.d. copies of $N(t)$, $t \in [0, T)$, with conditional intensity λ_1, as given by (1). Let these n copies be denoted by $N_i(t)$, $t \in [0, T)$, the spike times be written as $0 < w_1^{(i)} < \cdots < w_{N_i(T)}^{(i)} < T$ and $\mathbf{w}_T^{(i)} = \{w_1^{(i)}, \dots, w_{N_i(T)}^{(i)}\}$, $i = 1, \dots, n$.

PROPOSITION 3. *Let* $0 < \varepsilon < 1$ *and* $C^*_{\boldsymbol{\kappa},q}$ *be as in* (59). *Suppose that*

$$\int_{\varepsilon^2/2^8}^{\sqrt{2}\varepsilon}\left[C^*_{\boldsymbol{\kappa},q}\left(\frac{10}{x}\right)^{1/q}\right]^{1/2} dx \le \frac{n^{1/2}\varepsilon^2}{2^{13}\sqrt{2}}. \tag{60}$$

Then, letting $P^*_{s,r}$ *denote the outer measure corresponding to the density* $p_{s,r}$, *we have*

$$P^*_{s,r}\left\{\sup_{\|p^{1/2}_{s_1,r_1} - p^{1/2}_{s,r}\|_2 \ge \varepsilon,\, p_{s_1,r_1}\in\mathcal{F}_{\boldsymbol{\kappa},q,n}} \prod_{i=1}^{n}\frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s,r}(\mathbf{w}_T^{(i)})} \ge e^{-n\varepsilon^2/8}\right\} \le 4\exp\left[-\frac{n\varepsilon^2}{2^7(250)}\right].$$

The above proposition is motivated by (and the proof is similar to) Theorem 1 of Wong and Shen [30] (see also Theorem 3 of Shen and Wong [25]). As such, we shall refer the reader to Chan and Loh [6] for the proof. Next, we define nonnegative functions $s_n^\dagger$ and $r_n^\dagger$ on $t \in [0, T)$ by

$$\sqrt{s_n^\dagger(t)} = \sqrt{s(t)} + \delta_n, \qquad \sqrt{r_n^\dagger(t)} = \sqrt{r(t)} + \delta_n. \tag{61}$$

Since $s, r \in \Theta_{\boldsymbol{\kappa},q}$, we have $s_n^\dagger, r_n^\dagger \in \Theta_{\boldsymbol{\kappa},q,n}$ and $p_{s_n^\dagger, r_n^\dagger} \in \mathcal{F}_{\boldsymbol{\kappa},q,n}$ for sufficiently large n. We further observe from Lemma 8 of Wong and Shen [30] that

$$0 \le \delta_n^\dagger := E_{s,r}(p_{s,r} p_{s_n^\dagger, r_n^\dagger}^{-1} - 1) \le C_{\boldsymbol{\kappa}}^* \delta_n,$$

where $C_{\boldsymbol{\kappa}}^*$ is a constant depending only on $\boldsymbol{\kappa}$.

DEFINITION. Let η_n be a sequence of positive numbers converging to 0. We call an estimator $p_{\hat{s}_n, \hat{r}_n} : \mathcal{N}^n \to \mathbb{R}^+$ an *η_n-sieve MLE* of $p_{s,r}$ if $(\hat{s}_n, \hat{r}_n) \in \Theta_{\boldsymbol{\kappa},q,n}^2$ and

$$n^{-1} \sum_{i=1}^{n} \log[p_{\hat{s}_n,\hat{r}_n}(\mathbf{w}_T^{(i)})] \ge \sup_{p_{s_1,r_1} \in \mathcal{F}_{\boldsymbol{\kappa},q,n}} n^{-1} \sum_{i=1}^{n} \log[p_{s_1,r_1}(\mathbf{w}_T^{(i)})] - \eta_n.$$

$\hat{s}_n$ and $\hat{r}_n$ are called η_n-sieve MLE's of s and r, respectively. If $\delta_n = 0$ for $n = 1, 2, \dots$, then an η_n-sieve MLE is more simply called an *η_n-MLE*.

THEOREM 3. *Let $\varepsilon_n > 0$ be the smallest value of ε satisfying* (60), $q > 1/2$ *and* $0 < \eta_n < \varepsilon_n^2/16$. *If $p_{\hat{s}_n,\hat{r}_n}$ is an η_n-MLE of $p_{s,r}$, then*

$$E_{s,r} \| p_{\hat{s}_n,\hat{r}_n}^{1/2} - p_{s,r}^{1/2} \|_2 = O(n^{-q/(2q+1)}) \qquad \text{as } n \to \infty.$$

We now assume that there exists a refractory period in which the neuron cannot discharge another spike after a spike has been fired (cf. Brillinger [3] and Johnson and Swami [17]).

THEOREM 4. *Let $\varepsilon_n > 0$ be the smallest value of ε satisfying* (60), $q > 1/2$, $0 < \eta_n < \varepsilon_n^2/16$ *and* $\delta_n = n^{-\alpha}$ *for some constant* $\alpha \in (2q/(2q+1), 1)$. *Suppose that there exists a constant $\theta > 0$ such that $r(u) = 0, \forall u \in [0, \theta]$ and $\hat{s}_n, \hat{r}_n$ are η_n-sieve MLE's of s, r, respectively. Then*

$$E_{s,r}\left[\int_0^T |\hat{s}_n(t) - s(t)|\, dt\right] = O(n^{-q/(2q+1)} \log^{1/2} n) \qquad \text{as } n \to \infty.$$

If, in addition, $s(t) > 0$ for all $t \in [0, T]$, then

$$E_{s,r}\left[\int_0^{T^*} |\hat{r}_n(u) - r(u)|\, du\right] = O(n^{-q/(2q+1)} \log^{1/2} n) \qquad \text{as } n \to \infty,$$

where T^* is any constant satisfying $0 < T^* < T$.

We end this subsection by computing lower bounds on the convergence rate of estimators for s and r based on $N_1(t), \dots, N_n(t)$, $t \in [0, T)$. For $0 < \theta < T$, define

$$\Theta_{\theta,\boldsymbol{\kappa},q} = \Theta_{\boldsymbol{\kappa},q} \cap \{f = g^2 : g \in \mathcal{C}^{q_0}[0, T), g(t) = 0\ \forall t \in [0, \theta]\}.$$

Theorem 5 below is motivated by the lower bound results of Yatracos [31].

THEOREM 5. *Let $q > 0$ and θ, T^* be constants satisfying $0 < \theta < T^* < T$. Suppose that $\tilde{s}_n$ and $\tilde{r}_n$ are estimators for s and r, respectively, based on $N_1(t), \dots, N_n(t)$, $t \in [0, T)$. Then there exist strictly positive constants $C_{\boldsymbol{\kappa},q}$, $C_{\theta,\boldsymbol{\kappa},q}$ such that*

$$\sup\left\{E_{s,r}\left[\int_0^T |\tilde{s}_n(t) - s(t)|\,dt\right] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\right\} \geq C_{\boldsymbol{\kappa},q} n^{-q/(2q+1)},$$

$$\sup\left\{E_{s,r}\left[\int_0^{T^*} |\tilde{r}_n(u) - r(u)|\,du\right] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\right\} \geq C_{\theta,\boldsymbol{\kappa},q} n^{-q/(2q+1)}.$$

4.2. *Proofs.* In this subsection, we shall sketch the proofs of Theorems 3, 4 and 5. We refer the reader to Chan and Loh [6] for the details.

LEMMA 9. *Let $\varepsilon > 0$ and $\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}$ be as in (58). Then*

$$H^B(\varepsilon, \tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}, \rho_{\tilde{\mathcal{Z}}_{\boldsymbol{\kappa},q,n}}) \leq H^B\left(\frac{\varepsilon}{2e^{\tau/2}}, \mathcal{F}_{\boldsymbol{\kappa},q,n}, \rho_{\mathcal{F}_{\boldsymbol{\kappa},q,n}}\right) \leq C^*_{\boldsymbol{\kappa},q}\left(\frac{2e^{\tau/2}}{\varepsilon}\right)^{1/q},$$

where $C^*_{\boldsymbol{\kappa},q}$ is a constant depending only on $\boldsymbol{\kappa}$ and q.

PROOF. The first inequality follows from Lemma 3 of Wong and Shen [30]. The second inequality follows from (57) and Lemma 2.1 of Ossiander [24]. □

PROOF OF THEOREM 3. We observe that $\delta_n^\dagger = 0$ for $n = 1, 2, \dots$ (from the definition of an η_n-MLE) and that ε_n is exactly of order $n^{-q/(2q+1)}$ as $n \to \infty$. We further observe from Proposition 3 and Markov's inequality that

$$\begin{aligned}
&P_{s,r}(\|p^{1/2}_{\hat{s}_n,\hat{r}_n} - p^{1/2}_{s,r}\|_2 \geq \varepsilon_n) \\
&\quad\leq P^*_{s,r}\left\{\sup_{\|p^{1/2}_{s_1,r_1} - p^{1/2}_{s,r}\|_2 \geq \varepsilon_n,\, p_{s_1,r_1} \in \mathcal{F}_{\boldsymbol{\kappa},q,n}} \prod_{i=1}^n \frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s_n^\dagger,r_n^\dagger}(\mathbf{w}_T^{(i)})} \geq e^{-n\eta_n}\right\} \\
&\quad\leq P^*_{s,r}\left\{\sup_{\|p^{1/2}_{s_1,r_1} - p^{1/2}_{s,r}\|_2 \geq \varepsilon_n,\, p_{s_1,r_1} \in \mathcal{F}_{\boldsymbol{\kappa},q,n}} \prod_{i=1}^n \frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s_n^\dagger,r_n^\dagger}(\mathbf{w}_T^{(i)})} \geq e^{-n\varepsilon_n^2/16}\right\} \\
&\quad\leq P^*_{s,r}\left\{\sup_{\|p^{1/2}_{s_1,r_1} - p^{1/2}_{s,r}\|_2 \geq \varepsilon_n,\, p_{s_1,r_1} \in \mathcal{F}_{\boldsymbol{\kappa},q,n}} \prod_{i=1}^n \frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s,r}(\mathbf{w}_T^{(i)})} \geq e^{-n\varepsilon_n^2/8}\right\} \\
&\qquad + P_{s,r}\left\{\prod_{i=1}^n \frac{p_{s,r}(\mathbf{w}_T^{(i)})}{p_{s_n^\dagger,r_n^\dagger}(\mathbf{w}_T^{(i)})} \geq e^{n\varepsilon_n^2/16}\right\} \leq 4\exp\left[-\frac{n\varepsilon_n^2}{2^7(250)}\right] + \exp\left(-\frac{n\varepsilon_n^2}{16}\right).
\end{aligned}$$

We conclude that as $n \to \infty$,

$$E_{s,r} \| p_{\hat{s}_n,\hat{r}_n}^{1/2} - p_{s,r}^{1/2} \|_2 \le \varepsilon_n + 8 \exp\left[-\frac{n\varepsilon_n^2}{2^7(250)}\right] + 2\exp\left(-\frac{n\varepsilon_n^2}{16}\right)$$
$$= O(n^{-q/(2q+1)}). \qquad \square$$

We preface the proof of Theorem 4 by Lemmas 10 and 11, whose proofs can be found in Chan and Loh [6].

LEMMA 10. *Let $\varepsilon_n > 0$ be the smallest value of ε satisfying (60) and $0 < \eta_n < \varepsilon_n^2/16 \le (1 - e^{-1})^2/32$. If $r(u) = 0, \forall u \in [0, \theta]$, and $p_{\hat{s}_n,\hat{r}_n}$ is an η_n-sieve MLE of $p_{s,r}$, then*

$$P_{s,r}\left\{\sum_{j=0}^{n_\theta} \int_{0<w_1<\cdots<w_j<T} p_{s,r}(\mathbf{w}_j) \log\left[\frac{p_{s,r}(\mathbf{w}_j)}{p_{\hat{s}_n,\hat{r}_n}(\mathbf{w}_j)}\right] d\mathbf{w}_j \right.$$
$$\left. > \left[6 + \frac{2\log(2)}{(1-e^{-1})^2} + 8\max\left\{1, \log\left(\frac{e^{\bar{\kappa}^4(\bar{\kappa}^4+1)T/2}}{\varepsilon_n \delta_n^{2n_\theta}}\right)\right\}\right]\varepsilon_n^2\right\}$$
$$\le 4\exp\left[-\frac{n\varepsilon_n^2}{2^7(250)}\right] + \exp\left[-n\left(\frac{\varepsilon_n^2}{16} - \delta_n^\dagger\right)\right],$$
where $\bar{\kappa} = \kappa_0 \vee 1, n_\theta = \lceil T/\theta \rceil$.

LEMMA 11. *Let $N(t)$, $t \in [0, T)$, be a counting process with conditional intensity λ_1, as in (1). Suppose $r(u) = 0, \forall u \in [0, \theta]$, and*

$$\xi(t) := \lim_{\delta \downarrow 0} \frac{1}{\delta} P_{s,r}[N(t+\delta) - N(t) = 1] \qquad \forall t \in [0, T). \tag{62}$$

Then for $s_1, r_1 \in \Theta_{\boldsymbol{\kappa},q,n}$, we have

$$\sum_{j=0}^{n_\theta} \int_{0<w_1<\cdots<w_j<T} p_{s,r}(\mathbf{w}_j) \log\left[\frac{p_{s,r}(\mathbf{w}_j)}{p_{s_1,r_1}(\mathbf{w}_j)}\right] d\mathbf{w}_j$$
$$= \int_0^T \left\{\frac{s_1(t)}{s(t)} - 1 - \log\left[\frac{s_1(t)}{s(t)}\right]\right\} s(t) e^{-\int_0^t s(u)\,du}\, dt$$
$$+ \int_0^T \int_0^t \left\{\frac{s_1(t)r_1(u)}{s(t)r(u)} - 1 - \log\left[\frac{s_1(t)r_1(u)}{s(t)r(u)}\right]\right\}$$
$$\times \xi(t-u) s(t) r(u) e^{-\int_{t-u}^t s(v) r(v-t+u)\,dv}\, du\, dt.$$

Also, if $\sum_{j=0}^{n_\theta}\int_{0<w_1<\cdots<w_j<T} p_{s,r}(\mathbf{w}_j)\log[p_{s,r}(\mathbf{w}_j)/p_{s_1,r_1}(\mathbf{w}_j)]\,d\mathbf{w}_j \le 1$, *then*

$$\sum_{j=0}^{n_\theta}\int_{0<w_1<\cdots<w_j<T} p_{s,r}(\mathbf{w}_j)\log\biggl[\frac{p_{s,r}(\mathbf{w}_j)}{p_{s_1,r_1}(\mathbf{w}_j)}\biggr]d\mathbf{w}_j$$
$$\ge \min\biggl\{\frac{1}{20\int_0^T s(t)e^{-\int_0^t s(u)\,du}\,dt},\frac{1}{200}\biggr\}\biggl[\int_0^T |s_1(t)-s(t)|e^{-\int_0^t s(u)\,du}\,dt\biggr]^2$$

and

$$\sum_{j=0}^{n_\theta}\int_{0<w_1<\cdots<w_j<T} p_{s,r}(\mathbf{w}_j)\log\biggl[\frac{p_{s,r}(\mathbf{w}_j)}{p_{s_1,r_1}(\mathbf{w}_j)}\biggr]d\mathbf{w}_j$$
$$\ge \min\biggl\{\frac{1}{20\int_0^T\int_0^t \xi(t-u)s(t)r(u)e^{-\int_{t-u}^t s(v)r(v-t+u)\,dv}\,du\,dt},\frac{1}{200}\biggr\}$$
$$\times\biggl[\int_0^T\int_0^t |s_1(t)r_1(u)-s(t)r(u)|\xi(t-u)e^{-\int_{t-u}^t s(v)r(v-t+u)\,dv}\,du\,dt\biggr]^2.$$

PROOF OF THEOREM 4. We observe from Lemmas 10 and 11 that

$$P_{s,r}\biggl\{\int_0^T |\hat{s}_n(t)-s(t)|e^{-\int_0^t s(u)\,du}\,dt$$
$$\le \biggl\{\max\biggl\{20\int_0^T s(t)e^{-\int_0^t s(u)\,du}\,dt,200\biggr\}$$
$$\times\biggl[6+\frac{2\log(2)}{(1-e^{-1})^2}+8\max\biggl\{1,\log\biggl(\frac{e^{\bar\kappa^4(\bar\kappa^4+1)T/2}}{\varepsilon_n\delta_n^{2n_\theta}}\biggr)\biggr\}\biggr]\varepsilon_n^2\biggr\}^{1/2}\biggr\}$$
$$\ge 1-4\exp\biggl[-\frac{n\varepsilon_n^2}{2^7(250)}\biggr]-\exp\biggl[-n\biggl(\frac{\varepsilon_n^2}{16}-\delta_n^\dagger\biggr)\biggr].$$

This implies that

$$E_{s,r}\biggl[\int_0^T |\hat{s}_n(t)-s(t)|e^{-\int_0^t s(u)\,du}\,dt\biggr]$$
$$\le \biggl\{\max\biggl\{20\int_0^T s(t)e^{-\int_0^t s(u)\,du}\,dt,200\biggr\}$$
$$\times\biggl[6+\frac{2\log(2)}{(1-e^{-1})^2}+8\max\biggl\{1,\log\biggl(\frac{e^{\bar\kappa^4(\bar\kappa^4+1)T/2}}{\varepsilon_n\delta_n^{2n_\theta}}\biggr)\biggr\}\biggr]\varepsilon_n^2\biggr\}^{1/2}$$
$$+\kappa_0^2T\biggl\{4\exp\biggl[-\frac{n\varepsilon_n^2}{2^7(250)}\biggr]+\exp\biggl[-n\biggl(\frac{\varepsilon_n^2}{16}-\delta_n^\dagger\biggr)\biggr]\biggr\}$$

and consequently, $E_{s,r}[\int_0^T |\hat{s}_n(t)-s(t)|\,dt]=O(n^{-q/(2q+1)}\log^{1/2}n)$ as $n\to\infty$, since ε_n is exactly of the order $n^{-q/(2q+1)}$. Next, we assume, in addition, that

$s(t) > 0$ for all $t \in [0, T]$. Let $\xi(t), t \in [0, T)$, be as in (62). Since $s \in \Theta_{\boldsymbol{\kappa},q}$ and

$$s(t)e^{-\int_0^t s(u)\,du} \leq \xi(t) \leq \max\{s(t), s(t)r(u) : u \in [0, T)\},$$

we have $0 < \min_{0 \leq t < T} \xi(t) \leq \max_{0 \leq t < T} \xi(t) \leq \bar{\kappa}^4$. Thus, as in the previous case,

$$\begin{aligned}
&E_{s,r}\biggl[\int_0^T \int_0^t |\hat{s}_n(t)\hat{r}_n(u) - s(t)r(u)|\,du\,dt\biggr] \\
&\quad \leq \biggl\{\frac{\max\{20\bar{\kappa}^8 T^2 e^{2\bar{\kappa}^4 T}, 200 e^{2\bar{\kappa}^4 T}\}}{\min_{0 \leq t < T} \xi^2(t)} \\
&\qquad \times \biggl[6 + \frac{2\log(2)}{(1 - e^{-1})^2} + 8\max\biggl\{1, \log\biggl(\frac{e^{\bar{\kappa}^4(\bar{\kappa}^4+1)T/2}}{\varepsilon_n \delta_n^{2n_\theta}}\biggr)\biggr\}\biggr]\varepsilon_n^2\biggr\}^{1/2} \\
&\qquad + \frac{\bar{\kappa}^8 T^2 e^{\bar{\kappa}^4 T}}{\min_{0 \leq t < T} \xi(t)}\biggl\{4\exp\biggl[-\frac{n\varepsilon_n^2}{2^7(250)}\biggr] + \exp\biggl[-n\biggl(\frac{\varepsilon_n^2}{16} - \delta_n^{\dagger}\biggr)\biggr]\biggr\},
\end{aligned}$$

which is of order $n^{-q/(2q+1)} \log^{1/2} n$ as $n \to \infty$. Since

$$\begin{aligned}
&\Bigl[\min_{0 \leq t < T} s(t)\Bigr] E_{s,r}\biggl[\int_0^T |\hat{r}_n(u) - r(u)|(T - u)\,du\biggr] \\
&\quad \leq \bar{\kappa}^2 T\, E_{s,r}\biggl[\int_0^T |\hat{s}_n(t) - s(t)|\,dt\biggr] \\
&\qquad + E_{s,r}\biggl[\int_0^T \int_0^t |\hat{s}_n(t)\hat{r}_n(u) - s(t)r(u)|\,du\,dt\biggr],
\end{aligned}$$

we conclude that $E_{s,r}[\int_0^{T^*} |\hat{r}_n(u) - r(u)|\,du] = O(n^{-q/(2q+1)} \log^{1/2} n)$ as $n \to \infty$. $\square$

We precede the proof of Theorem 5 with the following lemma.

LEMMA 12. *Let* $\tilde{\Theta}_{\boldsymbol{\kappa},q,n} \subseteq \Theta_{\boldsymbol{\kappa},q}$ *such that* $\mathrm{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n}) < \infty$. *Suppose that* $\tilde{s}_n, \tilde{r}_n$ *are estimators for* s, r, *respectively, based on* $N_1(t), \ldots, N_n(t), t \in [0, T)$. *Then*

$$\begin{aligned}
&\sup\biggl\{E_{s,r}\biggl[\int_0^T |\tilde{s}_n(t) - s(t)|\,dt\biggr] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\biggr\} \\
&\quad \geq \frac{1}{2}\inf\biggl\{\int_0^T |s_1(t) - s_2(t)|\,dt : s_1 \neq s_2, s_1, s_2 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}\biggr\} \\
&\qquad \times \biggl\{1 - \frac{1}{\log[\mathrm{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n}) - 1]} \\
&\qquad\qquad \times \biggl[\log 2 + \frac{1}{[\mathrm{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n})]^2}
\end{aligned} \tag{63}$$

$$\times \sum_{s_1, s_2 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}} \sum_{i=1}^{n} E_{s_1,r_1} \log \frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s_2,r_1}(\mathbf{w}_T^{(i)})}\Bigg]\Bigg\},$$

for any $r_1 \in \Theta_{\theta,\boldsymbol{\kappa},q}$. *Next, let* T^* *be a constant satisfying* $\theta < T^* < T$ *and* $\tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n} \subset \Theta_{\theta,\boldsymbol{\kappa},q}$ *such that card* $(\tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n}) < \infty$ *and* $r_1(u) = r_2(u)$, $u \in [T^*, T)$, $\forall r_1, r_2 \in \tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n}$. *Then*

$$\begin{aligned}
&\sup\bigg\{E_{s,r}\bigg[\int_0^{T^*} |\tilde{r}_n(t) - r(t)|\,dt\bigg] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\bigg\}\\
&\quad \geq \frac{1}{2}\inf\bigg\{\int_0^{T^*} |r_1(t) - r_2(t)|\,dt : r_1 \neq r_2, r_1, r_2 \in \tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n}\bigg\}\\
&\qquad \times \bigg\{1 - \frac{1}{\log[\mathrm{card}(\tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n}) - 1]}\\
&\qquad\quad \times \bigg[\log 2 + \frac{1}{[\mathrm{card}(\tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n})]^2}\\
&\qquad\qquad \times \sum_{r_1, r_2 \in \tilde{\Theta}_{\theta,T^*,\boldsymbol{\kappa},q,n}} \sum_{i=1}^{n} E_{s_1,r_1} \log \frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s_1,r_2}(\mathbf{w}_T^{(i)})}\bigg]\bigg\},
\end{aligned} \tag{64}$$

for any $s_1 \in \Theta_{\boldsymbol{\kappa},q}$.

PROOF. Following Yatracos [31], page 1183, we observe that

$$\begin{aligned}
&\sup\bigg\{E_{s,r}\bigg[\int_0^{T} |\tilde{s}_n(t) - s(t)|\,dt\bigg] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\bigg\}\\
&\quad \geq \frac{1}{\mathrm{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n})} \sum_{s_1 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}} E_{s_1,r_1}\bigg[\int_0^T |\tilde{s}_n(t) - s_1(t)|\,dt\bigg],
\end{aligned} \tag{65}$$

for any $r_1 \in \Theta_{\theta,\boldsymbol{\kappa},q}$. Define $\tilde{s}_n^* \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}$ such that

$$\int_0^T |\tilde{s}_n(t) - \tilde{s}_n^*(t)|\,dt = \inf\bigg\{\int_0^T |\tilde{s}_n(t) - s_1(t)|\,dt : s_1 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}\bigg\}.$$

Then, for $s_1 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}$, we have $\int_0^T |\tilde{s}_n^*(t) - s_1(t)|\,dt \leq 2\int_0^T |\tilde{s}_n(t) - s_1(t)|\,dt$. So

$$\begin{aligned}
&\sum_{s_1 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}} E_{s_1,r_1}\bigg[\int_0^T |\tilde{s}_n(t) - s_1(t)|\,dt\bigg]\\
&\quad \geq \frac{1}{2}\inf\bigg\{\int_0^T |s_1(t) - s_2(t)|\,dt : s_1 \neq s_2, s_1, s_2 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}\bigg\}\\
&\qquad \times \sum_{s_1 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}} P_{s_1,r_1}(\tilde{s}_n^* \neq s_1).
\end{aligned} \tag{66}$$

We observe from Fano's lemma (cf. Ibragimov and Has'minskii [16], pages 323–325, or Yatracos [31], page 1182) that

$$\begin{aligned}
&\frac{1}{\operatorname{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n})} \sum_{s_1 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}} P_{s_1,r_1}(\tilde{s}_n^* \neq s_1) \\
&\qquad \geq 1 - \frac{1}{\log[\operatorname{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n}) - 1]} \\
&\qquad\qquad \times \left\{ \log 2 + \frac{1}{[\operatorname{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n})]^2} \right. \\
&\qquad\qquad\qquad \left. \times \sum_{s_1, s_2 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n}} E_{s_1,r_1} \log\left[\prod_{i=1}^{n} \frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s_2,r_1}(\mathbf{w}_T^{(i)})} \right] \right\}.
\end{aligned} \tag{67}$$

(63) now follows from (65), (66) and (67). (64) is proved in a similar manner. □

PROOF OF THEOREM 5. Let $\{b_n > 0 : n = 1, 2, \ldots\}$ be a sequence of constants that tend to 0 as $n \to \infty$ and such that each b_n^{-1} is an integer. For $i = 1, \ldots, b_n^{-1}$, define $\phi_{i,n} : [0, T) \to \mathbb{R}$ by

$$\phi_{i,n}(t) = \begin{cases} (b_n T)^q \left[1 - \left(\dfrac{2t - (2i-1) b_n T}{b_n T} \right)^2 \right]^q, & \text{if } (i-1) b_n T \leq t < i b_n T, \\ 0, & \text{otherwise.} \end{cases}$$

Writing $q = q_0 + q_1$, where q_0 is a nonnegative integer and $0 < q_1 \leq 1$, we have

$$\lim_{n\to\infty} \max_{t \in [0,T)} \left| \frac{d^j}{dt^j} \phi_{i,n}(t) \right| < \infty \qquad \forall j = 0, \ldots, q_0,$$

$$\lim_{n\to\infty} \max_{t_1 \neq t_2 \in [0,T)} \left| \frac{d^{q_0}}{dt^{q_0}} \phi_{i,n}(t_1) - \frac{d^{q_0}}{dt^{q_0}} \phi_{i,n}(t_2) \right| \Big/ |t_1 - t_2|^{q_1} < \infty.$$

Let $\Xi_{a,n}$ denote functions of the form $a[1 + \sum_{i=1}^{b_n^{-1}} \gamma_i \phi_{i,n}(t)]^2$, $\forall t \in [0, T)$, where $\gamma_i = 0$ or 1 and $a > 0$ is a suitably small constant such that $\Xi_{a,n} \subset \Theta_{\boldsymbol{\kappa},q}$. If $s_1, s_2 \in \Xi_{a,n}$ where $s_1 \neq s_2$, then writing

$$s_1(t) = a\left[1 + \sum_{i=1}^{b_n^{-1}} \gamma_{1,i} \phi_{i,n}(t) \right]^2, \qquad s_2(t) = a\left[1 + \sum_{i=1}^{b_n^{-1}} \gamma_{2,i} \phi_{i,n}(t) \right]^2, \tag{68}$$

with $\gamma_{1,i}, \gamma_{2,i}$ taking values 0 or 1, we have

$$\begin{aligned}
\int_0^T |s_1(t) - s_2(t)| \, dt &\geq a \int_0^{b_n T} [2\phi_{1,n}(t) + \phi_{1,n}^2(t)] \, dt \\
&= a (b_n T)^{q+1} J_q + \frac{a (b_n T)^{2q+1} J_{2q}}{2},
\end{aligned}$$

where $J_l = \int_{-1}^{1}(1-y^2)^l\,dy > 0$, $\forall l > 0$. Also, it follows from (68) that

$$\left|\frac{s_1(t)-s_2(t)}{s_1(t)}\right| \le 2a\phi_{i,n}(t) + a\phi_{i,n}^2(t) \le 2a(b_nT)^q + a(b_nT)^{2q} \qquad \forall t \in [0,T).$$

Let $r_1 \in \Theta_{\theta,\boldsymbol{\kappa},q}$. Now, using Lemma 11,

$$\begin{aligned} E_{s_1,r_1}\log\left[\frac{p_{s_1,r_1}(\mathbf{w}_T^{(i)})}{p_{s_2,r_1}(\mathbf{w}_T^{(i)})}\right] &\le \frac{1}{2}\int_0^T\left(\frac{s_1(t)-s_2(t)}{s_1(t)}\right)^2 s_1(t)e^{-\int_0^t s_1(u)\,du}\,dt \\ &\quad + \frac{1}{2}\int_0^T\int_0^t\left(\frac{s_1(t)-s_2(t)}{s_1(t)}\right)^2\xi(t-u)s_1(t)r_1(t) \\ &\qquad \times e^{-\int_{t-u}^t s_1(v)r_1(v-t+u)\,dv}\,du\,dt \\ &\le \frac{a^2(b_nT)^{2q}}{2}[2+(b_nT)^q]^2(1+\bar{\kappa}^8T^2), \end{aligned}$$

where $\bar{\kappa} = \kappa_0 \vee 1$. Finally, we observe from Proposition 3.8 of Birgé [2] that there exists a subset $\tilde{\Theta}_{\boldsymbol{\kappa},q,n}$ of $\Xi_{a,n}$ such that

$$\int_0^T |s_1(t)-s_2(t)|\,dt \ge \frac{1}{8b_n}\left[a(b_nT)^{q+1}J_q + \frac{a(b_nT)^{2q+1}J_{2q}}{2}\right] \qquad \forall s_1 \ne s_2 \in \tilde{\Theta}_{\boldsymbol{\kappa},q,n},$$

and $\log[\mathrm{card}(\tilde{\Theta}_{\boldsymbol{\kappa},q,n})-1] > 0.316/b_n$. Consequently, it follows from (63) that

$$\begin{aligned} &\sup\left\{E_{s,r}\left[\int_0^T|\tilde{s}_n(t)-s(t)|\,dt\right] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\right\} \\ &\quad \ge \frac{1}{16b_n}\left[a(b_nT)^{q+1}J_q + \frac{a(b_nT)^{2q+1}J_{2q}}{2}\right] \\ &\qquad \times\left\{1 - \frac{b_n}{0.316}\left[\log 2 + \frac{a^2n(b_nT)^{2q}}{2}[2+(b_nT)^q]^2(1+\bar{\kappa}^8T^2)\right]\right\}. \end{aligned}$$

Thus, we conclude that there exist strictly positive constants C_0 and $C_{\boldsymbol{\kappa},q}$ (depending only on $\boldsymbol{\kappa}$ and q) such that by taking $b_n = 1/\lceil C_0 n^{1/(2q+1)}\rceil$, we have

$$\sup\left\{E_{s,r}\left[\int_0^T|\tilde{s}_n(t)-s(t)|\,dt\right] : s \in \Theta_{\boldsymbol{\kappa},q}, r \in \Theta_{\theta,\boldsymbol{\kappa},q}\right\} \ge C_{\boldsymbol{\kappa},q}n^{-q/(2q+1)}.$$

The proof of the second part of Theorem 5 is similar and is omitted. □

Acknowledgments. Wei-Liem Loh would like to thank Professor Yannis Yatracos for the many discussions on metric entropy and minimum distance estimation and Professor Zhiyi Chi for introducing to him the field of neuroscience when he visited the University of Chicago in the Spring of 2003. We would also like to thank Professor Morris Eaton, an Associate Editor and a referee for their suggestions and comments on this manuscript.

REFERENCES

[1] BERRY, M. J. and MEISTER, M. (1998). Refractoriness and neural precision. *J. Neurosci.* **18** 2200–2211.

[2] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129

[3] BRILLINGER, D. R. (1992). Nerve cell spike train data analysis: A progression of technique. *J. Amer. Statist. Assoc.* **87** 260–271.

[4] BROWN, E. N., KASS, R. E. and MITRA, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neurosci.* **7** 456–461.

[5] CHAN, H. P. (2003). Upper bounds and importance sampling of p-values for DNA and protein sequence alignments. *Bernoulli* **9** 183–199. MR1997026

[6] CHAN, H. P. and LOH, W.-L. (2006). Some theoretical results on neural spike train probability models. Available at www.arxiv.org/abs/math.st/0703829.

[7] CHI, Z. (2005). Large deviations for template matching between point processes. *Ann. Appl. Probab.* **15** 153–174. MR2115040

[8] CHI, Z., RAUSKE, P. L. and MARGOLIASH, D. (2003). Pattern filtering for detection of neural activity, with examples from HVc activity during sleep in zebra finches. *Neural Comput.* **15** 2307–2337.

[9] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes* **1**. *Elementary Theory and Methods*, 2nd ed. Springer, New York. MR1950431

[10] DAVE, A. S. and MARGOLIASH, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290** 812–816.

[11] DAYAN, P. and ABBOTT, L. F. (2001). *Theoretical Neuroscience*: *Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge, MA. MR1985615

[12] DAYHOFF, J. E. and GERSTEIN, G. L. (1983). Favored patterns in spike trains. I. Detection. *J. Neurophysiol.* **49** 1334–1348.

[13] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Univ. Press. MR1720712

[14] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York. MR0270403

[15] GRÜN, S., DIESMANN, M. and AERTSEN, A. (2002). Unitary events in multiple single-neuron spiking activity. I. Detection and significance. *Neural Comput.* **14** 43–80.

[16] IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. Springer, New York. MR0620321

[17] JOHNSON, D. H. and SWAMI, A. (1983). The transmission of signals by auditory-nerve fiber discharge patterns. *J. Acoust. Soc. Am.* **74** 493–501.

[18] KASS, R. E. and VENTURA, V. (2001). A spike-train probability model. *Neural Comput.* **13** 1713–1720.

[19] KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Mat. Nauk* **14** 3–86 [in Russian; English transl. *Amer. Math. Soc. Transl. (2)* **17** 277–364 (1961)]. MR0112032, MR0124720

[20] LAI, T. L. and SHAN, J. Z. (1999). Efficient recursive algorithms for detection of abrupt changes in signals and control systems. *IEEE Trans. Automat. Control* **44** 952–966. MR1690539

[21] MILLER, M. I. (1985). Algorithms for removing recovery-related distortion from auditory-nerve discharge patterns. *J. Acoust. Soc. Am.* **77** 1452–1464.

[22] MOONEY, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. *J. Neurosci.* **20** 5420–5436.

[23] NÁDASDY, Z., HIRASE, H., CZURKÓ, A., CSICSVARI, J. and BUZSÁKI, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *J. Neurosci.* **19** 9497–9507.

[24] OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919. MR0893905

[25] SHEN, X. and WONG, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22** 580–615. MR1292531

[26] SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4** 673–684. MR0418369

[27] SIEGMUND, D. (1985). *Sequential Analysis*. *Tests and Confidence Intervals*. Springer, New York. MR0799155

[28] STONE, C. (1965). A local limit theorem for nonlattice multi-dimensional distribution functions. *Ann. Math. Statist.* **36** 546–551. MR0175166

[29] VENTURA, V., CARTA, R., KASS, R. E., GETTNER, S. N. and OLSON, C. R. (2002). Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics* **3** 1–20.

[30] WONG, W. H. and SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. MR1332570

[31] YATRACOS, Y. (1988). A lower bound on the error in nonparametric regression type problems. *Ann. Statist.* **16** 1180–1187. MR0959195

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
REPUBLIC OF SINGAPORE
E-MAIL: stachp@nus.edu.sg
stalohwl@nus.edu.sg