Finite-customer queues in series

Onno Boxma[*†], Offer Kella[‡§] and Michel Mandjes[¶†]

January 12, 2026

Abstract

In this paper we consider two models of queues in series with the special feature that they are fed by a finite pool of customers. For the case of infinite-server queues in series and various types of arrival processes which abort after a finite number of arrivals or after a finite time, we obtain the transient joint distribution of the numbers of customers at all queues. For the case of a $\cdot/G/1$ queue in series with a $\cdot/M/1$ queue we provide a recursive procedure to derive the transient joint queue-length distribution in the following setting: There is a finite number of arrivals, and the interarrival times are exponentially distributed with a rate that depends on the number of arrivals still due.

Keywords: Finite-customer models; queues in series; infinite-server queue; single-server queue; queue length

[*]EURANDOM and Department of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands. o.j.boxma@tue.nl

[†]Partly funded by the NWO Gravitation project Networks, grant number 024.002.003.

[‡]Department of Statistics, The Hebrew University of Jerusalem; Jerusalem 9190501, Israel. offer.kella@gmail.com

[§]Partially supported by ISF Grant 3336/24 and the Vigevani Chair in Statistics.

[¶]Mathematical Institute, Leiden University, Einsteinweg 55 2333 CC Leiden, The Netherlands, and Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. m.r.h.mandjes@uva.nl

1 Introduction

In queueing theory, there has traditionally been a strong focus on steady-state models, typically with Poisson or renewal input processes. In recent years, however, there has been growing interest in transient queueing behavior and in models with a finite customer input pool. The present paper, which may be viewed as a follow-up to [6], fits within this broader development. In [9], the belief was expressed that the $\cdot/G/1$ queue with a pool of $m < \infty$ customers, each having independent and identically distributed *arrival* times, would not be amenable to exact analysis. This claim was refuted in [6], albeit through an algorithmic approach based on Laplace transforms. In the present paper, we seek to investigate the extent to which *networks* of queues with a finite customer pool allow for an exact analysis. Specifically, we consider both a $\cdot/G/1$ queue in series with a $\cdot/M/1$ queue and a tandem configuration of infinite-server stations.

For the case of infinite-server stations in series, we show that the joint queue-length distribution in the successive queues at any time t can be derived for a wide range of finite input processes and for general service-time distributions at the stations. This is feasible because individual customers do not interfere with one another at any of the queues. For the case of two single-server queues in series, driven by a particular finite-source input process and with general (resp. exponential) service-time distributions in Q_1 (resp. Q_2), we extend the recursive approach of [6] to obtain the joint queue-length distribution at an exponentially distributed epoch. Through numerical Laplace transform inversion (cf. [1, 10]), this yields the transient joint queue-length distribution at any time t.

Next to our desire to investigate the extent to which a transient analysis of queueing systems with finite-source input is feasible, we were also motivated by the observation that, in many real-life settings, the assumption of an infinite stream of arrivals is unrealistic. In a wide range of service environments, such as banks, airports, retail shops, and medical facilities, it is often more natural to model the arrival times themselves, rather than the interarrival times, as independent and identically distributed (i.i.d.) random variables drawn from a finite customer pool. These examples illustrate situations in which customers decide independently when to show up, typically within a specific time window.

We refer to [7] for a comprehensive survey of finite-source queueing models, with particular emphasis on the strategic timing of arrivals. Some early contributions in this area include [2, 3, 9], which predominantly focus on scaling limits in the diffusion regime. In [11], an exact analysis is given for the workload at an exponentially distributed epoch in a single-server queue with i.i.d. arrival times; the paper also investigates several extensions, such as the presence of an additional Poisson arrival stream and the possibility of balking behavior. Hayashi et al. [8] establish an explicit relation between the transient queue-length distribution in queues with i.i.d. arrival times and the time-dependent joint distribution of the numbers of arrivals and departures in an auxiliary model driven by a non-homogeneous Poisson process. We also mention [12], which studies the ruin probability in an insurance risk model — conceptually dual to a single-server queue — with two types of clients: 'small' customers (arriving in the usual manner) and 'major' customers, the latter forming a finite pool. Each major customer submits exactly one claim and subsequently leaves the system.

The remainder of the paper is organized as follows. Sections 2 and 3 are devoted to infinite-server stations in series. In Section 2 customers arrive in the first station according to a Poisson process that stops as soon as m customers have arrived. We determine the joint distribution of the numbers of customers in all r stations in series, for any time $t \geq 0$. In Section 3 we again assume that there are only m arrivals at the first station, but this time the *arrival* times of these customers are i.i.d., and not the interarrival times. In Section 4 we consider the case of Poisson arrivals until some (fixed or random) time T. In each case, we again determine the joint queue-length distribution at any time $t \geq 0$. We exploit the fact that individual customers do not interfere with each other, while sums of independent Poisson processes again constitute a Poisson process. These were also essential elements in related papers about networks of infinite-server queues, like [5, 13]. In Section 5 we focus on the case of two single-server queues in series, with exponentially distributed service times in the second queue. At $t = 0$ there are already some customers in both queues, and m customers are yet to arrive in the first queue. Under the assumption that, if n customers are yet to arrive, then the next interarrival time is exponentially distributed with mean $1/\lambda_n$, $n = 1, \ldots, m$, we determine the probability generating function of the joint distribution of the numbers of customers in both queues, via a recursive procedure. An algorithm for this procedure is also presented.

Finally some notation: PGF denotes probability generating function and LST denotes Laplace-Stieltjes transform; $\mathbb{N}$ and $\mathbb{N}_0$ denote the sets of, respectively, positive and nonnegative integers; $\mathrm{Exp}(\zeta)$ denotes the exponential distribution with mean ζ^{-1}.

2 Infinite-server tandem queues with m Poisson arrivals

Let $N = \{N(t) \,|\, t \geq 0\}$ be a Poisson process with rate λ. Let $T_n := \inf\{t \,|\, N(t) = n\}$ for $n \in \mathbb{N}$ be the associated jump times and denote $T_0 := 0$. We fix the number of arrivals $m \in \mathbb{N}$. In this section we assume that the time of the last arrival is T_m, such that the arrival process is $N_m(t) := N(t) \wedge m$. In Subsection 2.1 we study a single infinite-server station in isolation, thus laying the groundwork that quickly allows us to generalize to an arbitrary number of infinite-server stations in series (Subsection 2.2), culminating in Theorem 2.

2.1 A single queue

Consider the $\cdot/G/\infty$ queue with arrival process $N_m(t)$ and i.i.d. service times independent of N, having some distribution F. Moreover, we assume that at time zero the number of customers in the system is zero. Denote by $L(t)$ the number of customers in the system at time t.

For $1 \leq k \leq m-1$ we have that $N_m(t) = k$ if and only if $N(t) = k$ and it is well known [14] that the conditional distribution of $L(t)$ given that $N(t) = k$ is $\mathrm{Bin}(k, p(t))$, where, with $\bar{F}(t) := 1 - F(t)$,

$$p(t) := \frac{1}{t} \int_0^t \bar{F}(s) \, \mathrm{d}s \,. \tag{1}$$

Let us now consider the conditional distribution of $L(t)$ given that $N_m(t) = m$, that is, given that $N(t) \geq m$. The conditional distribution of $L(t)$ given that $N(t) = m$ is $\mathrm{Bin}(m, p(t))$, so it remains to consider the conditional distribution of $L(t)$ given that $N(t) \geq m+1$, or equivalently, given that $T_{m+1} \leq t$. First we need to understand what is the conditional joint distribution of $T_1, \ldots, T_{m+1}$ given that $T_{m+1} \leq t$. Denoting $X_i = T_i - T_{i-1}$, we can find the conditional density of $T_1, \ldots, T_{m+1}$ given that $T_{m+1} \leq t$: for $0 = t_0 < t_1 <$

$\ldots < t_{m+1} < t$,

$$\begin{aligned} f_{T_1,\ldots,T_{m+1}\,|\,T_{m+1}\le t}(t_1,\ldots,t_{m+1}) &= \frac{f_{X_1,\ldots,X_{m+1}}(t_1,t_2-t_1,\ldots,t_{m+1}-t_m)}{\mathbb{P}(T_{m+1}\le t)} \\ &= \frac{\prod_{i=1}^{m+1}\left(\lambda e^{-\lambda(t_i-t_{i-1})}\right)}{\mathbb{P}(N(t)\ge m+1)} = \frac{e^{-\lambda t_{m+1}}\lambda^{m+1}}{1-\sum_{i=0}^{m} e^{-\lambda t}\frac{(\lambda t)^i}{i!}}\ ; \end{aligned} \tag{2}$$

whenever $t_1,\ldots,t_{m+1}$ does not satisfy $0 < t_1 < \ldots < t_{m+1} < t$ this conditional density is zero.

Recalling that $f_{T_{m+1}}(s) = \frac{\lambda^{m+1}}{m!}s^m e^{-\lambda s}$ for $s > 0$, we identify

$$\frac{e^{-\lambda t_{m+1}}\lambda^{m+1}}{1-\sum_{i=0}^{m} e^{-\lambda t}\frac{(\lambda t)^i}{i!}} = \frac{m!}{t_{m+1}^m}\, f_{T_{m+1}\,|\,T_{m+1}\le t}(t_{m+1})\,. \tag{3}$$

This means that this is the distribution where (i) we first pick a random variable having the conditional distribution of T_{m+1} given that $T_{m+1} \le t$, and (ii) then the conditional distribution of the remaining variables given T_{m+1} is equal to $t_{m+1} \in (0,t)$ is that of the order statistics of m i.i.d. Uniform$(0,t_{m+1})$ distributed random variables.

In order to compute the conditional distribution of $L(t)$ given that $T_{m+1} \le t$, we first sample a value s from the distribution with density $f_{T_{m+1}\,|\,T_{m+1}\le t}(\cdot)$. Then, the number of customers out of the first m that are still in the system at time t has the conditional distribution Bin$(m, p_t(t_m))$ where, for $0 < s < t$,

$$p_t(s) := \frac{1}{s}\int_0^s \bar F(t-u)\,\mathrm{d}u = \frac{1}{s}\int_{t-s}^t \bar F(u)\,\mathrm{d}u\,, \tag{4}$$

so that, recalling (1), $p_t(t) = p(t)$. Therefore we have that

$$\mathbb{P}(L(t)=\ell\,|\,N(t)\ge m+1) = \int_0^t \binom{m}{\ell} p_t^\ell(s)(1-p_t(s))^{m-\ell}\, f_{T_{m+1}\,|\,T_{m+1}\le t}(s)\,\mathrm{d}s\,. \tag{5}$$

We now note that

$$\begin{aligned} \lambda s p_t(s) &= \int_{t-s}^t \lambda \bar F(u)\,\mathrm{d}u\,, \\ \lambda s(1-p_t(s)) &= \int_{t-s}^t \lambda F(u)\,\mathrm{d}u\,, \\ \lambda s &= \int_{t-s}^t \lambda \bar F(u)\,\mathrm{d}u + \int_{t-s}^t \lambda F(u)\,\mathrm{d}u\,. \end{aligned} \tag{6}$$

Therefore, with $P_n(\mu) := e^{-\mu}\frac{\mu^n}{n!}$ for $n \ge 0$ and $\mu > 0$ (and the bottom equation in (6)), straightforward manipulations imply that

$$\mathbb{P}(L(t)=\ell,\,N(t)\ge m+1) = \lambda\int_0^t P_\ell\left(\int_{t-s}^t \lambda\bar F(u)\,\mathrm{d}u\right) P_{m-\ell}\left(\int_{t-s}^t \lambda F(u)\,\mathrm{d}u\right)\mathrm{d}s\,. \tag{7}$$

For $0 \le k \le m$ it is well known (and follows directly from the lines above (1)) that for $0 \le \ell \le k$,

$$\mathbb{P}(L(t)=\ell,\,N(t)=k) = P_\ell\left(\int_0^t \lambda\bar F(u)\,\mathrm{d}u\right) P_{k-\ell}\left(\int_0^t \lambda F(u)\,\mathrm{d}u\right)\,, \tag{8}$$

and therefore, summing for $\ell \le k \le m$, gives the following result.

Theorem 1. *The queue-length distribution in the initially empty $\cdot/G/\infty$ queue with arrival process $N_m(t)$ is given by: For $0 \le \ell \le m$,*

$$\mathbb{P}(L(t)=\ell) = P_\ell\left(\int_0^t \lambda \bar{F}(u)\,\mathrm{d}u\right) \sum_{i=0}^{m-\ell} P_i\left(\int_0^t \lambda F(u)\,\mathrm{d}u\right) + \lambda \int_0^t P_\ell\left(\int_{t-s}^t \lambda \bar{F}(u)\,\mathrm{d}u\right) P_{m-\ell}\left(\int_{t-s}^t \lambda F(u)\,\mathrm{d}u\right) \mathrm{d}s\,. \tag{9}$$

As a sanity check, we consider the limiting behavior as the number of customers tends to infinity. It is directly verified that, when $m \to \infty$, the right-hand side of (9) becomes $P_\ell\big(\int_0^t \lambda \bar{F}(u)\,\mathrm{d}u\big)$, which agrees with what is expected in the infinite-customer setting.

It is worthwhile observing that it also obviously holds that

$$\mathbb{P}(L(t)=\ell) = \mathbb{E}\binom{N_m(t)}{\ell} p_t^\ell(T_{m+1}\wedge t)(1-p_t(T_{m+1}\wedge t))^{N_m(t)-\ell} \tag{10}$$

and finally we note that, since $\lambda\int_0^t f(s)\,\mathrm{d}s = \mathbb{E}\int_{[0,t]} f(s)\,\mathrm{d}N(s) = \mathbb{E}\sum_{n=1}^{N(t)} f(T_n)$, we also have that

$$\begin{aligned}
&\lambda \int_0^t P_\ell\left(\int_{t-s}^t \lambda \bar{F}(u)\,\mathrm{d}u\right) P_{m-\ell}\left(\int_{t-s}^t \lambda F(u)\,\mathrm{d}u\right) \mathrm{d}s \\
&\quad= \lambda \int_0^t P_\ell\left(\int_0^s \lambda \bar{F}(t-u)\,\mathrm{d}u\right) P_{m-\ell}\left(\int_0^s \lambda F(t-u)\,\mathrm{d}u\right) \mathrm{d}s \qquad (11)\\
&\quad= \mathbb{E}\sum_{n=1}^{N(t)} P_\ell\left(\int_0^{T_n} \lambda \bar{F}(t-u)\,\mathrm{d}u\right) P_{m-\ell}\left(\int_0^{T_n} \lambda F(t-u)\,\mathrm{d}u\right),
\end{aligned}$$

where an empty sum is defined to be zero.

2.2 Queues in series

Once we understand how to treat the single-queue case, the case of two (or more) queues in series can be analyzed in an analogous manner. The only additional observation needed is the following. Let X and Y be (not necessarily independent) random variables distributed as the service times at queues Q_1 and Q_2, respectively. Then a customer arriving at time $u \in (0,t)$ will at time t be in Q_1 with probability $\bar{F}_X(t-u)$, in Q_2 with probability $F_X(t-u) - F_{X+Y}(t-u)$, and in neither queue (i.e., having already departed from the network) with probability $F_{X+Y}(t-u)$. Repeating the same arguments as in the single-queue case leads to the following conclusion, where $L_1(t)$ and $L_2(t)$ denote the numbers of customers in Q_1 and Q_2 at time t, respectively, for $\ell_1, \ell_2 \ge 0$ with $\ell_1 + \ell_2 \le m$:

$$\begin{aligned}
&\mathbb{P}(L_1(t)=\ell_1,\, L_2(t)=\ell_2) \qquad (12)\\
&\quad= P_{\ell_1}\left(\int_0^t \lambda \bar{F}_X(u)\,\mathrm{d}u\right) P_{\ell_2}\left(\int_0^t (\lambda F_X(u) - \lambda F_{X+Y}(u))\,\mathrm{d}u\right) \\
&\qquad\qquad \times \sum_{i=0}^{m-\ell_1-\ell_2} P_i\left(\int_0^t \lambda F_{X+Y}(u)\,\mathrm{d}u\right) \\
&\quad + \lambda \int_0^t P_{\ell_1}\left(\int_{t-s}^t \lambda \bar{F}_X(u)\,\mathrm{d}u\right) P_{\ell_2}\left(\int_{t-s}^t (\lambda F_X(u) - \lambda F_{X+Y}(u))\,\mathrm{d}u\right) \\
&\qquad\qquad \times P_{m-\ell_1-\ell_2}\left(\int_{t-s}^t \lambda F_{X+Y}(u)\,\mathrm{d}u\right) \mathrm{d}s\,.
\end{aligned}$$

The generalization to any number $r \geq 3$ of stations in series is straightforward. Namely, if S_i is the sum of the services times at queues $Q_1, \ldots, Q_i$ then a customer that arrives at time u will be in $Q_1, \ldots, Q_r$, and not in any queue, with probabilities

$$\bar{F}_{S_1}(t-u), F_{S_1}(t-u) - F_{S_2}(t-u), \ldots, F_{S_{r-1}}(t-u) - F_{S_r}(t-u), F_{S_r}(t-u)\,, \tag{13}$$

respectively, and therefore $\mathbb{P}(L_1(t) = \ell_1, \ldots, L_r(t) = \ell_r)$ is the following straightforward adaptation of (12).

Theorem 2. *For the model with r infinite-server stations in series, intially empty and with arrival process $N_m(t)$, the joint distribution of the numbers of customers $L_1(t), \ldots, L_r(t)$ in $Q_1, \ldots, Q_r$ at time t is given by: for $\ell_1, \ldots, \ell_r \geq 0$, $\sum_{i=1}^r \ell_i \leq m$,*

$$\begin{aligned}&\left(\prod_{j=1}^r P_{\ell_j}\left(\int_0^t (\lambda F_{S_{j-1}}(u) - \lambda F_{S_j}(u))\,du\right)\right) \sum_{i=0}^{m-\sum_{j=1}^r \ell_j} P_i\left(\int_0^t \lambda F_{S_r}(u)\,du\right) \\ &+ \lambda \int_0^t \left(\prod_{j=1}^r P_{\ell_j}\left(\int_{t-s}^t (\lambda F_{S_{j-1}}(u) - \lambda F_{S_j}(u))\,du\right)\right) P_{m-\sum_{j=1}^r \ell_j}\left(\int_{t-s}^t \lambda F_{S_r}(u)\,du\right) ds\,.\end{aligned} \tag{14}$$

3 Infinite-server tandem queues with m i.i.d. arrival times

In this section we again study r infinite-server stations in series. This time we assume that there will be m arrivals, and that their arrival times (not their *interarrival* times!) are i.i.d. In Subsection 3.2 we obtain the joint queue length distribution at some time t. This result builds upon the analysis in Subsection 3.1, where we consider the same queues in series, but *without* m future arrivals and with m customers who are initially present in the tandem queue.

3.1 The case of m initial customers and no future arrivals

Here we consider $r \in \mathbb{N}$ infinite-server stations in series, where there are no arrivals and $m \in \mathbb{N}$ customers initially present. Each customer has service demands from the various stations and, as before, we denote by S_i the sum of these service requirements from the first i stations ($1 \leq i \leq r$) and $S_0 = 0$. The sequences of service requirements for the m customers are i.i.d. random vectors. However, we do not require the service times of a given customer at successive stations to be independent.

The vector of parameters can be written as $\lambda \boldsymbol{p}(t)$, where the $(r+1)$-dimensional vector $\boldsymbol{p}(t)$ is given by

$$\boldsymbol{p}(t) = \left(p_1(t), \ldots, p_r(t), 1 - \sum_{i=1}^r p_i(t)\right)\,, \tag{15}$$

where, with $F_0(t) = 1$ for $t \geq 0$,

$$p_i(t) := F_{S_{i-1}}(t) - F_{S_i}(t)\,, \quad i = 1, \ldots, r\,, \tag{16}$$

noting that $1 - \sum_{i=1}^r p_i(t) = F_{S_r}(t)$ is the probability that a customer has left the network. If we denote by $L_1(t), \ldots, L_r(t)$ the number of customers in $Q_1, \ldots, Q_r$ at time t, then it is clear that, with MN denoting the multinomial distribution,

$$\left(L_1(t), \ldots, L_r(t), m - \sum_{i=1}^r L_i(t)\right) \sim \mathrm{MN}\left(m; \boldsymbol{p}(t)\right)\,. \tag{17}$$

Remark 1. We also note that if instead of m we assume that the initial number of customers M is Poisson(λ) distributed and independent of everything else, then

$$L_1(t), \dots, L_r(t), M - \sum_{i=1}^{r} L_i(t) \tag{18}$$

are independent random variables. Moreover, $L_i(t)$ is Poisson distributed with parameter $\lambda p_i(t)$ for $i = 1, \dots, r$, while $M - \sum_{i=1}^{r} L_i(t)$ is Poisson distributed with parameter $\lambda\big(1 - \sum_{i=1}^{r} p_i(t)\big)$. In fact, it is standard to show that in this case the departure process $D_i(t) := M - \sum_{j=1}^{i} L_j(t)$ from Q_i is nonhomogeneous Poisson with cumulative departure measure $\Lambda_i(t) := \lambda F_{S_i}(t)$ and that $\{D_i(s) \,|\, 0 \le s \le t\}, L_1(t), \dots, L_i(t)$ are independent. In particular this also holds for $i = r$, where $D_r(\cdot)$ is the departure process from the network. Clearly, for $1 \le i \le r-1$, $D_i(t)$ is the arrival process to Q_{i+1}. $\Diamond$

3.2 The case of m i.i.d. arrivals and no initial customers

We now aim to exploit the result of the previous subsection to analyze the case in which all r infinite-server stations are initially empty, while m customers are yet to arrive; their arrival times (not *inter*-arrival times!) are i.i.d. A generic arrival time is denoted by A, with distribution $A(\cdot)$. For each of m customers, consider the $(r+1)$-dimensional random vector

$$\{\text{time until arrival, service time at station } Q_1, \dots, \text{service time at station } Q_r\}.$$

These m random vectors are independent, while the entries of each of the vectors can be dependent. If we denote by $L_0(t)$ the number of customers who have yet to arrive and, as before, by $L_1(t), \dots, L_r(t)$ the numbers of customers at the various stations, then all arguments from the previous subsection may be repeated verbatim for

$$L_0(t), \dots, L_r(t)\,, \tag{19}$$

by interpreting index zero as an additional station in which the arrival time is identified with a zeroth service time. In particular, the arrival process to Q_1 is $D_0(t) = m - L_0(t)$. Hence we have found the following result.

Theorem 3. *For the model with r infinite-server stations in series, initially all empty and with m customers yet to arrive at Q_1, the joint distribution of the numbers of customers $L_0(t)$ yet to arrive and $L_1(t), \dots, L_r(t)$ in $Q_1, \dots, Q_r$ at time t is given by a multinomial distribution:*

$$\left(L_0(t), L_1(t), \dots, L_r(t), m - \sum_{i=0}^{r} L_i(t)\right) \sim \mathrm{MN}\left(m; \boldsymbol{p}(t)\right)\,, \tag{20}$$

where $p_0(t) := 1 - A(t)$ and where $\boldsymbol{p}(t)$ is given by (15) and (16) with now $S_{-1} := 0$, $S_0 := A$ and S_i is the sum of A and the service times in $Q_1, \dots, Q_i$, $i = 1, \dots, r$.

We finally note that Remark 1 applies here in the precise same manner as in Subsection 3.1.

4 Infinite-server tandem queues with Poisson arrivals only until time $T \in (0, \infty)$

We now assume that the system is initially empty, and that customers arrive at Q_1 according to a Poisson process with rate λ until some fixed time T, whereafter there are no more arrivals.

If arrivals occur according to a non-homogeneous Poisson process $N(\cdot)$ with a (sufficiently regular, i.e., Borel-measurable) rate function $\lambda(\cdot)$, then, conditioned on $N(t) = n$, the arrival epochs up to time t (before ordering) are i.i.d. random variables with density

$$\frac{\lambda(x)}{\int_0^t \lambda(s)\,\mathrm{d}s} 1_{[0,t]}(x)\,. \tag{21}$$

Therefore, the conditional distribution of $L_1(t), \dots, L_r(t), n - \sum_{i=1}^r L_i(t)$ is multinomial with probabilities $p_1(t), \dots, p_r(t), 1 - \sum_{i=1}^r p_i(t)$ where we now define, for $i = 1, \dots, r$,

$$p_i(t) := \frac{\int_0^t \lambda(s)(F_{S_{i-1}}(t-s) - F_{S_i}(t-s))\,\mathrm{d}s}{\int_0^t \lambda(s)\,\mathrm{d}s}\,. \tag{22}$$

As a consequence, $L_1(t), \dots, L_r(t)$ and $N(t) - \sum_{i=1}^r L_i(t)$ are independent Poisson random variables. The parameters of $L_1(t), \dots, L_r(t)$ are

$$\int_0^t \lambda(s)(F_{S_{i-1}}(t-s) - F_{S_i}(t-s))\,\mathrm{d}s, \quad 1 \le i \le r, \tag{23}$$

while the parameter of $N(t) - \sum_{i=1}^r L_i(t)$ is $\int_0^t \lambda(s) F_{S_r}(s)\,\mathrm{d}s$.

In the current case, we have that $\lambda(t) = \lambda 1_{[0,T]}(t)$. Therefore, the parameters for $i = 1, \dots, r$ become

$$\lambda \int_0^{T\wedge t} (F_{S_{i-1}}(t-s) - F_{S_i}(t-s))\mathrm{d}s = \big[\kappa_i(t) - \kappa_i(T-t)\big] - \big[\kappa_{i-1}(t) - \kappa_{i-1}(T-t)\big]\,, \tag{24}$$

where

$$\kappa_j(t) := \rho_j F_j^e(t), \quad F_j^e(t) := \frac{1}{\mathbb{E}S_j} \int_0^t (1 - F_{S_j}(s))\,\mathrm{d}s, \quad \rho_j := \lambda\,\mathbb{E}S_j, \tag{25}$$

and we observe that $F_j^e(t-T) = 0$ for $t \le T$. We summarize our findings in the following theorem.

Theorem 4. *For the model with r infinite-server stations in series, initially empty and with* Pois(λ) *arrivals at Q_1 up to a fixed time T, the numbers of customers $L_1(t), \dots, L_r(t)$ in $Q_1, \dots, Q_r$ and the number of departures from Q_r at time t are independent Poisson random variables. The parameters of $L_1(t), \dots, L_r(t)$ are given by (24), while the parameter for the number of departures from Q_r at time t is $\lambda\,(1 - \sum_{i=1}^r p_i(t)) = \lambda\,(1 - [\kappa_r(t) - \kappa_r(t-T)])$.*

Remark 2. When we assume that T is random, we replace $F_j^e(t-T)$ by $\mathbb{E}F_j^e(t-T)$, or equivalently by $\mathbb{P}(S_j^e + T > t)$ where $S_j^e \sim F_j^e$ and T are independent. One simple (random) case is where $T \sim \mathrm{Exp}(\mu)$, in which case the total number of arrivals has a geometric distribution with probability of success $\frac{\mu}{\lambda+\mu}$. In this particular case we also have that

$$\mathbb{E}F_j^e(t-T) = \mathbb{P}(T > t - S_j^e) = \mathbb{E}e^{-\mu(t-S_j^e)^+}\,, \tag{26}$$

for $t \ge 0$. $\Diamond$

Remark 3. We close these three sections on infinite-server stations in series by observing the following. Because individual customers do not interfere with each other, we can combine the results from Theorems 2, 3 and 4 in a straightforward manner, allowing one class of customers with arrival process $N_{m_1}(t)$, another with m_2 i.i.d. arrival times, and a third one with Poisson arrivals until time T. $\Diamond$

5 Single-server queues in series

In this section, we consider another tandem queueing system with a finite-source input: a $\cdot/G/1$ queue in series with a $\cdot/M/1$ queue, where some customers are already present at time zero and a finite number m of customers are still due to arrive. Our goal is to obtain the joint queue-length distribution at an independent random time that is $\text{Exp}(\gamma)$ distributed. This distribution, after division by γ, can be interpreted as the Laplace transform of the joint queue-length distribution at a fixed time t. Accordingly, one can apply numerical inversion of the Laplace transform (see, e.g., [10]) to recover the joint transient queue-length distribution.

In Subsection 5.1 we provide a detailed model description; in Subsection 5.2 we derive a finite recursive set of equations whose solution yields the probability generating function of the joint queue-length distribution at an $\text{Exp}(\gamma)$ time; and in Subsection 5.3 we present an algorithm implementing this recursive procedure.

Remark 4. It should be mentioned that Blanc et al. [4] have provided an exact steady-state analysis of the same tandem queue, but with a Poisson arrival process, i.e., the $M/G/1$ queue in series with a $\cdot/M/1$ queue. That analysis is mathematically quite involved; they determine the PGF of the joint steady-state queue-length distribution by solving a Fredholm integral equation of the second kind. Our transient analysis is simpler, as we consider only a finite number of arrivals and employ a recursive approach. $\diamondsuit$

5.1 Model description

Consider a queueing system consisting of two FCFS single-server queues Q_1 and Q_2 in series. Customers who have received service in Q_1 instantaneously move to Q_2; customers who have received service in Q_2 leave the system. Service times $B_{1,1}, B_{1,2}, \dots$ in Q_1 are i.i.d., with a general distribution $B_1(\cdot)$; a generic service time at Q_1 is denoted by B_1. Service times $B_{2,1}, B_{2,2}, \dots$ in Q_2 are i.i.d. $\text{Exp}(\beta)$ distributed; a generic service time in Q_2 is denoted by B_2. At time $t = 0$ there are already $(k_1, k_2) \in (\mathbb{N}_0, \mathbb{N}_0)$ customers present in (Q_1, Q_2), but their services have not yet started (this should at least hold for Q_1). In addition, $m \in \mathbb{N}_0$ customers are still to arrive at Q_1. The random variable $T_m \sim \text{Exp}(\lambda_m)$ is the time until the first arrival, $T_{m-1} \sim \text{Exp}(\lambda_{m-1})$ is the time between the first and second arrival, etc. All service times in both queues are independent of each other and of the interarrival times. We denote $Y_0 := 0$, $Y_j := \sum_{i=1}^{j} T_i$ for any $j \in \{1, \dots, m\}$, so that $Y_n - Y_{n-i} = T_n + \dots + T_{n-i+1}$ for $1 \le i \le n$, and $G_0 := 0$, $G_j := \sum_{i=1}^{j} B_{2,i}$ for any $j \in \{1, \dots, k_2 + m\}$.

Let $(L_1(t), L_2(t))$ denote the numbers of customers in (Q_1, Q_2) at time $t \ge 0$. Let $T \sim \text{Exp}(\gamma)$, independently of everything else. Finally, let $\mathbb{E}_{k_1,k_2,m}[\cdot]$ denote expectation when starting from (k_1, k_2) customers and m customers are yet to arrive according to the above-described mechanism (and similarly for $\mathbb{P}_{k_1,k_2,m}(\cdot)$). Our goal is to determine the two-dimensional PGF, with $\boldsymbol{z} = (z_1, z_2)$,

$$\mu_{k_1,k_2,m}(\boldsymbol{z}) := \mathbb{E}_{k_1,k_2,m}\left[z_1^{L_1(T)} z_2^{L_2(T)}\right], \quad z_1 \in (0,1],\ z_2 \in (0,1], \tag{27}$$

of the numbers of customers in (Q_1, Q_2) at T.

5.2 A recursive procedure

In this subsection in Step 1 we first focus on $\mu_{\ell_1,\ell_2,0}(\boldsymbol{z})$ terms, viz., there are no more arrivals. In Step 2 we consider $\mu_{0,\ell_2,n}(\boldsymbol{z})$: Q_1 is empty, but there are still n arrivals due. Finally, in

Step 3 we determine a recursive relationship between $\mu_{\ell_1,\ell_2,n}(\boldsymbol{z})$ and terms $\mu_{\ell_1+i-1,\ell_2-j,n-i}(\boldsymbol{z})$. We shall show that a combination of these three steps yields all $\mu_{\ell_1,\ell_2,n}(\boldsymbol{z})$. In both Step 1 and Step 3 we consider the queue lengths of the tandem queue immediately after two successive departures from Q_1. For an $M/G/1$ queue in isolation this is a classical approach, exploiting Markovian properties of the embedded queue-length process. In our tandem case, we also have such Markovian properties for the vector queue-length process, because the service times in Q_2 are memoryless.

Step 1. We first determine $\mu_{0,\ell_2,0}(\boldsymbol{z})$: There are no more arrivals, and Q_1 is already empty. Therefore, the number of customers present in Q_2 at time T is distributed like $(\ell_2 - N_p)^+$ where $N_p + 1$ is geometrically distributed with parameter $p := \frac{\gamma}{\gamma+\beta}$. This implies that, using that $\mathbb{P}(N_p \geq \ell_2) = (1-p)^{\ell_2}$, for $z_2 \neq 1-p$

$$\mu_{0,\ell_2,0}(\boldsymbol{z}) = \mathbb{E}z_2^{(\ell_2-N_p)^+} = \sum_{k=0}^{\ell_2-1} z_2^{\ell_2-k}(1-p)^k p + (1-p)^{\ell_2} = pz_2\,\frac{z_2^{\ell_2}-(1-p)^{\ell_2}}{z_2-(1-p)} + (1-p)^{\ell_2}\,, \quad (28)$$

whereas for $z_2 = 1-p$

$$\mu_{0,\ell_2,0}(\boldsymbol{z}) = (1-p)^{\ell_2}(\ell_2 p + 1)\,. \quad (29)$$

Next we turn to the case in which Q_1 is not empty. Distinguishing between $T > B_1$ and $T \leq B_1$ and noting that, with $A_\alpha(\cdot)$ a Poisson process with rate α independent of T, B_1, we have

$$\begin{aligned}
p_j(\alpha) &:= \mathbb{P}(A_\alpha(B_1) = j) = \mathbb{E}\,e^{-\alpha B_1}\frac{(\alpha B_1)^j}{j!} = \frac{(-\alpha)^j}{j!}\,\frac{\mathrm{d}^j}{\mathrm{d}\alpha^j}\underbrace{\mathbb{E}\,e^{-\alpha B_1}}_{p_0(\alpha)}\,,\\
q_j(\alpha) &:= \mathbb{P}(A_\alpha(T) = j) = \frac{\alpha^j\,\gamma}{(\alpha+\gamma)^{j+1}}\,.
\end{aligned} \quad (30)$$

As a consequence, we find

$$\begin{aligned}
\mathbb{P}(A_\alpha(B_1) = j, T > B_1) &= \mathbb{E}e^{-(\alpha+\gamma)B_1}\frac{(\alpha B_1)^j}{j!} = \left(\frac{\alpha}{\alpha+\gamma}\right)^j p_j(\alpha+\gamma)\,,\\
\mathbb{P}(A_\alpha(T) = j, T > t) &= e^{-\gamma t}\mathbb{P}(A_\alpha(T) = j | T > t) = e^{-\gamma t}\mathbb{P}(A_\alpha(t+T) = j)\\
&= \sum_{i=0}^{j} e^{-\gamma t}\,\mathbb{P}(A_\alpha(t) = i)\mathbb{P}(A_\alpha(T) = j-i)\\
&= \sum_{i=0}^{j}\left(\frac{\alpha}{\alpha+\gamma}\right)^i \mathbb{P}(A_{\alpha+\gamma}(t) = i) q_{j-i}(\alpha)\,, \qquad (31)\\
\mathbb{P}(A_\alpha(T) = j, T > B_1) &= \sum_{i=0}^{j}\left(\frac{\alpha}{\alpha+\gamma}\right)^i p_i(\alpha+\gamma)\,q_{j-i}(\alpha)\,,\\
\mathbb{P}(A_\alpha(T) = j, T \leq B_1) &= \mathbb{P}(A_\alpha(T) = j) - \mathbb{P}(A_\alpha(T) = j, T > B_1)\,.
\end{aligned}$$

Hence we have for $\ell_1 = 1, \dots, k_1$, with $\mathbb{P}(T > B_1) = 1 - \mathbb{P}(T \le B_1) = p_0(\gamma)$,

$$\begin{aligned}\mu_{\ell_1,\ell_2,0}(\boldsymbol{z}) = {} & \sum_{j=0}^{\ell_2-1} \mu_{\ell_1-1,\ell_2+1-j,0}(\boldsymbol{z}) \mathbb{P}(A_\beta(B_1) = j, T > B_1) + \\ & \mu_{\ell_1-1,1,0}(\boldsymbol{z}) \left(\mathbb{P}(T > B_1) - \sum_{j=0}^{\ell_2-1} \mathbb{P}(A_\beta(B_1) = j, T > B_1) \right) + \\ & z_1^{\ell_1} \left(\sum_{i=0}^{\ell_2-1} \mathbb{P}(A_\beta(T) = j, T \le B_1) z_2^{\ell_2-j} + \mathbb{P}(T \le B_1) - \sum_{i=0}^{\ell_2-1} \mathbb{P}(A_\beta(T) = j, T \le B_1) \right).\end{aligned} \tag{32}$$

Remark 5. Alternatively, we can rewrite (32) as

$$\begin{aligned}\mu_{\ell_1,\ell_2,0}(\boldsymbol{z}) &= \sum_{j=0}^{\ell_2-1} \mu_{\ell_1-1,\ell_2+1-j,0}(\boldsymbol{z}) \int_{t=0}^{\infty} e^{-\gamma t} e^{-\beta t} \frac{(\beta t)^j}{j!} \mathrm{d}\mathbb{P}(B_1 < t) + \\ & \quad \mu_{\ell_1-1,1,0}(\boldsymbol{z}) \int_{t=0}^{\infty} e^{-\gamma t} \sum_{j=\ell_2}^{\infty} e^{-\beta t} \frac{(\beta t)^j}{j!} \mathrm{d}\mathbb{P}(B_1 < t) + \\ & \quad z_1^{\ell_1} \int_{t=0}^{\infty} \gamma e^{-\gamma t}\, \mathbb{P}(B_1 > t) \left[\sum_{j=0}^{\ell_2-1} e^{-\beta t} \frac{(\beta t)^j}{j!} z_2^{\ell_2-j} + 1 - \sum_{j=0}^{\ell_2-1} e^{-\beta t} \frac{(\beta t)^j}{j!} \right] \mathrm{d}t\,,\end{aligned} \tag{33}$$

by working out the above probabilities. $\Diamond$

Notice that (28)–(29) gives us all terms $\mu_{0,\ell_2,0}(\boldsymbol{z})$, whereafter (32) (or, alternatively (33)) enables us to express all $\mu_{1,\ell_2,0}(\boldsymbol{z})$ into those terms, and subsequently all $\mu_{\ell_1,\ell_2,0}(\boldsymbol{z})$ into terms $\mu_{\ell_1-1,\cdot,0}(\boldsymbol{z})$ for $\ell_1 = 2, \dots, k_1$.

Step 2. Now consider $\mu_{0,\ell_2,n}(\boldsymbol{z})$: Q_1 is empty, but there are still n arrivals due. Distinguish between the two scenarios that interarrival time T_n is smaller than (the remaining time of) the $\mathrm{Exp}(\gamma)$ timer T (which has probability $\frac{\lambda_n}{\lambda_n+\gamma}$), and that it is larger:

$$\begin{aligned}\mu_{0,\ell_2,n}(\boldsymbol{z}) &= \frac{\lambda_n}{\lambda_n+\gamma} \sum_{j=0}^{\ell_2-1} \mu_{1,\ell_2-j,n-1}(\boldsymbol{z}) \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^j \frac{\lambda_n+\gamma}{\beta+\lambda_n+\gamma} + \\ & \quad \frac{\lambda_n}{\lambda_n+\gamma} \mu_{1,0,n-1}(\boldsymbol{z}) \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^{\ell_2} + \\ & \quad \frac{\gamma}{\lambda_n+\gamma} \left[\sum_{j=0}^{\ell_2-1} z_2^{\ell_2-j} \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^j \frac{\lambda_n+\gamma}{\beta+\lambda_n+\gamma} + \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^{\ell_2} \right] \\ &= \frac{\lambda_n}{\beta+\lambda_n+\gamma} \sum_{j=0}^{\ell_2-1} \mu_{1,\ell_2-j,n-1}(\boldsymbol{z}) \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^j + \\ & \quad \frac{\lambda_n}{\lambda_n+\gamma} \mu_{1,0,n-1}(\boldsymbol{z}) \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^{\ell_2} + \\ & \quad \frac{\gamma z_2}{(\beta+\lambda_n+\gamma) z_2 - \beta} \left[z_2^{\ell_2} - \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^{\ell_2} \right] + \frac{\gamma}{\lambda_n+\gamma} \Big(\frac{\beta}{\beta+\lambda_n+\gamma} \Big)^{\ell_2}.\end{aligned} \tag{34}$$

Here the first (resp. second) of the three terms in the righthand sides corresponds to $T_n < T$ and Q_2 does not (resp. does) empty completely before the end of T_n, and the third term corresponds to $T_n \geq T$. We have been using the memoryless property in all these calculations; e.g., $(\frac{\beta}{\beta+\lambda_n+\gamma})^{\ell_2}$ is the probability that all ℓ_2 customers in Q_2 are served before T_n or T expires.

Step 3. In order to establish a recursion for $\mu_{\ell_1,\ell_2,n}(\boldsymbol{z})$, we again distinguish between the two scenarios $T > B_1$ and $T \leq B_1$. In the $T > B_1$ scenario, there is a transition from some state (ℓ_1, ℓ_2, n) to state $(\ell_1 + i - 1, \ell_2 + 1 - j, n - i)$ between two successive service completions, corresponding to i new arrivals in Q_1 and j service completions in Q_2 during B_1; of course, $i = n$ and also $j = \ell_2$ require special attention. In the $T \leq B_1$ scenario, there is a transition from some state (ℓ_1, ℓ_2, n) to state $(\ell_1 + i, \ell_2 - j, n - i)$, corresponding to i new arrivals in Q_1 and j service completions in Q_2 during T; again, $i = n$ and also $j = \ell_2$ require special attention. We have, for any $\ell_1 \in \{1, \ldots, k_1\}$, $\ell_2 \in \{0, \ldots, k_2\}$, $n \in \{0, \ldots, m\}$:

$$\mu_{\ell_1,\ell_2,n}(\boldsymbol{z}) = \sum_{i=0}^{n}\sum_{j=0}^{\ell_2} \mu_{\ell_1+i-1,\ell_2+1-j,n-i}(\boldsymbol{z}) u_{n,i,l_2,j} + \sum_{i=0}^{n} z_1^{\ell_1+i} \sum_{j=0}^{\ell_2} z_2^{\ell_2-j} v_{n,i,\ell_2,j}\,, \tag{35}$$

the first double sum corresponding to the scenario $T > B_1$ and the second to the scenario $T \leq B_1$. The probabilities $u_{n,i,\ell_2,j}$ and $v_{n,i,\ell_2,j}$ are specified below. Recalling from Subsection 5.1 the definitions of Y_j (sums of interarrival times) and of G_j (sums of service times in Q_2), we can write:

$$u_{n,i,\ell_2,j} := \mathbb{P}(T > B_1, Y_n - Y_{n-i} \leq B_1 < Y_n - Y_{n-i-1}, G_j \leq B_1 < G_{j+1})\,, \tag{36}$$
$$v_{n,i,\ell_2,j} := \mathbb{P}(T \leq B_1, Y_n - Y_{n-i} \leq T < Y_n - Y_{n-i-1}, G_j \leq T < G_{j+1})\,, \tag{37}$$

for $\ell_2 = 0, \ldots, k_2,\;\; n = 1, \ldots, m,\;\; i = 0, \ldots, n-1,\;\; j = 0, \ldots, \ell_2 - 1$. The cases $i = n$ and/or $j = \ell_2$ can be dealt with similarly: for $i = n$ replace the condition $Y_n - Y_{n-i} \leq \ldots < Y_n - Y_{n-i-1}$ by $Y_n \leq \ldots$ and for $j = \ell_2$ replace the condition $G_j \leq \ldots < G_{j+1}$ by $G_{\ell_2} \leq \ldots$.

One can write the expressions (36) and (37) as follows:

$$u_{n,i,\ell_2,j} = \int_{t=0}^{\infty} e^{-\gamma t}\mathbb{P}(Y_n - Y_{n-i} \leq t < Y_n - Y_{n-i-1})\, e^{-\beta t}\frac{(\beta t)^j}{j!}\mathrm{d}\mathbb{P}(B_1 < t)\,, \tag{38}$$
$$v_{n,i,\ell_2,j} = \int_{t=0}^{\infty} \gamma e^{-\gamma t}\mathbb{P}(Y_n - Y_{n-i} \leq t < Y_n - Y_{n-i-1})\, e^{-\beta t}\frac{(\beta t)^j}{j!}\mathbb{P}(B_1 > t)\mathrm{d}t\,; \tag{39}$$

if $i = n$, then replace $\mathbb{P}(Y_n - Y_{n-i} \leq t < Y_n - Y_{n-i-1})$ by $\mathbb{P}(Y_n \leq t)$ and if $j = \ell_2$ then replace $e^{-\beta t}\frac{(\beta t)^j}{j!}$ by $1 - \sum_{k=0}^{\ell_2-1} e^{-\beta t}\frac{(\beta t)^k}{k!}$.

Remark 6. Just like in Step 1, we can alternatively write the above expressions in terms of expectations:

$$u_{n,i,\ell_2,j} = \mathbb{E}\Big[e^{-\gamma B_1}\, e^{-\lambda_{n-i}(B_1-(Y_n-Y_{n-i}))} 1_{\{Y_n-Y_{n-i}\leq B_1\}}\, e^{-\beta B_1}\frac{(\beta B_1)^j}{j!}\Big]\,, \tag{40}$$
$$v_{n,i,\ell_2,j} = \mathbb{E}\Big[1_{\{T\leq B_1\}} e^{-\lambda_{n-i}(T-(Y_n-Y_{n-i}))} 1_{\{Y_n-Y_{n-i}\leq T\}}\, e^{-\beta T}\frac{(\beta T)^j}{j!}\Big], \tag{41}$$

noting that $\mathbb{E}1_{\{T\leq B_1\}}f(T) = \mathbb{E}f(T) - \mathbb{E}1_{\{T>B_1\}}f(T)$ and that, for fixed t, $\mathbb{E}1_{\{T>t\}}f(T) = e^{-\gamma t}\,\mathbb{E}f(t+T)$ and hence $\mathbb{E}1_{\{T>B_1\}}f(T) = \mathbb{E}e^{-\gamma B_1}\, f(B_1+T)$. $\Diamond$

Remark 7. For special choices of the arrival time distribution and the service time distribution at Q_1, these expressions may be simplified. For example, assume that the arrival process is a Poisson $N_m(t)$ process that we introduced in Section 2, i.e., a Poisson(λ) process that is aborted after m arrivals. Then $\mathbb{P}(Y_n - Y_{n-i} \leq t < Y_n - Y_{n-i-1}) = e^{-\lambda t}\frac{(\lambda t)^i}{i!}$. If, moreover, $B_1 \sim \mathrm{Exp}(\eta)$, then we obtain that the expression in (38) simplifies to

$$u_{n,i,\ell_2,j} = \frac{\eta}{\eta+\gamma+\lambda+\beta}\binom{i+j}{i}\Big(\frac{\lambda}{\eta+\gamma+\lambda+\beta}\Big)^i\Big(\frac{\beta}{\eta+\gamma+\lambda+\beta}\Big)^j, \tag{42}$$

by either working out the integral in (38) or by a straightforward probabilistic reasoning. $\diamondsuit$

5.3 Algorithm

The above results can be used to devise a procedure that recursively determines $\mu_{k_1,k_2,m}(\boldsymbol{z})$. It is given by the following algorithm.

ALGORITHM to evaluate $\mu_{k_1,k_2,m}(\boldsymbol{z})$.
INPUT: k_1, k_2, m, $u_{n,i,\ell_2,j}$ and $v_{n,i,\ell_2,j}$ for all $n \in \{0,\dots,m\}$, $i \in \{0,\dots,n\}$, $\ell_2 \in \{0,\dots,k_2\}$ and $j \in \{0,\dots,\ell_2\}$.
OUTPUT: $\mu_{\ell_1,\ell_2,n}(\boldsymbol{z})$ for all $\ell_1 \in \{0,\dots,k_1\}$, $\ell_2 \in \{0,\dots,k_2\}$ and $n \in \{0,\dots,m\}$.

01: Compute $\mu_{0,\ell_2,0}(\boldsymbol{z})$ via (28)-(29);
02: FOR $\ell_1 = 1$ TO k_1 DO
03:   FOR $\ell_2 = 1$ TO k_2 DO
04:     Compute $\mu_{\ell_1,\ell_2,0}(\boldsymbol{z})$ via (33);
05:   END;
06: END;
07: FOR $n = 1$ TO m DO
08:   Compute $\mu_{0,\ell_2,n}(\boldsymbol{z})$ for all ℓ_2 via (34);
09:   FOR $\ell_1 = 1$ TO k_1 DO
10:     Compute $\mu_{\ell_1,\ell_2,n}(\boldsymbol{z})$ for all ℓ_2 via (35);
11:   END;
12: END;
13: RETURN $\mu_{k_1,k_2,m}(\boldsymbol{z})$.

Observe that in each step of this algorithm, one only needs objects that have been computed earlier. The complexity of the algorithm is $O(k_1k_2m^2)$, because the complexity of the for-loop in lines 07-12 is $O(k_1k_2n)$, while this loop has to be performed for $n = 1,\dots,m$.

Remark 8. By differentiating the above equations ℓ_1 times with respect to z_1 and ℓ_2 times with respect to z_2, the above algorithm results in the recursive computation of the factorial moments

$$\mathbb{E}_{k_1,k_2,m}[Z_1(T)(Z_1(T)-1)\dots(Z_1(T)-\ell_1+1)Z_2(T)(Z_2(T)-1)\dots(Z_2(T)-\ell_2+1)], \tag{43}$$

for any desired k_1, k_2, m and $\ell_1 \in \{1,\dots,k_1\}$, $\ell_2 \in \{1,\dots,k_2\}$. From this the moments $\mathbb{E}_{k_1,k_2,m}[Z_1(T)^{\ell_1}Z_2(T)^{\ell_2}]$ readily follow. Likewise, equating the coefficients of the polynomials in z_1 and z_2 on the two sides of (28)-(35) results in a similar algorithm for the recursive computation of the probabilities $\mathbb{P}_{k_1,k_2,m}(Z_1(T) = \ell_1, Z_2(T) = \ell_2)$. $\diamondsuit$

Remark 9. Several generalizations could be thought of, such as the ones discussed in [6]. Notably, in Remark 2 of [6] we also included the amount of work in the system; and in its Remark 3 we allowed service times to depend on n. $\diamondsuit$

References

[1] J. Abate and W. Whitt (1995). Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Comput.* **7**, 36-43.

[2] G. Bet (2020). An alternative approach to heavy-traffic limits for finite-pool queues. *Queueing Systems* **95**, 121-144.

[3] G. Bet, R. van der Hofstad and J. van Leeuwaarden (2019). Heavy-traffic analysis through uniform acceleration of queues with diminishing populations. *Math. Oper. Res.* **44**, 821-864.

[4] J.P.C. Blanc, R. Iasnogorodski and Ph. Nain (1988). Analysis of the $M/GI/1-\cdot/M/1$ queueing model. *Queueing Systems* **3**, 129-156.

[5] O.J. Boxma (1984). $M/G/\infty$ tandem queues. *Stoch. Proc. Appl.* **18**, 153-164.

[6] O.J. Boxma, O. Kella and M.R.H. Mandjes (2025). Finite customer-pool queues. *Oper. Res. Letters* **60**, 107267.

[7] M. Haviv and L. Ravner (2021). A survey of queueing systems with strategic timing of arrivals. *Queueing Systems* **99**, 163-198.

[8] K. Hayashi, Y. Inoue and T. Takine (2024). Time-dependent queue length distribution in queues fed by K customers in a finite interval. https://arxiv.org/abs/2412.06452. To appear in: *Queueing Systems.*

[9] H. Honnappa, R. Jain and A. Ward (2015). The $\Delta^{(i)}/G/1$ queueing model, and its fluid and diffusion approximations. *Queueing Systems* **80**, 71-103.

[10] P. den Iseger (2006). Numerical transform inversion using Gaussian quadrature. *Prob. Eng. Inform. Sci.* **20**, 1-44.

[11] M. Mandjes and D. Rutgers (2025). A queue with independent and identically distributed arrivals. *J. Appl. Probab.* **62**, 319-346.

[12] M. Mandjes and D. Rutgers (2025). Risk theory in a finite customer-pool setting. https://arxiv.org/abs/2505.11127.

[13] W.A. Massey and W. Whitt (1993). Networks of infinite-server queues with nonstationary Poisson input. *Queueing Systems* **13**, 183-250.

[14] L. Takács (1962). *Introduction to the Theory of Queues.* Oxford University Press.