COMBINED INVERSION AND THINNING METHODS FOR SIMULATING NONSTATIONARY NON-POISSON ARRIVAL PROCESSES

Ran Liu

SAS Institute
700 SAS Campus Drive
Cary, NC 27513, USA

Michael E. Kuhl

Department of Industrial and Systems Engineering
Rochester Institute of Technology
Rochester, NY 14623, USA

Yunan Liu
James R. Wilson

Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University
Raleigh, NC 27695, USA

ABSTRACT

We develop and evaluate SCIATA, a simplified combined inversion-and-thinning algorithm for simulating a nonstationary non-Poisson process (NNPP) over a finite time horizon, with the target arrival process having a given "rate" function and associated mean-value function together with a given variance-to-mean (dispersion) ratio. Designed for routine use when the dispersion ratio is at most two, SCIATA encompasses the following steps: (i) computing a piecewise-constant majorizing rate function that closely approximates the given rate function; (ii) computing the associated piecewise-linear majorizing mean-value function; (iii) generating an equilibrium renewal process (ERP) whose noninitial interrenewal times are Weibull distributed with mean one and variance equal to the given dispersion ratio; (iv) inverting the majorizing mean-value function at the ERP's renewal epochs to generate the associated majorizing NNPP; and (v) thinning the resulting arrival epochs to obtain an NNPP with the given rate function and dispersion ratio. Numerical examples illustrate the effectiveness of SCIATA in practice.

1 INTRODUCTION

Simulation modeling and analysis of a complex system often requires generation of an arrival process with the following characteristics:

- a given time-dependent instantaneous rate function over a given finite time horizon, where the rate function may have been estimated from data collected on the target arrival process;
- the associated mean-value function describing the buildup of the expected number of arrivals over the time horizon; and
- a given variance-to-mean (dispersion) ratio for the number of arrivals at each noninitial point in the time horizon, where the dispersion ratio also may have been estimated from data collected on the target arrival process.

For example, Figure 1 depicts the estimated mean-value function, the corresponding estimated variance function, and the estimated dispersion ratio that were observed for arrivals to an endocrinology outpatient

clinic of a major teaching hospital in South Korea (Kim et al. 2015). Figure 1 clearly shows that in each morning and afternoon period when the clinic is open, the buildup of patient arrivals exhibits time-dependent behavior as well as a variance-to-mean ratio close to 0.5. On the other hand, similar plots for arrivals to the emergency department of an Israeli hospital over a twenty-four period exhibit pronounced time-of-day and day-of-the-week effects as well as a dispersion ratio that was estimated to be about 1.5 after the day-of-the-week effects were properly taken into account (Whitt and Zhang 2015).

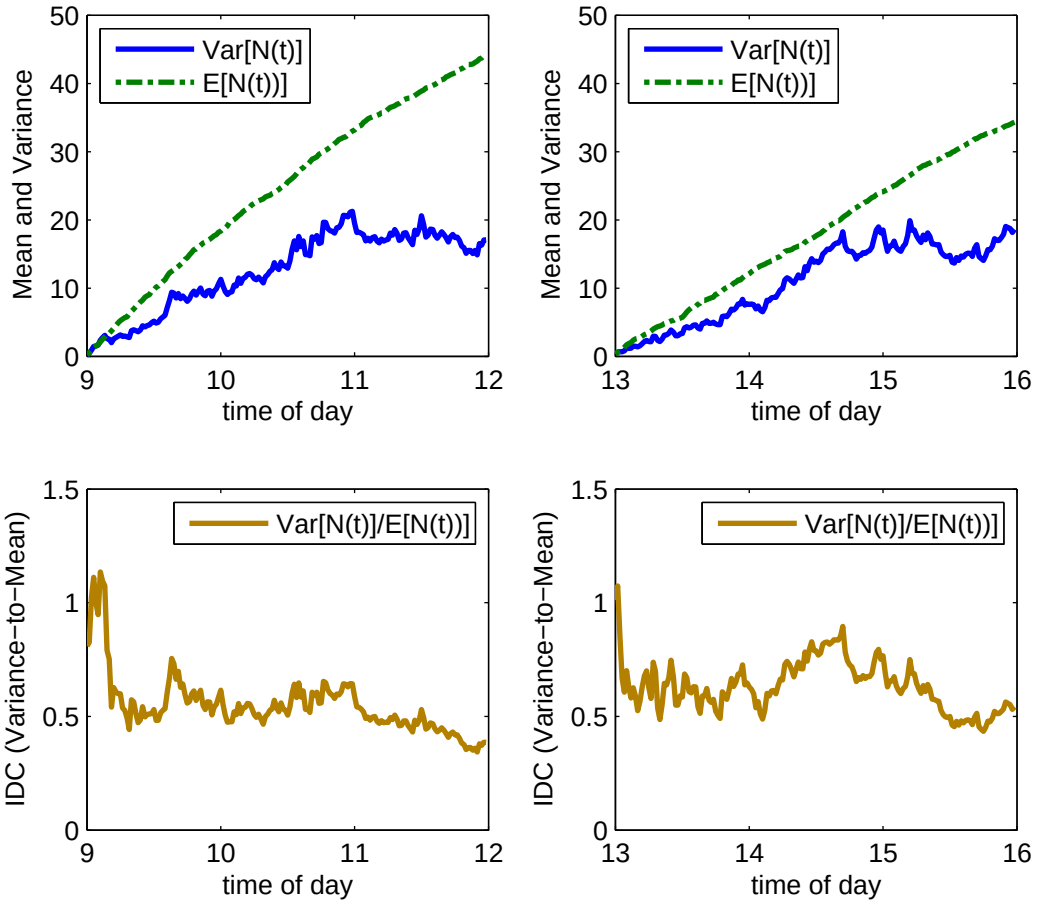

Figure 1: Estimated mean, variance, and variance-to-mean (dispersion) ratio for the number of arrivals to a Korean endocrinology outpatient clinic.

We examine the implications of these observations in some detail. For a nonstationary arrival process $\{N(t) : t \in [0,S]\}$ that is defined on a finite time horizon $[0,S]$, we let $N(t)$ denote the number of arrivals in the interval $[0,t]$ for $0 \le t \le S$; and we define the *mean-value function*,

$$\mathrm{E}[N(t)] \equiv \mu(t) \equiv \int_0^t \lambda(u)\,\mathrm{d}u \quad \text{for} \quad t \in [0,S], \tag{1}$$

as the expected number of arrivals in $[0,t]$, where the instantaneous *"rate" function* $\lambda(u)$ is nonnegative, bounded, and integrable over the given time horizon. We define the dispersion (variance-to-mean) ratio

$$C(t) \equiv \frac{\mathrm{Var}[N(t)]}{\mathrm{E}[N(t)]} \quad \text{for} \quad t \in (0,S]\,. \tag{2}$$

If $\{N(t) : t \in [0,S]\}$ is a nonhomogeneous Poisson process (NHPP) whose rate function is positive on $(0,S)$, then $C(t) = 1$ for all $t \in (0,S]$ (Çinlar 1975). NHPPs have been used by researchers and practitioners to represent time-dependent arrival processes exhibiting a wide variety of behaviors, including cyclic effects and long term trends—for example, see Lewis and Shedler (1976), Lee, Wilson, and Crawford (1991), Harper et al. (2000), Kuhl and Wilson (2001), Leemis (2004), and Kuhl, Deo, and Wilson (2008). Although NHPPs have been shown to model a large class of arrival processes effectively, there are systems such as the above-mentioned Korean endocrinology clinic and the Israeli emergency department for which NHPPs are not adequate, and alternative modeling approaches are needed.

If a time-dependent arrival process has a constant dispersion ratio so that $C(t) = C$ for all $t \in (0,S]$ but $C \neq 1$, then the resulting process is an example of a nonhomogeneous non-Poisson Process (NNPP). Liu et al. (2015) develop CIATA, a combined inversion-and-thinning algorithm for efficiently generating realizations of an NNPP $\{N(t) : t \in [0,S]\}$ with a given rate function $\lambda(t)$ for $t \in [0,S]$ and with a dispersion ratio $C(t)$ for $t \in (0,S]$ that rapidly approaches a given value C as the time t increases. In this paper we develop SCIATA, a simplified version of CIATA (Liu et al. 2015) for generating such NNPPs more easily when $C \in (0,2]$, a situation that in our experience is common in many practical applications. The slides for the oral presentation of this article are available online via www.ise.ncsu.edu/jwilson/wsc15sciata.pdf.

2 SUMMARY OF SCIATA: A SIMPLIFIED COMBINED INVERSION-AND-THINNING ALGORITHM FOR SIMULATING NNPPS WITH A CONSTANT DISPERSION RATIO

The main steps of SCIATA are detailed below.

Step 1. Construct a piecewise-constant majorizing rate function. We construct a piecewise-constant majorizing rate function $\widetilde{\lambda}_Q(t)$ that closely bounds the given rate function $\lambda(t)$ for $t \in [0,S]$ based on a partition of $[0,S]$ into Q subintervals, where Q is taken sufficiently large to ensure that $\lambda(t)$ is constant or unimodal in each subinterval. Let

$$\mathscr{L} \equiv \bigcup_{\ell=1}^{L} (\zeta_\ell^\dagger, \zeta_\ell^\ddagger)$$

represent the set of nonoverlapping subintervals of $[0,S]$ on which $\lambda(u)$ is constant with associated endpoints $\{\zeta_\ell^\dagger, \zeta_\ell^\ddagger : \ell = 1,\ldots,L\}$. Let $\{\overline{\zeta}_\ell \equiv \frac{1}{2}(\zeta_\ell^\dagger + \zeta_\ell^\ddagger) : \ell = 1,\ldots,L\}$ denote the corresponding subinterval midpoints. For each $Q \geq 1$, we take

$$\widetilde{\lambda}_Q(t) \equiv \left\{ \begin{array}{ll} \varepsilon/Q & \text{if } t \in (\zeta_\ell^\dagger, \zeta_\ell^\ddagger) \text{ and } \lambda(\overline{\zeta}_\ell) = 0 \\ \lambda(\overline{\zeta}_\ell) & \text{if } t \in (\zeta_\ell^\dagger, \zeta_\ell^\ddagger) \text{ and } \lambda(\overline{\zeta}_\ell) > 0 \end{array} \right\} \quad \text{for} \quad t \in \mathscr{L}, \tag{3}$$

where ε is an arbitrarily small positive number. Then, on $[0,S] \setminus \mathscr{L}$, for each of the subintervals $\{[\xi_j^\dagger, \xi_j^\ddagger] : j = 1,\ldots,M\}$ on which $\lambda(t)$ is nonconstant, we let

$$\mathscr{M} \equiv \bigcup_{j=1}^{M} [\xi_j^\dagger, \xi_j^\ddagger] = [0,S] \setminus \mathscr{L}$$

denote the subintervals of the time horizon on which the construction of $\widetilde{\lambda}_Q(t)$ is nontrivial. For $j = 1,\ldots,M$, we partition the subinterval $[\xi_j^\dagger, \xi_j^\ddagger]$ into equal-length subsubintervals $\{[z_{i-1,j}, z_{i,j}] : i = 1,\ldots,Q'; j = 1,\ldots,M\}$ of sufficiently small length so that $\widetilde{\lambda}_Q(t)$ is unimodal on each subsubinterval. Then for each $i = 1,\ldots,Q'$ and $j = 1,\ldots,M$, we seek to construct the majorizing rate function $\widetilde{\lambda}_Q(t)$ on $[z_{i-1,j}, z_{i,j}]$ by using the golden section search procedure to find the maximum value $\lambda_{i,j}^*$ of the given rate function $\lambda(t)$ on that subsubinterval. This construction yields a majorizing rate function defined on $\mathscr{M}$ by the relation

$$\widetilde{\lambda}_Q(t) = \begin{cases} \sum_{i=1}^{Q'} \sum_{j=1}^{M} \lambda_{i,j}^* \mathbb{I}_{(z_{i-1,j}, z_{i,j}]}(t) & \text{for } t \in \mathscr{M} \setminus \{\xi_j^\dagger : j = 1,\ldots,M\}, \\ \lambda_{1,j}^* & \text{if } t = \xi_j^\dagger \text{ for some } j \in \{1,\ldots,M\}, \end{cases} \tag{4}$$

where $\mathbb{I}_{(z_{i-1,j},z_{i,j}]}(t)$ is the indicator function for $(z_{i-1,j},z_{i,j}]$ so that we have $\mathbb{I}_{(z_{i-1,j},z_{i,j}]}(t)\equiv 1$ if $z_{i-1,j}<t\le z_{i,j}$ and otherwise $\mathbb{I}_{(z_{i-1,j},z_{i,j}]}(t)\equiv 0$. The formal statement of the process for constructing $\widetilde{\lambda}_Q(t)$ is given in Algorithm 1.

Algorithm 1 Constructing the piecewise-constant majorizing rate function (Liu et al. 2015)

Initialization: Set $\delta\leftarrow 10^{-4}$, $\varepsilon\leftarrow 10^{-20}$, and $Q^*\leftarrow 200$; then assign $Q'\ge Q^*$.
for $j=1,\ldots,M$ **do**
  Partition $[\xi_j^\dagger,\xi_j^\ddagger]$ into Q' equal-length subintervals $\{[z_{i-1,j},z_{i,j}]: i=1,\ldots,Q'\}$:
    Set $z_{i,j}\leftarrow \xi_j^\dagger+i(\xi_j^\ddagger-\xi_j^\dagger)/Q'$ for $i=0,1,\ldots,Q'$.
  for $i=1,\ldots,Q'$ **do**
    Set $a\leftarrow z_{i-1,j}$, $b\leftarrow z_{i,j}$, $x_1\leftarrow a+(2-\phi)(b-a)$, $x_2\leftarrow a+(\phi-1)(b-a)$.
    while $(b-a)\ge\delta$ **do**
      if $\lambda(x_1)<\lambda(x_2)$ **then**
        Set $\lambda^*_{i,j}\leftarrow\lambda(x_2)$, $a\leftarrow x_1$, $x_1\leftarrow x_2$, $x_2\leftarrow a+(\phi-1)(b-a)$;
      else
        Set $\lambda^*_{i,j}\leftarrow\lambda(x_1)$, $b\leftarrow x_2$, $x_2\leftarrow x_1$, $x_1\leftarrow a+(2-\phi)(b-a)$.
      end if
    end while
  end for
end for
Set $Q\leftarrow L+MQ'$, and define $\widetilde{\lambda}_Q(\cdot)$ on $\mathscr{L}$ and $\mathscr{M}$ by (3) and (4), respectively.

Step 2. Construct a piecewise-linear majorizing mean-value function. Using the majorizing rate function defined by (3) and (4), the majorizing mean-value function is defined as

$$\widetilde{\mu}_Q(t)\equiv\int_0^t\widetilde{\lambda}_Q(u)\big[1-\mathbb{I}_{\mathscr{Z}}(u)\big]\,\mathrm{d}u \text{ for } 0\le t\le S,$$

where $\mathscr{Z}\equiv\{u\in[0,S]:\lambda(u)=0\}$ and $\mathbb{I}_{\mathscr{Z}}(u)$ is the indicator function of $\mathscr{Z}$. Liu et al. (2015) show that the majorizing mean-value function $\widetilde{\mu}_Q(t)$ converges uniformly to the target $\mu(t)$ from above as the length of majorizing subintervals decreases.

Step 3. Generate an equilibrium renewal process (ERP). To obtain an ERP with the desired dispersion ratio, the noninitial interrenewal times are sampled from an appropriate Weibull distribution. The ERP $\big\{N^\circ(u):u\in[0,\widetilde{\mu}_Q(S)]\big\}$ has renewal epochs $\big\{S_n^\circ:n=1,\ldots,N^\circ[\widetilde{\mu}_Q(S)]\big\}$ and interrenewal times $\big\{X_n^\circ:n=1,\ldots,N^\circ[\widetilde{\mu}_Q(S)]\big\}$ satisfying $\mathrm{E}[X_n^\circ]=1$ and $\mathrm{Var}[X_n^\circ]=C$ for $n\ge 2$. Therefore we have the usual setup for a renewal process,

$$S_0^\circ=0 \text{ and } S_n^\circ=S_{n-1}^\circ+X_n^\circ \text{ for } n=1,2,\ldots, \text{ so that } N^\circ(t)=\sup\{n:S_n^\circ\le t\} \text{ for } t\in[0,S].$$

The probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of the required Weibull distribution are given by

$$f_{\mathrm{We}}(x;\alpha,\beta)=\begin{cases}0 & \text{if } x<0,\\ (\alpha/\beta)(x/\beta)^{\alpha-1}\exp\big[-(x/\beta)^\alpha\big] & \text{if } x\ge 0,\end{cases}$$

and

$$F_{\mathrm{We}}(x;\alpha,\beta)=\begin{cases}0 & \text{if } x<0\\ 1-\exp\big[-(x/\beta)^\alpha\big] & \text{if } x\ge 0,\end{cases}$$

respectively, where the required value of the shape parameter α is the unique positive root of the equation

$$\frac{\alpha\{2\Gamma(2/\alpha)-(1/\alpha)[\Gamma(1/\alpha)]^2\}}{[\Gamma(1/\alpha)]^2}=C; \tag{5}$$

and the associated scale parameter is given by

$$\beta=\alpha/\Gamma(1/\alpha)\,. \tag{6}$$

Next we formulate the initial interrenewal distribution that must be sampled to ensure that the ERP $\{N^\circ(u):u\in[0,\widetilde{\mu}_Q(S)]\}$ starts out in "steady-state" operation in the sense that the time to the first renewal X_1° has the same c.d.f. $G_e(x)$ as the steady-state forward recurrence-time (excess life) of the ERP; moreover the ERP's mean-value function is exactly linear at every point in time so that $\mathrm{E}[N^\circ(u)]=u/\mathrm{E}[X_2^\circ]=u$ for all $u\in[0,\widetilde{\mu}_Q(S)]$. In terms of the auxiliary gamma p.d.f. with shape parameter η and scale parameter 1,

$$f_{\mathrm{Ga}}(x;\eta)=\begin{cases}0, & \text{if } x\le 0\\ \dfrac{1}{\Gamma(\eta)}x^{\eta-1}e^{-x} & \text{if } x>0,\end{cases}$$

and the associated gamma c.d.f.,

$$F_{\mathrm{Ga}}(x;\eta)=\begin{cases}0, & \text{if } x\le 0\\ \displaystyle\int_0^x f_{\mathrm{Ga}}(u;\eta)\,\mathrm{d}u & \text{if } x>0\,,\end{cases}$$

we see that the initial interrenewal time (that is, the time to the first renewal) is sampled as follows:

$$X_1^\circ\sim G_e(x)\equiv F_{\mathrm{Ga}}\big[(x/\beta)^\alpha;1/\alpha\big]\,. \tag{7}$$

Equation (7) is easily seen to be equivalent the following:

$$\text{Generate } Y\sim F_{\mathrm{Ga}}(y;1/\alpha) \text{ and deliver } X_1^\circ\leftarrow\beta\big(Y^{1/\alpha}\big)\,.$$

All subsequent interrenewal times are generated from

$$\{X_i^\circ:i=2,3,\ldots\}\overset{\text{i.i.d.}}{\sim}G(x)\equiv F_{\mathrm{We}}(x;\alpha,\beta)\,, \tag{8}$$

where α and β are given by Equations (5) and (6), respectively.

Step 4. Generate the arrival epochs for the majorizing NNPP. The arrival epochs for the majorizing NNPP $\{\widetilde{S}_n:n=1,\ldots,\widetilde{N}_Q(S)\}$, where $\widetilde{N}_Q(S)=N^\circ[\widetilde{\mu}_Q(S)]$, are generated by inverting the majorizing mean-value function at the renewal epochs of the ERP so that we take $\widetilde{S}_n\equiv\widetilde{\mu}_Q^{-1}(S_n^\circ)$ for $n=1,\ldots,\widetilde{N}_Q(S)$.

Step 5. Apply thinning to the majorizing NNPP to yield the arrival epochs of the target NNPP. The method of thinning is applied to the arrival epochs $\{\widetilde{S}_n:n=1,\ldots,\widetilde{N}_Q(S)\}$ so that for $n=1,\ldots,\widetilde{N}_Q(S)$, the nth arrival epoch $\widetilde{S}_n$ of the majorizing NNPP is independently accepted as an arrival epoch of the target NNPP with probability $\lambda(\widetilde{S}_n)\big/\widetilde{\lambda}_Q(\widetilde{S}_n)$; and $\{S_j:j=1,\ldots,N_Q(S)\}$ denotes the resulting sequence of arrival epochs of the SCIATA-generated NNPP over the time horizon $[0,S]$.

The formal algorithmic statement of SCIATA is given in Algorithm 2.

3 PERFORMANCE EVALUATION OF SCIATA

In this section we evaluate the performance of SCIATA by conducting various simulation experiments. We first describe the experimental setup in §3.1 and the estimation procedure in §3.2. Next we provide simulation results for SCIATA in this section.

Algorithm 2 Using SCIATA to Generate an NNPP (Liu et al. 2015)

Construct the majorizing rate function $\widetilde{\lambda}_Q(t)$ given by (3) and (4) using Algorithm 1.
Define the c.d.f.'s $G_e(x)$ and $G(x)$ as in (7) and (8),
Set $n \leftarrow 1$, $j \leftarrow 0$, $S_0^\circ \leftarrow 0$, $\widetilde{S}_0 \leftarrow 0$, and $S_0 \leftarrow 0$. Generate $X_n^\circ \sim G_e$ and set $S_n^\circ \leftarrow S_{n-1}^\circ + X_n^\circ$. Set $\widetilde{S}_n \leftarrow \widetilde{\mu}^{-1}(S_n^\circ)$ and $\widetilde{X}_n \leftarrow \widetilde{S}_n - \widetilde{S}_{n-1}$.
while $\widetilde{S}_n \le S$ **do**
  Generate $U_n \sim$ Uniform[0,1].
  if $U_n \le \lambda(\widetilde{S}_n)/\widetilde{\lambda}(\widetilde{S}_n)$ **then**
    Set $j \leftarrow j+1$, $S_j \leftarrow \widetilde{S}_n$. Set the interarrival time $X_j \leftarrow S_j - S_{j-1}$.
  end if
  Set $n \leftarrow n+1$. Generate $X_n^\circ \sim G$. Set $S_n^\circ \leftarrow S_{n-1}^\circ + X_n^\circ$, $\widetilde{S}_n \leftarrow \widetilde{\mu}^{-1}(S_n^\circ)$, and $\widetilde{X}_n \leftarrow \widetilde{S}_n - \widetilde{S}_{n-1}$.
end while

3.1 Experimental Setup

For $0 \le t \le S$, we chose rate functions of the type *exponential-polynomial-trigonometric with multiple periodicities* (EPTMP), which means they have the form

$$\lambda(t) = \exp\{h(t;m,p,\Theta)\} \quad \text{with} \qquad h(t;m,p,\Theta) = \sum_{i=0}^{m} \alpha_i t^i + \sum_{k=1}^{p} \gamma_k \sin(\omega_k t + \phi_k), \tag{9}$$

where

$$\Theta = [\alpha_0, \alpha_1, \ldots, \alpha_m, \gamma_1, \ldots, \gamma_p, \omega_1, \ldots, \omega_p, \phi_1, \ldots, \phi_p,]$$

is the vector of continuous parameters of the designated rate function (Kuhl, Wilson, and Johnson 1997). The first $m+1$ terms in Equation (9) define a degree-m polynomial representing a possible long-term evolutionary trend in the arrival rate over time. The next p terms in Equation (9) define the trigonometric functions representing possible periodic effects exhibited by the arrival process. The use of an exponential rate function is a convenient means of ensuring that the instantaneous arrival rate is always positive. For specific applications, we are free to choose the appropriate degree m for the polynomial rate component together with the appropriate oscillation amplitude (γ_k), oscillation frequency (ω_k), and phase delay (ϕ_k) for each of the cyclic rate components. For more details on EPTMP-type rate functions, see Kuhl, Wilson, and Johnson (1997).

The parameters for three instances of EPTMP-type rate functions are shown in Table 1. All three cases contain two nested cyclic effects, with the first effect having a period of 1 time unit and the second effect having a period 0.5 time units. Case 1 does not contain a general trend over time. Cases 2 and 3 contain general trends that are represented by polynomials of degree 1 and 2, respectively. We then take $S = 8$, and the desired asymptotic variance-to-mean ratio is taken to be $C = 0.5, 0.75, 1.5$, and 2 in each of these three cases. To conserve space, we only present the results for case 1 in this article; but in the above-mentioned oral presentation of this work, we will provide the complete results for cases 2 and 3 as well as for case 1.

Table 1: Parameters of the NNPPs with EPTMP-type rate functions that are used in the experimental evaluation of SCIATA

Parameter	α_0	α_1	α_2	γ_1	ϕ_1	ω_1	γ_2	ϕ_2	ω_2
Case 1	3.6269	–	–	1.0592	-0.6193	6.2831	0.5000	0.5000	12.5664
Case 2	3.6269	0.1000	–	1.0592	-0.6193	6.2831	0.5000	0.5000	12.5664
Case 3	3.6269	-0.1000	0.0200	1.0592	-0.6193	6.2831	0.5000	0.5000	12.5664

To evaluate the performance of SCIATA, we performed a *metaexperiment* consisting of $R = 100$ independent basic experiments; and in each basic experiment, we performed $K = 200$ independent replications

of SCIATA to generate 200 realizations of an NNPP $\{N(t) : t \in [0,S]\}$ for one of the three cases in Table 1 with the dispersion ratio $C = 0.5, 0.75, 1.5$, or 2.0. Each basic experiment yielded estimates of the mean-value function $\mu(t) = \mathrm{E}[N_Q(t)]$ and the dispersion-ratio function $C_Q(t) \equiv \mathrm{Var}[N_Q(t)]/\mathrm{E}[N_Q(t)]$ for $t \in (0,S]$.

From 100 replications of each basic experiment, we computed 95% confidence interval (CI) estimators of $\mu(t_i)$ and $C_Q(t_i)$ at the times $\{t_i : t_i = \psi i, \text{ for } i = 1,\ldots,\mathbb{T}\}$, where ψ is the length of each subinterval and $\mathbb{T} = S/\psi$. For all the experiments in this section, we set $\psi = 0.2$. To validate SCIATA, we examined the following:

- Graphs of the desired rate function $\lambda(t_i)$ and the desired mean-value function $\mu(t_i)$ respectively superimposed on the majorizing rate function $\widetilde{\lambda}_Q(t_i)$ and on 95% CI estimators of the mean-value function $\mu_Q(t_i)$ for the SCIATA-generated NNPP at the times $\{t_i : t_i = \psi i, \text{ for } i = 1,\ldots,\mathbb{T}\}$;
- Graphs of the desired asymptotic dispersion ratios $C = 0.5, 0.75, 1.5$, and 2.0 superimposed on simulation-based point and CI estimators of $C_Q(t_i)$ for the SCIATA-generated NNPP at the times $\{t_i : t_i = \psi i, \text{ for } i = 1,\ldots,\mathbb{T}\}$; and
- Tabular summaries of the same results depicted in the graphs of items 1 and 2 above.

All our algorithms have been implemented in MATLAB, and all the experiments in this article were carried out on a 2.66GHz Intel 2 Quad processor with 3.25 GB RAM, running Windows XP Professional.

3.2 Formulation of Performance Estimators

On the kth realization of the SCIATA-generated NNPP in the rth basic experiment, let $N_Q(t;r,k)$ denote the number of accepted arrivals up to time $t \in [0,S]$, where $r = 1,\ldots,R$, and $k = 1,\ldots,K$. In the rth basic experiment, the estimators of the mean-value and variance functions are the sample statistics

$$\left.\begin{aligned} \overline{N}_Q(t;r) &= \frac{1}{K}\sum_{k=1}^{K} N_Q(t;r,k) \\ \widehat{\mathrm{Var}}\,[N_Q(t;r,1)] &= \frac{1}{K-1}\sum_{k=1}^{K}\left[N_Q(t;r,k) - \overline{N}_Q(t;r)\right]^2 \end{aligned}\right\} \quad \text{for } t \in (0,S] \text{ and } r = 1,\ldots,R.$$

The sample dispersion ratio at time t is

$$\widehat{C}_Q(t;r) = \frac{\widehat{\mathrm{Var}}\,[N_Q(t;r,1)]}{\overline{N}_Q(t;r)} \quad \text{for } t \in (0,S].$$

Based on the entire metaexperiment, the overall estimator of the desired mean-value function $\mu(t)$ is

$$\widehat{\mu}_Q(t) = \frac{1}{RK}\sum_{r=1}^{R}\sum_{k=1}^{K} N_Q(t;r,k) \quad \text{for } t \in (0,S];$$

and a pooled estimator of $\mathrm{Var}[N_Q(t)]$ is

$$\widehat{\mathrm{Var}}[N_Q(t)] = \frac{1}{R}\sum_{r=1}^{R}\widehat{\mathrm{Var}}\,[N_Q(t;r,1)] \quad \text{for } t \in (0,S].$$

The overall point estimator of $C_Q(t)$ is

$$\widehat{C}_Q(t) = \frac{1}{R}\sum_{r=1}^{R}\widehat{C}_Q(t;r) \quad \text{for } t \in (0,S];$$

and the overall estimator of $\mathrm{Var}[\widehat{C}_Q(t;1)]$ is

$$\widehat{\mathrm{Var}}[\widehat{C}_Q(t;1)] = \frac{1}{R-1}\sum_{r=1}^{R}\left[\widehat{C}_Q(t;r)-\widehat{C}_Q(t)\right]^2 \quad \text{for } t\in(0,S].$$

For a fixed time t, an approximate $100(1-\alpha)\%$ CI estimator for $C_Q(t)$ is

$$\widehat{C}_Q(t)\pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}[\widehat{C}_Q(t;1)]/R} \quad \text{for } t\in(0,S],$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Similarly, CIs are obtained for the mean-value function $\mu(t)$ at each fixed time t.

3.3 Performance of SCIATA

Here we report results for Case 1, with the target dispersion ratios $C = 0.5$, 0.75 1.5, and 2.0. For $t\in[0,8]$, we plot in Figure 2 (i) the given Case-1 rate function $\lambda(t)$ (dashed red curve) and the corresponding majorizing step rate function $\widetilde{\lambda}_Q(t)$ (solid blue step function); (ii) the 95% CI estimators of the mean-value function $\mu(t)$; and (iii) the 95% CI estimators for $C_Q(t)$. Here $Q = 160$ so the majorizing step size is $S/Q = 0.05$.

For the dispersion ratios $C = 1.5$ and 2.0, Figure 3 depicts the CIATA-generated results for the Case-1 arrival-rate function based on the same layout used in Figure 2.

These results show the effectiveness of SCIATA in generating an NNPP with the given mean-value function $\mu(t)$ (or equivalently the given rate function $\lambda(t)$) and the given asymptotic dispersion ratio C. Because it takes less than 0.2 time units for the time-dependent dispersion ratio to converge to C, we concluded that SCIATA could deliver fast convergence to the given dispersion ratio. We remark that the convergence time (or warm-up time) becomes relatively longer for large values of the dispersion ratios (e.g., $C = 10$).

To provide further validation of SCIATA's performance, we examine the relative errors of the estimators $\widehat{\mu}_Q(t)$ and $\widehat{C}_Q(t)$ (i.e., their respective percentage deviations from the target measures $\mu(t)$ and C), which we call the *closeness* measures; and we study the sensitivity of these closeness measures to the dispersion ratio C and to the number of majorizing intervals Q. We define the *average percentage discrepancy* (APD) $\Delta_{\mathbb{T}}(\cdot)$ and *maximum percentage discrepancy* (MPD) $\Delta^*_{\mathbb{T}}(\cdot)$ as two measures of closeness:

$$\Delta_{\mathbb{T}}(\widehat{\mu}_Q) = \frac{1}{\mathbb{T}}\sum_{i=1}^{\mathbb{T}} D_i(\widehat{\mu}_Q), \quad \Delta^*_{\mathbb{T}}(\widehat{\mu}_Q) = \max_{1\le i\le\mathbb{T}}\left\{D_i(\widehat{\mu}_Q)\right\},$$
$$\Delta_{\mathbb{T}}(\widehat{C}_Q) = \frac{1}{\mathbb{T}}\sum_{i=1}^{\mathbb{T}} D_i(\widehat{C}_Q), \quad \Delta^*_{\mathbb{T}}(\widehat{C}_Q) = \max_{1\le i\le\mathbb{T}}\left\{D_i(\widehat{C}_Q)\right\},$$

where the percentage discrepancy $D_i(\cdot)$ is the percentage difference of the estimators $\widehat{\mu}_Q(t_i)$ and $\widehat{C}_Q(t_i)$ from their respective target values $\mu(t_i)$ and $\widehat{C}_Q(t_i)$ at each t_i:

$$D_i(\widehat{\mu}_Q) = \left|\frac{\widehat{\mu}_Q(t_i)-\mu(t_i)}{\mu(t_i)}\right|\times 100\% \quad \text{and} \quad D_i(\widehat{C}_Q) = \left|\frac{\widehat{C}_Q(t_i)-C}{C}\right|\times 100\% \quad \text{for } i=1,\ldots,\mathbb{T}.$$

In addition, we compute the approximate 95% CI estimator for the expected value of $\Delta_{\mathbb{T}}(\widehat{\mu}_Q)$,

$$\Delta_{\mathbb{T}}(\widehat{\mu}_Q)\pm z_{1-\alpha/2}\frac{S_\Delta(\widehat{\mu}_Q)}{\sqrt{\mathbb{T}}}, \quad \text{where} \quad S_\Delta(\widehat{\mu}_Q) = \sqrt{\frac{1}{\mathbb{T}-1}\sum_{i=1}^{\mathbb{T}}\left[D_i(\widehat{\mu}_Q)-\Delta_{\mathbb{T}}(\widehat{\mu}_Q)\right]^2},$$

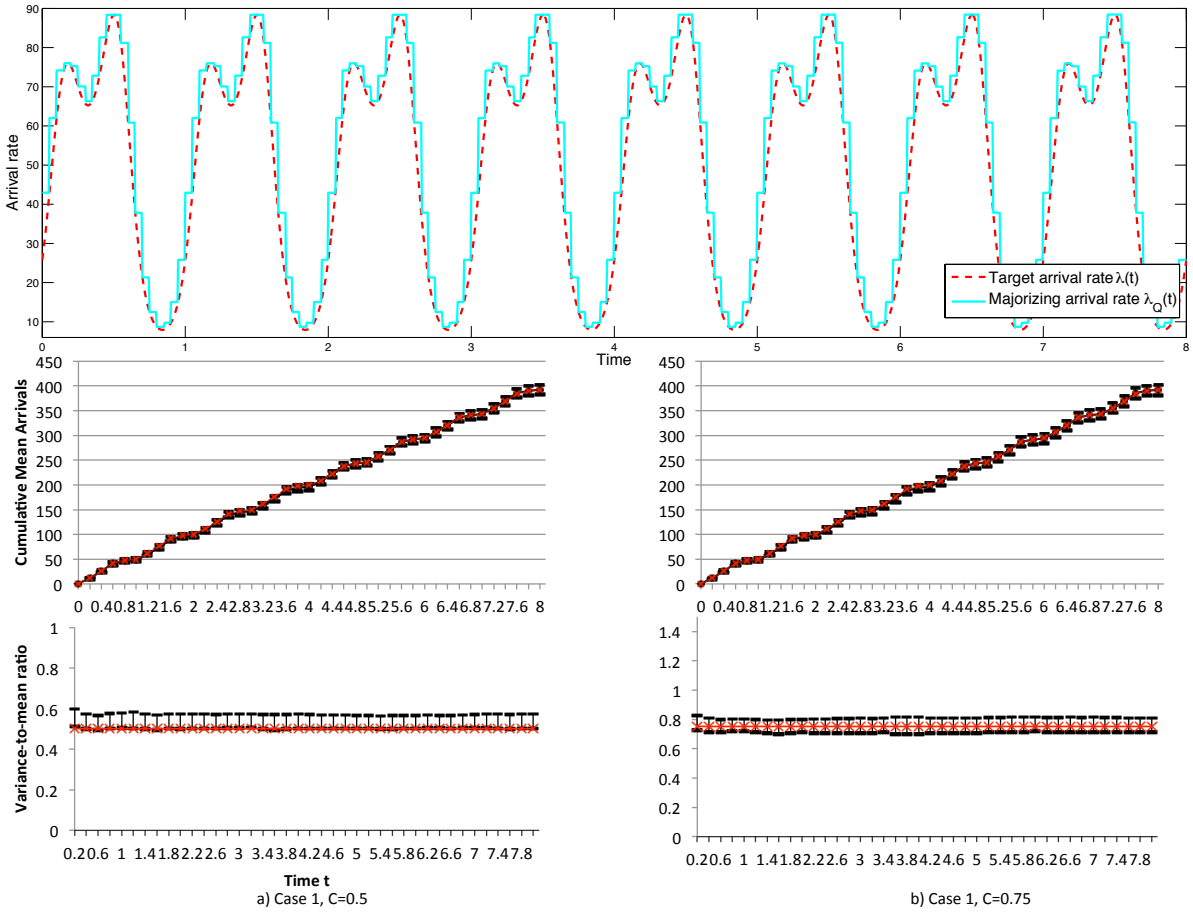

Figure 2: Performance of SCIATA for Case 1 with dispersion ratios $C = 0.5, 0.75$: (i) (top panel) the majorizing rate function $\widetilde{\lambda}_Q(t)$; (ii) (middle panel) 95% CIs for the mean-value function $\mu(t)$ with $C = 0.5$ (Left) and $C = 0.75$ (right); and (iii) (bottom panel) 95% CIs for $C_Q(t)$ with $C = 0.5$ (left) and $C = 0.75$ (right).

and we took $\alpha = 0.05$. For the APD estimator $\Delta_{\mathbb{T}}(\widehat{C}_Q)$, the standard-deviation estimator $S_{\Delta}(\widehat{C}_Q)$, and the CI estimator of $\mathrm{E}[\Delta_{\mathbb{T}}(\widehat{C}_Q)]$ are defined similarly.

Table 2 contains the closeness measures $\Delta_{\mathbb{T}}(\widehat{\mu}_Q)$, $\Delta^*_{\mathbb{T}}(\widehat{\mu}_Q)$, $\Delta_{\mathbb{T}}(\widehat{C}_Q)$, and $\Delta^*_{\mathbb{T}}(\widehat{C}_Q)$ and the approximate 95% CIs for $\mathrm{E}[\Delta_{\mathbb{T}}(\widehat{\mu}_Q)]$ and $\mathrm{E}[\Delta_{\mathbb{T}}(\widehat{C}_Q)]$ for the Case-1 arrival rate with $C = 0.5$, 0.75, 1.5, and 2. Table 2 shows that the APD and MPD for the mean-value function of the SCIATA-generated NNPP are less than 1% and 3%; and the APD and MPD for the dispersion ratio of the SCIATA-generated NNPP are less than 4% and 6%, respectively. We judged these result to provide good evidence of the effectiveness and efficiency of the CIATA-based estimators of the given rate and mean-value functions as well as the given dispersion ratio.

We remark that the adequacy of a CIATA-generated NNPP can also be impacted by Q, the number of subintervals that are used to generate the majorizing rate function $\widetilde{\lambda}_Q(t)$.

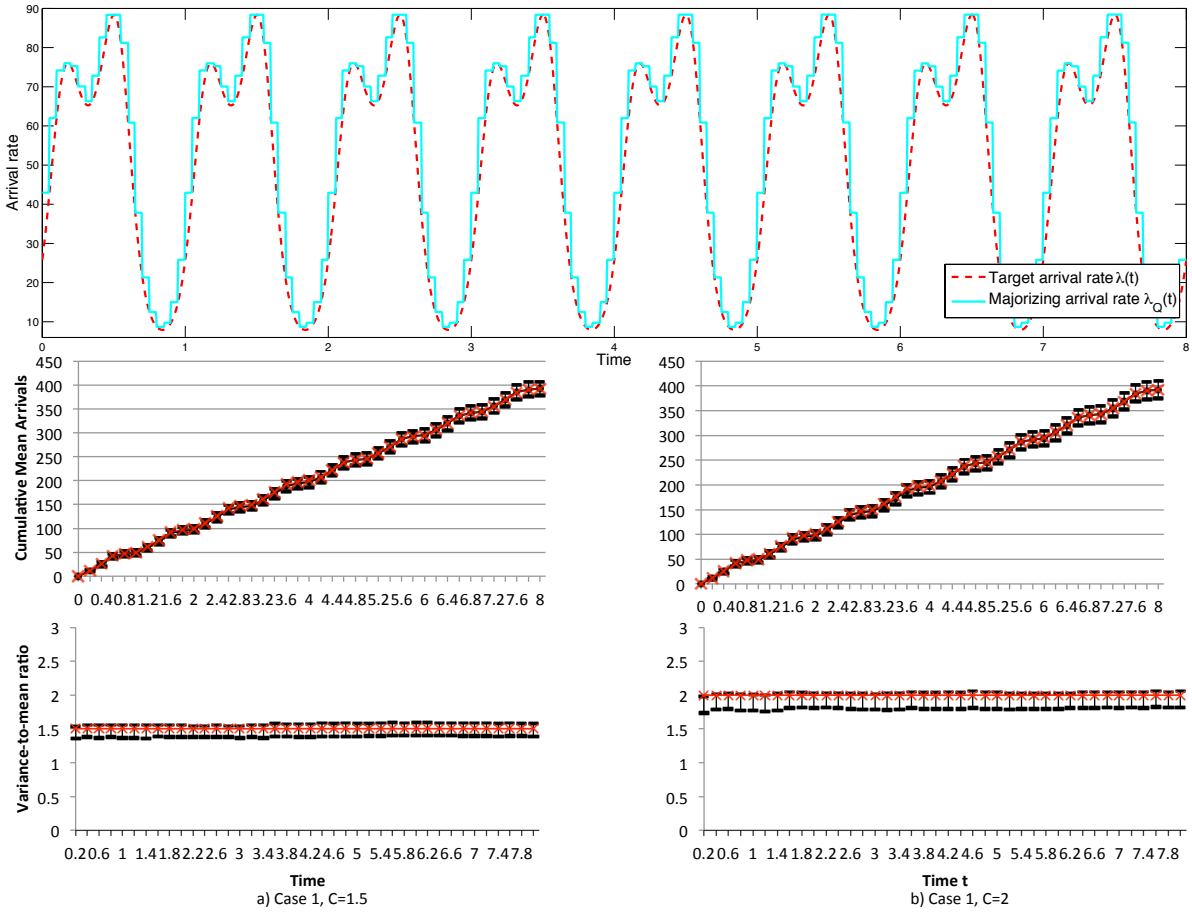

Figure 3: CIATA performance for Case 1 with dispersion ratio $C = 1.5, 2.0$: (i) (top panel) the majorizing rate function $\widetilde{\lambda}_Q(t)$; (ii) (middle panel) 95% CIs for the mean-value function $\mu(t)$ with $C = 1.5$ (left) and $C = 2.0$ (right); and (iii) (bottom panel) 95% CIs for $C_Q(t)$ with $C = 1.5$ (left) and $C = 2.0$ (right).

Table 2: SCIATA-based closeness measures for Case 1.

Scenario	$\Delta_{\mathbb{T}}(\widehat{\mu}_Q)$	$\Delta^*_{\mathbb{T}}(\widehat{\mu}_Q)$	$\Delta_{\mathbb{T}}(\widehat{C}_Q)$	$\Delta^*_{\mathbb{T}}(\widehat{C}_Q)$
$C = 0.5$	0.885% ± 0.193%	2.206%	7.324%± 0.178%	11.241%
$C = 0.75$	0.888%±0.194%	2.067%	1.255%±0.136%	3.447%
$C = 1.5$	0.895% ± 0.179%	2.176%	1.457%± 0.206%	3.397%
$C = 2.0$	0.863%±0.191%	2.017%	4.206%±0.177%	7.050%

4 CONCLUSIONS AND RECOMMENDATIONS

In this article we formulated and experimentally evaluated two effective procedures for modeling and simulation of nonstationary non-Poisson processes. More specifically, our contributions are as follows:

- We developed SCIATA for modeling and simulation of an NNPP with given rate and mean-value functions as well as a given asymptotic dispersion ratio. We provided detailed proofs that our method overcomes the disadvantages of the inversion and thinning methods of (Gerhardt and Nelson 2009)

while retaining the attractive features of those methods. More specifically, the key advantages of SCIATA can be summarized as follows:

- SCIATA can be applied to generate a wider range of NNPPs than other methods, especially for NNPPs whose given rate function $\lambda(t)$ corresponds to a mean-value function that cannot be inverted either analytically or numerically with sufficient accuracy and speed for large-scale simulation applications. SCIATA achieves this capability by constructing a step rate function $\widetilde{\lambda}_Q(t)$ that majorizes and closely approximates $\lambda(t)$.
- SCIATA generates NNPPs whose theoretical mean-value function exactly equals the given mean-value function $\mu(t)$ over the given time horizon $[0,S]$; and SCIATA's execution speed increases linearly with the length S of the time horizon.
- SCIATA generates NNPPs whose time-dependent dispersion ratio $C_Q(t)$ converges to the desired asymptotic dispersion ratio C with bias of the form $o[\mu(t)]/\mu(t)$, provided the majorizing step rate function $\widetilde{\lambda}_Q(t)$ is based on a sufficiently fine subdivision of the time horizon $[0,S]$ into Q subintervals. Thus if $\mu(t)$ grows without bound as t increases, then

$$C_Q(t) \approx C \quad \text{for sufficiently large } Q \text{ and } t. \tag{10}$$

  In our numerical experiments, we often found that taking $200 \leq Q \leq 300$ was sufficient to ensure the validity of the approximation (10) for values of $t \geq t_w$, the so-called warm-up time; in many test processes, we found that t_w was between 2.5% and 15% of length S of the time horizon.
- SCIATA is computationally efficient in generating NNPPs by utilizing a step rate function which approximates the given rate function closely.

- We conducted several computational experiments with different designated mean-value functions and asymptotic dispersion ratios. We showed that SCIATA can generate NNPPs with a given mean-value function and asymptotic dispersion ratio. The performance of SCIATA, expressed by the speed of its convergence and by the closeness of the estimated values $\widehat{\mu}_Q(t)$ and $\widehat{C}_Q(t)$ to their theoretical counterparts $\mu(t)$ and C, varies among test processes. The performance of SCIATA depends on the properties of the rate function $\lambda(t)$, on the value of asymptotic dispersion ratio C, and on the value of Q, the number of subintervals that is used to generate the majorizing rate function $\widetilde{\lambda}_Q(t)$.

There are several promising directions for future work on modeling and simulation of NNPPs. In particular we are exploring methods to minimize the "warm-up" period that is required for the $C(t)$ to stabilize at the target value C. This effort involves identifying alternatives to the Weibull distribution in Step 3 of SCIATA and to the hyperexponential and Erlang distributions used in Step 3 of CIATA. We are also exploring extensions of SCIATA and CIATA in which the target dispersion ratio is piecewise constant over the given time horizon. The latter problem appears to be difficult, and it is unclear that a satisfactory solution can be readily formulated and implemented.

REFERENCES

Çinlar, E. 1975. *Introduction to Stochastic Processes.* Englewood Cliffs, New Jersey: Prentice-Hall.

Gerhardt, I., and B. L. Nelson. 2009. "Transforming Renewal Processes for Simulation of Nonstationary Arrival Processes". *INFORMS Journal on Computing* 21 (4): 630–640.

Harper, A. M., S. E. Taranto, E. B. Edwards, and O. P. Daily. 2000. "Organ Transplantation Policies: An Update on a Successful Simulation Project: The Unos Liver Allocation Model". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 1955–1962. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Kim, S.-H., P. Vel, W. Whitt, and W. C. Cha. 2015. "Poisson and non-Poisson Properties in Appointment-Generated Arrival Processes: The Case of an Endocrinology Clinic". *Operations Research Letters* 43:247–253.

Kuhl, M. E., S. C. Deo, and J. R. Wilson. 2008. "Smooth Flexible Models of Nonhomogeneous Poisson Processes Using One or More Process Realizations". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 353–361. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Kuhl, M. E., and J. R. Wilson. 2001. "Modeling and Simulating Poisson Processes Having Trends or Nontrigonometric Cyclic Effects". *European Journal of Operational Research* 133 (3): 566–582.

Kuhl, M. E., J. R. Wilson, and M. A. Johnson. 1997. "Estimating and Simulating Poisson Processes Having Trends or Multiple Periodicities". *IIE Transactions* 29 (3): 201–211.

Lee, S., J. R. Wilson, and M. M. Crawford. 1991. "Modeling and Simulation of a Nonhomogeneous Poisson Process Having Cyclic Behavior". *Communications in Statistics—Simulation and Computation* 20 (2–3): 777–809.

Leemis, L. 2004. "Nonparametric Estimation and Variate Generation for a Nonhomogeneous Poisson Process from Event Count Data". *IIE Transactions* 36 (12): 1155–1160.

Lewis, P. A. W., and G. S. Shedler. 1976. "Statistical Analysis of Non-stationary Series of Events in a Data Base System". *IBM Journal of Research and Development* 20 (5): 465–482.

Liu, R., M. E. Kuhl, Y. Liu, and J. R. Wilson. 2015. "Modeling and Simulation of Nonstationary Non-Poisson Arrival Processes". *INFORMS Journal on Computing* in revision.

Whitt, W., and X. Zhang. 2015. "A Data-Generated Model of an Emergency Department". Technical report, Department of Industrial Enginering and Operations Research, Columbia University, New York, NY.

AUTHOR BIOGRAPHIES

RAN LIU is a consultant at SAS Institute. He earned his Ph.D. in Operations Research from North Carolina State University. His primary research interest includes simulation modeling and optimization and their applications in retail and call center systems. His email address is Ran.Liu@sas.com.

MICHAEL E. KUHL is a professor in the Department of Industrial and Systems Engineering at the Rochester Institute of Technology. He earned his Ph.D. in Industrial Engineering from North Carolina State University. His research interests include modeling and simulation of stochastic arrival processes, and the application of simulation and simulation-based optimization to systems including healthcare, manufacturing, cyber security, and project management. He has served WSC as *Proceedings* Editor (2005) and Program Chair (2013). His e-mail address is Michael.Kuhl@rit.edu.

YUNAN LIU is an assistant professor in the Department of Industrial and Systems Engineering at North Carolina State University. He earned his Ph.D. in Operations Research from Columbia University. His research interests include queueing theory, stochastic modeling, simulation, applied probability, and their applications in call centers, healthcare, and manufacturing systems.

JAMES R. WILSON is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. His current research interests are focused on probabilistic and statistical issues in the design and analysis of simulation experiments, with special emphasis on applications in healthcare and production. As a WSC participant, he served as *Proceedings* Editor (1986), Associate Program Chair (1991), and Program Chair (1992). During the period 1997–2004, he was a member of the WSC Board of Directors. He is a member of ACM, ASA, and ASEE; and he is a Fellow of IIE and INFORMS. His e-mail address is jwilson@ncsu.edu, and his Web page is www.ise.ncsu.edu/jwilson.