Nonparametric Estimation for a Nonhomogeneous Poisson Process

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

March 7, 2002

Abstract

In the setting of discrete-event simulation, 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 (known) intervals. We investigate the asymptotic (as the amount of data grows) behaviour of this estimator, 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 [16]), but not the most general possible (see, e.g., Resnick [15]).

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,*

*Work supported in part by National Science Foundation Grant No. DMI–9984717. Email: shane@orie.cornell.edu, www:http://www.orie.cornell.edu/∼shane

2. for all $t \geq 0$, $P(N(t+h) - N(t) >= 2) = o(h)$, *and*

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.) Associated with a NHPP is the cumulative rate (or cumulative intensity) function* $\Lambda = (\Lambda(t) : t \geq 0)$, *where* $\Lambda(t) = \int_0^t \lambda(s)\, ds$.

In this paper, we consider the problem of estimating the rate function λ, and the cumulative rate function Λ, over a finite interval $[0, T]$, for the purpose of generating realizations in a discrete-event simulation.

Leemis [12] gives a nonparametric estimator of Λ when several independent realizations of N are available 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 realization may not necessarily cover the full interval $[0, T]$. A different nonparametric estimator of the rate function was suggested in Lewis and Shedler [13] based on kernel estimation techniques, again assuming the availability of several independent realizations of N over $[0, T]$. 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. Kuhl and Bhairgond [6] give a nonparametric estimator using wavelets.

In addition to a nonparametric estimator, Lewis and Shedler [13] also propose a parametric estimator of the rate function. Other parametric estimators are developed in, for example, Lee et al. [11], Kuhl et al. [9], Kuhl and Wilson [7], and Kao and Chang [5]. Kuhl and Wilson [8] construct a hybrid parametric/nonparametric estimator.

The nonparametric estimators of the cumulative rate function developed in 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, realizations can be rapidly computed 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 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 [10] 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. However the estimator has several desirable properties.

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 increase with the number of observed realizations. 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. Realizations can be easily generated from the fitted rate function, although perhaps not quite as rapidly as in the case of the estimators developed in [12] and [1].

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 on individual event times. If one allows δ to decrease as a function of n, the number of observed realizations of N, then one can view the Law and Kelton estimator as a nonparametric estimator of either the rate function λ or the cumulative rate function Λ. By choosing $\delta = \delta_n$ appropriately, one obtains a consistent estimator of the rate function, without the need to store all event times in memory. We explore this estimator's asymptotic (as $n \to \infty$) properties, and compare it with the Leemis estimator.

2 Estimating the Cumulative Rate Function

Consider the problem of estimating the cumulative rate function Λ over the interval $[0, T]$. The cumulative rate function can be used to generate realizations of the NHPP through an inversion procedure; see, for example, page 486 of [10]. The inversion method is preferred to the thinning method introduced by Lewis and Shedler [14] if one wishes to employ variance reduction techniques that rely on monotonicity properties of the generated process as a function of the uniform random variables that serve as inputs to the procedure.

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

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

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

2. $n^{1/2}(\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t)) \Rightarrow \sigma\mathcal{N}(0,1)$ *as* $n \to \infty$*, where*

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

3. *For all* $n \geq 1$*,* $E\tilde{\Lambda}_n(t) = \tilde{\Lambda}(t)$*. If* λ *is continuously differentiable in a neighbourhood of* t*, then*

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

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

Proof:

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

$$\begin{aligned} |\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t)| &= \max[\tilde{\Lambda}_n(t) - \tilde{\Lambda}(t), \tilde{\Lambda}(t) - \tilde{\Lambda}_n(t)] \\ &\leq \max[\tilde{\Lambda}_n(u_{a(t)+1}) - \tilde{\Lambda}(u_{a(t)}), \tilde{\Lambda}(u_{a(t)+1}) - \tilde{\Lambda}_n(u_{a(t)})] \end{aligned}$$

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

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

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

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

The central limit theorem follows from the fact that for all $0 \le a < b$,

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

and the independent increments property of N. The fact that $E\tilde{\Lambda}_n(t) = \tilde{\Lambda}(t)$ is immediate from its definition. Finally,

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

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

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

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 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 continuously differentiable in a neighbourhood of t, then*

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

as $n \to \infty$.

Remark 1 *The hypotheses of the central limit theorem in Proposition 2 can be replaced by the assumption that λ is continuous in a neighbourhood of t, and $\delta_n = o(n^{-1/2})$. This is a weaker condition on λ but a stronger condition on δ_n.*

Proof:

The uniform strong law will follow 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 $n^{-1}\sum_{i=1}^n N_i(0,t)$.

Turning to the central limit theorem, note 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 $\xi_n \in [\ell_n(t),\ell_n(t)+\delta_n)$.

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

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

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

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

The first term in (4) is bounded in absolute value by $|u|n^{1/2}|\lambda'(\zeta_n)|\delta_n^2$ for some ζ_n lying between ξ_n and θ_n (and hence lying in the interval $[\ell_n(t),\ell_n(t)+\delta_n)$). 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. ■

The assumption that $\delta_n = o(n^{-1/4})$ is the "best possible" assumption in the following heuristic sense. The estimator $\Lambda_n(t)$ has expectation $\bar{\Lambda}_n(t)$ say, where $\bar{\Lambda}_n$ is given as in (2) with subinterval width $\delta=\delta_n$. From Proposition 1, $\bar{\Lambda}_n(t)-\Lambda(t)$ is typically of the order δ_n^2. Hence, if $n^{1/2}\delta_n^2$ does not converge to 0, then the bias in the estimator $\Lambda_n(t)$ does not converge to 0 as $n\to\infty$. Of course, it is possible for an estimator to be consistent and satisfy a central limit theorem while having non-vanishing bias, but under certain regularity conditions this is impossible. For example, if $\sqrt{n}(\Lambda_n(t)-\Lambda(t)) \Rightarrow \mathcal{N}(0,\Lambda(t))$, then

$$n(\Lambda_n(t)-\Lambda(t))^2 \Rightarrow \Lambda(t)\mathcal{N}(0,1)^2 \tag{5}$$

as $n \to \infty$. If the sequence of random variables on the left-hand side of (5) is uniformly integrable (in n), then by taking expectations through (5) we can conclude that the mean squared error in $\Lambda_n(t)$ converges to 0 at rate n^{-1}. But this implies that the bias in $\Lambda_n(t)$ should also converge to 0.

The two estimators $\tilde{\Lambda}_n$ and Λ_n developed in this section allow one to use an inversion procedure for generating realizations of the process. This generation process can be slow owing to the fact that the estimated cumulative rate function cannot be inverted in constant time without employing further "preprocessing"; see page 180 of Bratley et al. [2] for related comments. This stands in contrast to the estimator introduced in [12], which can be inverted in constant time. Partly for this reason, we next consider an alternative approach.

3 Estimating the Rate Function

Consider the problem of estimating the rate function λ over the interval $[0, T]$. This rate function can be used to generate realizations of the NHPP through a thinning procedure introduced in Lewis and Shedler [14]. The nature of the rate function estimator is such that thinning gives a fast generation procedure.

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 a 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, and so $\lambda(T^*)$ can be rapidly computed. This observation is analogous to one in [12] 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, let

$$\tilde{\lambda}_n(t) = \frac{1}{n\delta} \sum_{i=1}^{n} N_i(\ell(t), \ell(t) + \delta) \tag{6}$$

denote our estimator of the rate function λ. Observe that $\tilde{\lambda}$ and $\tilde{\lambda}_n$ are the right-hand derivatives of $\tilde{\Lambda}$ and $\tilde{\Lambda}_n$ respectively.

Proposition 3 mirrors Proposition 1 in that it describes the large-sample behaviour of λ_n for a fixed interval width δ. The proof is elementary and omitted.

Proposition 3

1. *We have that $\tilde{\lambda}_n(t) \to \tilde{\lambda}(t)$ 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 in a neighbourhood of t, then $E\tilde{\lambda}_n(t) = \lambda(\xi)$ for some ξ lying in the interval $[\ell(t), \ell(t)+\delta)$.*

We see that the bias in $\tilde{\lambda}_n$ depends on the local behaviour of $\lambda(\cdot)$ in a neighbourhood of t, as will become further evident below.

Now consider the case where the subinterval length $\delta = \delta_n$ depends on n, the number of realizations of the process N. Let $\lambda_n(t)$ be defined as in (6) 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 $n \to \infty$.*

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

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

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

   as $n \to \infty$.

3. *If λ is continuously differentiable in a neighbourhood of t, and $\lambda'(t) \neq 0$, then a bound on the mean squared error of $\lambda_n(t)$ is minimized by taking*

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

In this case where there are constants $a, b > 0$ with $an^{-1/3} < \delta_n < bn^{-1/3}$ for all n, the set of random variables $\{(n\delta_n)^{1/2}(\lambda_n(t) - \lambda(t)) : n \geq 1\}$ is tight.

Proof:

It is straightforward to establish that $\lambda_n(t)$ converges weakly to $\lambda(t)$ as $n \to \infty$. The strong law requires more effort. For $i \geq 1$ and $\theta > 0$, define

$$X(i,\theta) = \frac{N_i(\ell^\theta(t), \ell^\theta(t) + \theta)}{\theta},$$

where $\ell^\theta(t) = \lfloor t/\theta \rfloor \theta$. Then

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

From Proposition 3, for n sufficiently large there exists $\xi = \xi^\theta \in [\ell^\theta(t), \ell^\theta(t) + \theta)$ such that $EX(i,\theta) = \lambda(\xi^\theta)$. Define $X^c(i,\theta) = X(i,\theta) - EX(i,\theta)$ to be the centered $X(i,\theta)$s. Then $E|X^c(i,\theta)| \leq 2\lambda(\xi^\theta)$, and so for θ sufficiently small and positive, say $\theta \in A$, the continuity of λ in a neighbourhood of t establishes that expectations of $|X^c(i,\theta)|$ are uniformly bounded.

The uniform strong law established in Chung [3] then allows us to conclude that for all $\epsilon > 0$, there is an $n_1(\epsilon)$ not depending on θ such that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X^c(i,\theta)\right| \leq \epsilon \quad \forall n \geq n_1(\epsilon)\right) \geq 1 - \epsilon \tag{7}$$

for all $\theta \in A$. Since this holds for all $\theta \in A$, and $\delta_n \to 0$, it follows that (7) holds for $\theta = \theta_n = \delta_n$. But if $\theta = \delta_n$, then (7) is equivalent to stating that

$$P(|\lambda_n(t) - \lambda(\xi_n)| \leq \epsilon \quad \forall n \geq n_1(\epsilon)) \geq 1 - \epsilon,$$

where $\xi_n = \xi^{\delta_n}$.

Now, for $n \geq n_2(\epsilon)$ say, $|\lambda(\xi_n) - \lambda(t)| \leq \epsilon$. So if $n^*(\epsilon) = \max\{n_1(\epsilon), n_2(\epsilon)\}$, then

$$\begin{aligned} P(|\lambda_n(t) - \lambda(t)| \leq 2\epsilon \quad \forall n \geq n^*(\epsilon)) &\geq P(|\lambda_n(t) - \lambda(\xi_n)| \leq \epsilon \quad \forall n \geq n^*(\epsilon)) \\ &\geq 1 - \epsilon. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, it follows that $|\lambda_n(t) - \lambda(t)| \to 0$ almost surely as $n \to \infty$.

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.

The bias of $\lambda_n(t)$ is given by

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

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

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

$$\frac{\lambda(\xi_n)}{n\delta_n} + \lambda'(\theta_n)^2\delta_n^2, \tag{8}$$

which is minimized at

$$\begin{aligned} \delta_n &= \left(\frac{\lambda(\xi_n)}{2\lambda'(\theta_n)^2}\right)^{1/3} n^{-1/3} \\ &\sim \left(\frac{\lambda(t)}{2\lambda'(t)^2}\right)^{1/3} n^{-1/3}. \end{aligned} \tag{9}$$

To prove tightness, note that if $an^{-1/3} < \delta_n < bn^{-1/3}$, then from (8),

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

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

Thus, the optimal choice of δ_n from a mean squared error perspective is of the order $n^{-1/3}$. This choice of δ_n ensures that $\lambda_n(t)$ converges at rate $(n\delta_n)^{-1/2} = n^{-1/3}$ (through the tightness result above). A slower rate of convergence than $n^{-1/2}$ is representative of many nonparametric function estimators; see Wand and Jones [17] 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 [4].

Notice also that $\lambda'(t)^2$ appears in the denominator in (9), showing (as intuition would suggest) that when λ is changing rapidly one prefers small intervals, while larger intervals are better when λ is not changing rapidly. 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 $n^{-1/3}$.

4 Conclusions

The estimator of the rate function given in Section 3 is the right-hand derivative of the estimator of the cumulative rate function given in Section 2. The choice between rate function and cumulative rate function estimation depends partly on the method to be used for generating realizations of N.

If one wishes to use an inversion-type procedure, then one might use Λ_n. The bias involved in using intervals of fixed width δ is of the order δ^2. When $\delta = \delta_n$ varies with the number of realizations n, the memory requirements (for generation) are less when δ_n is large, but the mean squared error of Λ_n is minimized when δ_n is as small as possible. Proposition 2 shows that with an appropriate choice of δ_n, Λ_n converges at the canonical rate $n^{-1/2}$, which is the same rate as the Leemis estimator. In fact, the asymptotic variance constant is also the same, so that to second order, this estimator is comparable with the Leemis estimator. It has the advantage that it requires less storage, but generation may take slightly longer.

The rate function estimator in Section 3 can be used in a thinning procedure to generate realizations, and this should be quite fast because of the equal interval widths. In this setting, a small interval width leads to lower bias and higher variability than a larger interval width. The optimal interval width in terms of mean squared error is of the order $n^{-1/3}$, where n is the number of observed realizations of the Poisson process, and this leads to a rate of convergence of $n^{-1/3}$.

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] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*. Springer-Verlag, New York, 2nd edition, 1987.

[3] K. L. Chung. The strong law of large numbers. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 341–352, Berkeley and Los Angeles, California, 1951. University of California Press.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

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

[13] 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.

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

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

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

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