Adv. Appl. Prob. **48,** 1117–1138 (2016)
*doi:*10.1017/apr.2016.67

NONPARAMETRIC ESTIMATION OF THE SERVICE TIME DISTRIBUTION IN THE M/G/∞ QUEUE

ALEXANDER GOLDENSHLUGER,* *University of Haifa*

Abstract

The subject of this paper is the problem of estimating the service time distribution of the M/G/∞ queue from incomplete data on the queue. The goal is to estimate G from observations of the queue-length process at the points of the regular grid on a fixed time interval. We propose an estimator and analyze its accuracy over a family of target service time distributions. An upper bound on the maximal risk is derived. The problem of estimating the arrival rate is considered as well.

Keywords: M/G/∞ queue; nonparametric estimation; minimax risk; stationary process; covariance function; rates of convergence

2010 Mathematics Subject Classification: Primary 62G05
Secondary 60K25, 62M09

1. Introduction

Suppose that customers enter a system at time instances $\{\tau_j,\ j \in \mathbb{Z}\}$, obtain service upon arrival, and leave the system at time instances $\{y_j,\ j \in \mathbb{Z}\}$ after the service is completed. A jth customer arriving at τ_j requires service time σ_j, so that its departure epoch is $y_j = \tau_j + \sigma_j$. If $\{\tau_j,\ j \in \mathbb{Z}\}$ is a realization of a stationary Poisson process on $\mathbb{R}$, and $\{\sigma_j,\ j \in \mathbb{Z}\}$ are nonnegative independent random variables with common distribution G, independent of $\{\tau_j,\ j \in \mathbb{Z}\}$, then the above description corresponds to the stationary M/G/∞ queueing system. The subject of this paper is estimating the service time distribution G from incomplete data on the queue.

The M/G/∞ system is perhaps one of the most widely studied models in queueing theory; its probabilistic properties are fairly well understood. However, statistical inference in such models has attracted little attention.

The problem of estimating the service time distribution G in the M/G/∞ queue has been studied under different assumptions on the available data. In particular, the following three observation schemes have been considered in the literature:

(i) observation of arrival $\{\tau_j,\ j \in \mathbb{Z}\}$ and departure $\{y_j,\ j \in \mathbb{Z}\}$ epochs without their matchings;

(ii) observation of the queue-length (number-of-busy-servers) process $\{X(t)\}$;

(iii) observation of the busy-period process $\{\mathbf{1}(X(t) > 0)\}$, where $\mathbf{1}$ is the indicator function.

We note that observation schemes (i) and (ii) are equivalent up to initial conditions on the queue length. In particular, arrival and departure epochs are uniquely determined by the queue-

Received 25 August 2015; revision received 14 January 2016.
* Postal address: Department of Statistics, University of Haifa, Haifa, 31905, Israel.
Email address: goldensh@stat.haifa.ac.il

length process, while the queue length can be reconstructed from the input–output data provided that the initial state of the queue is known.

In setting (i), Brown (1970) proposed an estimator of G which is based on the idea of pairing every departure epoch with the closest arrival epoch to the left. Differences between these epochs constitute an ergodic stationary random sequence whose marginal distribution is related to the service time distribution G by a simple formula. Then estimation of G can be achieved by inverting the formula and substituting the empirical marginal distribution of the differences. Brown (1970) proved consistency of the proposed estimator. Recently, Blanghaps *et al.* (2013) extended the work of Brown; they showed that pairing of a departure epoch with the r-closest arrival epoch to the left can be worthwhile.

Nonparametric estimation of service time distribution G under observation schemes (ii) and (iii) was considered in Bingham and Pitts (1999). It is well known that in the steady-state the queue-length process $\{X(t)\}$ is stationary with the Poisson marginal distribution and correlation function

$$H(t) = 1 - G^*(t), \qquad G^*(t) := \left[\int_0^\infty [1 - G(x)]\,\mathrm{d}x\right]^{-1} \int_0^t [1 - G(x)]\,\mathrm{d}x; \tag{1}$$

see, e.g. Beneš (1957) and Reynolds (1975). This fact suggests that function G^* can be reconstructed by estimating the correlation function of the queue-length process. The work of Bingham and Pitts (1999) discusses this approach and provides standard results from the time series literature for estimators of G^*. The idea of reconstructing the service time distribution from correlation structure of the queue-length process was also exploited by Pickands and Stine (1997). The model considered in that paper assumes that a Poisson number of customers arrives at discrete times $1, 2, \ldots, T$, and the service times are independent and identically distributed random variables taking values in the set of nonnegative integer numbers. In this discrete setting, estimation of the service time distribution is equivalent to estimating a linear form of the correlation function of the queue-length process. For the latter problem, standard results from the time series literature are applicable. Other related work is reported in Brillinger (1974), Bingham and Dunham (1997), Hall and Park (2004), Moulines *et al.* (2007), Grübel and Wegener (2011), Schweer and Wichelhaus (2015); see Blanghaps *et al.* (2013) for additional references.

Although estimation of G under different observation schemes was considered in the literature, the most interesting and important statistical questions remain open. In particular, it is not clear what is the achievable estimation accuracy in such problems, and how to construct optimal estimators. The goal of this paper is to shed light on some of these issues.

In this work we adopt the minimax approach for measuring estimation accuracy. It is assumed that the estimated distribution G belongs to a given functional class, and accuracy of any estimator is measured by its worst case mean-squared error on the class. The functional class is defined in terms of restrictions on the smoothness and tail behavior of G (for precise definitions, see Section 4). We concentrate on the observation scheme (ii) when the queue-length process is observed on a fixed interval at the points of the regular grid. The goal is to estimate G at a fixed point from such observations. From now on we will refer to this setting as *the M/G/∞ estimation problem*.

We develop an estimator of G which is based on the relationship between distribution G and the covariance function of the queue-length process, as discussed in Bingham and Pitts (1999) and Pickands and Stine (1997) (see (1)). In particular, estimating G at a fixed point is reduced to estimating the first derivative of the covariance function of the queue-length process at this point. We analyze the accuracy of our estimator over a suitable class of target distributions and derive an upper bound on the maximal risk. The upper bound is expressed in terms of the functional class parameters and the observation horizon. The problem of estimating the arrival rate is discussed as well.

The rest of this paper is structured as follows. Section 2 contains formal statement of the M/G/∞ estimation problem. In Section 3 we present some results on properties of the queue-length process; these results are instrumental for subsequent developments in the paper. In Section 4 we consider the M/G/∞ estimation problem, define our estimator, and establish upper bounds on its maximal risk. In Section 5, we deal with the problem of estimating the arrival rate. Section 6 contains discussion; proofs and technical statements are presented in Section 7.

2. Problem formulation

Let $\{\tau_j,\ j \in \mathbb{Z}\}$ be arrival epochs constituting a realization of the stationary Poisson process point process of intensity λ on the real line. The service times $\{\sigma_j,\ j \in \mathbb{Z}\}$ are positive independent random variables with common distribution G, independent of $\{\tau_j,\ j \in \mathbb{Z}\}$. Assume that the system is in the steady-state; then the queue-length process $\{X(t),\ t \in \mathbb{R}\}$ is given by

$$X(t) = \sum_{j \in \mathbb{Z}} \mathbf{1}(\tau_j \le t,\ \sigma_j > t - \tau_j), \qquad t \in \mathbb{R}. \tag{2}$$

Suppose that $X(t)$ is observed on the time interval $[0, T]$ at the points of the regular grid $t_i = i\delta$, $i = 1, \dots, n$, where $\delta > 0$ is the sampling interval, and $T = n\delta$. Denote $X^n = (X(t_1), \dots, X(t_n)) \in \mathbb{R}^n_+$. Our goal is to estimate the distribution function G at a single given point $x_0 \in \mathbb{R}_+$ using observation X^n. In Section 5 we also discuss the problem of estimating the arrival rate λ from observation X^n.

Distribution of the observation X^n is fully characterized by the service time distribution G and by the arrival rate λ. From now on $\mathbb{P}_{G,\lambda}$ stands for the probability measure generated by $\{\tau_j,\ j \in \mathbb{Z}\}$ and $\{\sigma_j,\ j \in \mathbb{Z}\}$ when the σ_j are distributed G, and the arrival rate is λ. Correspondingly, $\mathbb{E}_{G,\lambda}$ is the expectation with respect to $\mathbb{P}_{G,\lambda}$. In the problem of estimating G when the arrival rate λ is known, we use notation $\mathbb{P}_G$ and $\mathbb{E}_G$ for the probability measure and expectation, respectively.

By estimator $\hat{G}(x_0) = \hat{G}(X^n; x_0)$ of $G(x_0)$ we mean any measurable function of the observation X^n. We adopt the minimax approach for measuring estimation accuracy. Let $\mathscr{G}$ be a class of distribution functions; then accuracy of $\hat{G}(x_0)$ is measured by the maximal mean-squared risk over the class, i.e.

$$\mathcal{R}_{x_0}[\hat{G}; \mathscr{G}] = \sup_{G \in \mathscr{G}} [\mathbb{E}_G |\hat{G}(x_0) - G(x_0)|^2]^{1/2}.$$

The minimax risk is defined by

$$\mathcal{R}^*_{x_0}[\mathscr{G}] = \inf_{\hat{G}} \mathcal{R}_{x_0}[\hat{G}; \mathscr{G}],$$

where inf is taken over all possible estimators. We want to develop a *rate-optimal* (*optimal in order*) estimator $\tilde{G}(x_0)$ such that

$$\mathcal{R}_{x_0}[\tilde{G}; \mathscr{G}] \le C \mathcal{R}^*_{x_0}[\mathscr{G}],$$

where C is a constant independent of the observation horizon T and the sampling interval δ.

In the problem of estimating the arrival rate λ from observation X^n the estimation accuracy is measured similarly. If $\hat{\lambda} = \hat{\lambda}(X^n)$ is an estimator of λ (a measurable function of X^n) then the maximal risk of $\hat{\lambda}$ is defined by

$$\mathcal{R}[\hat{\lambda}; \mathcal{G}] = \sup_{G\in\mathcal{G}} [\mathbb{E}_{G,\lambda}|\hat{\lambda} - \lambda|^2]^{1/2}.$$

We will consider functional classes $\mathcal{G}$ which impose restrictions on smoothness and tail behavior of the distribution functions. The corresponding definitions are given in Section 4.

3. Properties of the queue-length process

Let

$$\frac{1}{\mu} := \mathbb{E}_G[\sigma] = \int_0^\infty [1 - G(t)]\,\mathrm{d}t < \infty$$

with μ being *the service rate*, and let $\rho := \lambda/\mu$ be the *traffic intensity*. Define

$$H(t) := \mu \int_t^\infty [1-G(x)]\,\mathrm{d}x = \biggl[\int_0^\infty [1-G(x)]\,\mathrm{d}x\biggr]^{-1} \int_t^\infty [1-G(x)]\,\mathrm{d}x, \qquad t \in \mathbb{R}_+. \quad (3)$$

The function $G^* := 1 - H$ is often called *the stationary-excess cumulative distribution function* (see, e.g. Whitt (1985)). If G is a distribution function of an interval between points in a renewal process, then G^* represents a distribution function of the interval between arbitrary time and the next renewal point. In our context, the important role of H stems from the fact that it is the correlation function of the queue-length process $\{X(t),\ t \in \mathbb{R}\}$; see Proposition 1 below.

Observe that $H(0) = 1$, and H is nonnegative and monotone decreasing on the positive real line. Although function H is defined on $\mathbb{R}_+$ only, it will be convenient to extend its definition to the whole real line $\mathbb{R}$ by setting $H(t) = H(-t)$ for $t < 0$. For the sake of brevity, from now on we use the suffix notation $X_i = X(t_i) = X(i\delta)$, $H_i = H(t_i) = H(i\delta)$, $i = 1, \dots, n$, and so on.

Proposition 1. *The following statements hold.*

(i) *For any $t \in \mathbb{R}$ the distribution of $X(t)$ is Poisson with parameter ρ.*

(ii) *For any $t, s \in \mathbb{R}$, $\mathbb{E}_G[X(t)X(s)] = \rho^2 + \rho H(t - s)$.*

(iii) *For any $\theta = (\theta_1, \dots, \theta_n)$, $n \geq 1$, we have*

$$\log \mathbb{E}_G\biggl[\exp\biggl\{\sum_{i=1}^n \theta_i X_i\biggr\}\biggr] = \rho S_n(\theta), \quad (4)$$

$$S_n(\theta) := \sum_{m=1}^n (\exp\{\theta_m\} - 1) + \sum_{k=1}^{n-1} H_k \sum_{m=k}^{n-1} (\exp\{\theta_{m-k+1}\} - 1) \exp\biggl\{\sum_{i=m-k+2}^{m} \theta_i\biggr\} (\exp\{\theta_{m+1}\} - 1). \quad (5)$$

In particular, if $\theta^* := (\vartheta, \dots, \vartheta)$ for some $\vartheta \in \mathbb{R}$ then

$$S_n(\theta^*) = n(\exp\{\vartheta\} - 1) + n(\exp\{\vartheta\} - 1)^2 \sum_{k=1}^{n-1} \biggl(1 - \frac{k}{n}\biggr) \exp\{(k-1)\vartheta\} H_k.$$

Remark 1. (i) The statements in (i) and (ii) are well known; in fact, they are immediate consequences of (iii). The first statement can be found in many textbooks (see, e.g. Parzen (1962, p. 147) and Ross (1970, p. 19)), while the second one appears, e.g. in Beneš (1957) and Reynolds (1975). As for part (iii), Lindley (1956) considered the special case of $n = 3$ and discussed heuristically a derivation for general n. However, we could not find (4) and (5) in the literature, and, to the best of our knowledge, they are new. Equations (4) and (5) play an important role in subsequent derivations.

(ii) The joint distribution of X^n is the so-called multivariate Poisson; for details see, e.g. Lindley (1956, Section 2) and Milne (1970). The statements in (i) and (ii) show that H is the correlation function of the process $\{X(t),\ t \in \mathbb{R}\}$.

It is instructive to realize the form of (4) and (5) in the special case $n = 4$. Let $1 \le i \le j \le k \le m \le n$; then

$$\begin{aligned}
&\frac{1}{\rho} \log \mathbb{E}_G[\exp\{\theta_1 X_i + \theta_2 X_j + \theta_3 X_k + \theta_4 X_m\}] \\
&\quad = \sum_{l=1}^{4} (\exp\{\theta_l\} - 1) + H_{j-i}(\exp\{\theta_1\} - 1)(\exp\{\theta_2\} - 1) \\
&\qquad + H_{k-i}(\exp\{\theta_1\} - 1)\exp\{\theta_2\}(\exp\{\theta_3\} - 1) \\
&\qquad + H_{m-i}(\exp\{\theta_1\} - 1)\exp\{\theta_2 + \theta_3\}(\exp\{\theta_4\} - 1) + H_{k-j}(\exp\{\theta_2\} - 1)(\exp\{\theta_3\} - 1) \\
&\qquad + H_{m-j}(\exp\{\theta_2\} - 1)\exp\{\theta_3\}(\exp\{\theta_4\} - 1) + H_{m-k}(\exp\{\theta_3\} - 1)(\exp\{\theta_4\} - 1).
\end{aligned} \tag{6}$$

As is seen from the above equation, the first term on the right-hand side of (6) coincides with the cumulant generating function of independent Poisson random variables. The other terms are associated with all possible pairs of random variables. For every pair of random variables the corresponding term contains the correlation between the variables, and the factors $(1 - \exp\{\theta\})$ and $\exp\{\theta\}$, where the $(1 - \exp\{\theta\})$-factors correspond to the pair, and the $\exp\{\theta\}$-factors correspond to the random variables 'sandwiched' by the pair.

As a by-product of Proposition 1(iii), we can easily obtain the following Gaussian approximation to the finite-dimensional distributions of the queue-length process $\{X(t),\ 0 \le t \le T\}$.

Proposition 2. *Consider a sequence of the M/G/∞ queueing systems,* $\{\mathrm{M}_l/\mathrm{G}/\infty,\ l = 1, 2, \dots\}$*, with the fixed service time distribution G, and with the lth system characterized by the arrival rate* $\lambda_l = l\lambda$*,* $\lambda > 0$*. Let* $X_l^n = (X_{l,1}, \dots, X_{l,n}) = (X_l(t_1), \dots, X_l(t_n))$ *be the vector of observations of the queue-length process (2) in the lth system; then*

$$\frac{X_l^n - l\rho e_n}{\sqrt{l\rho}} \xrightarrow{\mathrm{D}} \mathcal{N}_n(0, \Sigma(H)), \qquad l \to \infty,$$

where $\rho = \lambda/\mu$*,* $e_n = (1, \dots, 1) \in \mathbb{R}^n$*,* $\Sigma(H) := \{H((i - j)\delta)\}_{i,j=1,\dots,n}$*, and '*$\xrightarrow{\mathrm{D}}$*' denotes convergence in distribution.*

The result of Proposition 2 is well known; it is in line with more general weak convergence results for queues in Borovkov (1967), Iglehart (1973), and Whitt (1974). The proof of Proposition 2 follows immediately from Proposition 1(iii), and it is omitted.

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

4. Estimation of service time distribution

According to Proposition 1(ii) the covariance function of the queue-length process is

$$R(t) := \mathrm{cov}_G\{X(s),\ X(s+t)\} = \rho H(t).$$

Therefore, differentiation yields

$$1 - G(t) = -\frac{1}{\lambda} R'(t), \qquad t \in \mathbb{R}_+. \tag{7}$$

This relationship is the basis for construction of our estimator of $G(x_0)$.

4.1. Estimator construction

Let

$$\hat{\rho}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} X_i, \qquad k = 0, 1, \dots, n-1,$$

and define

$$\hat{R}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} (X_i - \hat{\rho}_k)(X_{i+k} - \hat{\rho}_k), \qquad k = 0, 1, \dots, n-1. \tag{8}$$

Note that $\hat{R}_k$ is the empirical estimator of the covariance $R_k = R(k\delta) = \rho H(k\delta)$, $k = 0, 1, \dots, n-1$. For technical reasons, we use estimator $\hat{\rho}_k$ based on $n-k$ observations and not on n.

Let $h > 0$, and for every $x \in [0, T-\delta]$ define the segment

$$D_x := \begin{cases} [x-h, x+h], & h < x \le T - \delta - h, \\ [0, 2h], & 0 \le x \le h, \\ [T-\delta-2h, T-\delta], & T-\delta-h < x \le T-\delta. \end{cases}$$

Let M_{D_x} be the set of indexes $k \in \{1, \dots, n\}$ such that $k\delta \in D_x$, $M_{D_x} := \{k \colon k\delta \in D_x\}$, and let N_{D_x} be the cardinality of this set, $N_{D_x} := \#\{M_{D_x}\}$.

Fix positive integer ℓ, and assume that

$$h \ge \tfrac{1}{2}(\ell + 2)\delta. \tag{9}$$

For $x \in [0, T-\delta]$ let $\{a_k(x),\ k \in M_{D_x}\}$ denote the weights obtained as the solution to the following optimization problem $(\mathcal{P}_x)$:

$$\begin{gathered} \min \sum_{k \in M_{D_x}} a_k^2(x) \\ \text{subject to} \quad \sum_{k \in M_{D_x}} a_k(x) = 0, \qquad \sum_{k \in M_{D_x}} a_k(x)(k\delta)^j = j x^{j-1}, \qquad j = 1, \dots, \ell. \end{gathered} \tag{10}$$

We use the convention that if $x = 0$ and $j = 1$ then the right-hand side of the last constraint in (10) is equal to 1.

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

By definition, if (9) holds then the linear filter associated with the weights $\{a_k(x),\ k \in M_{D_x}\}$ has the following property: it reproduces without error the first derivative of any polynomial p of $\deg(p) \le \ell$ at point x,

$$\sum_{k \in M_{D_x}} a_k(x) p(k\delta) = p'(x) \quad \text{for all } p\colon \deg(p) \le \ell. \tag{11}$$

Now we are in a position to define our estimator of $G(x_0)$: it is given by

$$\hat{G}_h(x_0) = 1 + \frac{1}{\lambda} \sum_{k \in M_{D_{x_0}}} a_k(x_0) \hat{R}_k, \tag{12}$$

where $\hat{R}_k = \hat{R}(k\delta)$, $k = 0, \dots, n-1$, are defined in (8).

The expression under the summation sign on the right-hand side of (12) can be viewed as a local polynomial estimator of the derivative $R'(x_0)$ when the empirical covariances $\hat{R}_k$ are regarded as noisy observations of $R_k = R(k\delta)$. We refer the reader to Goldenshluger and Nemirovski (1997) for a similar construction of the local polynomial estimators of derivatives in the context of the nonparametric regression model.

The estimator $\hat{G}_h(x_0)$ depends on two design parameters, the *window width* h and the *degree of polynomial* ℓ; these parameters are specified in the sequel.

4.2. Upper bound on the maximal risk

Our current goal is to study the accuracy of $\hat{G}_h(x_0)$. For this purpose, we introduce the functional class of distributions G over which accuracy of estimator $\hat{G}_h(x_0)$ is assessed.

Definition 1. (i) Let $\beta > 0$ and $L > 0$ be real numbers, and let $I \subset (0, \infty)$ be a closed interval such that $x_0 \in I$. We define $\mathscr{H}_\beta(L, I)$ to be the class of all distribution functions G on $\mathbb{R}_+$ such that G is $\lfloor \beta \rfloor$ times continuously differentiable on I, and

$$|G^{(\lfloor \beta \rfloor)}(x) - G^{(\lfloor \beta \rfloor)}(y)| \le L|x-y|^{\beta - \lfloor \beta \rfloor} \quad \text{for all } x, y \in I;$$

here $\lfloor \beta \rfloor$ stands for the maximal integer number strictly less than β.

(ii) We say that the distribution function G on $\mathbb{R}_+$ belongs to the class $\mathscr{M}_p(K)$, $p \ge 1$, $K > 0$ if

$$\mathbb{E}_G[\sigma^p] = \int_0^\infty p x^{p-1}[1 - G(x)]\,\mathrm{d}x \le K < \infty.$$

(iii) Finally, we set

$$\mathscr{G}_\beta(L, I, K) := \mathscr{H}_\beta(L, I) \cap \mathscr{M}_2(K).$$

Remark 2. (i) The class $\mathscr{G}_\beta(L, I, K)$ imposes restrictions on the smoothness in the vicinity of x_0. In all what follows the point x_0 is assumed to be fixed. If x_0 is separated away from 0 then we always consider a symmetric interval I centered at x_0: $I = [x_0 - d,\ x_0 + d]$ for some $0 < d < x_0$. In the $x_0 = 0$ case we set $I = [0, 2d]$.

(ii) The definition of $\mathscr{G}_\beta(L, I, K)$ requires the boundedness of the second moment of the service time distribution. This condition implies that the correlation sequence $\{H(k\delta),\ k \in \mathbb{Z}\}$ is summable, which corresponds to the *short–term dependence* between the values of the sampled discrete-time queue-length process (see, e.g. Brockwell and Davis (1991, Section 13.2)). This assumption can be relaxed. However, we do not pursue the case of the long-term dependence in this paper.

Now we are in a position to state an upper bound on the maximal risk of our estimator.

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

Theorem 1. *Let x_0 be fixed, $I := [x_0 - d,\ x_0 + d] \subset [0, (1-\kappa)T]$ for some $\kappa \in (0, 1)$, and suppose that $G \in \mathcal{G}_\beta(L, I, K)$. Let $\hat{G}_*(x_0)$ be the estimator defined in (12) and associated with the degree $\ell \geq \lfloor \beta \rfloor + 1$ and the window width*

$$h = h_* := \left[\frac{K(\sqrt{K} \vee 1)}{L^2 \kappa T}\left(1 + \frac{1}{\lambda}\right)\right]^{1/(2\beta+2)}. \tag{13}$$

If

$$\frac{K(\sqrt{K} \vee 1)}{L^2 \kappa}\left(1 + \frac{1}{\lambda}\right) d^{-2\beta-2} \leq T \leq \frac{K(\sqrt{K} \vee 1)}{L^2 \kappa}\left(1 + \frac{1}{\lambda}\right)\left[\frac{2}{(\ell+2)\delta}\right]^{2\beta+2} \tag{14}$$

then we have

$$\mathcal{R}_{x_0}[\hat{G}_*; \mathcal{G}_\beta(L, I, K)] \leq C L^{1/(\beta+1)}\left[\frac{K(\sqrt{K} \vee 1)}{\kappa T}\left(1 + \frac{1}{\lambda}\right)\right]^{\beta/(2\beta+2)}, \tag{15}$$

where $C = C(\ell)$ depends on ℓ only.

Remark 3. (i) The upper bound in (14) originates in the requirement that the segment D_{x_0} contains at least $\ell + 1$ grid points. This inequality is fulfilled if sampling is fast enough, $\delta \leq O((\kappa T)^{-1/(2\beta+2)})$. Thus, if the asymptotics as $T \to \infty$ is considered then δ should tend to 0 so that (14) is fulfilled. The lower bound in (14) ensures that $D_{x_0} \subseteq I$.

(ii) The bound in (15) is nonuniform in x_0; it is established for fixed $x_0 \leq (1-\kappa)T$. The bound increases as κ gets closer to 0 (x_0 approaches T). This fact is not surprising: the empirical covariance estimator is not accurate for large lags. However, if x_0 is large in comparison with T then it is advantageous to use the trivial estimator $\tilde{G}(x_0) = 1$. The risk of $\tilde{G}(x_0)$ admits the following upper bound:

$$\mathcal{R}_{x_0}[\tilde{G}; \mathcal{G}_\beta(L, I, K)] \leq K x_0^{-2} \quad \text{for all } x_0 \in \mathbb{R}_+. \tag{16}$$

Indeed, it follows from $G \in \mathcal{M}_2(K)$ that, for any x,

$$1 - G(x) = \int_x^\infty \mathrm{d}G(t) \leq x^{-2} \int_x^\infty t^2 \,\mathrm{d}G(t) \leq K x^{-2}.$$

Thus, $G(x) \geq 1 - K x^{-2}$, which implies (16). Comparing (15) and (16), we see that for $x_0 \leq O(T^{\beta/(4\beta+4)})$ it is advantageous to use the estimator $\hat{G}_*(x_0)$; otherwise $\tilde{G}(x_0)$ is better. If more stringent conditions on the tail of G are imposed (e.g. $G \in \mathcal{M}_p(K)$ with $p > 2$) then the zone where $\hat{G}_*(x_0)$ is preferable becomes smaller.

5. Estimation of the arrival rate

The construction in Section 4.1 that led to $\hat{G}_h(x_0)$ can be used in order to estimate the arrival rate λ from discrete observations of the queue-length process.

Let $I = [0, 2d]$ and assume that $G \in \mathcal{G}_\beta(L, I, K)$. Under this condition, we can use (7) in order to construct an estimator of λ. Indeed, setting $t = 0$ in (7) and taking into account that $G(0) = 0$, we obtain $\lambda = -R'(0)$, where $R'(0)$ is understood here as the right-side derivative of R at 0. Therefore, we define the estimator for λ by

$$\hat{\lambda} = -\sum_{k \in M_{D_0}} a_k(0) \hat{R}_k, \tag{17}$$

where $D_0 := [0, 2h]$, $\{a_k(0),\ k \in M_{D_0}\}$ is the solution to $(\mathscr{P}_0)$ (i.e. (10) with $x = 0$), and $\hat{R}_k$, $k \in M_{D_0}$, are defined in (8).

The next statement provides an upper bound on the risk of $\hat{\lambda}$.

Theorem 2. *Let $I = [0, 2d]$ and suppose that $G \in \mathscr{G}_\beta(L, I, K)$. Let $\hat{\lambda}_*$ denote the estimator defined in (17) and associated with degree $\ell \geq \lfloor \beta \rfloor + 1$ and window width*

$$h = h_* := \left[\frac{K(\sqrt{K} \vee 1)}{L^2 T} \right]^{1/(2\beta+2)}. \tag{18}$$

If

$$K(\sqrt{K} \vee 1) L^{-2} d^{-2\beta-2} \leq T \leq K(\sqrt{K} \vee 1) L^{-2} \left[\frac{2}{(\ell+2)\delta} \right]^{2\beta+2} \tag{19}$$

then we have

$$\sup_{G \in \mathscr{G}_\beta(L,I,K)} [\mathbb{E}_{G,\lambda} |\hat{\lambda}_* - \lambda|^2]^{1/2} \leq C L^{1/(\beta+1)} (\lambda^2 + \lambda)^{1/2} \left[\frac{K(\sqrt{K} \vee 1)}{T} \right]^{\beta/(2\beta+2)}, \tag{20}$$

where $C = C(\ell)$ depends on ℓ only.

Remark 4. (i) The meaning of condition (19) is similar to that of (14); see Remark 3(i).

(ii) If the sampling interval δ is very small then one can build an estimator which is better than $\hat{\lambda}_*$. In particular, if the continuous-time observation $\{X(t),\ 0 \leq t \leq T\}$ is available then alternative estimators of λ can be constructed as follows:

$$\hat{\lambda}^{\uparrow} = \frac{1}{T} \#\{t \in (0, T] \colon X(t) - X(t-) = 1\}, \qquad \hat{\lambda}^{\downarrow} = \frac{1}{T} \#\{t \in (0, T] \colon X(t) - X(t-) = -1\}.$$

Since arrivals and departures constitute the Poisson process with intensity λ, the mean-squared errors of $\hat{\lambda}^{\uparrow}$ and $\hat{\lambda}^{\downarrow}$ are given by

$$\mathbb{E}_{G,\lambda} |\hat{\lambda}^{\uparrow} - \lambda|^2 = \mathbb{E}_{G,\lambda} |\hat{\lambda}^{\downarrow} - \lambda|^2 = \lambda T^{-1} \quad \text{for all } \lambda, \text{ for all } G.$$

Thus, in terms of dependence on the observation horizon T, the risks of $\hat{\lambda}^{\uparrow}$ and $\hat{\lambda}^{\downarrow}$ tend to 0 at the parametric rate $O(1/T)$. This rate is faster than the one in (20). However, if the estimator $\hat{\lambda}^{\uparrow}$ (or $\hat{\lambda}^{\downarrow}$) is applied with discrete observations X^n, its accuracy deteriorates very rapidly with the growth of δ.

6. Discussion

Theorem 1 indicates that under suitable relations between the observation horizon T and the sampling interval δ the service time distribution G can be estimated with the risk of the order $T^{-\beta/(2\beta+2)}$. In particular, for our estimator $\hat{G}_*(x_0)$, we have

$$\mathscr{R}_{x_0}[\hat{G}_*; \mathscr{G}_\beta(L, I, K)] \asymp O(T^{-\beta/(2\beta+2)}), \qquad T \to \infty,$$

provided that (14) holds. A natural question is if this rate of convergence is optimal in the minimax sense. This calls for a lower bound on the minimax risk $\mathscr{R}^*_{x_0}[\mathscr{G}_\beta(L, I, K)]$.

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

Although statement Proposition 1(iii) provides complete probabilistic characterization of finite-dimensional distributions of the queue-length process $\{X(t),\ t \in \mathbb{R}\}$, there is no explicit formula available for the distribution of X^n. Since all existing techniques for the derivation of lower bounds on minimax risks rely upon sensitivity analysis of the family of target distributions, such a derivation for the M/G/∞ estimation problem seems to be intractable. However, some understanding of accuracy limitations in estimating the service time distribution can be gained from consideration of a Gaussian approximating model.

In particular, in Proposition 2 we show that if the arrival rate λ is large, the finite-dimensional distributions of $\{X(t),\ 0 \le t \le T\}$ are close to Gaussian. Thus, for large arrival rates (heavy-traffic regime) we can regard the properly centered and scaled queue-length process as a zero mean stationary Gaussian process. Furthermore, from (7) we see that the service time distribution G is proportional to the first derivative of the covariance function of the queue-length process. This characterization suggests that in the heavy-traffic regime estimating G is as difficult as estimating the first derivative of the covariance function of a continuous-time stationary Gaussian process from discrete observations. Although we do not have a formal proof for the equivalence of the corresponding statistical experiments, this assumption seems plausible.

Goldenshluger (2015) discussed the relationship between the two settings, and studied the problem of estimating covariance function derivatives of a stationary Gaussian process from discrete observations. In particular, it was shown there that under conditions compatible with those of the M/G/∞ estimation problem, the minimax risk in estimating the first derivative of the covariance function cannot be less than $O(T^{-\beta/(2\beta+2)})$. This fact strongly suggests that the proposed estimator $\hat{G}_*$ is rate optimal in the heavy-traffic regime. However, in general, construction of rate optimal estimators in the original M/G/∞ estimation problem remains an open problem.

7. Proofs and auxiliary results

7.1. Proof of Proposition 1

For any $m > 1$, we write

$$\mathbb{E}_G \exp\bigg\{\sum_{i=1}^m \theta_i X_i\bigg\} = \mathbb{E}_G\bigg\{\mathbb{E}_G\bigg[\exp\bigg\{\sum_{i=1}^m \theta_i X_i\bigg\} \,\bigg|\, \{\tau_j,\ j \in \mathbb{Z}\}\bigg]\bigg\}. \tag{21}$$

By (2) and the independence of $\{\tau_j,\ j \in \mathbb{Z}\}$ and $\{\sigma_j,\ j \in \mathbb{Z}\}$, the conditional expectation in (21) takes the form

$$\begin{aligned}
&\mathbb{E}_G\bigg[\exp\bigg\{\sum_{i=1}^m \theta_i X_i\bigg\} \,\bigg|\, \{\tau_j,\ j \in \mathbb{Z}\}\bigg] \\
&\quad = \mathbb{E}_G\bigg[\exp\bigg\{\sum_{j\in\mathbb{Z}}\sum_{i=1}^m \theta_i\, \mathbf{1}(\tau_j \le t_i, \sigma_j > t_i - \tau_j)\bigg\} \,\bigg|\, \{\tau_j,\ j \in \mathbb{Z}\}\bigg] \\
&\quad = \prod_{j\in\mathbb{Z}} \mathbb{E}_G\bigg[\exp\bigg\{\sum_{i=1}^m \theta_i\, \mathbf{1}(\tau_j \le t_i, \sigma_j > t_i - \tau_j)\bigg\} \,\bigg|\, \{\tau_j,\ j \in \mathbb{Z}\}\bigg].
\end{aligned} \tag{22}$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

Given $x \in \mathbb{R}$, consider partition of the real line by the intervals $I_0(x) = (-\infty, t_1 - x]$, $I_k(x) = (t_k - x, t_{k+1} - x]$, $k = 1, \dots, m-1$, and $I_m(x) = (t_m - x, \infty)$. With this notation

$$\begin{aligned}\mathbb{E}_G\Bigg[\exp\Bigg\{\sum_{i=1}^m \theta_i\, \mathbf{1}(\tau_j \le t_i, \sigma_j > t_i - \tau_j)\Bigg\} \,\Bigg|\, \{\tau_j,\ j \in Z\}\Bigg] \\ = \mathbb{P}_G\{\sigma_j \in I_0(\tau_j)\} + \sum_{k=1}^m \exp\Bigg\{\sum_{i=1}^k \theta_i\, \mathbf{1}(\tau_j \le t_i)\Bigg\}\mathbb{P}_G\{\sigma_j \in I_k(\tau_j)\} \\ = 1 + \sum_{k=1}^m \Bigg[\exp\Bigg\{\sum_{i=1}^k \theta_i\, \mathbf{1}(\tau_j \le t_i)\Bigg\} - 1\Bigg]\mathbb{P}_G\{\sigma_j \in I_k(\tau_j)\}.\end{aligned}$$

If we let

$$f(x) = \log\Bigg(1 + \sum_{k=1}^m \Bigg[\exp\Bigg\{\sum_{i=1}^k \theta_i\, \mathbf{1}(x \le t_i)\Bigg\} - 1\Bigg]\mathbb{P}_G\{\sigma_j \in I_k(x)\}\Bigg),$$

then in view of (21), (22), and Campbell's theorem (see Kingman (1993, Section 3.2)), we obtain

$$\mathbb{E}_G \exp\Bigg\{\sum_{i=1}^m \theta_i X_i\Bigg\} = \mathbb{E}_G \exp\Bigg\{\sum_{j\in\mathbb{Z}} f(\tau_j)\Bigg\} = \exp\Bigg\{\lambda \int_{-\infty}^{\infty} [\exp\{f(x)\} - 1]\, \mathrm{d}x\Bigg\}.$$

Denote $S_m(\theta) = \mu \int_{-\infty}^{\infty} [\exp\{f(x)\} - 1]\, \mathrm{d}x$; our current goal is to compute this integral. We have

$$\begin{aligned}&\int_{-\infty}^{\infty} [\exp\{f(x)\} - 1]\, \mathrm{d}x \\ &\quad = \sum_{k=1}^{m-1} \int_{-\infty}^{\infty} \Bigg(\exp\Bigg\{\sum_{i=1}^k \theta_i\, \mathbf{1}(x \le t_i)\Bigg\} - 1\Bigg)[\bar G(t_k - x) - \bar G(t_{k+1} - x)]\, \mathrm{d}x \\ &\qquad + \int_{-\infty}^{\infty} \Bigg(\exp\Bigg\{\sum_{i=1}^m \theta_i\, \mathbf{1}(x \le t_i)\Bigg\} - 1\Bigg)\bar G(t_m - x)\, \mathrm{d}x \\ &\quad =: \sum_{k=1}^{m-1} J_k + L_m,\end{aligned}$$

where we denoted, for brevity, $\bar G = 1 - G$. For $k = 1, \dots, m-1$, we obtain

$$\begin{aligned}J_k &= \Bigg(\exp\Bigg\{\sum_{i=1}^k \theta_i\Bigg\} - 1\Bigg) \int_{-\infty}^{t_1} [\bar G(t_k - x) - \bar G(t_{k+1} - x)]\, \mathrm{d}x \\ &\quad + \sum_{j=1}^{k-1} \Bigg(\exp\Bigg\{\sum_{i=j+1}^k \theta_i\Bigg\} - 1\Bigg) \int_{t_j}^{t_{j+1}} [\bar G(t_k - x) - \bar G(t_{k+1} - x)]\, \mathrm{d}x\end{aligned}$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{1}{\mu}\bigg(\exp\bigg\{\sum_{i=1}^{k}\theta_i\bigg\}-1\bigg)[H(t_k-t_1)-H(t_{k+1}-t_1)] \\
&\quad + \frac{1}{\mu}\sum_{j=1}^{k-1}\bigg(\exp\bigg\{\sum_{i=j+1}^{k}\theta_i\bigg\}-1\bigg)[H(t_k-t_{j+1})-H(t_k-t_j)-H(t_{k+1}-t_{j+1}) \\
&\qquad\qquad + H(t_{k+1}-t_j)] \\
&= \frac{1}{\mu}\bigg(\exp\bigg\{\sum_{i=1}^{k}\theta_i\bigg\}-1\bigg)[H_{k-1}-H_k] \\
&\quad + \frac{1}{\mu}\sum_{j=1}^{k-1}\bigg(\exp\bigg\{\sum_{i=j+1}^{k}\theta_i\bigg\}-1\bigg)[H_{k-j-1}-2H_{k-j}+H_{k-j+1}]. \qquad (23)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
L_m &= \frac{1}{\mu}\bigg(\exp\bigg\{\sum_{i=1}^{m}\theta_i\bigg\}-1\bigg)H(t_m-t_1) \\
&\quad + \frac{1}{\mu}\sum_{j=1}^{m-1}\bigg(\exp\bigg\{\sum_{i=j+1}^{m}\theta_i\bigg\}-1\bigg)[H(t_m-t_{j+1})-H(t_m-t_j)] \\
&= \frac{1}{\mu}\bigg(\exp\bigg\{\sum_{i=1}^{m}\theta_i\bigg\}-1\bigg)H_{m-1}+\frac{1}{\mu}\sum_{j=1}^{m-1}\bigg(\exp\bigg\{\sum_{i=j+1}^{m}\theta_i\bigg\}-1\bigg)[H_{m-j-1}-H_{m-j}]. \qquad (24)
\end{aligned}
$$

The usual convention $\sum_{k=j}^{m}=0$ if $m<j$ is employed in (23) and (24) and from now on.

Note that, by definition, $S_m(\theta)=\mu\sum_{k=1}^{m-1}J_k+\mu L_m$, and we have the following recursive formula:

$$
S_{m+1}(\theta)=S_m(\theta)+\mu(J_m-L_m+L_{m+1}). \qquad (25)
$$

For any $m>1$, using (23) and (24), after straightforward algebraic manipulations, we obtain

$$
\begin{aligned}
&\mu(J_m-L_m+L_{m+1}) \\
&\quad= \bigg(\exp\bigg\{\sum_{i=1}^{n}\theta_i\bigg\}-1\bigg)(H_{m-1}-H_m) \\
&\qquad + \sum_{j=1}^{m-1}\bigg(\exp\bigg\{\sum_{i=j+1}^{m}\theta_i\bigg\}-1\bigg)(H_{m-j-1}-2H_{m-j}+H_{m-j+1}) \\
&\qquad - \bigg(\exp\bigg\{\sum_{i=1}^{m}\theta_i\bigg\}-1\bigg)H_{m-1}-\sum_{j=1}^{m-1}\bigg(\exp\bigg\{\sum_{i=j+1}^{m}\theta_i\bigg\}-1\bigg)(H_{m-j-1}-H_{m-j})
\end{aligned}
$$

$$+\left(\exp\left\{\sum_{i=1}^{m+1}\theta_i\right\}-1\right)H_m+\sum_{j=1}^{m}\left(\exp\left\{\sum_{i=j+1}^{m+1}\theta_i\right\}-1\right)(H_{m-j}-H_{m-j+1})$$
$$=(\exp\{\theta_{m+1}\}-1)+\sum_{k=1}^{m}H_k(\exp\{\theta_{m-k+1}\}-1)\exp\left\{\sum_{i=m-k+2}^{m}\theta_i\right\}(\exp\{\theta_{m+1}\}-1).$$

Taking into account the fact that $S_1(\theta)=\exp\{\theta_1\}-1$ and iterating (25), we obtain

$$\begin{aligned}S_{n+1}(\theta)&=(\exp\{\theta_1\}-1)+\sum_{m=1}^{n}(\exp\{\theta_{m+1}\}-1)\\&\quad+\sum_{m=1}^{n}\sum_{k=1}^{m}H_k(\exp\{\theta_{m-k+1}\}-1)\exp\left\{\sum_{i=m-k+2}^{m}\theta_i\right\}(\exp\{\theta_{m+1}\}-1)\\&=\sum_{m=1}^{n+1}(\exp\{\theta_m\}-1)\\&\quad+\sum_{k=1}^{n}H_k\sum_{m=k}^{n}(\exp\{\theta_{m-k+1}\}-1)\exp\left\{\sum_{i=m-k+2}^{m}\theta_i\right\}(\exp\{\theta_{m+1}\}-1).\end{aligned}$$

This completes the proof. □

7.2. Moments of the queue-length process

Equation (6) allows us to compute the fourth-order mixed moments of the queue-length process.

Proposition 3. *Let* $1\le i\le j\le k\le m\le n$*; then*

$$\begin{aligned}&\mathbb{E}_G[X_iX_jX_kX_m]\\&\quad=\rho^4+\rho^3(H_{j-i}+H_{k-i}+H_{m-i}+H_{k-j}+H_{m-j}+H_{m-k})\\&\qquad+\rho^2(H_{k-i}+H_{m-j}+2H_{m-i}+H_{j-i}H_{m-k}+H_{k-i}H_{m-j}+H_{k-j}H_{m-i})+\rho H_{m-i}.\end{aligned}$$

More generally, for any $i,j,k,m\in\{1,\dots,n\}$ *and any subset* I *of indexes* $I\subseteq\{i,j,k,m\}$ *define* $q_I=\max_{i_1,i_2\in I}|i_1-i_2|$*. Then*

$$\mathbb{E}_G[X_iX_jX_k]=\rho^3+\rho^2(H_{|i-j|}+H_{|k-j|}+H_{|k-i|})+\rho H_{q_{\{i,j,k\}}},$$

$$\begin{aligned}&\mathbb{E}_G[X_iX_jX_kX_m]\\&\quad=\rho^4+\rho^3(H_{|j-i|}+H_{|k-i|}+H_{|m-i|}+H_{|k-j|}+H_{|m-j|}+H_{|m-k|})+\rho H_{q_{\{i,j,k,m\}}}\\&\qquad+\rho^2[H_{q_{\{i,j,k\}}}+H_{q_{\{i,j,m\}}}+H_{q_{\{j,k,m\}}}+H_{q_{\{i,k,m\}}}\\&\qquad\quad+H_{|j-i|}H_{|m-k|}+H_{|k-i|}H_{|m-j|}+H_{|k-j|}H_{|m-i|}].\end{aligned}\tag{26}$$

Proof. The proof involves straightforward though tedious differentiation of (6).

Let $S(\theta)$ denote the right-hand side of (6), where $(\theta_1,\theta_2,\theta_3,\theta_4)$ is replaced by $(\theta_i,\theta_j,\theta_k,\theta_m)$ for convenience. Denote $\psi(\theta)=\mathbb{E}_G\exp\{\theta_iX_i+\theta_jX_j+\theta_kX_k+\theta_mX_m\}$. It is checked by direct calculation that

$$\frac{\partial^4\psi(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_m}=\exp\{-\rho S(\theta)\}[a_1(\theta)\rho+a_2(\theta)\rho^2+a_3(\theta)\rho^3+a_4(\theta)\rho^4],$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

where $a_1(\theta)$, $a_2(\theta)$, $a_3(\theta)$, and $a_4(\theta)$ are given by the following expressions:

$$
\begin{aligned}
a_1(\theta) &= S_{\theta_i\theta_j\theta_k\theta_m},\\
a_2(\theta) &= S_{\theta_i\theta_j\theta_k}S_{\theta_m} + S_{\theta_i\theta_j\theta_m}S_{\theta_k} + S_{\theta_j\theta_k\theta_m}S_{\theta_i} + S_{\theta_i\theta_k\theta_m}S_{\theta_j} + S_{\theta_i\theta_j}S_{\theta_k\theta_m} + S_{\theta_i\theta_k}S_{\theta_j\theta_m}\\
&\quad + S_{\theta_j\theta_k}S_{\theta_i\theta_m},\\
a_3(\theta) &= S_{\theta_i\theta_j}S_{\theta_k}S_{\theta_m} + S_{\theta_i\theta_k}S_{\theta_j}S_{\theta_m} + S_{\theta_i\theta_m}S_{\theta_j}S_{\theta_k} + S_{\theta_j\theta_k}S_{\theta_i}S_{\theta_m} + S_{\theta_j\theta_m}S_{\theta_i}S_{\theta_k}\\
&\quad + S_{\theta_k\theta_m}S_{\theta_i}S_{\theta_j},\\
a_4(\theta) &= S_{\theta_i}S_{\theta_j}S_{\theta_k}S_{\theta_m}.
\end{aligned}
$$

Here we set, for brevity, $S_{\theta_{j_1}\dots\theta_{j_k}} = S_{\theta_{j_1}\dots\theta_{j_k}}(\theta) := \partial^k S(\theta)/\partial\theta_{j_1}\cdots\partial\theta_{j_k}$.

In order to complete the proof, it is sufficient to note that

$$S(0) = 1, \qquad S_{\theta_j}(0) = 1 \quad \text{for all } j, \tag{27}$$

and, for any $j_1 \le j_2 \le j_3 \le j_4$,

$$S_{\theta_{j_1}\theta_{j_2}}(0) = H_{j_2-j_1}, \qquad S_{\theta_{j_1}\theta_{j_2}\theta_{j_3}}(0) = H_{j_3-j_1}, \qquad S_{\theta_{j_1}\dots\theta_{j_4}}(0) = H_{j_4-j_1}. \tag{28}$$

Although (28) is proved for $1 \le i \le j \le k \le m \le n$, a similar result holds more generally. With the introduced definition of q_I, (27) and (28) imply that

$$
\begin{aligned}
a_4(0) &= 1,\\
a_3(0) &= H_{|i-j|} + H_{|k-i|} + H_{|m-i|} + H_{|k-j|} + H_{|m-j|} + H_{|m-k|},\\
a_2(0) &= H_{q_{\{i,j,k\}}} + H_{q_{\{i,j,m\}}} + H_{q_{\{j,k,m\}}} + H_{q_{\{i,k,m\}}} + H_{|i-j|}H_{|k-m|} + H_{|k-i|}h_{|m-j|}\\
&\quad + H_{|k-j|}H_{|m-i|},\\
a_1(0) &= H_{q_{\{i,j,k,m\}}}.
\end{aligned}
$$

This completes the proof. □

7.3. Proof of Theorems 1 and 2

Throughout the proof $C_i, c_i, i = 1, 2, \dots,$ stand for constants depending on ℓ only, unless it is mentioned explicitly. The proofs of both theorems are almost identical. We first prove Theorem 1 and then indicate modifications needed for the proof of Theorem 2.

It follows from (7) and (12) that

$$
\begin{aligned}
\hat{G}_h(x_0) - G(x_0) &= \frac{1}{\lambda}\Big[\sum_{k\in M_{D_{x_0}}} a_k(x_0)\hat{R}_k - R'(x_0)\Big]\\
&= \frac{1}{\lambda}\Big[\sum_{k\in M_{D_{x_0}}} a_k(x_0)(\hat{R}_k - R_k) + \sum_{k\in M_{D_{x_0}}} a_k(x_0)R_k - R'(x_0)\Big].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&[\mathbb{E}_G|\hat{G}_h(x_0) - G(x_0)|^2]^{1/2}\\
&\quad \le \frac{1}{\lambda}\bigg\{\mathbb{E}_G\Big[\sum_{k\in M_{D_{x_0}}} a_k(x_0)(\hat{R}_k - R_k)\Big]^2\bigg\}^{1/2} + \frac{1}{\lambda}\Big|\sum_{k\in M_{D_{x_0}}} a_k(x_0)R_k - R'(x_0)\Big|.
\end{aligned} \tag{29}
$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

In the subsequent proof, we bound the expression on the right-hand side of (29). The result of the theorem will follow from the series of lemmas given below.

We begin with a well-known result on the properties of the local polynomial estimators; see, e.g. Nemirovski (2000, Lemma 1.3.1) and Tsybakov (2009, Section 1.6).

Lemma 1. *Let* $\{a_k(x_0),\ k \in M_{D_{x_0}}\}$ *be the solution to (10), and let (9) hold; then*

$$\Big[\sum_{k\in M_{D_{x_0}}} |a_k(x_0)|^2\Big]^{1/2} \le \frac{C_1}{h\sqrt{N_{D_{x_0}}}}, \qquad \sum_{k\in M_{D_{x_0}}} |a_k(x_0)| \le \frac{C_2}{h}, \tag{30}$$

where $C_1 = C_1(\ell)$ *and* $C_2 = C_2(\ell)$ *are constants depending on* ℓ *only.*

The next result establishes an upper bound on the accuracy of the empirical covariance estimator.

Lemma 2. *For any* $k = 0, \dots, n-1$*, we have*

$$\mathbb{E}_G|\hat{R}_k - R_k|^2 \le \frac{C_3}{n-k}(\rho^2+\rho)\sum_{i=1}^{n} H_i,$$

where C_3 *is an absolute constant.*

Proof. Let $\tilde{R}_k := (1/(n-k))\sum_{i=1}^{n-k}(X_i-\rho)(X_{i+k}-\rho)$; then $\mathbb{E}_G\tilde{R}_k = R_k$, and, by the definition of $\hat{\rho}_k$,

$$\hat{R}_k = \frac{1}{n-k}\sum_{i=1}^{n-k}(X_i-\hat{\rho}_k)(X_{i+k}-\hat{\rho}_k) = \tilde{R}_k - (\hat{\rho}_k-\rho)^2.$$

Therefore,

$$\begin{aligned}\mathbb{E}_G|\hat{R}_k - R_k|^2 &= \mathbb{E}_G|\tilde{R}_k - R_k|^2 - 2\mathbb{E}_G[\tilde{R}_k(\hat{\rho}_k-\rho)^2] + \mathbb{E}_G|\hat{\rho}_k-\rho|^4 \\ &=: J_1 - 2J_2 + J_3.\end{aligned} \tag{31}$$

Now we proceed with the computation of the terms on the right-hand side of (31).

Computation of J_1. Let $r_k := \mathbb{E}_G[X_iX_{i+k}] = R_k + \rho^2 = \rho H_k + \rho^2$ and $\tilde{r}_k := (1/(n-k))\sum_{i=1}^{n-k} X_iX_{i+k}$; then

$$\tilde{R}_k - R_k = \tilde{r}_k - r_k + 2\rho^2 - \frac{\rho}{n-k}\sum_{i=1}^{n-k}(X_i+X_{i+k}).$$

Thus,

$$\begin{aligned}J_1 &= \mathbb{E}_G|\tilde{R}_k - R_k|^2 \\ &= \mathbb{E}_G|\tilde{r}_k - r_k|^2 - 2\mathbb{E}_G\Big[(\tilde{r}_k - r_k)\frac{\rho}{n-k}\sum_{i=1}^{n-k}(X_i+X_{i+k})\Big] \\ &\quad + \mathbb{E}_G\Big[2\rho^2 - \frac{\rho}{n-k}\sum_{t=1}^{n-k}(X_i+X_{i+k})\Big]^2 \\ &=: J_1^{(1)} - J_1^{(2)} + J_1^{(3)}.\end{aligned} \tag{32}$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

Equation (26) implies that, for any $k = 0, \dots, n$ and $i, j = 1, \dots, n-k$, we have

$$\begin{aligned}
&\mathbb{E}_G[X_i X_{i+k} X_j X_{j+k}] \\
&\quad = \rho^4 + \rho^3[H_k + H_{|j-i|} + H_{|j-i+k|} + H_{|j-i-k|} + H_{|j-i|} + H_k] \\
&\qquad + \rho^2[2H_{k\vee|j-i|\vee|j-i-k|} + 2H_{k\vee|j-i|\vee|j-i+k|} + H_k^2 + H_{|j-i|}^2 + H_{|j-i+k|}H_{|j-i-k|}] \\
&\qquad + \rho H_{k\vee|j-i|\vee|j-i+k|\vee|j-i-k|}.
\end{aligned}$$

Since $r_k^2 = \rho^4 + 2\rho^3 H_k + \rho^2 H_k^2$, we have

$$\begin{aligned}
J_1^{(1)} = \mathbb{E}_G|\tilde{r}_k - r_k|^2 &= \frac{1}{(n-k)^2}\sum_{i,j=1}^{n-k}\mathbb{E}_G[X_i X_{i+k} X_j X_{j+k}] - r_k^2 \\
&= \frac{1}{(n-k)^2}\sum_{i,j=1}^{n-k}\{\rho^3[2H_{|j-i|} + H_{|j-i-k|} + H_{|j-i+k|}] + \rho H_{k\vee|j-i|\vee|j-i-k|\vee|j-i+k|} \\
&\qquad\qquad + \rho^2[2H_{k\vee|j-i|\vee|j-i-k|} + 2H_{k\vee|j-i|\vee|j-i+k|} + H_{|j-i|}^2 \\
&\qquad\qquad\qquad + H_{|j-i-k|}H_{|j-i+k|}]\}. \qquad (33)
\end{aligned}$$

Furthermore,

$$\begin{aligned}
J_1^{(3)} &= 4\rho^2 - \frac{4\rho^3}{n-k}\mathbb{E}_G\sum_{i=1}^{n-k}(X_i + X_{i+k}) + \mathbb{E}_G\frac{\rho^2}{(n-k)^2}\sum_{i,j=1}^{n-k}(X_i + X_{i+k})(X_j + X_{j+k}) \\
&= -4\rho^4 + \frac{\rho^2}{(n-k)^2}\sum_{i,j=1}^{n-k}[2r_{|j-i|} + r_{|j-i+k|} + r_{|j-i-k|}] \\
&= \frac{\rho^3}{(n-k)^2}\sum_{i,j=1}^{n-k}[2H_{|j-i|} + H_{|j-i-k|} + H_{|j-i+k|}]. \qquad (34)
\end{aligned}$$

Now we proceed with $J_1^{(2)}$, i.e.

$$\begin{aligned}
J_1^{(2)} &= \frac{2\rho}{n-k}\sum_{i=1}^{n-k}\mathbb{E}_G(\tilde{r}_k - r_k)(X_i + X_{i+k}) \\
&= \frac{2\rho}{n-k}\sum_{i=1}^{n-k}[\mathbb{E}_G(\tilde{r}_k X_i + \tilde{r}_k X_{i+k}) - 2\rho(\rho^2 + \rho H_k)].
\end{aligned}$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

We have

$$\mathbb{E}_G[\tilde{r}_k X_i] = \frac{1}{n-k}\sum_{j=1}^{n-k}\mathbb{E}_G[X_j X_{j+k} X_i]$$
$$= \frac{1}{n-k}\sum_{j=1}^{n-k}[\rho^3 + \rho^2 H_k + \rho^2 H_{|j-i|} + \rho^2 H_{|j-i-k|} + \rho H_{k\vee|j-i|\vee|i-j-k|}],$$
$$\mathbb{E}_G[\tilde{r}_k X_{i+k}] = \frac{1}{n-k}\sum_{j=1}^{n-k}\mathbb{E}_G[X_j X_{j+k} X_{i+k}]$$
$$= \frac{1}{n-k}\sum_{j=1}^{n-k}[\rho^3 + \rho^2 H_k + \rho^2 H_{|j-i|} + \rho^2 H_{|j-i+k|} + \rho H_{k\vee|j-i|\vee|i-j+k|}],$$

which yields

$$J_1^{(2)} = \frac{2}{(n-k)^2}\sum_{i,j=1}^{n-k}[2\rho^3 H_{|j-i|} + \rho^3 H_{|i-j+k|} + \rho^3 H_{|i-j-k|} + \rho^2 H_{|i-j|\vee k\vee|i-j-k|} + \rho^2 H_{|i-j|\vee k\vee|i-j+k|}]. \tag{35}$$

Combining (32)–(35), we obtain

$$J_1 = \frac{\rho^2}{(n-k)^2}\sum_{i,j=1}^{n-k}[H_{|i-j|}^2 + H_{|i-j+k|}H_{|i-j-k|}] + \frac{\rho}{(n-k)^2}\sum_{i,j=1}^{n-k} H_{k\vee|i-j|\vee|i-j-k|\vee|i-j+k|}.$$

Taking into account that H is a monotone decreasing function, and $H(0) = 1$, we obtain

$$J_1 \le \frac{c_1}{n-k}(\rho^2 + \rho)\sum_{i=1}^{n} H_i,$$

where c_1 is an absolute constant.

Computation of J_2. It follows from the definition of J_2 that

$$J_2 = \rho^3 H_k - 2\rho\mathbb{E}_G[\tilde{R}_k\hat{\rho}_k] + \mathbb{E}_G[\tilde{R}_k\hat{\rho}_k^2].$$

We have

$$E_G[\tilde{R}_k\hat{\rho}_k] = \frac{1}{(n-k)^2}\mathbb{E}_G\sum_{i,j=1}^{n-k}(X_i-\rho)(X_{i+k}-\rho)X_j$$
$$= \mathbb{E}_G\frac{1}{(n-k)^2}\sum_{i,j=1}^{n-k}[X_i X_{i+k} X_j - \rho X_{i+k} X_j - \rho X_i X_j + \rho^2 X_j]$$
$$= \rho^2 H_k + \frac{\rho}{(n-k)^2}\sum_{i,j=1}^{n-k} H_{k\vee|i-j|\vee|i-j+k|}\,.$$

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

Furthermore,

$$
\begin{aligned}
\mathbb{E}_G[\tilde{R}_k\hat{\rho}_k^2] &= \frac{1}{(n-k)^3}\sum_{i,j,l=1}^{n-k}\mathbb{E}_G[X_iX_{i+k}X_jX_l - \rho X_{i+k}X_jX_l - \rho X_iX_jX_l + \rho^2X_jX_l] \\
&= \frac{\rho^2}{(n-k)^3}\sum_{i,j,l=1}^{n-k}[H_{k\vee|i-j|\vee|i-j+k|} + H_{|i-l|\vee|i-l+k|\vee k} + H_kH_{|j-l|} \\
&\qquad + H_{|i-j|}H_{|i-l+k|} + H_{|i-l|}H_{|i-j+k|}] \\
&\quad + \rho^3H_k + \frac{\rho}{(n-k)^3}\sum_{i,j,l=1}^{n-k}H_{k\vee|i-j|\vee|i-l|\vee|i+k-j|\vee|i+k-l|\vee|j-l|} \\
&= \rho^3H_k + \frac{\rho^2}{(n-k)^2}\sum_{i,j=1}^{n-k}2H_{k\vee|i-j|\vee|i-j+k|} \\
&\quad + \frac{1}{(n-k)^3}\sum_{i,j,l=1}^{n-k}[\rho^2(H_kH_{|j-l|} + H_{|i-j|}H_{|i-l+k|} + H_{|i-l|}H_{|i-j+k|}) \\
&\qquad + \rho H_{k\vee|i-j|\vee|i-l|\vee|i+k-j|\vee|i+k-l|\vee|j-l|}].
\end{aligned}
$$

Combining these equalities, we obtain

$$
\begin{aligned}
J_2 &= \frac{1}{(n-k)^3}\sum_{i,j,l=1}^{n-k}[\rho^2(H_kH_{|j-l|} + H_{|i-j|}H_{|i-l+k|} + H_{|i-l|}H_{|i-j+k|}) \\
&\qquad + \rho H_{k\vee|i-j|\vee|i-l|\vee|i+k-j|\vee|i+k-l|\vee|j-l|}] \\
&\le \frac{c_2}{n-k}(\rho^2+\rho)\sum_{i=1}^{n}H_i,
\end{aligned}
$$

where c_2 is an absolute constant.

Computation of J_3. By definition, $J_3 = \mathbb{E}_G|(1/(n-k))\sum_{i=1}^{n-k}(X_i-\rho)|^4$. Using Proposition 3 and after routine calculation, we obtain, for all $i, j, l, m = 1, \dots, n-k$,

$$
\begin{aligned}
&\mathbb{E}_G[(X_i-\rho)(X_j-\rho)(X_l-\rho)(X_m-\rho)] \\
&\quad = \rho^2[H_{|i-j|}H_{|l-m|} + H_{|i-l|}H_{|j-m|} + H_{|l-j|}H_{|i-m|}] \\
&\qquad + \rho H_{|i-j|\vee|j-l|\vee|l-m|\vee|l-m|\vee|i-m|\vee|i-l|},
\end{aligned}
$$

so that

$$
\begin{aligned}
J_3 &= \frac{\rho^2}{(n-k)^4}\sum_{i,j,l,m=1}^{n-k}[H_{|i-j|}H_{|l-m|} + H_{|i-l|}H_{|j-m|} + H_{|l-j|}H_{|i-m|}] \\
&\quad + \frac{\rho}{(n-k)^4}\sum_{i,j,l,m=1}^{n-k}H_{|i-j|\vee|j-l|\vee|l-m|\vee|l-m|\vee|i-m|\vee|i-l|} \\
&\le \frac{c_3}{n-k}(\rho^2+\rho)\sum_{i=1}^{n}H_i,
\end{aligned}
$$

where c_3 is an absolute constant.

https://doi.org/10.1017/apr.2016.67 Published online by Cambridge University Press

Combining inequalities for J_1, J_2, and J_3 with (31), we complete the proof. □

Lemma 3. *For every $x_0 \in [0, T-\delta]$, we have*

$$\mathbb{E}_G\Bigg|\sum_{k\in M_{D_{x_0}}} a_k(x_0)(\hat{R}_k - R_k)\Bigg|^2 \le \frac{C_4\delta}{h^2\psi_{x_0}(T)}(\rho^2+\rho)\sum_{i=1}^n H_i,$$

where $C_4 = C_4(\ell)$ is a constant depending on ℓ only, and

$$\psi_{x_0}(T) = \psi_{x_0}(T,h,\delta) := \begin{cases} T - x_0 - h, & h \le x_0 < T-\delta-h, \\ T-2h, & 0 \le x_0 \le h, \\ \delta, & T-\delta-h \le x_0 \le T-\delta. \end{cases}$$

Proof. By Lemmas 1 and 2 and by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}\Bigg|\sum_{k\in M_{D_{x_0}}} a_k(x_0)(\hat{R}_k - R_k)\Bigg|^2 &\le \sum_{k\in M_{D_{x_0}}} a_k^2(x_0)\sum_{k\in M_{D_{x_0}}}\mathbb{E}_G(\hat{R}_k - R_k)^2 \\ &\le \frac{c_1^2}{h^2 N_{D_{x_0}}}(\rho^2+\rho)\bigg(\sum_{i=1}^n H_i\bigg)\sum_{k\in M_{D_{x_0}}}\frac{1}{n-k}. \end{aligned} \tag{36}$$

Let $\underline{k} = \min\{k \in (1,\dots,n-1) \colon k \in M_{D_{x_0}}\}$ and $\bar{k} = \max\{k \in (1,\dots,n-1) \colon k \in M_{D_{x_0}}\}$; then

$$\sum_{k\in M_{D_{x_0}}}\frac{1}{n-k} = \sum_{k=\underline{k}}^{\bar{k}}\frac{1}{n-k} \le \ln\bigg(\frac{n-\underline{k}}{n-\bar{k}}\bigg) = \ln\bigg(1+\frac{\bar{k}-\underline{k}}{n-\bar{k}}\bigg) \le \frac{\bar{k}-\underline{k}}{n-\bar{k}}.$$

First, assume that $D_{x_0} = [x_0-h, x_0+h]$. In this case, $\underline{k} = [(x_0-h)/\delta]+1$ and $\bar{k} = [(x_0+h)/\delta]$, where $[\cdot]$ is the integer part, and then $\sum_{k\in M_{D_{x_0}}} 1/(n-k) \le 2h/(T-x_0-h)$. If $D_{x_0} = [0, 2h]$ then $\underline{k} = 1$ and $\bar{k} = [2h/\delta]$, which leads to $\sum_{k\in M_{D_{x_0}}} 1/(n-k) \le 2h/(T-2h)$. Finally, if $D_{x_0} = [T-2h-\delta, T-\delta]$ then $\bar{k} = n-1$, $\underline{k} = (n-1)-[2h/\delta]$, and $\sum_{k\in M_{D_{x_0}}} 1/(n-k) \le 2h/\delta$.

Combining these bounds with (36) and taking into account that $(2h/\delta)-1 \le N_{D_{x_0}} \le (2h/\delta)+1$, we complete the proof. □

Lemma 4. *Let $G \in \mathscr{H}_\beta(L, I)$, $I = [x_0-d, x_0+d] \supseteq D_{x_0}$, and $\{a_k(x_0),\ k \in M_{D_{x_0}}\}$ be the weights defined by (10) with $\ell \ge \lfloor\beta\rfloor + 1$. Assume that (9) holds; then*

$$\Bigg|\sum_{k\in M_{D_{x_0}}} a_k(x_0)R_k - R'(x_0)\Bigg| \le C_2\lambda L h^\beta,$$

where $C_2 = C_2(\ell)$ is the constant appearing in (30).

Proof. Recall that $R(t) = \rho^2 + \rho h(t) = \rho^2 + \lambda\int_t^\infty [1-G(x)]\,\mathrm{d}x$; this implies that

$$R'(t) = -\lambda(1-G(t)), \qquad R^{(j)}(t) = \lambda G^{(j-1)}(t) \quad \text{for all } j = 2,\dots,\lfloor\beta\rfloor+1.$$

Thus, if $G \in \mathcal{H}_\beta(L, I)$ then $R \in \mathcal{H}_{\beta+1}(\lambda L, I)$. Since $D_{x_0} \subseteq I$, function R can be expanded in the Taylor series around x_0. In particular, for any $k \in M_{D_{x_0}}$,

$$R(k\delta) = R(x_0) + \sum_{j=1}^{\lfloor\beta\rfloor} \frac{1}{j!} R^{(j)}(x_0)(k\delta - x_0)^j + \frac{1}{(\lfloor\beta\rfloor + 1)!} R^{(\lfloor\beta\rfloor+1)}(\xi_k)(k\delta - x_0)^{\lfloor\beta\rfloor+1}, \quad (37)$$

where $\xi_k = \tau k\delta + (1 - \tau)x_0$ for some $\tau \in [0, 1]$. Denote

$$\bar{R}_{x_0}(y) := R(x_0) + \sum_{j=1}^{\lfloor\beta\rfloor+1} \frac{1}{j!} R^{(j)}(x_0)(y - x_0)^j, \qquad y \in D_{x_0}. \quad (38)$$

Since $\bar{R}_{x_0}(\cdot)$ is a polynomial of degree $\lfloor\beta\rfloor + 1$ and $\ell \geq \lfloor\beta\rfloor + 1$, we have, by (11),

$$\sum_{k \in M_{D_{x_0}}} a_k(x_0)\bar{R}_{x_0}(k\delta) = \bar{R}'_{x_0}(x_0) = R'(x_0).$$

Therefore,

$$\begin{aligned}
&\sum_{k \in M_{D_{x_0}}} a_k(x_0)R(k\delta) - R'(x_0) \\
&\qquad = \sum_{k \in M_{D_{x_0}}} a_k(x_0)[R(k\delta) - \bar{R}_{x_0}(k\delta)] \\
&\qquad = \sum_{k \in M_{D_{x_0}}} \frac{1}{(\lfloor\beta\rfloor + 1)!} a_k(x_0)[R^{(\lfloor\beta\rfloor+1)}(\xi_k) - R^{(\lfloor\beta\rfloor+1)}(x_0)](k\delta - x_0)^{\lfloor\beta\rfloor+1},
\end{aligned}$$

where we have used (37) and (38). This yields

$$\left| \sum_{k \in M_{D_{x_0}}} a_k(x_0)R(k\delta) - R'(x_0) \right| \leq \frac{\lambda L h^{\beta+1}}{(\lfloor\beta\rfloor + 1)!} \sum_{k \in M_{D_{x_0}}} |a_k(x_0)| \leq C_2 \lambda L h^\beta,$$

where the last inequality follows from (30). □

Now we complete the proof of Theorem 1. Since $G \in \mathcal{M}_2(K)$,

$$\begin{aligned}
\sum_{i=1}^n H_i &= \sum_{i=1}^n H(i\delta) \\
&\leq \frac{1}{\delta} \int_0^T H(t)\,\mathrm{d}t \\
&= \frac{\mu}{\delta} \int_0^T \int_t^\infty [1 - G(x)]\,\mathrm{d}x\,\mathrm{d}t \\
&\leq \frac{\mu}{\delta} \int_0^\infty x[1 - G(x)]\,\mathrm{d}x \\
&\leq \frac{\mu}{2\delta} K.
\end{aligned} \quad (39)$$

Moreover, $G \in \mathcal{M}_2(K)$ also implies that $1/\mu \leq \sqrt{K}$.

It can be easily verified that, under (14) and (13) for all large enough T, we have $T - x_0 \geq \kappa T$, and D_{x_0} contains at least $\ell + 1$ grid points. Therefore, by Lemmas 3 and 4 and (39), the chosen window width $h = h_*$ balances the upper bounds on the two terms on the right-hand side of (29). The result of Theorem 1 follows immediately by substitution of h_* in the bounds of Lemmas 3 and 4.

In order to prove Theorem 2, we note that the bias-variance decomposition in the problem of estimating λ takes the form

$$[\mathbb{E}_{G,\lambda}|\hat{\lambda} - \lambda|^2]^{1/2} \leq \bigg\{\mathbb{E}_{G,\lambda}\bigg[\sum_{k \in M_{D_{x_0}}} a_k(x_0)(\hat{R}_k - R_k)\bigg]^2\bigg\}^{1/2} + \bigg|\sum_{k \in M_{D_{x_0}}} a_k(x_0) R_k - R'(x_0)\bigg|;$$

see (29). The same upper bounds on the bias (Lemma 4) and the variance (Lemma 3) hold. The upper bound (20) follows by the special choice of the window width in (18). □

Acknowledgements

The author is grateful to Gideon Weiss for attracting his attention to the problem studied in this paper, and to Oleg Lepski for useful discussions and suggestions. Part of this work was carried out while the author was visiting NYU Shanghai. The work was supported by the United States Israel Binational Science Foundation (grant number 2010466) and the Israel Science Foundation (grant number 361/15).

References

Beneš, V. E. (1957). Fluctuations of telephone traffic. *Bell System Tech. J.* **36,** 965–973.

Bingham, N. H. and Dunham, B. (1997). Estimating diffusion coefficients from count data: Einstein–Smoluchowski theory revisited. *Ann. Inst. Statist. Math.* **49,** 667–679.

Bingham, N. H. and Pitts, S. M. (1999). Non-parametric estimation for the M/G/∞ queue. *Ann. Inst. Statist. Math.* **51,** 71–97.

Blanghaps, N., Nov, Y. and Weiss, G. (2013). Sojourn time estimation in an M/G/∞ queue with partial information. *J. Appl. Prob.* **50,** 1044–1056.

Borovkov, A. A. (1967). Limit laws for queueing processes in multichannel systems. *Siberian Math. J.* **8,** 746–763.

Brillinger, D. R. (1974). Cross-spectral analysis of processes with stationary increments including the stationary $G/G/\infty$ queue. *Ann. Prob.* **2,** 815–827.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd edn. Springer, New York.

Brown, M. (1970). An $M/G/\infty$ estimation problem. *Ann. Math. Statist.* **41,** 651–654.

Goldenshluger, A. (2015). Nonparametric estimation of service time distribution in the $M/G/\infty$ queue and related estimation problems. Preprint. Available at https://arxiv.org/abs/1508.00076.

Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Math. Meth. Statist.* **6,** 135–170.

Grübel, R. and Wegener, H. (2011). Matchmaking and testing for exponentiality in the M/G/∞ queue. *J. Appl. Prob.* **48,** 131–144.

Hall, P. and Park, J. (2004). Nonparametric inference about service time distribution from indirect measurements. *J. R. Statist. Soc. B* **66,** 861–875.

Iglehart, D. L. (1973). Weak convergence in queueing theory. *Adv. Appl. Prob.* **5,** 570–594.

Kingman, J. F. C. (1993). *Poisson Processes.* Oxford University Press.

Lindley, D. V. (1956). The estimation of velocity distributions from counts. In *Proceedings of the International Congress of Mathematicians (Amsterdam, 1954)*, Vol. III, Noordhoff, Groningen, pp. 427–444.

Milne, R. K. (1970). Identifiability for random translations of Poisson processes. *Z. Wahrscheinlichkeitsth.* **15,** 195–201.

Moulines, E., Roueff, F., Souloumiac, A. and Trigano, T. (2007). Nonparametric inference of photon energy distribution from indirect measurement. *Bernoulli* **13,** 365–588.

NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (Saint-Flour, 1998; Lecture Notes Math. **1738**), Springer, Berlin, pp. 85–277.

PARZEN, E. (1962). *Stochastic Processes.* Holden-Day, San Francisco, CA.

PICKANDS, J., III AND STINE, R. A. (1997). Estimation for an M/G/∞ queue with incomplete information. *Biometrika* **84,** 295–308.

REYNOLDS, J. F. (1975). The covariance structure of queues and related processes—a survey of recent work. *Adv. Appl. Prob.* **7,** 383–415.

ROSS, S. M. (1970). *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco, CA.

SCHWEER, S. AND WICHELHAUS, C. (2015). Nonparametric estimation of the service time distribution in the discrete-time $GI/G/\infty$ queue with partial information. *Stoch. Process. Appl.* **125,** 233–253.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York.

WHITT, W. (1985). The renewal-process stationary-excess operator. *J. Appl. Prob.* **22,** 156–167.

WHITT, W. (1974). Heavy traffic limit theorems for queues: a survey. In *Mathematical Methods in Queueing Theory* (Proc. Conf., Western Michigan Univ., Kalamazoo, MI, 1973; Lecture Notes Econom. Math. Systems **98**), Springer, Berlin, pp. 307–350.