Adv. Appl. Prob. **46,** 790–811 (2014)
Printed in Northern Ireland

NONPARAMETRIC INFERENCE FOR QUEUEING NETWORKS OF GEOM^X/G/∞ QUEUES IN DISCRETE TIME

DOMINIC EDELMANN * ** AND
CORNELIA WICHELHAUS,* *** *University of Heidelberg*

Abstract

We study nonparametric estimation problems for discrete-time stochastic networks of Geom^X/G/∞ queues. We assume that we are only able to observe the external arrival and external departure processes at the nodes over a stretch of time. Based on such incomplete information of the system, we aim to construct estimators for the unknown general service time distributions at the nodes without imposing any parametric condition. We propose two different estimation approaches. The first approach is based on the construction of a so-called sequence of differences, and a crucial relation between the expected number of external departures at a node and specific sojourn time distributions in the network. The second approach directly utilizes the structure of the cross-covariance functions between external arrival and departure processes at the nodes. Both methods lead to deconvolution problems which we solve explicitly. A detailed simulation study illustrates the numerical performances of our estimators and shows their advantages and disadvantages.

Keywords: Cross-covariance; discrete deconvolution; discrete-time queueing system; nonparametric statistics for queues; sojourn time estimation

2010 Mathematics Subject Classification: Primary 60K25
Secondary 62M05

1. Introduction

The overall primary aim of research in queueing theory is to improve the design and performance of given real-life stochastic systems. Since in practical applications there are usually processes or parameters which cannot be observed, there is great interest in statistical inference for system characteristics depending on incomplete information of the stochastic networks, such as, for example, idle or busy server periods, the workload processes, or sequences of the arrival and departure points of customers. In the focus of interest are the service time distributions at the nodes since they determine the performance, reliability, and efficiency of the networks. In the IT context Liu *et al.* (2006, p. 41) formulated this issue as follows: 'One of the biggest challenges in modeling complex IT systems using queueing models consists in the calibration of the queueing network parameters, such as the service requirements (…) at each station'. Moreover, the statistical analysis of given observations is essential for an appropriate probability modeling and analysis.

There is interest in both parametric and nonparametric statistical methods for the estimation of the service time distributions. From the mathematical point of view we are faced with

Received 12 June 2012; revision received 8 October 2013.
* Postal address: Institute of Applied Mathematics, University of Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany.
** Email address: dominic.edelmann@googlemail.com
*** Email address: wichelhaus@statlab.uni-heidelberg.de

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

complex identification and estimation problems which often lead to inverse problems of deconvolution. Despite the great need and use for the applications, the statistical theory for queueing systems has not been well developed yet. Therefore, most of the current applied practice is focused only on the estimation of the first- and second-order moments of the service processes; see, e.g. Liu *et al.* (2006) and Ke and Chu (2006).

To the best of our knowledge, the only published result for the important case of statistical inference for stochastic networks of two or more nodes where dependencies between the components have to be taken into consideration is Wichelhaus and Langrock (2012). All other results in the mathematical literature are for single nodes only. Important contributions to parametric estimation problems for queues are, among others, the papers Bhat and Subba Rao (1986/87), Conti (1999), Liu *et al.* (2006), and Pickands and Stine (1997). For nonparametric approaches, we cite Bingham and Pitts (1999), Brown (1970), Conti (2002), Hall and Park (2004), Hansen and Pitts (2006), and Pitts (1994). For a detailed literature overview, we refer the reader to Wichelhaus and Langrock (2012).

In this paper we consider discrete-time stochastic networks of a finite number of nodes with infinite server queues, general service time distributions and general batch arrival processes. We assume that we are only able to observe the external arrivals and external departures of the indistinguishable customers. Movements in the network are not observable. Direct matching of arrivals to departures is not possible. Our aim is to construct nonparametric estimators for the service time distributions at the nodes. These models, apart from being of theoretical interest, are of wide practical interest, since they can fairly adequately represent several real-life systems, such as, for example, transfer systems for data packages like integrated service digital networks, communication systems, nets of nerves in neural science, and production systems.

We propose two estimation approaches which lead to uniform, strongly consistent estimators of the service time distributions at the nodes. The fact that we can recover the basic characteristics of the networks from the external processes only is nontrivial. Note that, from the output process of a simple stationary M/M/N system, only the rate of the Poisson arrival process can be determined; the output process contains no information about the service times (not even the mean) or the number N of servers at the nodes (for details, see Daley (1976)). In our first approach we construct a so-called sequence of differences which leads to explicit relations between conditional sojourn time distributions in the network. We then arrive at discrete deconvolution problems for the service time distributions under study which we can solve explicitly. The underlying basic idea first appeared in the case of a single-node queue in continuous time in Brown (1970). For our second approach, we directly compute the cross-covariances between adequate external arrival and departure processes. Utilizing the structures behind, again leads to deconvolution problems which involve the service time distributions under study. In both cases we prove the uniform strong consistency of the estimators.

We first develop our statistical approaches and present results in the framework of stochastic networks of two nodes (Sections 2–5). Generalizations to networks with more than two nodes are shown in Section 6. In a detailed simulation study in Section 7 we compare the estimators and show their advantages and disadvantages.

We denote the set of positive integers by $\mathbb{N} := \{1, 2, 3, \dots\}$, and set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Throughout this paper, we will mostly study nonnegative integer-valued random variables with probability mass functions f on $\mathbb{N}_0$. For our purposes, it will be convenient to define the corresponding distribution functions F on $\mathbb{N}_0$, rather than on $\mathbb{R}$, by $F\colon \mathbb{N}_0 \to [0, 1], k \mapsto \sum_{i=0}^{k} f(i)$. Analogously, we call a function $F\colon \mathbb{N}_0 \to [0, 1]$ a distribution function on $\mathbb{N}_0$ if it is monotonically increasing and $\lim_{k\to\infty} F(k) = 1$.

A finite mass function h on $\mathbb{N}_0$ is defined by the two conditions $h(k) \geq 0$ for all $k \in \mathbb{N}_0$ and $\sum_{k\in\mathbb{N}_0} h(k) < \infty$. For two finite mass functions f and g on $\mathbb{N}_0$, we define the convolution $(f * g)(\cdot)$ on $\mathbb{N}_0$ by $(f * g)(k) := \sum_{l=0}^{k} f(k-l)g(l)$ for $k \in \mathbb{N}_0$. In the case of probability mass functions f and g, the convolution of the corresponding distribution functions F and G on $\mathbb{N}_0$ is defined by $(F * G)(k) := \sum_{i=0}^{k} (f * g)(i) = \sum_{l=0}^{k} F(k-l)g(l)$ for $k \in \mathbb{N}_0$.

2. Networks of two Geom^X/G/∞ queues

Let us consider a stationary, discrete-time stochastic network with two nodes of general topology. There are external arrival processes $\boldsymbol{A}_1 = (A_1(t))_{t\in\mathbb{Z}}$ and $\boldsymbol{A}_2 = (A_2(t))_{t\in\mathbb{Z}}$ at nodes 1 and 2, respectively, such that at each time slot t, $A_i(t) \geq 0$ indistinguishable customers arrive and enter node i from the outside, $i = 1, 2$. Both $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are nonnegative, integer-valued independent and identically distributed (i.i.d.) sequences. The distribution function of $A_i(0)$ on $\mathbb{N}_0$ will be denoted by F_i, $i = 1, 2$. We assume that $\mathbb{E}[A_i(0)] < \infty$ for $i = 1, 2$. At each node there are infinitely many servers so that no waiting occurs. The service times of customers at node i are distributed according to a distribution function G_i with related probability mass function g_i on $\mathbb{N}_0$. We assume that, with probability 1, mean sojourn times at the nodes are finite and customers stay at least one time slot in the network, i.e. $\sum_{k=0}^{\infty} (1 - G_i(k)) < \infty$ and $G_i(0) := 0$ for $i = 1, 2$.

Customers finishing their service at the first node either leave the system (with probability $1 - p \in (0, 1)$) or jump to the second node (with probability p). Analogously, customers finishing their service at the second node are redirected to the first node (with probability $q \in (0, 1)$) or leave the system (with probability $1 - q$). This means that a customer, once having entered the system, can possibly obtain an arbitrary number of consecutive service times at both nodes before leaving the system. We assume that the routeing probabilities p and q are known. We denote by $\boldsymbol{D}_i = (D_i(t))_{t\in\mathbb{Z}}$ the external departure process at node i, i.e. $D_i(t)$ gives the number of customers leaving the system from node i at time slot t. Note that $\boldsymbol{D}_i$ is in general not an i.i.d. sequence, $i = 1, 2$. We assume that at each node departures occur before arrivals take place and that there is no time needed for traveling from one node to another. We prescribe for all processes and random variables a common underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that all routeing decisions, service times, and arrivals are mutually independent.

We assume that we are able to observe the external input processes $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, and the external departure processes $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ over a stretch of time. The movements of customers from node 1 to node 2 and vice versa are not observable. Our aim is to estimate the service time distributions G_1 and G_2, and the input distributions F_1 and F_2. Since the arriving groups can be directly observed, the estimation of the input distributions F_1 and F_2 will be straightforward. However, this is not the case for the service time distributions G_1 and G_2, since we cannot directly assign the external departures to the external arrivals and measure the individual sojourn times of customers in the network. In particular, for a customer leaving the system via $\boldsymbol{D}_1$, we cannot tell if he/she has entered the system via $\boldsymbol{A}_1$ or $\boldsymbol{A}_2$. Thus, at first it is not clear if all parameters are identifiable from the given observations.

We will concentrate here on the estimation of the service time distribution function G_2 at the second node. Owing to the symmetry of the network considered, the estimation of the service time distribution function G_1 can be carried out in the same manner.

For the analysis of the system we introduce some further notation. For all $k \in \mathbb{Z}$, $j \in \mathbb{N}$, and $i = 1, 2$, we denote the node at which the jth customer, who arrives at node i at time slot k, leaves the network by the random variable $E_{k,j}^{(i)}$. Moreover, the random variable $S_{k,j}^{(i)}$ gives the

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

total sojourn time in the system of this specific customer. It is easy to compute

$$\mathbb{P}(E_{0,1}^{(1)} = 1) = \frac{1-p}{1-pq} \quad \text{and} \quad \mathbb{P}(E_{0,1}^{(2)} = 1) = \frac{(1-p)q}{1-pq}.$$

The output processes $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ can be expressed as follows:

$$\begin{aligned} D_1(t) &= \sum_{j=1}^{\infty}\sum_{l=1}^{A_1(t-j)} \mathbf{1}_{\{S_{t-j,l}^{(1)}=j,\ E_{t-j,l}^{(1)}=1\}} + \sum_{j=2}^{\infty}\sum_{l=1}^{A_2(t-j)} \mathbf{1}_{\{S_{t-j,l}^{(2)}=j,\ E_{t-j,l}^{(2)}=1\}}, \\ D_2(t) &= \sum_{j=2}^{\infty}\sum_{l=1}^{A_1(t-j)} \mathbf{1}_{\{S_{t-j,l}^{(1)}=j,\ E_{t-j,l}^{(1)}=2\}} + \sum_{j=1}^{\infty}\sum_{l=1}^{A_2(t-j)} \mathbf{1}_{\{S_{t-j,l}^{(2)}=j,\ E_{t-j,l}^{(2)}=2\}}. \end{aligned} \tag{1}$$

The distribution function $L_1\colon \mathbb{N} \to [0,1]$ for the sojourn time in the network of an arbitrary customer arriving at node 1 under the condition that he/she also leaves at node 1 is defined by

$$\begin{aligned} L_1(x) &:= \mathbb{P}(S_{0,1}^{(1)} \le x \mid E_{0,1}^{(1)} = 1) \\ &= \frac{\sum_{k=0}^{\infty}(pq)^k(1-p)(G_1 * (G_1 * G_2)^{*k})(x)}{(1-p)/(1-pq)} \\ &= (1-pq)\sum_{k=0}^{x}(pq)^k(G_1 * (G_1 * G_2)^{*k})(x). \end{aligned} \tag{2}$$

Analogously, the distribution function $L_2\colon \mathbb{N} \to [0,1]$ for the sojourn time of an arbitrary customer arriving at node 1 under the condition that he/she leaves the system at node 2 is defined by

$$L_2(x) := \mathbb{P}(S_{0,1}^{(1)} \le x \mid E_{0,1}^{(1)} = 2) = (1-pq)\sum_{k=0}^{x}(pq)^k((G_1 * G_2)^{*(k+1)})(x).$$

We define the corresponding probability mass functions $l_1, l_2\colon \mathbb{N} \to [0,1]$ by

$$l_1(x) := \mathbb{P}(S_{0,1}^{(1)} = x \mid E_{0,1}^{(1)} = 1) \quad \text{and} \quad l_2(x) := \mathbb{P}(S_{0,1}^{(1)} = x \mid E_{0,1}^{(1)} = 2) \quad \text{for } x \in \mathbb{N}. \tag{3}$$

We note that

$$L_2 = L_1 * G_2 \quad (\text{and } l_2 = l_1 * g_2,\ \text{respectively}). \tag{4}$$

This relation will turn out to be crucial for both approaches in our statistical analysis.

3. Estimation of the input distribution functions and the routeing probabilities

The estimation of the input distribution functions F_1 and F_2 is straightforward and the next result directly follows from the Glivenko–Cantelli theorem.

Theorem 1. *Let $i = 1, 2$. For every $x \in \mathbb{N}_0$, define $\hat{F}_n^i(x) := (1/n)\sum_{k=0}^{n-1}\mathbf{1}_{\{A_i(k)\le x\}}$. Then $\hat{F}_n^i$ converges almost surely (a.s.) uniformly to F_i, i.e. $\sup_{x\in\mathbb{N}_0}|\hat{F}_n^i(x) - F_i(x)| \to 0$ a.s. as $n \to \infty$.*

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

In Section 2 we prescribed for this paper that both routeing probabilities p and q are known. If only either p or q is known, the statistical analysis done in this paper is still possible since then an estimator for the other routeing probability can be derived. For example, Lemma 1 below implies that

$$q = \frac{\mathbb{E}(D_1(0)) - \mathbb{E}(A_1(0))(1-p)}{\mathbb{E}(D_1(0))p + \mathbb{E}(A_2(0))(1-p)}.$$

Knowing p, owing to Birkhoff's ergodic theorem, a strongly consistent estimator for q can then be constructed in a straightforward way using the empirical means of the observations of the processes $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, and $\boldsymbol{D}_1$. The resulting estimator for q can then be inserted in place of q in the estimators for G_2 which are constructed in this paper. However, if both probabilities p and q are unknown, the given observations of the external processes are not sufficient to identify the routeing probabilities and thus also the service time distributions cannot be estimated.

4. The sequence-of-differences estimator

In this section we present our first estimation method for the service time distribution function G_2 at the second node. For the applicability of the method, we assume that, with probability 1, there are time slots where no customers enter the system via $\boldsymbol{A}_1$, i.e. let $c_1 := \mathbb{P}(A_1(0) = 0) > 0$. The outline of the method is as follows. We define the sequence of differences $(Z_1(t))_{t\in\mathbb{Z}}$ as $Z_1(t) := t - \max\{n < t \mid A_1(n) > 0\}$ for $t \in \mathbb{Z}$. Here $Z_1(t)$ denotes the difference between the point in time t and the nearest external arrival before t at node 1, and can also be written as $Z_1(t) = \sum_{i=1}^{\infty} i\mathbf{1}_{\{A_1(t-i)\geq 1,\, A_1(t-i+1)=0,\ldots,A_1(t-1)=0\}}$. Note that the process $(Z_1(t))_{t\in\mathbb{Z}}$ can be directly computed from the observations. For an interpretation of $\boldsymbol{Z}_1$, we remark that, for a customer departing from the system via node j at time point t_0, $Z_1(t_0)$ gives his/her smallest possible sojourn time in the network under the assumption that he/she has entered the system via node 1, $j = 1, 2$. Thus, so to speak, the sequence $(Z_1(t))_{t\in\mathbb{Z}}$ assigns the observed departure points at both nodes to the observed arrival points at node 1 in an obviously wrong way. Surprisingly, it will turn out (see Theorem 2 below) that, for all $x \in \mathbb{N}$, the following explicit relations hold:

$$\frac{\mathbb{E}[D_1(0)\mathbf{1}_{\{Z_1(0)\leq x\}}]}{\mathbb{E}[D_1(0)]} = 1 - c_1^x\bigg(1 - \frac{1-p}{1-pq}\frac{\mathbb{E}[A_1(0)]}{\mathbb{E}[D_1(0)]}L_1(x)\bigg) \tag{5}$$

$$\text{and}\quad \frac{\mathbb{E}[D_2(0)\mathbf{1}_{\{Z_1(0)\leq x\}}]}{\mathbb{E}[D_2(0)]} = 1 - c_1^x\bigg(1 - \frac{(1-q)p}{1-pq}\frac{\mathbb{E}[A_1(0)]}{\mathbb{E}[D_2(0)]}L_2(x)\bigg). \tag{6}$$

Based on the observation of the processes $\boldsymbol{A}_1$, $\boldsymbol{D}_1$, and $\boldsymbol{D}_2$, consistent estimators for the ratios on the left-hand sides of the equations can be constructed. This will lead to consistent estimators for $L_1(x)$ and $L_2(x)$. Since $L_2 = L_1 * G_2$ (see (4)), deconvolution methods from Appendix A can be applied to derive estimators for G_2. We start with the expectation of $\boldsymbol{D}_1$.

Lemma 1. *The expectation of the external output process $\boldsymbol{D}_1$ at the first node is*

$$\mathbb{E}[D_1(0)] = \frac{1-p}{1-pq}(\mathbb{E}[A_1(0)] + q\mathbb{E}[A_2(0)]).$$

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

Proof. By Wald's equation we have

$$\begin{aligned}\mathbb{E}[D_1(0)] &= \mathbb{E}\Bigg[\sum_{k=1}^{\infty}\sum_{j=1}^{A_1(-k)}\mathbf{1}_{\{E^{(1)}_{-k,j}=1,\ S^{(1)}_{-k,j}=k\}}+\sum_{k=2}^{\infty}\sum_{j=1}^{A_2(-k)}\mathbf{1}_{\{E^{(2)}_{-k,j}=1,\ S^{(2)}_{-k,j}=k\}}\Bigg]\\
&=\sum_{k=1}^{\infty}\mathbb{E}[A_1(-k)]\mathbb{P}(E^{(1)}_{-k,1}=1,\ S^{(1)}_{-k,1}=k)\\
&\quad+\sum_{k=2}^{\infty}\mathbb{E}[A_2(-k)]\mathbb{P}(E^{(2)}_{-k,1}=1,\ S^{(2)}_{-k,1}=k)\\
&=\mathbb{E}[A_1(0)]\frac{1-p}{1-pq}\sum_{k=1}^{\infty}\mathbb{P}(S^{(1)}_{0,1}=k\mid E^{(1)}_{0,1}=1)\\
&\quad+\mathbb{E}[A_2(0)]\frac{(1-p)q}{1-pq}\sum_{k=2}^{\infty}\mathbb{P}(S^{(2)}_{0,1}=k\mid E^{(2)}_{0,1}=1)\\
&=\frac{1-p}{1-pq}\mathbb{E}[A_1(0)]+\frac{(1-p)q}{1-pq}\mathbb{E}[A_2(0)]\\
&=\frac{1-p}{1-pq}(\mathbb{E}[A_1(0)]+q\mathbb{E}[A_2(0)]).\end{aligned}$$

Remark 1. Owing to the symmetry of the system,

$$\mathbb{E}[D_2(0)]=\frac{1-q}{1-pq}(\mathbb{E}[A_2(0)]+p\mathbb{E}[A_1(0)]).$$

Theorem 2. *For every $x\in\mathbb{N}$, (5) and (6) hold.*

The proof of Theorem 2 is lengthy and postponed to Appendix B.

For formal convenience, we define, for every $x\in\mathbb{N}$, the abbreviations

$$H_1(x):=1-c_1^x\bigg(1-\frac{(1-p)\mathbb{E}[A_1(0)]}{(1-pq)\mathbb{E}[D_1(0)]}L_1(x)\bigg)$$

$$\text{and}\quad H_2(x):=1-c_1^x\bigg(1-\frac{(1-q)p\mathbb{E}[A_1(0)]}{(1-pq)\mathbb{E}[D_2(0)]}L_2(x)\bigg).$$

Then it follows that $\mathbb{E}[D_j(0)\mathbf{1}_{\{Z_1(0)\le x\}}]=\mathbb{E}[D_j(0)]H_j(x)$ for all $x\in\mathbb{N}$ and $j=1,2$.

We now construct estimators for the sojourn time distribution functions $L_1(\cdot)$ and $L_2(\cdot)$. The first step is the estimation of $H_1(\cdot)$ and $H_2(\cdot)$.

Lemma 2. *For every $x\in\mathbb{N}$, we define, for $i=1,2$,*

$$H_n^i(x):=\begin{cases}\dfrac{\sum_{j=0}^{n-1}D_i(j)\mathbf{1}_{\{Z_1(j)\le x\}}}{\sum_{k=0}^{n-1}D_i(k)} & \text{if } \sum_{k=0}^{n-1}D_i(k)\neq 0,\\ 0 & \text{otherwise.}\end{cases}$$

Then, for all $x\in\mathbb{N}$, $H_n^i(x)\to H_i(x)$ a.s. as $n\to\infty$ for $i=1,2$.

Proof. We fix an arbitrary $x\in\mathbb{N}$. First we remark that the sequence $(D_1(j)\mathbf{1}_{\{Z_1(j)\le x\}})_{j\in\mathbb{Z}}$ is stationary since, by the model assumptions, the sequence

$$(A_1(t),A_2(t),(E^{(1)}_{t,j})_{j\in\mathbb{N}},(E^{(2)}_{t,j})_{j\in\mathbb{N}},(S^{(1)}_{t,j})_{j\in\mathbb{N}},(S^{(2)}_{t,j})_{j\in\mathbb{N}})_{t\in\mathbb{Z}}$$

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

is i.i.d and, thus, the distribution of $(D_1(j)\mathbf{1}_{\{Z_1(j)\le x\}})_{j\in\mathbb{Z}}$ does not change by shifting. Furthermore, ergodicity of the sequence $(D_1(j)\mathbf{1}_{\{Z_1(j)\le x\}})_{j\in\mathbb{N}_0}$ can be proved in the same manner as Lemma 1 of Brown (1970). Owing to a generalization of Kolmogorov's 0–1 law, it is sufficient to show that each event of the tail σ-algebra of $(D_1(j)\mathbf{1}_{\{Z_1(j)\le x\}})_{j\in\mathbb{N}_0}$ is independent of $(D_1(j)\mathbf{1}_{\{Z_1(j)\le x\}})_{0\le j\le K}$ for all $K\in\mathbb{N}$. This is ensured by the following two facts. At a time point K there are only $\mathbb{P}$-a.s. finitely many customers in the network since mean sojourn times at the nodes are finite and the system is in its steady state. It takes only $\mathbb{P}$-a.s. finitely many time slots until there is an arrival of $\boldsymbol{A}_1$ after the time point K. Hence, Birkhoff's ergodic theorem and Theorem 2 yield:

$$\frac{1}{n}\sum_{j=0}^{n-1} D_1(j)\mathbf{1}_{\{Z_1(j)\le x\}} \to \mathbb{E}[D_1(0)]H_1(x) \quad \text{a.s. for } n\to\infty,$$

as $\mathbb{E}[D_1(0)]<\infty$ due to Lemma 1. Similarly, it follows that $(1/n)\sum_{k=0}^{n-1} D_1(k)\to\mathbb{E}[D_1(0)]$ a.s., since $(D_1(t))_{t\in\mathbb{N}_0}$ is ergodic. This implies that $H_n^1(x)\to H_1(x)$ a.s. For $H_n^2(x)$, the proof is similar.

For technical reasons, we now have to ensure that, for every $n\in\mathbb{N}$, the estimator for c_1 is greater than 0. Thus, we modify the estimator $\hat{F}_n^1(0)$ from Theorem 1 as follows.

Lemma 3. *Let $c_1>0$. For $n\in\mathbb{N}$, with some $a_0\in(0,1)$ define*

$$\hat{c}_n^1 := \begin{cases} a_0 & \text{if } \hat{F}_n^1(0)=0,\\ \hat{F}_n^1(0) & \text{if } \hat{F}_n^1(0)\ne 0.\end{cases}$$

Then, it holds that $\hat{c}_n^1\to c_1$ a.s. as $n\to\infty$.

Proof. Let $B:=\{\omega\in\Omega\colon \hat{F}_n^1(\omega)(0)\to c_1 \text{ as } n\to\infty\}$. By Theorem 1 we have $P(B)=1$. We fix an arbitrary $\omega\in B$. There is an $N(\omega)$ such that $|\hat{F}_n^1(\omega)(0)-c_1|<c_1$ for all $n\ge N(\omega)$. Hence, it holds that $\hat{F}_n^1(\omega)>0$ for all $n\ge N(\omega)$ and, thus, $\hat{c}_n^1(\omega)=\hat{F}_n^1(\omega)(0)$ for all $n\ge N(\omega)$.

We now note that, for every $x\in\mathbb{N}$, with $\hat{c}_n^1$ as defined in Lemma 3,

$$\hat{L}_n^1(x) := \frac{(1-pq)\sum_{i=0}^{n-1} D_1(i)}{(1-p)\sum_{i=0}^{n-1} A_1(i)}\left(1-(\hat{c}_n^1)^{-x}(1-H_n^1(x))\right)$$

converges a.s. to $L_1(x)$ as $n\to\infty$. Analogously, for every $x\in\mathbb{N}$, $\hat{L}_n^2(x)\to L_2(x)$ a.s., where

$$\hat{L}_n^2(x) := \frac{(1-pq)\sum_{i=0}^{n-1} D_2(i)}{(1-q)p\sum_{i=0}^{n-1} A_1(i)}\left(1-(\hat{c}_n^1)^{-x}(1-H_n^2(x))\right).$$

Since $L_1 * G_2 = L_2$ by (4), it remains to construct an estimator for the probability mass function g_2 by use of the deconvolution methods presented in Appendix A.

Definition 1. We define the estimator $\hat{l}_n^i$ for the probability mass function l_i for $x\ge 1$ by

$$\hat{l}_n^i(x) := \begin{cases} \hat{L}_n^i(x)-\hat{L}_n^i(x-1) & \text{if } \hat{L}_n^i(x)-\hat{L}_n^i(x-1)\ge 0,\\ 0 & \text{otherwise,}\end{cases}$$

where $\hat{l}_n^i(0):=\hat{L}_n^i(0):=0$ for $i=1,2$.

It follows that $\hat{l}_n^i(x)$ is a consistent estimator for $l_i(x)$ for all $x \in \mathbb{N}$ and $i = 1, 2$.

Lemma 4. *There is a nonnegative, integer-valued random function $\hat{M}_n$ which converges a.s. to $M := \min\{x \in \mathbb{N} : l_1(x) > 0\}$ as $n \to \infty$. Furthermore, it holds that $\hat{l}_n^1(\hat{M}_n) > 0$ for all $n \in \mathbb{N}$.*

Note that $M = \min\{x \in \mathbb{N} : l_1(x) > 0\} = \min\{x \in \mathbb{N} : g_1(x) > 0\}$. The following proof explicitly shows how the random function $\hat{M}_n$ can be constructed.

Proof of Lemma 4. We first note that, since mean service times at the nodes are finite by assumption, the mean of the conditional sojourn time distribution L_1 is finite and, thus, there is a finite $x_0 \in \mathbb{N}$ with $l_1(x_0) > 0$. It is obvious that $M \le x_0$. Furthermore, since the estimators $\hat{l}_n^1(x)$ are consistent for all $x \in \mathbb{N}$, we can choose an $\omega \in \Omega$ from a set with probability mass 1 such that, for all $x \in \mathbb{N}$, we have $\hat{l}_n^1(\omega)(x) \to l_1(x)$ as $n \to \infty$. We now fix an arbitrary $\delta > 0$. Then, for all $x \in \mathbb{N}$, there is an $N_{x,\delta}(\omega)$ such that, for all $m, n \ge N_{x,\delta}(\omega)$, we have $|\hat{l}_m^1(\omega)(x) - \hat{l}_n^1(\omega)(x)| < \delta$. With $N_\delta(\omega) := \max_{x \le x_0} N_{x,\delta}(\omega)$ we define $M_{N_\delta(\omega),\delta}(\omega) := \min\{y \le x_0 : \hat{l}_{N_\delta}^1(\omega)(\omega)(y) > \delta \text{ and } \hat{l}_{1/\delta}^1(\omega)(y) > 0\}$, where we set $\min \varnothing := x_0$. Then, $M_{N_\delta(\omega),\delta}(\omega) \ge M$, since $\hat{l}_{N_\delta}^1(\omega)(\omega)(x) \le \delta$ if $l_1(x) = 0$. As $\delta \to 0$, we have $N_\delta(\omega) \to \infty$ and, thus, $M_{N_\delta(\omega),\delta}(\omega) \to M$ as $\delta \to 0$. The proof is complete by setting $\hat{M}_n(\omega) := M_{N_{1/n}(\omega),1/n}(\omega)$.

Remark 2. The second condition $\hat{l}_{1/\delta}^1(\omega)(y) > 0$ in the definition of $M_{N_\delta(\omega),\delta}(\omega)$ in the proof of Lemma 4 is only needed to guarantee that $\hat{l}_n^1(\hat{M}_n) > 0$ for all $n \in \mathbb{N}$, which is necessary for Lemma 5 below.

We can now apply the method of Proposition 1 given in Appendix A to obtain the following result.

Lemma 5. *Recall the definition of $\hat{M}_n$ in the proof of Lemma 4. We further define, for $n \in \mathbb{N}$, $\hat{w}_n(1) := \hat{l}_n^2(\hat{M}_n + 1)/\hat{l}_n^1(\hat{M}_n)$, and $\hat{w}_n(k)$ iteratively as*

$$\hat{w}_n(k) = \frac{\hat{l}_n^2(\hat{M}_n + k) - \sum_{i=1}^{k-1} \hat{l}_n^1(\hat{M}_n + k - i)\hat{w}_n(i)}{\hat{l}_n^1(\hat{M}_n)}, \qquad k \in \mathbb{N}.$$

Then, it holds that, for all $x \in \mathbb{N}$, $\hat{w}_n(x) \to g_2(x)$ a.s. as $n \to \infty$.

Finally, we use Theorem 6 (see Appendix A) to construct an a.s. uniform convergent estimator.

Theorem 3. *Let $(\hat{w}_n)_{n\in\mathbb{N}}$ be defined as in the preceding lemma. Let $(\hat{W}_n)_{n\in\mathbb{N}}$ for all $x \in \mathbb{N}_0$ be defined as $\hat{W}_n(x) = 0$ for $x = 0$ and $\sum_{i=1}^{x} \hat{w}_n(i)$ for $x \in \mathbb{N}$. Then $\hat{W}_n(x)$ converges a.s. to $G_2(x)$ for all $x \in \mathbb{N}_0$. Moreover, define $\hat{G}_n^2$ for all $x \in \mathbb{N}_0$ as $\hat{G}_n^2(x) = \min(\max_{y \le x} \hat{W}_n(y), 1)$. Then $\hat{G}_n^2$ converges a.s. uniformly to G_2.*

Remark 3. For identifiability and estimation of the service time distribution G_2 at the second node, it is sufficient to observe the external processes $\boldsymbol{A}_1$, $\boldsymbol{D}_1$, and $\boldsymbol{D}_2$. Knowledge of the process $\boldsymbol{A}_2$ is not necessary here. In particular, the key relations (5) and (6) strongly depend via c_1 and $Z_1(0)$ on the distribution of the arrival process $\boldsymbol{A}_1$; however, there is no direct influence of the process $\boldsymbol{A}_2$.

5. The cross-covariance estimator

In this section we present our second estimation method for the service time distribution function G_2. We do not need here the assumption that $c_1 = \mathbb{P}(A_1(0) = 0) > 0$, but we

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

do have to impose the condition that $\mathrm{var}[A_1(0)] \in (0, \infty)$. The second estimation approach is based on the explicit calculation of the covariance functions between external arrival and external departure processes of the network. A comparison of the structure of these covariance functions then directly leads to a deconvolution setting which can be solved with the methods presented in Appendix A.

We start with the calculation of the cross-covariance functions.

Lemma 6. *The cross-covariance function* α_1 *between the external input process* $\boldsymbol{A}_1$ *and the external departure process* $\boldsymbol{D}_1$ *at the first node is, for all* $k \in \mathbb{N}$, *given by*

$$\alpha_1(k) = \mathrm{cov}[A_1(0), D_1(k)] = \mathrm{var}[A_1(0)] l_1(k) \frac{1-p}{1-pq},$$

where $l_1(k) = \mathbb{P}(S_{0,1}^{(1)} = k \mid E_{0,1}^{(1)} = 1)$, *as defined in (3).*

Proof. By (1) we have, owing to the mutual independence of the processes $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$,

$$\begin{aligned}\mathrm{cov}[A_1(0), D_1(k)] &= \mathrm{cov}\Bigg[A_1(0), \sum_{j=1}^{\infty} \sum_{l=1}^{A_1(k-j)} \mathbf{1}_{\{S_{k-j,l}^{(1)}=j,\, E_{k-j,l}^{(1)}=1\}}\Bigg] \\ &\quad + \mathrm{cov}\Bigg[A_1(0), \sum_{j=2}^{\infty} \sum_{l=1}^{A_2(k-j)} \mathbf{1}_{\{S_{k-j,l}^{(2)}=j,\, E_{k-j,l}^{(2)}=1\}}\Bigg] \\ &= \mathrm{cov}\Bigg[A_1(0), \sum_{l=1}^{A_1(0)} \mathbf{1}_{\{S_{0,l}^{(1)}=k,\, E_{0,l}^{(1)}=1\}}\Bigg].\end{aligned}$$

By Wald's equation, it follows that

$$\begin{aligned}\mathrm{cov}\Bigg[A_1(0), \sum_{l=1}^{A_1(0)} \mathbf{1}_{\{S_{0,l}^{(1)}=k,\, E_{0,l}^{(1)}=1\}}\Bigg] &= \mathrm{var}[A_1(0)] \mathbb{P}(S_{0,1}^{(1)} = k,\ E_{0,1}^{(1)} = 1) \\ &= \mathrm{var}[A_1(0)] \mathbb{P}(S_{0,1}^{(1)} = k \mid E_{0,1}^{(1)} = 1) \mathbb{P}(E_{0,1}^{(1)} = 1) \\ &= \mathrm{var}[A_1(0)] l_1(k) \frac{1-p}{1-pq}.\end{aligned}$$

Lemma 7. *The cross-covariance function* α_2 *between the external input process* $\boldsymbol{A}_1$ *at the first node and the external departure process* $\boldsymbol{D}_2$ *at the second node is, for all* $k \in \mathbb{N}$, *given by*

$$\alpha_2(k) = \mathrm{cov}[A_1(0), D_2(k)] = \mathrm{var}[A_1(0)] l_2(k) \frac{(1-q)p}{1-pq},$$

where $l_2(k) = \mathbb{P}(S_{0,1}^{(1)} = k \mid E_{0,1}^{(1)} = 2)$ *as defined in (3).*

Proof. The proof is analogous to that of Lemma 6.

The following sample means are strongly consistent estimates due to Birkhoff's ergodic theorem (note that $\boldsymbol{A}_1$ is an i.i.d. sequence by assumption, and ergodicity of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ can be shown as in the proof of Lemma 2):

$$\bar{A}_n^1 := \frac{1}{n} \sum_{i=0}^{n-1} A_1(i), \qquad \bar{D}_n^1 := \frac{1}{n} \sum_{i=0}^{n-1} D_1(i), \qquad \bar{D}_n^2 := \frac{1}{n} \sum_{i=0}^{n-1} D_2(i).$$

With the help of these means, we define the estimators for the cross-covariance functions:

$$\hat{\alpha}_n^j(k) := \frac{1}{n}\sum_{i=0}^{n-k-1}(A_1(i)-\bar{A}_n^1)(D_j(i+k)-\bar{D}_n^j), \qquad k\in\mathbb{N},\ j=1,2.$$

We have $\hat{\alpha}_n^1(k)\to\alpha_1(k)$ a.s. and $\hat{\alpha}_n^2(k)\to\alpha_2(k)$ a.s. for all $k\in\mathbb{N}$.

Now, since $l_2 = l_1 * g_2$, we derive from Lemma 6 and Lemma 7 that

$$\frac{1-p}{(1-q)p}\alpha_2(k) = \alpha_1 * g_2(k), \qquad k\in\mathbb{N}.$$

Since $\alpha_1(k)\ge 0$ and $\sum_{k\in\mathbb{N}}\alpha_1(k) = \operatorname{var}[A_1(0)](1-p)/(1-pq) < \infty$, we can apply the deconvolution techniques from Appendix A. Hence, we can directly derive a uniformly, strongly consistent estimator for G_2 by performing analogous steps as in Lemma 4, Lemma 5, and Theorem 3.

We first note that $M = \min\{x\in\mathbb{N}\colon l_1(x)>0\} = \min\{x\in\mathbb{N}\colon \alpha_1(x)>0\}$. Therefore, in analogy to Lemma 4 we can show the following result.

Lemma 8. *There is a nonnegative, integer-valued random function $\tilde{M}_n$ which converges a.s. to $M = \min\{x\in\mathbb{N}\colon \alpha_1(x)>0\}$ as $n\to\infty$. Furthermore, it holds that $\hat{\alpha}_n^1(\tilde{M}_n)>0$ for all $n\in\mathbb{N}$.*

Next we use the deconvolution techniques (see Proposition 1 given in Appendix A) to obtain a strongly consistent estimator for g_2.

Lemma 9. *Define*

$$\hat{v}_n(1) := \frac{(1-p)\hat{\alpha}_n^2(\tilde{M}_n+1)/(1-q)p}{\hat{\alpha}_n^1(\tilde{M}_n)},$$

and $\hat{v}_n(k)$ iteratively as

$$\hat{v}_n(k) = \frac{(1-p)\hat{\alpha}_n^2(\tilde{M}_n+k)/(1-q)p - \sum_{i=1}^{k-1}\hat{\alpha}_n^1(\tilde{M}_n+k-i)\hat{v}_n(i)}{\hat{\alpha}_n^1(\tilde{M}_n)}, \qquad k\in\mathbb{N}.$$

Then it holds that, for all $x\in\mathbb{N}$, $\hat{v}_n(x)\to g_2(x)$ a.s. as $n\to\infty$.

Finally, we construct a uniform, strongly consistent estimator for G_2 according to Theorem 6 (see Appendix A).

Theorem 4. *Let $(\hat{v}_n)_{n\in\mathbb{N}}$ be defined as in the preceding lemma. Let $(\hat{V}_n)_{n\in\mathbb{N}}$ for all $x\in\mathbb{N}_0$ be defined as $\hat{V}_n(x)=0$ for $x=0$ and $\hat{V}_n(x)=\sum_{i=1}^{x}\hat{v}_n(i)$ for $x\in\mathbb{N}$. Then $\hat{V}_n(x)$ converges a.s. to $G_2(x)$ for all $x\in\mathbb{N}_0$. Moreover, define $\tilde{G}_n^2$ for all $x\in\mathbb{N}_0$ as $\tilde{G}_n^2(x) = \min(\max_{y\le x}\hat{V}_n(y), 1)$. Then $\tilde{G}_n^2$ converges a.s. uniformly to G_2.*

Remark 4. As in the case of the sequence-of-differences estimator for the estimation of the service time function G_2, observations of only the processes $\boldsymbol{A}_1$, $\boldsymbol{D}_1$, and $\boldsymbol{D}_2$ are sufficient. Note that $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ depend on the distribution function F_1, but are independent of F_2.

6. Generalizations to networks of more than two nodes

We consider here a stochastic network of J nodes of general topology. Let $\bar{J} := \{1, 2, \ldots, J\}$. Assume that we are given the routeing matrix and observations of the external arrival and external departure processes at the nodes. The question of this section is: under which conditions is

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

the service time distribution at node k identifiable, i.e. under which conditions on the network topology can we construct a sequence of estimators based on one of the two concepts of this paper (sequence-of-differences or cross-covariance estimator)?

Let us first specify the model. We consider a discrete-time stochastic network with J nodes of general topology. As in the preceding sections, we assume that all processes of the system are defined on a common underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. At each node $i \in \bar{J}$ there is an external arrival process $\boldsymbol{A}_i = (A_i(t))_{t\in\mathbb{Z}}$, where $\boldsymbol{A}_i$ is a nonnegative, integer-valued i.i.d. sequence with the meaning that at each time slot t, $A_i(t) \geq 0$ customers arrive and enter node i from the outside. The distribution function of $A_i(0)$ on $\mathbb{N}_0$ will be denoted by F_i. For technical reasons, with respect to the definition of the service time estimators we further assume that, for all $i \in \bar{J}$, $\mathbb{E}[A_i(0)] < \infty$, $\text{var}[A_1(0)] \in (0, \infty)$, and $c_i := \mathbb{P}(A_i(0) = 0) > 0$. At each node there are infinitely many servers so that no waiting occurs. The service times of customers at node $i \in \bar{J}$ are distributed according to a distribution function G_i on $\mathbb{N}_0$ satisfying $\sum_{s=0}^{\infty}(1 - G_i(s)) < \infty$ and $G_i(0) = 0$. After having been served at node i a customer jumps to node j with probability $r(i, j)$ and leaves the network with probability $r(i, 0) := 1 - \sum_{k=1}^{J} r(i, k)$. We assume that at each node departures occur before arrivals take place and that there is no time needed for traveling from one node to another. We denote by $\boldsymbol{D}_i = (D_i(t))_{t\in\mathbb{Z}}$ the external departure process at node i, i.e. $D_i(t)$ gives the number of customers leaving the system from node i at time t. We assume that all routeing decisions, service times, and arrivals are mutually independent.

Assume that we are able to observe the external input processes $\boldsymbol{A}_i$ $(i = 1, 2, \dots, J)$ and the external departure processes $\boldsymbol{D}_i$ $(i = 1, 2, \dots, J)$ over a stretch of time. Fix some node $k \in \bar{J}$. The input distribution function F_k may be estimated along the same lines as in Section 3. We concentrate on the estimation of the service time distribution function G_k here. Denote by $S_{0,1}^{(i)}$ the total sojourn time of the first customer arriving via A_i at node i at time slot $t = 0$ and denote by $E_{0,1}^{(i)}$ the node at which this customer leaves the network.

Theorem 5. *Consider a general stochastic network of J nodes. Assume that we have observations of the external arrival and external departure processes at the nodes over a stretch of time. The service time distribution at node $k \in \bar{J}$ is identifiable if there exist, for some $r \in \mathbb{N}_0$, nodes $a_0, a_1, \dots, a_r, d_0, d_1, \dots, d_r \in \bar{J}$ with $c_{a_0} < 1$, $c_{a_v} < 1$ and $r(d_0, 0) > 0$, $r(d_v, 0) > 0$ for all $v = 1, \dots, r$, and nodes $i_1, \dots, i_s \in \bar{J} \setminus \{k\}$ for some $s \in \mathbb{N}_0$ such that the service time distributions at the nodes $i_1, \dots, i_s$ are identifiable and, for $x \in \mathbb{N}$,*

$$\mathbb{P}(S_{0,1}^{(a_0)} \leq x \mid E_{0,1}^{(a_0)} = d_0) = K(R) G_{i_1} * \cdots * G_{i_s} * G_k * \mathbb{P}(S_{0,1}^{(a_1)} \leq \cdot \mid E_{0,1}^{(a_1)} = d_1) * \cdots$$
$$* \, \mathbb{P}(S_{0,1}^{(a_r)} \leq \cdot \mid E_{0,1}^{(a_r)} = d_r)(x), \tag{7}$$

where $K(R)$ is a factor just depending on the entries of the routeing matrix $(r(i, j))_{i,j\in\bar{J}\cup\{0\}}$. Moreover, estimators for the service time distribution at node k can then be constructed based on the concept of sequence-of-differences estimators as well as of cross-covariance estimators.

Proof. First, since we can construct estimators for all G_{i_l}, $l = 1, \dots, s$, the same is true for the discrete convolution $G_{i_1} * \cdots * G_{i_s}$. Then, on the one hand, analogously to Lemma 6, it can be shown that, for $v = 0, 1, \dots, r$,

$$\text{cov}[A_{a_v}(0), D_{d_v}(x)] = \text{var}[A_{a_v}(0)] \mathbb{P}(S_{0,1}^{(a_v)} = x \mid E_{0,1}^{(a_v)} = d_v) \mathbb{P}(E_{0,1}^{(a_v)} = d_v) \quad \text{for } x \in \mathbb{N}.$$

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

The probabilities $\mathbb{P}(E_{0,1}^{(a_v)} = d_v)$, $v = 0, 1, \dots, r$, can be calculated from the routeing topology. The quantities $\operatorname{var}[A_{a_v}(0)]$ and $\operatorname{cov}[A_{a_v}(0), D_{d_v}(x)]$, $v = 0, 1, \dots, r$, can be estimated from the observations of the processes A_{a_v} and D_{d_v} in a straightforward way. Thus, all terms in (7) except for G_k can be directly estimated. Furthermore, a consistent estimator for $N := \min\{x \in \mathbb{N} \colon h(x) > 0\}$ with $h(x)$ as the probability mass function of

$$G_{i_1} * \cdots * G_{i_s} * \mathbb{P}(S_{0,1}^{(a_1)} \leq \cdot \mid E_{0,1}^{(a_1)} = d_1) * \cdots * \mathbb{P}(S_{0,1}^{(a_r)} \leq \cdot \mid E_{0,1}^{(a_r)} = d_r)(x)$$

can be constructed as in the proof of Lemma 4. Clearly, $N \geq s + r$. Then, the concept of discrete deconvolution (see Proposition 1 given in Appendix A) yields an estimator for G_k based on cross-covariances.

On the other hand, for the estimation using sequence-of-differences estimators similarly to Theorem 2, it can be shown that, for all $x \in \mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}[D_{d_v}(0)\mathbf{1}_{\{Z_{a_v}(0) \leq x\}}] \\
&\quad = \mathbb{E}[D_{d_v}(0)]\biggl[1 - c_{a_v}^x \biggl(1 - \mathbb{P}\biggl(E_{0,1}^{(a_v)} = d_v\biggr) \frac{\mathbb{E}[A_{a_v}(0)]}{\mathbb{E}[D_{d_v}(0)]} \mathbb{P}(S_{0,1}^{(a_v)} \leq x \mid E_{0,1}^{(a_v)} = d_v)\biggr)\biggr],
\end{aligned}$$

where $(Z_{a_v}(t))_{t \in \mathbb{Z}}$ is the sequence of differences given by $Z_{a_v}(t) := t - \max\{n < t \mid A_{a_v}(n) > 0\}$, $t \in \mathbb{Z}$, $v = 0, 1, \dots, r$. Thus, analogously to Section 5, the sojourn times $\mathbb{P}(S_{0,1}^{(a_v)} \leq x \mid E_{0,1}^{(a_v)} = d_v)$, $v = 0, 1, \dots, r$, can be estimated and then, again, the concept of discrete deconvolution yields an estimator for G_k due to (7).

Example 1. (a) In the case of a single-node system, i.e. $J = 1$, the crucial relation (7) is just $\mathbb{P}(S_{0,1}^{(1)} \leq x \mid E_{0,1}^{(1)} = 1) = G_1(x)$, i.e. $k = 1$, $s = 0$, $r = 0$, $a_0 = d_0 = 1$, and $K(R) = 1$.

(b) In the case of the network of two nodes considered in Sections 2–5, the crucial relation (7) is $L_2(x) = L_1 * G_2(x)$ from (4). Thus, here $k = 2$, $s = 0$, $r = 1$, $a_0 = a_1 = 1$, $d_0 = 2$, $d_1 = 1$, and $K(R) = 1$.

Example 2. Consider a network of two nodes with just one external arrival and departure process given by the parameters $c_1 < 1$, $c_2 = 1$, $r(1, 0) = 1 - p$, $r(1, 2) = p$, and $r(2, 1) = 1$. Then a relation like (7) cannot be established. Also, the service time distributions at both nodes are not identifiable. This is plausible since, based on the observation of the external arrival and departure process only, there is no way to distinguish between the sojourn times at the two different nodes.

The preceding example shows that the question of identifiability of nodes in a general network is highly complex. We were unable to find a general all-encompassing topology of a network such that a relation like (7) holds. The next corollary gives sufficient conditions on the routeing topology such that (7) can be established and the service time distribution at node k is identifiable. These conditions entail rather general routeing topologies of the underlying networks for which our estimation methods may be applied.

Corollary 1. *Consider for a discrete-time network of J nodes the estimation of the service time distribution function G_k at node $k \in \bar{J}$. In the following cases there exist appropriate nodes $a_0, a_1, \dots, a_r, d_0, d_1, \dots, d_r, i_1, \dots, i_s \in \bar{J}$ such that relation (7) for G_k holds and, thus, the distribution function G_k can be estimated.*

1. *$c_k < 1$ and $r(k, 0) > 0$, and node k can only be visited once during a customer's stay in the network.*

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

2. *$r(k,0) > 0$, node k can only be visited once during a customer's stay in the network, and there is a node $l \in \bar{J} \setminus \{k\}$ such that $c_l < 1$, $r(l,k) > 0$, and $r(l,0) > 0$. Furthermore, there is no way from node l to node k other than the direct path.*

3. *$c_k < 1$ and there are nodes $l_1, \dots, l_n \in \bar{J}$ with $n \in \mathbb{N}$ such that $r(k,l_1)r(l_1,l_2)\cdots r(l_{n-1},l_n) > 0$ and $r(l_n,0) > 0$ and all $G_{l_1}, \dots, G_{l_n}$ are identifiable. Furthermore, there is no other path through the network starting at node k and leaving from node l_n.*

4. *$r(k,0) > 0$ and there are nodes $s_1, \dots, s_m \in \bar{J}$ with $m \in \mathbb{N}$ such that $c_{s_1} < 1$ and $r(s_1,s_2)r(s_2,s_3)\cdots r(s_m,k) > 0$ and all $G_{s_1}, \dots, G_{s_m}$ are identifiable. Furthermore, there is no other path through the network starting at node s_1 and leaving from node k.*

5. *There are $m, n \in \mathbb{N}$ and nodes $s_1, \dots, s_m, l_1, \dots, l_n \in \bar{J}$ such that $c_{s_1} < 1$,*

$$r(s_1,s_2)r(s_2,s_3)\cdots r(s_m,k)r(k,l_1)r(l_1,l_2)\cdots r(l_{n-1},l_n) > 0$$

and $r(l_n,0) > 0$ and all $G_{s_1}, \dots, G_{s_m}, G_{l_1}, \dots, G_{l_n}$ are identifiable. Furthermore, there is no other path through the network starting at node s_1 and leaving from node l_n.

Proof. The proof is straightforward. For example, in case 2 we directly derive

$$\mathbb{P}(S_{0,1}^{(l)} \le x \mid E_{0,1}^{(l)} = k) = G_k * \mathbb{P}(S_{0,1}^{(l)} \le x \mid E_{0,1}^{(l)} = l),$$

and in case 5,

$$\mathbb{P}(S_{0,1}^{(s_1)} \le x \mid E_{0,1}^{(s_1)} = l_n) = G_{s_1} * \cdots * G_{s_m} * G_k * G_{l_1} * \cdots * G_{l_n},$$

which are special cases of (7).

Example 3. Consider the network according to Figure 1.

We apply Corollary 1. First, G_1, G_6, and G_9 are identifiable due to criterion 1. According to 2 or 4 the distributions G_2 and G_{10} are identifiable. Next, G_5 and then G_4 are identifiable due to 3, and then G_3 is identifiable using criterion 5. The service time distributions at nodes 7 and 8 are not identifiable.

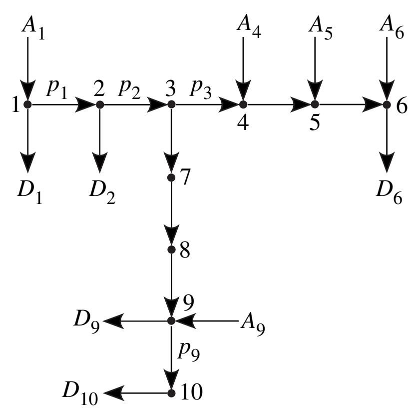

FIGURE 1: Feedforward network of ten nodes.

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

7. Simulations and discussions

To outline the efficiency of our proposed estimators and to compare the exactness of their estimation, we have performed a simulation study using R. In particular, we considered a two-node network with routeing probabilities $p = 0.7$ for jumping from node 1 to node 2, and $q = 0.4$ for jumping from node 2 to node 1. We concentrated on the estimation of the service time distribution G_2 at the second node. We always set a forerun of $5 \cdot 10^5$ time units, so that we could assume that our system had practically reached its steady state. The next 10^6 time units were used to compute the estimates. Our simulations show that our estimation procedures yield good results for a wide range of models. We focus here on Bernoulli and Poisson arrivals as well as on the case of the geometric service time distribution and as a generalization the negative binomial service distribution. These are the most relevant discrete-time models for applications and the most frequently studied ones in the literature. First we illustrate the behavior of the estimators by prototype examples with a balanced ratio between arrival and service rates. Later on, we study the advantages and disadvantages of our two estimation methods with respect to variations of the arrival and service parameters involved. Of particular importance here are the cases of a high arrival rate at node 1 and of large possible service times at node 2. At the end, we compare our two methods with the so-called B-customer estimator, an elementary estimator proposed in Ross (1970). We start with two prototype simulation examples presented in Figure 2. In the left-hand diagram of Figure 1 we present results for arriving Poisson sequences and geometrically distributed service times at both nodes; in the right-hand diagram we present results for arriving Poisson sequences at both nodes, negative binomial service times at the second node, and geometric service times at the first node. For the chosen parameters, see the figure captions. The results show good performance of both methods. Furthermore, it can be deduced that the cross-covariance estimator gives more precise estimation results than the sequence-of-differences estimator.

Next we examine the situation of a high arrival rate at node 1. Note that the arrival process $\boldsymbol{A}_2$ has no direct influence on the estimators (compare with Remark 3 and Remark 4). For the left-hand diagram of Figure 3 we chose Bernoulli input distributions and geometric service distributions. It is well-known that the cross-covariance estimator remains very stable when there are customers arriving at nearly every time slot at node 1. On the contrary, the sequence-

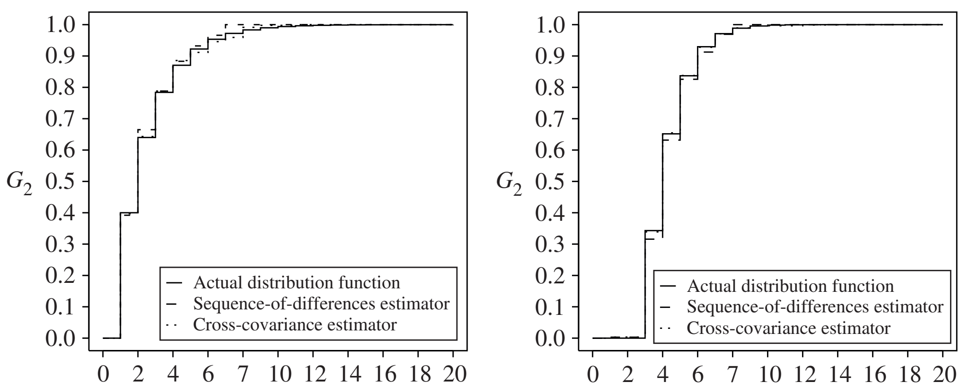

FIGURE 2: *Left:* Poisson inputs $\lambda_1 = 0.25$ and $\lambda_2 = 0.4$, and geometric service times $s_1 = 0.3$ and $s_2 = 0.4$. *Right:* Poisson inputs $\lambda_1 = 0.3$ and $\lambda_2 = 0.3$, with geometric service times $s_1 = 0.3$ at the first node and negative binomial service times $k_2 = 3$ and $s_2 = 0.7$ at the second node.

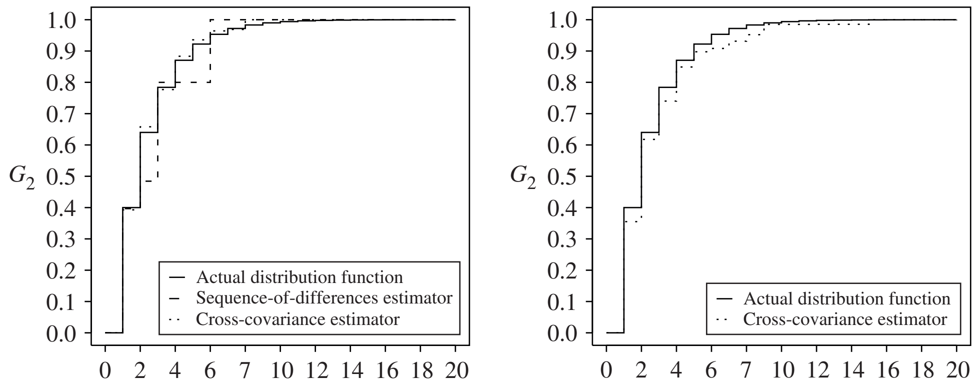

FIGURE 3: *Left:* Bernoulli inputs $b_1 = 0.9$ and $b_2 = 0.2$, and geometric service times $s_1 = 0.3$ and $s_2 = 0.4$. *Right:* input distributions $\mathbb{P}(A_1(0) = 1) = \mathbb{P}(A_2(0) = 1) = 0.7$ and $\mathbb{P}(A_1(0) = 2) = \mathbb{P}(A_2(0) = 2) = 0.3$, and geometric service times $s_1 = 0.3$ and $s_2 = 0.4$.

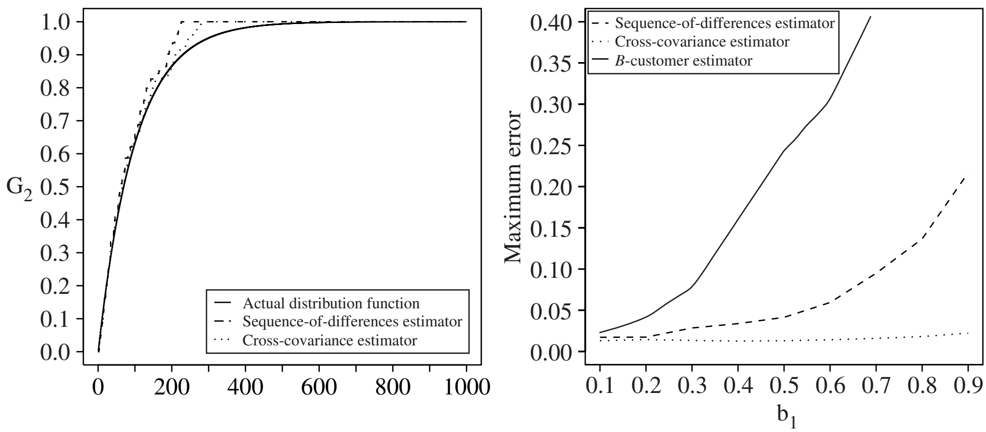

FIGURE 4: *Left:* Bernoulli inputs $b_1 = 0.01$ and $b_2 = 0.01$, and geometric service times $s_1 = 0.3$ and $s_2 = 0.01$. *Right:* graphical illustration of Table 1.

of-differences estimator does not perform in a satisfactory way here. The cross-covariance estimator can even be used in those scenarios where there are customers arriving at every time slot with probability 1 as the right-hand diagram of Figure 3 shows. There is at each of the nodes one external arriving customer with probability 0.7 and two external arriving customers with probability 0.3 per time slot. The service times are geometric. We now turn to the case of large possible service times at node 2. For the left-hand diagram of Figure 4, we chose the service times at the second node to be geometric with the small parameter $s_2 = 0.01$. Both external input sequences are Bernoulli. We see that both estimators give satisfactory results. However, the computation of the sequence-of-differences estimator was more than five times faster. The reason for this is that, for the cross-covariance estimator, we have to compute two cross-covariances for the calculation of each estimated value of the service time distribution function G_2, while the calculation for the sequence-of-differences estimator is practically based on the computation of two empirical conditional distribution functions. Thus, especially when calculating the estimators for a large number of values (which is necessary

TABLE 1: Bernoulli inputs: b_1 = as below, $b_2 = 0.2$, geometric service times: $s_1 = 0.3$, $s_2 = 0.4$.

Estimator	b_1								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
B-customer	0.0235	0.0410	0.0779	0.1595	0.2435	0.3043	0.4190	NaN	NaN
Sequence-of-differences	0.0172	0.0174	0.0285	0.0338	0.0414	0.0594	0.0949	0.1365	0.2168
Cross-covariance	0.0130	0.0144	0.0134	0.0126	0.0130	0.0142	0.0159	0.0181	0.0223

when the service times at node 2 tend to be rather high), it takes much more computing time to calculate the cross-covariance estimator than the sequence-of-differences estimator. Hence, the sequence-of-differences estimator should be preferred in settings which are used to approximate continuous-time systems.

Next we compare our estimators with the so-called B-customer estimator. The idea of this estimator goes back to Ross (1970). The construction is as follows. Only customers entering at the beginning of busy periods are used for estimation (so-called B-customers). Here, an idle period of a network is the time where there are no customers at all nodes of the system. To estimate the values $L_1(x)$ and $L_2(x)$, only those B-customers are used for which the next arrival into the system (after their arrival) is larger than or equal to x (so-called B_x-customers) for $x \in \mathbb{N}$. Based on (2), estimation for $L_1(x)$ is then done in the following way:

$$\frac{1-pq}{1-p} \times \frac{\#\{B_x\text{-customers entering system at node 1 and leaving at node 1 after} \leq x \text{ time units}\}}{\#\{B_x\text{-customers}\}}.$$

The estimation of L_2 is defined analogously. An estimator for G_2 is now constructed by deconvolution using the relation $L_1 * G_2 = L_2$ as in case of the sequence-of-differences estimator. Obviously, the B-customer estimator is the naive estimator one would think of right away. Clearly, its biggest disadvantage is that it only uses a fraction of the observed data. However, the B-customer estimator can be defined for nearly every class of queueing networks, e.g. for $\mathrm{GI}^X/\mathrm{G}/k$ queues with a finite number k of parallel servers.

To compare the accuracy of the three estimators, we calculate the maximal difference between the respective estimated values and the true distribution function (i.e. the $||\cdot||_\infty$ distance) in the case of Bernoulli arrivals. To study in particular the influence of the external Bernoulli arrival parameter b_1 at the first node on the estimators, we examined the exactness of the estimators for different choices of b_1 by calculating the mean of the respective maximal differences in ten runs. As can be seen from the results in Table 1 and Figure 4 (right-hand diagram), all three estimators give satisfactory results for small arrival parameters b_1. This can be explained by a high number of idle periods, and clear differences between external departure and external arrival points. We further note that the sequence-of-differences estimator and the B-customer estimator perform considerably worse when increasing the arrival parameter b_1, whereas the cross-covariance estimator is only slightly influenced. For $b_1 = 0.8$ and $b_1 = 0.9$, we even obtained no B-customer in the simulated 10^6 time units.

In summary, we have seen that the cross-covariance estimator is the most accurate estimator, in particular for frequent external arrivals. The sequence-of-differences estimator is considerably faster in computing than the cross-covariance estimator, especially when approximating continuous-time settings. Moreover, it is still much more accurate than the

B-customer estimator. The B-customer estimator can be used in an all-purpose way. However, much data is needed for its computation and its accuracy is much worse than for the two more sophisticated estimators.

Appendix A. Mathematical basics

A.1. Discrete deconvolution on $\mathbb{N}$

We assume here that we are given consistent estimators for a finite mass function f, as well as for the convolution $f * g$ of f with a second finite mass function g. The aim is to find conditions under which it is possible to uniquely identify and determine g. The following proposition is related to techniques for discrete deconvolution in the deterministic setting used in, e.g. digital signal processing (see Proakis and Manolakis (1992, pp. 374–376)).

Proposition 1. *Let f and g be two finite mass functions on $\mathbb{N}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(f_n)_{n\in\mathbb{N}}$ and $(h_n)_{n\in\mathbb{N}}$ be sequences of random functions defined on that space such that, for all $x \in \mathbb{N}$, $f_n(x) \to f(x)$ a.s. as $n \to \infty$ and $h_n(x) \to (f*g)(x)$ a.s. as $n \to \infty$. Moreover, let M_n be a random sequence on $(\Omega, \mathcal{F}, \mathbb{P})$, such that $M_n(\omega) \in \mathbb{N}_0$, $f_n(M_n(\omega)) > 0$ for all $\omega \in \Omega$ and $M_n \to M$ a.s. as $n \to \infty$, where $M := \min\{x \in \mathbb{N} \mid f(x) > 0\} < \infty$. Define $g_n(1) := h_n(M_n + 1)/f_n(M_n)$, and $g_n(k)$ iteratively for $k \geq 2$ as*

$$g_n(k) := \frac{h_n(M_n + k) - \sum_{i=1}^{k-1} f_n(M_n + k - i) g_n(i)}{f_n(M_n)}.$$

Then it holds that, for all $x \in \mathbb{N}$, $g_n(x) \to g(x)$ a.s. as $n \to \infty$.

Proof. We choose an arbitrary $\omega \in \Omega$ such that the functions $M_n(\omega)$, $f_n(\omega)(z)$, and $h_n(\omega)(z)$ converge to the respective values M, $f(z)$, and $(f * g)(z)$ for all $z \in \mathbb{N}$. The set of all such ω has probability 1 since $\mathbb{P}(\bigcap_{i\in\mathbb{N}} A_i) = 1$ when, for all $i \in \mathbb{N}$, it holds that $\mathbb{P}(A_i) = 1$. Since, for $n \to \infty$, $M_n(\omega) \to M < \infty$ and $M_n(\omega) \in \mathbb{N}_0$ for all $n \in \mathbb{N}$, there is an $N(\omega) \in \mathbb{N}$ such that $M_n(\omega) = M$ for all $n \geq N(\omega)$. Thus, we can conclude that, for every $j \in \mathbb{N}_0$ with $n \geq N(\omega)$,

$$\begin{aligned} f_n(\omega)(M_n(\omega) + j) &= f_n(\omega)(M + j) \to f(M + j) \quad \text{as } n \to \infty, \\ h_n(\omega)(M_n(\omega) + j) &= h_n(\omega)(M + j) \to (f * g)(M + j) \quad \text{as } n \to \infty. \end{aligned} \tag{8}$$

Hence,

$$g_n(\omega)(1) = \frac{h_n(\omega)(M_n(\omega) + 1)}{f_n(\omega)(M_n(\omega))} \to \frac{(f * g)(M + 1)}{f(M)} = g(1) \quad \text{as } n \to \infty.$$

The convergence of $g_n(x)$ for $x \in \mathbb{N}$ is shown by induction. For this purpose, we pick an arbitrary $x \in \mathbb{N}$ and assume that, for all $y < x$, $g_n(y) \to g(y)$ a.s. as $n \to \infty$. Hence, with (8) we have $\sum_{i=1}^{x-1} f_n(M_n + x - i) g_n(i) \to \sum_{i=1}^{x-1} f(M + x - i) g(i)$ a.s. as $n \to \infty$, and the convergence of $g_n(x)$ to $g(x)$ directly follows.

A.2. Uniform convergence of distribution functions on $\mathbb{N}_0$

In this subsection we derive a construction principle for random functions that converge a.s. uniformly to a distribution function on $\mathbb{N}_0$. In particular, we show how to construct a.s. uniformly convergent estimators from a.s. pointwise convergent estimators for a distribution function on $\mathbb{N}_0$. We start with sufficient conditions under which an a.s. pointwise convergent sequence of random functions is also a.s. uniformly convergent on $\mathbb{N}_0$.

Lemma 10. *Let F be a distribution function on $\mathbb{N}_0$, and let $(F_n)_{n\in\mathbb{N}}$ be a sequence of random functions on $\mathbb{N}_0$ defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that*

1. $F_n(x) \le 1$ *for all* $n \in \mathbb{N}$ *and* $x \in \mathbb{N}_0$, $\mathbb{P}$*-a.s.,*
2. $F_n(\cdot)$ *is monotonically increasing for all* $n \in \mathbb{N}$, $\mathbb{P}$*-a.s.,*
3. $F_n(x) \to F(x)$ *a.s. as* $n \to \infty$ *for all* $x \in \mathbb{N}_0$.

Then the sequence $(F_n)_{n\in\mathbb{N}}$ converges $\mathbb{P}$-a.s. uniformly to F, i.e.

$$\mathbb{P}\Big(\sup_{x\in\mathbb{N}_0} |F_n(x) - F(x)| \to 0 \text{ as } n \to \infty\Big) = 1.$$

Proof. We choose an arbitrary $\omega \in \Omega$ for which the conditions 1–3 are fulfilled. Moreover, let ε be arbitrary in $\mathbb{R}_+$. We prove the assertion by showing that there is an $N \in \mathbb{N}$ such that $\sup_{x\in\mathbb{N}_0} |F_n(\omega)(x) - F(x)| \le \varepsilon$ for all $n \ge N$. First, we find a $C \in \mathbb{N}_0$ with $F(C) > 1 - \varepsilon/2$, since F is a distribution function on $\mathbb{N}_0$. Then, for each $x \in \{0, 1, \dots, C\}$, there is, due to condition 3, a number $N_x \in \mathbb{N}$ with $|F_n(\omega)(x) - F(x)| < \varepsilon/2$ for all $n \ge N_x$. With $N := \max_{x\in\{0,1,\dots,C\}} N_x$, it follows that, for all $x \in \{0, 1, \dots, C\}$, $|F_n(\omega)(x) - F(x)| < \varepsilon/2$ for all $n \ge N$. Then, for all $n \ge N$, $F_n(\omega)(C) = F(C) - F(C) + F_n(\omega)(C) \ge F(C) - |F(C) - F_n(\omega)(C)| > 1 - \varepsilon/2 - \varepsilon/2 = 1 - \varepsilon$. Thus, due to conditions 1 and 2, $1 - \varepsilon < F_n(\omega)(x) \le 1$ for all $x \ge C$ and $n \ge N$. Since $1 - \varepsilon < 1 - \varepsilon/2 < F(x) \le 1$ for all $x \ge C$, it is now easy to see that, for all $n \ge N$, $|F_n(\omega)(x) - F(x)| < \varepsilon$ for all $x \ge C$. It follows that $\sup_{x\in\mathbb{N}_0} |F_n(\omega)(x) - F(x)| \le \varepsilon$ for all $n \ge N$. Since $\mathbb{P}(\bigcap_{i\in\mathbb{N}} A_i) = 1$, if, for all $i \in \mathbb{N}$, we have $\mathbb{P}(A_i) = 1$, the proof is complete.

Theorem 6. *Let F be a distribution function on $\mathbb{N}_0$, and let $(F_n)_{n\in\mathbb{N}}$ be a sequence of random functions on $\mathbb{N}_0$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying $F_n(x) \to F(x)$ a.s. as $n \to \infty$ for all $x \in \mathbb{N}_0$. Let $G_n : \mathbb{N}_0 \to [0, 1]$ be defined as $G_n(x) := \min(\max_{y\le x} F_n(y), 1)$. Then the sequence $(G_n)_{n\in\mathbb{N}}$ converges $\mathbb{P}$-a.s. uniformly to F.*

Proof. We show that $(G_n)_{n\in\mathbb{N}}$ satisfies the assumptions of Lemma 10. The first two conditions are obvious. For the proof of 3, we observe that, for all $x \in \mathbb{N}_0$,

$$\Big|\max_{y\le x} F_n(y) - F(x)\Big| = \Big|\max_{y\in\{0,1,\dots,x\}} F_n(y) - \max_{y\in\{0,1,\dots,x\}} F(y)\Big| \le \max_{y\in\{0,1,\dots,x\}} |F_n(y) - F(y)|.$$

By assumption we have $|F_n(\omega)(y) - F(y)| \to 0$ as $n \to \infty$ for all ω in some set A_y with $\mathbb{P}(A_y) = 1$. We conclude that, for all $\omega \in A_0 \cap A_1 \cap \cdots \cap A_x$: $\max_{y\le x} |F_n(\omega)(y) - F(y)| \to 0$ as $n \to \infty$. Thus, for all $x \in \mathbb{N}_0$, $G_n(x) \to \min(F(x), 1) = F(x)$ a.s. as $n \to \infty$.

Appendix B. Proof of Theorem 2

We show (5). The proof of (6) proceeds analogously. We begin by splitting the following expectation into two terms:

$$\mathbb{E}[D_1(0)\mathbf{1}_{\{Z_1(0)\le x\}}] = \mathbb{E}\bigg[\bigg(\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S^{(1)}_{-j,k}=j,\, E^{(1)}_{-j,k}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)\le x\}}\bigg] + \mathbb{E}\bigg[\bigg(\sum_{l=2}^{\infty}\sum_{m=1}^{A_2(-l)} \mathbf{1}_{\{S^{(2)}_{-l,m}=l,\, E^{(2)}_{-l,m}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)\le x\}}\bigg]. \tag{9}$$

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

We consider the two terms in (9) separately. Owing to the mutual independence of $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, and the sequences of service times, the second term of (9) can be written as

$$\mathbb{E}\bigg[\bigg(\sum_{l=2}^{\infty}\sum_{m=1}^{A_2(-l)}\mathbf{1}_{\{S^{(2)}_{-l,m}=l,\,E^{(2)}_{-l,m}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)\le x\}}\bigg]$$
$$=\mathbb{E}\bigg[\sum_{l=2}^{\infty}\sum_{m=1}^{A_2(-l)}\mathbf{1}_{\{S^{(2)}_{-l,m}=l,\,E^{(2)}_{-l,m}=1\}}\bigg]\mathbb{P}(Z_1(0)\le x). \tag{10}$$

Furthermore, we obtain $\mathbb{P}(Z_1(0)\le x)=1-\mathbb{P}(A_1(-x)=0,\ A_1(-x+1)=0,\dots,A_1(-1)=0)=1-c_1^x$. Moreover, by the monotone convergence theorem and Wald's equation,

$$\mathbb{E}\bigg[\sum_{l=2}^{\infty}\sum_{m=1}^{A_2(-l)}\mathbf{1}_{\{S^{(2)}_{-l,m}=l,\,E^{(2)}_{-l,m}=1\}}\bigg]=\mathbb{E}[A_2(0)]\mathbb{P}(E^{(2)}_{0,m}=1)\sum_{l=2}^{\infty}\mathbb{P}(S^{(2)}_{-l,m}=l\mid E^{(2)}_{-l,m}=1)$$
$$=\frac{(1-p)q}{1-pq}\mathbb{E}[A_2(0)].$$

Hence, (10) reads

$$\mathbb{E}\bigg[\bigg(\sum_{l=2}^{\infty}\sum_{m=1}^{A_2(-l)}\mathbf{1}_{\{S^{(2)}_{-l,m}=l,\,E^{(2)}_{-l,m}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)\le x\}}\bigg]=\frac{(1-p)q}{1-pq}\mathbb{E}[A_2(0)](1-c_1^x). \tag{11}$$

For the first term in (9), we fix an arbitrary $1\le i\le x$ and calculate

$$\mathbb{E}\bigg[\bigg(\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)}\mathbf{1}_{\{S^{(1)}_{-j,k}=j,\,E^{(1)}_{-j,k}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)=i\}}\bigg]$$
$$=\mathbb{P}(Z_1(0)=i)$$
$$\times\mathbb{E}\bigg[\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)}\mathbf{1}_{\{S^{(1)}_{-j,k}=j,\,E^{(1)}_{-j,k}=1\}}\;\bigg|\;A_1(-1)=0,\dots,A_1(-i+1)=0,\ A_1(-i)>0\bigg]. \tag{12}$$

Obviously, it holds that $\mathbb{P}(Z_1(0)=i)=(1-c_1)c_1^{i-1}$. We split the second factor of (12):

$$\mathbb{E}\bigg[\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)}\mathbf{1}_{\{S^{(1)}_{-j,k}=j,\,E^{(1)}_{-j,k}=1\}}\;\bigg|\;A_1(-1)=0,\dots,A_1(-i+1)=0,\ A_1(-i)>0\bigg]$$
$$=\mathbb{E}\bigg[\sum_{j=1}^{i-1}\sum_{k=1}^{A_1(-j)}\mathbf{1}_{\{S^{(1)}_{-j,k}=j,\,E^{(1)}_{-j,k}=1\}}\;\bigg|\;A_1(-1)=0,\dots,A_1(-i+1)=0,\ A_1(-i)>0\bigg]$$
$$+\mathbb{E}\bigg[\sum_{k=1}^{A_1(-i)}\mathbf{1}_{\{S^{(1)}_{-i,k}=i,\,E^{(1)}_{-i,k}=1\}}\;\bigg|\;A_1(-1)=0,\dots,A_1(-i+1)=0,\ A_1(-i)>0\bigg]$$
$$+\mathbb{E}\bigg[\sum_{j=i+1}^{\infty}\sum_{k=1}^{A_1(-j)}\mathbf{1}_{\{S^{(1)}_{-j,k}=j,\,E^{(1)}_{-j,k}=1\}}\;\bigg|\;A_1(-1)=0,\dots,A_1(-i+1)=0,\ A_1(-i)>0\bigg]$$
$$=\mathbb{E}\bigg[\sum_{k=1}^{A_1(-i)}\mathbf{1}_{\{S^{(1)}_{-i,k}=i,\,E^{(1)}_{-i,k}=1\}}\;\bigg|\;A_1(-i)>0\bigg]+\mathbb{E}\bigg[\sum_{j=i+1}^{\infty}\sum_{k=1}^{A_1(-j)}\mathbf{1}_{\{S^{(1)}_{-j,k}=j,\,E^{(1)}_{-j,k}=1\}}\bigg]. \tag{13}$$

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

We calculate the second term in (13):

$$\mathbb{E}\Bigg[\sum_{k=1}^{A_1(-i)} \mathbf{1}_{\{S^{(1)}_{-i,k}=i, E^{(1)}_{-i,k}=1\}} \,\Bigg|\, A_1(-i)>0\Bigg] = \frac{\mathbb{E}\Big[\sum_{k=1}^{A_1(-i)} \mathbf{1}_{\{S^{(1)}_{-i,k}=i,\, E^{(1)}_{-i,k}=1\}} \mathbf{1}_{\{A_1(-i)>0\}}\Big]}{\mathbb{P}(A_1(-i)>0)}$$
$$= \frac{\mathbb{E}[A_1(0)]\mathbb{P}(S^{(1)}_{-i,1}=i,\, E^{(1)}_{-i,1}=1)}{1-c_1}$$
$$= \frac{1}{1-c_1}\mathbb{E}[A_1(0)]\frac{1-p}{1-pq}\mathbb{P}(S^{(1)}_{0,1}=i \mid E^{(1)}_{0,1}=1). \tag{14}$$

Moreover, the third term in (13) can be written as

$$\mathbb{E}\Bigg[\sum_{j=i+1}^{\infty}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S^{(1)}_{-j,k}=j,\, E^{(1)}_{-j,k}=1\}}\Bigg]$$
$$= \mathbb{E}\Bigg[\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S^{(1)}_{-j,k}=j,\, E^{(1)}_{-j,k}=1\}}\Bigg] - \mathbb{E}\Bigg[\sum_{j=1}^{i}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S^{(1)}_{-j,k}=j,\, E^{(1)}_{-j,k}=1\}}\Bigg]$$
$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)] - \mathbb{E}[A_1(0)]\frac{1-p}{1-pq}\sum_{j=1}^{i}\mathbb{P}(S^{(1)}_{-j,1}=j \mid E^{(1)}_{-j,1}=1)$$
$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)](1-\mathbb{P}(S^{(1)}_{0,1}\le i \mid E^{(1)}_{0,1}=1)). \tag{15}$$

Inserting (14) and (15) into (13) yields

$$\mathbb{E}\Bigg[\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S^{(1)}_{-j,k}=j,\, E^{(1)}_{-j,k}=1\}} \,\Bigg|\, A_1(-1)=0,\ldots,A_1(-i+1)=0,\ A_1(-i)>0\Bigg]$$
$$= \frac{1}{1-c_1}\mathbb{E}[A_1(0)]\frac{1-p}{1-pq}\mathbb{P}(S^{(1)}_{0,1}=i \mid E^{(1)}_{0,1}=1)$$
$$+ \frac{1-p}{1-pq}\mathbb{E}[A_1(0)](1-\mathbb{P}(S^{(1)}_{0,1}\le i \mid E^{(1)}_{0,1}=1))$$
$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)]\bigg(1-\mathbb{P}(S^{(1)}_{0,1}\le i \mid E^{(1)}_{0,1}=1) + \frac{1}{1-c_1}\mathbb{P}(S^{(1)}_{0,1}=i \mid E^{(1)}_{0,1}=1)\bigg).$$

Furthermore, inserting the preceding equation into (12) gives

$$\mathbb{E}\Bigg[\bigg(\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S^{(1)}_{-j,k}=j,\, E^{(1)}_{-j,k}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)=i\}}\Bigg]$$
$$= (c_1^{i-1}(1-c_1))\frac{1-p}{1-pq}\mathbb{E}[A_1(0)]$$
$$\times\bigg(1-\mathbb{P}(S^{(1)}_{0,1}\le i \mid E^{(1)}_{0,1}=1) + \frac{1}{1-c_1}\mathbb{P}(S^{(1)}_{0,1}=i \mid E^{(1)}_{0,1}=1)\bigg)$$
$$= c_1^{i-1}\frac{1-p}{1-pq}\mathbb{E}[A_1(0)]$$
$$\times((1-c_1)-(1-c_1)\mathbb{P}(S^{(1)}_{0,1}\le i \mid E^{(1)}_{0,1}=1) + \mathbb{P}(S^{(1)}_{0,1}=i \mid E^{(1)}_{0,1}=1))$$

https://doi.org/10.1239/aap/1409319560 Published online by Cambridge University Press

$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)]$$
$$\times (c_1^{i-1}(1-\mathbb{P}(S_{0,1}^{(1)} \le i-1 \mid E_{0,1}^{(1)}=1)) - c_1^i(1-\mathbb{P}(S_{0,1}^{(1)} \le i \mid E_{0,1}^{(1)}=1))).$$

Summing over all $1 \le i \le x$ yields

$$\mathbb{E}\bigg[\bigg(\sum_{j=1}^{\infty}\sum_{k=1}^{A_1(-j)} \mathbf{1}_{\{S_{-j,k}^{(1)}=j,E_{-j,k}^{(1)}=1\}}\bigg)\mathbf{1}_{\{Z_1(0)\le x\}}\bigg]$$
$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)]$$
$$\times \bigg(\sum_{i=0}^{x-1} c_1^i(1-\mathbb{P}(S_{0,1}^{(1)} \le i \mid E_{0,1}^{(1)}=1)) - \sum_{i=1}^{x} c_1^i(1-\mathbb{P}(S_{0,1}^{(1)} \le i \mid E_{0,1}^{(1)}=1))\bigg)$$
$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)][1-c_1^x(1-\mathbb{P}(S_{0,1}^{(1)} \le x \mid E_{0,1}^{(1)}=1))].$$

Inserting the preceding equation and (11) into (9) gives the result with Lemma 1:

$$\mathbb{E}[D_1(0)\mathbf{1}_{\{Z_1(0)\le x\}}]$$
$$= \frac{1-p}{1-pq}\mathbb{E}[A_1(0)][1-c_1^x(1-\mathbb{P}(S_{0,1}^{(1)} \le x \mid E_{0,1}^{(1)}=1))]$$
$$+ \frac{1-p}{1-pq}q\mathbb{E}[A_2(0)](1-c_1^x)$$
$$= (1-c_1^x)\bigg[\frac{1-p}{1-pq}\mathbb{E}[A_1(0)] + \frac{1-p}{1-pq}q\mathbb{E}[A_2(0)]\bigg]$$
$$+ \frac{1-p}{1-pq}\mathbb{E}[A_1(0)]c_1^x\mathbb{P}(S_{0,1}^{(1)} \le x \mid E_{0,1}^{(1)}=1)$$
$$= \mathbb{E}[D_1(0)](1-c_1^x) + \mathbb{E}[D_1(0)]\frac{(1-p)\mathbb{E}[A_1(0)]}{(1-pq)\mathbb{E}[D_1(0)]}c_1^x\mathbb{P}(S_{0,1}^{(1)} \le x \mid E_{0,1}^{(1)}=1)$$
$$= \mathbb{E}[D_1(0)]\bigg[1-c_1^x\bigg(1-\frac{(1-p)\mathbb{E}[A_1(0)]}{(1-pq)\mathbb{E}[D_1(0)]}L_1(x)\bigg)\bigg].$$

Acknowledgement

We thank an anonymous referee for her/his constructive comments which significantly improved this paper.

References

Bhat, U. N. and Subba Rao, S. (1986/87). Statistical analysis of queueing systems. *Queueing Systems* **1,** 217–247.

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

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

Conti, P. L. (1999). Large sample Bayesian analysis for Geo/*G*/1 discrete-time queueing models. *Ann. Statist.* **27,** 1785–1807.

Conti, P. L. (2002). Nonparametric statistical analysis of discrete-time queues, with applications to ATM teletraffic data. *Stoch. Models* **18,** 497–527.

Daley, D. J. (1976). Queueing output processes. *Adv. Appl. Prob.* **8,** 395–415.

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

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

Ke, J.-C. and Chu, Y.-K. (2006). Nonparametric and simulated analysis of intensity for a queueing system. *Appl. Math. Comput.* **183,** 1280–1291.

Liu, Z., Wynter, L., Xia, C. H. and Zhang, F. (2006). Parameter inference of queueing models for IT systems using end-to-end measurements. *Performance Evaluation* **63,** 36–60.

Pickands, J., III and Stine, R. A. (1997). Estimation for an $M/G/\infty$ queue with incomplete information. *Biometrika* **84,** 295–308.

Pitts, S. M. (1994). Nonparametric estimation of the stationary waiting time distribution function for the $GI/G/1$ queue. *Ann. Statist.* **22,** 1428–1446.

Proakis, J. G. and Manolakis, D. G. (1992). *Digital Signal Processing: Principles, Algorithms, and Applications*. Macmillan, New York.

Ross, S. M. (1970). Identifiability in $GI/G/k$ queues. *J. Appl. Prob.* **7,** 776–780.

Wichelhaus, C. and Langrock, R. (2012). Nonparametric inference for stochastic feedforward networks based on cross-spectral analysis of point processes. *Electron. J. Statist.* **6,** 1670–1714.