

EXPERIMENTAL EVALUATION OF A PROCEDURE FOR ESTIMATING NONHOMOGENEOUS POISSON PROCESSES HAVING CYCLIC BEHAVIOR

Mary A. Johnson

Department of Mechanical and Industrial Engineering
University of Illinois-Urbana/Champaign
Urbana, Illinois 61801

Sanghoon Lee

Department of Industrial Engineering
Kyung Won University
Seongnam, KOREA

James R. Wilson

Department of Industrial Engineering
North Carolina State University
Raleigh, North Carolina 27695

ABSTRACT

This paper summarizes an experimental performance evaluation of a procedure for estimating a nonhomogeneous Poisson process having an exponential rate function whose exponent includes both polynomial and trigonometric components. Maximum likelihood estimates of the unknown continuous parameters of the rate function are obtained numerically, and the degree of the polynomial rate component is determined by a likelihood ratio test. Although this procedure can estimate the oscillation frequency of the trigonometric rate component, the experimental evaluation of this procedure is limited to the case of a known frequency. A "piecewise thinning" algorithm is used to generate independent replications of selected arrival processes having the postulated type of rate function. To evaluate the accuracy of the resulting estimators of the true rate and mean-value functions, we tabulate summary statistics for the maximum absolute error and average absolute error observed in estimating each of these functions on each replication of the estimation procedure. We also plot tolerance intervals for the rate and mean-value functions based on the sample estimates of these functions.

1 INTRODUCTION

In this paper we present a Monte Carlo performance evaluation of procedures recently developed by Lee, Wilson, and Crawford (1991) for identification, estimation, and simulation of a nonhomogeneous Poisson process (NHPP) whose rate function may contain

either a cyclic component, a long-term evolutionary trend, or both. Specifically, the instantaneous arrival rate of the NHPP is assumed to be an exponential function of time whose exponent is the sum of polynomial and trigonometric components—that is, an exponential-polynomial-trigonometric function (EPTF). Such a stochastic model is sufficiently general to handle a broad range of input modeling situations that have been previously reported in the literature. Lewis (1970) used spectral methods to analyze an NHPP with an exponential-trigonometric rate function. To model the arrival of patients at an intensive-care unit, Lewis (1972) used an NHPP with an EPTF-type rate function that included a quadratic trend. Lewis and Shedler (1976) modeled transaction-initiation times in a data base system as an NHPP with an exponential-polynomial rate function. To represent the arrival pattern for storms at an off-shore drilling site, Lee (1985) used an NHPP with an EPTF-type rate function.

The estimation procedure of Lee, Wilson, and Crawford (1991) can be applied to an NHPP with an EPTF-type rate function whether the oscillation frequency of the cyclic rate component is known or unknown. Given sample data from such a process, we compute maximum likelihood estimates of the unknown continuous parameters of the rate function using a Newton-Raphson scheme; and the degree of the polynomial component is determined by a likelihood ratio test. The success of this approach depends critically on the initial estimates the unknown continuous parameters being reasonably close to the true maximum of the associated likelihood function. If the oscillation frequency is unknown,

then an initial estimate of this parameter is obtained via spectral analysis of the observed series of events; and initial estimates of the remaining trigonometric (respectively, polynomial) parameters are computed from a standard maximum likelihood (respectively, moment-matching) procedure for an exponential-trigonometric (respectively, exponential-polynomial) rate function. Although Lee, Wilson, and Crawford (1991) reported a successful application of this estimation procedure to the storm-arrival data originally studied by Lee (1985), they did attempt an analytical or empirical evaluation of the performance of this procedure in repeated applications. In this paper we report some initial results of a large-scale, simulation-based performance evaluation of this procedure.

In our simulation experiments, we used the “piecewise thinning” algorithm of Lee, Wilson, and Crawford (1991) to generate independent replications of an NHPP with an EPTF-type rate function. The initial step in this procedure is to approximate the target rate function as closely as possible using a piecewise linear majorizing function—that is, a piecewise linear function which provides a tight upper bound for the target rate function. At each event epoch of the target process, a series of events is generated from the majorizing function by the method of inversion; these events are then screened by a thinning (rejection) procedure so that the next interevent time for the target process is finalized when the first acceptable event is generated for the majorizing process. With this approach, the computation time can be greatly reduced in simulating a process with a complicated cyclic rate function such as an EPTF.

This paper is organized as follows. In Section 2 we summarize the procedure of Lee, Wilson, and Crawford (1991) for estimating the parameters of an EPTF-type rate function. Section 3 describes the simulation experiments used to evaluate the performance of this estimation procedure, including a brief informal description of the “piecewise thinning” algorithm. In Section 4 we summarize and discuss our experimental results. Finally in Section 5 we present our conclusions and recommendations for future work. The Appendix contains some computing formulas that are used in the estimation procedure.

2 IDENTIFICATION AND ESTIMATION OF NHPPs

An NHPP $\{N(t) : t \geq 0\}$ is a generalization of a Poisson process in which the instantaneous arrival rate $\lambda(t)$ at time t is a nonnegative integrable function of time. The mean-value function (or the integrated

rate function) of the NHPP is defined by

$$\mu(t) \equiv E[N(t)] = \int_0^t \lambda(z) dz \text{ for all } t \geq 0.$$

In this study, an NHPP displaying cyclic behavior is assumed to have an EPTF-type rate function. An EPTF of degree m has the form

$$\lambda(t) = \exp\{h_\Theta(m, t)\} \quad \text{with} \quad (1)$$

$$h_\Theta(m, t) = \sum_{i=0}^m \alpha_i t^i + \gamma \sin(\omega t + \phi),$$

where: $\Theta = [\alpha_0, \alpha_1, \dots, \alpha_m, \gamma, \phi, \omega]$ is the vector of unknown parameters; the first term in $h_\Theta(m, t)$ is an ordinary polynomial function representing the general trend over time; and the second term in $h_\Theta(m, t)$ is a periodic function representing a cyclic effect exhibited by the process. In this section we summarize the method of Lee, Wilson, and Crawford (1991) for identifying an appropriate value of m and for estimating the corresponding parameter vector Θ .

The analysis of an NHPP with a rate function of the form (1) is substantially more difficult when the oscillation frequency ω (expressed in radians per unit time) is unknown. Although ω is often known from prior information about the mechanism generating the events of interest, there is a large class of simulation applications for which such prior information is unavailable so that ω must be estimated from sample data. To develop a completely general technique for modeling and simulation of an NHPP with an EPTF-type rate function, Lee, Wilson, and Crawford (1991) assumed that the oscillation frequency is unknown and must be estimated along with all of the other parameters of the rate function. If ω is known, then we simply drop the last component of Θ before applying the parameter estimation technique described below.

Consider a sequence of n events occurring at the epochs $t_1 < t_2 < \dots < t_n$ in a fixed time interval $(0, S]$ according to an NHPP with a rate function of the form (1). Then the log-likelihood function of Θ , given $N(S) = n$ and $\mathbf{t} = (t_1, t_2, \dots, t_n)$, is

$$\mathcal{L}(\Theta | n, \mathbf{t}) = \sum_{i=0}^m \alpha_i T_i \quad (2)$$

$$+ \gamma \sum_{j=1}^n \sin(\omega t_j + \phi) - \int_0^S \exp\{h_\Theta(m, z)\} dz,$$

where $T_i = \sum_{j=1}^n t_j^i$ for $i = 0, 1, \dots, m$; see Cox and Lewis (1966). Strictly speaking, we observe that the degree m of the polynomial component of $h_\Theta(m, t)$ is also an unknown parameter which could in prin-

ciple be estimated along with Θ from the given sequence of events by the method of maximum likelihood. However, since m is constrained to be a non-negative integer, we cannot estimate m by solving the usual likelihood equations that are applicable to continuous parameters; moreover the usual regularity conditions ensuring the asymptotic efficiency of maximum likelihood estimators do not apply to the estimation of m .

In view of the fundamental problems inherent in maximum likelihood estimation of the degree m of the polynomial rate component, Lee, Wilson, and Crawford (1991) chose to condition the estimation of Θ on a fixed value of m and then to determine the appropriate value of m by a likelihood ratio test to be described later (see equation (14) and the accompanying discussion given below). Thus for a given value of m where $m \geq 0$, we obtain $m + 4$ likelihood equations involving the continuous parameter vector Θ

$$\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \alpha_i} = T_i \tag{3}$$

$$- \int_0^S z^i \exp\{h_\Theta(m, z)\} dz = 0, \quad i = 0, 1, \dots, m,$$

$$\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \omega} = \sum_{j=1}^n t_j \cos(\omega t_j + \phi) \tag{4}$$

$$- \int_0^S z \cdot \cos(\omega z + \phi) \exp\{h_\Theta(m, z)\} dz = 0$$

$$\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \gamma} = \sum_{j=1}^n \sin(\omega t_j + \phi) \tag{5}$$

$$- \int_0^S \sin(\omega z + \phi) \exp\{h_\Theta(m, z)\} dz = 0,$$

$$\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \phi} = \sum_{j=1}^n \cos(\omega t_j + \phi) \tag{6}$$

$$- \int_0^S \cos(\omega z + \phi) \exp\{h_\Theta(m, z)\} dz = 0.$$

The solution to the system (3)–(6) of nonlinear equations can be obtained numerically, yielding the maximum likelihood estimates of the parameters. Unfortunately, general numerical techniques such as the Newton-Raphson method have proven to be unstable when they are applied to (3)–(6) outside of a fairly small neighborhood of the optimal solution.

2.1 Initial Parameter Estimates

If the oscillation frequency ω is unknown, then it is important to choose the initial value of ω sufficiently close to the true maximum likelihood estimate because the log-likelihood function (2) has multiple local maxima due to the trigonometric rate component. In this situation, an initial estimate of ω can be obtained from a preliminary spectral analysis of the observed series of events; see pp. 361–363 of Lewis (1970). The periodogram of a point process having cyclic behavior should display peaks in the vicinity of the corresponding frequency, even when the process also possesses a long-term evolutionary trend. Let ω_0 denote either the known frequency or an initial estimate of the unknown frequency inferred from the periodogram of the observed series of event times $\{t_j : j = 1, 2, \dots, n\}$.

We obtain the corresponding initial estimates of the amplitude γ and the phase ϕ as follows. If we assume that the long-term evolutionary trend is nearly constant over the observation interval and that ω_0 is the actual frequency of the cyclic component, then the rate function for the process has the exponential-trigonometric form

$$\lambda_0(t) = \exp\{\alpha + \gamma \sin(\omega_0 t + \phi)\} \tag{7}$$

for all $t \in (0, S]$. Under the simplifying assumption (7), the initial estimates γ_0 and ϕ_0 (for γ and ϕ respectively) are obtained from

$$\phi_0 = \tan^{-1} \left[\frac{A(\omega_0)}{B(\omega_0)} \right], \tag{8}$$

γ_0 is the solution of

$$\frac{1}{n_0} \sqrt{A^2(\omega_0) + B^2(\omega_0)} = \frac{I_1(\gamma_0)}{I_0(\gamma_0)}. \tag{9}$$

In (8) and (9), we define n_0 to be the number of events in the time interval

$$\left(0, \left\lfloor \frac{\omega_0 S}{2\pi} \right\rfloor \frac{2\pi}{\omega_0} \right] \tag{10}$$

and $\lfloor z \rfloor$ represents the greatest integer $\leq z$ for all real z ; moreover

$$A(\omega_0) \equiv \sum_{j=1}^{n_0} \cos(\omega_0 t_j), \quad B(\omega_0) \equiv \sum_{j=1}^{n_0} \sin(\omega_0 t_j), \tag{11}$$

and $I_u(\cdot)$ denotes a modified Bessel function of the first kind of order u for $u = 0, 1$. (Note that the time interval (9) corresponding to the event count n_0 is generally taken to be somewhat shorter than the

original observation interval $(0, S]$ to ensure that ω_0 is one of the frequencies included in the spectral computations.) For most cases in which the evolutionary trend changes slowly over time, the estimates γ_0 and ϕ_0 provided by (8)–(11) will be good initial values of the corresponding trigonometric parameters. Equations (46) and (48) of Lewis (1970) provide the basis for this approach to the computation of γ_0 and ϕ_0 .

To determine initial values for the coefficients $\{\alpha_k : k = 0, 1, \dots, m\}$ of the polynomial rate component, Lee, Wilson, and Crawford (1991) used a variant of MacLean’s (1974) procedure for estimating the parameters of an exponential-polynomial rate function by moment matching. Suppose that $C(m, t) = \sum_{k=0}^m c_k t^k$ is an ordinary polynomial function of degree m whose first $m + 1$ moments over the interval $(0, S]$ match those of $\exp\{h_\Theta(m, t)\}$. Then by equation (3), we have

$$\begin{aligned} T_i &= \int_0^S z^i \exp\{h_\Theta(m, z)\} dz \\ &= \int_0^S z^i C(m, z) dz = \sum_{k=0}^m \frac{c_k S^{i+k+1}}{i+k+1} \end{aligned} \quad (12)$$

for $i = 0, 1, \dots, m$. The values of the $\{c_k\}$ can be obtained from this linear equation system by matrix inversion. Next, $h_\Theta(m, t)$ can be approximated by matching its first $m + 1$ moments over the interval $(0, S]$ with the corresponding moments of $\log[C(m, t)]$. Thus for $i = 0, 1, \dots, m$, we have

$$\begin{aligned} \int_0^S z^i \log\left(\sum_{k=0}^m c_k z^k\right) dz - \gamma \mathcal{M}_{\sin}(i, S; \omega, \phi) \\ = \sum_{k=0}^m \frac{\alpha_k S^{i+k+1}}{i+k+1} \end{aligned} \quad (13)$$

where $\mathcal{M}_{\sin}(i, S; \omega, \phi) \equiv \int_0^S z^i \sin(\omega z + \phi) dz$, the i th moment of $\sin(\omega t + \phi)$ over the interval $(0, S]$. Using the initial values of γ , ω , and ϕ based on (7)–(11), we can evaluate the second term on the left-hand side of equation (13) from the computational formulas for $\mathcal{M}_{\sin}(i, S; \omega, \phi)$ given in the Appendix; and by using numerical integration, we can also evaluate the first term on the left-hand side of (13). A single matrix inversion for equation system (13) yields initial estimates of the $\{\alpha_k\}$. This elaborate procedure for assembling an initial estimate of the complete parameter vector Θ is designed to ensure that the $(m + 4)$ -dimensional Newton-Raphson scheme for solving the full likelihood equation system (3)–(6) will start reasonably close to the true maximum of the log-likelihood function (2).

2.2 Final Parameter Estimates

Corresponding to each trial value of the degree m for the fitted EPTF-type rate function, the procedure described above is used to obtain an initial estimate of the associated parameter vector Θ_m ; then the log-likelihood function $\mathcal{L}_m(\Theta_m|n, \mathbf{t})$ is optimized by the Newton-Raphson procedure to yield the maximum likelihood estimator $\tilde{\Theta}_m$. (Here the subscript “ m ” is used to emphasize the dependence of the subscripted quantities on the degree of the fitted EPTF.) To determine the appropriate value of m , Lee, Wilson, and Crawford (1991) used a likelihood ratio test (Cox and Hinkley 1974) as follows. Under the null hypothesis that m is the true degree of the underlying EPTF-type rate function, the difference

$$2\left[\mathcal{L}_{m+1}(\tilde{\Theta}_{m+1}|n, \mathbf{t}) - \mathcal{L}_m(\tilde{\Theta}_m|n, \mathbf{t})\right] \quad (14)$$

is asymptotically chi-square with 1 degree of freedom as $n \rightarrow \infty$. Successive differences of this form are evaluated, and the smallest value of m yielding a non-significant difference (14) is taken as the degree of the fitted EPTF. The corresponding $\tilde{\Theta}_m$ is taken as the final estimator of the parameters of the underlying NHPP.

2.3 Model Adequacy Checking

To perform diagnostic checking of the adequacy of the fitted NHPP, Lee, Wilson, and Crawford (1991) recommended informal visual assessment of the uniformity of the observed point process after it has been “detrended” using the final estimate of the mean-value function. Consider the observed series of event epochs $t_1 < t_2 < \dots < t_n$ for the original process in the interval $(0, S]$. Let $\tilde{\mu}(t)$ denote the mean-value function of the fitted NHPP corresponding to the final estimate $\tilde{\Theta}_m$. If the original process $\{N(t) : t \in (0, S]\}$ is in fact an NHPP whose mean value function $\mu(t)$ has been accurately estimated by $\tilde{\mu}(t)$, then the detrended process $\{\tilde{M}(s) \equiv N[\tilde{\mu}^{-1}(s)] : s \in (0, \tilde{\mu}(S))\}$ with the event epochs $\tilde{s}_1 = \tilde{\mu}(t_1) < \tilde{s}_2 = \tilde{\mu}(t_2) < \dots < \tilde{s}_n = \tilde{\mu}(t_n)$ should closely approximate a homogeneous Poisson process with unit rate. Thus the detrended interevent times $\{\tilde{X}_i \equiv \tilde{s}_i - \tilde{s}_{i-1} : i = 1, \dots, n; \tilde{s}_0 \equiv 0\}$ should closely resemble a random sample from an exponential distribution with mean $1/\tilde{\lambda} = 1$; and this should be reflected in approximate linearity of the corresponding exponential probability plot (Hahn and Shapiro 1967, pp. 292–294). If $\{\tilde{X}_i\}$ is a random sample from an ex-

ponential distribution with mean $1/\tilde{\lambda}$ and if $\tilde{X}_{(k)}$ is the k th order statistic of the sample, then

$$E[\tilde{X}_{(k)}] = \frac{1}{\tilde{\lambda}} \sum_{i=0}^{k-1} \frac{1}{n-i} \text{ for } k = 1, \dots, n$$

(Feller, 1971; pp. 19–20); and in this case a plot of $\tilde{X}_{(k)}$ versus $\sum_{i=0}^k 1/(n-i)$ should be nearly linear.

As a final visual check on the adequacy of the fitted NHPP, Lee, Wilson, and Crawford (1991) also recommended comparing the plot of the fitted mean-value function $\tilde{\mu}(t)$ against the plot of the cumulative number of events $N(t)$ observed up to time t for all $t \in (0, S]$. Since these visual diagnostic checks are not easily automated and currently there are no formal statistical tests for uniformity of the detrended event epochs in which the null distribution of the test statistic is known, we did not attempt to perform any diagnostic checking of the adequacy of the fitted NHPP on each replication of the estimation procedure. However, the plots given in §4 allow visual evaluation of the adequacy of the fitted rate and mean-value functions over all 100 replications of the procedure.

3 EXPERIMENTAL PROTOCOL

3.1 Processes to be Simulated

To carry out an experimental performance evaluation for the estimation procedure detailed in §2, we simulated $K = 100$ independent replications of an NHPP with a third-degree EPTF

$$\lambda(t) = \exp\left[\sum_{i=0}^3 \alpha_i t^i + \gamma \sin(\omega t + \phi)\right] \quad (15)$$

for all $t \in (0, S]$, where the actual parameter values are displayed in Table 1. This is a variant of the process used by Lee, Wilson, and Crawford (1991) to model the arrival pattern for storms at an off-shore drilling site in the Arctic Sea. Note that the time unit is a year, and the oscillation frequency $\omega = 2\pi$ so there is exactly one cycle per year. To study both the small- and large-sample properties of the estimation procedure, we considered separately the cases in which $S = 4$ and $S = 12$.

3.2 Simulation Algorithm

To generate independent replications of the target NHPP having rate function (15), we used the thinning procedure of Lewis and Shedler (1979) with the

Table 1: Parameters of NHPP Used in the Experimental Evaluation

Parameter	Value
α_0	3.6269
α_1	-0.4743
α_2	0.0873
α_3	-0.0041
γ	1.0592
ϕ	-0.6193
ω	6.2831

piecewise linear majorizing function

$$\lambda^*(t) = \sum_{j=1}^r (a_j t + b_j) I_{(L_{j-1}, L_j]}(t) \quad (16)$$

for all $t \in (0, S]$, where the points $L_0 \equiv 0 < L_1 < L_2 < \dots < L_r \equiv S$ constitute a partition of $(0, S]$ and $I_{(L_{j-1}, L_j]}(t)$ is the indicator function for the j th subinterval $(L_{j-1}, L_j]$. Procedure `maxline()` of Lee, Wilson, and Crawford (1991) was used to determine the parameters $\{a_j, b_j, L_j\}$ of a function $\lambda^*(t)$ which majorizes the target rate function $\lambda(t)$ so that $\lambda^*(t) \geq \lambda(t)$ for all $t \in (0, S]$. Then the “piecewise thinning” procedure `nhpp()` of Lee, Wilson, and Crawford (1991) was used to simulate the target NHPP.

For completeness we provide a brief description of procedure `nhpp()`. To simplify the notation in this discussion, we will write the increment of the majorizing mean-value function over the subinterval $(t_1, t_2]$ as follows:

$$\mu^*(t_1, t_2) = \int_{t_1}^{t_2} \lambda^*(z) dz, \quad 0 \leq t_1 < t_2 \leq S. \quad (17)$$

For the majorizing process with the rate function (16) and mean-value function (17), the cumulative distribution function of the i th event time τ_i^* conditioned on the value of the previous event time τ_{i-1}^* can then be expressed as

$$F_{\tau_i^*}(t|\tau_{i-1}^*) = \Pr\{\tau_i^* \leq t|\tau_{i-1}^*\} \quad (18)$$

$$= \begin{cases} 0, & t < \tau_{i-1}^* \\ 1 - \exp\{-\mu^*(\tau_{i-1}^*, t)\}, & t \geq \tau_{i-1}^* \end{cases}$$

with $\tau_0^* \equiv 0$.

Given the $(i-1)$ st event time τ_{i-1}^* for the majorizing process, we generate the next event time τ_i^* for

this process via the method of inversion by sampling a random number U , solving for t in the equation $F_{\tau_i^*}(t|\tau_{i-1}^*) = U$, and setting $\tau_i^* = t$. This amounts to solving for τ_i^* in the equation

$$\mu^*(\tau_{i-1}^*, \tau_i^*) = -\log(1 - U). \tag{19}$$

Suppose that the $(i - 1)$ st event time τ_{i-1}^* falls in the j th subinterval $(L_{j-1}, L_j]$. We compute the i th event time τ_i^* from (16)–(19) using the sampled random number U as follows. If $-\log(1 - U) \leq \mu^*(\tau_{i-1}^*, L_j)$ so that both τ_{i-1}^* and τ_i^* occur in the j th subinterval, then we have

$$\tau_i^* = \frac{-b_j + \sqrt{(a_j \tau_{i-1}^* + b_j)^2 - 2a_j \log(1 - U)}}{a_j} \quad \text{if } a_j \neq 0; \tag{20}$$

and

$$\tau_i^* = \tau_{i-1}^* - \frac{\log(1 - U)}{b_j} \quad \text{if } a_j = 0. \tag{21}$$

On the other hand, if $-\log(1 - U) > \mu^*(\tau_{i-1}^*, L_j)$ so that τ_i^* occurs in the k th subinterval for some $k \geq j + 1$, then from (16)–(19), we have

$$\tau_i^* = \frac{-b_k + \sqrt{(a_k L_{k-1} + b_k)^2 - 2a_k [Q + \log(1 - U)]}}{a_k} \tag{22}$$

if $a_k \neq 0$, and

$$\tau_i^* = L_{k-1} - \frac{Q + \log(1 - U)}{b_k} \quad \text{if } a_k = 0. \tag{23}$$

As the final step on each iteration of the logic of procedure `nhpp()`, the next event epoch τ_i^* of the majorizing process is “thinned” with probability $1 - \lambda(\tau_i^*)/\lambda^*(\tau_i^*)$ to yield the next event epoch τ_ℓ of the target process. This means that another random number U' is generated and if

$$U' \leq \lambda(\tau_i^*)/\lambda^*(\tau_i^*), \tag{24}$$

then τ_i^* is delivered as the next event epoch τ_ℓ of the target NHPP. Otherwise, i is incremented by one and the steps described in equations (19)–(24) are repeated by sampling another pair of random numbers (U, U') . The preceding sequence of steps is reiterated until condition (24) is satisfied.

3.3 Performance Measures

To characterize the accuracy of the approximations to the target NHPP that were obtained by the estimation procedure detailed in §2, we used several numerical and graphical techniques. On the k th replication of the estimation procedure, let $\tilde{\lambda}_k(t)$ (respectively, $\tilde{\mu}_k(t)$) denote the estimated rate function (respectively, the estimated mean-value function). The *average absolute error* in the estimation of the rate function $\lambda(t)$ on the k th replication is

$$\delta_k \equiv \frac{1}{S} \int_0^S \left| \tilde{\lambda}_k(t) - \lambda(t) \right| dt,$$

and the *maximum absolute error* is

$$\delta_k^* \equiv \max \left\{ \left| \tilde{\lambda}_k(t) - \lambda(t) \right| : 0 \leq t \leq S \right\}$$

for $k = 1, \dots, K$. We computed similar performance measures with respect to estimation of the mean-value function $\mu(t)$:

$$\Delta_k \equiv \frac{1}{S} \int_0^S \left| \tilde{\mu}_k(t) - \mu(t) \right| dt$$

and

$$\Delta_k^* \equiv \max \left\{ \left| \tilde{\mu}_k(t) - \mu(t) \right| : 0 \leq t \leq S \right\}$$

for $k = 1, \dots, K$.

To provide a visual assessment of the quality of the estimates of the rate and mean-value functions, we also plotted tolerance intervals for these functions. For a fixed time $t \in (0, S]$, let

$$\tilde{\lambda}_{(1)}(t) < \tilde{\lambda}_{(2)}(t) < \dots < \tilde{\lambda}_{(K)}(t)$$

denote the ordered estimates of $\lambda(t)$ obtained on all $K = 100$ replications of the estimation procedure. Then we have the following approximate $100(1 - \beta)\%$ tolerance interval for the quantity $\lambda(t)$ when $t \in (0, S]$:

$$\left[\tilde{\lambda}_{(\lfloor K\beta/2 \rfloor + 1)}(t), \tilde{\lambda}_{(\lfloor K\{1-\beta/2\} \rfloor + 1)}(t) \right], \tag{25}$$

where $\lfloor z \rfloor$ denotes the greatest integer strictly less than z . Thus in particular when $K = 100$ and $\beta = 0.10$, the 90% tolerance interval for $\lambda(t)$ at a *single fixed* time $t \in (0, S]$ is $[\tilde{\lambda}_{(5)}(t), \tilde{\lambda}_{(95)}(t)]$. Similarly we obtained tolerance intervals for the mean-value function $\mu(t)$ at an arbitrary fixed time $t \in (0, S]$. By superimposing plots of true rate function and the

Table 2: Statistics Describing the Errors in Estimating $\lambda(t)$ and $\mu(t)$ for $t \in (0, S]$ when $S = 4$

Data Set	$\lambda(t)$		$\mu(t)$	
	$\{\delta_k\}$	$\{\delta_k^*\}$	$\{\Delta_k\}$	$\{\Delta_k^*\}$
Mean	5.03	20.58	6.85	11.32
Std Dev	1.58	9.19	4.40	5.68
Min	2.24	8.37	1.29	3.19
Max	8.96	57.26	18.79	25.51

corresponding the upper and lower tolerance limits in (25), we get an overview of the accuracy of the estimation procedure. Similar remarks apply to the mean value function.

The estimation procedure detailed in §2 and the simulation algorithm discussed in §3.2 have been implemented in portable FORTRAN 77 subprograms which are in the public domain and are available upon request. These subprograms make extensive use of NETLIB routines (Dongarra and Grosse 1987) and the portable random number generator UNIF() of Bratley, Fox, and Schrage (1987). For all of the experiments reported in the next section, we used a significance level of 10% in the likelihood ratio test (14) to determine the degree of the polynomial rate component.

4 EXPERIMENTAL RESULTS

First we consider the “small-sample” results for the case in which $S = 4$ so that the cyclic behavior in the target arrival process can only be observed for four cycles. On each replication of the estimation procedure, the number of observations n is Poisson distributed with mean $\mu(S) = 118.51$. Table 2 summarizes the results for this case.

Figures 1 and 2 display 90% tolerance intervals for the rate function and the mean value function in the case that $S = 4$. The results presented in Table 1 taken in conjunction with the tolerance intervals in Figures 1 and 2 provide some evidence of the ability of the estimation procedure to accurately estimate the NHPP having rate function (12) based on small to moderate sample sizes.

Next we consider the “large-sample” results for the case $S = 12$ so that the cyclic behavior of the target process can be observed for twelve cycles. Table 3 summarizes the results for this case. On each replication of the estimation procedure, the number of observations n is Poisson distributed with mean $\mu(S) = 405.5$. Although the sample mean of the

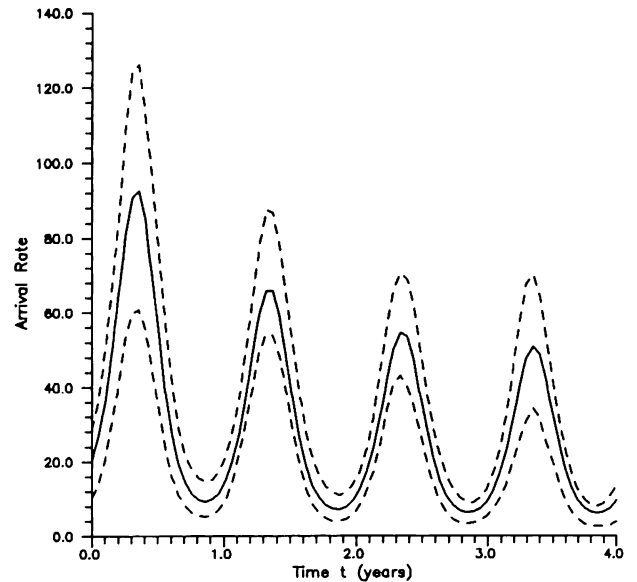


Figure 1: 90% Tolerance Intervals for $\lambda(t)$, $t \in (0, S]$, when $S = 4$

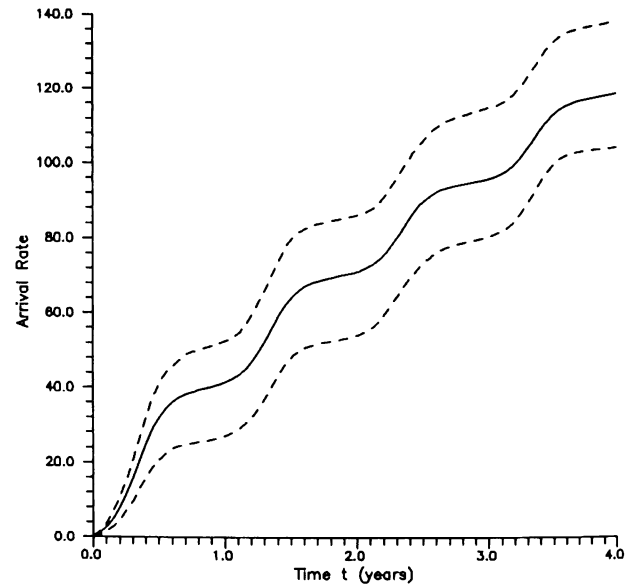


Figure 2: 90% Tolerance Intervals for $\mu(t)$, $t \in (0, S]$, when $S = 4$

Table 3: Statistics Describing the Errors in Estimating $\lambda(t)$ and $\mu(t)$ for $t \in (0, S]$ when $S = 12$

Data Set	$\lambda(t)$		$\mu(t)$	
	$\{\delta_k\}$	$\{\delta_k^*\}$	$\{\Delta_k\}$	$\{\Delta_k^*\}$
Mean	3.99	21.52	10.21	20.46
Std Dev	1.63	9.60	6.48	11.13
Min	1.05	4.78	1.74	4.47
Max	8.37	54.57	31.17	55.25

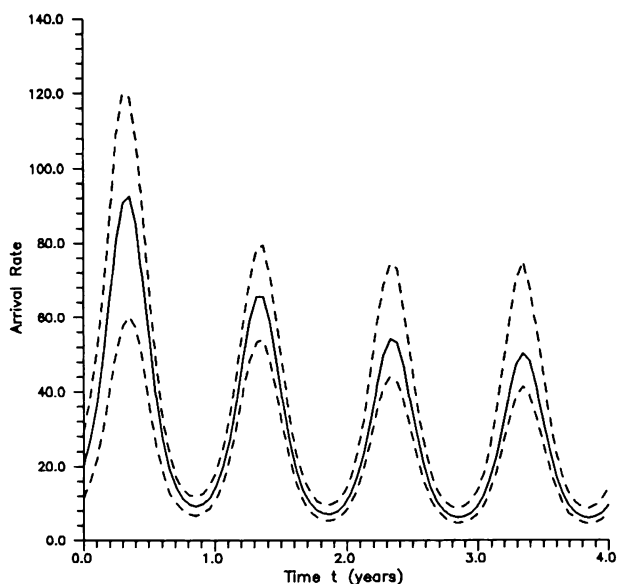


Figure 3: 90% Tolerance Intervals for $\lambda(t)$, $t \in (0, 4]$, when $S = 12$

maximum absolute estimation errors $\{\delta_k^*\}$ is roughly the same for the small-sample case as for the large-sample case, we also observe that the sample mean of the time-averaged absolute estimation errors $\{\delta_k\}$ declines by about 20% in the latter case. Figures 3–5 depict this phenomenon graphically—although the estimation errors remain relatively stable in magnitude within a small neighborhood of each local minimum and maximum of the underlying rate function, everywhere else the estimation errors decline in magnitude as the size of the data set increases.

5 CONCLUSIONS AND RECOMMENDATIONS

It is generally difficult to characterize periodic behavior in nonstationary point processes. In our ex-

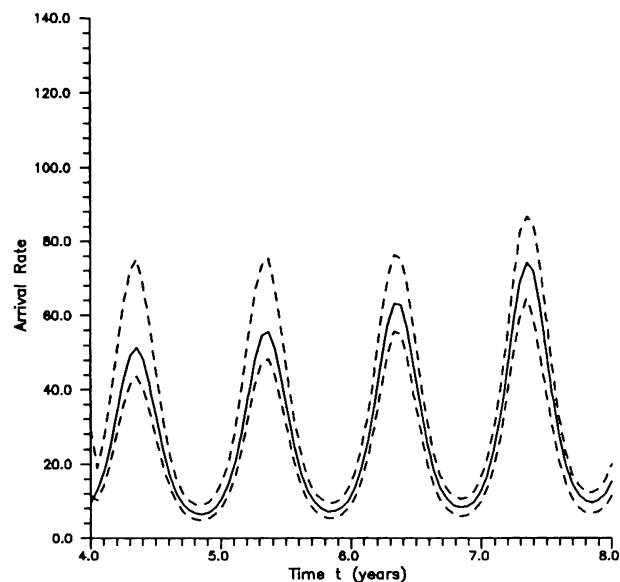


Figure 4: 90% Tolerance Intervals for $\lambda(t)$, $t \in (4, 8]$, when $S = 12$

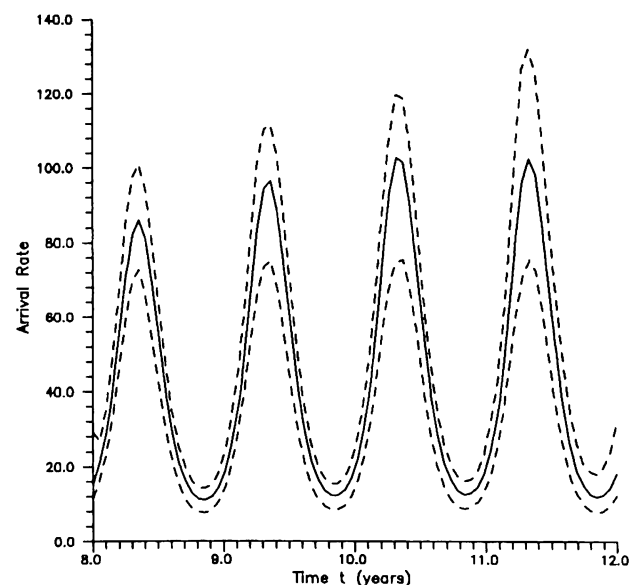


Figure 5: 90% Tolerance Intervals for $\lambda(t)$, $t \in (8, 12]$, when $S = 12$

perience, an NHPP whose rate function includes a long-term trend or a cyclic rate component can be adequately modeled with an EPTF-type rate function. Moreover, the parameters of an EPTF-type rate function can be estimated efficiently by (a) using a likelihood ratio test to identify the degree of the polynomial rate component; and (b) solving the usual likelihood equations to derive the maximum likelihood estimates of the continuous parameters of the rate function. The scheme for obtaining initial parameter estimates using equations (7)–(13) has been found to work reliably in practice. In our experience, the Newton-Raphson scheme for solving the likelihood equations (3)–(6) requires much more computing time than the scheme for determining the initial parameter estimates; and yet the Newton-Raphson scheme often yields relatively little additional improvement in the accuracy of the fit. Thus the scheme for determining initial parameter estimates may be used as a “quick” method for fitting a sample data set using an NHPP with an EPTF-type rate function.

An NHPP with an EPTF-type rate function can be generated exactly and efficiently by the piecewise thinning algorithm `nhpp()` of Lee, Wilson, and Crawford (1991). The supplemental procedure `maxline()` provides a piecewise linear majorizing function which is close enough to the original rate function to achieve high efficiency. To facilitate the use of the statistical-estimation and simulation procedures discussed in this paper, we have implemented all of these procedures in portable, public-domain subprograms written in FORTRAN 77.

The experimental performance evaluation reported in this paper is the first step in a more comprehensive Monte Carlo study of the estimation procedure of Lee, Wilson, and Crawford (1991). To support truly general conclusions about the performance of the estimation procedure, we must experiment with a much broader diversity of NHPPs whose rate functions exhibit different types of periodic behavior or evolutionary trends. We also need to examine situations in which the oscillation frequency is unknown and must be estimated. Finally, we need to test the estimation procedure using NHPPs with non-EPTF-type rate functions. On the basis of the preliminary results reported in this paper, there is some reason to believe that satisfactory performance can be expected in a wide range of input-modeling applications.

Several extensions of the estimation procedure should also be considered. Of particular interest is the case of an NHPP whose rate function exhibits multiple periodicities. Issues of numerical stability should also be investigated. Although we have developed special computation techniques to avoid numeri-

cal problems in the estimation procedure, it is unclear whether these techniques are absolutely necessary or whether these techniques can be made substantially faster.

ACKNOWLEDGMENTS

The authors thank Professor David Kelton of the University of Minnesota and Mr. Robert Klein of the Regenrief Institute for their comments and suggestions. This work was partially work supported by the National Science Foundation under Grant No. DMS-8717799.

APPENDIX

If $k = 0$, then we have

$$\mathcal{M}_{\sin}(0, S; \omega, \phi) = \frac{[\cos(\phi) - \cos(\omega S + \phi)]}{\omega}$$

If $k > 0$, then repeated integration by parts yields

$$\mathcal{M}_{\sin}(k, S; \omega, \phi) = \tag{A-1}$$

$$\begin{aligned} & \cos(\omega S + \phi) \sum_{r=0}^{\lfloor k/2 \rfloor} \frac{(-1)^{r+1} k! S^{k-2r}}{(k-2r)! \omega^{2r+1}} \\ & + \sin(\omega S + \phi) \sum_{r=0}^{\lfloor k-1/2 \rfloor} \frac{(-1)^r k! S^{k-2r-1}}{(k-2r-1)! \omega^{2r+2}} \\ & - (-1)^{\lfloor (k-1)/2 \rfloor} \frac{k!}{\omega^{k+1}} \cdot R_k, \end{aligned} \tag{A-2}$$

where $\lfloor z \rfloor$ represents the greatest integer $\leq z$ and

$$R_k = \begin{cases} \cos(\phi) & \text{if } k \text{ is even} \\ \sin(\phi) & \text{if } k \text{ is odd} \end{cases} \tag{A-3}$$

Note that (A-1) and (A-3) correct analogous formulas given in Lee, Wilson, and Crawford (1991).

REFERENCES

Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A guide to simulation*. 2d ed. New York: Springer-Verlag.

Cox, D. R. and D. V. Hinkley. 1974. *Theoretical statistics*. London: Chapman and Hall.

Cox, D. R. and P. A. W. Lewis. 1966. *The statistical analysis of series of events*. London: Chapman and Hall.

Dongarra, J. J. and E. Grosse. 1987. Distribution of mathematical software via electronic mail. *Communications of the ACM* 30:403–407.

- Feller, W. 1971. *An introduction to probability theory and its applications*, Volume II, Second Edition. New York: John Wiley & Sons.
- Hahn, G. J. and S. S. Shapiro. 1967. *Statistical models in engineering*. New York: John Wiley & Sons.
- Lee, S. 1985. Modeling and simulation of sea-state variables as nonhomogeneous Poisson processes and bilinear time series. Master's Thesis, Mechanical Engineering Department, The University of Texas, Austin, Texas.
- Lee, S., J. R. Wilson, and M. M. Crawford. 1991. Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. *Communication in Statistics—Simulation and computation* B20: forthcoming.
- Lewis, P. A. W. 1970. Remarks on the theory, computation, and application of the spectral analysis of series of events. *Journal of Sound and Vibration* 12: 353–375.
- Lewis, P. A. W. 1972. Recent results in the statistical analysis of univariate point processes. In *Stochastic point processes*, ed. P. A. W. Lewis, 1–54. New York: John Wiley & Sons.
- Lewis, P. A. W. and G. S. Shedler. 1976. Statistical analysis of non-stationary series events in a data base system. *IBM Journal of Research and Development* 20: 465–482.
- Lewis, P. A. W. and G. S. Shedler. 1979. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* 26: 403–413.
- MacLean, C. J. 1974. Estimation and testing of an exponential polynomial rate function within the nonstationary Poisson process. *Biometrika* 61: 81–85.

AUTHOR BIOGRAPHIES

MARY A. JOHNSON is an Assistant Professor in the Department of Mechanical and Industrial Engineering at the University of Illinois-Urbana/Champaign. She received a B.S. degree in mathematics from Adrian College in 1981, and she received M.S. and Ph.D. degrees in industrial engineering from Purdue University in 1986 and 1988 respectively. Her research interests are focused on computational aspects of stochastic models, particularly fitting phase-type distributions, modeling nonstationary point processes, and determining error bounds for queueing performance measures.

SANGHOON LEE is an Assistant Professor in the Department of Industrial Engineering at Kyung Won University, Korea. He received a B.S. in mechanical engineering from Sung Kyun Kwan University (Ko-

rea) in 1981, and he received M.S. and Ph.D. degrees in operations research from The University of Texas at Austin in 1985 and 1990 respectively. His research interests are focused on probabilistic optimization, resampling, and simulation. He is also involved in statistical pattern recognition and stochastic modeling for remotely-sensed images.

JAMES R. WILSON is a Professor in the Department of Industrial Engineering at North Carolina State University. His current research interests are focused on the design and analysis of simulation experiments, including modeling and generation of probabilistic input processes, output analysis, variance reduction techniques, and simulation optimization. He also has an active interest in applications of operations research techniques to all areas of industrial engineering. He currently serves as Associate Editor of *IIIE Transactions*, Departmental Editor of *Management Science* for Simulation, and Associate Program Chairman of the 1991 Winter Simulation Conference.