FUNCTIONAL DATA ANALYSIS FOR POINT PROCESSES WITH RARE EVENTS

Shuang Wu, Hans-Georg Müller and Zhen Zhang

University of Rochester, University of California, Davis and Abbott Vascular Inc.

Abstract: In various applications one encounters samples of objects, where each object consists of a small number of repeated event times observed over a fixed time interval. For such rare event data there are no flexible methods available that can be applied when the shapes of the intensity functions that generate the observed event times are not known, or vary substantially between objects. We model the underlying intensity functions as nonparametric object-specific random functions. Applying a novel functional method to obtain the covariance structure of the associated random densities, we reconstruct object-specific density functions that reflect the distribution of event times. We demonstrate in simulations that the proposed functional approach is superior to conventional nonparametric methods, as it borrows strength from the entire sample of objects rather than aiming at the estimation of each object's density separately. Our method is based on a key relationship that allows one to reduce the covariance estimation problem for random densities to the simpler problem of estimating a non-random joint density from pooled event times. We describe an application to model bid arrivals for a sample of online auctions, and also include asymptotic justifications of the methodology.

Key words and phrases: Bid arrival times, Cox process, density function, functional principal components, on-line auction, repeated events.

1. Introduction

The analysis of repeated event data generated by a point process is of interest in many fields. In medical studies, severity of a disease can be reflected by the frequency of recurrent symptoms, for example, asthma patients experience recurrent events of coughing and sputum (Sears et al. (1990)) and the quantitative analysis of the occurrence of such events can be an important measure of treatment efficacy. In behavioral science, smoking behaviors can be represented by the occurrence of repeated events (Shiffman et al. (2002)). The analysis of point processes has been studied extensively in parametric settings as well as in situations where one considers one realization of the underlying point process at a time (Cox and Isham (1980); Cook and Lawless (2002)).

Our methodology focuses on the commonly observed case of low intensity point processes that give rise to rare series of events. This situation will be

illustrated with the analysis of bid arrival times in on-line auctions. As bidders who participate in a particular auction submit bids on-line, bid arrival events are recorded and may be thought of as being generated by an auction-specific realization of a point process. As one observes bid arrival times over many auctions, the data reflect a sample of realizations of the point process. It is then of interest to determine the characteristics of this process, for example its variation from auction to auction. On-line auctions with their sparse bidding patterns constitute a rich source of data for such studies, for example, millions of eBay auctions take place every day.

Using nonparametric methods for rare repeated event data requires overcoming methodological challenges. We address them by developing suitable methods of functional data analysis (FDA) designed for the analysis of a sample of random functions, or data that can be thought of as having been generated by underlying random trajectories. A key objective for rare events is to pool information across the sample from all objects, and a core methodology to achieve this is functional principal component analysis (FPCA). Under mild conditions, random trajectories can be expanded in the eigenbasis of its associated covariance operator. This has led to the popularity of FPCA as a versatile tool to identify functional variation and to model random trajectories (Castro, Lawton, and Sylvestre (1986); Rice and Silverman (1991); Silverman (1996)).

If the data obtained for each random trajectory are densely sampled and noise can be ignored, a common approach (Ramsay and Silverman (2005)) is to separately smooth the data sampled for each trajectory, with the aim of generating a sample of smooth function estimates. Once the available data have been converted into a sample of functional data by such a presmoothing technique, various established FDA methods can be easily applied. This approach, obtaining in a first pre-smoothing step spline smoothed local intensity functions for point processes, has been proposed by Bouzas et al. (2006a,b) in the framework of doubly stochastic Poisson processes, in the context of an analysis of Spanish mortgage data. Similarly, Illian et al. (2006) applied FPCA to smoothed second-order statistics for the purpose of classifying spatial point process data.

The repeated event data that one encounters in on-line bidding and similar repeatedly observed series of relatively rare events are typically not dense enough for this pre-smoothing approach to be feasible. Often only few events are recorded per object (throughout, we use "object" synonymously for event series, e.g., an auction for which a series of bids is recorded). For objects with rare events, the nonparametric estimation of object-specific local intensity functions based on the data for just one object does not work; we substantiate this point with simulations reported in Section 5. Commonly used smoothing methods for point processes require large numbers of events per object to be effective. This

motivates to develop a specific functional approach for event and especially rare event data, which does not rely on the pre-smoothing of individual trajectories. The general plan is to decompose the local intensity function into an overall intensity factor that is a scalar random variable and a shape function that corresponds to a random density function, then to aim at conditional expectations of object-specific random densities. A key finding is that the covariances of these random density functions can be easily evaluated by straightforward bivariate density estimation. These estimates then lead to an eigenfunction expansion with random effects, the functional principal components. We demonstrate how to obtain predicted functional principal components for individual objects, leading to predicted object-specific random densities.

The article is organized as follows. In Section 2, we introduce the functional modeling of point process data, followed by a description of corresponding estimation procedures in Section 3. Consistency results are the theme of Section 4, followed by a report on simulation studies in Section 5 and an illustration of the methods with ebay online auction data in Section 6. Concluding remarks can be found in Section 7 and auxiliary results and proofs are provided in an appendix.

2. Functional Modeling of Event Data

2.1. Intensities and densities of point processes as functional data

We view bid arrival times as realizations of an underlying point process. Point processes can be characterized by their *local intensity function* (or rate function). Let $N(a,b)$ denote the observed number of events in the interval $(a,b) \subset [0,T]$, with an endpoint $T > 0$. The local intensity function is defined as

$$\lambda(x) = \lim_{\Delta x \to 0^+} \frac{\mathrm{E}\{N(x, x + \Delta x)\}}{\Delta x}. \qquad (2.1)$$

In our study, we interpret these intensities as object-specific functional data, i.e., as random trajectories. Random intensities are commonly considered in the framework of a doubly stochastic Poisson process (Cox and Isham (1980)). This point process is generated from a non-negative valued stochastic intensity process $\Lambda(x)$. Conditional on a realization $\Lambda = \lambda$, it is an inhomogeneous Poisson process with local intensity function $\lambda(x)$ that satisfies

$$\mathrm{E}\{N(a,b)|\Lambda = \lambda\} = \int_a^b \lambda(u)du \qquad (2.2)$$

and $\mathrm{var}\{N(a,b)|\Lambda = \lambda\} = \int_a^b \lambda(u)du.$

We consider here more general point processes with random intensity functions, for which we only require property (2.2). A stochastic density process

reflecting the shapes of the intensity functions can be defined via the random density trajectories

$$f_\lambda(x) = \frac{\lambda(x)}{\int_0^T \lambda(u)du}. \tag{2.3}$$

Defining

$$\tau = \int_0^T \lambda(u)du, \tag{2.4}$$

there exists a one-to-one mapping between the random intensity functions λ and $\{f_\lambda, \tau\}$. Given f_λ and τ, one obtains λ via $\lambda(x) = \tau f_\lambda(x)$.

2.2. Event data generated from a functional density model

Our basic assumption is that the event data are generated by an intensity process Λ. We target the induced random densities f_λ (2.3). Conditional on a realization f_i and N_i for the i-th auction, the observed event times $x_{i1}, \ldots, x_{iN_i}$ are assumed to be generated as an i.i.d. sample $x_{ij} \sim f_i$, $j = 1, \ldots, N_i$. Thus, the observed event times are assumed to be generated through a sample of object-specific density functions f_i, $i = 1, \ldots, n$, $x \in [0, T]$, which are realizations of a stochastic process $\mathcal{F}$ that produces trajectories corresponding to smooth density functions. The random number of events N in $[0, T]$ is assumed to be generated by an arbitrary integer-valued distribution with finite moments, for example a Poisson distribution, where N is assumed to be independent of f.

The bidding times x_{ij} are usually observed as ordered sequences $x_{i(1)} \leq x_{i(2)} \leq \ldots \leq x_{i(N_i)}$, or may be recorded as differences, $\triangle_1 = x_{i(1)}, \triangle_2 = x_{i(2)} - x_{i(1)}, \ldots, \triangle_{N_i} = x_{i(N_i)} - x_{i(N_i-1)}$. In the second formulation they correspond to waiting times for the next event, assuming that the starting value is at time 0. The random densities $\{f_1, \ldots, f_n\}$, which generate the sample of i.i.d. event times observed for an auction, are not observed. The sparsity of observed event times x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, makes it difficult to infer the individual densities with classical density estimation methods and calls for a functional approach. Given random densities f, we define a mean density function $\mathrm{E}f(x) = f_\mu(x)$ and covariance function $\mathrm{cov}\,[f(x), f(y)] = G(x, y)$. Assuming that there is an orthogonal expansion (in the L^2 sense) of G in terms of eigenfunctions ϕ_k and non-increasing eigenvalues ρ_k, one obtains the representation

$$G(x, y) = \sum_k \rho_k \phi_k(x)\phi_k(y), \quad x, y \in [0, T]. \tag{2.5}$$

Applying functional principal component analysis (FPCA), one may then express the functional density by the Karhunen-Loève representation,

$$f(x) = f_\mu(x) + \sum_k \xi_k \phi_k(x), \quad x \in [0, T]. \tag{2.6}$$

Here the ξ_k are uncorrelated random variables, the functional principal component (FPC) scores. They satisfy $\mathrm{E}\xi_k = 0$, $\mathrm{E}\xi_k^2 = \rho_k$, and $\sum_k \rho_k < \infty$, where $\rho_1 \geq \rho_2 \geq \ldots$ are the associated eigenvalues. The ξ_k play the role of object-specific random effects. In contrast to the model in Yao, Müller, and Wang (2005) that aims at the functional data analysis of random trajectories, for which one has sparse and noisy measurements, in the case of event times, one does not have any direct observations of the target densities $f(x)$. Therefore, a new approach is needed.

3. Principle of Borrowing Strength Across the Sample

3.1. Estimating the moments of functional densities

We aim at predicting object-specific densities f_i from observed sparse event data based on the expansion in (2.6), borrowing strength from all objects to infer the object-specific densities. The overall density function $f_\mu = \mathrm{E}(f_i)$ can be estimated by pooling all observed event times and applying kernel density estimation. Pooling $N = \sum_{i=1}^{n} N_i$ observed event times, the estimate for the density f_μ is

$$\hat{f}_\mu(x) = \frac{1}{Nh_\mu} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \kappa_1 \left(\frac{x - x_{ij}}{h_\mu} \right), \tag{3.1}$$

where h_μ is a window width that we select by generalized cross-validation (GCV), and κ_1 is a kernel function, e.g., a truncated Gaussian probability density function (Silverman (1986)).

To study estimation of the covariance function $G(x, y)$, $x, y \in [0, T]$, conditional on f, consider independent random variables X and Y, both with density function f. As f is random, X and Y are unconditionally dependent with joint density function g on $[0, T] \times [0, T]$, $(X, Y) \sim g(., .)$. Denote by x_ε and y_ε ε-neighborhoods of x and y, $x_\varepsilon = (x - \varepsilon, x + \varepsilon)$ and $y_\varepsilon = (y - \varepsilon, y + \varepsilon)$. Then, using boundedness assumptions,

$$\mathrm{E}[f(x)f(y)] = \mathrm{E}\left[\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} P(X \in x_\varepsilon | f) \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} P(Y \in y_\varepsilon | f) \right]$$

$$= \mathrm{E}\left[\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} P(X \in x_\varepsilon, Y \in y_\varepsilon | f) \right]$$

$$= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} P(X \in x_\varepsilon, Y \in y_\varepsilon) = g(x, y),$$

$$G(x, y) = \mathrm{cov}[f(x), f(y)] = \mathrm{E}[f(x)f(y)] - \mathrm{E}f(x)\mathrm{E}f(y)$$

$$= g(x, y) - f_\mu(x)f_\mu(y). \tag{3.2}$$

This key relationship reduces estimation of the covariance function G to the familiar step of estimating the joint density g of any two event times. For this, we pool all pairs of events observed for the same object across all objects and then apply a two-dimensional kernel density estimator, noting that the number of pairs is $N' = \sum_{i=1, N_i \geq 2}^{n} N_i(N_i - 1)$,

$$\hat{g}(x,y) = \frac{1}{N' h_{g_1} h_{g_2}} \sum_{i=1, N_i \geq 2}^{n} \sum_{j=1}^{N_i} \sum_{l=1, l \neq j}^{N_i} \kappa_2 \left(\frac{x - x_{ij}}{h_{g_1}}, \frac{x - x_{il}}{h_{g_2}} \right). \tag{3.3}$$

Here h_{g_1}, h_{g_2} are window widths and the kernel κ_2 is usually chosen as a symmetric 2-dimensional probability density function. We implement this method by choosing common bandwidths $h_{g_1} = h_{g_2} = \hat{h}_g$, in practical applications selected by GCV, and κ_2 as standard bivariate Gaussian kernel.

The estimate of the covariance function $G(x, y)$ is then

$$\widehat{G}(x,y) = \hat{g}(x,y) - \hat{f}_\mu(x) \hat{f}_\mu(y), \tag{3.4}$$

from which one obtains estimated eigenfunctions $\hat{\phi}_k(x)$ and eigenvalues $\hat{\rho}_k$ by numerical spectral decomposition of the covariance for a suitably discretized version. A projection on the space of all positive definite covariance functions can be easily constructed (Hall, Müller, and Yao (2008)).

3.2. Estimating object-specific density and intensity functions

Once estimates for the mean function f_μ and eigenfunctions ϕ_k have been obtained, individual trajectories are determined by using an object's functional principal components (FPCs) ξ_k, $k = 1, \ldots, K$, if K components are included in the expansion (2.6). From (2.6), we observe $\xi_{ik} = \int (f_i(x) - f_\mu(x))\phi_k(x)dx$, suggesting the estimates

$$\hat{\xi}_{ik} = \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{\phi}_k(x_{ij}) - \int \hat{f}_\mu(x) \hat{\phi}_k(x)dx. \tag{3.5}$$

An important auxiliary parameter is the number of eigen-components K to be included at (2.6). Once K has been chosen, the prediction for the object-specific density f_i is given by

$$\hat{f}_i(x) = \hat{f}_\mu(x) + \sum_{k=1}^{K} \hat{\xi}_{ik} \hat{\phi}_k(x), \; x \in [0, T]. \tag{3.6}$$

Note that $\hat{f}_i$ may not be a bona fide density, as it is neither restricted to be positive nor to integrate to 1. To obtain final estimates that are densities,

we therefore apply an adjustment by projecting $\hat{f}_i$ onto the space of density functions by truncating the function estimate when it is negative at 0, followed by multiplying with a factor so that the integral becomes 1, a standard adjustment procedure in density estimation (Gajek (1986); Glad, Hjort, and Ushakov (2003)). Although the density estimates $\hat{f}_i$ are motivated by the need to model sparse event times, these estimates can be used in any situation where one wishes to model a sample of random densities, and they are also of interest for the case with many observed events per object.

Since by (2.2), $\mathrm{E}\{N_i(0,T)|\Lambda_i = \lambda_i\} = \tau_i$, and as we only observe one realization of each object-specific process with $N_i(0,T) = N_i$, we use N_i to substitute for τ_i. Given the density estimates $\hat{f}_i$, intensity functions are then estimated by

$$\hat{\lambda}_i(x) = N_i \hat{f}_i(x), \quad i = 1, \ldots, n. \tag{3.7}$$

3.3. Choice of the number of functional principal components

While we do not make a Poisson assumption, we use a pseudo-Poisson information criterion to select the number of included functional components, similarly to the Gaussian pseudo-likelihood method for selection of components that has been shown to work well also for non-Gaussian cases in Yao, Müller, and Wang (2005). Denoting the estimated density (3.6), obtained when using K FPCs to approximate the density f_i of the i-th object by $\hat{f}_{iK}$, the Poisson deviance for the observed data x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, is approximated by

$$D(K) = 2 \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[f_i(x_{ij}) \log \left(\frac{f_i(x_{ij})}{\hat{f}_{iK}(x_{ij})} \right) - (f_i(x_{ij}) - \hat{f}_{iK}(x_{ij})) \right].$$

Since we do not observe the density functions f_i directly, we substitute a smoothed version, where we start with histograms constructed with small bin size, as follows: Given a random sample $x_1, \ldots, x_m \in (a,b)$ that is sampled from a density function f, define a partition of (a,b) by $(t_0, \ldots, t_P)$, where $a = t_0 < t_1 < \ldots < t_P = b$. The bin width Δ_j of bin $B_j = [t_{j-1}, t_j)$, $j = 1, \ldots, P$, is $\Delta_j = t_j - t_{j-1}$. Denote the count of event times falling within bin B_j by y_j, $j = 1, \ldots, P$, so that $\sum_{j=1}^{P} y_j = m$. The histogram density estimate $\tilde{f}_H$ of f is given by

$$\tilde{f}_H(x) = \frac{y_j}{m\Delta_j}, \ x \in B_j, \ j = 1, \ldots, P. \tag{3.8}$$

Applying local linear regression smoothing (Fan and Gijbels (1996)) to $\tilde{f}_H$ (3.8) with bandwidth selected by GCV leads to smoothed histogram estimates

$\hat{f}_{iH}$ of f_i and deviance estimates

$$\hat{D}(K) = 2 \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[\hat{f}_{iH}(x_{ij}) \log \left(\frac{\hat{f}_{iH}(x_{ij})}{\hat{f}_{iK}(x_{ij})} \right) - (\hat{f}_{iH}(x_{ij}) - \hat{f}_{iK}(x_j)) \right]. \quad (3.9)$$

The proposed pseudo-Poisson information criterion (PPIC) is then

$$\text{PPIC}(K) = \hat{D}(K) + 2K, \quad (3.10)$$

with penalty analogous to AIC. Other penalties, e.g., of BIC type, could be easily substituted. The selected dimension is $\hat{K} = \text{argmin}_K \text{PPIC}(K)$.

If desired, the PPIC can be adapted for choosing the number of principal components for each object separately. Let $\hat{K}_i = \text{argmin}_K \text{PPIC}_i(K)$, where $\text{PPIC}_i(K) = \hat{D}_i(K) + 2K$ and $\hat{D}_i(K)$ is (3.9) adapted for the ith object only. The prediction for the object-specific density is then $\hat{f}_i(x) = \hat{f}_\mu(x) + \sum_{k=1}^{\hat{K}_i} \hat{\xi}_{ik} \hat{\phi}_k(x)$. Since individual tuning parameter choices for functional data often lead to increased variance of the estimates, which makes it harder to interpret individually predicted density functions, and overemphasizes differences between process estimates, we use a common number of functional components across all objects in the analysis.

4. Theoretical Results

To provide theoretical justifications for our approaches, we establish the uniform convergence of the estimated mean function, covariance function, and eigenfunctions, as well as individual density estimates, and the consistency of eigenvalues and estimated FPCs, under suitable regularity conditions. Proofs are in the Appendix.

The observed event times x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, are assumed to be i.i.d. samples with density f_i, conditional on a realization of $f_i \in \mathcal{F}$ and $N_i = n_i$. A basic assumption throughout is that densities f_i, mean density $f_\mu(x)$, joint density $g(x, y)$, and eigenfunctions ϕ_k are twice continuously differentiable on the domain $[0, T]$ on which the repeated events are observed. It is also assumed that density functions and their first two derivatives are bounded, where the bounds hold uniformly across the set of random densities $\mathcal{F}$. Further assumptions are as follows; we provide additional discussions of (A1) in Section 7.

(A1) The numbers of observations N_i for the i-th object, $i = 1, \ldots, n$, are i.i.d. r.v.'s that are independent of the densities f_i and satisfy

$$\text{E}\Big(\frac{n}{\sum_{i=1}^{n} N_i} \Big) < \infty, \quad \text{E}\Big(\frac{n}{\sum_{i=1}^{n} N_i(N_i - 1)} \Big) < \infty.$$

(A2) The univariate and bivariate kernel functions $\kappa_1(\cdot)$ and $\kappa_2(\cdot,\cdot)$ for the smoothing steps for the mean f_μ and covariance G are compactly supported symmetric density functions with Fourier transforms $\gamma_1(t) = \int e^{-iut}\kappa_1(u)du$ and $\gamma_2(s,t) = \int e^{-(ius+ivt)}\kappa_2(u,v)dudv$ such that $\gamma_1(t)$ and $\gamma_2(s,t)$ are absolutely integrable.

(A3) The bandwidths h_μ and h_g used for estimating $\hat{f}_\mu$ and $\hat{g}$ satisfy
$$h_\mu \to 0 \text{ and } nh_\mu^6 < \infty; \quad h_g \to 0 \text{ and } nh_g^8 < \infty, \quad \text{as } n \to \infty.$$

We obtain the uniform convergence rates for the estimators of the mean function, joint density and covariance function. These are quantities that benefit directly from the pooling of data across objects, and therefore are well behaved.

Theorem 1. *Assume $N_i \geq 1$, $1 \leq i \leq n$. Under* (A1)$-$(A3),

$$\sup_x |\hat{f}_\mu(x) - f_\mu(x)| = O_P(\frac{1}{\sqrt{nh_\mu}}), \quad \sup_{x,y} |\hat{g}(x,y) - g(x,y)| = O_P(\frac{1}{\sqrt{nh_g^2}}), \quad (4.1)$$

$$\sup_{x,y} |\hat{G}(x,y) - G(x,y)| = O_P(\frac{1}{\sqrt{nh_g^2}} + \frac{1}{\sqrt{nh_\mu}}). \quad\quad\quad\quad (4.2)$$

Combining this result with perturbation arguments, as laid out in the proof of Theorem 2 of Yao, Müller, and Wang (2005), leads to the consistency of eigenvalue and eigenfunction estimates.

Theorem 2. *Under the assumptions of Theorem* 1,

$$|\hat{\rho}_k - \rho_k| = O_P(\frac{1}{\sqrt{nh_g^2}} + \frac{1}{\sqrt{nh_\mu}}), \quad \sup_x |\hat{\phi}_k(x) - \phi_k(x)| = O_P(\frac{1}{\sqrt{nh_g^2}} + \frac{1}{\sqrt{nh_\mu}}). \quad (4.3)$$

While the proposed method is applicable to general point processes, we proceed to study convergence rates of the estimated FPCs $\hat{\xi}_{ik}$ and object-specific density estimators $\hat{f}_i$ for the case of Poisson processes. This ubiquitous scenario serves as a model case of how to obtain such rates. Assumptions are needed.

(A4) The number of observations N_i made for the ith object is a positive r.v. from a truncated-Poisson(τ_n) distribution,
$P(N_i = 0) = 0$, $P(N_i = k) = \tau_n^k e^{-\tau_n}/(k!(1 - e^{-\tau_n}))$, $k \geq 1$.

(A5) The intensity parameters τ_n form a sequence of positive r.v.'s such that τ_n and the densities f_i, $1 \leq i \leq n$, are independent; $E\tau_n^{-1/2} = O(\alpha_n)$, $\alpha_n \to 0$, as $n \to \infty$.

(A6) The number of eigenfunctions and functional principal components K_i for each object is a random variable with $K_i \sim K$, where K is a positive finite discrete r.v. such that, for any $\epsilon > 0$, there is a $K_\epsilon^* < \infty$ with $P(K > K_\epsilon^*) < \epsilon/2$.

In the context of the application in Section 6, (A5) means that the average number of bids per auction is assumed to increase asymptotically. Without such an assumption, consistency for individual density estimates is not achievable, in analogy with the case of Gaussian trajectories (Yao, Müller, and Wang (2005)).

Nevertheless, even under this assumption, if many of the N_i are small or events are very unevenly distributed, the observed event data may be considered sparse. For example, if $\tau_n = n^{3/7}$, $\mathrm{P}(N_i \leq \sqrt{n}) = 0.89$ for $n = 100$, and 0.98 for $n = 500$. So for $n = 100$, there is a sizable proportion of observed event process objects that are very sparse; for $n = 500$, all objects with few exceptions would still be somewhat sparse. In the eBay application, most of the bids occur during the last day, while the first six days have relatively few bids, leading to subdomains with very sparse events. Situations where events are rare within some subdomains may also be viewed as sparse event data.

Assumption (A6) allows for varying numbers of eigen-components across objects. In practice, we nevertheless adopt a common number of selected components K for all objects; this is a common practice in other versions of FDA. We believe that this is a good strategy to balance individual variability and stability, as one cannot expect stable choices of individual numbers of components K_i when the observed data are sparse, since this would mean selection of one tuning parameter per object. We remark that (A6) can be relaxed at the cost of substantial additional complexity and other conditions, by employing arguments as in Müller and Yao (2008).

We use the fact that the rates of convergence of $\hat{\phi}_k(x)$ and $\hat{\rho}_k$ in Theorem 2 are uniform over a finite range of components K; uniform convergence rate of $\hat{f}_i$ on $\{K : 1 \leq K \leq K_\epsilon^*\}$ emerges as a consequence.

Theorem 3. *For any $\epsilon > 0$, there exists an event A_ϵ that has probability $P(A_\epsilon) \geq 1 - \epsilon$, such that under* (A1)$-$(A6), *on A_ϵ it holds that*

$$|\hat{\xi}_{ik} - \xi_{ik}| = O_P(\alpha_n + \frac{1}{\sqrt{n}h_g^2} + \frac{1}{\sqrt{n}h_\mu}), \qquad (4.4)$$

$$\sup_x |\hat{f}_i(x) - f_i(x)| = O_P(\alpha_n + \frac{3}{\sqrt{n}h_g^2} + \frac{2}{\sqrt{n}h_\mu}), \qquad (4.5)$$

where α_n is as in (A5).

5. Simulation Studies

Simulations were designed to compare the proposed functional approach with two straightforward alternatives: smoothing histograms separately for each object, based on just the observed event times for the object in question; pooling all observed event times, using the resulting density estimate for the pooled data as

predictor for individual object densities. The smoothed histogram is a commonly used nonparametric method for density estimation; the pooled density estimate corresponds to the estimate $\hat{f}_\mu$ of the mean density function f_μ in the expansion (2.6). Since the smoothed histograms are based on the observed data for each object separately, we would expect large deviations from the true densities for objects with rare events. Although the pooled estimate $\hat{f}_\mu$ of the mean density function borrows information across the objects, it fails to account for the object-specific features of the random densities, and does not reflect individual event counts.

In each simulation, we generated recurrent events for $n = 200$ objects. For the number of events N_i recorded for the i-th object, we considered three cases: N_i uniformly selected from 2 to 5, uniformly from 5 to 10, and N_i fixed at 25, representing rare, relatively rare and dense events, respectively. The performance measure is mean integrated squared error

$$\text{MISE} = \frac{1}{n} \sum_{i=1}^{n} \int_0^T \left[f_i(t) - \hat{f}_i(t) \right]^2 dt, \tag{5.1}$$

obtained by averaging over 400 simulations.

In a *Gaussian scenario*, sample densities were generated as Gaussian densities with random means $\mu_i \sim N(8, 1)$ and random variances $\sigma_i^2 \sim U(0.5, 1.5)$, $i = 1, \ldots, n$, choosing $T = 16$. In a *Gaussian Mixture scenario*, object-specific densities f_i were generated as $f_i(x) = \varphi_{i1}(x)/2 + \varphi_{i2}(x)/2$. Here φ_{i1}, φ_{i2} are Gaussian densities with parameters (μ_{i1}, σ) for φ_{i1} and (μ_{i2}, σ) for φ_{i2}, where $\mu_{i1} \sim U(1, 2)$ and $\mu_{i1} \sim U(2, 3)$, $i = 1, \ldots, n$, and $\sigma = 0.15$ is fixed, choosing $T = 4$. In a *Beta-Gaussian Mixture scenario*, we simulated bimodal densities, where $f_i(x) = \kappa_i \psi_i(x) + (1 - \kappa_i)\varphi_i(x)$. Here $\kappa_i \sim U(0.3, 0.7)$ and the ψ_i are Beta densities with random parameters (a_i, b_i), where $a_i \sim U(2, 4)$ and $b_i \sim U(0.2, 0.8)$; the φ_i are Gaussian densities with random means $\mu_i \sim U(0.2, 0.8)$ and fixed standard deviation $\sigma = 0.05$, $i = 1, \ldots, n$, choosing $T = 1$. These mixture densities have similar shapes as the densities observed for on-line auction example discussed in the following section.

The simulation results are listed in Table 1. One finds there that the proposed functional approach consistently performs better than individually smoothed densities, even for moderately large sample sizes $N_i = 25$. For objects with dense events, one would have expected individual histograms to be quite competitive, and it is surprising that they are not. The functional approach also improves considerably upon the estimates obtained by pooling all data, especially when the observed events are dense. These results support the idea of borrowing strength across the sample of rare event data to improve object-specific estimates.

Table 1.   Mean integrated squared errors for three different simulation scenarios.

Simulation Setting	Sample Size N_i	Functional		Smoothed Histogram		Pooled Estimate	
		Median	Mean	Median	Mean	Median	Mean
Gaussian	$2 \sim 5$	0.0611	0.0700	0.1567	0.1566	0.0934	0.0932
	$5 \sim 10$	0.0357	0.0412	0.0785	0.0787	0.0925	0.0927
	25	0.0150	0.0162	0.0274	0.0273	0.0919	0.0919
Gaussian	$2 \sim 5$	0.3906	0.4045	0.6806	0.6837	0.5301	0.5297
Mixture	$5 \sim 10$	0.2762	0.2846	0.3668	0.3675	0.5259	0.5259
	25	0.1237	0.1299	0.1346	0.1345	0.5224	0.5220
Beta-Gaussian	$2 \sim 5$	1.1662	1.1841	2.2430	2.2670	1.2167	1.2160
Mixture	$5 \sim 10$	0.7641	0.7644	1.2948	1.2915	1.2052	1.2060
	25	0.3243	0.3241	0.4974	0.4977	1.1981	1.1979

6. Functional Modeling of Bid Arrival Times for Online Auctions

While the analysis of the on-line auction price process and its dynamics has been well studied, much less is known about the processes generating the events that correspond to bidding, referred to as bid arrival times. For eBay online auctions, the bidding events are relatively rare, with more intensive bidding near the end of an auction, sometimes referred to as "bid sniping", as potential buyers try to outbid each other (Bapna, Jank, and Shmueli (2008)). Previous studies of bid arrivals were primarily descriptive (Ockenfels and Roth (2002)) or focused on parametric approaches, introducing the "Barista" model (Shmueli, Russo, and Jank (2007)) and the concept of self-similarity of the bid arrival process (Russo, Shmueli, and Shyamalkumar (2008)). We aim at a fully nonparametric analysis, letting the data speak for themselves.

We consider 156 Palm M515 PDA online auctions from e-Bay, which were all 7-day second price auctions that took place between March and May, 2003, using data collected by Wolfgang Jank. For each auction, the bid arrival times were recorded with accuracy in seconds. We converted the time unit into hours and recalibrated the start of all auctions to time 0, so the time domain is always between 0 and 168 hours. The total number of bids recorded per auction ranged from 8 to 51, with a median of 22. Figure 1 displays the bid arrival times for five randomly selected auctions, indicating substantial variation in bidding patterns. The median bid count per day is relatively low during the first six days, while on day 7, the last day, it increases to 9, with a maximum of 32. This uneven distribution of the daily number of bids reflects a typically more intense start, followed by much less activity during the middle stage and then intense bidding near the end.

The intensity of last-minute bidding activity is demonstrated by the histograms of the aggregated bids obtained from all auctions for the last day and
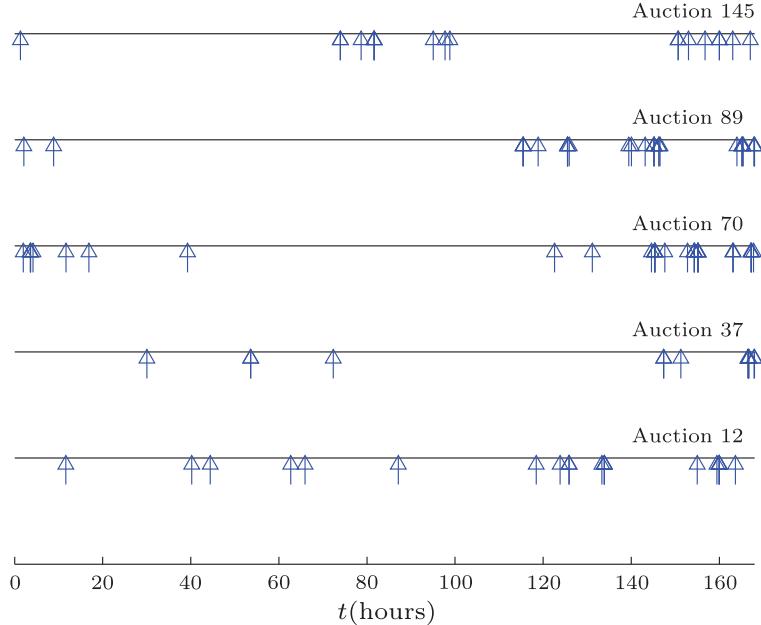
Figure 1. Bid arrival times (arrows) for 5 randomly selected auctions.

the last hour (Figure 2). During the last hour, the aggregated number of bids in the last 6 minutes was 190, while the median number of bids in the aggregated sample for the first 54 minutes was only 20. The huge difference in scale between the bidding activities in the last 0.1 hour and the rest of the auction suggests that bid arrivals in the last 0.1 hour are generated from a different process and modeling it together with the rest of the auction may lead to biased estimates for the earlier phases, since the variation of the last 0.1 hour of bidding would dominate. We therefore restricted the illustration of our method to the bids that were placed between 0 and 167.9 hours. The last 0.1 hours were analyzed separately with the same methodology (results not reported).

The mean density estimate $\hat{f}_\mu$ of bidding activity, overlaid with the histogram for the pooled data, is shown in Figure 3. As expected, the mean density has a small peak around 3hr and a fast rising right tail. We note that, instead of directly plugging the estimates $\hat{g}$ and $\hat{f}_\mu$ into (3.2), an alternative approach is to first construct the histogram estimates $\tilde{g}_H$ and $\tilde{f}_{H\mu}$ for densities g and f_μ, respectively, and then to obtain the smoothed covariance function $\hat{G}$ by applying a two-dimensional smoothing step to the values of $\tilde{G}(x,y) = \tilde{g}_H(x,y) - \tilde{f}_{H\mu}(x)\tilde{f}_{H\mu}(y)$ on the grid defined by the midpoints of the histogram bins. We use this simple implementation, obtaining the fitted covariance displayed in the left panel of Figure 4; it is the projection of the initial smooth $\hat{G}$ onto the space of positive definite surfaces. The corresponding correlation surface (right panel) shows
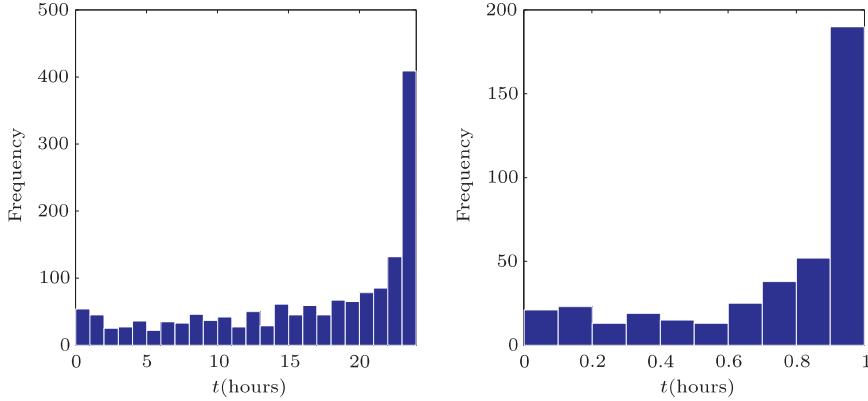
Figure 2. Histograms of the number of bids placed on the last day (left panel) and in the last hour (right panel).
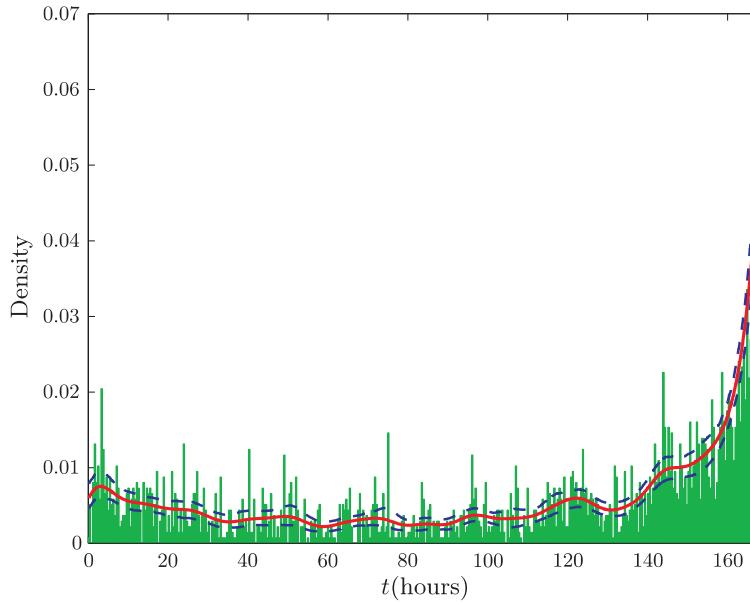

Figure 3. Histogram of the pooled bid arrival times for 156 auctions, overlaid with the estimated mean density function $\hat{f}_\mu$ (solid) and 95% pointwise confidence bands (dashed), $t \in [0, 167.9]$ (hours).

negative correlations between early and late bidding intensities, indicating that there is a trade-off in bidding activity for the different auction phases.

The proposed criterion PPIC led to the selection of $K = 3$ components for modeling the auction-specific densities in the main auction phase, explaining
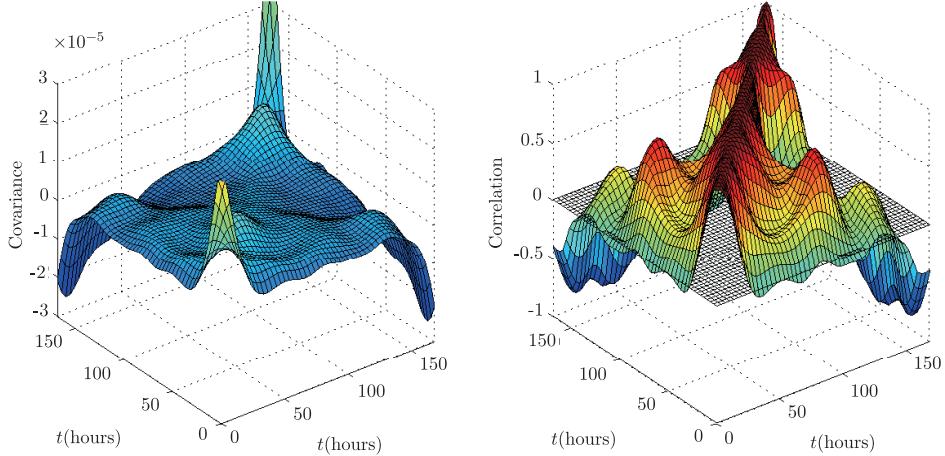
Figure 4. The estimated covariance surface (left panel) and the corresponding correlation surface (right panel), $t \in [0, 167.9]$ (hours). A zero plane is overlaid with the correlation surface for reference of sign changes.

86.9% of the total variation; see Figure 5. The sign change of the first eigenfunction (explaining 50.6% of total variation) reflects the negative correlation between early and late bidding. The first eigen-component is seen to be associated with larger variability in the bid counts during the initial and close-to-end auction phases. Both the second and the third eigenfunctions (explaining 24.63% and 11.70%, respectively) have a peak around 144hr and are aligned with trends where increased (decreased) bidding about one day before the end of the auction is associated with decreased (increased) bidding closer to the end. Auction-specific densities were estimated via (3.6) with $K = 3$. We display the fits, overlaid with the corresponding histograms, for nine randomly selected auctions in Figure 6.

Inference for the estimates of model components was obtained by the bootstrap. A natural bootstrap sampling scheme is to resample from the objects with replacement and then to resample for a selected object from the bid events x_{ij}, $1 \leq j \leq N_i$ within the selected auction, also with replacement. The bootstrap estimates of the mean, covariance, and eigenfunctions were used to construct pointwise confidence bands for the corresponding quantities, by locating the corresponding empirical quantiles in the bootstrap distributions (Figures 3 and 5). This double bootstrap resampling scheme reflects that the variability in the individual density estimates comes from two sources: The variability of the mean and eigenfunction estimates and, in addition, the variability in the principal component estimates. The quantiles for the confidence intervals were constructed from the resulting B density estimates for each auction of interest, where $B = 500$
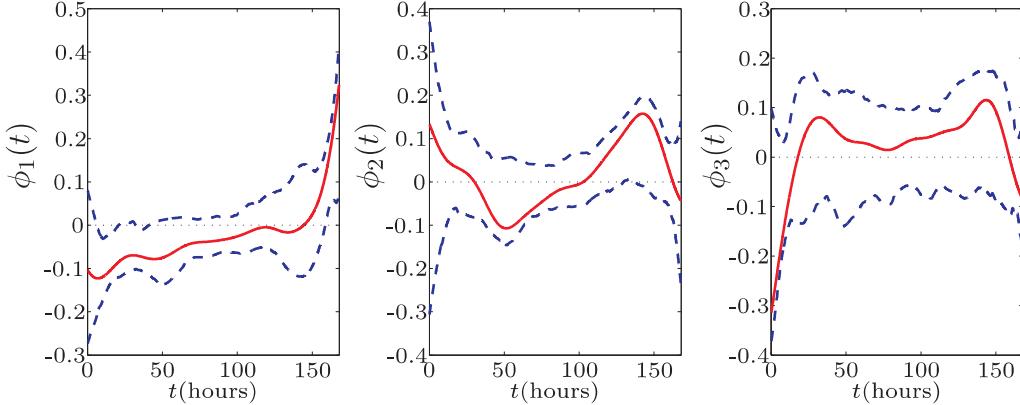
Figure 5. First three eigenfunctions (solid lines; 1st in the left panel, 3rd in the right panel) and their 95% pointwise confidence bands (dashed lines), $t \in [0, 167.9]$ (hours).

bootstrap samples were used to construct the 95% pointwise confidence bands; these are shown in Figure 6.

7. Discussion and Concluding Remarks

For the well-studied case where data are obtained from a single realization of a point process, various nonparametric approaches have been developed over the years. For example, Diggle (1985) proposed a nonparametric kernel estimate for the local intensity function of stationary Cox processes and discussed properties of this estimate and bandwidth selection for uniform kernel functions; Diggle and Marron (1988) noted the similarity between kernel intensity and density estimation and demonstrated the utility of kernel density estimates in the point process setting; kernel-type and smoothing based intensity estimators and their properties were also discussed in Ramlau-Hansen (1983); Yandell (1983); Guan (2008). Of particular interest in this context is Chiang, Wang, and Huang (2005), where a sample of point processes was considered. The objective of the work of Chiang, Wang, and Huang (2005) was to obtain estimators for the population average of the occurrence probability of recurrent events. In contrast, we focus on object-specific estimation, where we face specific challenges, due to the randomness of the intensity functions and the sparsity of the repeated events.

One can ask whether assumption (A1), which is central to our developments and requires that the numbers of observations N_i for the i-th object and the densities f_i be independent, is satisfied for our application. We checked this (results not shown) and found no evidence that the shapes of the random densities are related to the numbers of events. Generally, small violations of this assumption would not matter, but if there are major violations the method would need to
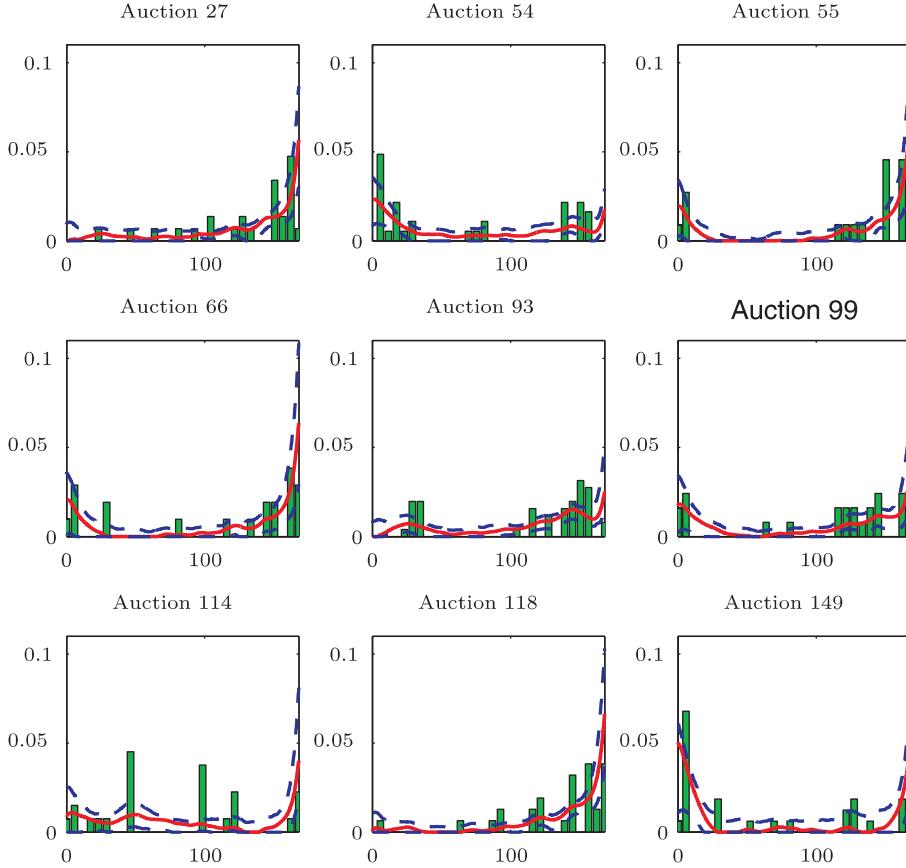
Figure 6. The estimated auction-specific densities (solid) and 95% pointwise bands (dashed) for 9 randomly selected auctions, overlaid with the corresponding histograms, $t \in [0, 167.9]$ (hours).

be modified. This could be done either by stratifying the sample into subgroups with more homogeneous numbers of observed events or, more formally, by implementing a moving window functional principal component analysis, similar to the proposal in Chiou and Müller (2009), where the number of observed events can be chosen as the covariate defining the moving windows.

We demonstrate that applying functional data analysis methods for rare repeated events is useful for predicting random intensity or density functions for individual objects, without resorting to parametric assumptions. This object oriented approach leads to qualitative assessment of the variability of individual density functions; comparisons of object-specific and mean density functions; compact graphical displays describing the sample, including mean and eigenfunctions, covariance, and correlation functions; improved estimation by borrowing strength across the sample, as demonstrated in simulations. The latter is a key

feature of the functional approach that is particularly useful for the analysis of objects that consist of rare repeated events.

Acknowledgement

Appendix: Auxiliary Results and Proofs

We require approximations of the Poisson(λ) distribution with a normal distribution with mean λ and variance λ. The proof of Lemma 1 follows from Makabe and Morimura (1955); Corollary 1 and Lemma 2 are obtained by straightforward calculations.

Lemma A.1. If $N \sim Poisson(\lambda)$, $\lambda > 0$, then for small $\varepsilon > 0$,

$$\mathrm{P}\big(\min(N, \lambda) \leq \varepsilon \lambda^{2/3}\big) = \mathrm{P}\big(0 \leq N \leq [\varepsilon \lambda^{2/3}]\big) \leq D(\lambda),$$

$$
\begin{aligned}
D(\lambda) =\ & \frac{1}{\sqrt{2\pi\lambda}}([\varepsilon\lambda^{2/3}]+1)e^{-\frac{([\varepsilon\lambda^{2/3}]-\lambda+1/2)^2}{2\lambda}} \\
& + \frac{1}{6\sqrt{2\pi\lambda}}\Big\{\Big(1-\frac{([\varepsilon\lambda^{2/3}]-\lambda+1/2)^2}{\lambda}\Big)e^{-\frac{([\varepsilon\lambda^{2/3}]-\lambda+1/2)^2}{2\lambda}} \\
& - \Big(1-\frac{(\lambda+1/2)^2}{\lambda}\Big)e^{-\frac{(\lambda+1/2)^2}{2\lambda}}\Big\} \\
& + \frac{0.0544}{\lambda} + \frac{0.0108}{\lambda^{3/2}} + \frac{0.2743}{\lambda^2} + \frac{0.0065}{\lambda^{5/2}} + \Big(1+\frac{1}{2\sqrt{\lambda}}\Big)e^{-2\sqrt{\lambda}}.
\end{aligned}
$$

Corollary A.1. If N is a r.v. from a truncated-Poisson(λ), $\lambda > 0$, $N \geq 1$, for $D(\lambda)$ as in Lemma 1 and a small $\varepsilon > 0$,

$$\mathrm{P}\Big(\min\Big(N, \frac{\lambda}{1-e^{-\lambda}}\Big) \leq \varepsilon\lambda^{2/3}\Big) \leq D(\lambda).$$

Lemma A.2. If $N \geq 1$ is a r.v. from a truncated-Poisson(λ), $\lambda > 0$,

$$\mu_N = \mathrm{E}(N) = \frac{\lambda}{1-e^{-\lambda}}, \quad \mathrm{var}(N) = \frac{\lambda(1-e^{-\lambda}-\lambda e^{-\lambda})}{(1-e^{-\lambda})^2}.$$

Since $N_1, \ldots, N_n$ are i.i.d., for any function $q : \mathbb{N} \mapsto \mathbb{N}$, $q(N_1), \ldots, q(N_n)$ are also i.i.d. and the conditional distribution of $(q(N_1), \ldots, q(N_n))$ given

$\tilde{N} = \sum_{i=1}^{n} q(N_i)$ is Multinomial$(\tilde{N}, 1/n, \ldots, 1/n)$. Therefore,

$$\mathrm{E}\left(\frac{q(N_i)}{\tilde{N}}\right) = \mathrm{E}\left[\mathrm{E}\left(\frac{q(N_i)}{\tilde{N}}\Big|\tilde{N}\right)\right] = \frac{1}{n}, \tag{A.1}$$

$$\mathrm{E}\left(\frac{q(N_i)}{\tilde{N}}\right)^2 = \mathrm{E}\left[\mathrm{E}\left(\frac{q(N_i)^2}{\tilde{N}^2}\Big|\tilde{N}\right)\right] = \frac{1}{n^2} + \frac{n-1}{n^2}\mathrm{E}(\frac{1}{\tilde{N}}), \tag{A.2}$$

$$\mathrm{E}\left(\frac{q(N_i)q(N_j)}{\tilde{N}^2}\right) = \mathrm{E}\left[\mathrm{E}\left(\frac{q(N_i)q(N_j)}{\tilde{N}^2}\Big|\tilde{N}\right)\right] = \frac{1}{n^2} - \frac{1}{n^2}\mathrm{E}(\frac{1}{\tilde{N}}), \ i \neq j. \tag{A.3}$$

Proof of Theorem 1. Note that $|\hat{f}_\mu(x) - f_\mu(x)| \leq |\hat{f}_\mu(x) - \mathrm{E}\hat{f}_\mu(x)| + |\mathrm{E}\hat{f}_\mu(x) - f_\mu(x)|$. From (3.1), using a Taylor expansion with a suitable mean value δ_u, one obtains

$$\mathrm{E}\hat{f}_\mu(x) = \sum_{i=1}^{n} \mathrm{E}(\frac{N_i}{N}) \int \kappa_1(u) f_\mu(x - uh_\mu)du$$

$$= \int \kappa_1(u)\left(f_\mu(x) - uh_\mu f'_\mu(x) + \frac{1}{2}u^2 h_\mu^2 f''_\mu(\delta_u)\right)du$$

$$= f_\mu(x) + \frac{1}{2}\left(\int \kappa_1(u) f''_\mu(\delta_u) u^2 du\right) h_\mu^2,$$

which implies $\sup_x |\mathrm{E}\hat{f}_\mu(x) - f_\mu(x)| = O(h_\mu^2)$. For the random part $|\hat{f}_\mu(x) - \mathrm{E}\hat{f}_\mu(x)|$, we insert the inverse Fourier transform $\kappa_1(u) = (1/(2\pi))\int e^{iut}\gamma_1(t)dt$ into $\hat{f}_\mu$,

$$\hat{f}_\mu(x) = \frac{1}{2\pi}\sum_{k=1}^{n}\frac{1}{N}\sum_{j=1}^{N_k}\int e^{iu(x-X_{kj})}\gamma_1(uh_\mu)du = \frac{1}{2\pi}\int \varphi(u)e^{iux}\gamma_1(uh_\mu)du,$$

where $\varphi(u) = \sum_{k=1}^{n}(1/N)\sum_{j=1}^{N_k} e^{-iuX_{kj}}$. Thus $\sup_x |\hat{f}_\mu(x) - \mathrm{E}\hat{f}_\mu(x)| \leq (1/(2\pi))\int |\varphi(u) - \mathrm{E}\varphi(u)||\gamma_1(uh_\mu)|du$. With $\mathrm{E}|\varphi(u) - \mathrm{E}\varphi(u)| \leq \sqrt{\mathrm{E}(\varphi(u) - \mathrm{E}\varphi(u))^2}$ and conditioning,

$$\mathrm{var}(\varphi(u)) \leq \mathrm{E}\left(\frac{1}{N}\right) + \mathrm{var}\left(\sum_{k=1}^{n}\frac{N_k}{N}\int e^{-iut}f_k(t)dt\right).$$

Let $P_k = N_k/N$, $Q_k = \int e^{-iut}f_k(t)dt$. By (A1), P_k and Q_k are independent, and the Q_k are i.i.d. with $\alpha = \mathrm{E}Q_k = \int e^{-iut}f_\mu(t)dt$, $\beta = \mathrm{E}Q_k^2 \leq 1$. Applying (A.1)−(A.3) with $g(N_k) = N_k$, $\tilde{N} = N$, we obtain

$$\mathrm{var}\left(\sum_{k=1}^{n}\frac{N_k}{N}\int e^{-iut}f_k(t)dt\right) = \mathrm{var}\left(\sum_{k=1}^{n} P_k Q_k\right)$$

$$\leq \sum_{k=1}^{n}\left[\beta\{\frac{1}{n^2} + \frac{n-1}{n^2}\mathrm{E}(\frac{1}{N})\}\right] + \sum\sum_{k\neq j}\left[\alpha^2\{-\frac{1}{n^2}\mathrm{E}(\frac{1}{N})\}\right] \leq \frac{1}{n}\left[1 + \mathrm{E}(\frac{n}{N})\right].$$

Therefore, $\text{var}(\varphi(u)) \le (1/n)[1 + 2\text{E}(n/N)]$, and by (A1) and (A2), $\text{E}\{\sup_x |\hat{f}_\mu(x) - \text{E}\hat{f}_\mu(x)|\} = O(1/(\sqrt{n}h_\mu))$, which implies $\sup_x |\hat{f}_\mu(x) - \text{E}\hat{f}_\mu(x)| = O_P(1/\sqrt{n}h_\mu)$. Since $nh_\mu^6 < \infty$ implies $h_\mu^2 = O(1/\sqrt{n}h_\mu)$, the first result in (4.1) follows. The proof of $\hat{g}(x, y)$ can be done in similar fashion and (4.2) is an immediate consequece of (4.1).

Proof of Theorem 3. Write $f_\mu(x) = \sum_{k=1}^\infty \varsigma_k \phi_k(x)$ and $f_i(x) = \sum_{k=1}^\infty \zeta_{ik}\phi_k(x)$, where $\zeta_{ik} = \varsigma_k + \xi_{ik}$. From (4.3), we have

$$|\hat{\xi}_{ik} - \xi_{ik}| \le \left| \int f_i(x)\phi_k(x)dx - \frac{1}{N_i}\sum_{j=1}^{N_i} \hat{\phi}_k(X_{ij}) \right|$$

$$+ \left| \int f_\mu(x)\phi_k(x)dx - \int \hat{f}_\mu(x)\hat{\phi}_k(x)dx \right|$$

$$\le \left| \zeta_{ik} - \frac{1}{N_i}\sum_{j=1}^{N_i} \phi_k(X_{ij}) \right| + O_P\left(\frac{1}{\sqrt{n}h_g^2} + \frac{1}{\sqrt{n}h_\mu}\right),$$

Let

$$\Delta = \frac{1}{N_i}\sum_{j=1}^{N_i} \{\phi_k(X_{ij}) - \zeta_{ik}\} = \left(\sum_{j=1}^{N_i} \frac{\phi_k(X_{ij}) - \zeta_{ik}}{\sqrt{\tau_n/(1 - e^{-\tau_n})}}\right) \frac{\tau_n/(1 - e^{-\tau_n})}{N_i} \frac{1}{\sqrt{\tau_n/(1 - e^{-\tau_n})}}.$$

By a conditioning argument, we can show that

$$\frac{\sum_{j=1}^{N_i} \{\phi_k(X_{ij}) - \zeta_{ik}\}}{\sqrt{\tau_n/(1 - e^{-\tau_n})}}$$

has zero mean and finite variance, and thus is $O_P(1)$. We also note that $\text{E}(\|\{\tau_n/(1 - e^{-\tau_n}\|)\}^{-1/2}) < \text{E}\tau_n^{-1/2} = O(\alpha_n)$, as $n \to \infty$, implying $|\Delta| = O_P(\alpha_n)$ as long as

$$\frac{\tau_n/(1 - e^{-\tau_n})}{N_i} = O_P(1). \tag{A.4}$$

We expand $(\tau_n/(1 - e^{-\tau_n}))/N_i$ around $\tau_n^* = \text{E}(N_i|\tau_n) = \tau_n/(1 - e^{-\tau_n})$, with a suitable choice of mean value $\tilde{\tau}$,

$$\frac{\tau_n/(1 - e^{-\tau_n})}{N_i} = \frac{\tau_n/(1 - e^{-\tau_n})}{\tau_n^*} - \frac{\tau_n/(1 - e^{-\tau_n})}{\tau_n^{*2}}(N_i - \tau_n^*) + \frac{2\tau_n/(1 - e^{-\tau_n})}{\tilde{\tau}^3}(N_i - \tau_n^*)^2$$

$$\le 1 - \frac{N_i - \tau_n/(1 - e^{-\tau_n})}{\tau_n/(1 - e^{-\tau_n})}$$

$$+ \frac{\tau_n^2}{[\min(N_i, \tau_n/(1 - e^{-\tau_n}))]^3} \frac{2}{(1 - e^{-\tau_n})^2} \frac{(N_i - \tau_n/(1 - e^{-\tau_n}))^2}{\tau_n/(1 - e^{-\tau_n})}.$$

By Lemma 2 and (A4), we find that

$$\left|\frac{N_i - \tau_n/(1 - e^{-\tau_n})}{\tau_n/(1 - e^{-\tau_n})}\right| = O_P(1), \qquad \frac{(N_i - \tau_n/(1 - e^{-\tau_n}))^2}{\tau_n/(1 - e^{-\tau_n})} = O_P(1),$$

and

$$\mathrm{E}\left(\frac{1}{(e^{\tau_n} - 1)^2}\right) < \mathrm{E}(\tau_n^{-1/2}) = O(\alpha_n),$$

as $n \to \infty$. Therefore, (A.4) holds as long as

$$\mathrm{P}\left(\min(N_i, \tau_n/(1 - e^{-\tau_n}))\right) \le \varepsilon\tau_n^{2/3}) \to 0, \text{ as } n \to 0, \ \forall\varepsilon > 0. \tag{A.5}$$

Writing $\tau_n = \tau_n(\omega_1)$ and $N_i = N_i(\omega_2)$, where $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, and Ω_1, Ω_2 are suitable probability spaces, one obtains by Fubini's Theorem and Corollary 1, $\mathrm{P}(\min(N_i, \tau_n/(1 - e^{-\tau_n}))) \le \varepsilon\tau_n^{2/3}) = \mathrm{E}_{\Omega_1}[\mathrm{P}_{\Omega_2}(\min(N_i, \tau_n/(1 - e^{-\tau_n}))) \le \varepsilon\tau_n^{2/3}|\tau_n(\omega_1))] \le \mathrm{E}_{\Omega_1}\{D(\tau_n(\omega_1))\}$, where $D(\tau_n(\omega_1))$ is a sum of finitely many terms of the form $\tau_n^{-a}(\omega_1)$, $a \ge 1$ and $\tau_n^b(\omega_1)e^{-\tau_n^c(\omega_1)}$, $b > 0$, $c \ge 1/2$. Since $\mathrm{E}\tau_n^{-a} < \mathrm{E}(\tau_n^{-1/2}) = O(\alpha_n)$, and $\mathrm{E}_{\Omega_1}(\tau_n^b e^{-\tau_n^c}) = \mathrm{E}_{\Omega_1}(1/\sum_{j=0}^{\infty}(1/j!)\tau_n^{cj-b}) \le \mathrm{E}_{\Omega_1}\tau_n^{-1/2} = O(\alpha_n)$, as $n \to \infty$, we have $\mathrm{E}_{\Omega_1}\{D(\tau_n(\omega_1))\} = O_P(\alpha_n)$, which implies (A.4) and (4.4).

Let $f_i^K(x) = f_\mu(x) + \sum_{k=1}^{K} \xi_{ik}\phi_k(x)$, then $|\hat{f}_i(x) - f_i(x)| \le |\hat{f}_i(x) - f_i^K(x)| + |f_i^K(x) - f_i(x)|$. By Mercer's Theorem

$$\sup_x \mathrm{E}\left[f_i^K(x) - f_i(x)\right]^2 = \sup_x \mathrm{var}\left(\sum_{k=K+1}^{\infty} \xi_{ik}\phi_k(x)\right) = \sup_x \sum_{k=K+1}^{\infty} \rho_k\phi_k^2(x) \longrightarrow 0$$

Thus, $|f_i^K(x) - f_i(x)| = o_P(1)$.

We now define an event A_ϵ as $A_\epsilon^c = \{K > K_\epsilon^*\} \cup \{N_i = 1, i = 1, \ldots, n\}$. Since $\mathrm{P}(\{N_i = 1, i = 1, \ldots, n\}) = (\tau_n/e^{\tau_n} - 1)^n = (1/\sum_{j=1}^{\infty}(1/j!)\tau_n^{j-1})^n < \epsilon/2$ for sufficiently large n, by (A6) we have $\mathrm{P}(A_\epsilon) \ge 1 - \epsilon$. On A_ϵ, (A1) is valid and so are (4.1) and (4.3). In addition, for a finite range of K, the rates in (4.3) and (4.4) are uniform in K. This implies that on the event A_ϵ,

$$\sup_x |\hat{f}_i(x) - f_i^K(x)| \le \sup_x |\hat{f}_\mu(x) - f_\mu(x)| + \sum_{k=1}^{K} \sup_x |\hat{\xi}_{ik}\hat{\phi}_k(x) - \xi_{ik}\phi_k(x)|$$

$$= O_P\left(\frac{1}{\sqrt{n}h_g^2}\right) + O_P\left(\alpha_n + \frac{2}{\sqrt{n}h_g^2} + \frac{2}{\sqrt{n}h_\mu}\right),$$

establishing (4.5).

References

Bapna, R., Jank, W. and Shmueli, G. (2008). Price formation and its dynamics in online auctions. *Decis. Support Syst.* **44**, 641-656.

Bouzas, P. R., Aguilera, A. M., Valderrama, M. J. and Ruiz-Fuentes, N. (2006a). On the structure of the stochastic process of mortgages in Spain. *Comput. Statist.* **21**, 73-89.

Bouzas, P. R., Valderrama, M., Aguilera, A. M. and Ruiz-Fuentes, N. (2006b). Modeling the mean of a doubly stochastic Poisson process by functional data analysis. *Comput. Statist. Data Anal.* **50**, 2655-2667.

Castro, P. E., Lawton, W. H. and Sylvestre, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28**, 329-337.

Chiang, C.-T., Wang, M.-C. and Huang, C.-Y. (2005). Kernel estimation of rate function for recurrent event data. *Scand. J. Statist. Theory and Applications* **32**, 77-91.

Chiou, J.-M. and Müller, H.-G. (2009). Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *J. Amer. Statist. Assoc.* **104**, 572-585.

Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Statist. Meth. Medical Res.* **11**, 141-166.

Cox, D. R. and Isham, V. (1980). *Point Processes*. Chapman & Hall, London.

Diggle, P. and Marron, J. (1988). Equivalence of smoothing parameter selections in density and intensity estimation. *J. Amer. Statist. Assoc.* **83**, 793-800.

Diggle, P. J. (1985). A kernel method for smoothing point process data. *Appl. Statist.* **34**, 138-147.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.

Gajek, L. (1986). On improving density estimators which are not bona fide functions. *Ann. Statist.* **14**, 1612-1618.

Glad, I. K., Hjort, N. L. and Ushakov, N. G. (2003). Correction of density estimators that are not densities. *Scand. J. Statist. Theory and Applications* **30**, 415-427.

Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *J. Amer. Statist. Assoc.* **103**, 1238-1247.

Hall, P., Müller, H.-G. and Yao, F. (2008). Modeling sparse generalized longitudinal observations with latent gaussian processes. *J. Roy. Statist. Soc. Ser. B* **70**, 730-723.

Illian, J., Benson, E., Crawford, J. and Staines, H. (2006). Principal component analysis for spatial point processes—assessing the appropriateness of the approach in an ecological context. In *Case Studies in Spatial Point Process Modeling*, 135-150. Springer, New York.

Makabe, H. and Morimura, H. (1955). A normal approximation to poisson distribution. *Reports of Statistical Applications Research Union of Japanese Scientists and Engineers* **4**, 31-40.

Müller, H.-G. and Yao, F. (2008). Functional additive models. *J. Amer. Statist. Assoc.* **103**, 1534-1544.

Ockenfels, A. and Roth, A. B. (2002). The timing of bids in internet auctions: Market design, bidder behavior, and artificial agents. *AI Magazine* **23**.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd edition. Springer, New York.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53**, 233-243.

Russo, R. P., Shmueli, G. and Shyamalkumar, N. D. (2008). Models of bidder activity consistent with self-similar bid arrivals. In *Statistical Methods in eCommerce Research* (Edited by W. Jank and G. Shmueli). Wiley, New York.

Sears, M., Taylor, D., Print, C., Lake, D., Li, Q., Flannery, E., Yates, D., Lucas, M. and Herbison, G. (1990). Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* **336**, 1391-1396.

Shiffman, S., Gwaltney, C., Baladanis, M. H., Liu, K., Paty, J., Kassel, J. D., Hickcox, M. and Gnys, M. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology* **111**, 531-545.

Shmueli, G., Russo, R. P. and Jank, W. (2007). The barista: a model for bid arrivals in online auctions. *Ann. Appl. Statist.* **1**, 412-441.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24**, 1-24.

Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Statist.* **11**, 1119-1135.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Department of Biostatistics and Computational Biology, University of Rochester, NY 14642, USA.

E-mail: shuang_wu@urmc.rochester.edu

Department of Statistics, University of California, Davis, CA 95616, USA.

E-mail: hgmueller@ucdavis.edu

Abbott Vascular Inc., 3200 Lakeside Drive, Santa Clara, CA 95054, USA.

E-mail: zhen.zhang@av.abbott.com