Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

FITTING CONTINUOUS PIECEWISE LINEAR POISSON INTENSITIES VIA MAXIMUM LIKELIHOOD AND LEAST SQUARES

Zeyu Zheng
Peter W. Glynn

Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, USA

ABSTRACT

We investigate maximum likelihood (ML) and ordinary least squares (OLS) methods to fit a continuous piecewise linear (PL) intensity function for non-homogeneous Poisson processes. The estimation procedures are formulated as convex optimization problems that are highly tractable. We also study the model misspecification issues in settings where the point process is non-Poisson or the underlying intensity is not piecewise linear. The performances of ML and OLS estimators are exhibited through a computational study, with both simulated data and real data from a large U.S. bank call center.

1 INTRODUCTION

Many applications settings require simulation models that are fed by point processes that exhibit non-stationarities. In particular, in many service contexts, the point process describes customer, job, or order arrival epochs, and the non-stationarity is induced by time-of-day or day-of-week effects. Such applications lead naturally to consideration of non-homogeneous Poisson processes (NHPPs). A NHPP is a point process $N = (N(t) : t \geq 0)$ having independent and Poisson-distributed increments, for which $EN(t) = \Lambda(t)$, where $\Lambda(\cdot)$ is a non-decreasing deterministic function. When $\Lambda(\cdot)$ can be expressed in the form $\Lambda(t) = \int_0^t \lambda(s)ds$ $\lambda(\cdot)$ is called the *intensity (or rate) function* of the process, and $\Lambda(\cdot)$ is known as the *integrated (or cumulative) intensity function*.

In this paper, we consider settings in which we have available observed data from which to estimate the intensity function of the NHPP. Specifically, we assume we have observed n independent realizations of the point process over a day (when time-of-day effects are the key non-stationarity to be modeled) or a week (when weekly effects need to be incorporated). When n is very large, one can statistically estimate the intensity via non-parametric methods; see Henderson (2003) for a specific implementation that is reasonably easy to simulate. For small or moderate values of n, parametric models may be more suitable. The most widely used such models in practice involve subdividing the time interval in question (a day or a week) into sub-intervals, say p of them, and then fitting a constant rate intensity over each of the p sub-intervals. Such a fitted intensity seems aesthetically unreasonable, given its implausible discontinuities, and is likely to predict poorly at times that are close to the sub-interval boundaries.

This suggests consideration of the next simplest parametric model, namely one in which the intensity is linear over each sub-interval, and continuous at the sub-interval boundaries. We call such an intensity a *piecewise linear* (PL) intensity. Such a model comes at a modest cost in terms of the number of parameters to be estimated relative to the piecewise constant version. In fact, if the intensity is assumed to be periodic (as may occur when the facility is open 24 hours per day or is open 7 days per week), then the PL model involves precisely the same number of parameters as for the commonly used piecewise constant model, namely p. We also note that the simulation of NHPPs with PL intensities is only marginally more complicated than for piecewise constant intensities; see Klein and Roberts (1984) for a discussion of an efficient implementation.

One challenge with the PL model is that there is no closed form for the estimators of the parameters, because it can be that the required non-negativity constraints will become active when computing the estimated intensity that best fits the observed data. Thus, the problem of computing the PL estimator that best fits a given data set will involve numerically solving a constrained optimization problem.

This paper studies the optimization problem that arises when maximum likelihood (ML) is applied to estimation of a PL intensity. We consider the problem both when the point process arrivals are fully observed (so that each "customer" arrival is observed) and when only interval count data is available. Section 2 shows that the optimization formulation leads to a very tractable convex programming problem. In Section 3, we discuss the use of ordinary least squares (OLS) as an alternative to ML estimation of PL intensities, and discuss and contrast the estimation properties of these different estimators. Both of these methods are based on specific statistical assumptions, and carry the advantages that a statistical formulation implies. In particular, such models permit the construction of confidence intervals, goodness-of-fit tests, and hypothesis tests.

One can also consider PL intensity fitting methods that start from observed data, but are not based upon a statistical formulation of the problem. For example, one can choose to compute the PL estimator without any non-negativity constraints, and to dynamically increase the number of sub-intervals (if needed) until the unconstrained problem solves with a non-negative intensity. In such approaches, one is free to also add a penalty term to the objective function that measures the model fit. Such penalty terms can serve to enforce "smoothness" of the solution (by, for example, penalizing large "second derivatives," as evidenced by large changes in the slope of the PL intensity from one interval to the next). The papers of Chen and Schmeiser (2013) and Nicol and Leemis (2014) follow this approach. Methods on piecewise linear cumulative intensity fitting are discussed in Leemis (1991) and Leemis (2004).

Our main contributions in this paper are:

1. Development of a convex programming formulation of the maximum likelihood estimation problem for estimating a PL intensity, as well as associated code for computing it (Section 2);
2. Development of an ordinary least squares method for estimating a PL intensity (Section 3), as well as related code;
3. Study of model mis-specification issues for ML and OLS in settings where the point process is non-Poisson or the true underlying intensity is not PL (Section 4).
4. A computational study of the ML and OLS estimators, and their performance relative to previous PL estimators introduced by Chen and Schmeiser (2013) and Nicol and Leemis (2014) (Section 5).

As noted earlier, a principal reason for use of statistically formulated models is the ability to construct confidence intervals, goodness-of-fit tests, and hypothesis tests. These ideas will be explored in greater detail in Zheng and Glynn (2017).

2 FITTING PIECEWISE LINEAR INTENSITIES VIA MAXIMUM LIKELIHOOD

Suppose that we wish to estimate the true intensity $\lambda^*(\cdot)$ underlying the NHPP N over the interval $[0,t]$. We presume that λ^* is linear over each of the intervals $(t_0,t_1],(t_1,t_2],\cdots,(t_{p-1},t_p]$, where $0=t_0<\cdots<t_p=t$, and is continuous over $[0,t]$. In particular, if $y_i^*=\lambda^*(t_i)$, it follows that $\lambda^*(s)=\lambda(y^*;s)$, where $y^*=(y_0^*,y_1^*,\ldots,y_p^*)$ and

$$\lambda(y;s)=y_{i-1}+\left(\frac{y_i-y_{i-1}}{t_i-t_{i-1}}\right)(s-t_{i-1}) \tag{1}$$

for $t_{i-1}\leq s\leq t_i$, $1\leq i\leq p$ and any $y=(y_0,\ldots,y_p)$. We start by considering the case where λ^* is periodic, so that $y_0^*=y_p^*$ and assuming that the t_i's are known. In this case, the model for λ^* has p parameters $y_1^*,\ldots,y_p^*$ that must be estimated from the data.

We presume that we have available n independent and identically distributed (iid) observations of the point process N, call them $N_1, N_2, \ldots, N_n$. We start with the case in which the N_i's are fully observed, so that the corresponding arrival times are known. The ML optimization problem can be expressed in terms of the (normalized) log-likelihood $\mathscr{L}_n(\cdot)$ given by

$$\mathscr{L}_n(y) = \frac{1}{n}\sum_{i=1}^{n}\int_{(0,t]}\log\lambda(y;s)N_i(ds) - \int_0^t \lambda(y;s)ds. \tag{2}$$

The ML optimization problem can then be expressed as

$$\begin{aligned} \max_{y_0, y_1, \ldots, y_p} \quad & \mathscr{L}_n(y_0, y_1, \ldots, y_p) \\ \text{s.t.} \quad & y_0 = y_p \\ & y_i \geq 0, 0 \leq i \leq p. \end{aligned} \tag{3}$$

The ML estimator $\hat{y}_n$ for y^* is, of course, the maximizer of (3).

It is easily seen that $-\mathscr{L}_n(\cdot)$ is convex, subject to a linear equality constraint and $p+1$ linear inequality constraints. This structure makes the problem computationally tractable, even for large data sets and moderate values of p. We have written an R/MATLAB code implementation that is available online; see http://poisson.zhengzeyu.com. In our experiments, this code can successfully solve problems involving tens of thousands of observations with up to hundreds of parameters; see also Grant and Boyd (2014).

To gain some insight into the convex programming problem (3), we note that we can eliminate the variable y_0, since $y_0 = y_p$. It follows that when $\hat{y}_n = (\hat{y}_{n1}, \ldots, \hat{y}_{np})$ is an interior solution, so that $\hat{y}_{ni} > 0$ for $1 \leq i \leq p$, then $\hat{y}_n$ is a root of

$$\frac{\partial}{\partial y_i}\mathscr{L}(\hat{y}_n) = 0$$

for $1 \leq i \leq p$. This implies that

$$\frac{1}{n}\sum_{j=1}^{n}\int_{(t_i,t_{i+1}]}\left(\frac{t_{i+1}-s}{t_{i+1}-t_i}\right)\left(\frac{N_j(ds)}{\lambda(\hat{y}_n;s)} - ds\right) + \frac{1}{n}\sum_{j=1}^{n}\int_{(t_{i-1},t_i]}\left(\frac{s-t_{i-1}}{t_i-t_{i-1}}\right)\left(\frac{N_j(ds)}{\lambda(\hat{y}_n;s)} - ds\right) = 0 \tag{4}$$

for $1 \leq i \leq p-1$ and

$$\frac{1}{n}\sum_{j=1}^{n}\int_{(t_0,t_1]}\left(\frac{t_1-s}{t_1-t_0}\right)\left(\frac{N_j(ds)}{\lambda(\hat{y}_n;s)} - ds\right) + \frac{1}{n}\sum_{j=1}^{n}\int_{(t_{p-1},t_p]}\left(\frac{s-t_{p-1}}{t_p-t_{p-1}}\right)\left(\frac{N_j(ds)}{\lambda(\hat{y}_n;s)} - ds\right) = 0. \tag{5}$$

We note that the p equations above then uniquely determine the p values $\hat{y}_{n1}, \ldots, \hat{y}_{np}$ when the number of observed arrivals in each sub-interval is non-zero. This uniqueness holds more generally, even in the setting where the constraints are active.

PROPOSITION 1 Suppose that the number of observed arrivals in each sub-interval is non-zero (i.e., $\sum_{k=1}^{n} N_k(t_{i-1}, t_i] > 0$ for $i = 1, 2, \ldots, p$) and that the total number of observed arrivals is larger than p (i.e., $\sum_{k=1}^{n} N_k(0,t] > p$). Then, the ML optimization problem (3) solves uniquely.

For the proof, see Zheng and Glynn (2017).

We further observe that we can multiply equation i in (4) by $\hat{y}_{ni}$ and (5) by $\hat{y}_{np}$ and sum them. This yields the conclusion that

$$\frac{1}{n}\sum_{j=1}^{n}\int_{(0,t]}N_j(ds) = \int_{(0,t]}\lambda(\hat{y}_n;s)ds, \tag{6}$$

so that the estimated intensity has the same integral (or associated area) as does the data. We call this the global *equal area property* of the ML estimator. This is the sample analog of the "mean-constrained" intensity function concept introduced by Chen and Schmeiser (2013).

The global equal area property (EAP) may be generalized to settings in which the non-negativity constraints are binding at $\hat{y}_n$.

PROPOSITION 2 Any maximizer $\hat{y}_n$ of (3) satisfies the global EAP.

Proof. We note that $r\hat{y}_n$ is feasible for $r \geq 0$, since $\hat{y}_n$ is. Furthermore, $\mathscr{L}_n(\hat{y}_n) \geq \mathscr{L}_n(r\hat{y}_n)$ for $r \geq 0$, since $\hat{y}_n$ is a maximizer. Note that $\lambda(r\hat{y}_n;\cdot) = r\lambda(\hat{y}_n;\cdot)$, so

$$\mathscr{L}_n(\hat{y}_n) - \mathscr{L}_n(r\hat{y}_n) = -\log r \cdot \frac{1}{n}\sum_{i=1}^{n}\int_{(0,t]} N_i(ds) + (r-1)\int_{(0,t]} \lambda(\hat{y}_n;s)ds.$$

Since $r^* = 1$ is a minimizer of $\mathscr{L}(\hat{y}_n) - \mathscr{L}(r\hat{y}_n)$ over $r \geq 0$, $\frac{d}{dr}\mathscr{L}(r\hat{y}_n)|_{r=1} = 0$, yielding (6). □

We turn next to the setting in which the observed data are interval count data. Here, the observations from N_j on $(t_{i-1}, t_i]$ consist of interval counts $N_j(s_{i0}, s_{i1}], \ldots, N_j(s_{im_{i-1}}, s_{im_i}]$ for $t_{i-1} = s_{i0} < s_{i1} < \cdots < s_{im_i} = t_i$, where $N_j(a,b]$ is the number of observations associated with N_j on the interval $(a,b]$. Note that m_i is the number of subinterval observations between slope changes $(t_{i-1}, t_i]$ and $t_{i-1} = s_{i0} < s_{i1} < \cdots < s_{im_i} = t_i$ are the boundaries of these subintervals. This changes the log-likelihood that enters the ML optimization to the interval log-likelihood $\mathscr{L}_n^I(\cdot)$ given by

$$\mathscr{L}_n^I(y) = \sum_{i=1}^{p}\sum_{j=1}^{m_i}\log\left(\frac{\lambda(y;s_{i,j-1}) + \lambda(y;s_{ij})}{2}\right)\frac{1}{n}\sum_{k=1}^{n} N_k(s_{i,j-1}, s_{ij}] - \int_{(0,t]} \lambda(y;s)ds.$$

Of course, here the ML optimization problem becomes

$$\begin{aligned} \max_{y_0,y_1,\ldots,y_p} \quad & \mathscr{L}_n^I(y_0, y_1, \ldots, y_p) \\ \text{s.t.} \quad & y_0 = y_p \\ & y_i \geq 0,\, 0 \leq i \leq p. \end{aligned} \tag{7}$$

Again, the ML estimator $\hat{y}_n$ is the maximizer of (7). As in the fully observed setting, this problem can be viewed as a convex optimization problem with linear constraints, and the R/MATLAB code discussed earlier in this section is designed to handle this type of interval count data as well.

A particularly simple case is when $m_i = 1$ for $1 \leq i \leq p$, so that the count data corresponds precisely to the PL sub-intervals that are used to define $\lambda(y;\cdot)$. In this case, $\lambda(y;s_{i0}) = y_{i-1}$ and $\lambda(y;s_{im_i}) = y_i$. The analogs to the gradient conditions (4) and (5) are then

$$\frac{1}{\hat{y}_{n,i-1} + \hat{y}_{ni}} \cdot \frac{1}{n}\sum_{k=1}^{n} N_k(t_{i-1}, t_i] - \frac{1}{2}(t_i - t_{i-1}) + \frac{1}{\hat{y}_{ni} + \hat{y}_{n,i+1}} \cdot \frac{1}{n}\sum_{k=1}^{n} N_k(t_i, t_{i+1}] - \frac{1}{2}(t_{i+1} - t_i) = 0 \tag{8}$$

for $1 \leq i \leq p-1$ and

$$\frac{1}{\hat{y}_{n,p-1} + \hat{y}_{n0}} \cdot \frac{1}{n}\sum_{k=1}^{n} N_k(t_{p-1}, t_p] - \frac{1}{2}(t_p - t_{p-1}) + \frac{1}{\hat{y}_{n0} + \hat{y}_{n1}} \cdot \frac{1}{n}\sum_{k=1}^{n} N_k(t_0, t_1] - \frac{1}{2}(t_1 - t_0) = 0. \tag{9}$$

When the interval counts are all positive, (8) and (9) together imply that

$$\frac{1}{\hat{y}_{n,i-1} + \hat{y}_{ni}} \cdot \frac{1}{n}\sum_{k=1}^{n} N_k(t_{i-1}, t_i] - \frac{1}{2}(t_i - t_{i-1}) = 0$$

for $1 \le i \le p$, so that

$$\frac{1}{n}\sum_{k=1}^{n} N_k(t_{i-1},t_i] = \frac{\hat{y}_{n,i-1}+\hat{y}_{ni}}{2}(t_i - t_{i-1}) = \int_{t_{i-1}}^{t_i} \lambda(\hat{y}_n;s)ds$$

for $1 \le i \le p$. In other words, the ML estimator $\hat{y}_n$ then has the property that the integral of $\lambda(\hat{y}_n,\cdot)$ over each sub-interval $(t_{i-1},t_i]$ agrees with the observed count intensity for that sub-interval. When this holds, we call this the *local equal area property* (or local EAP).

We conclude this section by noting that in some service contexts, one need not require periodicity. For example, if a service facility is only open for part of each day, the intensity need not be periodic. In this setting, the linear equality constraint $y_0 = y_p$ is dropped from (3) in the fully observed context and (7) in the interval count setting, so that the model now has $p+1$ free statistical parameters (rather than p).

Again, the ML optimization problems can be solved as convex optimization problems subject to linear inequality constraints; our R/MATLAB code covers these settings as well. The question of whether these problems have unique maximizers is again of importance.

PROPOSITION 3 Suppose that the number of observed arrivals in each sub-interval is non-zero (i.e., $\sum_{k=1}^{n} N_k(t_{i-1},t_i] > 0$ for $i = 1,2,\ldots,p$) and that $\sum_{i=1}^{p}\sum_{j=1}^{m_i} I(\sum_{k=1}^{n} N_k(s_{i,j-1},s_{ij}] > 0) > p$. Then, the ML optimization problem (7) in the interval count setting solves uniquely.

For the proof, see Zheng and Glynn (2017).

3 FITTING PIECEWISE LINEAR INTENSITIES BY ORDINARY LEAST SQUARES

In this section, we discuss an alternative estimation approach, based on ordinary least squares (OLS), for fitting a PL intensity to observed data. We focus exclusively on the setting in which only interval count data are available. Specifically, we consider the following objective function

$$\beta_n(y) \triangleq \sum_{i=1}^{p}\sum_{j=1}^{m_i}\frac{1}{n}\sum_{k=1}^{n}\left(N_k(s_{i,j-1},s_{ij}] - \int_{(s_{i,j-1},s_{ij}]}\lambda(y;s)ds\right)^2.$$

Assuming that the intensity $\lambda^*(\cdot)$ is periodic, the OLS optimization problem is then given by

$$\begin{aligned} \min_{y_0,y_1,\ldots,y_p} \quad & \beta_n(y_0,y_1,\ldots,y_p) \\ \text{s.t.} \quad & y_0 = y_p \\ & y_i \ge 0,\ 0 \le i \le p. \end{aligned} \tag{10}$$

The OLS estimator $\hat{y}_n^{LS}$ is the minimizer of (10), assuming it is unique. An important special case arises when $m_i = 1$ for $1 \le i \le p$. In that case, one can make $\nabla_y \beta_n(\tilde{y}) = 0$ at any $\tilde{y}$ for which the local EAP

$$\left(\frac{\tilde{y}_{i-1}+\tilde{y}_i}{2}\right)(t_i - t_{i-1}) = \frac{1}{n}\sum_{k=1}^{n} N_k(t_{i-1},t_i] \tag{11}$$

holds for $1 \le i \le p$. If we view (11) as a linear system in the unknown $\tilde{y}_i$'s, it turns out that the y_i's are uniquely determined when p is odd. Specifically,

$$\tilde{y}_i = \frac{1}{2}\sum_{j=i+1}^{p} b_{nj}(-1)^{j-i-1} + \frac{1}{2}\sum_{j=1}^{i} b_{nj}(-1)^{p+j-i-1} I(i \ge 1)$$

for $0 \le i \le p-1$ with $\tilde{y}_p = \tilde{y}_0$, where $b_{nj} = \frac{2}{n}\sum_{k=1}^{n} N_k(t_{i-1}, t_i]/(t_i - t_{i-1})$. Hence, if $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_p)$ is feasible for (10), the OLS estimator $\hat{y}_n^{LS}$ coincides with this $\tilde{y}$ and is unique. Since the local EAP uniquely determines $\tilde{y}$ in this case, it follows that $\hat{y}_n^{LS}$ must then coincide with the ML estimator $\hat{y}_n$.

But when p is even, (11) has an infinite number of solutions (unlike the ML setting), and it can be that there are then infinitely many feasible minimizers of (10). In this case, one may need to add further requirements on the minimizer (e.g. smoothness of the solution) so as to select the OLS estimator.

As in the ML setting, we are able to provide a sufficient condition guaranteeing a unique minimizer (and hence a unique $\hat{y}_n^{LS}$) to (10).

PROPOSITION 4 Suppose that the number of observed arrivals in each sub-interval is non-zero (i.e., $\sum_{k=1}^{n} N_k(t_{i-1}, t_i] > 0$ for $i = 1, 2, \ldots, p$) and that $\sum_{i=1}^{p} m_i > p$. Then, the OLS optimization problem (10) in the interval count setting solves uniquely.

For the proof, see Zheng and Glynn (2017).

Again, the discussion above generalizes in a natural way to the setting where $\lambda^*(\cdot)$ is not periodic. We also note that our R code implementation includes an OLS solver that numerically minimizes (10).

4 MODEL MIS-SPECIFICATION

In this section, we briefly discuss model mis-specification issues for our ML and OLS estimators. We start by noting that the OLS estimator does not require that N be a NHPP. In fact, the essential requirements for the consistency of $\hat{\lambda}_n^{LS}$ are only that $N(\cdot)$ be square-integrable and that

$$EN(s_{i,j-1}, s_{ij}] = \int_{(s_{i,j-1}, s_{ij}]} \lambda(y^*; s)\,ds$$

for $1 \le j \le m_i$, $1 \le i \le p$. In particular, consistency does not rest on the fact that $N(s,u]$ follows a Poisson distribution for $0 \le s \le u \le t$. Our fully observed ML estimator also does not require the Poisson structure in order to exhibit consistency.

Suppose that we demand only that

$$EN(0,s] = \int_{(0,s]} \lambda^*(u)\,du \tag{12}$$

for $0 \le s \le t$. It follows that

$$\frac{1}{n}\sum_{k=1}^{n} N_k(0,s] \to \int_{(0,s]} \lambda(y^*; u)\,du \quad a.s. \tag{13}$$

for $s \in [0,t]$. Furthermore, in view of the global EAP enjoyed by $\hat{\lambda}_n$, we find that

$$\int_{(0,t]} \hat{\lambda}_n(s)\,ds \to \int_{(0,t]} \lambda(y^*; s)\,ds \quad a.s.$$

as $n \to \infty$. Also, for each feasible y,

$$\int_{(0,t]} \log \lambda(y; s) \frac{1}{n}\sum_{k=1}^{n} N_k(ds) \to \int_{(0,t]} \log \lambda(y; s)\lambda(y^*; s)\,ds \quad a.s.$$

as $n \to \infty$. For feasible $y \ne y^*$ for which

$$\int_{(0,t]} \lambda(y; s)\,ds = \int_{(0,t]} \lambda(y^*; s)\,ds,$$

Jensen's inequality implies that

$$\int_{(0,t]} \log \lambda(y;s)\lambda(y^*;s)ds < \int_{(0,t]} \log \lambda^*(y;s)\lambda^*(y;s)ds.$$

It follows that $\lambda(\hat{y}_n;s) \to \lambda(y^*;s)$ a.s. for $s \in (0,t]$. Again, the consistency argument does not require Poisson structure.

We turn next to studying the behavior of $\hat{\lambda}_n$ in the presence of (12) when λ^* is not piecewise linear. The global EAP and (13) suggest that $\hat{y}_n$ converges to $\tilde{y}$, where $\tilde{y}$ is the maximizer of

$$\begin{aligned}
\max_{y_0,y_1,\ldots,y_p} \quad & \int_{(0,t]} \log\left(\frac{\lambda(y;s)}{\lambda^*(s)}\right)\lambda^*(s)ds \\
\text{s.t.} \quad & \int_{(0,t]} \lambda(y;s)ds = \int_{(0,t]} \lambda^*(s)ds \\
& y_0 = y_p \\
& y_i \geq 0,\, 0 \leq i \leq p.
\end{aligned} \tag{14}$$

In other words, $\hat{\lambda}_n$ then converges to the feasible function $\lambda(\tilde{y};\cdot)$ that minimizes the relative entropy of λ^* with respect to $\lambda(y;\cdot)$. For a full proof, see Zheng and Glynn (2017).

5 COMPUTATIONAL COMPARISON

In this section, we demonstrate the performances of ML and OLS methods on simulated data and real data. For simulated data, we consider one case in which the true intensity is PL and a second case in which the true intensity is a smooth non-PL function. For real data, we study the customer arrival data from a large US bank call center.

There are four methods under consideration. The first two are ML and OLS methods presented in Sections 2 and 3. The third method is a "mean-constrained" method introduced by Chen and Schmeiser (2013) and (Nicol and Leemis 2014) where the local EAP is required. The fourth method is the piecewise constant (PC) intensity estimator; see Law and Kelton (1991). In order to measure the performance of a fitted intensity function $\hat{\lambda}(\cdot)$ in comparison with the true intensity function $\lambda^*(\cdot)$, we introduce L_1 and L_∞ differences, defined respectively by

$$d_1(\hat{\lambda},\lambda^*) = \int_0^t |\hat{\lambda}(s) - \lambda^*(s)|ds$$

and

$$d_\infty(\hat{\lambda},\lambda^*) = \max_{0\leq s\leq t} |\hat{\lambda}(s) - \lambda^*(s)|.$$

In the first case with simulation data, the true intensity function is PL on $(0,24]$, as shown by the black line in Figure 1. We run 1000 independent replications where in each replication, we simulate respectively $n = 10, 40, 160$ copies of the NHPP with the given PL intensity. We collect count data from $m = 24$ equal-spaced sub-intervals and use them as input data for all fourth methods. For the ML and OLS methods, we fit a PL intensity with $p = 6$ equal-spaced intervals. For mean-constrained and PC methods, we use the best available intensity estimators with the count data available.

In the second case with simulation data, the true intensity function is non-PL but Lipschitz continuous, as shown by the black line in Figure 2. We run 1000 independent replications and in each replication we simulate respectively $n = 10, 40, 160$ copies of the NHPP with the given non-PL intensity. We collect count data from $m = 12$ equal-spaced sub-intervals and use them as input data for all fourth methods. For ML and OLS, we fit a PL intensity with $p = 12$ equal-spaced pieces. For mean-constrained and PC estimators, we use the best available intensity estimators given the count data.

Tables 1 and 2 present the results when the true intensity function is PL and non-PL, respectively. The fitting performances of the fourth methods (named by “PL-ML”, “PL-OLS”, “PL-local EAP” and “PC” in the table) are provided, measured by the L_1 and L_∞ distance. As shown by the tables, the ML method is the best. The OLS method is nearly as good as ML. The mean-constrained local EAP method is moderately worse, since requiring local EAP is heuristic and may lead to overfitting in some settings. The PC estimator is significantly worse than the other three, since it fails to capture the slope information of the intensity function, particularly when the intensity is changing rapidly.

We conclude this section by exhibiting an experiment with real data. We fit a PL intensity function to the customer arrival process of a large US bank call center. This call center had about 50,000 customer arrivals each day during 2001-2003. To avoid the day-of-week effect and long-term trends, we only use data from all Mondays of the year 2001. The performance of the ML method is demonstrated where a PL intensity function with 12 equal-spaced pieces is fitted. In Figure 3, the black line indicates the trace arrivals histogram and the red line depicts the fitted PL intensity function. As shown in Figure 3, a simple 12-piece model performs well and successfully captures the rapid change in the customer arrival intensity from 7-9 AM in the morning and 4-6 PM in the afternoon, which may help enhance managerial decision making compared to a piecewise constant intensity model.

Table 1: Methods performance I.

Case			PL-ML		PL-OLS		PL-local EAP		PC	
n	p	m	L_1	L_∞	L_1	L_∞	L_1	L_∞	L_1	L_∞
10	6	24	42.6	5.4	45.1	6.3	277.7	58.2	356.9	79.6
40	6	24	20.0	2.4	21.9	2.7	129.1	26.7	349.9	73.3
160	6	24	9.7	1.50	9.9	1.54	57.3	11.5	348.1	71.2

Table 2: Methods performance II.

Case		PL-ML		PL-OLS		PL-local EAP		PC	
n	p	L_1	L_∞	L_1	L_∞	L_1	L_∞	L_1	L_∞
10	12	46.6	12.1	50.0	12.4	70.0	14.1	102.2	35.8
40	12	30.6	11.42	32.6	11.45	42.5	12.0	94.0	35.2
160	12	24.6	11.1	23.9	11.2	33.18	11.6	91.25	34.9

REFERENCES

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.

Grant, M., a., and S. Boyd. 2014, March. “CVX: Matlab Software for Disciplined Convex Programming, version 2.1”.

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

Klein, R. W., and S. D. Roberts. 1984. “A Time-varying Poisson Arrival Process Generator”. *Simulation* 43 (4): 193–195.

Law, A. M., and W. D. Kelton. 1991. *Simulation Modeling and Analysis*, Volume 2. McGraw-Hill New York.

Leemis, L. M. 1991. “Nonparametric Estimation of the Cumulative Intensity Function for a Nonhomogeneous Poisson Process”. *Management Science* 37 (7): 886–900.

Leemis, L. M. 2004. “Nonparametric estimation and variate generation for a nonhomogeneous Poisson process from event count data”. *IIE Transactions* 36 (12): 1155–1160.

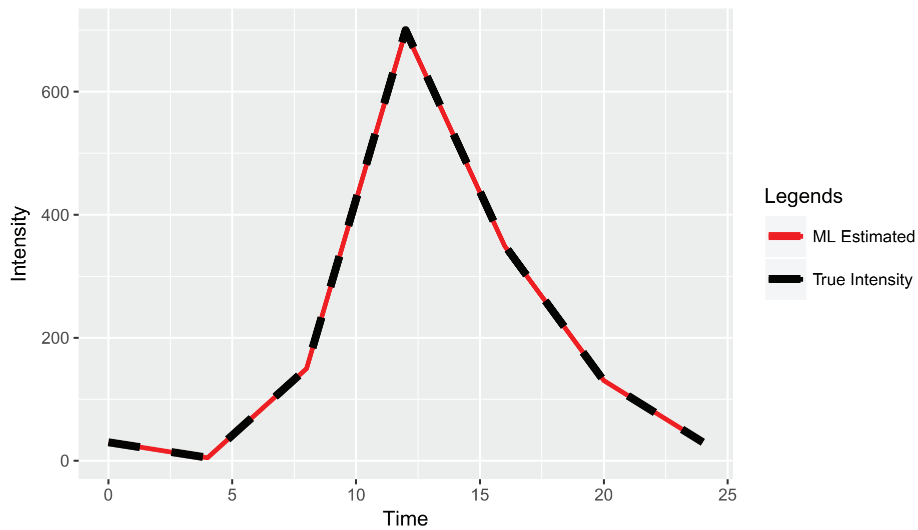

Figure 1: Fitting performance when true intensity is PL.

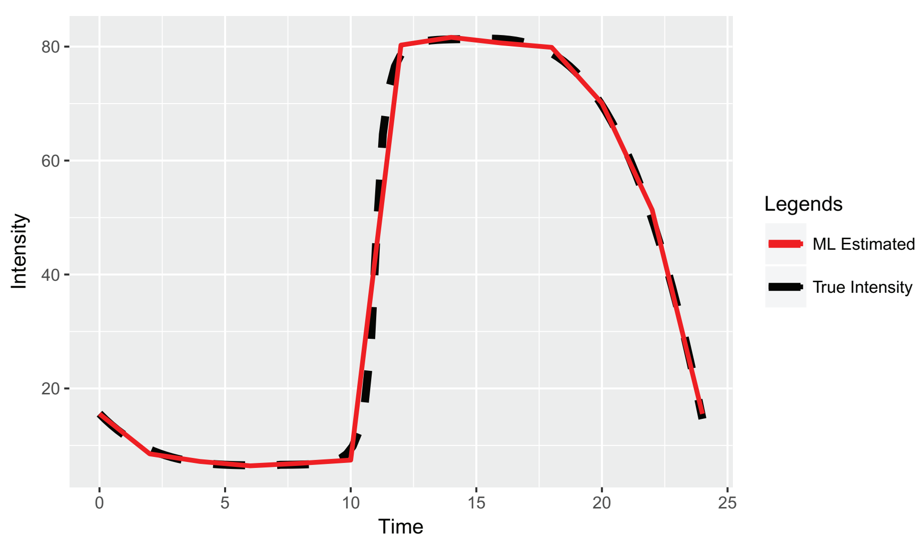

Figure 2: Fitting performance when true intensity is non-PL.

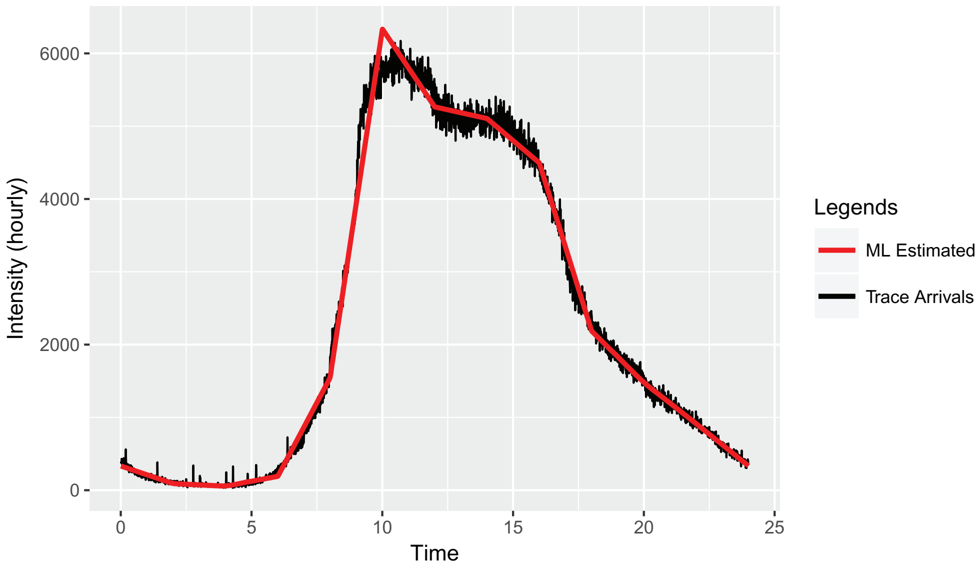

Figure 3: Fitting a PL intensity to call center customer arrivals.

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.D. Diallo, I.O. Ryzhov, L. Yilmaz, S. Buckley, and J.A. Miller, 498-509. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Zheng, Z., and P. W. Glynn. 2017. "Modeling and Testing Nonhomogeneous Poisson Processes". Working paper, Stanford University.

AUTHOR BIOGRAPHIES

ZEYU ZHENG is currently a Ph.D. candidate in the Department of Management Science and Engineering (MS&E) at Stanford University, specializing in Operations Research. He holds a Ph.D. minor in Statistics and an M.A. in Economics from Stanford University and a B.Sc. in Mathematics from Peking University. His research interests lie in stochastic simulation, high volume traffic modeling, statistical inference and market microstructure. His email address is zyzheng@stanford.edu.

PETER W. GLYNN is the Thomas Ford Professor in the Department of Management Science and Engineering (MS&E) at Stanford University. He is a Fellow of INFORMS and of the Institute of Mathematical Statistics, has been co-winner of Best Publication Awards from the INFORMS Simulation Society in 1993, 2008, and 2016, and was the co-winner of the John von Neumann Theory Prize from INFORMS in 2010. In 2012, he was elected to the National Academy of Engineering. His research interests lie in stochastic simulation, queueing theory, and statistical inference for stochastic processes. His email address is glynn@stanford.edu.