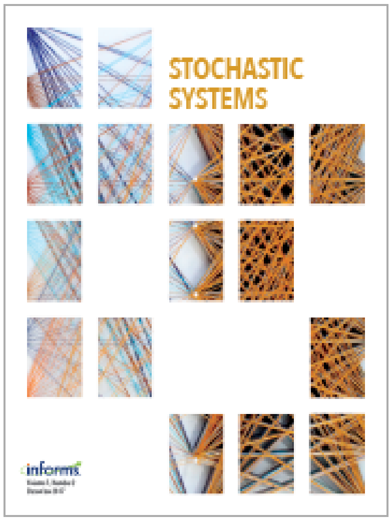


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Nonparametric Estimation of Service Time Characteristics in Infinite-Server Queues with Nonstationary Poisson Input

Alexander Goldenshluger, David T. Koops

To cite this article:
Alexander Goldenshluger, David T. Koops (2019) Nonparametric Estimation of Service Time Characteristics in Infinite-Server Queues with Nonstationary Poisson Input. Stochastic Systems 9(3):183-207. https://doi.org/10.1287/stsy.2018.0026

informs
http://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 9, No. 3, September 2019, pp. 183–207
ISSN 1946-5238 (online)

Nonparametric Estimation of Service Time Characteristics in Infinite-Server Queues with Nonstationary Poisson Input

Alexander Goldenshluger,[a] **David T. Koops**[b]

[a] Department of Statistics, University of Haifa, 31905 Haifa, Israel; [b] Korteweg–De Vries Institute for Mathematics, University of Amsterdam, 1098 XG Amsterdam, Netherlands

Contact: goldensh@stat.haifa.ac.il (AG); d.t.koops@uva.nl, http://orcid.org/0000-0002-3691-4600 (DTK)

Received: May 25, 2018
Revised: September 14, 2018
Accepted: November 14, 2018
Published Online in Articles in Advance: August 5, 2019

https://doi.org/10.1287/stsy.2018.0026

Abstract. This paper provides a mathematical framework for estimation of the service time distribution and expected service time of an infinite-server queueing system with a nonhomogeneous Poisson arrival process in the case of partial information, where only the numbers of busy servers are observed over time. The problem is reduced to a statistical deconvolution problem, which is solved by using Laplace transform techniques and kernels for regularization. Upper bounds on the mean squared error of the proposed estimators are derived. Some concrete simulation experiments are performed to illustrate how the method can be applied and provide some insight in the practical performance.

History: Former designation of this paper was SSy-2018-018.R1.

Funding: This research is supported by the Israel Science Foundation [Grant 361/15] and by the Netherlands Organisation for Scientific Research Gravitation Project NETWORKS [Grant 024.002.003].

Keywords: $M_t/G/\infty$ queue • nonparametric estimation • deconvolution • minimax risk • rate of convergence • upper bound

1. Introduction

The advance of more and larger data sets leads to new questions in operations research and statistics. This paper can be placed at the intersection of these two fields. In particular, we study the statistical problems of estimating the distribution function and expectation of service times in an infinite-server queueing model in the case of partial information. The information is incomplete in the sense that the number of busy servers is observed, but the individual customers cannot be tracked.

1.1. Infinite-Server Queue

We first provide some background on the $M/G/\infty$ queueing model, which is well studied and could be considered as a standard model. In such queues, there are arrivals according to a homogeneous Poisson process, each customer is served independent of all other customers, and customers do not have to queue for service because there is an infinite number of servers. The model has a wide variety of applications in, for example, telecommunications, road traffic, and hospital modeling. It can be used as an approximation to $M/G/n$ systems, where n is relatively large with respect to the arrival rate, but the model is also interesting in its own right. For example, it can be interpreted outside queueing theory as a model for the size of a population. In this paper, we will use queueing terminology (servers, customers, etc.), but these terms could be adjusted according to the application. For example, "customers" could be cars traveling between two locations, and their "service time" could be the travel time. In many applications, it is seen that the arrival rate is not constant, but it varies over time. We will provide some examples below. This observation motivates studying the $M_t/G/\infty$ queue, where the arrival process is assumed to be a *nonhomogeneous* Poisson process. This model is still particularly tractable (compare with Eick et al. 1993) and amenable for statistical analysis as shown in this paper.

1.2. Statistical Queueing Problems

Queueing theory studies probabilistic properties of random processes in service systems on the basis of a complete specification of system parameters. In this paper, we are interested in inverse problems when unknown characteristics of a system should be inferred from observations of the associated random processes. Typically, such observations are incomplete in the sense that individual customers cannot be tracked as they go through the service system. The importance of such statistical inverse problems with incomplete observations was emphasized in Baccelli et al. (2009).

The service time distribution and its expected value are important performance metrics of the infinite-server queue. Our goal is to estimate these characteristics of the $M_t/G/\infty$ system from observations of the queue-length process. In case of complete data, where it can be seen when each customer arrives and leaves, the above estimation problems are trivial. The difficulties of the incomplete data problems are in the fact that only macro-level data are observed (i.e., only the numbers of busy servers are recorded).

1.3. Applications

A nonexhaustive list of example applications of our analysis is given below.

- Traffic: Toll plazas or sensors often count the number of cars entering and leaving a certain highway segment. Cars can overtake each other and therefore, do not leave in the same order as they arrive. The problem is to determine the distribution of the speed of the cars over that segment. A solution that would make the problem trivial is to identify the car by recording its license plate. This comes with several downsides. For example, it could be costly to implement, and there could be privacy concerns. There have been several attempts in the literature to solve this problem in another way, usually by trying to match vehicles at the upstream and downstream points based on nonunique signatures (e.g., vehicle length), which can be obtained by a particular type of dual-loop detectors (compare with, e.g., Coifman and Cassidy 2002). Our approach applies to the harder case where single-loop detectors are used (i.e., where only the vehicle counts at the upstream and downstream points are known).
- Communication systems: If two internet routers only track the timestamp of when a packet arrived, then what is the distribution of the packet flow duration between those routers? The importance of such statistical inverse problems was emphasized by Baccelli et al. (2009). In particular, the recent paper by Antunes et al. (2017) proposes a sampling framework for measuring internet traffic based on the $M/G/\infty$ model.
- Biology: As noted in Jansen et al. (2016), the production of molecules may be "bursty" in the sense that periods of high production activity can be followed by periods of low activity. In Dobrzyński and Bruggeman (2009), this is modeled using an interrupted Poisson process (i.e., a Poisson process that is modulated by a stochastic on/off switch). The number of active molecules can then be modeled as the number of jobs in a modulated infinite-server queue. By expectation maximization and maximum likelihood techniques, it is possible to filter the most likely on/off arrival rate sample path (compare with Krishnamurthy and Elliot 1997). Subsequently, using such filtered paths, our methods can provide estimates of the lifetime distribution of molecules when the molecules cannot be tracked separately but only the aggregate amount.

In all of these examples, it would be unreasonable to assume that the arrival rate is constant. It is known that traffic is busier during rush hours, that production of molecules is bursty, and that internet activity is higher during the day than at night. Thus, inhomogeneity of the Poisson process is an important feature in a variety of applications.

1.4. Our Contributions

In this paper, we deal with the statistical inversion problem of estimating the service time distribution G and the expected service time μ_G in the $M_t/G/\infty$ queue using only observations of the queue-length process. We approach the problem as a statistical deconvolution problem and develop nonparametric kernel estimators based on Laplace transform techniques. Their accuracy is analyzed over suitable nonparametric classes of service time distributions. In particular, we derive upper bounds on the worst case root mean squared error (RMSE) of the proposed estimators and show how properties of the arrival rate function $\{\lambda(t), t \geq 0\}$ and service time distribution G affect the estimation accuracy. Furthermore, we provide details on the implementation of the estimators. For example, we describe an adaptive estimation procedure for the distribution function and confirm its efficiency by a simulation study.

Our results are based on a formula for the joint moment-generating function of the queue-length process at different time instances. We provide a derivation of this result, which to the best of our knowledge has not appeared in the literature before and is of independent interest.

1.5. Current Literature

Our contribution is related to two different strands of research. First, similar types of statistical inference questions in queueing theory were studied before but for the homogeneous case (i.e., the $M/G/\infty$ system; see, e.g., Brown 1970, Bingham and Pitts 1999, Schweer and Wichelhaus 2015, Goldenshluger 2016, 2018 and references therein). The analysis in the case of the $M_t/G/\infty$ system is vastly different. This is because of the nonstationary nature of the $M_t/G/\infty$ queue, whereas the analysis for the $M/G/\infty$ system relied on stationary measures. Second, similar deconvolution problems, such as density deconvolution, have been studied in statistical literature (see, e.g., Zhang 1990, Fan 1991). However, this strand of research typically considers models with independent observations and advocates the use of Fourier transform techniques. In contrast, our setting is completely different: the queue-length process in the $M_t/G/\infty$ model is intrinsically dependent, and Fourier-based techniques are not applicable because the arrival rate $\{\lambda(t), t \geq 0\}$ does not need to be (square) integrable. A connection can be drawn with Bigot et al. (2013), in which a deconvolution problem of estimating intensity of a nonhomogeneous Poisson process on a finite interval was considered. We also mention recent work (Abramovich et al. 2013) where Laplace transform techniques were applied for signal deconvolution in a Gaussian model.

1.6. Outline

The rest of this paper is organized as follows. In Section 2, we present the formula for the joint moment-generating function of the queue-length process at different time instances, from which known results in Eick et al. (1993) can be derived. In particular, the covariance of the queue length at different points in time can be found, which plays an important role in the subsequent statistical analysis. In Section 3, we formulate the estimation problems and introduce necessary notation and assumptions. Our main statistical results (Theorems 2, 3, and 4) that provide upper bounds on the risk of estimators of G and μ_G are given in Sections 4 and 5. Section 6 discusses numerical implementation of the proposed estimators and presents simulation results. Proofs of main results of this paper are given in Section 7.

2. Properties of the Queue-Length Process

In this section, we derive some probabilistic results on properties of the queue-length process of the $M_t/G/\infty$ queue; these results provide a basis for construction of our estimators. For the sake of convenience in derivations of this section, we assume that the $M_t/G/\infty$ queue operates an infinite amount of time (i.e., the arrival and queue-length processes are defined on the whole real line). Such a model was considered, for example, in Eick et al. (1993). The setting where the system starts empty at $t = 0$ is an easy specific case of the considered one.

Let $\{\tau_j, j \in \mathbb{Z}\}$ be arrival epochs constituting a realization of a nonhomogeneous Poisson process on the real line with intensity $\{\lambda(t), t \in \mathbb{R}\}$. 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 operates for infinite time; then the queue-length process $\{X(t), t \in \mathbb{R}\}$ is given by

$$X(t) = \sum_{j \in \mathbb{Z}} \mathbf{1}(\tau_j \leq t, \sigma_j > t - \tau_j), \quad t \in \mathbb{R}.$$

The next result provides formulas for the Laplace transform of finite-dimensional distributions of $\{X(t), t \in \mathbb{R}\}$.

Theorem 1. *Let $\bar{G}(t) := 1 - G(t)$, let*

$$H(t) := \int_0^\infty \bar{G}(u)\lambda(t-u)\mathrm{d}u < \infty, \quad \forall t \in \mathbb{R}, \tag{1}$$

and for $0 \leq a < b \leq \infty$, let

$$H_{a,b}(t) := \int_a^b \bar{G}(u)\lambda(t-u)\mathrm{d}u.$$

Let $t_1 < \cdots < t_m$ be fixed points, and let $t_0 = -\infty$ by convention; then, for any $\theta = (\theta_1, \ldots, \theta_m)$, $m \geq 1$, one has

$$\ln \mathbb{E}_G\left[\exp\left\{\sum_{i=1}^m \theta_i X(t_i)\right\}\right] = \sum_{k=1}^m (e^{\theta_k} - 1)H(t_k) + \sum_{k=1}^{m-1}(e^{\theta_{k+1}} - 1)\sum_{j=0}^{k-1}\left(e^{\sum_{i=j+1}^k \theta_i} - 1\right)H_{t_{k+1}-t_{j+1}, t_{k+1}-t_j}(t_{k+1}), \tag{2}$$

where $\mathbb{E}_G$ stands for the expectation with respect to the probability measure $\mathbb{P}_G$ generated by the queue-length process $\{X(t), t \in \mathbb{R}\}$ when the service time distribution is G.

The following statement is an immediate consequence of the result of Theorem 1 for specific cases $m = 1$ and $m = 2$.

Corollary 1. *For any $t \in \mathbb{R}$,*

$$\ln \mathbb{E}_G\big[\exp\{\theta X(t)\}\big] = (e^{\theta} - 1)H(t)$$

(i.e., $X(t)$ is a Poisson random variable with parameter $H(t)$). Moreover, for any $t_2 > t_1$,

$$\ln \mathbb{E}_G\big[\exp\{\theta_1 X(t_1) + \theta_2 X(t_2)\}\big] = (e^{\theta_1} - 1)H(t_1) + (e^{\theta_2} - 1)H(t_2) + (e^{\theta_2} - 1)(e^{\theta_1} - 1)H_{t_2-t_1,\infty}(t_2),$$

and therefore,

$$\mathrm{cov}_G[X(t_1), X(t_2)] = H_{t_2-t_1,\infty}(t_2) = \int_{t_2-t_1}^{\infty} \bar{G}(u)\lambda(t_2 - u)\mathrm{d}u. \tag{3}$$

The results in Corollary 1 are well known; they are proved (e.g., in section 1 of Eick et al. 1993) using a completely different technique. We were not able, however, to locate Equation (2) in the existing literature.

3. Estimation Problems, Notation, and Assumptions

In this section, we formulate estimation problems, introduce necessary notation and assumptions, and present a general idea for the construction of estimators of linear functionals of service time distribution G.

3.1. Formulation of Estimation Problems

Consider an $M_t/G/\infty$ queueing system empty at $t = 0$, with Poisson arrivals of intensity $\{\lambda(t), t \geq 0\}$. Assume that n independent realizations $X_1(t), \ldots, X_n(t)$ of the queue-length process $\{X(t), t \geq 0\}$ are observed on the interval $[0, T]$. Using the observations $\mathscr{X}_{n,T} := \{X_k(t), 0 \leq t \leq T, k = 1, \ldots, n\}$, we want to estimate the service time distribution G and the expected service time $\mu_G := \int_0^\infty [1 - G(t)]\mathrm{d}t$. In what follows, we will be interested in estimating the value $G(x_0)$ of G at a single given point $x_0 > 0$.

It is worth noting that the observation scheme of this paper, where n independent copies of the queue-length process are given, is quite standard in statistics of nonstationary processes. We refer, for example, to Kutoyants (1998), where nonparametric estimation of the intensity of a nonhomogeneous Poisson process was considered. Moreover, in what follows, we impose conditions on the arrival intensity $\lambda(\cdot)$, which stipulate that $\lambda(\cdot)$ is scaled with an amplitude parameter $\lambda_0 > 0$. As we will show, the estimation accuracy improves as the product $\lambda_0 n$ grows so that good estimations can be obtained even for $n = 1$, provided that the arrival rate amplitude λ_0 is large enough.

By estimators $\tilde{G}(x_0) = \tilde{G}(x_0; \mathscr{X}_{n,T})$ and $\tilde{\mu}_G = \tilde{\mu}_G(\mathscr{X}_{n,T})$ of $G(x_0)$ and μ_G, respectively, we mean measurable functions of $\mathscr{X}_{n,T}$. Their accuracy will be measured by the worst case risk over a set $\mathscr{G}$ of distributions G. In particular, for a functional class $\mathscr{G}$, the risk of $\tilde{G}(x_0)$ is defined by

$$\mathscr{R}_{x_0}[\tilde{G}; \mathscr{G}] := \sup_{G \in \mathscr{G}} \big[\mathbb{E}_G|\tilde{G}(x_0) - G(x_0)|^2\big]^{1/2},$$

whereas the risk of $\tilde{\mu}_G$ is defined by

$$\mathscr{R}[\tilde{\mu}_G; \mathscr{G}] := \sup_{G \in \mathscr{G}} \big[\mathbb{E}_G|\tilde{\mu}_G - \mu_G|^2\big]^{1/2}.$$

Our goal is to construct estimators of $G(x_0)$ and μ_G with provable accuracy guarantees over natural classes $\mathscr{G}$ of service time distributions.

3.2. General Idea for Estimator Construction

It follows from Theorem 1 that expectation $H(t) = \mathbb{E}_G[X(t)]$ of the queue-length process $X(t)$ is related to the service time distribution G via the convolution equation (1):

$$H(t) = \mathbb{E}_G[X(t)] = \int_0^\infty \bar{G}(u)\lambda(t - u)\mathrm{d}u, \quad \bar{G}(t) = 1 - G(t).$$

Therefore, estimating a linear functional $\theta(G)$ of G, such as $G(x_0)$ or μ_G, from observation of $X(t)$ is a statistical inverse problem of the deconvolution type.

We will base construction of our estimators on the linear functional strategy that is frequently used for solving ill-posed inverse problems. The main idea of this strategy is to find a pair of kernels (say, $K(\cdot)$ and $L(\cdot)$) such that the following two conditions are fulfilled:

(i) The integral $\int K(t)\bar{G}(t)\mathrm{d}t$ approximates the value $\theta(G)$.
(ii) The kernel L is related to K via the equation $\int K(t)\bar{G}(t)\mathrm{d}t = \int L(t)H(t)\mathrm{d}t$.

In view of Corollary 1 and by condition (ii), the statistic $\int L(t)X(t)\mathrm{d}t$ is an unbiased estimator for the integral $\int K(t)\bar{G}(t)\mathrm{d}t$, which by condition (i) approximates the value $\theta(G)$. Thus, $\int L(t)X(t)\mathrm{d}t$ is a reasonable estimator of $\theta(G)$.

3.3. Notation and Assumptions

To construct estimators in our settings, we will use the Laplace transform techniques. For this purpose, we require the following notation.

3.3.1. Notation. For a generic locally integrable function ψ on $\mathbb{R}$, the bilateral Laplace transform is defined by

$$\widehat{\psi}(z) := \mathscr{L}[\psi; z] = \int_{-\infty}^{\infty} e^{-zt}\psi(t)\mathrm{d}t.$$

The region of convergence of the integral on the right-hand side is a vertical strip in the complex plane; it will be denoted by

$$\Sigma_\psi := \left\{z \in \mathbb{C} : \sigma_\psi^- < \mathrm{Re}(z) < \sigma_\psi^+\right\}$$

for some constants $-\infty \le \sigma_\psi^- < \sigma_\psi^+ \le \infty$. If ψ is supported on $[0,\infty)$, then $\widehat{\psi}(z) := \int_0^\infty e^{-zt}\psi(t)\mathrm{d}t$, and the corresponding region of convergence is a half-plane

$$\Sigma_\psi := \{z \in \mathbb{C} : \mathrm{Re}(z) > \sigma_\psi\}, \quad \sigma_\psi \in \mathbb{R}.$$

The inversion formula for the Laplace transform is

$$\psi(t) = \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty} \widehat{\psi}(z)e^{zt}\mathrm{d}z = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\psi}(s+i\omega)e^{(s+i\omega)t}\mathrm{d}\omega, \quad \sigma_\psi^- < s < \sigma_\psi^+.$$

3.3.2. Assumptions on the Arrival Rate. In what follows, we assume that the arrival intensity is scaled by a parameter $\lambda_0 > 0$ (i.e., it can be represented in the form $\lambda(t) = \lambda_0\ell(t)$, where $\{\ell(t), t \ge 0\}$ is a fixed baseline arrival intensity that determines the shape of function $\lambda(t)$). This representation is useful for the study of the heavy traffic limit $\lambda_0 \to \infty$. As we will show, accuracy of our estimators depends on the product $N := \lambda_0 n$ and is determined by the growth rate of $\lambda(t)$ at infinity and the rate of decay of the Laplace transform $\widehat{\lambda}(z)$ over vertical lines in the region of convergence Σ_λ. These properties of the arrival rate are quantified in the next two assumptions.

Assumption 1. *The intensity function $\{\lambda(t), t \ge 0\}$ is a nonnegative locally integrable function on $[0,\infty)$ with the abscissa of convergence of the Laplace transform $\sigma_\lambda \ge 0$.*

(a) *If $\sigma_\lambda > 0$, then*

$$\lambda(t) \le \lambda_0 e^{\sigma_\lambda t}, \quad t \ge 0. \tag{4}$$

(b) *If $\sigma_\lambda = 0$, then there exist real numbers $p \ge 0$, and $a_1 \ge 0$, $a_2 > 0$, $\max\{a_1, a_2\} = 1$ such that*

$$\lambda(t) \le \lambda_0(a_1 + a_2 t^p), \quad t \ge 0. \tag{5}$$

Assumption 2. *The Laplace transform $\widehat{\lambda}(z)$ does not have zeros in $\{z \in \mathbb{C} : \mathrm{Re}(z) > \sigma_\lambda\}$, and*

$$|\widehat{\lambda}(\sigma + i\omega)| \ \ge\ \frac{\lambda_0 k_0}{[(\sigma - \sigma_\lambda)^2 + \omega^2]^{\gamma/2}}, \qquad \omega \in \mathbb{R}, \ \ \sigma > \sigma_\lambda, \tag{6}$$

where $\gamma \ge 0$ and $k_0 > 0$.

Several remarks on these assumptions are in order.

Assumption 1 states growth conditions on λ: Assumption 1(a) allows exponential growth, whereas Assumption 1(b) is a polynomial growth condition. The important case of bounded λ is included in Assumption 1(b) and corresponds to $p = 0$. Here note that $a_2 > 0$, whereas $a_1 \ge 0$, and a convenient normalization $\max\{a_1, a_2\} = 1$ is used. The growth conditions (4) and (5) guarantee that $\widehat{\lambda}$ is absolutely convergent in the half-plane $\Sigma_\lambda = \{z \in \mathbb{C} : \mathrm{Re}(z) > \sigma_\lambda\}$ with $\sigma_\lambda > 0$ and $\sigma_\lambda = 0$, respectively.

Assumption 2 states that $\widehat{\lambda}$ does not have zeros in Σ_λ. By the Riemann–Lebesgue lemma, $\widehat{\lambda}$ decreases along vertical lines in Σ_λ (see, e.g., section 23 in Doetsch 1974), and Assumption 2 stipulates the decrease rate. Note that if $\gamma > 1$ in (6), then $\int_{-\infty}^{\infty} |\widehat{\lambda}(\sigma + i\omega)| d\omega < \infty$ for all $\sigma > \sigma_\lambda$, and $\lambda(t)$ can be inverted from $\widehat{\lambda}$ by integrating over any vertical line $\mathrm{Re}(z) = \sigma > \sigma_\lambda$. If $\gamma > 1/2$, then $\int_{-\infty}^{\infty} |\widehat{\lambda}(\sigma + i\omega)|^2 d\omega < \infty$ for any $\sigma > \sigma_\lambda$, and the Laplace inversion formula should be understood in the sense of the $\mathbb{L}_2$ convergence (compare with chapter 2, section 10 in Widder 1946). As mentioned earlier, parameter $\lambda_0 > 0$ in Assumptions 1 and 2 is a scaling parameter that plays the role of the amplitude of the arrival intensity $\{\lambda(t), t \geq 0\}$.

The next examples show that Assumptions 1 and 2 hold in many cases of interest.

Example 1 (Constant Arrival Rate). Let $a \geq 0$ and $\lambda(t) = \lambda_0 \mathbf{1}_{[a,\infty)}(t)$. It is obvious that Assumption 1 holds with $\sigma_\lambda = 0$ and $p = 0$. In addition, because

$$\widehat{\lambda}(\sigma + i\omega) = \frac{\lambda_0 e^{a(\sigma + i\omega)}}{\sigma + i\omega}, \quad \sigma > 0,$$

and $|\widehat{\lambda}(\sigma + i\omega)| = \lambda_0 e^{a\sigma}(\sigma^2 + \omega^2)^{-1/2}$, Assumption 2 holds with $k_0 = 1$ and $\gamma = 1$. Note that $\widehat{\lambda}(z)$ has a unique singularity point at $z = 0$.

Example 2 (Polynomial Arrival Rate). Let $\lambda(t) = \lambda_0 t^p \mathbf{1}_{[0,\infty)}(t)$, $p > 0$; here Assumption 1 holds with $\sigma_\lambda = 0$, $a_1 = 0$, and $a_2 = 1$. Moreover, $\widehat{\lambda}(\sigma + i\omega) = \lambda_0 \Gamma(p+1)/(\sigma + i\omega)^{p+1}$, $\sigma > 0$; hence,

$$|\widehat{\lambda}(\sigma + i\omega)| = \frac{\lambda_0 \Gamma(p+1)}{(\sigma^2 + \omega^2)^{(p+1)/2}}, \quad \sigma > 0.$$

Thus, Assumption 2 is valid with $\sigma_\lambda = 0$, $\gamma = p + 1$, and $k_0 = \Gamma(p+1)$. The function $\widehat{\lambda}(z)$ has a unique singularity point at $z = 0$.

Example 3 (Sinusoidal Arrival Rate). Let $\lambda(t) = \lambda_0[1 + b\sin(t)]$ for $t \geq 0$, $0 < b \leq 1$; then

$$\widehat{\lambda}(\sigma + i\omega) = \frac{\lambda_0}{\sigma + i\omega} + \frac{\lambda_0 b}{\sigma^2 + \omega^2 + 1}, \quad \sigma > 0.$$

Here

$$|\widehat{\lambda}(\sigma + i\omega)| = \frac{\lambda_0}{\sigma^2 + \omega^2}\left(\omega^2 + \frac{[\sigma(\sigma^2 + \omega^2 + 1) + b(\sigma^2 + \omega^2)]^2}{(\sigma^2 + \omega^2 + 1)^2}\right)^{1/2}.$$

It is evident that

$$\frac{\lambda_0}{\sqrt{\sigma^2 + \omega^2}} \leq |\widehat{\lambda}(\sigma + i\omega)| \leq \frac{\lambda_0}{\sqrt{\sigma^2 + \omega^2}}\left(1 + \frac{b}{\sqrt{\sigma^2 + \omega^2}}\right).$$

Thus, Assumption 2 holds with $\sigma_\lambda = 0$, $\gamma = 1$, and $k_0 = 1$. Three singularity points of $\widehat{\lambda}$ are located on the imaginary axis: $z = 0$, $z = \pm i$.

Example 4 (Exponential Arrival Rate). Let $\lambda(t) = \lambda_0 e^{\theta t}$, $t \geq 0$, for some $\theta > 0$; then $\widehat{\lambda}(\sigma + i\omega) = \lambda_0(\sigma + i\omega - \theta)^{-1}$, $\sigma > \sigma_\lambda := \theta$. Here

$$|\widehat{\lambda}(\sigma + i\omega)| = \frac{\lambda_0}{[(\sigma - \theta)^2 + \omega^2]^{1/2}}, \quad \sigma > \theta.$$

Thus, Assumption 2 is fulfilled with $\sigma_\lambda = \theta$, $\gamma = 1$, and $k_0 = 1$.

4. Estimation of the Service Time Distribution

In this section, we consider the problem of estimating the service time distribution.

4.1. Estimator Construction

The estimator construction follows the linear functional strategy described in Section 3.2. Let K be a fixed bounded function supported on $[0, \infty)$. For real number $h > 0$, we define

$$L_h(t) := \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty} \frac{\widehat{K}(zh)}{\widehat{\lambda}(-z)} e^{zt} dz = \frac{e^{st}}{2\pi}\int_{-\infty}^{\infty} \frac{\widehat{K}((s+i\omega)h)}{\widehat{\lambda}(-s-i\omega)} e^{i\omega t} d\omega, \qquad t \in \mathbb{R}. \tag{7}$$

Because the convergence region of $\widehat{\lambda}$ is $\Sigma_\lambda = \{z \in \mathbb{C} : \mathrm{Re}(z) > \sigma_\lambda\}$, $\widehat{\lambda}(-z)$ is well defined in $\{z \in \mathbb{C} : \mathrm{Re}(z) < -\sigma_\lambda\}$. The kernel K will always be chosen so that $\Sigma_K = \mathbb{C}$, and if Assumption 2 holds, then $\widehat{\lambda}(z)$ does not have zeros in Σ_λ. Then function $\widehat{K}(zh)/\widehat{\lambda}(-z)$ is analytic in $\Sigma_L := \{z \in \mathbb{C} : \mathrm{Re}(z) < -\sigma_\lambda\}$, and kernel L_h does not depend on real number s, provided that $s < -\sigma_\lambda$.

The next result reveals a relationship between kernels L_h and $K_h(\cdot) := (1/h)K(\cdot/h)$ and provides motivation for construction of our estimator.

Lemma 1. *Let $H(t) = \mathbb{E}_G[X(t)]$ as defined in (1), and assume that*

$$\int_{-\infty}^{\infty} \left| \frac{\widehat{K}((s+i\omega)h)}{\widehat{\lambda}(-s-i\omega)} \right| \mathrm{d}\omega < \infty, \quad \int_0^\infty |L_h(t-x_0)| H(t)\mathrm{d}t < \infty, \quad \forall x_0. \tag{8}$$

Then

$$\int_0^\infty L_h(t-x_0)H(t)\mathrm{d}t = \int_{-\infty}^{\infty} \frac{1}{h} K\Big(\frac{x-x_0}{h}\Big) \bar{G}(x)\mathrm{d}x, \quad \forall x_0 > 0. \tag{9}$$

Proof. The first condition in Equation (8) ensures that $L_h(t)$ is well defined and finite for each t. In view of the second condition, by Fubini's theorem,

$$\int_{-\infty}^{\infty} L_h(t-x_0)H(t)\mathrm{d}t = \int_0^\infty \bar{G}(x) \int_{-\infty}^{\infty} L_h(t-x_0)\lambda(t-x)\mathrm{d}t\mathrm{d}x.$$

Now we show that the definition (7) implies that function L_h solves the equation

$$\int_{-\infty}^{\infty} L_h(t-x_0)\lambda(t-x)\mathrm{d}t = \frac{1}{h}K\Big(\frac{x-x_0}{h}\Big), \quad \forall x. \tag{10}$$

Indeed, the bilateral Laplace transform of the left-hand side is

$$\int_{-\infty}^{\infty} e^{-zx}\Big[\int_{-\infty}^{\infty} L_h(t-x_0)\lambda(t-x)\mathrm{d}t\Big]\mathrm{d}x = \widehat{\lambda}(-z)\int_{-\infty}^{\infty} e^{-zt}L_h(t-x_0)\mathrm{d}t = \widehat{\lambda}(-z)\widehat{L}_h(z)e^{-zx_0}.$$

However,

$$\int_{-\infty}^{\infty} e^{-zx}\frac{1}{h}K\Big(\frac{x-x_0}{h}\Big)\mathrm{d}x = e^{-zx_0}\widehat{K}(zh).$$

It follows from the definition of $L_h(t)$ and the inversion formula for the bilateral Laplace transform (compare with chapter 7, section 5 in Widder 1946) that $\widehat{L}_h(z) = \widehat{K}(zh)/\widehat{\lambda}(-z)$ for all z in a region where $\widehat{K}(\cdot h)/\widehat{\lambda}(-\cdot)$ is analytic. Then L_h is indeed a solution to Equation (10). □

Now we are in a position to define the estimator of $G(x_0)$ based on n independent realizations $X_1(t), \ldots, X_n(t)$ of the queue-length process $\{X(t), t \geq 0\}$ observed on the interval $[0, T]$. Let $\{t_j^{(k)}\}$ be an ordered set of the departure and arrival epochs of realization $X_k(t)$, $k = 1, \ldots, n$, so that $X_k(\cdot)$ is constant in between any two sequential epochs. The estimator of $G(x_0)$ is defined as follows:

$$\tilde{G}_h(x_0) := 1 - \frac{1}{n}\sum_{k=1}^{n}\int_0^T L_h(t-x_0)X_k(t)\mathrm{d}t = 1 - \frac{1}{n}\sum_{k=1}^{n}\sum_j X_k(t_j^{(k)})\mathbf{1}_{[0,T]}\Big(t_j^{(k)}\Big)\int_{t_j^{(k)}}^{t_{j+1}^{(k)}\wedge T} L_h(t-x_0)\mathrm{d}t. \tag{11}$$

Note that $\frac{1}{n}\sum_{k=1}^{n}\int_0^T L_h(t-x_0)X_k(t)\mathrm{d}t$ is an empirical estimator of the integral on the left-hand side of Equation (9) based on the observation $\mathscr{X}_{n,T}$. The construction depends on the design parameter h that will be specified in what follows.

4.2. Upper Bound on the Risk

Now we study the accuracy of the estimator $\tilde{G}_h(x_0)$ defined in Equation (11). The risk of the estimator $\tilde{G}_h(x_0)$ will be analyzed under local smoothness and moment assumptions on the probability distribution G. In particular, we will assume that G belongs to a local Hölder class and has a bounded second moment.

Definition 1. Let $A > 0$, $\beta > 0$, and $d > 0$. We say that a function f belongs to the class $\mathscr{H}_{\beta,x_0}(A)$ if f is $\lfloor\beta\rfloor := \max\{k \in \mathbb{N}_0 : k < \beta\}$ times continuously differentiable on $(x_0 - d, x_0 + d)$ and

$$|f^{(\lfloor\beta\rfloor)}(x) - f^{(\lfloor\beta\rfloor)}(x')| \leq A|x - x'|^{\beta - \lfloor\beta\rfloor}, \quad \forall x, x' \in (x_0 - d, x_0 + d).$$

Definition 2. Let $M \geq 0$. We say that a distribution function G supported on $[0,\infty)$ belongs to the class $\mathscr{G}(M)$ if

$$\max_{k=1,2}\left\{\int_0^\infty kt^{k-1}[1-G(t)]\mathrm{d}t\right\} \leq M.$$

We denote also that

$$\bar{\mathscr{H}}_{\beta,x_0}(A,M) := \mathscr{H}_{\beta,x_0}(A) \cap \mathscr{G}(M).$$

Let K be a kernel supported on $[0,1]$ satisfying the following conditions:

(K1) For a fixed positive integer m,

$$\int_0^1 K(t)\mathrm{d}t = 1, \quad \int_0^1 t^k K(t)\mathrm{d}t = 0, \quad k = 1,\ldots,m.$$

(K2) For a positive integer number r, kernel K is r times continuously differentiable on $\mathbb{R}$ and

$$\max_{x\in[0,1]} |K^{(j)}(x)| \leq C_K < \infty, \quad \forall j = 0,1,\ldots,r.$$

Note that conditions (K1) and (K2) are standard in nonparametric kernel estimation (see, e.g., Tsybakov 2009).

Theorem 2. *Suppose that Assumptions 1 and 2 hold and $G \in \bar{\mathscr{H}}_{\beta,x_0}(A,M)$. Let K be a kernel satisfying conditions (K1) and (K2) with $m \geq \lfloor\beta\rfloor + 1$ and $r > \gamma + 1$. Let $\tilde{G}_*(x_0)$ be the estimator (11) associated with kernel K and parameter $h = h_* := (M\varkappa A^{-2}N^{-1})^{1/(2\beta+2\gamma+1)}$, where*

$$N := \lambda_0 n, \qquad \varkappa := \begin{cases} \sigma_\lambda^{-1} e^{2\sigma_\lambda x_0}, & \sigma_\lambda > 0\,, \\ a_1 + a_2 x_0^p, & \sigma_\lambda = 0. \end{cases}$$

If $T \to \infty$ so that $\frac{1}{T}\ln N \to 0$ as $N \to \infty$, then

$$\limsup_{N\to\infty}\left\{\varphi_N^{-1}\mathcal{R}_{x_0}[\tilde{G}_*; \bar{\mathscr{H}}_{\beta,x_0}(A,M)]\right\} \leq C_1, \tag{12}$$

where C_1 may depend on β, γ, and p only, and

$$\varphi_N := A^{(2\gamma+1)/(2\beta+2\gamma+1)}\left(M\varkappa N^{-1}\right)^{\beta/(2\beta+2\gamma+1)}.$$

Remark 1. The risk of $\tilde{G}_*$ converges to zero at the nonparametric rate $O(N^{-\beta/(2\beta+2\gamma+1)})$ as $N \to \infty$. The "ill-posedness index" γ is determined by the smoothness of function $e^{-\sigma t}\lambda(t)\mathbf{1}_{[0,\infty)}(t)$ with appropriate σ on the entire real line. In particular, if λ is continuously differentiable on $(0,\infty)$ and $\lambda(0) > 0$, then $\gamma = 1$, and the resulting rate is $O(N^{-\beta/(2\beta+3)})$. The deconvolution problem is much harder if $e^{-\sigma t}\lambda(t)\mathbf{1}_{[0,\infty)}(t)$ is smooth on $\mathbb{R}$: in this case, the rate at which the risk converges to 0 slows down. For instance, under conditions of Example 2 with $p > 0$, the resulting rate is $O(N^{-\beta/(2\beta+2p+3)})$.

Remark 2. The asymptotic regime $N = \lambda_0 n \to \infty$ of Theorem 2 includes two natural cases of interest: (i) the number of realizations n goes to infinity, whereas the arrival intensity $\lambda(\cdot)$ is fixed, and (ii) in the heavy traffic regime, the number n of realizations is fixed, but the amplitude λ_0 goes to infinity. The result of Theorem 2 shows that, even if only one realization of the queue-length process is observed, the estimator of the service time distribution can be accurate, provided that the arrival rate amplitude λ_0 is large enough.

Remark 3. In general, the rate of convergence φ_N may depend on x_0, and this dependence is primarily determined by behavior of the arrival rate at infinity. In particular, if $\lambda(\cdot)$ increases exponentially, then the accuracy is proportional to $(e^{2\sigma_\lambda x_0})^{\beta/(2\beta+2\gamma+1)}$ and deteriorates with growth of x_0. Note, however, that if $\lambda(\cdot)$ is bounded (here $\sigma_\lambda = 0$ and $p = 0$), then the upper bound does not depend on x_0.

Remark 4. In cases $\sigma_\lambda > 0$ and $\sigma_\lambda = 0$, $p = 0$, the statement of the theorem remains intact if boundedness of the second moment of G in the definition of $\bar{\mathscr{H}}_{\beta,x_0}(A,M)$ is replaced by the boundedness of the first moment. This is not so in the case $\sigma_\lambda = 0$, $p > 0$: if boundedness of the first moment only is assumed, then the dependence on x_0 is worse for large x_0 as the bound becomes proportional to x_0^{p+1}.

5. Estimation of the Expected Service Time

In this section, we consider the problem of estimating the expected service time $\mu_G = \int_0^\infty [1 - G(t)]\mathrm{d}t$ from observations $\mathscr{X}_{n,T} := \{X_k(t), 0 \le t \le T, k = 1, \ldots, n\}$.

5.1. Estimator Construction

For real number $b > 1/4$ (where $1/4$ is chosen arbitrarily), let $\psi_b(t)$ be an infinitely differentiable function on $\mathbb{R}$ such that

$$\psi_b(t) = \begin{cases} 1, & t \in [0, b], \\ 0, & t \notin [-1/4, b + 1/4]. \end{cases}$$

To define $\psi_b(t)$ on the intervals $[-1/4, 0]$ and $[b, b + 1/4]$, we use standard construction. Let $\psi_0(t) = e^{-\frac{1}{t(1-t)}}$, $t \in [0, 1]$; then, on the interval $[-1/4, 0]$ where ψ_b climbs from 0 to 1, we put

$$\psi_b(t) = 4c_0 \int_{-1/4}^{t} \psi_0(4x + 1)\mathrm{d}x, \quad -\tfrac{1}{4} \le t \le 0,$$

whereas on the interval $[b, b + 1/4]$ where $\psi_b(t)$ descends from 1 to 0, we let

$$\psi_b(t) = 1 - 4c_0 \int_{b}^{t} \psi_0(4(x - b))\mathrm{d}x, \quad b \le t \le b + \tfrac{1}{4}.$$

Because ψ_b is infinitely differentiable, the Laplace transform $\widehat{\psi}_b$ is an entire function.

Define

$$J_b(t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\widehat{\psi}_b(s + i\omega)}{\widehat{\lambda}(-s - i\omega)} e^{(s+i\omega)t}\mathrm{d}t, \quad s < -\sigma_\lambda, \tag{13}$$

where σ_λ is the abscissa of convergence of the Laplace transform $\widehat{\lambda}$ of λ. The following statement is a key result on properties of the function $J_b(t)$.

Lemma 2. *If*

$$\int_{-\infty}^{\infty} \left| \frac{\widehat{\psi}_b(s + i\omega)}{\widehat{\lambda}(-s - i\omega)} \right| \mathrm{d}t < \infty, \quad \int_0^\infty |J_b(t)| H(t)\mathrm{d}t < \infty,$$

then

$$\int_0^\infty J_b(t) H(t)\mathrm{d}t = \int_0^\infty \psi_b(x) \bar{G}(x)\mathrm{d}x.$$

The proof of Lemma 2 is omitted, because it goes along the same lines as the proof of Lemma 1. By construction, ψ_b is an infinitely differentiable function, which can be viewed as an approximation to the indicator function $\mathbf{1}_{[0,b]}(t)$. Therefore, for large values of b, the integral on the right-hand side of the preceding display equation approximates μ_G; this fact underlies the construction of our estimator.

The estimator of μ_G is defined as follows:

$$\tilde{\mu}_G = \frac{1}{n} \sum_{k=1}^{n} \int_0^T J_b(t) X_k(t)\mathrm{d}t. \tag{14}$$

The estimator depends on the tuning parameter b, which will be specified in what follows.

5.2. Upper Bounds on the Risk

The next two statements establish upper bounds on the risk of $\tilde{\mu}_G$ under different growth conditions on the arrival rate (Assumptions 1(a) and 1(b), respectively).

Theorem 3. *Let Assumptions 1(a) and 2 hold, and let $G \in \mathscr{G}(M)$. Let $\tilde{\mu}_G$ be the estimator (14) associated with parameter*

$$b = b_* := \frac{1}{2\sigma_\lambda} \left\{ \ln\left(\sigma_\lambda^{-2\gamma+1} MN\right) - 3 \ln\left[\frac{1}{2\sigma_\lambda} \ln\left(\sigma_\lambda^{-2\gamma+1} MN\right)\right] \right\} - \frac{1}{4},$$

where $N := \lambda_0 n$. *If* $T \to \infty$ *and* $\frac{1}{T}[\ln N]^2 \to 0$ *as* $N \to \infty$, *then*

$$\limsup_{N\to\infty}\{\varphi_N^{-1}\mathcal{R}[\tilde{\mu}_G;\mathcal{G}(M)]\} \le C_1, \quad \varphi_N := M\sigma_\lambda\left[\ln\left(\frac{MN}{\sigma_\lambda^{2\gamma-1}}\right)\right]^{-1},$$

and C_1 *is a positive constant depending on* γ *and* k_0 *only.*

Theorem 4. *Let Assumptions 1(b) and 2 hold, and let* $G \in \mathcal{G}(M)$. *Let* $\tilde{\mu}_G$ *be the estimator (14) associated with parameter*

$$b = b_* := (MN)^{1/(p+2)}, \quad N := \lambda_0 n.$$

If $\frac{1}{T}N^{1/(p+2)}\ln N \to 0$ *as* $N \to \infty$, *then*

$$\limsup_{N\to\infty}\{\varphi_N^{-1}\mathcal{R}[\tilde{\mu}_G;\mathcal{G}(M)]\} \le C_2, \quad \varphi_N := M^{(p+1)/(p+2)}N^{-1/(p+2)},$$

where C_2 *is a positive constant depending on* γ, k_0, *and* p *only.*

Remark 5. The results of Theorems 3 and 4 hold without any conditions on the smoothness of G. In particular, G may be a discrete distribution with a bounded second moment.

Remark 6. Theorems 3 and 4 show that the accuracy of $\tilde{\mu}_G$ is primarily determined by the growth of the arrival rate at infinity. In particular, if the arrival rate increases exponentially (i.e., $\sigma_\lambda > 0$), then the risk of $\tilde{\mu}_G$ converges to 0 at a slow logarithmic rate. At the same time, if the arrival rate is bounded (i.e., $\sigma_\lambda = 0$ and $p = 0$), then the risk tends to 0 at the parametric rate. A close inspection of the proof shows that growth of the arrival rate manifests in growth of the variance of $X(\cdot)$, which, in turn, affects the rate of convergence.

6. Implementation and Numerical Examples

In this section, we provide details on implementation of the proposed estimators. In particular, we discuss numerical methods for calculating kernels L_h and J_b and an adaptive scheme for bandwidth selection. We also conduct a small simulation study to illustrate the practical behavior of the proposed estimators.

6.1. Implementation Issues

In our simulations, we consider a Gaussian kernel: that is,

$$K(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}, \quad \text{and thus,} \quad \widehat{K}(z) = e^{z^2/2}$$

for all $z \in \mathbb{C}$. The reason that we use this kernel is that, in some cases, L_h can be computed explicitly. If such analytical expressions are not available, we resort to numerical integration and inversion techniques. For consistency, we still use the Gaussian kernel in the numerical cases, although this is obviously not necessary.

It is computationally somewhat expensive to implement the theoretical kernel J_b in Equation (13) as covered in Theorems 3 and 4 because the integral of ψ_0 does not have an explicit form, and also, its transform does not have an explicit form. From that point of view, it is more attractive to work with a slightly different kernel J_b, in particular

$$\widehat{J}_b(z) := \frac{\widehat{\psi}(z)}{\widehat{\lambda}(-z)} = \frac{(1-e^{-zb})e^{z^2h^2/2}}{z\widehat{\lambda}(-z)},$$

where ψ is a convolution of the indicator function $\mathbf{1}_{[0,b]}(\cdot)$ and $(1/h)K(\cdot/h)$, and K is a standard Gaussian kernel. In some cases, we let ψ to be a convolution of the indicator function $\mathbf{1}_{[0,b+x_1]}(\cdot)$ and $(1/h)K(\cdot - x_1/h)$ for some $x_1 < 0$. The shift to the left with $-x_1$ is there to get rid of some bias caused by the boundary effects as h approaches zero.

6.1.1. Exact Calculation of Kernels L_h and J_b. Before we consider cases where only numerical calculation of kernels is possible, we present the following examples where explicit expressions for kernels L_h and J_b are available.

Example 5 (Constant Arrival Rate). Suppose that $\lambda(t) = \lambda_0 \mathbf{1}_{[0,\infty)}(t)$; then $\widehat{\lambda}(-z) = -\lambda_0/z$ for $\mathrm{Re}(z) < 0$. It can be checked that the assumptions in Lemma 1 are satisfied; hence, L_h is well defined for $c < 0$ by

$$L_h(t) = -\frac{1}{2\pi\lambda_0}\int_{-\infty}^{\infty}(c+i\omega)e^{(c+i\omega)t+\frac{1}{2}h^2(c+i\omega)^2}\mathrm{d}\omega = \frac{te^{-\frac{t^2}{2h^2}}}{\sqrt{2\pi}\lambda_0 h^3}.$$

To estimate the expected value of service times, we consider the kernel J_b, which in this case equals

$$J_b(t) = -\frac{1}{2\pi\lambda_0}\int_{-\infty}^{\infty}(1-e^{-(c+i\omega)b})e^{(c+i\omega)t+\frac{1}{2}h^2(c+i\omega)^2}\mathrm{d}\omega = \frac{1}{\sqrt{2\pi}\lambda_0 h}\left(e^{-\frac{1}{2h^2}(b-t)^2} - e^{-\frac{1}{2h^2}t^2}\right).$$

This can be seen as the difference of two Gaussian kernels: one centered at b and one at zero. As $h \downarrow 0$, it converges to the difference of Dirac delta functions centered on b and 0. In other words, for $h \downarrow 0$, the estimator of the mean service time is

$$\tilde{\mu}_G = \lim_{h\downarrow 0}\frac{1}{n\lambda_0}\sum_{i=1}^{n}\int_0^{\infty}J_b(t)X_i(t)\mathrm{d}t = \frac{1}{n\lambda_0}\sum_{i=1}^{n}X_i(b).$$

In this special case, it is also obvious how such an estimator could be obtained directly by observing that

$$\mathbb{E}_G[X(b) - X(0)] = \int_0^b \lambda(t-u)\bar{G}(u)\mathrm{d}u = \lambda_0\int_0^b \bar{G}(u)\mathrm{d}u.$$

Example 6 (High/Low Switching). Suppose that the arrival rate repetitively switches between a high and low arrival rate λ_1, λ_2 after each time unit. Mathematically, we can write this as

$$\lambda(t) = \lambda_1\sum_{j=0}^{\infty}\mathbf{1}_{(2j,2j+1]}(t) + \lambda_2\sum_{j=0}^{\infty}\mathbf{1}_{(2j+1,2j+2]}(t).$$

Elementary calculus yields $\widehat{\lambda}(z) = (\lambda_1 + \lambda_2 e^{-z})z^{-1}(1+e^{-z})^{-1}$. In the special case that the low rate is equal to zero, the kernels L_h and J_b can be calculated explicitly. Suppose that $\lambda_1 > 0$ and $\lambda_2 = 0$; then

$$L_h(t) = \frac{1}{\sqrt{2\pi}\lambda_1 h^3}e^{-\frac{(t+1)^2}{2h^2}}\left(1 + t + te^{\frac{2t+1}{2h^2}}\right),$$

from which the estimator of $G(x_0)$ follows. Furthermore, it can be calculated that

$$J_b(t) = \frac{1}{\sqrt{2\pi}\lambda_1 h}\left(e^{-\frac{t^2}{2h^2}} + e^{-\frac{(t-1)^2}{2h^2}} - e^{-\frac{(t-(1+b))^2}{2h^2}} - e^{-\frac{(t-b)^2}{2h^2}}\right).$$

Following the same logic as in the previous example and by shifting the kernel J_b to the left by 1 unit to get rid of bias, we see that (as $h \downarrow 0$)

$$\tilde{\mu}_G = \frac{1}{n\lambda_1}\sum_{i=1}^{n}(X_i(b-1) + X_i(b)). \tag{15}$$

Example 7 (Linearly Increasing Arrival Rate). Suppose that $\lambda(t) = \lambda_0 t\,\mathbf{1}_{[0,\infty)}(t)$; then $\widehat{\lambda}(z) = \lambda_0 z^{-2}$, and therefore,

$$L_h(t) = \frac{1}{\sqrt{2\pi}\lambda_0 h^5}e^{-\frac{t^2}{2h^2}}(t^2 - h^2).$$

Furthermore, with similar calculations as before, we find that

$$J_b(t) = -\frac{1}{\sqrt{2\pi}h^3}e^{-\frac{b^2+t^2}{2h^2}}\left(te^{\frac{b^2}{2h^2}} + (b-t)e^{\frac{bt}{h^2}}\right).$$

In other words, J_b is a differentiation kernel centered around b minus a differentiation kernel centered at zero. It can also be argued intuitively that differentiation plays a role for the estimator. Indeed, note that

$$\frac{\mathrm{d}}{\mathrm{d}b}\mathbb{E}X(b) = \frac{\mathrm{d}}{\mathrm{d}b}\left[\int_0^b (b-u)\bar{G}(u)\mathrm{d}u\right] = \int_0^b \bar{G}(u)\mathrm{d}u,$$

and for b large, the last integral is an approximation for μ_G.

6.1.2. Numerical Evaluation of Kernels L_h and J_b. For some arrival rates $\lambda(\cdot)$, there is no analytic form for L_h and J_b. In that case, one needs to resort to numerical inversion. Following Durbin (1973), we discretize the Bromwich integral with a trapezoidal rule. Although it is assumed in Durbin (1973) that the inverse transform is zero on the negative half-line, the algorithm still works if this assumption does not hold. For completeness, we provide the following inversion formula (L_h and $\widehat{L}_h$ can be replaced by J_b and $\widehat{J}_b$, respectively):

$$L_h(t) = \frac{2e^{ct}}{\tilde{T}}\left(\frac{1}{2}\mathrm{Re}(\widehat{L}_h(c)) + \sum_{k=1}^{n_{\max}}\mathrm{Re}\Big(\widehat{L}_h\big(c+ik\pi/\tilde{T}\big)\Big)\cos(k\pi t/\tilde{T}) - \sum_{k=0}^{n_{\max}}\mathrm{Im}\Big(\widehat{L}_h\big(c+ik\pi/\tilde{T}\big)\Big)\sin(k\pi t/\tilde{T})\right),$$

where the error is controlled by the parameters $n_{\max}, c, \tilde{T}$ and the machine precision. When applying this formula, it is important that c is chosen in the strip of convergence of $\widehat{L}$. For example, if $\widehat{L}(z)$ is defined for z such that $\mathrm{Re}(z) < 0$, we take c small but $c < 0$. We stuck to the guideline that $|c\tilde{T}| \approx 30$. In principle, picking c closer to the edge of the strip should improve accuracy. However, when c is too close to the edge of the strip, the solution will become numerically unstable. For more details on picking the right parameters and bounds on the error, see Durbin (1973).

6.2. Adaptive Scheme for Bandwidth Selection

The bandwidth h controls the trade-off between bias and variance when estimating the distribution function: lower h corresponds to high variance and low bias, and vice versa. In applications, the smoothness of G is unknown, and therefore, the bandwidth choice given in Theorem 2 is not feasible. In our simulations, we implemented a variant of the adaptive procedure of Lepskiĭ (1990) proposed in Goldenshluger and Nemirovski (1997).

The adaptive bandwidth scheme is as follows:

1. Pick a minimum bandwidth $h_{\min} > 0$, and define $h_i := (1+\alpha)^i h_{\min}$ for $\alpha > 0$.
2. Estimate the variance of $\tilde{G}_{h_i}(x_0)$, and denote the estimate by $v_{h_i}^2$.
3. Define I_{h_i} as the interval $\big[\tilde{G}_{h_i}(x_0) - 2\varkappa v_{h_i}, \tilde{G}_{h_i}(x_0) + 2\varkappa v_{h_i}\big]$, with $\varkappa = \frac{1}{4}\sqrt{\ln(n)}$.
4. The estimator will be the middle of the last interval I_{j^*} that is nonempty: that is,

$$j^* := \max\left\{j \ : \ \bigcap_{i=1}^{j} I_{h_i} \neq \emptyset\right\}.$$

Note that the variance of $\tilde{G}_h(x_0)$ is given by

$$\mathrm{var}_G\big[\tilde{G}_h(x_0)\big] = \frac{1}{n}\int_0^T \int_0^T L_h(t-x_0)\overline{L_h(\tau-x_0)}R(t,\tau)\mathrm{d}t\mathrm{d}\tau,$$

where $R(t,\tau) := \mathrm{cov}_G[X(t), X(\tau)]$ is determined by Equation (3). We discretize the double integral and estimate the covariance function by its empirical counterpart. In particular, we define a uniform time grid $\{t_i : 1 \le i \le N\}$ with grid size Δ, a vector $L \in \mathbb{R}^N$ of which the ith component is $L_h(t_i - x_0)$, a matrix $X \in \mathbb{R}^{N\times n}$ such that $X_{ij} := X^{(i)}(t_j)$, and a covariance matrix $R \in \mathbb{R}^{N\times N}$ such that element ij gives the sample covariance between $X(t_i)$ and $X(t_j)$. More precisely,

$$R = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i\cdot} - \bar{X})(X_{i\cdot} - \bar{X})^{\mathrm{T}},$$

where $X_{ij} = X_i(t_j)$, and $\bar{X}$ is such that $\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$ for $1 \le j \le N$. With this notation, the variance of the estimator is approximated by

$$v_{h_i}^2 = \frac{1}{n-1}\sum_{k=1}^{N}\sum_{\ell=1}^{N} L_{h_i}(t_k - x_0)\overline{L_h(t_\ell - x_0)}R(t_k, t_\ell)\Delta^2 = \frac{\Delta^2}{n-1}L^* R L.$$

6.3. Simulation Results for Estimating the Distribution Function

Now we present a small simulation study on some particular examples. The goal is to verify consistency and compare accuracies in two cases where $\lambda(\cdot)$ is varied and G is kept constant. Consider the following examples:

1a. Take $G(x) = 1 - e^{-x}$, $\lambda(t) = 10(1+\cos(t))$ for $t \ge 0$. Set parameters $h_{\min} = 0.025$ and $\alpha = 0.25$.
1b. Take $G(x) = 1 - e^{-x}$, $\lambda(t) = 10t$ for $t \ge 0$. Set parameters $h_{\min} = 0.05$ and $\alpha = 0.15$.

Figure 1 corresponds to Case 1a, and Figure 2 corresponds to Case 1b. In these figures (from left to right and from upper to lower), we plotted the RMSE as a function of n, which is based on 50 runs of $1-\tilde{G}(1)$; the estimator of $1-G(1)=e^{-1}\approx 0.368$, which is plotted next to it; and the kernel $L_h(t)$ for several values of h, as indicated in the captions. For Case 1, we also included a boxplot to show the performance of these 50 runs for x_0 ranging from 0.5 to 3.5.

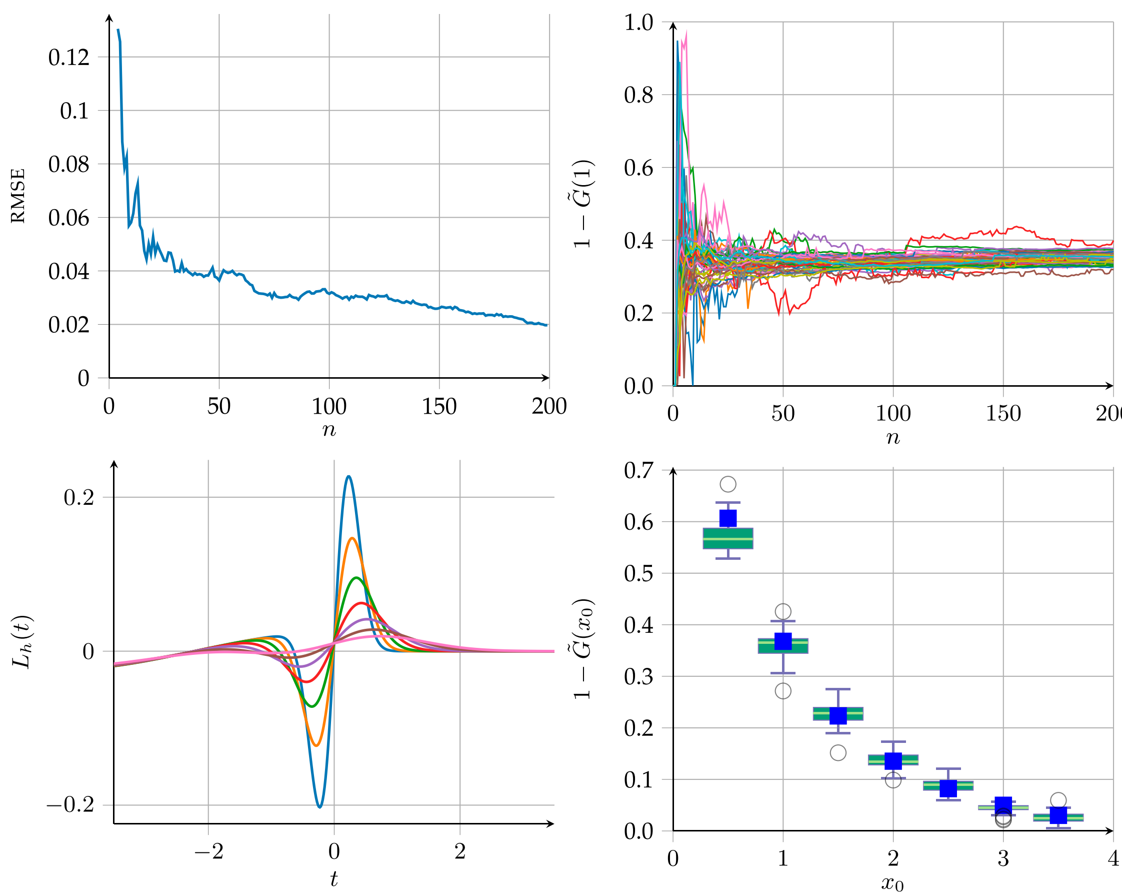
From the figures, we note the following. First, the RMSE is clearly smaller in Case 1a than in Case 1b. This was to be expected, because $\gamma=1$ in Case 1a and $\gamma=2$ in Case 1b, which lead to inferior worst case minimax convergence rates in theory. Note that in Case 1b, the estimators are often not even within $[0,1]$. In Case 1a, we see some slight bias for smaller x_0: it seems that estimation for x small in this case is relatively hard, which could be owing to the fact that $G(x)$ is steeper for small x. The L_h kernel in Case 1a seems quite similar to the standard Gaussian differentiation kernel that would be used in the case of constant $\lambda(\cdot)$ (when $\lambda(\cdot)\equiv\lambda_0$ is constant, the derivative of $\int_0^t \lambda_0\bar{G}(s)ds$ equals $\lambda_0\bar{G}(t)$). However, there is an additional "bump" that compensates for the fact that $\lambda(\cdot)$ is a cosine function rather than a constant.

6.4. Simulation Results for Estimating the Mean

We consider the following numerical settings:

2a. Take $\lambda(t)=1+\sin(t)$, $G(x)=1-e^{-x}$, $x_1=-1$, and $h=0.08$ (as small as possible so that the numerical inversion algorithm is still stable), with parameters $\tilde{T}=30$, $c=-1$, and $b=25$.

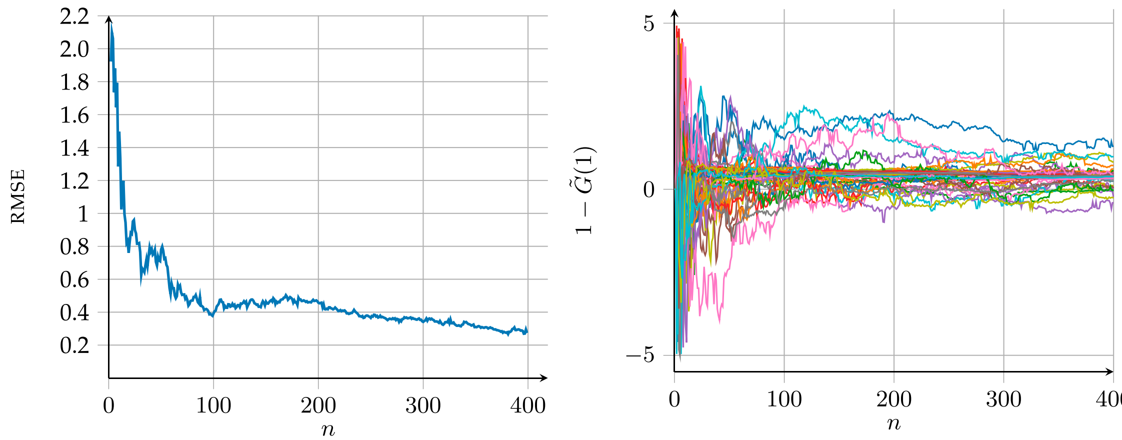
Figure 1. Results Corresponding to Case 1a

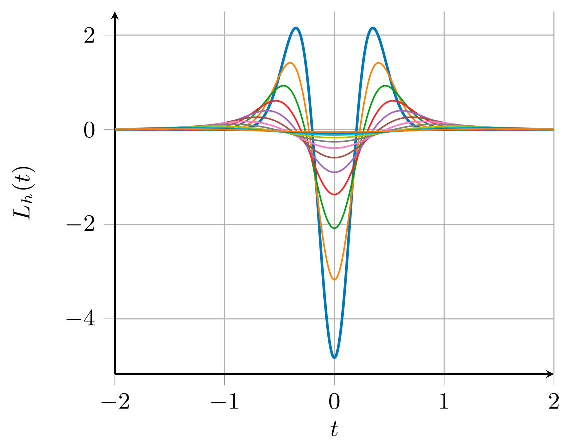


x_0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
$1-G(x_0)$	0.607	0.368	0.223	0.135	0.0821	0.0414	0.0302
Mean $1-\tilde{G}(x_0)$	0.570	0.361	0.228	0.136	0.0976	0.0441	0.0261
St. dev.	0.028	0.025	0.021	0.015	0.0140	0.0076	0.0097

Notes. (Top left) Plot of the RMSE versus the number of observations n averaged over 50 runs. (Top right) Fifty runs for $n=0$ to $n=200$ for $x_0=1$. (Middle left) L_h for $h\in[0.23,0.89]$. More precisely, $h=0.025\cdot(1.25)^q$, with $q=10,11,\ldots,16$. Smaller h corresponds to steeper bumps. (Middle right) Boxplot of 50 runs with $n=200$. (Bottom) The table corresponds to the boxplot. St. dev., standard deviation.

Figure 2. Results Corresponding to Case 1b





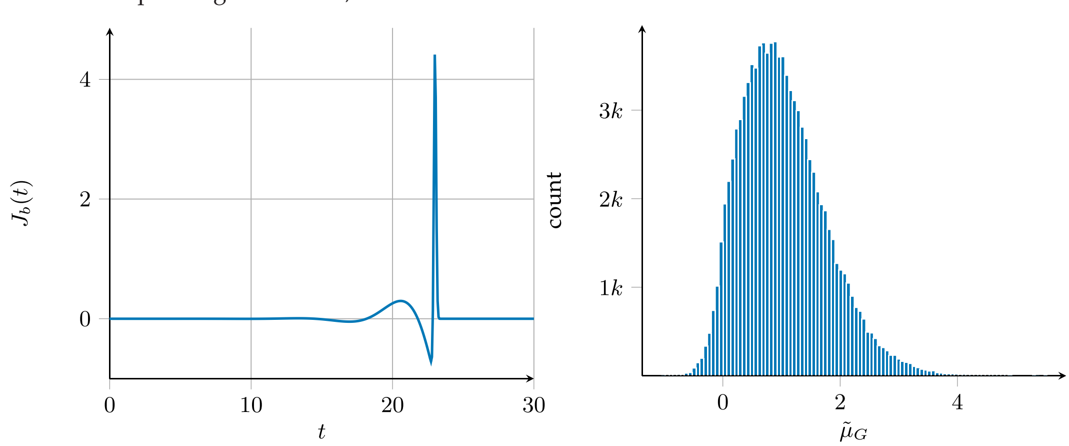
n	100	200	300	400
Mean	0.43	0.42	0.35	0.37
St. dev.	0.39	0.45	0.34	0.28

Notes. (Upper left) RMSE versus n. (Upper right) Fifty runs for $n = 0$ to $n = 400$ for $x_0 = 1$. (Lower left) L_h for $h \in [0.23, 0.94]$. More precisely, $h = 0.05 \cdot (1.15)^q$, with $q = 10, 11, \ldots, 21$. Smaller h corresponds to steeper bumps. (Lower right) Descriptive statistics of estimator $1 - \tilde{G}(1)$. St. dev., standard deviation.

2b. Let λ be the on/off rate of Example 2, and suppose that service times are uniformly distributed on $[0, 2]$ such that $\mu_G = 1$. We take $b = 25$ in estimator (15).

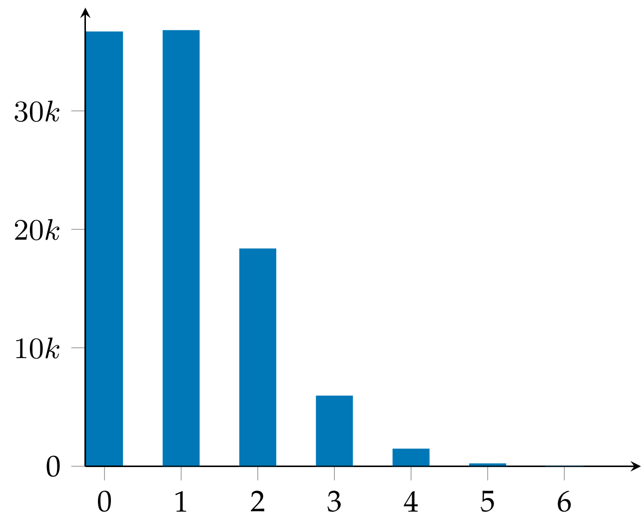
Consider Case 2a, the results of which are shown in Figure 3. The histogram shows the distribution of $\hat{\mu}_G$, each based on a single observation. Note that the distribution is approximately normal, with some skew to the left, mean of 1.03, and standard deviation of 0.73. Although the standard deviation seems quite large, one must take into account that the kernel J_b is almost 0 outside of [15, 25]. During those 10 time units, there are

Figure 3. Results Corresponding to Case 2a, Where We Denote $k = 10^3$



Note. In the left panel, we plotted the kernel J_b, and in the right panel, we plotted $\tilde{\mu}_G$ based on a single observation for 10^5 simulation runs.

Figure 4. Results Corresponding to Case 2b, Where We Denote $k = 10^3$



Notes. We plotted the histogram of estimators based on a single observation. All estimates based on a single observation are obviously integer valued. The mean of this sample is 0.999, with a standard deviation of 0.998.

approximately 10 arrivals. Effectively, the amount of useful data is relatively small. Furthermore, note that the kernel J_b in Case 2a is quite similar to kernel J_b in the case of constant $\lambda(\cdot)$ (as in Example 5): there is a Dirac delta function at $b + x_1$, and in addition to that, there is a wave preceding the Dirac delta function to somehow compensate for the fact that λ is not constant.

In Case 2b (Figure 4), we used the explicit kernels as calculated in Example 6. The bias turns out to be negligible, and the standard deviation is approximately $\frac{1}{\sqrt{n}}$ after n observations.

6.5. A Concluding Remark

We close this section with a remark on practical applicability of the proposed estimators. Throughout this paper, we have assumed that the arrival intensity $\{\lambda(t), t \geq 0\}$ is known, but in practice, it may be unknown. We note, however, that the proposed estimators can be modified to accommodate this problem setting. In particular, based on the observations of the queue-length process $\{X(t), t \geq 0\}$ on the interval $[0, T]$, we can estimate the value $\widehat{\lambda}(-z)$ by $\widetilde{\ell}(-z) = \sum_j e^{z\tau_j}\mathbf{1}_{[0,T]}(\tau_j)$, where $\{\tau_j, j \in \mathbb{Z}\}$ are the arrival epochs. Then $\widehat{\lambda}(-z)$ should be substituted by $\widetilde{\ell}(-z)$ (if necessary with a truncation that separates $|\widetilde{\ell}(-z)|$ away from zero, effectively enforcing the requirement of no roots in Assumption 2) in Equations (7) and (13). Because $\widetilde{\ell}(-z)$ is an accurate estimator of $\widehat{\lambda}(-z)$, we expect that the convergence rates of the modified estimators will coincide with those presented in Theorems 2, 3, and 4. In general, the problem setting with unknown arrival intensity deserves an independent study, and we leave it for future research.

7. Proofs

7.1. Proof of Theorem 1

The proof is similar to the one for proposition 1 in Goldenshluger (2016). By conditioning on $\{\tau_j, j \in \mathbb{Z}\}$, we obtain that

$$\mathbb{E}_G \exp\left\{\sum_{i=1}^{m} \theta_i X(t_i)\right\} = \mathbb{E}_G \exp\left\{\sum_{j\in\mathbb{Z}} f(\tau_j)\right\},$$

where

$$f(x) = \ln\left(1 + \sum_{k=1}^{m}\left[e^{\sum_{i=1}^{k} \theta_i \mathbf{1}(x \leq t_i)} - 1\right]\mathbb{P}_G\{\sigma_1 \in I_k(x)\}\right),$$

$I_k(x) = (t_k - x, t_{k+1} - x]$, $k = 1, \ldots, m-1$, and $I_m(x) = (t_m - x, \infty)$. Then Campbell's formula yields

$$\mathbb{E}_G \exp\left\{\sum_{j\in\mathbb{Z}} f(\tau_j)\right\} = \exp\left\{\int_{-\infty}^{\infty} [e^{f(x)} - 1]\lambda(x)\mathrm{d}x\right\},$$

and our current goal is to evaluate the integral on the right-hand side of the last equality. Denote this integral by S_m:

$$\begin{aligned} S_m &:= \int_{-\infty}^{\infty} [e^{f(x)} - 1]\lambda(x)\mathrm{d}x \\ &= \sum_{k=1}^{m-1} \int_{-\infty}^{\infty} \left(\exp\left\{\sum_{i=1}^{k} \theta_i \mathbf{1}(x \le t_i)\right\} - 1\right)\left[\bar{G}(t_k - x) - \bar{G}(t_{k+1} - x)\right]\lambda(x)\mathrm{d}x \\ &\quad + \int_{-\infty}^{\infty} \left(\exp\left\{\sum_{i=1}^{m} \theta_i \mathbf{1}(x \le t_i)\right\} - 1\right)\bar{G}(t_m - x)\lambda(x)\mathrm{d}x =: \sum_{k=1}^{m-1} J_k + L_m. \end{aligned} \tag{16}$$

We derive Equation (2) by iterating Equation (16); we have

$$S_{m+1} = S_m + J_m + L_{m+1} - L_m = S_1 + \sum_{k=1}^{m} [J_k + L_{k+1} - L_k]. \tag{17}$$

For any $k \ge 1$,

$$\begin{aligned} J_k &= \left(e^{\sum_{i=1}^{k} \theta_i} - 1\right) \int_{t_k - t_1}^{\infty} \left[\bar{G}(u) - \bar{G}(t_{k+1} - t_k + u)\right]\lambda(t_k - u)\mathrm{d}u \\ &\quad + \sum_{j=1}^{k-1} \left(e^{\sum_{i=j+1}^{k} \theta_i} - 1\right) \int_{t_k - t_{j+1}}^{t_k - t_j} \left[\bar{G}(u) - \bar{G}(t_{k+1} - t_k + u)\right]\lambda(t_k - u)\mathrm{d}u, \\ L_k &= \left(e^{\sum_{i=1}^{k} \theta_i} - 1\right) \int_{t_k - t_1}^{\infty} \bar{G}(u)\lambda(t_k - u)\mathrm{d}u + \sum_{j=1}^{k-1} \left(e^{\sum_{i=j+1}^{k} \theta_i} - 1\right) \int_{t_k - t_{j+1}}^{t_k - t_j} \bar{G}(u)\lambda(t_k - u)\mathrm{d}u; \end{aligned}$$

this yields

$$\begin{aligned} J_k - L_k &= -\left(e^{\sum_{i=1}^{k} \theta_i} - 1\right) \int_{t_k - t_1}^{\infty} \bar{G}(t_{k+1} - t_k + u)\lambda(t_k - u)\mathrm{d}u - \sum_{j=1}^{k-1} \left(e^{\sum_{i=j+1}^{k} \theta_i} - 1\right) \int_{t_k - t_{j+1}}^{t_k - t_j} \bar{G}(t_{k+1} - t_k + u)\lambda(t_k - u)\mathrm{d}u \\ &= -\left(e^{\sum_{i=1}^{k} \theta_i} - 1\right) \int_{t_{k+1} - t_1}^{\infty} \bar{G}(u)\lambda(t_{k+1} - u)\mathrm{d}u - \sum_{j=1}^{k-1} \left(e^{\sum_{i=j+1}^{k} \theta_i} - 1\right) \int_{t_{k+1} - t_{j+1}}^{t_{k+1} - t_j} \bar{G}(t_{k+1} - t_k + u)\lambda(t_k - u)\mathrm{d}u. \end{aligned}$$

Then we obtain

$$\begin{aligned} J_k - L_k + L_{k+1} &= \left(e^{\theta_{k+1}} - 1\right)H(t_{k+1}) + \left(e^{\theta_{k+1}} - 1\right)\left(e^{\sum_{i=1}^{k} \theta_i} - 1\right)H_{t_{k+1} - t_1, \infty}(t_{k+1}) \\ &\quad + \left(e^{\theta_{k+1}} - 1\right) \sum_{j=1}^{k-1} \left(e^{\sum_{i=j+1}^{k} \theta_i} - 1\right)H_{t_{k+1} - t_{j+1}, t_{k+1} - t_j}(t_{k+1}). \\ &= \left(e^{\theta_{k+1}} - 1\right)\left[H(t_{k+1}) + \sum_{j=0}^{k-1} \left(e^{\sum_{i=j+1}^{k} \theta_i} - 1\right)H_{t_{k+1} - t_{j+1}, t_{k+1} - t_j}(t_{k+1})\right], \end{aligned}$$

where, by convention, we put $t_0 = -\infty$. Taking into account that $S_1 = (e^{\theta_1} - 1)H(t_1)$ and using Equation (17), we complete the proof.

7.2. Auxiliary Results

We state two auxiliary results that are used in the subsequent proofs.

Lemma 3. *Let* $\{X(t), t \ge 0\}$ *be the queue-length process of the* $M_t/G/\infty$ *queue with* $G \in \mathscr{G}(M)$, *and let* $R(t, \tau) := \mathrm{cov}_G[X(t), X(\tau)]$. *Suppose that Assumption 1 holds.*

(a) *If* $\sigma_\lambda > 0$, *then for any* $\nu \ge 0$,

$$\int_0^{\infty} R(t, \tau)\mathrm{d}\tau \le 2M\lambda_0 e^{(\sigma_\lambda + \nu)t}(\sigma_\lambda + \nu)^{-1}. \tag{18}$$

(b) *If* $\sigma_\lambda = 0$, *then*

$$\int_0^{\infty} R(t, \tau)\mathrm{d}\tau \le 3M\lambda_0(a_1 + a_2 t^p).$$

Proof. By Corollary 1,

$$R(t,\tau) := \mathrm{cov}_G[X(t), X(\tau)] = \int_{\tau-t}^{\tau} \bar{G}(u)\lambda(\tau-u)\mathrm{d}u, \quad \forall \tau \ge t \ge 0. \tag{19}$$

(a) First, assume that $\sigma_\lambda > 0$. By Equation (4), we have, for any $\nu \ge 0$,

$$\begin{aligned}\int_0^\infty R(t,\tau)\mathbf{1}(\tau \ge t)\mathrm{d}\tau &= \int_0^\infty \mathbf{1}(\tau \ge t)\int_{\tau-t}^{\tau}\bar{G}(u)\lambda(\tau-u)\mathrm{d}u\mathrm{d}\tau \\ &\le \int_0^\infty \bar{G}(u)\mathrm{d}u \int_0^t \lambda(u)\mathrm{d}u \le M\lambda_0(\sigma_\lambda+\nu)^{-1}e^{(\sigma_\lambda+\nu)t},\end{aligned} \tag{20}$$

and

$$\begin{aligned}\int_0^\infty R(t,\tau)\mathbf{1}(\tau < t)\mathrm{d}\tau &= \int_0^\infty \mathbf{1}(\tau < t)\int_{t-\tau}^{t}\bar{G}(u)\lambda(t-u)\mathrm{d}u\mathrm{d}\tau \\ &= \int_0^t u\bar{G}(u)\lambda(t-u)\mathrm{d}u \le e^{(\sigma_\lambda+\nu)t}\int_0^t ue^{-(\sigma_\lambda+\nu)u}\bar{G}(u)\mathrm{d}u \le M\lambda_0(\sigma_\lambda+\nu)^{-1}e^{(\sigma_\lambda+\nu)t}.\end{aligned} \tag{21}$$

Summing up Equations (20) and (21), we obtain Equation (18).

(b) Second, consider the case $\sigma_\lambda = 0$. Using Equations (5) and (19), we obtain

$$\begin{aligned}\int_0^\infty R(t,\tau)\mathbf{1}(\tau \ge t)\mathrm{d}\tau &= \int_0^\infty\int_0^\infty \bar{G}(u)\mathbf{1}(t\vee u \le \tau \le t+u)\lambda(\tau-u)\mathrm{d}\tau\mathrm{d}u \\ &\le \lambda_0 a_1\int_0^\infty \bar{G}(u)(t\wedge u)\mathrm{d}u + \lambda_0 a_2\int_0^\infty \bar{G}(u)\int_{t\vee u}^{t+u}(\tau-u)^p\mathrm{d}\tau\mathrm{d}u \\ &\le \lambda_0 M a_1 + \frac{\lambda_0 a_2}{p+1}\left[\int_0^t \bar{G}(u)[t^{p+1}-(t-u)^{p+1}]\mathrm{d}u + t^{p+1}\int_t^\infty \bar{G}(u)\mathrm{d}u\right].\end{aligned}$$

Moreover, because $G \in \mathscr{G}(M)$, it holds that $t^{p+1}\int_t^\infty \bar{G}(u)\mathrm{d}u \le Mt^p$ and

$$\int_0^t \bar{G}(u)[t^{p+1}-(t-u)^{p+1}]\mathrm{d}u \le t^{p+1}\int_0^t \bar{G}(u)(p+1)\tfrac{u}{t}\mathrm{d}u \le (p+1)Mt^p,$$

where in the first inequality we have used that $1-[1-(u/t)]^p \le (p+1)u/t$ for $0 \le u \le t$. This yields

$$\int_0^\infty R(t,\tau)\mathbf{1}(\tau \ge t)\mathrm{d}\tau \le 2M\lambda_0(a_1+a_2t^p).$$

Similarly,

$$\int_0^\infty R(t,\tau)\mathbf{1}(\tau < t)\mathrm{d}\tau = \int_0^t u\bar{G}(u)\lambda(t-u)\mathrm{d}u \le M\lambda_0(a_1+a_2t^p).$$

Summing up the last two inequalities, we complete the proof. □

Lemma 4. *Let $\delta > 0$ and $p \ge 0$; then*

$$\int_T^\infty e^{-\delta t}t^p\mathrm{d}t \le 2\delta^{-2}T^{p-1}e^{-\delta T},$$

provided that $\delta T \ge 2p$.

Proof. The statement is an immediate consequence of corollary 2.5 in Borwein and O-Yeat (2009). □

7.3. Proof of Theorem 2

Proof. Throughout the proof, $c_1, c_2, \ldots$ stand for positive constants that do not depend on n, T, or the parameters x_0, A, M, σ_λ, and λ_0. The proof is divided in steps. The numbering of constants at each step of the proof starts anew so that $c_1, c_2, \ldots$ are different on different appearances.

1^0. We show that, under the premise of the theorem, the estimator is well defined. Because $\mathrm{supp}(K)=[0,1]$, $\widehat{K}$ is an entire function. For any $\sigma\in\mathbb{R}$,

$$\widehat{K}(\sigma+i\omega)=\int_0^1 K(t)e^{-(\sigma+i\omega)t}\mathrm{d}t,$$

and by condition (K2), kernel K is r times continuously differentiable. Hence, repeatedly integrating by parts, we obtain that, for any ω and σ,

$$|\widehat{K}(\sigma+i\omega)|\le c_1e^{|\sigma|}(\sigma^2+\omega^2)^{-r/2}. \tag{22}$$

In addition, in view of (K2) and for any ω and σ,

$$|\widehat{K}(\sigma+i\omega)|\le c_1e^{|\sigma|}. \tag{23}$$

By Assumption 2, $\widehat{\lambda}(z)$ does not have zeros in the half-plane $\{z\in\mathbb{C}:\mathrm{Re}(z)>\sigma_\lambda\}$. This implies that $\widehat{K}(zh)/\widehat{\lambda}(-z)$ is an analytic function in $\{z\in\mathbb{C}:\mathrm{Re}(z)<-\sigma_\lambda\}$, and integration in Equation (7) can be performed on any vertical line $\mathrm{Re}(z)=s$ with $s<-\sigma_\lambda$.

Put $s=-\sigma_\lambda-\delta$, where $\delta>0$ is a parameter to be specified. In the subsequent proof, we assume that $\delta<1/h$ and $h<1$; this does not lead to loss of generality because h will be chosen tending to 0 as $\lambda_0 n\to\infty$, and δ will always satisfy $\delta<1/h$ for large $\lambda_0 n$. Thus, for $s=-\sigma_\lambda-\delta$, it follows from Equation (7) and Assumption 2 that

$$|L_h(t)|\le\frac{e^{-(\sigma_\lambda+\delta)t}}{2\pi}\int_{-\infty}^{\infty}\frac{|\widehat{K}((-\sigma_\lambda-\delta+i\omega)h)|}{|\widehat{\lambda}(\sigma_\lambda+\delta-i\omega)|}\mathrm{d}\omega\le\frac{e^{-(\sigma_\lambda+\delta)t}}{2\pi\lambda_0k_0}\int_{-\infty}^{\infty}|\widehat{K}((-\sigma_\lambda-\delta+i\omega)h)|(\delta^2+\omega^2)^{\gamma/2}\mathrm{d}\omega.$$

We continue bounding this quantity:

$$\begin{aligned}|L_h(t)|&\le\frac{c_2e^{-(\sigma_\lambda+\delta)t}}{\lambda_0}\left[\delta^\gamma\int_{|\omega|\le\delta}|\widehat{K}((-\sigma_\lambda-\delta+i\omega)h)|\mathrm{d}\omega+\int_{|\omega|>\delta}|\widehat{K}((-\sigma_\lambda-\delta+i\omega)h)|\,|\omega|^\gamma\mathrm{d}\omega\right]\\&\le\frac{c_3e^{-(\sigma_\lambda+\delta)t}}{\lambda_0}\left[\delta^{\gamma+1}e^{(\sigma_\lambda+\delta)h}+h^{-\gamma-1}\int_{|\xi|>\delta h}|\widehat{K}(-(\sigma_\lambda+\delta)h+i\xi)|\,|\xi|^\gamma\mathrm{d}\xi\right]\\&\le c_4e^{-(\sigma_\lambda+\delta)(t-h)}(\lambda_0h^{\gamma+1})^{-1},\end{aligned} \tag{24}$$

where in the first inequality above we use that $(\delta^2+\omega^2)^{\gamma/2}\le 2^{(\frac{\gamma}{2}-1)_+}(\delta^\gamma+\omega^\gamma)$ and the last inequality follows from Equations (22) and (23) and $r>\gamma+1$.

Because $G\in\mathscr{H}_{\beta,x_0}(A,M)$, it holds that $H(t)\le M\lambda_0e^{\sigma_\lambda t}$ when $\sigma_\lambda>0$ and that $H(t)\le M\lambda_0(a_1+a_2t^p)$ when $\sigma_\lambda=0$ (compare with Equation (1)). Combining this with (24), we obtain

$$\int_0^\infty|L_h(t-x_0)|H(t)\mathrm{d}t\le c_5\varkappa_1(\delta)Me^{(\sigma_\lambda+\delta)(x_0+h)}h^{-\gamma-1}<\infty,$$

where we put

$$\varkappa_1(\delta):=\begin{cases}\delta^{-1}, & \sigma_\lambda>0,\\ a_1\delta^{-1}+a_2\delta^{-p-1}, & \sigma_\lambda=0.\end{cases}$$

Thus, the estimator $\tilde{G}_h(x_0)$ is well defined, and condition (8) of Lemma 1 holds.

2^0. Now we establish an upper bound on the bias of $\tilde{G}_h(x_0)$. By Lemma 1 and conditions (K1) and (K2), we have

$$\begin{aligned}\mathbb{E}_G\big[\tilde{G}_h(x_0)\big]&=1-\mathbb{E}_G\int_0^T L_h(t-x_0)X(t)\mathrm{d}t=1-\int_0^T L_h(t-x_0)H(t)\mathrm{d}t\\&=\int_{-\infty}^{\infty}\frac{1}{h}K\Big(\frac{x-x_0}{h}\Big)G(x)\mathrm{d}x+\int_T^\infty L_h(t-x_0)H(t)\mathrm{d}t.\end{aligned}$$

It follows from Taylor expansion of G and conditions (K1) and (K2) that

$$\left|\int_{-\infty}^{\infty}\frac{1}{h}K\Big(\frac{x-x_0}{h}\Big)G(x)\mathrm{d}x-G(x_0)\right|=\left|\int_0^1K(u)[G(x_0+uh)-G(x_0)]\mathrm{d}u\right|\le C_KAh^\beta.$$

Moreover, if $\sigma_\lambda > 0$, then in view of Equation (24),

$$\left|\int_T^\infty L_h(t-x_0)H(t)\mathrm{d}t\right| \le c_1 M e^{(\sigma_\lambda+\delta)(x_0+h)} e^{-\delta T}\delta^{-1}h^{-\gamma-1}.$$

If $\sigma_\lambda = 0$, $p>0$, and $\delta T \ge 2p$, then, by Lemma 4,

$$\left|\int_T^\infty L_h(t-x_0)H(t)\mathrm{d}t\right| \le c_2 M e^{\delta(x_0+h)} e^{-\delta T} h^{-\gamma-1}(a_1\delta^{-1}+a_2\delta^{-2}T^{p-1}).$$

Thus,

$$\left|\int_T^\infty L_h(t-x_0)H(t)\mathrm{d}t\right| \le c_3 \varkappa_2(\delta,T) M e^{(\sigma_\lambda+\delta)(x_0+h)} e^{-\delta T} h^{-\gamma-1},$$

where

$$\varkappa_2(\delta,T) := \begin{cases} \delta^{-1} & \sigma_\lambda > 0 \ \text{ and } \ \sigma_\lambda = 0, p=0, \\ a_1\delta^{-1}+a_2\delta^{-2}T^{p-1}, & \sigma_\lambda = 0, p>0. \end{cases}$$

Combining these inequalities, we obtain

$$\left|\mathbb{E}_G\left[\tilde{G}_h(x_0)-G(x_0)\right]\right| \le c_4\left[Ah^\beta + \varkappa_2(\delta,T) M e^{(\sigma_\lambda+\delta)(x_0+h)} e^{-\delta T} h^{-\gamma-1}\right]. \tag{25}$$

3^0. The next step is to bound the variance of $\tilde{G}_h(x_0)$. Let $R(t,\tau) := \mathrm{cov}_G[X(t),X(\tau)]$; then

$$\begin{aligned}\mathrm{var}_G\left[\tilde{G}_h(x_0)\right] &= \frac{1}{n}\int_0^T\int_0^T L_h(t-x_0)\overline{L_h(\tau-x_0)}R(t,\tau)\mathrm{d}t\mathrm{d}\tau \le \frac{1}{n}\int_0^T\int_0^T |L_h(t-x_0)|^2 R(t,\tau)\mathrm{d}t\mathrm{d}\tau \\ &\le \frac{1}{n}\int_0^\infty |L_h(t-x_0)|^2\int_0^\infty R(t,\tau)\mathrm{d}\tau\mathrm{d}t, \end{aligned} \tag{26}$$

where the first inequality is owing to Cauchy–Schwarz and symmetry of the covariance function in its arguments. Our current goal is to bound the expression on the right-hand side of Equation (26) from above. We consider the cases $\sigma_\lambda > 0$ and $\sigma_\lambda = 0$ separately.

(a) If $\sigma_\lambda > 0$, then using Lemma 3(a) with $\nu = \sigma_\lambda + 2\delta$, $\delta > 0$, we obtain

$$\mathrm{var}_G\left[\tilde{G}_h(x_0)\right] \le \frac{\lambda_0 M}{n(\sigma_\lambda+\delta)}\int_0^\infty |L_h(t-x_0)|^2 e^{2(\sigma_\lambda+\delta)t}\mathrm{d}t \le \frac{\lambda_0 M e^{2(\sigma_\lambda+\delta)x_0}}{4\pi^2 n(\sigma_\lambda+\delta)}\int_{-\infty}^\infty \left|\frac{\widehat{K}(-(\sigma_\lambda+\delta)h+i\omega h)}{\widehat{\lambda}(\sigma_\lambda+\delta-i\omega)}\right|^2\mathrm{d}\omega,$$

where in the last line we have used Parseval's identity. By Assumption 2 and in view of Equations (22) and (23),

$$\begin{aligned}\int_{-\infty}^\infty \left|\frac{\widehat{K}(-(\sigma_\lambda+\delta)h+i\omega h)}{\widehat{\lambda}(\sigma_\lambda+\delta-i\omega)}\right|^2\mathrm{d}\omega &\le \frac{1}{\lambda_0^2 k_0^2}\int_{-\infty}^\infty |\widehat{K}(-\sigma h+i\omega h)|^2\left[\delta^2+\omega^2\right]^\gamma\mathrm{d}\omega \\ &\le c_1\lambda_0^{-2}e^{2\sigma h}\left[\delta^{2\gamma}h^{-1}+h^{-2\gamma-1}\right] \le c_2\lambda_0^{-2}e^{2\sigma h}h^{-2\gamma-1},\end{aligned}$$

where we have taken into account that $\delta < 1/h$. This leads to the following bound on the variance:

$$\mathrm{var}_G\left[\tilde{G}_h(x_0)\right] \le c_3(\sigma_\lambda+\delta)^{-1} M e^{2(\sigma_\lambda+\delta)(x_0+h)}(\lambda_0 n h^{2\gamma+1})^{-1}. \tag{27}$$

(b) Now consider the case $\sigma_\lambda = 0$. In this case, Equation (26) and Lemma 3 show that

$$\mathrm{var}_G\left[\tilde{G}_h(x_0)\right] \le \frac{c_4\lambda_0 M}{n}\int_0^\infty |L_h(t-x_0)|^2(a_1+a_2t^p)\mathrm{d}t, \tag{28}$$

and our current goal is to bound the integral on the right-hand side.

It follows from Assumption 1(b) that $\widehat{\lambda}(z)$ is bounded in its region of analyticity; hence, the limit $\lim_{\sigma\downarrow 0}\widehat{\lambda}(\sigma+i\omega) = \widehat{\lambda}(i\omega)$ exists for almost all ω (see, e.g., chapter 19 in Hille 1962). Moreover, Assumption 2 implies that $\widehat{\lambda}(z)$ does not vanish on the imaginary axis. Therefore, $|\widehat{\lambda}(\sigma+i\omega)|^{-1}$ is finite for all $\sigma \ge 0$ and $\omega \in \mathbb{R}$, and function $\widehat{K}(zh)/\widehat{\lambda}(-z)$

can be analytically continued to the imaginary axis. Thus, the integral in the definition (7) of L_h can be computed over the line $\mathrm{Re}(z)=0$, and by Parseval's identity and Assumption 2, we obtain

$$\int_{-\infty}^{\infty}|L_h(t-x_0)|^2\mathrm{d}t=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\left|\frac{\widehat{K}(i\omega h)}{\widehat{\lambda}(-i\omega)}\right|^2\mathrm{d}\omega\le\frac{1}{4\pi^2}\int_{-\infty}^{\infty}|\widehat{K}(i\omega h)|^2|\omega|^{2\gamma}\mathrm{d}\omega\le c_5h^{-2\gamma-1},$$

where the last inequality holds in view of Equations (22) and (23) and $r>\gamma+1$. Furthermore, for any real $\delta>0$,

$$\begin{aligned}\int_0^{\infty}|L_h(t-x_0)|^2t^p\mathrm{d}t&=\frac{1}{4\pi^2}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\widehat{K}((-\delta+i\omega)h)}{\widehat{\lambda}(\delta-i\omega)}\frac{\overline{\widehat{K}((-\delta+i\nu)h)}}{\widehat{\lambda}(\delta-i\nu)}e^{-2\delta(t-x_0)+i(\omega-\nu)(t-x_0)}t^p\mathrm{d}\omega\mathrm{d}\nu\mathrm{d}t\\&=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\widehat{K}((-\delta+i\omega)h)}{\widehat{\lambda}(\delta-i\omega)}\frac{\overline{\widehat{K}((-\delta+i\nu)h)}}{\widehat{\lambda}(\delta-i\nu)}e^{2\delta x_0-i(\omega-\nu)x_0}\frac{\Gamma(p+1)}{(2\delta+i(\nu-\omega))^{p+1}}\mathrm{d}\omega\mathrm{d}\nu\\&\le\frac{\Gamma(p+1)e^{2\delta x_0}}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\frac{\widehat{K}((-\delta+i\omega)h)}{\widehat{\lambda}(\delta-i\omega)}\right|^2\left[4\delta^2+(\omega-\nu)^2\right]^{-\frac{1}{2}(p+1)}\mathrm{d}\omega\mathrm{d}\nu\\&\le c_6e^{2\delta x_0}\delta^{-p}\int_{-\infty}^{\infty}\left|\frac{\widehat{K}((-\delta+i\omega)h)}{\widehat{\lambda}(\delta-i\omega)}\right|^2\mathrm{d}\omega\le c_7\lambda_0^{-2}\delta^{-p}e^{2\delta(x_0+h)}h^{-2\gamma-1},\end{aligned}$$

which together with Equation (28) yields

$$\mathrm{var}_G[\tilde{G}_h(x_0)]\le c_8(a_1+a_2\delta^{-p})Me^{2\delta(x_0+h)}(\lambda_0nh^{2\gamma+1})^{-1}.\tag{29}$$

Combining Equations (27) and (29), we obtain the final bound on the variance that holds for any $\sigma_\lambda\ge0$:

$$\mathrm{var}_G[\tilde{G}_h(x_0)]\le c_9\varkappa_3(\sigma_\lambda,\delta)Me^{2(\sigma_\lambda+\delta)(x_0+h)}\left(\lambda_0nh^{2\gamma+1}\right)^{-1},\tag{30}$$

where

$$\varkappa_3(\sigma_\lambda,\delta):=\begin{cases}(\sigma_\lambda+\delta)^{-1}, & \sigma_\lambda>0,\\ a_1+a_2\delta^{-p}, & \sigma_\lambda=0.\end{cases}$$

4^0. Now we are in a position to finish the proof of Equation (12). For this purpose, we specify the choice of parameters h and δ and substitute the corresponding expressions in the derived bounds on bias (25) and variance (30).

(a) Consider first the case $\sigma_\lambda>0$. Here we put

$$h=h_*=\left(\frac{Me^{2\sigma_\lambda x_0}}{A^2\sigma_\lambda\lambda_0n}\right)^{1/(2\beta+2\gamma+1)},$$

and

$$\delta=\delta_*:=\frac{c_1}{T-x_0}\ln\left\{M^{\frac{\beta+\gamma}{2\beta+2\gamma+1}}(Ae^{-\sigma_\lambda x_0})^{\frac{1}{2\beta+2\gamma+1}}(\sigma_\lambda\lambda_0n)^{\frac{\beta+\gamma+1}{2\beta+2\gamma+1}}\right\},$$

with some appropriately large absolute constant c_1. Note that $\delta_*\to0$ as $\lambda_0n\to\infty$ because $\frac{1}{T}\ln(\lambda_0n)\to0$ as $\lambda_0n\to\infty$. With the choice of $h=h_*$ and $\delta=\delta_*$, it is straightforward to verify in Equation (25) that the bias is bounded above by a multiple of Ah_*^{β}. Then, substituting the expressions for h_* and δ_* in Equations (25) and (27) and taking the limit as $\lambda_0n\to\infty$, we obtain Equation (12) for the case $\sigma_\lambda>0$.

(b) Now let $\sigma_\lambda=0$, and assume first that $p=0$ (i.e., the intensity function $\lambda(t)$ is bounded). Here we choose

$$h=h_*=\left(\frac{M}{A^2\lambda_0n}\right)^{1/(2\beta+2\gamma+1)}$$

and

$$\delta=\delta_*=\frac{c_2}{T-x_0}\ln\left\{M^{\frac{\beta+\gamma}{2\beta+2\gamma+1}}A^{\frac{1}{2\beta+2\gamma+1}}(\lambda_0n)^{\frac{\beta+\gamma+1}{2\beta+2\gamma+1}}\right\},$$

with appropriate constant c_2. Substituting these values in Equations (25) and (29), we obtain Equation (12).

If $p>0$, then we set

$$h = h_* := \left[\frac{M(a_1 + a_2 x_0^p)}{A^2 \lambda_0 n}\right]^{1/(2\beta+2\gamma+1)},$$

and $\delta = \delta_* := \frac{p}{2x_0}$. It is easily checked that, under this choice, Equation (12) holds with

$$\varphi_n = A^{(2\gamma+1)/(2\beta+2\gamma+1)}\left[\frac{M(a_1 + a_2 x_0^p)}{\lambda_0 n}\right]^{\beta/(2\beta+2\gamma+1)}. \quad \square$$

7.4. Proof of Theorems 3 and 4

We provide a unified proof of Theorems 3 and 4. Each step of the proof considers two cases: $\sigma_\lambda > 0$ (corresponds to the proof of Theorem 3) and $\sigma_\lambda = 0$ (corresponds to the proof of Theorem 4).

Throughout the proof, $c_1, c_2, \ldots$ stand for constants depending on γ, k_0, and p only. In particular, they do not depend on M, λ_0, σ_λ, n, and T. To avoid additional technicalities in what follows, we assume that γ is an integer.

1^0. We bound the bias of $\tilde{\mu}_G$. By Lemma 2,

$$\mathbb{E}_G[\tilde{\mu}_G] = \int_0^T J_b(t)H(t)\mathrm{d}t = \int_0^\infty \psi_b(x)\bar{G}(x)\mathrm{d}x - \int_T^\infty J_b(t)H(t)\mathrm{d}t.$$

Therefore,

$$\left|\mathbb{E}_G[\tilde{\mu}_G] - \mu_G\right| \le \int_b^\infty \bar{G}(x)\mathrm{d}x + \int_T^\infty |J_b(t)|H(t)\mathrm{d}t. \tag{31}$$

Because $G \in \mathscr{G}(M)$,

$$\int_b^\infty \bar{G}(x)\mathrm{d}x \le \int_b^\infty \frac{x}{b}\bar{G}(x)\mathrm{d}x \le Mb^{-1}. \tag{32}$$

Our current goal is to bound the second integral on the right-hand side of Equation (31).

Let $s = -\sigma$, where $\sigma := \sigma_\lambda + \delta$ with some parameter $\delta > 0$ to be specified. By definition of $J_b(t)$ and by Assumption 2,

$$\begin{aligned}|J_b(t)| &\le \frac{e^{-\sigma t}}{2\pi\lambda_0 k_0}\int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)|\left[(\sigma - \sigma_\lambda)^2 + \omega^2\right]^{\gamma/2}\mathrm{d}\omega \\ &\le c_1 e^{-\sigma t}\lambda_0^{-1}\left\{\delta^\gamma \int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)|\mathrm{d}\omega + \int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)| \cdot |\omega|^\gamma \mathrm{d}\omega\right\}.\end{aligned}$$

Repeatedly integrating by parts the integral $\int_{-1/4}^{b+1/4}\psi_b(t)e^{(\sigma - i\omega)t}\mathrm{d}t$, we obtain for any $r = 1, 2, \ldots$ that

$$|\widehat{\psi}_b(-\sigma + i\omega)| \le c_2(r)e^{\sigma(b+1/4)}(\sigma^2 + \omega^2)^{-r/2},$$

where constant $c_2(r)$ depends on r. This inequality with $r > 1$ yields

$$\int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)|\mathrm{d}\omega \le c_3(r)e^{\sigma(b+1/4)}\sigma^{-r+1},$$

and with $r > \gamma + 1$,

$$\int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)| \cdot |\omega|^\gamma \mathrm{d}\omega \le c_4(r)e^{\sigma(b+1/4)}\sigma^{-r+\gamma+1}.$$

Therefore, the following bound holds:

$$|J_b(t)| \le c_5(r)\lambda_0^{-1}e^{-\sigma(t-b-1/4)}\left\{\delta^\gamma\sigma^{-r+1} + \sigma^{-r+\gamma+1}\right\},$$

where $r > \gamma + 1$.

(a) First, consider the case $\sigma_\lambda > 0$. By Assumption 1(a), $H(t) \le M\lambda_0 e^{\sigma_\lambda t}$; hence,

$$\int_T^\infty |J_b(t)| H(t)\mathrm{d}t \le \frac{c_5(r)M}{\sigma^{r-1}\delta}(\delta^\gamma + \sigma^\gamma)e^{\sigma(b+1/4)}e^{-\delta T}. \tag{33}$$

Combining Equations (33), (32), and (31), we obtain

$$\left|\mathbb{E}_G[\tilde{\mu}_G] - \mu_G\right| \le Mb^{-1} + c_5(r)M\delta^{-1}(\sigma_\lambda + \delta)^{-r+\gamma+1}e^{(\sigma_\lambda+\delta)(b+1/4)}e^{-\delta T}.$$

In particular, if we set $r = \gamma + 2$ and if

$$T - b - \tfrac{1}{4} \ge \frac{c_6}{\delta}\ln\left(\frac{be^{\sigma_\lambda(b+1/4)}}{\sigma_\lambda}\right), \tag{34}$$

then

$$\left|\mathbb{E}_G[\hat{\mu}_G] - \mu_G\right| \le c_7 M b^{-1}. \tag{35}$$

(b) Now let $\sigma_\lambda = 0$. In view of Assumption 1(b), $H(t) \le M\lambda_0(a_1 + a_2 t^p)$ so that

$$\int_T^\infty |J_b(t)|H(t)\mathrm{d}t \le c_5(r)\delta^{\gamma-r+1}Me^{\delta(b+1/4)}\left[a_1\delta^{-1}e^{-\delta T} + a_2\int_T^\infty t^p e^{-\delta t}\mathrm{d}t\right].$$

If $\delta T \ge 2p$, then using Lemma 4, we obtain

$$\int_T^\infty |J_b(t)|H(t)\mathrm{d}t \le c_8(r)\delta^{\gamma-r}Me^{-\delta(T-b-1/4)}\left(a_1 + a_2\delta^{-1}T^{p-1}\right). \tag{36}$$

It follows from Equation (36) that if

$$T - b - \frac{1}{4} \ge \frac{c_9}{\delta}\ln\left\{b\delta^{\gamma-r}(a_1 + a_2\delta^{-1}T^{p-1})\right\},\quad T \ge \frac{2p}{\delta}, \tag{37}$$

then again Equation (35) holds.

We conclude that the bound (35) on the bias holds, provided that b and δ are chosen to satisfy Equation (34) in the case $\sigma_\lambda > 0$ and Equation (37) in the case $\sigma_\lambda = 0$.

2^0. Now we proceed with bounding the variance of $\tilde{\mu}_G$. We have

$$\mathrm{var}_G[\tilde{\mu}_G] = \frac{1}{n}\int_0^T\int_0^T J_b(t)\overline{J_b(\tau)}R(t,\tau)\mathrm{d}t\mathrm{d}\tau \le \frac{1}{n}\int_0^\infty |J_b(t)|^2\int_0^\infty R(t,\tau)\mathrm{d}\tau\mathrm{d}t, \tag{38}$$

where $R(t,\tau) = \mathrm{cov}_G[X(t), X(\tau)]$ is given in Equation (19).

(a) Let $\sigma_\lambda > 0$; then, applying Lemma 3(a) with $\nu = \sigma_\lambda + 2\delta$, we have

$$\mathrm{var}_G[\tilde{\mu}_G] \le \frac{M\lambda_0}{n(\sigma_\lambda+\delta)}\int_0^\infty |J_b(t)|^2 e^{2(\sigma_\lambda+\delta)}\mathrm{d}t \le \frac{M\lambda_0}{4\pi^2 n(\sigma_\lambda+\delta)}\int_{-\infty}^\infty \left|\frac{\widehat{\psi}_b(-\sigma_\lambda - \delta + i\omega)}{\widehat{\lambda}(\sigma_\lambda + \delta - i\omega)}\right|^2\mathrm{d}\omega. \tag{39}$$

We recall that by Assumption 2 and for $\sigma = \sigma_\lambda + \delta$, $\delta > 0$, one has

$$\int_{-\infty}^\infty \left|\frac{\widehat{\psi}_b(-\sigma + i\omega)}{\widehat{\lambda}(\sigma - i\omega)}\right|^2\mathrm{d}\omega \le \frac{1}{\lambda_0^2 k_0^2}\int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)|^2[\delta^2 + \omega^2]^\gamma\mathrm{d}\omega,$$

and the integral on the right-hand side is bounded as follows. First, by Parseval's identity,

$$\int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)|^2\mathrm{d}\omega = 2\pi\int_{-1/4}^{b+1/4} e^{2\sigma t}\psi_b^2(t)\mathrm{d}t \le \pi e^{2\sigma(b+1/4)}\sigma^{-1}. \tag{40}$$

Second, for integer $\gamma \ge 0$, we have

$$\begin{aligned}\int_{-\infty}^\infty |\widehat{\psi}_b(-\sigma + i\omega)|^2|\omega|^{2\gamma}\mathrm{d}\omega &= 2\pi\int_{-1/4}^{b+1/4}\left|\frac{\mathrm{d}^\gamma}{\mathrm{d}t^\gamma}[e^{\sigma t}\psi_b(t)]\right|^2\mathrm{d}t \\ &= 2\pi\int_{-1/4}^{b+1/4}\left|\sum_{j=0}^\gamma \binom{\gamma}{j}\sigma^j e^{\sigma t}\psi_b^{(\gamma-j)}(t)\right|^2\mathrm{d}t \le c_1\sum_{j=0}^\gamma \sigma^{2j}\int_{-1/4}^{b+1/4} e^{2\sigma t}|\psi_b^{(\gamma-j)}(t)|^2\mathrm{d}t.\end{aligned}$$

By construction, derivatives of ψ_b are nonzero only on the intervals $[-1/4, 0]$ and $[b, b+1/4]$; therefore,

$$\int_{-\infty}^{\infty} |\widehat{\psi}_b(-\sigma+i\omega)|^2|\omega|^{2\gamma}\mathrm{d}\omega \le c_2 e^{2\sigma(b+1/4)}\left[\sigma^{2\gamma}(b+\tfrac{1}{2}) + \sum_{j=0}^{\gamma-1}\sigma^{2j}\right]. \tag{41}$$

Combining Equations (40) and (41), we obtain

$$\int_{-\infty}^{\infty}\left|\frac{\widehat{\psi}_b(-\sigma+i\omega)}{\widehat{\lambda}(\sigma-i\omega)}\right|^2\mathrm{d}\omega \le c_3\lambda_0^{-2}e^{2\sigma(b+1/4)}\left(\delta^{2\gamma}\sigma^{-1} + \sigma^{2\gamma}(b+\tfrac{1}{2}) + \sum_{j=0}^{\gamma-1}\sigma^{2j}\right),$$

which along with Equation (39) yields

$$\mathrm{var}_G[\tilde{\mu}_G] \le \frac{c_4 M e^{2(\sigma_\lambda+\delta)(b+1/4)}}{\lambda_0 n(\sigma_\lambda+\delta)}\left[\frac{\delta^{2\gamma}}{\sigma_\lambda+\delta} + (\sigma_\lambda+\delta)^{2\gamma}(b+\tfrac{1}{2}) + \sum_{j=0}^{\gamma-1}(\sigma_\lambda+\delta)^{2j}\right].$$

We set

$$\delta = \delta_* := \frac{1}{b};$$

also, b will be always chosen so that $b \ge \sigma_\lambda^{-1}$ for large $\lambda_0 n$. Under these conditions, we finally obtain

$$\mathrm{var}_G[\tilde{\mu}_G] \le c_5 M\sigma_\lambda^{2\gamma-1} b e^{2\sigma_\lambda(b+1/4)}(\lambda_0 n)^{-1}.$$

We note also that with the choice $\delta = \delta_*$, the bound on the bias (35) holds, provided that

$$T \ge c_6[b + b\sigma_\lambda(b+1/4) + b\ln(b/\sigma_\lambda)];$$

see Equation (34).

(b) Now consider the case $\sigma_\lambda = 0$. It follows from Equation (38) and Lemma 3 that

$$\mathrm{var}_G[\tilde{\mu}_G] \le \frac{c_7 M\lambda_0}{n}\int_0^T |J_b(t)|^2(a_1 + a_2 t^p)\mathrm{d}t.$$

Because $|\widehat{\lambda}(\sigma+i\omega)|^{-1}$ is finite for all $\sigma \ge 0$ and $\omega \in \mathbb{R}$, function $\widehat{\psi}_b(z)/\widehat{\lambda}(-z)$ can be analytically continued to the imaginary axis. Therefore, by Parseval's identity and Assumption 2, we obtain

$$\int_{-\infty}^{\infty}|J_b(t)|^2\mathrm{d}t = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\left|\frac{\widehat{\psi}_b(i\omega)}{\widehat{\lambda}(-i\omega)}\right|^2\mathrm{d}\omega \le \frac{1}{4\pi^2\lambda_0^2 k_0^2}\int_{-\infty}^{\infty}|\widehat{\psi}_b(i\omega)|^2|\omega|^{2\gamma}\mathrm{d}\omega \le c_8\lambda_0^{-2}\int_{-1/4}^{b+1/4}|\psi_b^{(\gamma)}(t)|^2\mathrm{d}t \le c_9\lambda_0^{-2},$$

and for $\delta > 0$,

$$\begin{aligned}
\int_0^{\infty}|J_b(t)|^2 t^{p+1}\mathrm{d}t &= \frac{1}{4\pi^2}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\widehat{\psi}_b(-\delta+i\omega)}{\widehat{\lambda}(\delta-i\omega)}\frac{\overline{\widehat{\psi}_b(-\delta+i\nu)}}{\widehat{\lambda}(\delta-i\nu)}e^{-2\delta t+i(\omega-\nu)t}t^p\mathrm{d}\omega\mathrm{d}\nu\mathrm{d}t \\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\widehat{\psi}_b(-\delta+i\omega)}{\widehat{\lambda}(\delta-i\omega)}\frac{\overline{\widehat{\psi}_b(-\delta+i\nu)}}{\widehat{\lambda}(\delta-i\nu)}\frac{\Gamma(p+1)}{(2\delta+i(\nu-\omega))^{p+1}}\mathrm{d}\omega\mathrm{d}\nu \\
&\le \frac{\Gamma(p+1)}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\frac{\widehat{\psi}_b(-\delta+i\omega)}{\widehat{\lambda}(\delta-i\omega)}\right|^2\left[4\delta^2+(\omega-\nu)^2\right]^{-\frac{1}{2}(p+1)}\mathrm{d}\omega\mathrm{d}\nu \\
&\le \frac{c_{10}}{\delta^p}\int_{-\infty}^{\infty}\left|\frac{\widehat{\psi}_b(-\delta+i\omega)}{\widehat{\lambda}(\delta-i\omega)}\right|^2\mathrm{d}\omega.
\end{aligned}$$

Therefore,

$$\mathrm{var}_G[\tilde{\mu}_G] \le c_{11}\frac{M\lambda_0}{n}(a_1 + a_2\delta^{-p})\int_{-\infty}^{\infty}\left|\frac{\widehat{\psi}_b(-\delta+i\omega)}{\widehat{\lambda}(\delta-i\omega)}\right|^2\mathrm{d}\omega.$$

Using the same reasoning as before,

$$\int_{-\infty}^{\infty}\left|\frac{\widehat{\psi}_b(-\delta+i\omega)}{\widehat{\lambda}(\delta-i\omega)}\right|^2 \mathrm{d}\omega \le c_{12}\lambda_0^{-2}e^{2\delta(b+1/4)}\left[\delta^{2\gamma}(b+\tfrac{1}{2})+\sum_{j=0}^{\gamma-1}\delta^{2j}+\delta^{2\gamma-1}\right].$$

We again set $\delta=\delta_* := b^{-1}$; with this choice,

$$\mathrm{var}_G[\tilde{\mu}_G] \le c_{13}M(a_1+a_2b^p)(\lambda_0 n)^{-1},$$

and the bound on the bias is given in Equation (35), provided that

$$T \ge c_{14}\left[b+\frac{1}{4}+b\ln b+b\ln(a_1+a_2bT^{p-1})\right], \quad T \ge 2p\ln b;$$

see Equation (37).

3^0. Now we complete the proof by selecting b to balance the obtained bounds on the bias and variance.
(a) Consider the case $\sigma_\lambda > 0$. Let

$$b=b_*=\frac{1}{2\sigma_\lambda}\left\{\ln\left(\frac{M\lambda_0 n}{\sigma_\lambda^{2\gamma-1}}\right)-3\ln\left[\frac{1}{2\sigma_\lambda}\ln\left(\frac{M\lambda_0 n}{\sigma_\lambda^{2\gamma-1}}\right)\right]\right\}-\frac{1}{4}.$$

We note that because $[\ln(\lambda_0 n)]^2/T \to 0$ as $\lambda_0 n \to \infty$, Equation (34) is fulfilled for large $\lambda_0 n$ with $b=b_*$ and $\delta=\delta_*=1/b_*$. Therefore, substituting the expression for b_* in the bounds for the bias and variance, we obtain

$$\limsup_{N\to\infty}\{\varphi_N^{-1}\mathcal{R}[\tilde{\mu}_G;\mathcal{G}(M)]\} \le C_1, \quad \varphi_N := M\sigma_\lambda\left[\ln\left(\frac{M\lambda_0 n}{\sigma_\lambda^{2\gamma-1}}\right)\right]^{-1}.$$

(b) Now let $\sigma_\lambda = 0$, and assume that $a_2 = 0$. In this case, the bound on the variance is $\mathrm{var}_G[\tilde{\mu}_G] \le c_1 M(\lambda_0 n)^{-1}$ and on the bias Mb^{-1}. If we choose $b \ge b_* \ge c_2\sqrt{M\lambda_0 n}$ with appropriate absolute constant c_2, then under condition $\sqrt{\lambda_0 n}\ln(\lambda_0 n)/T \to 0$ as $\lambda_0 n \to \infty$, we obtain that

$$\limsup_{N\to\infty}\left\{\sqrt{\frac{\lambda_0 n}{M}}\mathcal{R}[\tilde{\mu}_G;\mathcal{G}(M)]\right\} \le C_2.$$

If $a_2 > 0$, then we set $b=b_*=[M\lambda_0 n]^{1/(p+2)}$. If $\frac{1}{T}(\lambda_0 n)^{1/(p+2)}\ln(\lambda_0 n) \to 0$ as $\lambda_0 n \to \infty$, then Equation (37) is fulfilled for large $\lambda_0 n$, and

$$\varphi_N := M^{(p+1)/(p+2)}(\lambda_0 n)^{-1/(p+2)}. \quad \square$$

References

Abramovich F, Pensky M, Rozenholc Y (2013) Laplace deconvolution with noisy observations. *Electronic J. Statist.* 7:1094–1128.
Antunes N, Pipiras V, Veitch D (2017) Skampling for the flow duration distribution. Bolla R, Ciucu F, eds. *Proc. 29th Internat. Teletraffic Congress (ITC 29), Genoa, Italy*, vol. 1, 63–71.
Baccelli F, Kauffman B, Veitch D (2009) Inverse problems in queueing theory and internet probing. *Queueing Systems* 63(1–4):59–107.
Bigot J, Gadat S, Klein T, Marteau C (2013) Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Electronic J. Statist.* 7:881–931.
Bingham NH, Pitts SM (1999) Non–parametric estimation for the $M/G/\infty$ queue. *Ann. Inst. Statist. Math.* 51(1):71–97.
Borwein JM, O-Yeat C (2009) Uniform bounds for the complementary incomplete gamma function. *Math. Inequaities Appl.* 12(1):115–121.
Brown M (1970) An $M/G/\infty$ estimation problem. *Ann. Math. Statist.* 41(2):651–654.
Coifman B, Cassidy M (2002) Vehicle reidentification and travel time measurement on congested freeways. *Transportation Res. Part A: Policy Practice* 36(10):899–917.
Dobrzyński M, Bruggeman FJ (2009) Elongation dynamics shape bursty transcription and translation. *Proc. Natl. Acad. Sci. USA* 106(8): 2583–2588.
Doetsch G (1974) *Introduction to the Theory and Application of the Laplace Transformation* (Springer-Verlag, Berlin).
Durbin F (1973) Numerical inversion of Laplace transforms: An efficient improvement to dubner and abate's method. *Comput. J.* 17(4):371–376.
Eick S, Massey W, Whitt W (1993) The physics of the $M_t/G/\infty$ queue. *Management Sci.* 39(2):241–252.
Fan J (1991) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* 19(3):1257–1272.
Goldenshluger A (2016) Nonparametric estimation of the service time distribution in the $M/G/\infty$ queue. *Adv. Appl. Probab.* 48(4):1117–1138.
Goldenshluger A (2018) The $M/G/\infty$ estimation problem revisited. *Bernoulli* 24(4A):2531–2568.

Goldenshluger A, Nemirovski A (1997) On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.* 6(2):135–170.
Hille E (1962) *Analytic Function Theory*, vol. II (American Mathematical Society Chelsea Publishing, Providence, RI).
Jansen HM, Mandjes MRH, De Turck K, Wittevrongel S (2016) A large deviations principle for infinite-server queues in a random environment. *Queueing Systems* 82(1):199–235.
Krishnamurthy V, Elliot R (1997) Filters for estimating Markov modulated Poisson processes and image-based tracking. *Automatica* 33(5): 821–833.
Kutoyants YA (1998) *Statistical Inference for Spatial Poisson Processes*, Lecture Notes in Statistics, vol. 134 (Springer-Verlag, New York).
Lepskiĭ OV (1990) A problem of adaptive estimation in Gaussian white noise [Russian]. *Theory Probab. Appl*. 35(3): 454–466.
Schweer S, Wichelhaus C (2015) Nonparametric estimation of the service time distribution in the discrete-time $GI/G/\infty$ queue with partial information. *Stochastic Processes Appl.* 125(1):233–253.
Tsybakov AB (2009) *Introduction to Nonparametric Estimation*, Springer Series in Statistics (Springer, New York).
Widder D (1946) *The Laplace Transform* (Princeton University Press, Princeton, NJ).
Zhang C-H (1990) Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* 18(2):806–831.