Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

ESTIMATING SPLINE-BASED NONHOMOGENEOUS POISSON INTENSITIES USING CONSTRAINED QUADRATIC PROGRAMMING

Siqi Chen
Jing Yang (Sunny) Xi
Wai Kin (Victor) Chan

Tsinghua-Berkeley Shenzhen Institute, Shenzhen International Graduate School
Tsinghua University
Xili University Town
Shenzhen, Guangdong, 518071, China

ABSTRACT

This paper estimates the intensity function of a nonhomogeneous Poisson process (NHPP) using a spline-based method with constrained quadratic programming (CQP). Based on the property of B-splines, we transform the estimation problem into an optimization problem and apply CQP to obtain the estimated intensity function with low computational expense. Numerical experiments are conducted to verify the performance of our method. In addition, the impacts of the number of intervals from event-count data and the number of knots in B-splines are also discussed to explore the properties of spline-based models.

1 INTRODUCTION

Nonhomogeneous Poisson processes (NHPP) are widely used for modeling stochastic systems that exhibit non-steady event occurrences that are time-dependent. Their advantages over simple Poisson processes with constant rates are preferred in various scenarios, such as modeling incoming traffic stream (Bowman and Miller 2016) and simulating call center and hospital arrivals (Kim and Whitt 2014). An NHPP can be generalized from a homogeneous Poisson process by replacing the invariant arrival rate λ with an intensity function $\lambda(t)$ that varies with time t. Its integral function $\Lambda(t) = \int_0^t \lambda(s)ds$ where $t > 0$ represents the nondecreasing cumulative intensity function. The number of events that occur during the time interval $(t, t+\Delta t]$ follows a Poisson distribution with expected mean of $\Lambda(t+\Delta t) - \Lambda(t)$.

Building a realistic model necessitates the utilization of data-based stochastic simulations that strike a delicate balance between achieving high accuracy and minimizing computational costs. These simulations leverage extensive datasets to accurately capture the complexities and stochastic nature of real-world systems, allowing for the generation of realistic and representative models. By carefully optimizing the simulation techniques, researchers can obtain accurate results while ensuring computational efficiency, enabling practical implementation and analysis of the model. In order to improve the performance of simulation methods, efficient estimation strategies for the intensify function are necessary and have been extensive investigated. Parametric estimations for NHPP mainly relies on the maximum likelihood and least-squares approaches (Lee et al. 1991; Zheng and Glynn 2017; Morgan et al. 2019). These methods can provide efficient estimations with fewer data points and offer interpretable parameter estimates. However, they heavily rely on the correctness of the assumed parametric form and may struggle to capture complex patterns or irregularities in the intensity function. On the other hand, nonparametric methods provide flexibility and robustness by making fewer assumptions about the form of the intensity function. Although they may have higher computation complexity and produce less interpretable estimates, nonparametric models can capture complex patterns well and are less sensitive to model misspecification, which has been

the focus of several recent studies. Henderson (2003) estimates the piecewise-constant intensity function using event-count data and presents its asymptotic properties. Such theoretical results are developed to a systematic nonparametric estimation procedure and a simulation algorithm in Leemis (2004). Based on the mean-constrained premise (Chen and Schmeiser 2013), Nicol and Leemis (2014) introduced I-SMOOTH to modify the intensity function from the piecewise constant form to the piecewise linear form, which ensures the continuity of $\lambda(t)$. Moreover, the derivability and smoothness have also gained increasing attention. Max nonnegativity ordering piecewise-quadratic rate smoothing (MNO-PQRS) is proposed by Chen and Schmeiser (2017) to estimate $\lambda(t)$ in piecewise quadratic forms.

In some applications, higher degree of smoothness for the intensity function, for instance, twice differentiable everywhere, may be required. One of the ways to achieve this is to use splines, which are higher degree piecewise polynomials, but determining the optimal coefficients for higher degree polynomials can be computationally expensive. Morgan et al. (2019) consider spline functions as the basic functions to estimate the intensity function for NHPP. To simplify calculations, a trust region algorithm is employed to obtain near optimal coefficients; however, obtaining the hessian matrix and its inverse version of the penalised likelihood function still necessitates extensive computations. Therefore, the spline-based estimation model for NHPP intensities lack efficient techniques to reduce computation complexity.

This paper focuses on fitting the spline-based Poisson intensities using constrained quadratic programming (CQP), which can derive the optimal estimated intensity function with high accuracy and efficiency. We assume that the interval count data is given prior to estimation since the event-count data usually require less storage and can be more easily obtained in comparison with the arrival time data (the time data also can be aggregated to the event-count data). The rest of the paper is organized as follows. Section 2 describes the procedures for formulating the estimation problem for NHPP based on B-splines. In Section 3, the process of computing the optimal coefficients of the spline-based intensity function is transformed into a tractable CQP. Section 4 details the calculation of the integrals. Numerical experiments are implemented to verify the effectiveness of our methods in Section 5 and Section 6 concludes the paper.

2 FORMULATING THE B-SPLINE INTENSITIES

Suppose z independent and identically distributed (i.i.d.) nonhomogenous Poisson processes $N_1, N_2, \cdots, N_z$ over the time interval $[T_{init}, T_{end}]$ have been observed. To improve the smoothness of the estimated intensity function, we consider a spline-based intensity $\lambda(t)$ based on the sequence of ordered knots $\mathbf{T}_k = \{T_{-k}, T_{-k+1}, \cdots, T_{m-2}, T_{m-1}\}$, where k represents the degree of the spline function and $0 = T_{-k} \leq T_{-k+1} \leq \cdots \leq T_{m-2} \leq T_{m-1} = T_{end}$. The positions of these knots depend on the types of B-splines. For open uniform B-splines, all knots are distributed uniformly with a constant interval $\delta = T_i - T_{i-1}, i = -k+1, -k+2, \cdots, m-1$ (cf. Morgan et al. (2019)). As the uniform B-splines can exhibit overshoot or undershoot behavior at the boundaries, the clamped B-spline can be adopted instead to ensure the function reaches zero beyond the endpoints of the data smoothly. Clamped B-splines repeat the knots at the end points $k-1$ times, which is denoted by $T_{init} = T_{-k} = T_{-k+1} = \cdots = T_{-\lfloor k/2 \rfloor - 1} \leq T_{-\lfloor k/2 \rfloor} \leq \cdots \leq T_{m-\lfloor k/2 \rfloor + 1} = \cdots = T_{m-1} = T_{end}$ where $\lfloor \cdot \rfloor$ is the floor function. In this case, the interval becomes $\delta = T_i - T_{i-1}, i = -\lfloor k/2 \rfloor + 1, -\lfloor k/2 \rfloor + 2, \cdots, m - \lfloor k/2 \rfloor + 1$. For a normalized B-spline function $B_k(t)$, the intensity function $\lambda(t)$ is given by

$$\lambda(t) = \sum_{j=-k}^{m-1} c_j B_k\left(\frac{t - T_j}{\delta}\right),$$

where $c_i \in \mathbf{R}, i = -k, -k+1, \cdots, m-1$ are the weighting coefficients. More specifically, the normalized B-spline function of degree k is $B_k(t) = S_{k-l,k}(t-l)$ when $l \leq t < l+1$ for $l = 0, 1, \cdots, k$; otherwise it is equal to 0. In the spline function, $S_{l,k}(t-l)$ denotes the basis element. Based on the initialization

$S_{0,0}(t) = 1$, $S_{l,j}(t)$ can be calculated recursively using the following function

$$S_{l,j}(t) = \frac{j-l+t}{j} S_{l-1} j-1(t) + \frac{1+l-t}{j} S_{l,j-1}(t),$$

where $l = 1,2,\cdots,j-1$ and $j = 0,1,\cdots,k$. For $l = 0$ and $l = j$, the basic elements are obtained by $S_{0,j}(t) = \frac{1-t}{j} S_{0,j-1}(t)$ and $S_{j,j}(t) = \frac{t}{j} S_{j-1,j-1}(t)$, respectively.

Once the knot sequence $\mathbf{T}_k$ is determined, the $m+k$ B-splines are fixed. Since it is non-conducive to calculate each B-spline at a certain time point t through the traditional recursive algorithm, Kano et al. (2011) propose a transformation of the recursive algorithm into a matrix multiplication between a constant matrix M_k and a vector comprising polynomial terms $q_k(t)$. The basic element can be written by

$$S_k(t) = M_k q_k(t),$$

where

$$q_k(t) = [t^k, t^{k-1}, \cdots, 1]^T, S_k(t) = [S_{0,k}(t), S_{1,k}(t), \cdots, S_{k,k}(t)]^T.$$

Based on Kano et al. (2011), M_k can be obtained by

$$M_i = \frac{1}{i}([0_{i+1} \quad \Gamma_i M_{i-1}] + [\Delta_i M_{i-1} \quad 0_{i+1}]),$$

where $M_0 = 1$, and 0_{i+1} denotes a $(i+1)$-dimension zero vector. $\Gamma_i, \Delta_i \in R^{(i+1)\times i}$ are defined as

$$\Gamma_i = \begin{bmatrix} 1 & & & & & & \\ i-1 & 2 & & & & & \\ & i-2 & 3 & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & 1 & i \\ & & & & 0 & 0 \end{bmatrix}, \Delta_i = \begin{bmatrix} -1 & & & & & & \\ 1 & -1 & & & & & \\ & 1 & -1 & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & 1 & -1 \\ & & & & & 1 \end{bmatrix}.$$

Note that here the results from Kano et al. (2011) are directly applied to the estimation problem of NHPP. Since the $m+k$ B-splines are fixed after the knot sequence $\mathbf{T}_k$ is determined, the shape of a spline-based $\lambda(t)$ merely depends on the weighting coefficients. The following section will describe the detailed procedures of choosing optimal weighting coefficients c_i in our model according to the given event-count data.

3 ESTIMATING THE WEIGHTING COEFFICIENTS

Suppose that the numbers of arrivals over the time period $(T_{i-1}, T_i]$, $i = -k, -k+1, \cdots, m-1$ in h-th realization are recorded as $N_h(t_{i,0}, t_{i,1}], N_h(t_{i,1}, t_{i,2}], \cdots, N_h(t_{i,\zeta_{i-1}}, t_{i,\zeta_i}]$ where $N_h(t_{i,j-1}, t_{i,j}]$ represents the event counts in $(t_{i,j-1}, t_{i,j}], j = 1,2,\cdots,\zeta_i$. We use ζ_i to denote the number of equal time intervals from the dataset during $(T_{i-1}, T_i]$. Note that $(t_{i,j-1}, t_{i,j}]$ is one of the subintervals in $(T_{i-1}, T_i]$ and $T_{i-1} = t_{i,0} \leq t_{i,1} \leq \cdots \leq t_{i,\zeta_i-1} \leq t_{i,\zeta_i} = T_i$.

Since the weighting coefficient c_i should ensure the accuracy of the estimated intensity function, this section takes the ordinary least squares (OLS) into account to establish the cost function $W(c)$. The mean square error loss function is defined based on the mean constraint property of NHPP, which is $E(N_h(t_{i,j-1}, t_{i,j}]) = \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s) ds$. Hence, the first term in the objective function is

$$\beta_1'(c) = \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \frac{1}{z} \sum_{h=1}^{z} \left(N_h(t_{i,j-1}, t_{i,j}] - \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds \right)^2.$$

This term is inspired by the objective function in Zheng and Glynn (2017) that utilizes OLS on the piecewise linear intensities. We modify $\beta_1^{'}(c)$ to facilitate the construction of the quadratic cost function in our model.

Corollary 1 Minimizing $\beta_1^{'}(c)$ is equivalent to minimizing $\beta_1(c)$ which is

$$\beta_1(c) = \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \left(\frac{1}{z} \sum_{h=1}^{z} N_h(t_{i,j-1}, t_{i,j}] - \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds \right)^2.$$

Proof. See Appendix A □

To ensure smoothness and to prevent abrupt changes in the estimated intensity function, the square of second derivative is used to avoid over-fitting. This kind of penalization has been also adopted in spline-based intensities of NHPP with penalized likelihood method (Morgan et al. 2019). Thus, the second term in our objective function is

$$\beta_2(c) = \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \alpha \int_{t_{i,j-1}}^{t_{i,j}} \left(\lambda^{(2)}(s;c)\right)^2 ds,$$

where α represents the smoothing parameter. Considering both the accuracy and smoothness of the estimated intensities, we use the summation of the two terms as the cost function $W(c) = \beta_1(c) + \beta_2(c)$ where $c = [c_{-k}, c_{-k+1}, \cdots, c_{m-2}, c_{m-1}]^T$. For the purpose of computing the optimal c, this cost function can be rewritten as a quadratic function

$$W(c) = c^T G c - 2 g^T c + \gamma, \tag{1}$$

where

$$G = I_b \Phi I_b^T + \alpha Q,\ g = I_b \Phi \bar{N},\ \gamma = \bar{N}^T \Phi \bar{N},$$

with

$$\begin{aligned}
I_b &= \left[Ib^{(-k,1)}, Ib^{(-k,2)}, \cdots, Ib^{(-k,\zeta_{-k})}, \cdots, Ib^{(m-1,1)}, Ib^{(m-1,2)}, \cdots, Ib^{(m-1,\zeta_{m-1})} \right], \\
Ib^{(i,j)} &= \left[\int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(x - T_{-k}))dx, \int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(x - T_{-k+1}))dx, \cdots, \int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(x - T_{m-1}))dx \right]^T \\
\bar{N} &= \left[\frac{1}{z}\sum_{h=1}^{z} N_h(t_{-k,0}, t_{-k,1}], \frac{1}{z}\sum_{h=1}^{z} N_h(t_{-k,1}, t_{-k,2}], \cdots, \frac{1}{z}\sum_{h=1}^{z} N_h(t_{-k,\zeta_{-k}-1}, t_{-k,\zeta_{-k}}], \cdots, \right. \\
&\quad \left. \frac{1}{z}\sum_{h=1}^{z} N_h(t_{m-1,0}, t_{m-1,1}], \cdots, \frac{1}{z}\sum_{h=1}^{z} N_h(t_{m-1,1}, t_{m-1,2}], \cdots, \frac{1}{z}\sum_{h=1}^{z} N_h(t_{m-1,\zeta_{m-1}-1}, t_{m-1,\zeta_{m-1}}] \right].
\end{aligned}$$

Notice that Φ in (1) is a square diagonal matrix with $\sum_{i=-k}^{m-1} \zeta_i$ rows, and its diagonal element equals to $1/z$. The smoothness penalty term is represented by the gramian $Q = \int_{t_0}^{t_n} \frac{d^2 b(t)}{dt^2} \frac{d^2 b^T(t)}{dt^2} dt$. It can be approximated by using the results in Kano et al. (2005).

Based on the above definitions, each term of $W(c)$ except c is a constant when the interval count data is given. Quadratic programming techniques can be employed by minimizing the quadratic cost function to obtain the optimal value of c efficiently. With the optimal coefficient c^*, the estimated intensity function is expressed by

$$\lambda(t) = (c^*)^T b(t),$$

where

$$b(t) = [B_k(\frac{1}{\delta}(t - T_{-k})), B_k(\frac{1}{\delta}(t - T_{-k+1})), \cdots, B_k(\frac{1}{\delta}(t - T_{m-1}))]^T.$$

Since the intensity function is nonnegative everywhere, the solution searching range for c is limited to the interval $[0, \infty)$. Therefore, the whole estimation problem can be viewed as a constrained quadratic programming problem, which can be solved with relative ease.

4 CALCULATING INTEGRALS

We see that $Ib^{(i)}$ consists of the integrals of B-splines over $(t_{i,j-1}, t_{i,j}]$. Using the matrix multiplication version for computing the basic elements of B-splines, we can also denote their integral in matrix form by modifying the vector comprised of polynomial terms. The integral of each element in $q_k(t))$ is

$$\int_{t_{i,j-1}}^{t_{i,j}} q_k(t)dt = \begin{bmatrix} \frac{1}{k+1} & & & & \\ & \frac{1}{k} & & & \\ & & \ddots & & \\ & & & \frac{1}{2} & \\ & & & & 1 \end{bmatrix} \begin{bmatrix} t^k \\ t^{k-1} \\ \vdots \\ t \\ 1 \end{bmatrix} t|_{t_{i,j-1}}^{t_{i,j}} = D_k t q_k(t)|_{t_{i,j-1}}^{t_{i,j}}, \tag{2}$$

From (2), the integral of the basic element of B-splines can be derived by $\int_{t_{i,j-1}}^{t_{i,j}} S_k(t)dt = M_k D_k t q_k(t)|_{t_{i,j-1}}^{t_{i,j}}$. Each basis function has non-zero values over only a small interval of the curve, which is identified as the local support property in B-splines. Consequently, the integral of $B_k(\frac{1}{\delta}(s-T_l))$ is affected by different event-count datasets and selections of knots. The calculations of the B-spline integrals for $t \in (t_{i,j-1}, t_{i,j}]$ are divided into three cases.

1. **Case 1:** $0 \le \frac{1}{\delta}(t_{i,j-1} - T_l) < k+1$.
   Suppose that j_0 satisfies $j_0 \le \frac{1}{\delta}(t_{i,j-1} - T_l) < j_0 + 1$ and $\Delta T = \min\{\lfloor \frac{t_{i,j} - t_{i,j-1}}{\delta} \rfloor, m-l-j_0, k-j_0\}$. When $\Delta T > 0$, we have

$$\int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(s-T_l))ds = \sum_{\Delta T-1}^{\xi=0} [\delta M_k D_k q_k(1)]_{k-j_0+1-\xi} + [\delta M_k D_k(\frac{t_{i,j}-T_l}{\delta} - j_0 - \Delta T)$$
$$\times q_k(\frac{t_{i,j}-T_l}{\delta} - j_0 - \Delta T)]_{k-j_0-\Delta T+1} - [\delta M_k D_k(\frac{t_{i,j}-T_l}{\delta} - j_0) q_k(\frac{t_{i,j}-T_l}{\delta} - j_0)]_{k-j_0+1};$$

   When $\Delta T = 0$, $\lfloor \frac{t_{i,j} - t_{i,j-1}}{\delta} \rfloor = 0$ and

$$\begin{aligned} \int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(s-T_l))ds &= [\delta M_k D_k\big((\frac{t_{i,j}-T_l}{\delta} - j_0) q_k(\frac{t_{i,j}-T_l}{\delta} - j_0) \\ &\quad -(\frac{t_{i,j-1}-T_l}{\delta} - j_0) q_k(\frac{t_{i,j-1}-T_l}{\delta} - j_0)\big)]_{k-j_0+1}; \end{aligned}$$

   When $T = 0$, $\lfloor \frac{t_{i,j} - t_{i,j-1}}{\delta} \rfloor \neq 0$ and

$$\int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(s-T_l))ds = [\delta M_k D_k\big(q_k(1) - (\frac{t_{i,j-1}-T_l}{\delta} - j_0) q_k(\frac{t_{i,j-1}-T_l}{\delta} - j_0)\big)]_{k-j_0+1}.$$

2. **Case 2:** $\frac{1}{\delta}(t_{i,j-1} - T_l) \ge k+1$ or $\frac{1}{\delta}(t_{i,j} - T_l) \le 0$. We have

$$\int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(s-T_l))ds = 0.$$

3. **Case 3:** $\frac{1}{\delta}(t_{i,j-1} - T_l) < 0$.
   Suppose that $\Delta T = \min\{\lfloor \frac{t_{i,j} - t_{i,j-1}}{\delta} \rfloor, k, m-l\}$. When $\frac{t_{i,j} - T_l}{\delta} < k+1$,

$$\int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(s-T_l))ds = [\delta M_k D_k(\frac{t_{i,j}-T_l}{\delta} - \Delta T) q_k(\frac{t_{i,j}-T_l}{\delta} - \Delta T)]_{k-\Delta T+1} + \sum_{\xi=0}^{\Delta T-1} [\delta M_k D_k q_k(1)]_{k-\xi+1};$$

When $\frac{t_{i,j}-T_l}{\delta} \geq k+1$,

$$\int_{t_{i,j-1}}^{t_{i,j}} B_k(\frac{1}{\delta}(s-T_l))ds = \sum_{\xi=0}^{\Delta T} [\delta M_k D_k q_k(1)]_{k-\xi+1}$$

5 PERFORMANCE

This section presents numerical experiments to validate the effectiveness of our method and compare its performance with that of two alternative estimation algorithms, I-SMOOTH (Nicol and Leemis 2014) and MNO-PQRS (Chen and Schmeiser 2017), based on the interval counts data. I-SMOOTH considers a piecewise linear function as the basic form of intensities, but ignores the smoothness of the intensity functions. MNO-PQRS focuses on tackling the drawback of I-SMOOTH, and considers smoothing intensities on the basis of a piecewise quadratic function.

For spline-based model with CQP, different types of splines with different knot positions are also investigated. In the experiments, we consider both uniform splines and clamped splines. The former predetermines the knots embedded in splines as $T_{-3} \leq T_{-2} \leq T_{-1} \cdots \leq T_{m-2} \leq T_{m-1}$, where $T_i - T_{i-1}$ remains constant for any $i = -2, \cdots, m-1$. The latter designs the knots by $T_{-3} = T_{-2} \leq T_{-1} \cdots \leq T_{m-2} = T_{m-1}$, where all $T_i - T_{i-1}$'s have the same values for any $i = -1, \cdots, m-2$; there are $k-1$ points equal to the boundary points. For simplicity, the uniform splines and clamped spline-based methods are denoted by UBspline-CQP and CBspline-CQP, respectively.

To generate event-count data in numerical experiments, NHPPs are simulated using a thinning algorithm based on a true intensity function $\lambda^*(t)$. Four methods, UBspline-CQP, CBspline-CQP, I-SMOOTH and MNO-PQRS, are applied to the generated data to compute the estimated intensity function $\hat{\lambda}(t)$. Compared to other estimation methods, spline-based strategies are able to estimate intensities with higher continuity and smoothness. To validate this advantage, we consider a smooth cosinusoidal function $\lambda_1^*(t)$ given by

$$\lambda_1^*(t) = 8 - 5\cos(\frac{\pi}{4}t), t \in [0,16].$$

Furthermore, our simulations adopt a piecewise linear function $\lambda_2^*(t)$ to evaluate the performance of UBspline-CQP and CBspline-CQP on nonsmooth functions. This function is given by

$$\lambda_2^*(t) = \begin{cases} 5 + \frac{5}{4}t, & t \in [0,4), \\ 10 - \frac{5}{4}(t-4), & t \in [4,8), \\ 5 - \frac{5}{8}(t-8), & t \in [8,12), \\ \frac{5}{2} + \frac{5}{8}(t-12), & t \in [12,16). \end{cases}$$

Notice that $t_{-k,1} = 0$ and $t_{m-1,\zeta_{-k}} = 16$ in both cases.

The differences between the estimated function and the true intensity function are applied as metrics to measure the performance of a method. For the purpose of this study, we use the integrated absolute difference $L_1 = d_1(\hat{\lambda}(t), \lambda^*(t)) = \int_{T_{-k}}^{T_{m-1}} |\hat{\lambda}(t) - \lambda^*(t)| dt$ and the maximum absolute difference $L_\infty = d_2(\hat{\lambda}(t), \lambda^*(t)) = \max_{t \in [T_{-k}, T_{m-1}]} |\hat{\lambda}(t) - \lambda^*(t)|$. Fifty independent replications are conducted in experiments, and the final results are determined by calculating the averages of $L_1(\hat{\lambda}(t), \lambda^*(t))$ and $L_2(\hat{\lambda}(t), \lambda^*(t))$. Each replication is comprised of five realizations which can be regarded as five NHPP samples. The smoothing parameter is set to be $\alpha = 0.01$ in all simulations.

5.1 Comparison of the Four Methods

We first generate event-count data with 20 intervals by the thinning algorithm; then, the cubic spline-based methods: UBspline-CQP and CBspline-CQP (based on 20 knots), as well as the classical methods from

the literature: I-SMOOTH and MNO-PQRS, are all applied to estimate the intensity function. In Figure 1, each subplot showcases the true intensity function $\lambda_1^*(t)$ (black) and 50 estimated functions (colored) based on the generated data. The differences L_1 and L_∞ are presented in the titles of each subplot. The spline-based strategies perform better because of the smoothness and stability of their estimated functions. Both the averaged metrics for both the uniform splines and clamped splines are very similar and further illustrate the accuracy of our methods. On the other hand, the estimated intensities using I-SMOOTH and MNO-PQRS are more easily influenced by noise due to more frequent and intense fluctuations. This results in worse accuracy as seen in the larger averaged differences.

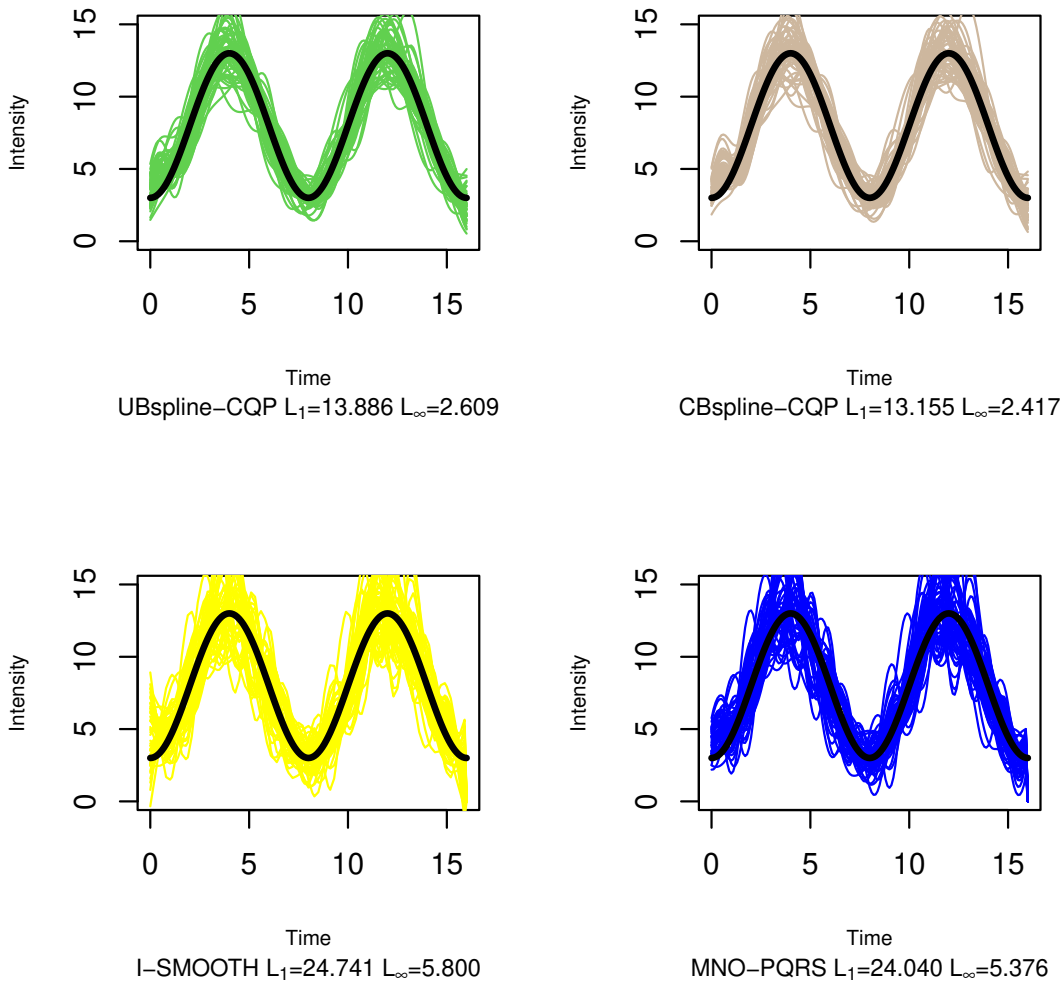

Figure 1: Estimated Rate Functions for $\lambda_1^*(t)$ and Averaged Metrics with Four Methods.

Next, we adopt the linear piecewise function $\lambda_2^*(t)$ as the true intensity function. Figure 2 illustrates the effectiveness of each method by simulation. Again, we see that both cubic spline-based intensities using CQP still outperform the two classical methods, that is, the results are consistent with those analyzed previously in Fig 1. This suggests that our methods are flexible and can be applied to different scenarios. Lastly, for both $\lambda_1^*(t)$ and $\lambda_2^*(t)$, CBspline-CQP is able to control the estimated curve at the boundary

points better than the UBspline-CQP, thereby preventing dramatic fluctuations. To be more specific, in the simulations of estimating $\lambda_2^*(t)$, the estimated intensities using UBspline-CQP ranges in $(3.51, 7.94)$ with a standard deviation of 1.15 at the initial time, and ranges in $(1.25, 9.51)$ with a standard deviation of 1.90 at the end time. CBspline-CQP method can have narrower ranges and smaller deviations at both boundaries compared to those of UBspline-CQP. Its estimated values are bounded in $(3.37, 7.35)$ with a standard deviation of 0.89 at the initial point, and bounded in $(1.27, 9.45)$ with a standard deviation of 1.89 at the end point. Similar findings can be observed in simulations when estimating $\lambda_1^*(t)$.

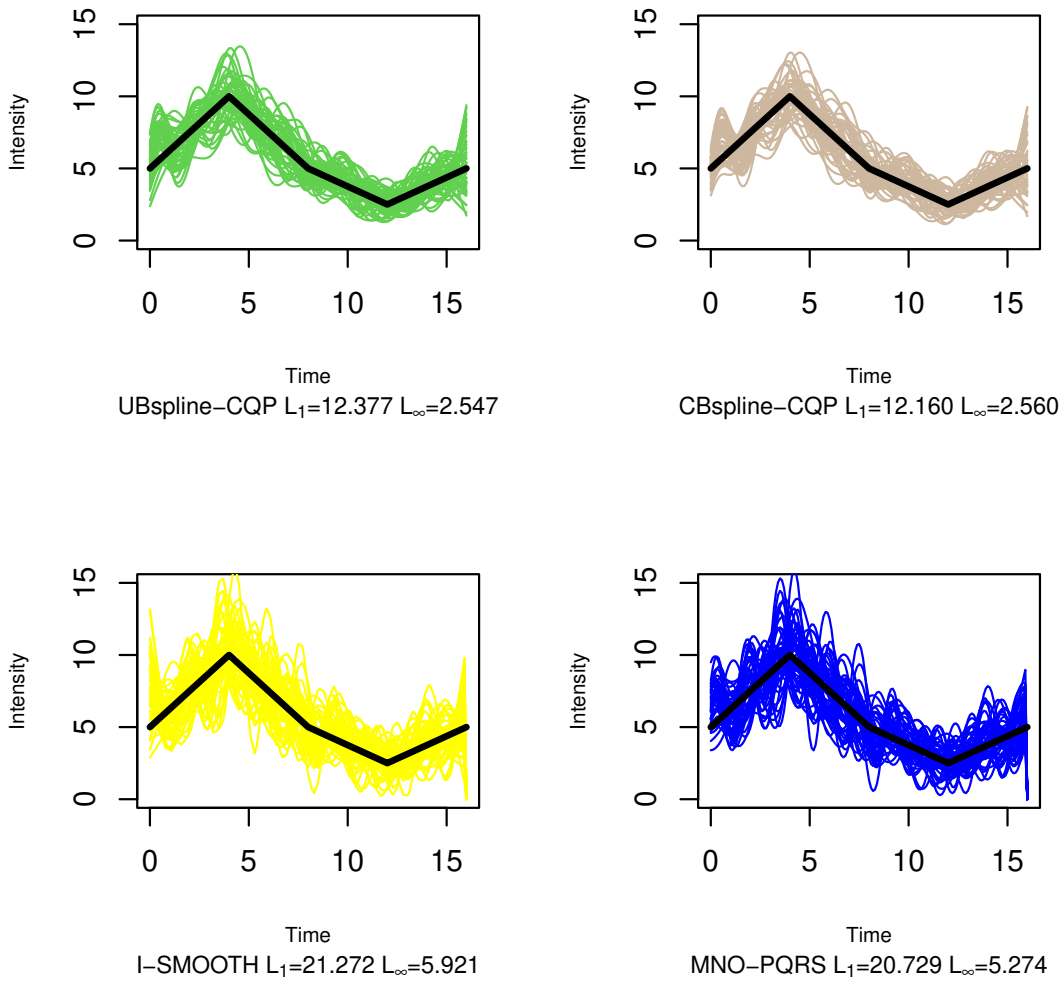

Figure 2: Estimated Rate Functions for $\lambda_2^*(t)$ and Averaged Metrics under Four Strategies.

To compare the computational efficiency of existing methods for intensity estimations, we need to first examine them in detail. I-SMOOTH and MNO-PQRS algorithms apply the matrix inversion to obtain the optimal estimated intensity functions. The time complexity of matrix inversion methods, such as Gaussian elimination or LU decomposition, is approximately $O(n^3)$, where n is the size of event-count data in one realization. Our spline-based model transforms the whole estimation into quadratic programming. The time complexity can vary depending on the problem size, desired accuracy, and the specific algorithm

employed. For example, interior-point methods and active-set methods typically have time complexities of around $O((n^3)log(\frac{1}{\varepsilon}))$ and $O(n^3)$ respectively, where ε is the desired accuracy. The existing spline-based method proposed by Morgan et al. (2019) uses the trust region algorithm to tackle a nonlinear optimization problem. Since our objective function is quadratic, the spline-CQP is one of the sub-problems in the trust region algorithm, which indicates our improvement of computational efficiency in spline-based methods. Although the time complexity in spline-CQP is higher than that of I-SMOOTH and MNO-PQRS, we can obtain better performance with similar time in medium- or small-scale problems. In our simulations, the averaged implementation times of UBspline-CQP, CBspline-CQP, I-SMOOTH and MNO-PQRS are 2.13s, 1.33s, 0.04s and 0.03s, respectively. But for large-scale dataset, there exists trade off between accuracy and computational efficiancy among all models.

5.2 Effects of NTI and NK

From previous results, we infer that inappropriate bin-size of the event-count data may causes the I-SMOOTH and MNO-PQRS algorithms to estimate intensity functions with erratic fluctuations and poor smoothness. In this subsection, we investigate how each estimation method is affected by the bin-size. When the number of knots in the splines is NK = 10, we set the number of intervals to be NTI = 10, 15 and 20. Table 1 reports the integrated absolute difference and the maximum absolute difference for each parameter combination. In most cases, our spline-based methods demonstrate better performance than the two other methods regardless of which intensity function is used; moreover, thinner bin-size or more intervals in data can improve the accuracy of estimation when NK is fixed to 10. In contrast, the performances of I-SMOOTH and MNO-PQRS are very sensitive to NTI. Excessively thin intervals of the event count data can result in significant deviation from the true intensity function. Recently, some studies (cf. Chen and Schmeiser (2018) and Chen and Schmeiser (2019)) have observed this phenomenon when using nonparametric techniques, i.e., I-SMOOTH and MNO-PQRS, and discuss the optimal choice of bin-size in interval counts data.

Table 1: The Averaged Metrics under Different Numbers of Bins in Interval Counts Data.

			UBspline-CQP		CBspline-CQP		I-SMOOTH		MNO-PQRS	
NK	$\lambda^*(t)$	NTI	L_1	L_∞	L_1	L_∞	L_1	L_∞	L_1	L_∞
10	$\lambda_1^*(t)$	10	13.765	2.449	16.275	2.590	15.767	3.291	14.946	3.109
		15	12.728	2.405	15.481	2.527	18.629	3.987	17.990	3.529
		20	12.085	2.340	15.470	2.472	20.237	4.307	19.550	4.145
	$\lambda_2^*(t)$	10	15.725	3.999	16.832	4.562	11.985	5.026	11.683	5.000
		15	14.361	4.015	15.961	4.557	14.598	5.093	14.198	5.000
		20	13.796	4.059	16.236	4.546	16.661	5.178	16.262	5.020

Although the number of intervals has little impact on the accuracy of UBspline-CQP and CBspline-CQP, different numbers of knots might create considerable variations in the averaged metrics. Therefore, we study the effects of NK and the interactions between NTI and NK. Figure 3 and Figure 4 show the integrated difference with varying NTI and NK using UBspline-CQP and CBspline-CQP, respectively. In the NHPP simulations based on the true intensity $\lambda_2^*(t)$, the NTI changes from 34 to 70, while the NK varies within $[12,30]$. We see that the selection of NK can affect the performances of spline-based approaches in that too few knots lack the flexibility to estimate relatively intricate intensity functions, while too many knots may result in enormous deviations due to noise points. It can also be observed that NTI has a relatively small impact on the accuracy of estimators. Both UBspline-CQP and CBspline-CQP can find an optimal and several suboptimal combinations of NTI and NK, but they produce different patterns of interactions between the two variables. For example, UBspline-CQP performs best when (NTI,NK)=(42,16), while CBspline-CQP obtains the optimal accuracy when (NTI,NK)=(70,18). In addition, considering the larger

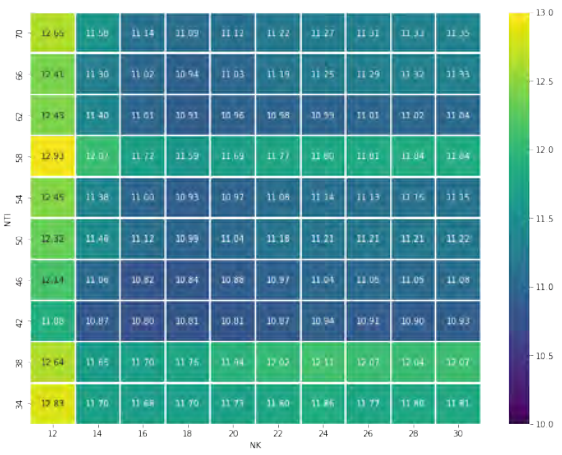

Figure 3: The Heatmap of L_1 Using UBspline-CQP with Varying NTI and NK.

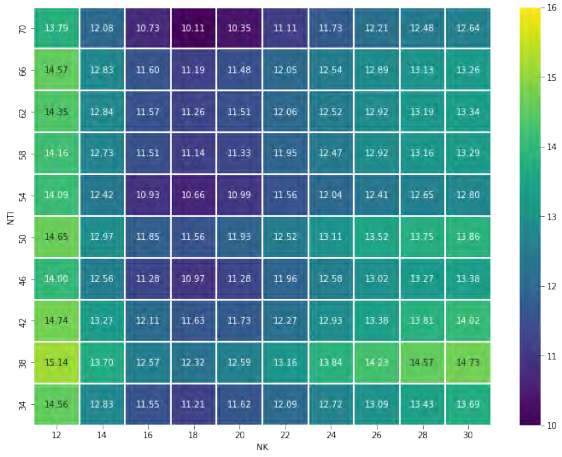

Figure 4: The Heatmap of L_1 Using CBspline-CQP with Varying NTI and NK.

variation span of integrated distance in CBspline-CQP, we find that its performance can be affected by the choices of NTI and NK more easily compared to UBspline-CQP.

5.3 Discussions with Existing Spline-Based Method

Spline-based strategies increase the flexibility and continuity of the intensity function in NHPP. With reasonable smoothing coefficient, NTI and NK, the accuracy and robustness of intensity estimations can be improved. To the best of our knowledge, the existing spline-based method is the one introduced by Morgan et al. (2019). We will summarize the differences and connections between it and our proposed approach.

First, our spline-CQP model has higher computation efficiency than the existing spline-based method. By minimizing the square error and considering some transformations, our objective function becomes quadratic and can be easily tackled by quadratic programming algorithms. Morgan et al. (2019) formulates the objective function as a nonlinear log-likelihood function without closed-form solutions. They apply trust region algorithm which approximates the objective function as a quadratic form locally in each step; yet the computational complexity of our model is only that of a subproblem in the very same algorithm. Second, compared to the existing spline-based method (with only arrival time data), our spline-CQP can be implemented based on more general-form data (both event-count data and arrival time data). Finally, Morgan et al. (2019) puts forward an algorithm to compute the optimal parameters of spline-based intensities and the smoothing coefficient by minimizing the RIC score. However, our model does not discuss the optimal choice of the smoothing coefficient, which will be investigated in future work.

6 CONCLUSION

In this paper, we apply constrained quadratic programming to estimate the spline-based intensity function of NHPPs and present detailed theoretical derivations to transform the estimation problem into an optimization problem. Through numerical experiments, we show that this method is computationally inexpensive and provides accurate estimations of the intensity function, even outperforming the I-SMOOTH and MNO-PQRS algorithms in most cases. We also observe that there exist optimal combinations of NTI and NK in both UBspline-CQP and CBspline-CQP, and the performance of CBspline-CQP can be affected more easily by the two variables. Lastly, the differences and connections between our spline-CQP model and the existing spline-based method on computational efficiency, input data types and optimization strategies are also discussed.

ACKNOWLEDGMENTS

This research was funded by the Science and Technology Innovation Committee of Shenzhen-Platform and Carrier (International Science and Technology Information Center), the Science and Technology Innovation Commission of Shenzhen (JCYJ20210324135011030), Guangdong Pearl River Plan (2019QN01X890), National Natural Science Foundation of China (Grant No. 71971127), and Shenzhen International Graduate School, Tsinghua University (JC2021004).

A PROOFS

The proof of corollary 1 is as follows.

$$
\begin{aligned}
& \min_c \beta_1^{'}(c) \\
= & \min_c \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \frac{1}{z} \sum_{h=1}^{z} \left(N_h(t_{i,j-1}, t_{i,j}] - \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds \right)^2 \\
= & \min_c \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \frac{1}{z} \sum_{h=1}^{z} \left(N_h^2(t_{i,j-1}, t_{i,j}] - 2N_h(t_{i,j-1}, t_{i,j}] \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds + (\int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds)^2 \right) \\
= & \min_c \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \left(\frac{1}{z} \sum_{h=1}^{z} N_h^2(t_{i,j-1}, t_{i,j}] - 2 \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds (\frac{1}{z} \sum_{h=1}^{z} N_h(t_{i,j-1}, t_{i,j}]) + (\int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds)^2 \right) \\
= & \min_c \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \left((\frac{1}{z} \sum_{h=1}^{z} N_h(t_{i,j-1}, t_{i,j}])^2 - 2 \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds (\frac{1}{z} \sum_{h=1}^{z} N_h(t_{i,j-1}, t_{i,j}]) + (\int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds)^2 \right) \\
= & \min_c \sum_{i=-k}^{m-1} \sum_{j=1}^{\zeta_i} \left(\frac{1}{z} \sum_{h=1}^{z} N_h(t_{i,j-1}, t_{i,j}] - \int_{t_{i,j-1}}^{t_{i,j}} \lambda(s;c) ds \right)^2 \\
= & \min_c \beta_1(c).
\end{aligned}
$$

Therefore, minimizing $\beta_1^{'}(c)$ is equivalent to minimizing $\beta_1(c)$. It is reasonable to construct a cost function including $\beta_1(c)$ and obtain the optimal parameter vector c by minimizing it.

REFERENCES

Bowman, C. N., and J. A. Miller. 2016. "Modeling Traffic Flow using Simulation and Big Data Analytics". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 1206–1217. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, H., and B. Schmeiser. 2013. "I-SMOOTH: Iteratively Smoothing Mean-Constrained and Nonnegative Piecewise-Constant Functions". *INFORMS Journal on Computing* 25(3):432–445.

Chen, H., and B. Schmeiser. 2017. "MNO-PQRS: Max Nonnegativity Ordering - Piecewise-Quadratic Rate Smoothing". *ACM Transactions on Modeling and Computer Simulation* 27(3):1–19.

Chen, H., and B. Schmeiser. 2018. "MISE-Optimal Grouping of Point-Process Data with a Constant Dispersion Ratio". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1563–1574. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, H., and B. W. Schmeiser. 2019. "MISE-Optimal Intervals for MNO–PQRS Estimators of Poisson Rate Functions". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo., P. Haas, and Y.-J. Son, 368–379. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Henderson, S. G. 2003. "Estimation for Nonhomogeneous Poisson Processes from Aggregated Data". *Operations Research Letters* 31(5):375–382.

Kano, H., H. Fujioka, and C. F. Martin. 2011. "Optimal Smoothing and Interpolating Splines with Constraints". *Applied Mathematics and Computation* 218(5):1831–1844.

Kano, H., H. Nakata, and C. F. Martin. 2005. "Optimal Curve Fitting and Smoothing using Normalized Uniform B-splines: A Tool for Studying Complex Systems". *Applied Mathematics and Computation* 169(1):96–128.

Kim, S.-H., and W. Whitt. 2014. "Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes?". *Manufacturing & Service Operations Management* 16(3):464–480.

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.

Morgan, L. E., B. L. Nelson, A. C. Titman, and D. J. Worthington. 2019. "A Spline-Based Method for Modelling and Generating a Nonhomogeneous Poisson Process". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo., P. Haas, and Y.-J. Son, 356–367. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nicol, D. M., and L. M. Leemis. 2014. "A Continuous Piecewise-Linear NHPP Intensity Function Estimator". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. J. Buckley, and J. A. Miller, 498–509. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Zheng, Z., and P. W. Glynn. 2017. "Fitting Continuous Piecewise Linear Poisson Intensities via Maximum Likelihood and Least Squares". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. H. Page, 1740–1749. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

SIQI CHEN is a master's student at the Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, China. Her primary research interests include input modeling, stochastic process, parameter estimations and missing data effects on social networks. Her email address is Chensq22@mails.tsinghua.edu.cn.

JING YANG (SUNNY) XI is a Ph.D. Candidate at the Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, China. His primary research interests include agent-based simulation, discrte-event simulation, and their applications in operations management. His email address is sunnyx@berkeley.edu.

WAI KIN (VICTOR) CHAN is Professor of the Tsinghua-Berkeley Shenzhen Institute (TBSI) and Shenzhen International Graduation School (SIGS), Tsinghua University, China. His research interests include discrete-event simulation, agent-based simulation, big-data analytics and their applications in social networks, business big-data, service systems, healthcare, transportation, energy markets, and manufacturing. His email address is chanw@sz.tsinghua.edu.cn. (corresponding author)