Estimation for Nonhomogeneous Poisson Processes from Aggregated Data

Shane G. Henderson*
School of Operations Research and Industrial Engineering,
Cornell University, Ithaca, NY 14853.

November 22, 2002

Abstract

A well-known heuristic for estimating the rate function or cumulative rate function of a nonhomogeneous Poisson process assumes that the rate function is piecewise constant on a set of data-independent intervals. We investigate the asymptotic (as the amount of data grows) behavior of this estimator in the case of equal interval widths, and show that it can be transformed into a consistent estimator if the interval lengths shrink at an appropriate rate as the amount of data grows.

1 Introduction

Nonhomogeneous Poisson processes (NHPPs) are widely used to model time-dependent arrivals in a multitude of stochastic models. Their widespread use is perhaps a consequence of the fact that they may be defined in terms of very natural assumptions about the mechanism through which events occur. The following definition is standard (see, e.g., Ross [14]), but not the most general possible (see, e.g., Resnick [13]).

Definition 1 *Let $N = (N(t) : t \geq 0)$ be an integer-valued, nondecreasing process with $N(0) = 0$. We say that N is a NHPP with rate (or intensity) function $\lambda = (\lambda(t) : t \geq 0)$ if*

1. *the process N has independent increments,*
2. *for all $t \geq 0$, $P(N(t + h) - N(t) \geq 2) = o(h)$, and*

*Email: shane@orie.cornell.edu, www:http://www.orie.cornell.edu/∼shane

3. for all $t \geq 0$, $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$.

(We say that a function $f(h) = o(g(h))$ if $f(h)/g(h) \to 0$ as $h \to 0$ from above, and $f(h) = O(g(h))$ if $f(h)/g(h)$ is bounded in h.) Associated with a NHPP is the cumulative rate (or cumulative intensity) function $\Lambda = (\Lambda(t) : t \geq 0)$, where $\Lambda(t) = \int_0^t \lambda(s)\, ds$ for all $t \geq 0$.

In this paper, we consider the problem of estimating the rate function λ, and the cumulative rate function Λ, over a finite interval $[0, T)$ (T is a fixed constant), for the purpose of generating realizations in a discrete-event simulation. Since we assume that Λ has the density λ, it is immaterial whether we include the endpoints of intervals or not. We use intervals of the form $[a, b)$ for definiteness. We assume that several i.i.d. observed realizations of N are available over $[0, T)$, although some of our results rely only on an aggregated version of this data (i.e., the data consists of several realizations, each consisting of the number of events in each of equal-sized intervals or bins).

A nonparametric estimator of the rate function was suggested in Lewis and Shedler [11] based on kernel estimation techniques. The kernel estimator is not often considered within the discrete-event simulation community, perhaps because of the computational effort required to generate realizations from a fitted kernel estimator of the rate function. Lewis and Shedler [11] also propose a parametric estimator of the rate function. Other parametric estimators are developed in, for example, Lee et al. [9], Kuhl et al. [7], Kuhl and Wilson [5], and Kao and Chang [3]. Kuhl and Wilson [6] construct a hybrid parametric/nonparametric estimator, and Kuhl and Bhairgond [4] give a nonparametric estimator using wavelets.

Leemis [10] gives a nonparametric estimator of Λ when several i.i.d. realizations of N are observed over the time interval $[0, T)$. He provides an efficient algorithm for generating the NHPP from the fitted cumulative intensity function, and establishes the asymptotic (in the number of observed realizations) behavior of the estimator through a strong law and central limit theorem. Arkin and Leemis [1] extend this work to allow for "partial" realizations where each observed realization may not necessarily cover the full interval $[0, T)$. The nonparametric estimators of the cumulative rate function developed by Leemis, and Arkin and Leemis, are easily computed from the data, and their asymptotic (as the number of observed realizations increases) behavior is well-understood. Furthermore, simulated realizations can be rapidly generated from the estimated rate functions. An important disadvantage is that they require that every event time in every observed realization be retained in memory to allow the generation of simulated realizations.

In contrast, most of the parametric estimators mentioned earlier are based on a fixed number of parameters, and so their storage requirements do not increase as the number of observed realizations increases. With careful implementation, rapid generation is possible. Furthermore, the parametric forms can be chosen so as to incorporate prior information about the rate function. They have the disadvantage that their asymptotic behavior is, in general, not well understood. Another, perhaps less important, objection is that they will not converge to the true rate function if the true rate function does not lie within the assumed parametric class. But perhaps most importantly, their estimation through maximum likelihood or other techniques represents a nontrivial computational task.

Law and Kelton [8] describe a heuristic for estimating the rate function of a NHPP. The method assumes that the rate function is piecewise constant, with known breakpoints where the rate changes value. The rate within each subinterval is then easily estimated from the data. This approach is heuristic in that the breakpoints are user-specified and it will not necessarily converge to the true rate function as the number of observed realizations increases. However the estimator has several desirable properties. First, it is easily computed from the data, and its storage requirements are determined by the number of chosen subintervals, so that the storage requirements do not necessarily increase with the number of observed realizations. Second, its asymptotic properties are easily derived, and we do so in this paper. These properties are important, as they provide an understanding of the error in the estimator. Third, simulated realizations can be easily generated from the fitted rate function.

In this paper we consider a specialization of the Law and Kelton estimator where the subintervals are assumed to be of equal length. Our motivation for doing so is twofold.

1. Many database systems employed in service systems to track performance do not record individual transactions. Instead, they track aggregate performance over fixed increments of time δ say, where common choices of δ are 60 minutes, 30 minutes, or 15 minutes. When data is in this form, individual event times are not available, and the Leemis estimator cannot be employed. Most of the parametric estimators can still be employed, but they suffer from the

disadvantages alluded to earlier including computationally-intensive estimation procedures. The Law and Kelton estimator is easily computed given this data, and is often used in practice, so it is important to understand its properties.

2. Suppose that one does in fact have data consisting of individual event times for each of n observed realizations. If one allows δ to decrease as a function of n, then one can view the Law and Kelton estimator as a nonparametric estimator. By choosing $\delta = \delta_n$ appropriately, one obtains a consistent estimator of the rate function. Note that the entire data set may be stored in a database off-line, but the simulation program need only store the aggregated data, which may represent a considerable savings. The assumption of equal interval widths leads to simplicity in both theory and implementation.

The tradeoff in selecting δ_n is essentially the same as that involved in kernel smoothing; see, e.g., Wand and Jones [15]. If δ_n is selected too large, then the time-dependent nature of the arrival rate function is lost through over-smoothing. If δ_n is selected too small, then "overfitting to the data" occurs, and it is possible for many intervals to be assigned a rate of 0. One will never see a generated event in such intervals.

There are basically two methods for generating NHPPs. The first is a form of "inversion" procedure (see, e.g., Law and Kelton [8], p. 486) that relies on an estimate of the cumulative rate function Λ. We explore an estimator Λ_n of Λ in Section 2, and find that the bias involved in using intervals of (fixed) width δ is typically of the order δ^2. When $\delta = \delta_n$ varies with the number of observed realizations n appropriately, the estimator Λ_n converges to Λ at rate $n^{-1/2}$ as $n \to \infty$, which is the same rate as the Leemis estimator. In fact, for each fixed t, $\Lambda_n(t)$ satisfies a central limit theorem (CLT) identical to that of the Leemis estimator, so that to second order the estimators are statistically identical, even though the estimator Λ_n requires significantly less storage.

The second method for generating a NHPP is the thinning method introduced by Lewis and Shedler [12] that relies on an estimate of the rate function λ. We discuss an estimator of the rate function in Section 3. In the case where $\delta = \delta_n$ is allowed to vary with n, we show that in terms of a bound on mean squared error, the optimal interval width is of the order $n^{-1/3}$, and in this case the mean squared error converges to 0 at rate $n^{-2/3}$. This corresponds to a convergence rate of $n^{-1/3}$, which is slower than the rate $n^{-1/2}$ obtained for the estimator Λ_n. Intuitively speaking, the slower convergence rate arises because $\Lambda(t)$ can be expressed as the expected number of events in $[0,t)$ and thus can be estimated by a sample average, while no such statement is true for $\lambda(t)$.

We view the choice of method (inversion or thinning) for generating a NHPP as application dependent and therefore beyond the scope of this paper, since our focus is on the asymptotics of estimators of the cumulative rate function and rate function.

2 Estimating the Cumulative Rate Function

Consider the problem of estimating the cumulative rate function Λ over the interval $[0,T)$. We begin by considering the case where the subinterval width $\delta > 0$ is fixed, and the number of observed realizations n of the NHPP N over the interval $[0,T)$ grows without bound. In this case it is immaterial whether the data is in aggregate form or not, so long as the data gives, for each observed realization, the number of events in each interval $[(k-1)\delta, k\delta)$, for $k = 1, \ldots, \lceil T/\delta \rceil$. Let $\tilde{\Lambda}_n(t)$ denote the estimator of $\Lambda(t)$ based on n independent observed realizations of N over the interval $[0,T)$. To define $\tilde{\Lambda}_n(t)$ we need some notation.

Let $N_i(a,b)$ denote the number of events falling in the interval $[a,b)$ in the ith independent observed realization of N. For $t \geq 0$, let

$$\ell(t) = \left\lfloor \frac{t}{\delta} \right\rfloor \delta$$

so that t belongs to the subinterval $[\ell(t), \ell(t)+\delta)$. Then, for $t \in [0,T)$,

$$\tilde{\Lambda}_n(t) = \frac{1}{n}\sum_{i=1}^{n} N_i(0,\ell(t)) + \frac{t-\ell(t)}{\delta}\frac{1}{n}\sum_{i=1}^{n} N_i(\ell(t),\ell(t)+\delta). \tag{1}$$

For $t \in [0,T)$, define

$$\tilde{\Lambda}(t) = \Lambda(\ell(t)) + \frac{t-\ell(t)}{\delta}[\Lambda(\ell(t)+\delta) - \Lambda(\ell(t))] \tag{2}$$

to be a piecewise-linear approximation of Λ. Specifically, $\tilde{\Lambda}(t)$ equals $\Lambda(t)$ at the breakpoints $\{t : t = \ell(t)\}$, and linearly interpolates between these values at other points. Let $\Rightarrow$ denote convergence in distribution, and $\mathcal{N}(\mu,\sigma^2)$ denote a normally distributed random variable with mean μ and variance σ^2.

Proposition 1 *With the definitions above,*

1. $\sup_{t\in[0,T)} |\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t)| \to 0$ *almost surely as* $n \to \infty$,

2. *for all* $t \in [0,T)$, $\sqrt{n}(\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t)) \Rightarrow \sigma(t)\mathcal{N}(0,1)$ *as* $n \to \infty$, *where*

$$\sigma^2(t) = \Lambda(\ell(t)) + \left(\frac{t-\ell(t)}{\delta}\right)^2 [\Lambda(\ell(t)+\delta) - \Lambda(\ell(t))],$$

   and

3. *for all* $n \geq 1$, $E\tilde{\Lambda}_n(t) = \tilde{\Lambda}(t)$. *If* λ *is continuously differentiable on an open interval containing* $[\ell(t), \ell(t)+\delta]$, *then*

$$|E\tilde{\Lambda}_n(t) - \Lambda(t)| \leq |\lambda'(\zeta)|\delta^2,$$

   for some $\zeta \in [\ell(t), \ell(t)+\delta]$.

Proof:

Recall that $N_i(a,b)$ is Poisson with mean (and variance) $\Lambda(b) - \Lambda(a)$ for all $0 \leq a < b$ and all i. Applying the strong law of large numbers to each of the averages in (1) gives the strong law for each t. It remains to establish the *uniform* part of the result. Note that $\tilde{\Lambda}$ is a continuous nondecreasing function. Therefore, for all $\epsilon > 0$, there exists $m(\epsilon) < \infty$ and points $u_0 = 0, u_1, \ldots, u_{m(\epsilon)} = T$ such that $\tilde{\Lambda}(u_i) - \tilde{\Lambda}(u_{i-1}) \leq \epsilon$ for all $i = 1, \ldots, m(\epsilon)$. For $t \in [0,T)$, let $a(t) = a_\epsilon(t)$ denote the value i such that $t \in [u_i, u_{i+1})$. Then

$$\begin{aligned}
|\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t)| &= \max[\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t), \tilde{\Lambda}(t) - \tilde{\Lambda}_n(t)] \\
&\leq \max[\tilde{\Lambda}_n(u_{a(t)+1}) - \tilde{\Lambda}(u_{a(t)}), \tilde{\Lambda}(u_{a(t)+1}) - \tilde{\Lambda}_n(u_{a(t)})] \\
&\leq \max[|\tilde{\Lambda}_n(u_{a(t)+1}) - \tilde{\Lambda}(u_{a(t)+1})| + \tilde{\Lambda}(u_{a(t)+1}) - \tilde{\Lambda}(u_{a(t)}), \\
&\qquad \tilde{\Lambda}(u_{a(t)+1}) - \tilde{\Lambda}(u_{a(t)}) + |\tilde{\Lambda}(u_{a(t)}) - \tilde{\Lambda}_n(u_{a(t)})|] \\
&\leq \epsilon + \max_{i=1,\ldots,m(\epsilon)} |\tilde{\Lambda}_n(u_i) - \tilde{\Lambda}(u_i)|, \qquad (3)
\end{aligned}$$

where we have used the monotonicity of $\tilde{\Lambda}_n$ and $\tilde{\Lambda}$. The bound (3) is the same for all $t \in [0,T)$. Taking limit suprema in the above, we obtain that, almost surely,

$$\limsup_{n\to\infty} \sup_{t\in[0,T)} |\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t)| \leq \epsilon.$$

Since ϵ was arbitrary this completes the proof of the uniform strong law.

Turning to the second result, the standard CLT allows us to conclude that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} N_i(a,b) - (\Lambda(b) - \Lambda(a))\right) \Rightarrow \mathcal{N}(0, \Lambda(b) - \Lambda(a))$$

as $n \to \infty$. The CLT for $\tilde{\Lambda}_n(t)$ then follows from this observation and the independent increments property of N.

The fact that $E\tilde{\Lambda}_n(t) = \tilde{\Lambda}(t)$ is immediate from the definitions of $\tilde{\Lambda}_n(t)$ and $\Lambda(t)$, since $EN_i(0,t) = \Lambda(t)$ for all $t \geq 0$ and all i. Finally,

$$\begin{aligned} E\tilde{\Lambda}_n(t) - \Lambda(t) &= \tilde{\Lambda}(t) - \Lambda(t) \\ &= [\Lambda(\ell(t)) - \Lambda(t)] + \frac{t - \ell(t)}{\delta}[\Lambda(\ell(t) + \delta) - \Lambda(\ell(t))] \\ &= -\lambda(\xi)(t - \ell(t)) + \frac{t - \ell(t)}{\delta}\lambda(\theta)\delta \end{aligned}$$

for some $\xi \in [\ell(t), t]$ and $\theta \in [\ell(t), \ell(t) + \delta]$ by the mean value theorem. Thus

$$\begin{aligned} |E\tilde{\Lambda}_n(t) - \Lambda(t)| &= (t - \ell(t))|\lambda(\theta) - \lambda(\xi)| \\ &\leq \delta|\lambda'(\zeta)|\delta \end{aligned}$$

where $\zeta \in [\ell(t), \ell(t) + \delta]$. ■

Proposition 1 sheds light on the performance of the Law and Kelton estimator with fixed subinterval widths. It therefore gives an indication of the performance of this estimator in the setting where only aggregate data is available. The estimator $\tilde{\Lambda}_n(t)$ converges to the true value $\Lambda(t)$ at breakpoints, and in general converges to $\tilde{\Lambda}(t)$ at rate $n^{-1/2}$. Furthermore its bias is independent of n, typically of the order δ^2, and small when $|\lambda'|$ is small.

One might then ask whether it is possible to obtain a consistent estimator of $\Lambda(t)$ if one chooses the subinterval length $\delta = \delta_n$ to be a function of n, the number of observed realizations of N. Let us now consider an estimator Λ_n defined as in (1), but with the difference that now δ_n is permitted to vary with n. We now write $\ell_n(t) = \lfloor t/\delta_n \rfloor \delta_n$ instead of $\ell(t)$ to reflect the fact that breakpoints now change with n.

Proposition 2 *If $\delta_n \to 0$ as $n \to \infty$, then $\sup_{t\in[0,T)} |\Lambda_n(t) - \Lambda(t)| \to 0$ almost surely as $n \to \infty$. If $\delta_n = o(n^{-1/4})$ and the rate function λ is Lipschitz continuous in a neighborhood of t, then*

$$\sqrt{n}(\Lambda_n(t) - \Lambda(t)) \Rightarrow \mathcal{N}(0, \Lambda(t))$$

as $n \to \infty$.

Proof:

The uniform strong law follows as in Proposition 1 once we show pointwise almost sure convergence. Observe that $\Lambda_n(t) = n^{-1}\sum_{i=1}^n N_i(0,t) + R_n$, where

$$R_n = -\frac{1}{n}\sum_{i=1}^n N_i(\ell_n(t), t) + \frac{t - \ell_n(t)}{\delta_n}\frac{1}{n}\sum_{i=1}^n N_i(\ell_n(t), \ell_n(t) + \delta_n).$$

Since $\ell_n(t) \le t \le \ell_n(t) + \delta_n$,

$$|R_n| \le \frac{2}{n}\sum_{i=1}^n N_i(\ell_n(t), \ell_n(t) + \delta_n).$$

Now $\ell_n(t) \to t$ and $\delta_n \to 0$ as $n \to \infty$, so for all $\epsilon > 0$, $t - \epsilon \le \ell_n(t) \le \ell_n(t) + \delta_n \le t + \epsilon$ for n sufficiently large. It immediately follows that

$$\limsup_{n\to\infty} |R_n| \le \limsup_{n\to\infty} \frac{2}{n}\sum_{i=1}^n N_i(t-\epsilon, t+\epsilon) = 2\int_{t-\epsilon}^{t+\epsilon} \lambda(s)\, ds$$

almost surely. Since ϵ was arbitrary, $R_n \to 0$ almost surely as $n \to \infty$. The (pointwise) strong law now follows by the strong law of large numbers applied to $n^{-1}\sum_{i=1}^n N_i(0,t)$.

We now establish the central limit theorem. Fix $t \in [0,T)$ and suppose that n is sufficiently large so that λ is Lipschitz continuous in $[\ell_n(t), \ell_n(t) + \delta_n]$. Observe that

$$\begin{aligned} n^{1/2}(\Lambda_n(t) - \Lambda(t)) &= n^{-1/2}\sum_{i=1}^n N_i(0, \ell_n(t)) + n^{-1/2}\frac{t - \ell_n(t)}{\delta_n}\sum_{i=1}^n N_i(\ell_n(t), \ell_n(t) + \delta_n) \\ &\quad - n^{1/2}\Lambda(t) \\ &\stackrel{\mathcal{D}}{=} n^{-1/2}X_n + n^{-1/2}\frac{t - \ell_n(t)}{\delta_n}Y_n - n^{1/2}\Lambda(t), \end{aligned}$$

where X_n and Y_n are independent Poisson random variables with respective means $n\Lambda(\ell_n(t))$ and $n[\Lambda(\ell_n(t) + \delta_n) - \Lambda(\ell_n(t))] = n\lambda(\xi_n)\delta_n$ for some $\xi_n \in [\ell_n(t), \ell_n(t) + \delta_n]$. The notation $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution.

Hence, if ϕ_n is the moment generating function of $n^{1/2}(\Lambda_n(t) - \Lambda(t))$, then

$$\begin{aligned}
\ln \phi_n(u) &= n\Lambda(\ell_n(t))[e^{n^{-1/2}u} - 1] + n\lambda(\xi_n)\delta_n \left[\exp\left(\frac{t-\ell_n(t)}{\delta_n}\frac{u}{\sqrt{n}}\right) - 1\right] - n^{1/2}\Lambda(t)u \\
&= n\Lambda(\ell_n(t))[n^{-1/2}u + n^{-1}u^2/2 + O(n^{-3/2})] \\
&\quad + n\lambda(\xi_n)\delta_n \left[\frac{t-\ell_n(t)}{\delta_n}\frac{u}{\sqrt{n}} + O\left(\frac{(t-\ell_n(t))^2}{n\delta_n^2}\right)\right] - n^{1/2}\Lambda(t)u \\
&= un^{1/2}[\Lambda(\ell_n(t)) - \Lambda(t)] + \Lambda(\ell_n(t))u^2/2 + O(n^{-1/2}) \\
&\quad + un^{1/2}\lambda(\xi_n)[t - \ell_n(t)] + O\left(\frac{(t-\ell_n(t))^2}{\delta_n}\right).
\end{aligned}$$

Now, $\Lambda(\ell_n(t)) - \Lambda(t) = -\lambda(\theta_n)(t - \ell_n(t))$ for some $\theta_n \in [\ell_n(t), t]$, and $t - \ell_n(t) \leq \delta_n$. So

$$\ln \phi_n(u) = un^{1/2}(t - \ell_n(t))(\lambda(\xi_n) - \lambda(\theta_n)) + \Lambda(\ell_n(t))u^2/2 + O(n^{-1/2} + \delta_n). \tag{4}$$

Lipschitz continuity of $\lambda(\cdot)$ ensures that the first term in (4) is bounded in absolute value by $|u|n^{1/2}K\delta_n^2$ for some positive (Lipschitz) constant K. Since $\delta_n^2 = o(n^{-1/2})$, this first term converges to 0 as $n \to \infty$. As for the second term, note that $\ell_n(t) \to t$ as $n \to \infty$, and Λ is continuous at t. This completes the proof. ■

Remark 1 The hypotheses of the central limit theorem in Proposition 2 can be replaced by the assumption that λ is continuous in a neighborhood of t, and $\delta_n = o(n^{-1/2})$. (Note that the first term in (4) still converges to 0 under these conditions.) This is a weaker condition on λ but a stronger condition on δ_n.

3 Estimating the Rate Function

Now consider the problem of estimating the rate function λ over the interval $[0, T)$. This rate function can be used to generate simulated realizations of the NHPP through a thinning procedure introduced in Lewis and Shedler [12]. The nature of the rate function estimator is such that thinning gives a fast generation procedure. (One could also use inversion, but we believe that thinning will typically lead to faster generation of simulated realizations.)

To see why, let λ_n be the rate function estimator. Recall that thinning first generates a candidate event time T^* say, and then accepts the event time with probability $\lambda_n(T^*)/\lambda^*$, where

λ^* is an upper bound on λ_n. So thinning requires rapid calculation of $\lambda_n(\cdot)$. Since λ_n is defined piecewise on equal-sized intervals, the interval containing a given time T^* can be computed in $O(1)$ time as n grows, and so $\lambda_n(T^*)$ can be rapidly computed. This observation is analogous to one by Leemis [10] related to the efficiency of generating the process N based on the nonparametric cumulative intensity function estimator given there.

Again, let us first consider the case where the subinterval length δ is fixed. Define

$$\tilde{\lambda}(t) = \frac{1}{\delta} \int_{\ell(t)}^{\ell(t)+\delta} \lambda(s)\, ds$$

to be the "aggregated" rate function that is constant on each interval of the form $[(k-1)\delta, k\delta)$ for $k \geq 1$. Also, denote our estimator of the rate function by

$$\tilde{\lambda}_n(t) = \frac{1}{n\delta} \sum_{i=1}^{n} N_i(\ell(t), \ell(t)+\delta). \tag{5}$$

Proposition 3 mirrors Proposition 1 in that it describes the large-sample behaviour of $\tilde{\lambda}_n$ for a fixed interval width δ. The proof is elementary and omitted.

Proposition 3 *For all* $t \in [0, T)$,

1. $\tilde{\lambda}_n(t) \to \tilde{\lambda}(t)$ *almost surely as* $n \to \infty$, *and*

2. $$n^{1/2}(\tilde{\lambda}_n(t) - \tilde{\lambda}(t)) \Rightarrow \delta^{-1}\mathcal{N}(0, \Lambda(\ell(t)+\delta) - \Lambda(\ell(t)))$$

   as $n \to \infty$.

3. *For all* $n \geq 1$, $E\tilde{\lambda}_n(t) = \tilde{\lambda}(t)$. *If* λ *is continuous on an open interval containing* $[\ell(t), \ell(t)+\delta]$, *then* $E\tilde{\lambda}_n(t) = \lambda(\zeta)$ *for some* ζ *lying in the interval* $[\ell(t), \ell(t)+\delta]$.

We see that the bias in $\tilde{\lambda}_n$ depends on the local behavior of $\lambda(\cdot)$ in a neighborhood of t, as will become further evident in the proof of Proposition 4.

Now consider the case where the subinterval length $\delta = \delta_n$ depends on n, the number of observed realizations of the process N. Let $\lambda_n(t)$ be defined as in (5) where δ is taken to equal δ_n. We say that a random sequence $(V_n : n \geq 1)$ is tight if for all $\epsilon > 0$, there exists a deterministic constant $M = M(\epsilon) > 0$ such that $P(|V_n| > M) \leq \epsilon$ for all n.

Proposition 4 *Suppose that $\delta_n \to 0$ and $n\delta_n \to \infty$ as $n \to \infty$. Fix $t \in [0, T)$.*

1. *If λ is continuous in a neighborhood of t, then $\lambda_n(t) \Rightarrow \lambda(t)$ as $n \to \infty$.*

2. *If λ is Lipschitz continuous in a neighborhood of t, and $\delta_n = o(n^{-1/3})$, then*

$$(n\delta_n)^{1/2}(\lambda_n(t) - \lambda(t)) \Rightarrow \mathcal{N}(0, \lambda(t))$$

   as $n \to \infty$.

3. *If λ is continuously differentiable in a neighborhood of t and $\lambda'(t) \neq 0$, then for n sufficiently large a bound on the mean squared error of $\lambda_n(t)$ is minimized by taking $\delta_n = \delta_n^*$ where*

$$\delta_n^* \sim \left(\frac{\lambda(t)}{2\lambda'(t)^2}\right)^{1/3} n^{-1/3},$$

   and $c_n \sim d_n$ if $\lim_{n\to\infty} c_n/d_n = 1$. In the case where there are constants $a, b > 0$ with $an^{-1/3} < \delta_n < bn^{-1/3}$ for all n, the sequence of random variables $\{(n\delta_n)^{1/2}(\lambda_n(t) - \lambda(t)) : n \geq 1\}$ is tight.

Proof:

Fix $t \geq 0$ and let $\ell_n(t)$ be defined as in the previous section. Define $X(i,n) = N_i(\ell_n(t), \ell_n(t) + \delta_n)$, so that

$$\lambda_n(t) = \frac{1}{n\delta_n} \sum_{i=1}^{n} X(i,n).$$

Note that $\sum_{i=1}^n X(i,n)$ is a Poisson random variable with mean $n[\Lambda(\ell_n(t) + \delta_n) - \Lambda(\ell_n(t))]$. This mean is given, for n sufficiently large (say $n \geq n_1$), by $n\delta_n\lambda(\xi_n)$ for some $\xi_n \in [\ell_n(t), \ell_n(t) + \delta_n]$. Let $\varphi_n(\cdot)$ be the moment generating function of $\lambda_n(t)$. Then for $n \geq n_1$,

$$\begin{aligned}
\varphi_n(u) &= E \exp\left(\frac{u}{n\delta_n}\sum_{i=1}^{n} X(i,n)\right) \\
&= \exp\left(n\delta_n\lambda(\xi_n)(e^{u/n\delta_n} - 1)\right) \\
&\to e^{\lambda(t)u}
\end{aligned}$$

as $n \to \infty$, where the convergence follows since $\lambda(\xi_n) \to \lambda(t)$ as $n \to \infty$ and $y(e^{u/y} - 1) \to u$ as $y \to \infty$. This proves the required weak law of large numbers.

The central limit theorem can be proved using the same approach as in Proposition 2. We omit the details. It remains to establish the mean squared error result.

Note that $\lambda'(t) \neq 0$ implies that $\lambda(t) > 0$ so that the coefficient of $n^{-1/3}$ as given in the statement of the proposition is positive. The bias of $\lambda_n(t)$ is given, for n sufficiently large, by

$$E\lambda_n(t) - \lambda(t) = \lambda(\xi_n) - \lambda(t) = \lambda'(\theta_n)(\xi_n - t),$$

for some $\theta_n \in [\ell_n(t), \ell_n(t) + \delta_n]$, so that the bias in $\lambda_n(t)$ is at most $|\lambda'(\theta_n)|\delta_n$.

Similarly, for large enough n we can compute Var $\lambda_n(t)$ to equal $\lambda(\xi_n)/n\delta_n$. The mean squared error of $\lambda_n(t)$ for large enough n is then bounded by

$$\frac{\lambda(\xi_n)}{n\delta_n} + \lambda'(\theta_n)^2\delta_n^2, \tag{6}$$

which is minimized at

$$\begin{aligned} \delta_n &= \left(\frac{\lambda(\xi_n)}{2\lambda'(\theta_n)^2}\right)^{1/3} n^{-1/3} \\ &\sim \left(\frac{\lambda(t)}{2\lambda'(t)^2}\right)^{1/3} n^{-1/3}. \end{aligned} \tag{7}$$

To prove tightness, note that if $an^{-1/3} < \delta_n < bn^{-1/3}$, then from (6),

$$n^{2/3}E(\lambda_n(t) - \lambda(t))^2$$

is bounded in n. Chebyshev's inequality then establishes the result. ■

Remark 2 The weak convergence established in Part 1 of Proposition 4 can be strengthened to almost sure convergence if we require that $n\delta_n \to \infty$ as $n \to \infty$ *fast enough*. Suppose that, in addition to the conditions in Part 1 of Proposition 4, $\sum_{n=1}^{\infty} \rho^{n\delta_n} < \infty$ for all $\rho \in (0,1)$. (This condition holds, for example, if $\delta_n = n^{-a}$ for any $a \in (0,1)$.) Then $\lambda_n(t) \to \lambda(t)$ almost surely as $n \to \infty$. A proof of this result is similar to the proof of the strong law of large numbers for i.i.d. random variables with finite fourth moment given in Ross [14, p. 56]. A sketch of the proof is as follows. One first establishes that for given $\epsilon > 0$,

$$P(|\lambda_n(t) - \lambda(t)| > \epsilon) \leq 2\rho^{n\delta_n}$$

for $n \geq n(\epsilon)$ say using Chernoff bounds (see, e.g., Ross [14, p. 39]), and some $\rho < 1$. We then obtain that

$$\sum_{n=1}^{\infty} P(|\lambda_n(t) - \lambda(t)| > \epsilon) \leq n(\epsilon) - 1 + 2 \sum_{n=n(\epsilon)}^{\infty} \rho^{n\delta_n} < \infty$$

and the Borel-Cantelli lemma then ensures that $P(|\lambda_n(t) - \lambda(t)| > \epsilon$ infinitely often$) = 0$, establishing the strong law.

Proposition 4 shows that choosing δ_n to be of the order $n^{-1/3}$ asymptotically minimizes a bound on the mean squared error. This choice of δ_n ensures that $\lambda_n(t)$ converges at rate $(n\delta_n)^{-1/2} = n^{-1/3}$. A slower rate of convergence than $n^{-1/2}$ is representative of many nonparametric function estimators; see Wand and Jones [15] for example. It is also consistent with results for density estimation via histograms, which is a subject that is closely related to the discussion in this section; see Freedman and Diaconis [2].

Notice that $\lambda'(t)^2$ appears in the denominator in (7), showing (as intuition would suggest) that when λ is changing rapidly one prefers small intervals. Note that these are "local" results, in that they only apply to the choice of subinterval width near t. Based on this observation, one might consider an estimator of λ with varying interval widths, as is the case for the Leemis estimator, but that is beyond the scope of this paper.

One might ask how large the interval widths should be when $\lambda'(t) = 0$. In this case, one can use extra terms in the Taylor expansion used to derive the bias estimate, and the optimal choice of δ_n is then larger than $n^{-1/3}$.

Acknowledgments

The author would like to thank the referees and the associate editor for their careful reading and extremely helpful comments. This work was supported by National Science Foundation Grant No. DMI–0230528 (previously numbered DMI–9984717).

References

[1] B. L. Arkin and L. M. Leemis. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process from overlapping realizations. *Management Science*, 46(7):989–998, 2000.

[2] D. Freedman and P. Diaconis. On the histogram as a density estimator: L_2 theory. *Z. Wahr. Gebiete*, 57:453–476, 1981.

[3] E. P. C. Kao and S. L. Chang. Modeling time-dependent arrivals to service systems - a case in using a piecewise-polynomial rate-function in a nonhomogeneous Poisson-process. *Management Science*, 34(11):1367–1379, 1988.

[4] M. E. Kuhl and P. S. Bhairgond. Nonparametric estimation of nonhomogeneous Poisson processes using wavelets. In J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, editors, *Proceedings of the 2000 Winter Simulation Conference*, pages 562–571, Piscataway NJ, 2000. IEEE.

[5] M. E. Kuhl and J. R. Wilson. Least squares estimation of nonhomogeneous Poisson processes. *Journal of Statistical Computation and Simulation*, 67(1):75–108, 2000.

[6] M. E. Kuhl and J. R. Wilson. Modeling and simulating Poisson processes having trends or nontrigonometric cyclic effects. *European Journal of Operational Research*, 133(3):566–582, 2001.

[7] M. E. Kuhl, J. R. Wilson, and M. A. Johnson. Estimating and simulating Poisson processes having trends or multiple periodicities. *IIE Transactions*, 29(3):201–211, 1997.

[8] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, 3rd edition, 2000.

[9] S. H. Lee, J. R. Wilson, and M. M. Crawford. Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. *Communications in Statistics: Simulation and Computation*, 20:777–809, 1991.

[10] L. M. Leemis. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process. *Management Science*, 37(7):886–900, 1991.

[11] P. A. W. Lewis and G. S. Shedler. Statistical analysis of nonstationary series of events in a database system. *IBM Journal of Research and Development*, 20(5):465–482, 1976.

[12] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics*, 26(3):403–413, 1979.

[13] S. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.

[14] S. M. Ross. *Stochastic Processes*. Wiley, New York, 2nd edition, 1996.

[15] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman & Hall, London, 1995.