Convergence rates of Laplace-transform based estimators

ARNOUD V. DEN BOER[1,3] and MICHEL MANDJES[2,3]

[1]*University of Twente, Drienerlolaan 5, 7522 NB Enschede. E-mail: a.v.denboer@utwente.nl*
[2]*University of Amsterdam, Science Park 904, 1098 XH Amsterdam. EURANDOM, P.O. Box 513, 5600 MB Eindhoven. E-mail: m.r.h.mandjes@uva.nl*
[3]*Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam*

This paper considers the problem of estimating probabilities of the form $\mathbb{P}(Y \leq w)$, for a given value of w, in the situation that a sample of i.i.d. observations $X_1, \ldots, X_n$ of X is available, and where we explicitly know a functional relation between the Laplace transforms of the non-negative random variables X and Y. A plug-in estimator is constructed by calculating the Laplace transform of the empirical distribution of the sample $X_1, \ldots, X_n$, applying the functional relation to it, and then (if possible) inverting the resulting Laplace transform and evaluating it in w. We show, under mild regularity conditions, that the resulting estimator is weakly consistent and has expected absolute estimation error $O(n^{-1/2} \log(n+1))$. We illustrate our results by two examples: in the first we estimate the distribution of the workload in an M/G/1 queue from observations of the input in fixed time intervals, and in the second we identify the distribution of the increments when observing a compound Poisson process at equidistant points in time (usually referred to as "decompounding").

Keywords: decompounding; estimation in queues; Laplace transform

1. Introduction

The estimation problem considered in this paper is the following. Suppose we have independent observations of the (nonnegative) random variable X, but we are interested in estimating the distribution of the (nonnegative) random variable Y. The crucial element in our set up is that we explicitly know the relation between the Laplace transforms of the random variables X and Y, that is, we have a mapping Ψ which maps Laplace transforms of random variables to complex-valued functions defined on the right-half complex plane, and which maps the Laplace transform of X to the Laplace transform of Y.

A straightforward estimation procedure could be the following. (i) Estimate the Laplace transform of X by its evident empirical estimator; denote this estimate by $\tilde{X}_n$; (ii) estimate the Laplace transform of Y by $\Psi\tilde{X}_n$; (iii) apply Laplace inversion on $\Psi\tilde{X}_n$, so as to obtain an estimate of the distribution of Y. To justify this procedure, there are several issues that need to be addressed. First, $\tilde{X}_n$ may not lie in the domain of the mapping Ψ, and second, $\Psi\tilde{X}_n$ may not be a Laplace transform, and thus not amenable for Laplace inversion.

The main contribution of this paper is that we specify a procedure in which the above caveats are addressed, leading to the result that the plug-in estimator described above converges, in probability, to the true value as n grows large. In addition we have bounds on its performance: the

expected absolute estimation error is $O(n^{-1/2} \log(n+1))$. Perhaps surprisingly, the techniques used primarily rely on an appropriate combination of standard proof techniques. Our result is valid under three mild regularity conditions: two of them are essentially of a technical nature, whereas the third can be seen as a specific continuity property that needs to be imposed on the mapping Ψ.

In this paper, two specific examples are treated in greater detail. In the first, an M/G/1 queueing system is considered: jobs of random size arrive according to a Poisson process with rate $\lambda > 0$, the job sizes are i.i.d. samples from a nonnegative random variable B, and the system is drained at unit rate. Suppose that we observe the amount of work arriving in intervals of fixed length, say $\delta > 0$; these observations are compound Poisson random variables, distributed as

$$X \stackrel{\mathrm{d}}{=} \sum_{i=1}^{N} B_i,$$

with N Poisson distributed with mean $\lambda\delta$, independent of the job sizes, and with $B_1, B_2, \ldots$ mutually independent and distributed as B. We show how our procedure can be used to estimate the distribution of the workload Y from the compound Poisson observations; the function Ψ follows from the Pollaczek–Khinchine formula. As we demonstrate, the regularity conditions mentioned above are met. In the second example, often referred to as "decompounding," the goal is to determine the job size distribution from compound Poisson observations.

Literature

Related work can be found in various branches of the literature. Without aiming at giving an exhaustive overview, we discuss some of the relevant papers here. The first branch consists of papers on estimating the probability distribution of a non-observed random variable by exploiting a given functional relation between the Laplace transforms of X and Y. The main difficulty that these papers circumvent is the issue of "ill-posedness": a sequence of functions $(f_n)_{n \in \mathbb{N}}$ may not converge to a function f, as n grows large, even if the corresponding Laplace transforms of f_n do converge to the Laplace transform of f. Remedies, based on "regularized Laplace inversion" have been proposed, in a compound Poisson context, by Shimizu [23] (including Gaussian terms as well) and Mnatsakanov *et al.* [20]; the rate of convergence is typically just $1/\log n$ in an appropriately chosen L_2-norm. Hansen and Pitts [17] use the Pollaczek–Khinthcine formula to construct estimators for the service-time distribution and its stationary excess distribution in an M/G/1 queue, and show that the estimated stationary excess distribution is asymptotically Normal.

Some related papers that use Fourier instead of Laplace inversion are [8,9,24] and [16]. Van Es *et al.* [24] estimate the density of B_i by inverting the empirical Fourier transform associated with a sample of X, and prove that this estimator is weakly consistent and asymptotically normal. Comte *et al.* [8] also estimate the density of B_i using the empirical Fourier transform of X, by exploiting an explicit relation derived by Duval [14] between the density of X and B_i. They show that this estimator achieves the optimal convergence rate in the minimax sense over Sobolev balls. Comte *et al.* [9] extend this to the case of mixed compound Poisson distributions (where

the intenstiy λ of the Poisson process is itself a random variable), and provide bounds on the L^2-norm of the density estimator. Finally, Hall and Park [16] estimate service-time characteristics from busy period data in an infinite-server queueing setting, and prove convergence rates in probability.

A second branch of research concerns methods that do not involve working with transforms and inversion. Buchmann and Grübel [4] develop a method for decompounding: in the case the underlying random variables have a discrete distribution by relying on the so-called Panjer recursion, and in the case of continuous random variables by expressing the distribution function of the summands B_i in terms of a series of alternating terms involving convolutions of the distribution of X. The main result of this paper concerns the asymptotic Normality of specific plug-in estimators. This method having the inherent difficulty that probabilities are not necessarily estimated by positive numbers, an advanced version (for the discrete case only) has been proposed by the same authors in [5]. This method has been further extended by Bøgsted and Pitts [3] to that of a general (but known) distribution for the number of terms N. Duval [14] estimates the probability density of B_i by exploiting an explicit relation between the densities of X and B_i, which however is only valid if $\lambda \delta < \log 2$. She shows that minimax optimal convergence rates are achieved in an asymptotic regime where the sampling rate δ converges to zero. The introduction of [3] gives a compact description of the state-of-the-art of this branch of the literature.

A third body of work concentrates on the specific domain of queueing models, and develops techniques to efficiently estimate large deviation probabilities. Bearing in mind that estimating small tail probabilities directly from the observations may be inherently slow and inefficient [15], techniques have been developed that exploit some structural understanding of the system. Assuming exponential decay in the exceedance level, the pioneering work of Courcoubetis *et al.* [10] provide (experimental) backing for an extrapolation technique. The approach proposed by Zeevi and Glynn [25] has provable convergence properties; importantly, their results are valid in great generality, in that they cover for example, exponentially decaying as well as Pareto-type tail probabilities. Mandjes and van de Meent [19] consider queues with Gaussian input; it is shown how to accurately estimate the characteristics of the input stream by just measuring the buffer occupancy; interestingly, and perhaps counter-intuitively, relatively crude periodic measurements are sufficient to estimate fine time-scale traffic characteristics.

As it is increasingly recognized that probing techniques may play a pivotal role when designing distributed control algorithms, there is a substantial number of research papers focusing on applications in communication networks. A few examples are the procedure of Baccelli *et al.* [2] that infers input characteristics from delay measurements, and the technique of Antunes and Pipiras [1] that estimates the interrenewal distribution based on probing information. This paper contributes to this line of research by showing how a Laplace-transform based estimator, using samples of the workload obtained by probing, can be used to estimate the workload in an M/G/1 queue; cf. Section 4.1 and 6.

Organization

The rest of this paper is organized as follows. In Section 2, we formally define our Laplace-transform based estimator, and Section 3 shows that the expected absolute estimation error is

$O(n^{-1/2} \log(n + 1))$. In Section 4.1, we apply this result to an estimation problem in queueing theory, and in Section 4.2 to a decompounding problem. Section 5 contains a number of auxiliary lemmas used to prove the main theorems in this paper. A numerical illustration is provided in Section 6.

Notation

We finish this introduction by introducing notation that is used throughout this paper. The real and imaginary part of a complex number $z \in \mathbb{C}$ are denoted by $\Re(z)$ and $\Im(z)$; we use the symbol i for the imaginary unit. We write $\mathbb{C}_+ := \{z \in \mathbb{C} | \Re(z) \geq 0\}$ and $\mathbb{C}_{++} := \{z \in \mathbb{C} | \Re(z) > 0\}$. For a function $f : [0, \infty) \to \mathbb{R}$, let $\bar{f}(s) = \int_0^\infty f(x)e^{-sx} \, dx$ denote the Laplace transform of f, defined for all $s \in \mathbb{C}$ where the integral is well-defined. For any nonnegative random variable X, let $\tilde{X}(s) := \mathbb{E}[\exp(-sX)]$ denote the Laplace transform of X, defined for all $s \in \mathbb{C}_+$. (Although $\tilde{X}(s)$ may be well-defined for s with $\Re(s) < 0$, we restrict ourselves without loss of generality to $\mathbb{C}_+$, which is contained in the domain of $\tilde{X}(s)$ for each nonnegative random variable X.) For $t \in (0, \infty)$, as usual, $\Gamma(t) := \int_0^\infty x^{t-1}e^{-x} \, dx$ denotes the Gamma function. The complement of an event A is written as A^c; the indicator function corresponding to A is given by $\mathbf{1}_A$.

2. Laplace-transform based estimator

In this section, we formally define our plug-in estimator. The setting is as sketched in the introduction: we have n i.i.d. observations $X_1, \ldots, X_n$ of the random variable X at our disposal, and we wish to estimate the distribution of Y, where we know a functional relation between the transforms of X and Y.

Let $\mathcal{X}$ be a collection of (single-dimensional) nonnegative random variables, and let the collection $\tilde{\mathcal{X}} = \{\tilde{X}(\cdot) | X \in \mathcal{X}\}$ represent their Laplace transforms. Let

$$\Psi : \tilde{\mathcal{X}} \to \{g : \mathbb{C}_+ \to \mathbb{C}\}$$

map each Laplace transform in $\tilde{\mathcal{X}}$ to a complex-valued function on $\mathbb{C}_+$. Finally, let Y be a nonnegative random variable such that $\tilde{Y}(s) = (\Psi\tilde{X})(s)$ for some unknown $X \in \mathcal{X}$ and all $s \in \mathbb{C}_+$, that is, Ψ maps the Laplace transform of X onto the Laplace transform of Y. We are interested in estimating the cumulative distribution function $F^Y(w)$ of Y at a given value $w > 0$, based on the sample $X_1, \ldots, X_n$. The distributions of both X and Y are assumed to be unknown, but the mapping Ψ *is* known (and will be exploited in our estimation procedure).

A natural approach to this estimation problem is to (i) estimate the Laplace transform of Y by $\Psi\tilde{X}_n$, where $\tilde{X}_n$ is the "naïve" estimator

$$\tilde{X}_n(s) = \frac{1}{n} \sum_{i=1}^n \exp(-sX_i), \qquad (s \in \mathbb{C}_+);$$

observe that $\tilde{X}_n$ can be interpreted as the Laplace transform of the empirical distribution of the sample $X_1, \ldots, X_n$; then (ii) estimate the Laplace transform corresponding to the distribution

function F^Y by $s \mapsto s^{-1}(\Psi \tilde{X}_n)(s)$; and finally (iii) apply an inversion formula for Laplace transforms and evaluate the resulting expression in w. Note that in step (ii) we relied on the standard identity

$$\int_0^\infty e^{-sw} F^Y(w) \, dw = \frac{\mathbb{E}[e^{-sY}]}{s}.$$

There are two caveats, however, with this approach: first, the transform $\tilde{X}_n$ is not necessarily an element of $\tilde{\mathcal{X}}$, in which case $\Psi \tilde{X}_n$ is undefined, and second, the function $s^{-1}(\Psi \tilde{X}_n)(s)$ is not necessarily a Laplace transform and thus not amenable for inversion.

To overcome the first issue, we let E_n be the event that $\tilde{X}_n \in \tilde{\mathcal{X}}$. We assume that E_n lies in the natural filtration generated by $X_1, \ldots, X_n$, is not defined in terms of characteristics of the (unknown) X, and also that $\mathbb{P}(E_n^c) \to 0$ as $n \to \infty$. For the main result of this paper, Theorem 1 in Section 3, it turns out to be irrelevant how $F^Y(w)$ is estimated on E_n^c (as long as the estimate lies in $[0, 1]$); we could, for example, estimate it by zero on this event. In concrete situations, it is typically easy to determine a suitable choice for the sets E_n; for both applications considered in Section 4, we explicitly identify the E_n.

On the event E_n, we estimate the Laplace transform of F^Y by the plug-in estimator

$$\bar{F}_n^Y(s) := \frac{1}{s}(\Psi \tilde{X}_n)(s), \qquad (s \in \mathbb{C}_{++}). \tag{1}$$

To overcome the second issue, of $\bar{F}_n^Y(s)$ not necessarily being a Laplace transform, we estimate $F^Y(w)$ by applying a *truncated* version of Bromwich' Inversion formula [13]:

$$F_n^Y(w) = \int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \bar{F}_n^Y(c + \mathrm{i}y) \, dy, \tag{2}$$

where c is an arbitrary positive real number. In the "untruncated" version of Bromwich' Inversion formula the integration in (2) is over the whole real line. Since that integral may not be well-defined if $\bar{F}_n^Y$ is not a Laplace transform, we integrate over a finite interval (which grows in the sample size n).

The thus constructed estimator has remedied the two complications that we identified above. The main result of this paper, which describes the performance of this estimator as a function of the sample size n, is given in the next section.

3. Main result: Convergence rate

In this section, we show that the expected absolute estimation error of our estimator $F_n^Y(w)$, as defined in the previous section, is bounded from above by a constant times $n^{-1/2} \log(n + 1)$.

Our result is proven under the following assumptions.

(A1) For each $n \in \mathbb{N}$ there is an event $A_n \subset E_n$, such that $\mathbb{P}(A_n^c) \le \kappa_1 n^{-1/2}$ for some $\kappa_1 > 0$ independent of n;

(A2) $F^Y(y)$ is continuously differentiable on $[0, \infty)$, and twice differentiable at $y = w$;

(A3) There are constants $\kappa_2 \geq 0$, $\kappa_3 \geq 0$ and (nonnegative and random) Z_n, $n \in \mathbb{N}$, such that $\sup_{p \in (1,2)} \mathbb{E}[|Z_n|^p] \leq \kappa_3 n^{-1/2}$ for all $n \in \mathbb{N}$, and such that, on the event A_n,

$$\left|(\Psi \tilde{X}_n)(s) - (\Psi \tilde{X})(s)\right| \leq \kappa_2 \left|\tilde{X}_n(s) - \tilde{X}(s)\right| + Z_n \qquad \text{a.s.,}$$

for all $s = c + \mathrm{i}y$, $n \in \mathbb{N}$, and $-\sqrt{n} \leq y \leq \sqrt{n}$.

These assumptions are typically "mild;" we proceed with a short discussion of each of them.

Assumption (A1) ensures that the contribution of the complement of A_n (and also that of the complement of E_n) to the expected absolute estimation error is sufficiently small. The difference between A_n and E_n is that the definition of E_n does not involve the unknown $X \in \mathcal{X}$ (which enables us to define the estimator $F_n^Y(w)$ *without* knowing the unknown X), whereas A_n may actually depend on X. It is noted that in specific applications, this helps when checking whether the assumptions (A1)–(A3) are satisfied; cf. the proofs of Theorems 2 and 3. If $\mathbb{P}(A_n^c) \sim n^{-a}$ for some $a \in (-1/2, 0)$, then (A1) does not hold and Theorem 1 is not valid.

Assumption (A2) is a smoothness condition on F^Y that we use to control the error caused by integrating in (2) over a finite interval, instead of integrating over $\mathbb{R}$. The twice-differentiability assumption is only used to apply Lemma 3 with $f = F^Y$ in the proof of Theorem 1. It can be replaced by any other condition that guarantees that, for all $n \in \mathbb{N}$ and some $\kappa_4 > 0$ independent of n,

$$\left| \int_{|y| > \sqrt{n}} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \bar{F}^Y(c + \mathrm{i}y) \, \mathrm{d}y \right| \leq \frac{\kappa_4}{\sqrt{n}}.$$

Continuous differentiability of F^Y makes sure that Bromwich' Inversion formula applied to $\bar{F}^Y$ yields F^Y again, i.e., $F^Y(w) = \int_{-\infty}^{\infty} (2\pi)^{-1} e^{(c+\mathrm{i}y)w} \bar{F}^Y(c + \mathrm{i}y) \, \mathrm{d}y$, cf. [22], Chapter 4. This equality is still true if the derivative of $F^Y(y)$ with respect to y is piecewise continuous with finitely many discontinuity points, and in addition continuous at $y = w$.

Assumption (A3) can be seen as a kind of Lipschitz-continuity condition on Ψ that guarantees that $\Psi \tilde{X}_n$ is "close to" $\Psi \tilde{X}$ if $\tilde{X}_n$ is "close to" $\tilde{X}$. This condition is necessary to prove the weak consistency of our estimator. The formulation with the random variables Z_n allows for a more general setting than with $Z_n = 0$, and is used in both applications in Section 4.

A straightforward example that satisfies assumptions (A1)–(A3) is the case where $X \overset{d}{=} Y + W$, where W is a known nonnegative random variable. If the cdf of Y satisfies the smoothness condition (A2), then, with $\tilde{Y}(s) = (\Psi \tilde{X})(s) := \tilde{X}(s)/\tilde{W}(s)$, $A_n^c = E_n^c = \varnothing$, $Z_n = 0$ a.s., $c > 0$ arbitrary, and $\kappa_2 := \sup_{s=c+\mathrm{i}y, -\sqrt{n} \leq y \leq \sqrt{n}} 1/|\tilde{Z}(s)|$, it is easily seen that assumptions (A1)–(A3) are satisfied. More involved examples that satisfy the assumptions are presented in Section 4.

Theorem 1. *Let $w > 0$, $c > 0$, and assume (A1)–(A3). Then $F_n^Y(w)$ converges to $F^Y(w)$ in probability, as $n \to \infty$, and there is a constant $C > 0$ such that, for all $n \in \mathbb{N}$,*

$$\mathbb{E}\left[\left|F_n^Y(w) - F^Y(w)\right|\right] \leq C n^{-1/2} \log(n + 1). \tag{3}$$

Proof. It suffices to prove (3), since this implies weak consistency of $F_n^Y(w)$. Fix $n \in \mathbb{N}$.

The proof consists of three steps. In Step 1, we bound the estimation error on the event A_n, in Step 2 we consider the complement A_n^c, and in Step 3 we combine Step 1 and 2 to arrive at the statement of the theorem. Some of the intermediate steps in the proof rely on auxiliary results that are presented in Section 5.

Step 1. We show that there are positive constants κ_4 and κ_5, independent of n, such that, for all $n \in \mathbb{N}$ and $p \in (1, 2)$,

$$\mathbb{E}\big[\big|F_n^Y(w) - F^Y(w)\big| \cdot \mathbf{1}_{A_n}\big] \le \kappa_4 n^{-1/2} + \kappa_5 (p-1)^{-1/p} n^{1/2 - 1/p}. \tag{4}$$

To prove the inequality (4), consider the following elementary upper bound:

$$\mathbb{E}\big[\big|F_n^Y(w) - F^Y(w)\big| \cdot \mathbf{1}_{A_n}\big]$$

$$= \mathbb{E}\bigg[\bigg|\int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{(c+iy)w} \bar{F}_n^Y(c+iy)\,\mathrm{d}y - \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{(c+iy)w} \bar{F}^Y(c+iy)\,\mathrm{d}y\bigg| \cdot \mathbf{1}_{A_n}\bigg]$$

$$= \frac{1}{2\pi}\mathbb{E}\bigg[\bigg|\int_{-\sqrt{n}}^{\sqrt{n}} e^{(c+iy)w}\big(\bar{F}_n^Y(c+iy) - \bar{F}^Y(c+iy)\big)\,\mathrm{d}y$$

$$\qquad - \int_{|y|>\sqrt{n}} e^{(c+iy)w} \bar{F}^Y(c+iy)\,\mathrm{d}y\bigg| \cdot \mathbf{1}_{A_n}\bigg]$$

$$\le \mathbb{E}\bigg[\bigg|\int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{(c+iy)w}\big(\bar{F}_n^Y(c+iy) - \bar{F}^Y(c+iy)\big)\,\mathrm{d}y\bigg| \cdot \mathbf{1}_{A_n}\bigg] \tag{5}$$

$$+ \bigg|\int_{|y|>\sqrt{n}} \frac{1}{2\pi} e^{(c+iy)w} \bar{F}^Y(c+iy)\,\mathrm{d}y\bigg|. \tag{6}$$

We now treat the terms (5) and (6) separately, starting with the latter. By assumption (A2) and the observation

$$\int_w^\infty \bigg(\frac{\mathrm{d}}{\mathrm{d}y}F^Y(y+w)\bigg)\frac{e^{-cy}}{y}\frac{w}{e^{-cw}}\,\mathrm{d}y \le \int_w^\infty \bigg(\frac{\mathrm{d}}{\mathrm{d}y}F^Y(y+w)\bigg)\,\mathrm{d}y = F^Y(\infty) - F^Y(2w) < 1,$$

we conclude that

$$\int_w^\infty \bigg|\frac{\mathrm{d}}{\mathrm{d}y}F^Y(y+w)\bigg|\frac{e^{-cy}}{y}\,\mathrm{d}y < \frac{e^{-cw}}{w} < \infty,$$

and therefore F^Y satisfies the conditions of Lemma 3. As a result, (6) satisfies

$$\bigg|\int_{|y|>\sqrt{n}} \frac{1}{2\pi} e^{(c+iy)w} \bar{F}^Y(c+iy)\,\mathrm{d}y\bigg| \le \frac{\kappa_4}{\sqrt{n}}, \tag{7}$$

for some constant $\kappa_4 > 0$ independent of n.

We now bound the term (5). It is obviously majorized by

$$\mathbb{E}\bigg[\int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{cw} \big|\bar{F}_n^Y(c+\mathrm{i}y) - \bar{F}^Y(c+\mathrm{i}y)\big| \,\mathrm{d}y \cdot \mathbf{1}_{A_n}\bigg].$$

Let $p \in (1,2)$ and $q > 1$, with $p^{-1} + q^{-1} = 1$. By subsequent application of Hölder's inequality, this expression is further bounded by

$$\mathbb{E}\bigg[\bigg(\int_{-\sqrt{n}}^{\sqrt{n}} \Big(\frac{1}{2\pi} e^{cw}\Big)^q \,\mathrm{d}y\bigg)^{1/q} \bigg(\int_{-\sqrt{n}}^{\sqrt{n}} \big|\bar{F}_n^Y(c+\mathrm{i}y) - \bar{F}^Y(c+\mathrm{i}y)\big|^p \,\mathrm{d}y\bigg)^{1/p} \cdot \mathbf{1}_{A_n}\bigg].$$

By computing the first integral, and an application of Jensen's inequality, this is not larger than

$$e^{cw} \frac{(2\sqrt{n})^{1/q}}{2\pi} \bigg(\mathbb{E}\bigg[\int_{-\sqrt{n}}^{\sqrt{n}} \big|\bar{F}_n^Y(c+\mathrm{i}y) - \bar{F}^Y(c+\mathrm{i}y)\big|^p \,\mathrm{d}y \cdot \mathbf{1}_{A_n}\bigg]\bigg)^{1/p}.$$

Finally applying Fubini's theorem, we arrive at the upper bound

$$e^{cw} \frac{(2\sqrt{n})^{1/q}}{2\pi} \bigg(\int_{-\sqrt{n}}^{\sqrt{n}} \mathbb{E}\big[\big|\bar{F}_n^Y(c+\mathrm{i}y) - \bar{F}^Y(c+\mathrm{i}y)\big|^p \cdot \mathbf{1}_{A_n}\big] \,\mathrm{d}y\bigg)^{1/p}. \tag{8}$$

We now study the behavior of (8), being an upper bound to (5), as a function of n. To this end, we first derive a bound on the integrand. Assumption (A3) implies that there exists a sequence of nonnegative random variables Z_n, $n \in \mathbb{N}$, such that

$$\begin{aligned}
\big|\bar{F}_n^Y(s) - \bar{F}^Y(s)\big| \cdot \mathbf{1}_{A_n} &= \big|s^{-1}(\Psi\tilde{X}_n)(s) - s^{-1}(\Psi\tilde{X})(s)\big| \cdot \mathbf{1}_{A_n} \\
&\leq \big(\kappa_2\big|\tilde{X}(s) - \tilde{X}_n(s)\big| + Z_n\big) \cdot \big|s^{-1}\big| \cdot \mathbf{1}_{A_n} \qquad \text{a.s.,}
\end{aligned} \tag{9}$$

for all $s = c + \mathrm{i}y$ with $-\sqrt{n} \leq y \leq \sqrt{n}$.

Now recall the so-called c_r-inequality

$$\mathbb{E}\big[|X+Y|^p\big] \leq 2^{p-1}\big(\mathbb{E}\big[|X|^p\big] + \mathbb{E}\big[|Y|^p\big]\big),$$

and the obvious inequality $\mathbf{1}_{A_n} \leq 1$ a.s. As a consequence of Lemma 1, we thus obtain

$$\begin{aligned}
\mathbb{E}\big[&\big|\bar{F}_n^Y(c+\mathrm{i}y) - \bar{F}^Y(c+\mathrm{i}y)\big|^p \cdot \mathbf{1}_{A_n}\big] \\
&\leq 2^{p-1}\big(\kappa_2^p \mathbb{E}\big[\big|\tilde{X}(c+\mathrm{i}y) - \tilde{X}_n(c+\mathrm{i}y)\big|^p\big] + \mathbb{E}\big[|Z_n|^p\big]\big)|c+\mathrm{i}y|^{-p} \\
&\leq 2^{p-1}\big(2^p\kappa_2^p + \kappa_3\big)n^{-1/2}|c+\mathrm{i}y|^{-p}.
\end{aligned}$$

From (8) and the straightforward inequality

$$\begin{aligned}
\int_{-\infty}^{\infty} |c+\mathrm{i}y|^{-p} \,\mathrm{d}y &\leq \int_{-\infty}^{\infty} \frac{1}{(c^2+y^2)^{p/2}} \,\mathrm{d}y = c^{1-p} \int_0^{\infty} \frac{z^{-1/2}}{(1+z)^{p/2}} \,\mathrm{d}z \\
&= C_0(p) := c^{1-p}\pi^{1/2}\frac{\Gamma((p-1)/2)}{\Gamma(p/2)},
\end{aligned} \tag{10}$$

it follows that

$$\mathbb{E}\left[\left|\int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{(c+iy)w} \left(\bar{F}_n^Y(c+iy) - \bar{F}^Y(c+iy)\right) dy\right| \cdot \mathbf{1}_{A_n}\right]$$

$$\leq e^{cw} \frac{(2\sqrt{n})^{1/q}}{2\pi} \left(2^{p-1}\left(2^p \kappa_2^p + \kappa_3\right) n^{-1/2} \int_{-\sqrt{n}}^{\sqrt{n}} |c+iy|^{-p} dy\right)^{1/p} \quad (11)$$

$$\leq C_1(p) n^{1/2 - 1/p},$$

where

$$C_1(p) := e^{cw} \frac{2^{2-2/p}}{2\pi} \left(2^p \kappa_2^p + \kappa_3\right)^{1/p} \left(C_0(p)\right)^{1/p}.$$

It follows from $\Gamma((p-1)/2) = 2\Gamma((p+1)/2)/(p-1)$ that

$$\lim_{p \downarrow 1} (p-1)^{1/p} C_1(p) < \infty.$$

This implies that there is a $\kappa_5 > 0$ such that

$$C_1(p) \leq \kappa_5 (p-1)^{-1/p} \qquad \text{for all } p \in (1, 2). \quad (12)$$

Upon combining the results presented in displays (5), (6), (7), (11), and (12), we obtain inequality (4), as desired.

Step 2. On the complement of the event A_n we have, by assumption (A1),

$$\mathbb{E}\left[\left|F_n^Y(w) - F^Y(w)\right| \cdot \mathbf{1}_{A_n^c}\right] \leq \mathbb{P}\left(A_n^c\right) \leq \kappa_1 n^{-1/2}. \quad (13)$$

Step 3. When combining inequalities (4) and (13), we obtain that

$$\mathbb{E}\left[\left|F_n^Y(w) - F^Y(w)\right|\right] = \mathbb{E}\left[\left|F_n^Y(w) - F^Y(w)\right| \cdot \mathbf{1}_{A_n}\right] + \mathbb{E}\left[\left|F_n^Y(w) - F^Y(w)\right| \cdot \mathbf{1}_{A_n^c}\right]$$

$$\leq \kappa_4 n^{-1/2} + \kappa_5 (p-1)^{-1/p} n^{1/2-1/p} + \kappa_1 n^{-1/2}.$$

Now realize that we have the freedom to pick in the above inequality any $p \in (1, 2)$. In particular, the choice $p = p_n := 1 + 1/(2\log(n+1)) \in (1, 2)$ yields the bound

$$\mathbb{E}\left[\left|F_n^Y(w) - F^Y(w)\right|\right] \leq \kappa_4 n^{-1/2} + \kappa_5 \left(2\log(n+1)\right)^{1/p_n} n^{1/2-1/p_n} + \kappa_1 n^{-1/2}$$

$$\leq \left(\kappa_4 + 2\kappa_5 e^{1/2} + \kappa_1\right) n^{-1/2} \log(n+1),$$

using

$$\frac{1}{2} - \frac{1}{p_n} = -\frac{1}{2} + \frac{1}{1 + 2\log(n+1)}$$

and

$$n^{1/(1+2\log(n+1))} = \exp\left(\frac{\log n}{1 + 2\log(n+1)}\right) \leq e^{1/2}.$$

This finishes the proof of Theorem 1. $\qquad\qquad\square$

Remark 1. Contrary to some of the literature mentioned in Section 1 (e.g. [20] and [23]), we are not estimating a density but a cumulative distribution function. This difference translates into an additional $|s^{-1}|$ term in equation (9), which enables us to bound the integral in equation (10). This appears to be an crucial step in the proof of Theorem 1, because it means that the ill-posedness of the inversion problem (the fact that the inverse Laplace transform operator is not continuous) does not play a rôle: convergence of $\bar{F}_n^Y(\cdot)$ to $\bar{F}^Y(\cdot)$ implies convergence of $F_n^Y(w)$ to $F^Y(w)$. As a result, we do not need regularization techniques as in [20] and [23].

4. Applications

In this section, we discuss two examples that have attracted a substantial amount of attention in the literature. In both examples, the verification of the assumptions (A1)–(A3) can be done, as we will demonstrate now.

4.1. Workload estimation in an M/G/1 queue

In our first example, we consider the so-called M/G/1 queue: jobs arrive at a service station according to a Poisson process with rate $\lambda > 0$, where these jobs are i.i.d. samples with a service time distribution B; see, for example, [7] for an in-depth account of the M/G/1 queue, and [21] for an annotated bibliography on inference in queueing models. Under the stability condition $\rho := \lambda \mathbb{E}[B] \in (0, 1)$ the queue's stationary workload is well defined. Our objective, motivated by the setup described in [11], is to estimate $\mathbb{P}(Y > w)$, where Y is the stationary workload, and $w > 0$ is a given threshold. The idea is that this estimate is based on samples of the queue's input process.

In more detail, the procedure works as follows. By the Pollaczek–Khintchine formula [18], the Laplace transform of the stationary workload distribution Y satisfies the relation

$$\tilde{Y}(s) = \frac{s(1-\rho)}{s - \lambda + \lambda \tilde{B}(s)}, \qquad s \in \mathbb{C}_+. \tag{14}$$

For subsequent time intervals of (deterministic) length $\delta > 0$, the amount of work arriving to the queue is measured. These observations are i.i.d. samples from a compound distribution $X \overset{\mathrm{d}}{=} \sum_{i=1}^N B_i$, with N Poisson distributed with parameter $\lambda\delta$, and the random variables $B_1, B_2, \ldots$ independent and distributed as B (independent of N). By Wald's equation, we have $\mathbb{E}[X] = \delta\rho$, and a direct computation yields $\tilde{X}(s) = \exp(-\lambda\delta + \lambda\delta \tilde{B}(s))$. Combining this with (14), we obtain the following relation between the Laplace transforms of X and Y:

$$\tilde{Y}(s) = \frac{s(1 - \delta^{-1}\mathbb{E}[X])}{s + \delta^{-1}\mathrm{Log}(\tilde{X}(s))}. \tag{15}$$

Here Log is the distinguished logarithm of $\tilde{X}(s)$ [6], which is convenient to work with in this context [24]. Our goal is to estimate $\mathbb{P}(Y \leq w) = F^Y(w)$, for a given $w > 0$, based on an independent sample $X_1, \ldots, X_n$. We use the estimator $F_n^Y(w)$ defined in Section 2, for an arbitrary $c > 0$, and with

(i) $\mathcal{X}$ the collection of all random variables X' of the form $\sum_{i=1}^{N'} B_i'$ with N' Poisson distributed with strictly positive mean, $\{B_i'\}_{i\in\mathbb{N}}$ i.i.d., independent of N', nonnegative, and with $0 < \mathbb{E}[X'] < \delta$;

(ii) the sets

$$E_n := \left\{ 0 \le \frac{1}{n} \sum_{i=1}^{n} X_i < \delta \right\},$$

so as to ensure that the "empirical occupation rate" of the queue, $\delta^{-1} n^{-1} \sum_{i=1}^{n} X_i$, is strictly smaller than one, and that therefore the Pollaczek–Khintchine formula holds;

(iii) Ψ defined through

$$(\Psi \tilde{X})(s) = \frac{s(1 + \delta^{-1} \tilde{X}'(0))}{s + \delta^{-1} \mathrm{Log}(\tilde{X}(s))}, \qquad s \in \mathbb{C}_+,$$

where $\tilde{X}'(t)$ denotes the derivative of $\tilde{X}(t)$ in $t \in (0, \infty)$ and $\tilde{X}'(0) = \lim_{t\downarrow 0} \tilde{X}'(t) = -\mathbb{E}[X]$.

Theorem 2. *Consider the estimation procedure outlined above. Suppose F^Y is continuously differentiable, twice differentiable in w, and $\mathbb{E}[B^2] < \infty$. Then there is a constant $C > 0$ such that*

$$\mathbb{E}\big[\big| F_n^Y(w) - F^Y(w) \big|\big] \le C n^{-1/2} \log(n+1)$$

for all $n \in \mathbb{N}$.

Proof. Let $n \in \mathbb{N}$ be arbitrary, and define the events

$$A_{n,1} := \left\{ \sup_{-\sqrt{n} \le y \le \sqrt{n}} \left| \frac{\tilde{X}_n(c+iy)}{\tilde{X}(c+iy)} - 1 \right| \le \min\left\{ \frac{1}{2}, \frac{c\delta(1 - \delta^{-1}\mathbb{E}[X])}{2\log 4} \right\} \right\},$$

$$A_{n,2} := \left\{ \delta^{-1} \left| \mathbb{E}[X] - \frac{1}{n} \sum_{i=1}^{n} X_i \right| \le \delta^{-1} \mathbb{E}[X] \big(1 - \delta^{-1}\mathbb{E}[X] \big) \right\},$$

and $A_n := A_{n,1} \cap A_{n,2}$. We have $A_{n,2} \subset E_n$ (because, using $\rho = \delta^{-1}\mathbb{E}[X]$, the event $A_{n,2}$ implies $\delta^{-1} n^{-1} \sum_{i=1}^{n} X_i \in [\rho^2, \rho(2-\rho)] \subset (0, 1)$) and thus $A_n \subset E_n$. We show that assumptions (A1)–(A3), as defined in Section 3, are satisfied. To this end, we only need to show (A1) and (A3), since (A2) is assumed in the statement of the theorem.

Assumption (A1). Let

$$\beta := \exp(-2\lambda\delta) \min\left\{ \frac{1}{2}, \frac{c\delta(1 - \delta^{-1}\mathbb{E}[X])}{2\log 4} \right\}.$$

Then, for $s \in \mathbb{C}_+$,

$$\big| \tilde{X}(s) \big| = \big| \exp\big(-\lambda\delta + \lambda\delta \tilde{B}(s) \big) \big| \ge \exp\big(-\lambda\delta + \lambda\delta \Re\big(\tilde{B}(s) \big) \big) \ge \exp(-2\lambda\delta), \qquad (16)$$

which implies that

$$\mathbb{P}\big(A_{n,1}^c\big) \le \mathbb{P}\Big(\sup_{-\sqrt{n} \le y \le \sqrt{n}} \big|\tilde{X}_n(c+\mathrm{i}y) - \tilde{X}(c+\mathrm{i}y)\big| > \beta \Big),$$

so that Lemma 2 then yields

$$\mathbb{P}\big(A_{n,1}^c\big) < 4\Big(1 + \frac{8\sqrt{n}\mathbb{E}[|X|]}{\beta}\Big)e^{-n\beta^2/18} + \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i\Big| \ge \frac{4}{3}\mathbb{E}[|X|]\Big).$$

Since $\sqrt{n}\exp(-n\beta^2/18) = O(n^{-1/2})$ and

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i\Big| \ge \frac{4}{3}\mathbb{E}[|X|]\Big) \le \mathbb{P}\Big(\Big|\mathbb{E}[X] - \frac{1}{n}\sum_{i=1}^{n} X_i\Big| \ge \frac{1}{3}\mathbb{E}[X]\Big) \le n^{-1}\frac{9\mathbb{E}[(X - \mathbb{E}[X])^2]}{\mathbb{E}[X]^2},$$

using the nonnegativity of X and Chebyshev's inequality, it follows that $\mathbb{P}(A_{n,1}^c) = O(n^{-1/2})$. It also follows easily from Chebyshev's inequality that $\mathbb{P}(A_{n,2}^c) = O(n^{-1/2})$. It follows immediately that $\mathbb{P}(A_n^c) \le \mathbb{P}(A_{n,1}^c) + \mathbb{P}(A_{n,2}^c) = O(n^{-1/2})$, which implies that assumption (A1) is satisfied.

Assumption (A3). Fix $y \in [-\sqrt{n}, \sqrt{n}]$ and $s = c + \mathrm{i}y$. Then

$$\big|(\Psi\tilde{X}_n)(s) - (\Psi\tilde{X})(s)\big|$$

$$\le \left| \frac{s(1 + \delta^{-1}\tilde{X}_n'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}_n(s))} - \frac{s(1 + \delta^{-1}\tilde{X}_n'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}(s))} \right|$$

$$+ \left| \frac{s(1 + \delta^{-1}\tilde{X}_n'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}(s))} - \frac{s(1 + \delta^{-1}\tilde{X}'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}(s))} \right|$$

$$\le \left| \frac{\delta^{-1}\mathrm{Log}(\tilde{X}(s)) - \delta^{-1}\mathrm{Log}(\tilde{X}_n(s))}{s(1 + \delta^{-1}\tilde{X}'(0))} \cdot \frac{s(1 + \delta^{-1}\tilde{X}_n'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}_n(s))} \cdot \frac{s(1 + \delta^{-1}\tilde{X}'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}(s))} \right| \qquad (17)$$

$$+ \left| \frac{\delta^{-1}(\tilde{X}_n'(0) - \tilde{X}'(0))}{(1 + \delta^{-1}\tilde{X}'(0))} \cdot \frac{s(1 + \delta^{-1}\tilde{X}'(0))}{s + \delta^{-1}\mathrm{Log}(\tilde{X}(s))} \right|$$

$$\le \frac{\delta^{-1}}{1 + \delta^{-1}\mathbb{E}[X]}\big|\mathrm{Log}(\tilde{X}(s)) - \mathrm{Log}(\tilde{X}_n(s))\big| \cdot \big|s^{-1}(\Psi\tilde{X}_n)(s)\big| \cdot \big|(\Psi\tilde{X})(s)\big|$$

$$+ \frac{\delta^{-1}}{1 + \delta^{-1}\mathbb{E}[X]}\Big|\mathbb{E}[X] - \frac{1}{n}\sum_{i=1}^{n} X_i\Big| \cdot \big|(\Psi\tilde{X})(s)\big|.$$

If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function with $f(0) = 1$ and $f(t) \neq 0$ for all $t \in \mathbb{R}$, then for all t such that $|f(t) - 1| \leq \frac{1}{2}$ we have $\text{Log}(f(t)) = L(f(t))$, where, for $z \in \mathbb{C}$, $|z - 1| < 1$,

$$L(z) = \sum_{j \geq 1} \frac{(-1)^{j-1}}{j} (z-1)^j; \tag{18}$$

this follows from the construction of the distinguished logarithm [6]. In addition, if $|z - 1| \leq \frac{1}{2}$, then

$$\left| L(z) \right| \leq \sum_{j \geq 1} \frac{1}{j} |z-1|^j = \log\left(\frac{1}{1 - |z-1|} \right) \leq |z-1| \log 4. \tag{19}$$

This implies that, on A_n, we have

$$\left| \text{Log}\big(\tilde{X}_n(c+iy)\big) - \text{Log}\big(\tilde{X}(c+iy)\big) \right|$$
$$= \left| \text{Log}\left(\frac{\tilde{X}_n(c+iy)}{\tilde{X}(c+iy)} \right) \right| = \left| L\left(\frac{\tilde{X}_n(c+iy)}{\tilde{X}(c+iy)} \right) \right| \tag{20}$$
$$\leq \left| \frac{\tilde{X}_n(c+iy)}{\tilde{X}(c+iy)} - 1 \right| \log 4 \leq \left| \tilde{X}_n(c+iy) - \tilde{X}(c+iy) \right| (\log 4) \exp(2\lambda\delta),$$

where the last inequality follows from (16).

Furthermore, we have on A_n that

$$\left| s^{-1}(\Psi \tilde{X}_n)(s) \right|$$
$$= \left| \frac{1 - \delta^{-1} 1/n \sum_{i=1}^n X_i}{s + \delta^{-1} \text{Log}(\tilde{X}_n(s))} \right|$$
$$\leq \left| \frac{1 - \delta^{-1} 1/n \sum_{i=1}^n X_i}{1 - \delta^{-1} \mathbb{E}[X]} \right| \cdot \left| \frac{s + \delta^{-1} \text{Log}(\tilde{X}_n(s))}{s + \delta^{-1} \text{Log}(\tilde{X}(s))} \right|^{-1} \cdot \left| \frac{1 - \delta^{-1} \mathbb{E}[X]}{s + \delta^{-1} \text{Log}(\tilde{X}(s))} \right|$$
$$\leq \left| 1 + \delta^{-1} \frac{\mathbb{E}[X] - 1/n \sum_{i=1}^n X_i}{1 - \delta^{-1} \mathbb{E}[X]} \right| \cdot \left| 1 + \frac{\delta^{-1} \text{Log}(\tilde{X}_n(s)/\tilde{X}(s))}{s + \delta^{-1} \text{Log}(\tilde{X}(s))} \right|^{-1} \cdot \left| s^{-1} \tilde{Y}(s) \right| \tag{21}$$
$$\leq \left(1 + \frac{\delta^{-1} \mathbb{E}[X](1 - \delta^{-1} \mathbb{E}[X])}{1 - \delta^{-1} \mathbb{E}[X]} \right) \cdot \left| 1 + (\Psi \tilde{X})(s) \frac{\delta^{-1} L(\tilde{X}_n(s)/\tilde{X}(s))}{s(1 - \delta^{-1} \mathbb{E}[X])} \right|^{-1} \cdot c^{-1}$$
$$\leq \left(1 + \frac{\delta^{-1} \mathbb{E}[X](1 - \delta^{-1} \mathbb{E}[X])}{1 - \delta^{-1} \mathbb{E}[X]} \right) \cdot 2 \cdot c^{-1},$$

since $|1 + z|^{-1} \leq (1 - |z|)^{-1} \leq (1 - 1/2)^{-1}$ for all $z \in \mathbb{C}$ with $|z| \leq \frac{1}{2}$; bear in mind that, in particular,

$$\left| (\Psi \tilde{X})(s) \frac{\delta^{-1} L(\tilde{X}_n(s)/\tilde{X}(s))}{s(1 - \delta^{-1} \mathbb{E}[X])} \right| \leq \frac{\delta^{-1} c^{-1} |L(\tilde{X}_n(s)/\tilde{X}(s))|}{1 - \delta^{-1} \mathbb{E}[X]} \leq \frac{\delta^{-1} c^{-1} \log 4}{1 - \delta^{-1} \mathbb{E}[X]} \left| \frac{\tilde{X}_n(s)}{\tilde{X}(s)} - 1 \right| \leq \frac{1}{2}$$

on $A_{n,1}$. Finally, writing

$$Z_n = \frac{\delta^{-1}}{1 + \delta^{-1}\mathbb{E}[X]} \left| \mathbb{E}[X] - \frac{1}{n}\sum_{i=1}^{n} X_i \right| \cdot \left| (\Psi\tilde{X})(s) \right|,$$

and noting that $\mathbb{E}[X^2] < \infty$, it follows from Lemma 1 and $|(\Psi\tilde{X})(s)| \leq 1$ that $\mathbb{E}[|Z_n|^p] \leq \kappa_3 n^{-1/2}$ for all $p \in (1, 2)$ and some $\kappa_3 > 0$ independent of n and p. Combining this with equations (17), (20), and (21), implies that assumption (A3) holds. □

Remark 2. An important problem in [11] is to develop heuristics for choosing δ, in order to minimize the expected estimation error. In the proof of Theorem 1 we show the following upper bound

$$\mathbb{E}\left[\left| F_n^Y(w) - F^Y(w) \right|\right] \leq \kappa_4 n^{-1/2} + C_1(p)n^{1/2-1/p} + \mathbb{P}\left(A_n^c\right), \tag{22}$$

where $p = p_n = 1 + 1/(2\log(n+1))$. A close look at the proof reveals that $\lim_{p\downarrow 1}(p - 1)^{1/p} C_1(p) = \exp(cw)\pi^{-1}(2\kappa_2 + \kappa_3)$, and for the M/G/1 example it is not difficult to show that $\kappa_2 = \kappa_2(\delta) \leq 2c^{-1}\delta^{-1}$, $\kappa_3 = \kappa_3(\delta) \leq (1 + \text{Var}[X])\delta^{-1}(1 + \rho)^{-1}$, $\text{Var}[X] = \delta\lambda\mathbb{E}[B^2]$, and $\mathbb{P}(A_n^c) = O(n^{-1})$. This means that, for large n, the right-hand side of (22) can be approximated by

$$(\alpha + \beta\delta^{-1})e^{1/2}n^{-1/2}\log(n+1), \tag{23}$$

where

$$\alpha := \kappa_4 + e^{cw}\pi^{-1}\lambda\mathbb{E}\left[B^2\right](1+\rho)^{-1}, \qquad \beta := e^{cw}\pi^{-1}\left(4c^{-1} + (1+\rho)^{-1}\right).$$

If we neglect the $\log(n+1)$ term, then, on a fixed time horizon of length $T = \delta n$, the upper bound (23) equals

$$\left(\alpha\delta^{1/2} + \beta\delta^{-1/2}\right)e^{1/2}T^{-1/2},$$

which suggests that δ should be chosen that minimizes $\alpha\delta^{1/2} + \beta\delta^{-1/2}$. In the application [11] α and β are unknown (because they depend on e.g., λ and $\mathbb{E}[B^2]$), but if they can be replaced by known upper bounds α_u and β_u, then a heuristic choice for δ is to pick a minimizer of $\alpha_u\delta^{1/2} + \beta_u\delta^{-1/2}$ (yielding $\delta = \beta_u/\alpha_u$).

Remark 3. Interestingly, the technique described above enables a fast and accurate estimation of rare-event probabilities (i.e., $1 - F^Y(w)$ for w large), even in situations in which the estimation is based on input $X_1, \ldots, X_n$ for which the corresponding queue would not have exceeded level w. This idea, which resonates the concepts developed in [19], has been worked out in detail in [11]. A numerical illustration of our estimator in this setting, and a comparison to the empirical estimator, is provided in Section 6.

4.2. Decompounding

Our second application involves decompounding a compound Poisson distribution, a concept that has been studied in the literature already (see the remarks on this in the introduction).

We start by providing a formal definition of the problem. Let $\mathcal{X}$ denote the collection of random variables of the form $\sum_{i=1}^{N'} Y_i'$, with N' Poisson distributed with $\mathbb{E}[N'] > 0$, and $(Y_i')_{i \in \mathbb{N}}$ i.i.d. nonnegative random variables, independent of N', and with $\mathbb{P}(Y_1' = 0) = 0$ (which can be assumed without loss of generality). For each $\tilde{X} \in \tilde{\mathcal{X}}$, let, for $s \in \mathbb{C}_+$,

$$(\Psi \tilde{X})(s) = 1 + \frac{1}{-\log(\tilde{X}(\infty))} \mathrm{Log}(\tilde{X}(s)),$$

where Log denotes the distinguished logarithm of $\tilde{X}$, and

$$\tilde{X}(\infty) := \lim_{s \to \infty, s \in \mathbb{R}} \tilde{X}(s) = \lim_{s \to \infty, s \in \mathbb{R}} e^{\mathbb{E}[N](-1+\tilde{Y}_1(s))} = e^{-\mathbb{E}[N]}$$

if $X = \sum_{i=1}^{N} Y_i$; here the last equality follows from $\mathbb{P}(Y_1 = 0) = 0$.

Let $X = \sum_{i=1}^{N} Y_i$ be an element of $\mathcal{X}$, for some particular $Y \stackrel{\mathrm{d}}{=} Y_1$ and a Poisson distributed random variable N with mean $\lambda > 0$. Since $-\log(\tilde{X}(\infty)) = \lambda$ and $\tilde{X}(s) = \exp(-\lambda + \lambda \mathbb{E}[-sY])$, we have $\tilde{Y} = \Psi \tilde{X}$. The idea is to estimate $F^Y(w)$, for $w > 0$, based on a sample $X_1, \ldots, X_n$ of $n \in \mathbb{N}$ independent copies of X, using the estimator $F_n^Y(w)$ of Section 2, with, for $n \in \mathbb{N}$,

$$E_n := \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i = 0\}} \in (0, 1) \right\}$$

and arbitrary $c > 0$.

Theorem 3. *Consider the estimation procedure outlined above. Suppose F^Y is continuously differentiable, twice differentiable in w, and suppose $\mathbb{E}[|X|^2] < \infty$. Then there is a constant $C > 0$ such that*

$$\mathbb{E}\left[\left| F_n^Y(w) - F^Y(w) \right|\right] \le C n^{-1/2} \log(n + 1)$$

for all $n \in \mathbb{N}$.

Proof. Write

$$\lambda_n = -\log(\tilde{X}_n(\infty)) = -\log\left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i = 0\}} \right)$$

(being well-defined on E_n), and define

$$A_{n,1} := \left\{ \sup_{-\sqrt{n} \le y \le \sqrt{n}} \left| \tilde{X}_n(c + iy) - \tilde{X}(c + iy) \right| \le \exp(-2\lambda)/2 \right\},$$

$$A_{n,2} := \left\{ \frac{\lambda}{2} \leq \lambda_n \leq 2\lambda \right\},$$

and $A_n = A_{n,1} \cap A_{n,2}$. Note that $A_{n,2} \subset E_n$ and thus $A_n \subset E_n$. We show that assumptions (A1)–(A3) are valid. Because we explicitly assumed (A2), we are left with verifying (A1) and (A3). These verification resemble those of the M/G/1 example.

Assumption (A1). $\mathbb{P}(A_{n,1}^{\complement}) = O(\sqrt{n} \exp(-n\beta^2/18)) = O(n^{-1/2})$ follows from Lemma 2, with $\beta = \exp(-2\lambda)/2$, together with Chebyshev's inequality and the assumption $\mathbb{E}[|X|^2] < \infty$. $\mathbb{P}(A_{n,2}^{\complement}) = O(n^{-1/2})$ follows from Hoeffding's inequality, and thus

$$\mathbb{P}(A_n^{\complement}) \leq \mathbb{P}(A_{n,1}^{\complement}) + \mathbb{P}(A_{n,2}^{\complement}) = O(n^{-1/2}).$$

Assumption (A3). On A_n, for $s = c + iy$, $-\sqrt{n} \leq y \leq \sqrt{n}$, we have

$$\left| \frac{\tilde{X}_n(s)}{\tilde{X}(s)} - 1 \right| \leq |\tilde{X}_n(s) - \tilde{X}(s)| e^{2\lambda} \leq \frac{1}{2},$$

where $|\tilde{X}(s)|^{-1} \leq \exp(2\lambda)$ follows as in (16), and thus

$$\left| \text{Log}(\tilde{X}_n(s)) - \text{Log}(\tilde{X}(s)) \right| = \left| \text{Log}(\tilde{X}_n(s)/\tilde{X}(s)) \right| = \left| L(\tilde{X}_n(s)/\tilde{X}(s)) \right|$$

$$\leq |\tilde{X}_n(s) - \tilde{X}(s)| (\log 4) e^{2\lambda},$$

using (18) and (19). This implies

$$\left| (\Psi \tilde{X}_n)(s) - (\Psi \tilde{X})(s) \right| \cdot \mathbf{1}_{A_n}$$

$$\leq \left| \lambda_n^{-1} \text{Log}(\tilde{X}_n(s)) - \lambda_n^{-1} \text{Log}(\tilde{X}(s)) \right| \cdot \mathbf{1}_{A_n} + \left| \lambda_n^{-1} \text{Log}(\tilde{X}(s)) - \lambda^{-1} \text{Log}(\tilde{X}(s)) \right| \cdot \mathbf{1}_{A_n}$$

$$\leq \left| \lambda_n^{-1} \right| \cdot \left| \text{Log}(\tilde{X}_n(s)) - \text{Log}(\tilde{X}(s)) \right| \cdot \mathbf{1}_{A_n} + \left| \lambda_n^{-1} - \lambda^{-1} \right| \cdot \left| \text{Log}(\tilde{X}(s)) \right| \cdot \mathbf{1}_{A_n}$$

$$\leq \frac{2 \log 4}{\lambda} e^{2\lambda} \cdot \left| \tilde{X}_n(s) - \tilde{X}(s) \right| \cdot \mathbf{1}_{A_n} + Z_n \qquad \text{a.s.,}$$

with $Z_n = 2\lambda^{-2} |\lambda_n - \lambda| \cdot \mathbf{1}_{A_n}$. By definition of $A_{n,2}$, Z_n is bounded, and it follows from Hoeffding's inequality that there is a $\kappa_3 > 0$ independent of n such that, for all $1 < p < 2$, $\mathbb{E}[|Z_n|^p] \leq \kappa_3 n^{-1/2}$. This shows that (A3) is valid. $\qquad \square$

Remark 4. The decompounding example above can also be carried out with distributions other than Poisson. For example, if N is $\text{Bin}(M, p)$ distributed, for known $M \in \mathbb{N}$ and unknown $p \in (0, 1)$, then $\tilde{X}(s) = (p\tilde{Y}(s) + 1 - p)^M$, $\tilde{X}(\infty) = (1 - p)^M$, and thus

$$\tilde{Y}(s) = (\Psi \tilde{X})(s) := \frac{\tilde{X}(s)^{1/M} - \tilde{X}(\infty)^{1/M}}{1 - \tilde{X}(\infty)^{1/M}}.$$

Or, if N is negative binomially distributed, that is,

$$\mathbb{P}(N = n) = \binom{n + M - 1}{n}(1 - p)^M p^n, \qquad (n = 0, 1, 2, \ldots),$$

for some known $M \in \mathbb{N}$ and unknown $p \in (0, 1)$, then $\tilde{X}(s) = (1 - p)^M (1 - p\tilde{Y}(s))^{-M}$, $\tilde{X}(\infty) = (1 - p)^M$, and thus

$$\tilde{Y}(s) = (\Psi\tilde{X})(s) := \frac{1 - \tilde{X}(\infty)^{1/M}\tilde{X}(s)^{-1/M}}{1 - \tilde{X}(\infty)^{1/M}}.$$

For both examples it is not difficult to construct A_n and Z_n, in the same spirit as in the proof of Theorem 3, such that the convergence rates $\mathbb{E}[|F_n^Y(w) - F^Y(w)|] = O(n^{-1/2}\log(n + 1))$ hold. The key requirement on N to obtain these rates is that the relation $\tilde{X}(s) = \mathbb{E}[\tilde{Y}(s)^N]$ can be inverted, such that we can write $\tilde{Y}(s) = (\Psi\tilde{X})(s)$ for some mapping Ψ.

5. Auxiliary lemmas

This section contains a number of auxiliary lemmas that are used in the proofs of Theorems 1, 2, and 3.

Lemma 1. *Let $c > 0$, $n \in \mathbb{N}$ and let $X_1, \ldots, X_n$ be i.i.d. nonnegative random variables distributed as X. For all $p \in (1, 2)$ and $s \in c + i\mathbb{R}$,*

$$\mathbb{E}\left[\left|\tilde{X}(s) - \frac{1}{n}\sum_{i=1}^n \exp(-sX_i)\right|^p\right] \leq 2^p n^{-1/2},$$

and

$$\mathbb{E}\left[\left|\mathbb{E}[X] - \frac{1}{n}\sum_{i=1}^n X_i\right|^p\right] \leq (1 + \mathrm{Var}[X])n^{-1/2},$$

where the last inequality is only informative if $\mathrm{Var}[X] < \infty$.

Proof. Let $s \in c + i\mathbb{R}$. Since $X_i \geq 0$ a.s. for all $i = 1, \ldots, n$, we have

$$\left|\tilde{X}(s) - \frac{1}{n}\sum_{i=1}^n \exp(-sX_i)\right|^{p-1} \leq \left(\left|\tilde{X}(s)\right| + \frac{1}{n}\sum_{i=1}^n \left|\exp(-sX_i)\right|\right)^{p-1} \leq 2^{p-1} \qquad \text{a.s.}$$

Jensen's inequality then implies

$$\mathbb{E}\left[\left|\tilde{X}(s) - \frac{1}{n}\sum_{i=1}^n \exp(-sX_i)\right|^p\right] \leq 2^{p-1}\sqrt{\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n (\exp(-sX_i) - \mathbb{E}[\exp(-sX)])\right|^2\right]}$$

$$= 2^{p-1}\sqrt{\frac{1}{n}\mathbb{E}\left[\left|\exp(-sX) - \mathbb{E}[\exp(-sX)]\right|^2\right]} \leq 2^p n^{-1/2}.$$

Furthermore, we have, again by Jensen's inequality,

$$\mathbb{E}\left[\left|\mathbb{E}[X] - \frac{1}{n}\sum_{i=1}^{n} X_i\right|^p\right] \le n^{-p}\mathbb{E}\left[\left|\sum_{i=1}^{n}(X_i - \mathbb{E}[X])\right|^2\right]^{p/2}$$

$$\le n^{-p}n^{p/2}\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]^{p/2} \le n^{-1/2}\operatorname{Var}[X]^{p/2}$$

$$\le n^{-1/2}(1 + \operatorname{Var}[X]).$$

This proves the claims. □

Lemma 2. *Let $n \in \mathbb{N}$, and let $X_1, \ldots, X_n$ be i.i.d. nonnegative random variables distributed as X. Let $\alpha > 0$, $\beta > 0$, $c > 0$, and*

$$\tilde{X}_n(s) := \frac{1}{n}\sum_{i=1}^{n}\exp(-sX_i),$$

for $s \in \mathbb{C}_+$. Then

$$\mathbb{P}\left(\sup_{|t|\le\alpha}\left|\tilde{X}(c + \mathrm{i}t) - \tilde{X}_n(c + \mathrm{i}t)\right| > \beta\right) < 4\left(1 + \frac{8\alpha\mathbb{E}[|X|]}{\beta}\right)\exp(-n\beta^2/18)$$

$$+ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \ge \frac{4}{3}\mathbb{E}[|X|]\right).$$

Proof. One can show that, for all $t, s \in [-\alpha, \alpha]$,

$$\left|\tilde{X}(c + \mathrm{i}t) - \tilde{X}(c + \mathrm{i}s)\right| \le \mathbb{E}\left[\left|1 - \exp(\mathrm{i}(t - s)X)\right|\right],$$

and

$$\left|\tilde{X}_n(c + \mathrm{i}t) - \tilde{X}_n(c + \mathrm{i}s)\right| \le |t - s|\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right|,$$

whereas, for each $t_i \in [-\alpha, \alpha]$,

$$\mathbb{P}\left(\left|\tilde{X}(c + \mathrm{i}t_i) - \tilde{X}(c + \mathrm{i}t_i)\right| > \tfrac{1}{3}\beta\right)$$

$$\le \mathbb{P}\left(\left|\Re\left(\tilde{X}(c + \mathrm{i}t_i) - \tilde{X}(c + \mathrm{i}t_i)\right)\right| > \tfrac{1}{6}\beta\right) + \mathbb{P}\left(\left|\Im\left(\tilde{X}(c + \mathrm{i}t_i) - \tilde{X}(c + \mathrm{i}t_i)\right)\right| > \tfrac{1}{6}\beta\right)$$

$$\le 4\exp(-2n\beta^2/36),$$

using Hoeffding's inequality. The claim then follows along precisely the same lines as the proof of [12], Theorem 1. □

Lemma 3. *Let $w > 0$, $c > 0$, and let $f : [0, \infty) \to [0, 1]$ be a continuously differentiable function, twice differentiable in the point w, and such that $\int_w^\infty |f'(y+w)|e^{-cy}y^{-1}\,dy < \infty$. There exists a $\kappa_4 > 0$ such that, for all $m > 0$,*

$$\left| \int_{|y|>m} \frac{1}{2\pi} e^{(c+iy)w} \bar{f}(c+iy)\,dy \right| \le \frac{\kappa_4}{m}.$$

Proof. Fix $m > 0$. Observe that

$$\int_{|y|\le m} \frac{1}{2\pi} e^{(c+iy)w} \bar{f}(c+iy)\,dy$$

$$= \int_{|y|\le m} \frac{1}{2\pi} e^{(c+iy)w} \int_0^\infty e^{-(c+iy)x} f(x)\,dx\,dy$$

$$= \int_0^\infty \frac{1}{2\pi} f(x) \int_{|y|\le m} e^{(c+iy)(w-x)}\,dy\,dx \qquad (24)$$

$$= \int_0^\infty \frac{1}{\pi} f(x) e^{c(w-x)} \frac{\sin(m(w-x))}{w-x}\,dx$$

$$= \int_{-w}^\infty \frac{1}{\pi} f(y+w) e^{-cy} \frac{\sin(my)}{y}\,dy,$$

using Fubini's theorem and the variable substitution $y := x - w$, together with the obvious identity $\sin(-my)/(-y) = \sin(my)/y$.

We consider the integral (24) separately over the domain $[w, \infty)$ and $[-w, w]$. For the interval $[w, \infty)$, we have

$$\left| \int_w^\infty \frac{1}{\pi} \frac{f(y+w)e^{-cy}}{y} \sin(my)\,dy \right|$$

$$\le \left| \left[\frac{1}{\pi} \frac{f(y+w)e^{-cy}}{y} \frac{\cos(my)}{-m} \right]_{y=w}^\infty \right| + \left| \int_w^\infty \frac{1}{\pi} \frac{\partial}{\partial y} \left[\frac{f(y+w)e^{-cy}}{y} \right] \frac{\cos(my)}{m}\,dy \right|$$

$$\le \left| \frac{1}{\pi} \frac{f(w+w)e^{-cw}}{w} \frac{\cos(mw)}{m} \right| \qquad (25)$$

$$+ \int_w^\infty \frac{1}{\pi} |f'(y+w)|e^{-cy}y^{-1}\frac{1}{m}\,dy + \int_w^\infty \frac{1}{\pi} f(y+w)(ce^{-cy}y^{-1}+e^{-cy}y^{-2})\frac{1}{m}\,dy$$

$$\le \frac{1}{m} \cdot \left(\frac{e^{-cw}}{w\pi} + \frac{1}{\pi} \int_w^\infty |f'(y+w)|e^{-cy}y^{-1}\,dy + \frac{e^{-cw}}{\pi w} + \frac{e^{-cw}}{\pi c w^2} \right).$$

We now consider the integral (24) on the interval $[-w, w]$. Write $\phi(y) := f(y+w)e^{-cy}$ and $g(y) := (\phi(y) - \phi(0) - \phi'(0)y)/y$, and observe that g is continuously differentiable on the

interval $[-w, w]$ (which follows from the fact that $f''(w)$ exists). We have

$$\left| \phi(0) - \int_{-w}^{w} \frac{1}{\pi} f(y+w) e^{-cy} \frac{\sin(my)}{y} \, dy \right|$$

$$= \left| \phi(0) - \int_{-w}^{w} \frac{1}{\pi} \big(\phi(0) + \phi'(0)y + g(y)y \big) \frac{\sin(my)}{y} \, dy \right| \tag{26}$$

$$\leq \phi(0) \left| 1 - \int_{-w}^{w} \frac{1}{\pi} \frac{\sin(my)}{y} \, dy \right| + \left| \int_{-w}^{w} \frac{1}{\pi} g(y) \sin(my) \, dy \right|;$$

realize that $\int_{-w}^{w} \pi^{-1} \phi'(0) \sin(my) \, dy = 0$.

We first bound the first term of (26)

$$\left| 1 - \int_{-w}^{w} \frac{\sin(my)}{\pi y} \, dy \right| \leq \left| 1 - \int_{-\infty}^{\infty} \frac{\sin(my)}{\pi y} \, dy \right| + \left| \int_{w}^{\infty} \frac{2\sin(my)}{\pi y} \, dy \right| = \left| \int_{w}^{\infty} \frac{2\sin(my)}{\pi y} \, dy \right|.$$

Write $h(a) := \int_{w}^{\infty} e^{-ay} y^{-1} \sin(my) \, dy$, $a \geq 0$. Then $\lim_{a \to \infty} h(a) = 0$,

$$h'(a) = \int_{w}^{\infty} -e^{-ay} \sin(my) \, dy$$

$$= -e^{-aw} \int_{0}^{\infty} e^{-ax} \sin\big(m(x+w) \big) \, dx = -e^{-aw} \frac{m\cos(wm) + a\sin(wm)}{a^2 + m^2},$$

and thus

$$\left| \int_{w}^{\infty} \frac{\sin(my)}{y} \, dy \right| = \big| h(0) \big| = \left| \lim_{a \to \infty} h(a) - \int_{0}^{\infty} h'(a) \, da \right|$$

$$= \left| \int_{0}^{\infty} e^{-aw} \frac{m\cos(wm) + a\sin(wm)}{a^2 + m^2} \, da \right|$$

$$\leq \int_{0}^{\infty} e^{-aw} \frac{m+a}{a^2 + m^2} \, da \leq \frac{2}{m} \int_{0}^{\infty} e^{-aw} \, da = \frac{2}{mw},$$

which implies

$$\left| 1 - \int_{-w}^{w} \frac{\sin(my)}{\pi y} \, dy \right| \leq \frac{4}{w\pi m}. \tag{27}$$

The second term of (26) is bounded by

$$\left| \int_{-w}^{w} \frac{1}{\pi} g(y) \sin(my) \, dy \right|$$

$$= \left| \frac{1}{\pi} g(w) \frac{\cos(-mw)}{m} - \frac{1}{\pi} g(-w) \frac{\cos(mw)}{m} - \int_{-w}^{w} \frac{1}{\pi} g'(y) \frac{\cos(-my)}{m} \, dy \right| \tag{28}$$

$$\leq \frac{|g(w) - g(-w)|}{\pi m} + \frac{1}{\pi m} \int_{-w}^{w} \big| g'(y) \big| \, dy.$$

Combining (24), (25), (26), (27) and (28), using $f(w) = \phi(0)$, it follows that

$$
\left| \int_{|y|>m} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \overline{f}(c+\mathrm{i}y) \, \mathrm{d}y \right|
$$

$$
= \left| f(w) - \int_{|y|\leq m} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \overline{f}(c+\mathrm{i}y) \, \mathrm{d}y \right|
$$

$$
= \left| f(w) - \int_{-w}^{w} \frac{1}{\pi} f(y+w) e^{-cy} \frac{\sin(my)}{y} \, \mathrm{d}y - \int_{w}^{\infty} \frac{1}{\pi} f(y+w) e^{-cy} \frac{\sin(my)}{y} \, \mathrm{d}y \right|
$$

$$
\leq f(w) \frac{4}{\pi m w} + \frac{|g(w) - g(-w)|}{\pi m} + \frac{1}{\pi m} \int_{-w}^{w} |g'(y)| \, \mathrm{d}y
$$

$$
+ \frac{1}{m} \cdot \left(\frac{e^{-cw}}{w\pi} + \frac{1}{\pi} \int_{w}^{\infty} |f'(y+w)| e^{-cy} y \, \mathrm{d}y + \frac{e^{-cw}}{\pi w} + \frac{e^{-cw}}{\pi c w^2} \right).
$$

Defining

$$
\kappa_4 := f(w) \frac{4}{\pi w} + \frac{|g(w) - g(-w)|}{\pi} + \frac{1}{\pi} \int_{-w}^{w} |g'(y)| \, \mathrm{d}y
$$

$$
+ \frac{e^{-cw}}{w\pi} + \frac{1}{\pi} \int_{w}^{\infty} |f'(y+w)| e^{-cy} y \, \mathrm{d}y + \frac{e^{-cw}}{\pi w} + \frac{e^{-cw}}{\pi c w^2},
$$

this implies the stated of the lemma. □

6. Numerical illustration

We provide a numerical illustration of the performance of our estimator, inspired by an application of estimating high-load probabilities in communication links [11]. In particular, we consider an M/G/1 queue in stationarity that serves jobs at unit speed, and whose (unknown) service time distribution is exponential with mean $1/20$. We choose the (unknown) arrival rate λ from $\{10, 18, 19\}$; this corresponds to load factors ρ of 0.50, 0.90, and 0.95. For $n = 10\,000$ consecutive time intervals of length $\delta = 0.10$, the amount of work arriving to the queue in each interval is recorded. Based on these samples, we estimate the tail probabilities $\mathbb{P}(Y > w)$ of the workload distribution Y for different values of w, using the Laplace-transform based estimator outlined in Section 4.1. We test values of w corresponding to the 90th, 99th, and 99.9th percentile of Y; the particular values, denoted by $w_{0.9}$, $w_{0.99}$, and $w_{0.999}$, are given in Table 1.

For each $\rho \in \{0.50, 0.90, 0.95\}$ and each of the three corresponding values of w, we run 1000 simulations and record the relative estimation error

$$
\left| \frac{(1 - F_n^Y(w)) - \mathbb{P}(Y > w)}{\mathbb{P}(Y > w)} \right|, \tag{29}
$$

Table 1. 90th, 99th, and 99.9th percentiles of W, for different values of ρ

ρ	$w_{0.9}$	$w_{0.99}$	$w_{0.999}$
0.50	0.1609	0.3912	0.6215
0.90	1.0986	2.2499	3.4012
0.95	2.2513	4.5539	6.8565

where $F_n^Y(w)$ denotes the outcome of the Laplace-transform based estimator. The simulation average of (29), for different values of ρ and w, is reported in Table 2, at the lines starting with "Laplace."

We compare the performance of the Laplace-transform based estimator to that of the empirical estimator that samples the workload $W(i\delta)$ at time points $i\delta$, $i = 1, \ldots, n$, and estimates the tail probability $\mathbb{P}(Y > w)$ by the fraction $n^{-1} \sum_{i=1}^{n} \mathbf{1}_{Y(i\delta) > w}$. The corresponding simulation average of the relative estimation error is reported in Table 2, at the lines starting with "Empirical."

Table 2 shows that the Laplace-transform based estimator has a lower relative error than the empirical estimator, for all-but-one tested instances of ρ and w. This is perhaps not surprising, since the "Laplace" estimator is based on i.i.d. samples (of the amount of work arriving to the queue in δ time units), whereas the "Empirical" estimator is based on correlated samples (of the workload in the queue).

A third estimator, that is based on the same samples as the "Empirical" estimator, can be constructed as follows: consider the samples of the workload process $Y(i\delta)$, $i = 1, \ldots, n$, and let $Q = \{Y(i\delta) - (Y((i-1)\delta) - \delta) | Y((i-1)\delta) \geq \delta, 2 \leq i \leq n\}$. If, for some i, $Y((i-1)\delta) \geq \delta$, then the amount of work arrived in the δ time units prior to time point $i\delta$ is precisely equal to $Y(i\delta) - (Y((i-1)\delta) - \delta)$. (If $Y((i-1)\delta) < \delta$, then the exact amount of work arrived between

Table 2. Average relative estimation error

Estimator	$w = w_{0.9}$	$w = w_{0.99}$	$w = w_{0.999}$
		$\rho = 0.50$	
Laplace	0.05	0.13	0.25
Empirical	0.50	0.50	0.67
Laplace, censored	0.15	0.39	0.67
		$\rho = 0.90$	
Laplace	0.19	0.40	0.65
Empirical	0.29	0.96	1.82
Laplace, censored	0.23	0.49	0.81
		$\rho = 0.95$	
Laplace	0.39	0.96	2.09
Empirical	0.52	1.36	1.83
Laplace, censored	0.43	1.07	2.34

time points $(i - 1)\delta$ and $i\delta$ can not be inferred from the workload samples). If we apply the Laplace-transform based estimator on the samples in the set Q (which are independent samples from the amount of work arriving to the queue in δ time units), then we obtain an estimate of $\mathbb{P}(Y > w)$ that is based on the same samples as the "Empirical" estimator. The relative estimation error of this third estimator is reported in Table 2, at the lines starting with "Laplace, censored."

Table 2 shows that the "Laplace, censored" estimator still outperforms the "Empirical" estimator, in all-but-one instances. Both these estimators are based on the same samples of the workload process. A notable disadvantage of "Empirical" estimator is that it requires the system to reach high load in order to obtain informative estimates. In practice, particularly in the context of operated communication links, this is not desirable: network operators would certainly intervene if the network load reaches exceedingly high levels. These interventions hamper the estimation of the probability that this high load occurs. In contrast, both the Laplace-transform based estimators produce informative estimates of $\mathbb{P}(Y > w)$, even if all sampled values of the workload process are below w.

7. Discussion, concluding remarks

In this paper, we have discussed a technique to estimate the distribution of a random variable Y, focusing on the specific context in which we have i.i.d. observations $X_1, \ldots, X_n$, distributed as a random variable X, where the relation between the Laplace transforms of X and Y is known. Our problem was motivated from a practical question of an internet service provider, who wished to develop statistically sound techniques to estimate the packet delay distribution based on various types of probe measurements; specific quantiles of the delay distribution are mutually agreed upon by the service provider and its customers, and posted in the service level agreement. To infer whether these service level agreements are met, the internet provider estimates several tail probabilities of the delay distribution. This explains why we have focused on the setup presented in our paper, concentrating on estimating the distribution function $F^Y(w)$ and bounding the error $\mathbb{E}[|F_n^Y(w) - F^Y(w)|]$ for this w. It is noted that various other papers focus on estimating the density, and often use different convergence metrics; some establish asymptotic Normality.

A salient feature of our analysis is that the ill-posedness of Laplace inversion, that is, the fact that the inverse Laplace transform operator is not continuous, does not play a rôle. Our estimate $F_n^Y(w)$ is "close" to $F^Y(w)$ if the Laplace transform $\bar{F}_n^Y$ is "close" to the Laplace transform $\bar{F}^Y$, measuring "closeness" of these Laplace transforms by the integral (8). Our assumptions (A1)–(A3) ensure that this integral converges to zero (as n grows large), and Section 4 shows that these conditions are met in practical applications. We therefore do not need regularized inversion techniques as in [20] and [23], with convergence rates of just $1/\log(n)$. (See further Remark 1).

Acknowledgments

literature references. The constructive comments and suggestions of the anonymous referees have improved the paper, and are kindly acknowledged. Part of this work was done while the first author was affiliated with Eindhoven University of Technology and University of Amsterdam.

References

[1] Antunes, N. and Pipiras, V. (2011). Probabilistic sampling of finite renewal processes. *Bernoulli* **17** 1285–1326. MR2854773

[2] Baccelli, F., Kauffmann, B. and Veitch, D. (2009). Inverse problems in queueing theory and Internet probing. *Queueing Syst*. **63** 59–107. MR2576007

[3] Bøgsted, M. and Pitts, S.M. (2010). Decompounding random sums: A nonparametric approach. *Ann. Inst. Statist. Math*. **62** 855–872. MR2669741

[4] Buchmann, B. and Grübel, R. (2003). Decompounding: An estimation problem for Poisson random sums. *Ann. Statist*. **31** 1054–1074. MR2001642

[5] Buchmann, B. and Grübel, R. (2004). Decompounding Poisson random sums: Recursively truncated estimates in the discrete case. *Ann. Inst. Statist. Math*. **56** 743–756. MR2126809

[6] Chung, K.L. (2001). *A Course in Probability Theory*, 3rd ed. San Diego, CA: Academic Press. MR1796326

[7] Cohen, J.W. (1982). *The Single Server Queue*, 2nd ed. *North-Holland Series in Applied Mathematics and Mechanics* **8**. Amsterdam: North-Holland. MR0668697

[8] Comte, F., Duval, C. and Genon-Catalot, V. (2014). Nonparametric density estimation in compound Poisson processes using convolution power estimators. *Metrika* **77** 163–183. MR3152023

[9] Comte, F., Duval, C., Genon-Catalot, V. and Kappus, J. (2015). Estimation of the jump size density in a mixed compound Poisson process. *Scand. J. Stat*. **42** 1023–1044. MR3426308

[10] Courcoubetis, C., Kesidis, G., Ridder, A., Walrand, J. and Weber, R. (1995). Admission control and routing in ATM networks using inferences from measured buffer occupancy. *IEEE Trans. Commun*. **43** 1778–1784.

[11] den Boer, A.V., Mandjes, M.R.H., Núñez Queija, R. and Żuraniewski, P.W. (2014). Efficient model-based estimation of network load. Unpublished manuscript.

[12] Devroye, L. (1994). On the non-consistency of an estimate of Chiu. *Statist. Probab. Lett*. **20** 183–188. MR1294102

[13] Doetsch, G. (1974). *Introduction to the Theory and Application of the Laplace Transformation*. New York: Springer. MR0344810

[14] Duval, C. (2013). Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl*. **123** 3963–3986. MR3091096

[15] Glynn, P.W. and Torres, M. (1996). Nonparametric estimation of tail probabilities for the single-server queue. In *Stochastic Networks* (*New York*, 1995) (P. Glasserman, K. Sigman and D.D. Yao, eds.). *Lecture Notes in Statist*. **117** 109–138. New York: Springer. MR1466785

[16] Hall, P. and Park, J. (2004). Nonparametric inference about service time distribution from indirect measurements. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **66** 861–875. MR2102469

[17] Hansen, M.B. and Pitts, S.M. (2006). Nonparametric inference from the $M/G/1$ workload. *Bernoulli* **12** 737–759. MR2248235

[18] Kendall, D.G. (1951). Some problems in the theory of queues. *J. Roy. Statist. Soc. Ser. B*. **13** 151–173; discussion: 173–185. MR0047944

[19] Mandjes, M. and van de Meent, R. Resource dimensioning through buffer sampling. *IEEE/ACM Transactions on Networking* **17** 1631–1644.

[20] Mnatsakanov, R., Ruymgaart, L.L. and Ruymgaart, F.H. (2008). Nonparametric estimation of ruin probabilities given a random sample of claims. *Math. Methods Statist.* **17** 35–43. MR2400363

[21] Nazarathy, Y. and Pollet, P.K. (2012). Parameter and state estimation in queues and related stochastic models: A bibliography. Available at http://www.maths.uq.edu.au/ pkp/papers/Qest/QEstAnnBib.pdf.

[22] Schiff, J.L. (1999). *The Laplace Transform*: *Theory and Applications*. *Undergraduate Texts in Mathematics*. New York: Springer. MR1716143

[23] Shimizu, Y. (2012). Non-parametric estimation of the Gerber–Shiu function for the Wiener–Poisson risk model. *Scand. Actuar. J.* **1** 56–69. MR2881947

[24] van Es, B., Gugushvili, S. and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli* **13** 672–694. MR2348746

[25] Zeevi, A. and Glynn, P.W. (2004). Estimating tail decay for stationary sequences via extreme values. *Adv. in Appl. Probab.* **36** 198–226. MR2035780