EURANDOM PREPRINT SERIES

2018-016

November 23, 2018

Parameter Mixing in Infinite-Server Queues

L. van Kreveld, O.Boxma

ISSN 1389-2355

Parameter Mixing in Infinite-Server Queues

Lucas van Kreveld[1*] & Onno Boxma[2]

November 23, 2018

[1*] Korteweg-de Vries Institute, University of Amsterdam, Amsterdam, The Netherlands
Corresponding author, l.r.vankreveld@uva.nl
[2] Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

We consider two infinite-server queueing models with a so-called mixed arrival process. The arrival process is Poisson, but the arrival intensity is resampled from some distribution at exponentially distributed time intervals. First we study the case of Coxian service times. For this infinite-server model we show how to compute all joint moments of the arrival rate and the number of customers, both in transient and steady state. Secondly we consider a Markov-modulated infinite-server queue with general service times. The arrival intensity is resampled at the state change epochs of the Markov background process. We develop a procedure to obtain all transient moments of the number of customers given the initial state and the initial arrival rate.

1 Introduction

In queueing theory it is often assumed that the arrival process is a Poisson process with a constant rate. This is not always a realistic assumption, and hence various adaptations have been suggested to better reflect reality. We mention three types of adaptation, in which the arrival process still maintains essential elements and properties of the Poisson process. Firstly, there have been studies of queues with time-inhomogeneous Poisson arrivals [9], and of queues with Markov-modulated arrivals in which the arrival process is Poisson with rate λ_i when some underlying Markov process is in state i ([4], Ch. XI). Secondly, there is a growing interest in queues with Cox arrival processes. These are Poisson processes in which the time-dependent arrival intensity, say $\Lambda(t)$, itself is a stochastic process. The variance of the number of arrivals of a Cox process in a given interval is larger than the mean (whereas they are equal for Poisson); this phenomenon is usually called overdispersion. In a few papers, the process $\{\Lambda(t), t \geq 0\}$ was taken to be a shot-noise process. We refer to [2] for an example in insurance mathematics and to [14] for a queueing example. Thirdly, so called *mixture* models have been suggested. In such models, the arrival rate is itself a random variable with some distribution. In the queueing literature, it is not so common to consider mixing distributions [12], but this is different in the finance and insurance literature; e.g., Bühlmann [6] already considered them in 1972 in the context of credibility based dynamic premium rules. Recent studies of single server queues with mixing are [18] (in which the random parameter is sampled once and for all) and [13] (in which the random parameter is resampled in each new busy period).

One class of queueing models for which generalizations of the classical Poisson arrival process have been studied is the class of infinite-server models. Websites provide interesting applications of such systems. The arrival rate of website visits typically is not constant in time, and could depend on an environment (Markov modulation), or vary according to some stochastic process that takes external events into account (e.g., shot-noise driven), or take a new random value after some time (mixing). Numerous variants of stochastic background processes in infinite-server queues have been studied after the pioneering paper of O'Cinneide and Purdue [17]. Jansen *et al.* [11] have considered a general càdlàg stochastic process as background process, whereas [7], [3] and [5] were restricted to the case where the background process is Markov. Next to the already mentioned case of an infinite-server model with a Coxian, shot-noise driven, arrival process [14], also infinite-server systems with the more general Hawkes, or self-exciting, input process have recently been analyzed [8, 15].

The present paper is devoted to the study of infinite-server queues with a mixed Poisson arrival process. We assume that, at exponential intervals, a new value for the arrival rate is sampled. We focus on two different models.

The first model is an $M_\Lambda/Cox_n/\infty$ queue. Here the notation M_Λ is used to emphasize that the arrival process is a Poisson process with a *random* arrival rate. Cox_n indicates that the service times have a Coxian distribution with n phases. The second model allows service times to have a fully general distribution, and furthermore allows Markov modulation: when a Markov background process enters some state i, then a new arrival rate is sampled from some distribution (which is clearly more general than the ordinary Markov modulation, in which one always has rate λ_i if the background state is i). For both models we consider transient and steady state moments of the queue length process; in the slightly simpler first model, we even present a procedure to obtain all joint moments of the arrival rate and the numbers of customers in each Coxian service phase.

The paper is organized as follows. Section 2 is devoted to an analysis of the $M_\Lambda/Cox_n/\infty$ queue. We show how all joint moments of the arrival rate and the number of customers in each of the n Coxian phases can be computed. In Section 3 we consider a Markov-modulated infinite-server queue, with generally distributed service times and with mixing of the arrival rate at modulation epochs. We demonstrate how successive queue length moments can be determined iteratively. Section 4 contains some suggestions for further research.

2 The $M_\Lambda/\mathrm{Cox}_n/\infty$ queue

In this section, we consider an infinite-server queue where the arrival parameter repeatedly resamples after i.i.d. (independent, identically distributed) exponential amounts of time. We shall analyze the behavior of this queue and make comparisons to "standard" infinite-server queues with a fixed deterministic arrival parameter.

Let us first describe our model in detail. We consider an $M/\mathrm{Cox}_n/\infty$-queue where the arrival intensity $\Lambda(t)$ at any time $t \geq 0$ is a random variable. The arrival intensity is resampled at random times generated by a Poisson process with rate γ, and is drawn according to some distribution G_Λ.

The service time distribution we consider is a Coxian with n phases (cf. [4], p. 85). Any customer spends an $\exp(\mu_1)$ amount of time in phase 1. With probability $1-p_1$, the service is thereafter completed. Otherwise the service continues with the next phase, which takes $\exp(\mu_2)$. Again, with probability $1-p_2$ the service is completed, and otherwise the service goes on. This procedure continues up to a maximum of n phases.

The Coxian distribution possesses useful exponential properties, while still remaining relatively general. The parameters of a Coxian distribution can be chosen such that it can approximate any distribution D on $[0,\infty)$ arbitrarily closely; for any such D, there exists a sequence of Coxian distributions weakly converging to D. See Section III.4 of [4] for a more detailed discussion.

Our main object of study is the distribution of the number of customers $X(t)$. Throughout the paper we assume that $X(0) = 0$ (empty system at time 0), which (as shown later) extends to any initial condition on $X(0)$.

More specifically, we are interested in the distribution of the random vector $(\Lambda(t), X_1(t), ..., X_n(t))$, where $X_i(t)$ represents the number of customers in phase i of their service at time t. To uniquely identify this distribution, we consider here the joint transform $E\left(e^{-s\Lambda(t)}\prod_{i=1}^{n} z_i^{X_i(t)}\right)$. For the analysis of our model we use a differential equation method introduced in [15]. We observe $(\Lambda(t), X_1(t), ..., X_n(t))$ during a small interval and thus derive a partial differential equation (PDE) for its joint transform. Unfortunately, we are not able to solve the PDE; however, we are able to extract from this PDE a recursive equation in the various moments of $\Lambda(t)$ and the numbers of customers $X_1(t), \ldots, X_n(t)$. Below, we first derive the PDE. We then discuss its solvability and look at special cases. Finally, we manipulate the equation to find a recursion that allows one to iteratively retrieve all moments of the vector $(\Lambda(t), X_1(t), ..., X_n(t))$. Special attention will be given to the steady state vector $(\Lambda, X_1, ..., X_n)$ and the special Cox_1 case of exponential service times.

2.1 The differential equation

Let z be a shorthand notation for the vector $(z_1, ..., z_n)^t$. The following theorem describes a PDE for $f(s, z, t) := E\left(e^{-s\Lambda(t)}\prod_{i=1}^{n} z_i^{X_i(t)}\right)$.

Theorem 2.1. *The joint transform $f(s, z, t)$ satisfies the following partial* *differential equation with boundary condition* $f(0,1,t)=1$*:*

$$\begin{aligned}-\gamma f(s,z,t) &- \frac{\partial}{\partial t}f(s,z,t)+(1-z_1)\frac{\partial}{\partial s}f(s,z,t) \qquad (2.1)\\ &+ \sum_{i=1}^{n-1}\mu_i(p_iz_{i+1}-z_i+1-p_i)\frac{\partial}{\partial z_i}f(s,z,t)\\ &+ \mu_n(1-z_n)\frac{\partial}{\partial z_n}f(s,z,t)+\gamma f(s,1,t)f(0,z,t)=0.\end{aligned}$$

Proof. We exploit the fact that $\{(\Lambda(t),X_1(t),\dots,X_n(t)),t\geq 0\}$ is a Markov process. At some given time t, let us assume $X_i(t)=k_i$ for each i, and $\Lambda(t)=\lambda$. In $[t,t+h),h\downarrow 0$, three things can happen: a customer completes phase i of its service, at rate $k_i\mu_i$; a customer arrives, at rate λ; the arrival parameter resamples, at rate γ. Hence

$$\begin{aligned}&E\Big(e^{-s\Lambda(t+h)}\prod_{i=1}^n z_i^{X_i(t+h)}\Big|\Lambda(t)=\lambda,X_1(t)=k_1,...,X_n(t)=k_n\Big)\\ &=(1-(\lambda+\gamma+\sum_{i=1}^n k_i\mu_i)h)e^{-s\lambda}\prod_{i=1}^n z_i^{k_i}\\ &+\sum_{i=1}^{n-1}k_i\mu_ih\left(p_ie^{-s\lambda}\frac{z_{i+1}}{z_i}\prod_{i=1}^n z_i^{k_i}+(1-p_i)e^{-s\lambda}\frac{1}{z_i}\prod_{i=1}^n z_i^{k_i}\right)\\ &+k_n\mu_nhe^{-s\lambda}\frac{1}{z_n}\prod_{i=1}^n z_i^{k_i}+\lambda he^{-s\lambda}z_1\prod_{i=1}^n z_i^{k_i}\\ &+\gamma hE\left(e^{-s\Lambda}\right)\prod_{i=1}^n z_i^{k_i}+o(h),\quad h\downarrow 0.\end{aligned}$$

Straightforward calculations now yield

$$\begin{aligned}&f(s,z,t+h)-f(s,z,t)=h\Big(-\gamma f(s,z,t)+(1-z_1)\frac{\partial}{\partial s}f(s,z,t)\\ &+ \sum_{i=1}^{n-1}\mu_i(p_iz_{i+1}-z_i+1-p_i)\frac{\partial}{\partial z_i}f(s,z,t)\\ &+ \mu_n(1-z_n)\frac{\partial}{\partial z_n}f(s,z,t)+\gamma f(s,1,t)f(0,z,t)\Big)+o(h),\quad h\downarrow 0.\end{aligned}$$

Dividing by h and letting $h\downarrow 0$ leads to the PDE (2.1). □

Let the vector $(\Lambda, X_1, ..., X_n)$ be the steady state version of the time-dependent $(\Lambda(t), X_1(t), ..., X_n(t))$, and write $f(s,z) := E(e^{-s\Lambda}\prod_{i=1}^{n} z_i^{X_i})$. The following corollary then immediately follows from Theorem 2.1.

Corollary.

$$\begin{aligned}-\gamma f(s,z) + (1-z_1)\frac{\partial}{\partial s}f(s,z) + \sum_{i=1}^{n-1}\mu_i(p_i z_{i+1} - z_i + 1 - p_i)\frac{\partial}{\partial z_i}f(s,z)\\ + \mu_n(1-z_n)\frac{\partial}{\partial z_n}f(s,z) + \gamma f(s,1)f(0,z) = 0,\end{aligned} \tag{2.2}$$

with $f(0,1) = 1$.

Solving either of the equations (2.1) or (2.2) is no easy task. Both are in the class of semilinear first order PDE's. The usual solution approach would therefore be the method of characteristics (see for example [20]). However, the appearance of $f(0,z,t)$ (and $f(0,z)$) also makes it a delay equation. Hence we cannot use standard techniques to compute an exact solution.

Numerical approaches for solving partial delay differential equations have been considered, though most often for very specific types. Finding a suitable method seems hard, and it remains a problem to solve (2.1) and (2.2) either analytically or numerically.

Remark 1. Taking $z_1 = ... = z_n = 1$, Equation (2.1) becomes $\frac{\partial}{\partial t}f(s,1,t) = 0$. Therefore the marginal transform $E\left(e^{-s\Lambda(t)}\right)$ is independent of t – and hence so is the distribution of $\Lambda(t)$. For each point in time t, $\Lambda(t)$ has distribution G_Λ. Of course, considering $\Lambda(t)$ is still valuable to analyze correlation with $\Lambda(t')$ or $X_i(t')$ for some $t' \in \mathbb{R}$. From now on, when we are not concerned with such correlations, we simply write Λ instead of $\Lambda(t)$.

Remark 2. Taking $s = 0$ in (2.2) yields

$$\begin{aligned}(1-z_1)\frac{\partial}{\partial s}f(s,z)\Big|_{s=0} + \sum_{i=1}^{n-1}\mu_i(p_i z_{i+1} - z_i + 1 - p_i)\frac{\partial}{\partial z_i}f(0,z)\\ + \mu_n(1-z_n)\frac{\partial}{\partial z_n}f(0,z) = 0.\end{aligned}$$

Although this equality does not seem to have a direct interpretation, it does when we take $z_1 = ... = z_n$ and subsequently take $z = 1$. We then get

$$E(\Lambda) = \sum_{i=1}^{n-1} \mu_i(1-p_i)E(X_i) + \mu_n E(X_n). \tag{2.3}$$

With S representing the sojourn time, one could compare this to Little's formula $\lambda = E(X)/E(S)$ for any standard queue. Since the sojourn time equals the service time in our infinite-server setting, $\frac{1}{E(S)}$ is the rate at which a service is completed. Note that

$$\begin{aligned}\frac{E(X)}{E(S)} &= E(X)\sum_{i=1}^{n} P(\text{customer is in phase } i) \times (\text{completion rate in state } i)\\ &= E(X)\sum_{i=1}^{n-1}\frac{E(X_i)}{E(X)}\cdot\mu_i(1-p_i) + E(X)\frac{E(X_n)}{E(X)}\cdot\mu_n\\ &= \sum_{i=1}^{n-1}\mu_i(1-p_i)E(X_i) + \mu_n E(X_n),\end{aligned}$$

so that Little's formula indeed holds with λ replaced by $E(\Lambda)$.

2.2 Calculating moments

One important feature of a generating function or Laplace-Stieltjes transform is the ability to quickly extract any moments for the corresponding random variable. In this section we show that the PDE (2.1) provides enough information to do just that. More specifically, it allows us to calculate any moment of $(\Lambda(t), X_1(t), ..., X_n(t))$ by solving a recursion.

Let k denote the vector $(k_1, ..., k_n)^t$. If we would have an expression for $f(s, z, t)$, then the $(l, k)^{\text{th}}$ moment could be calculated by

$$E\Big(\Lambda(t)^l \prod_{i=1}^{n} \frac{X_i(t)!}{(X_i(t)-k_i)!}\Big) = (-1)^l \left[\frac{d^l}{ds^l}\prod_{i=1}^{n}\frac{d^{k_i}}{dz_i^{k_i}} f(s,z,t)\right]_{s=0,z=1}. \tag{2.4}$$

Remark 3. Formally the moment should be defined as $E\Big(\Lambda(t)^l \prod\limits_{i=1}^{n} X_i(t)^{k_i}\Big)$ rather than $E\Big(\Lambda(t)^l \prod\limits_{i=1}^{n} \frac{X_i(t)!}{(X_i(t)-k_i)!}\Big)$. However, the former can easily be obtained from the latter.

Applying these same differentiations to PDE (2.1) yields, with d_x^m denoting the m-th derivative w.r.t. x:

$$\begin{aligned}-\gamma d_s^l d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) - \frac{\partial}{\partial t} d_s^l d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) \\ + (1-z_1) d_s^{l+1} d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) - k_1 d_s^{l+1} d_{z_1}^{k_1-1} d_{z_2}^{k_2} \cdots d_{z_n}^{k_n} f(s,z,t) \\ + \sum_{i=1}^{n-1} \mu_i (p_i z_{i+1} - z_i + 1 - p_i) \frac{\partial}{\partial z_i} d_s^l d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) \\ - \sum_{i=1}^{n-1} k_i \mu_i d_s^l d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) \\ + \sum_{i=1}^{n-1} k_{i+1} p_i \mu_i d_s^l d_{z_1}^{k_1} \cdots d_{z_{i-1}}^{k_{i-1}} d_{z_i}^{k_i+1} d_{z_{i+1}}^{k_{i+1}-1} d_{z_{i+2}}^{k_{i+2}} \cdots d_{z_n}^{k_n} f(s,z,t) \\ + \mu_n (1-z_n) \frac{\partial}{\partial z_n} d_s^l d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) - k_n \mu_n d_s^l d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(s,z,t) \\ + \gamma d_s^l f(s,1,t) d_{z_1}^{k_1} \cdots d_{z_n}^{k_n} f(0,z,t) = 0.\end{aligned}$$

Now define $E(l, k_1, ..., k_n, t) := E\Big(\Lambda(t)^l \prod_{i=1}^{n} \frac{X_i(t)!}{(X_i(t)-k_i)!}\Big)$ as a shorthand notation for the joint moment. When we take $s = 0$, $z = 1$ and use (2.4) we obtain

$$\begin{aligned}\frac{d}{dt} E(l, k_1, ..., k_n, t) = &-\left(\gamma + \sum_{i=1}^{n} k_i \mu_i\right) E(l, k_1, ..., k_n, t) \\ &+ k_1 E(l+1, k_1 - 1, k_2, ..., k_n, t) \\ &+ \sum_{j=1}^{n-1} p_j \mu_j k_{j+1} E(l, k_1, ..., k_{j-1}, k_j + 1, k_{j+1} - 1, k_{j+2}, ..., k_n, t) \\ &+ \gamma E(l, 0, ..., 0, t) E(0, k_1, ..., k_n, t).\end{aligned} \tag{2.5}$$

Our aim is to show that the recursive formula (2.5) allows one to iteratively compute all moments of $(\Lambda(t), X_1(t), ..., X_n(t))$, assuming all moments of G_Λ are known. Let us first make the assumption that the system is empty at $t = 0$. If not, we split $X_j(t)$ into the customers that arrived before or after $t = 0$. The contribution to $X_j(t)$ of arrivals before time 0 can be calculated if we know $(X_1(0), ..., X_n(0))$, by finding the probabilities $p_{ij}(t)$ that a customer who was in phase i at time 0 is in phase j at time t. Conditioning on the case that it takes exactly u time to move from phase i to phase j,

$$\begin{aligned} p_{ij}(t) &= p_i \cdots p_{j-1} \int\limits_0^t \sum_{k=i}^{j-1} C_{k,j-1} \mu_k e^{-\mu_k u} e^{-\mu_j (t-u)} du \\ &= p_i \cdots p_{j-1} \sum_{k=i}^{j-1} C_{k,j-1} \frac{\mu_k}{\mu_k - \mu_j} \left(e^{-\mu_j t} - e^{-\mu_k t} \right). \end{aligned}$$

Here $C_{k,j-1} = \prod\limits_{m \in \{i,...,j-1\} \setminus \{k\}} \frac{\mu_m}{\mu_m - \mu_k}$ and $\sum\limits_{k=i}^{j-1} C_{k,j-1} \mu_k e^{-\mu_k u}$ is the density of a hypo-exponential distribution (see [19], p. 310). Recall that a hypo-exponential random variable is a sum of independent exponentials (in our case $j - i$ exponentials with rates μ_i,...,μ_{j-1} respectively). It can also be seen as a Coxian random variable for which all continuation probabilities $p_k = 1$.

Let $Y_{ij}(t) \sim \text{Bernoulli}(p_{ij}(t))$ for $i \le j$, and let $Y_{ij,m}(t)$, $m = 1, 2, \dots$ be i.i.d. copies of $Y_{ij}(t)$. Then the fraction of customers from $X_j(t)$ that arrived before $t = 0$ equals

$$\sum_{i=1}^{j} \sum_{m=1}^{X_i(0)} Y_{ij,m}(t).$$

Now that we have found the fraction of customers from $X_j(t)$ that arrived before $t = 0$, let us assume for the remainder of the section that the system is empty at time 0.

Theorem 2.2. *All moments of* $(\Lambda(t), X_1(t), ..., X_n(t))$ *can be computed using (2.5), and have the form*

$$E\Big(\Lambda(t)^l \prod_{i=1}^{n} \frac{X_i(t)!}{(X_i(t) - k_i)!}\Big) = \sum_{j=1}^{N} a_j e^{-b_j t}, \tag{2.6}$$

for some constants $a_1, ..., a_N$ *and nonnegative constants* $N, b_1, ..., b_N$.

Proof. We provide an inductive argument. In Remark 1 we have seen that the distribution of $\Lambda(t)$ is independent of t. In particular, $E(\Lambda^l(t))$ is constant in t, so the statement is true when taking $k_1 = ... = k_n = 0$.

Let us take a look at the structure of the recursion (2.5). The derivative of the moment with $(l, k_1, ..., k_n)$ equals a linear function consisting of four unknown components: (1) the moment itself, (2) the same moment with $k_1 - 1$ and $l + 1$, (3) the same moment with $k_j - 1$ and $k_{j-1} + 1$ for some $j = 1, ..., n - 1$, and (4) the same moment with $l = 0$.

Let $K = \sum_{i=1}^{n} k_i$. We have shown above that the statement is true for $K = 0$. The inductive step is split into two parts. First assume we know all moments of order $(l, K - 1)$ and we want to calculate all moments of order $(0, K)$. Note that when substituting $(0, K)$, component 4 coincides with component 1, and component 2 is known by assumption. From $E(1, K - 1, 0, ..., 0, t)$, we can calculate an arbitrary moment $E(0, k_1, ..., k_n, t)$ of order $(0, K)$ with the following scheme:

$$\begin{aligned} &(1, K-1, 0, ..., 0) \to (0, K, 0, ..., 0) \to (0, K-1, 1, 0, ..., 0) \\ \to &(0, K-2, 2, 0, ..., 0) \to ... \to (0, k_1, K-k_1, 0, ..., 0) \to ... \to (0, k_1, ..., k_n). \end{aligned}$$

Note that for each arrow, a differential equation needs to be solved.

For the second part of the induction step, we assume that all moments of order $(l', K-1)$ and $(0, K)$ are known, and want to calculate all moments of order (l, K). In this case, both components 2 and 4 are known by assumption. The arbitrary moment $E(l, k_1, ..., k_n, t)$ of order (l, K) can now be calculated using the following scheme:

$$\begin{aligned} &(l+1, K-1, ..., 0) \to (l, K, 0, ..., 0) \to (l, K-1, 1, 0, ..., 0) \\ \to &(l, K-2, 2, 0, ..., 0) \to ... \to (l, k_1, K-k_1, 0, ..., 0) \to ... \to (l, k_1, ..., k_n). \end{aligned}$$

Combining both schemes, we can iteratively find an expression for any moment $E(l, k_1, ..., k_n, t)$. All that remains is to show that the moments take the form of (2.6). This will be done by induction on the number of iterations.

Assume all moments calculated thus far take the form of (2.6). Note that each iteration requires the differential equation (2.5) to be solved. This differential equation has the form

$$\frac{d}{dt} g(t) = -C g(t) + \sum_{m=1}^{M} \tilde{a}_m g_m(t),$$

where $C, M, \tilde{a}_m$ are constants of which C, M are nonnegative and $g_m(t)$ are known functions from previous iterations. Since the $g_m(t)$ have the required form by assumption, it holds that

$$\frac{d}{dt}g(t) = -Cg(t) + \sum_{m=1}^{M} \tilde{a}_m \sum_{j=1}^{N_m} a_{j,m} e^{-b_{j,m}t} = -Cg(t) + \sum_{m=1}^{N'} a'_m e^{-b'_m t}, \quad (2.7)$$

for some constants C, N', a'_m, b'_m of which C, N', b'_m are nonnegative. Solving this differential equation quickly leads to the result of the theorem. □

Remark 4. It is an interesting question how many different terms a moment consists of, given its order $(l, k_1, ..., k_n)$. In other words: how large is N in (2.6)? To answer this, note that when we solve the differential equation (2.7) the only exponent we have not seen in previous iterations is $-Ct$. So N is at most the number of possible values of C. Checking with (2.5), we conclude that $N = O\left(\left(\sum_{i=1}^{n} k_i\right)^n\right)$.

In terms of computational complexity, observe that (2.5) has $O(n)$ terms, each one consisting of one moment. Since a moment has $O\left(\left(\sum_{i=1}^{n} k_i\right)^n\right)$ terms, one iteration requires $O\left(n \cdot \left(\sum_{i=1}^{n} k_i\right)^n\right)$ operations. Note also that we need one iteration for each moment of lesser order. Therefore, the calculation of a moment of order $(l, k_1, ..., k_n)$ requires $O\left(n \cdot \left(\sum_{i=1}^{n} k_i\right)^{2n}\right)$ operations.

Now that we have seen that all moments can be retrieved, and which form they have, let us calculate some of these moments. This gives an idea how the system behaves in general, and presents a quantitative way to compare our queue to other models.

One can check that expressions from higher moments become very large and complicated, requiring much computational effort and making analysis hard. However, for some lower moments and special cases, we can derive "nicer" expressions. This enables us to give intuitive explanations and perform proper analysis on the formulas we find.

We start with an expression for the mean number of customers.

Theorem 2.3. *Let* $C_{i,j} = \prod_{m \neq i}^{j} \frac{\mu_m}{\mu_m - \mu_i}$. *Given that the system is empty at*

$t = 0$, the mean number of customers in phase j at time t equals

$$E(X_j(t)) = \left(\prod_{i=1}^{j-1} p_i\right) \frac{E(\Lambda)}{\mu_j} \left(1 - \sum_{i=1}^{j} C_{i,j} e^{-\mu_i t}\right), \qquad j = 1, ..., n. \tag{2.8}$$

Proof. Let us first consider the case $j = 1$. Substituting $k_1 = 1$ and $l = k_2 = ... = k_n = 0$ gives the differential equation

$$\begin{aligned}\frac{d}{dt}E(X_1(t)) &= -(\gamma + \mu_1)E(X_1(t)) + E(\Lambda(t)) + \gamma E(X_1(t)) \\ &= -\mu_1 E(X_1(t)) + E(\Lambda).\end{aligned}$$

With $E(X_1(0)) = 0$, its solution is

$$E(X_1(t)) = \frac{E(\Lambda)}{\mu_1}\left(1 - e^{-\mu_1 t}\right). \tag{2.9}$$

Therefore (2.8) holds for $j = 1$.
Now let $j \in \{2, ..., n\}$. The differential equation corresponding to $k_j = 1$ is

$$\begin{aligned}\frac{d}{dt}E(X_j(t)) &= -(\gamma + \mu_j)E(X_j(t)) + \mu_{j-1}p_{j-1}E(X_{j-1}(t)) + \gamma E(X_j(t)) \\ &= -\mu_j E(X_j(t)) + \mu_{j-1}p_{j-1}E(X_{j-1}(t)). \qquad (2.10)\end{aligned}$$

To show that (2.8) is indeed a solution to this equation, we differentiate (2.8) and rearrange terms, using that $\frac{\mu_j - \mu_i}{\mu_j} C_{i,j} = C_{i,j-1}$:

$$\begin{aligned}\frac{d}{dt}E(X_j(t)) &= \left(\prod_{i=1}^{j-1} p_i\right) E(\Lambda) \sum_{i=1}^{j} \frac{\mu_i}{\mu_j} C_{i,j} e^{-\mu_i t} \\ &= \left(\prod_{i=1}^{j-1} p_i\right) E(\Lambda) \left(\sum_{i=1}^{j-1} \left(1 - \frac{\mu_j - \mu_i}{\mu_j}\right) C_{i,j} e^{-\mu_i t} + C_{j,j} e^{-\mu_j t}\right) \\ &= \left(\prod_{i=1}^{j-1} p_i\right) E(\Lambda) \left(-\left(1 - \sum_{i=1}^{j} C_{i,j} e^{-\mu_i t}\right) + \left(1 - \sum_{i=1}^{j-1} C_{i,j-1} e^{-\mu_i t}\right)\right) \\ &= -\mu_j E(X_j(t)) + \mu_{j-1}p_{j-1}E(X_{j-1}(t)).\end{aligned}$$

To verify that solution (2.8) to differential equation (2.10) is the desired solution, we need to show that (2.8) satisfies the boundary condition $E(X_j(0)) =$

0. So what remains is to check that $\sum\limits_{i=1}^{j} C_{i,j} = 1$. We can do this by observing that

$$\sum_{i=1}^{j} C_{i,j} = \sum_{i=1}^{j} C_{i,j} \cdot \int\limits_{0}^{\infty} \mu_i e^{-\mu_i t} dt = \int\limits_{0}^{\infty} \sum_{i=1}^{j} \mu_i C_{i,j} e^{-\mu_i t} dt.$$

As is proven in ([19], p. 309-310), the integrand is a density function of a random variable on $[0, \infty)$. It follows that $\sum\limits_{i=1}^{j} C_{i,j} = 1$, which concludes the proof. □

For further moments, it is necessary to find $E\left(\Lambda(t)X_1(t)\right)$ first. So with $l = k_1 = 1$ and $k_2 = ... = k_n = 0$, (2.5) gives

$$\begin{aligned} \frac{d}{dt}E(\Lambda(t)X_1(t)) &= -(\gamma+\mu_1)E\left(\Lambda(t)X_1(t)\right) + E(\Lambda^2) + \gamma E(\Lambda)E(X_1(t)) \\ &= -(\gamma+\mu_1)E\left(\Lambda(t)X_1(t)\right) + \frac{\mu_1 E(\Lambda^2) + \gamma E(\Lambda)^2}{\mu_1} - \frac{\gamma E(\Lambda)^2}{\mu_1} e^{-\mu_1 t}. \end{aligned}$$

Here we used the $E(X_1(t))$ we found in (2.9). Combined with the boundary condition $E(\Lambda(0)X_1(0)) = 0$, the solution equals

$$\begin{aligned} E(\Lambda(t)X_1(t)) &= \frac{\text{Var}(\Lambda)}{\gamma+\mu_1} + \frac{E(\Lambda)^2}{\mu_1} - \frac{E(\Lambda)^2}{\mu_1}e^{-\mu_1 t} - \frac{\text{Var}(\Lambda)}{\gamma+\mu_1}e^{-(\gamma+\mu_1)t} \\ &= \frac{E(\Lambda)^2}{\mu_1}\left(1-e^{-\mu_1 t}\right) + \frac{\text{Var}(\Lambda)}{\gamma+\mu_1}\left(1-e^{-(\gamma+\mu_1)t}\right). \end{aligned} \tag{2.11}$$

Next, we use the same trick for $E\left(X_1(t)\left(X_1(t)-1\right)\right)$. The differential equation for $k_1 = 2$ and $l = k_2 = ... = k_n = 0$ is

$$\frac{d}{dt}E\left(X_1(t)\left(X_1(t)-1\right)\right) = -2\mu_1 E\left(X_1(t)\left(X_1(t)-1\right)\right) + 2E(\Lambda(t)X_1(t)),$$

with $E\left(X_1(0)\left(X_1(0)-1\right)\right) = 0$. This has solution

$$\begin{aligned} E\left(X_1(t)\left(X_1(t)-1\right)\right) &= \frac{\mu_1 E(\Lambda^2) + \gamma E(\Lambda)^2}{\mu_1^2(\gamma+\mu_1)} - \frac{2E(\Lambda)^2}{\mu_1^2}e^{-\mu_1 t} \\ + \quad \frac{2\text{Var}(\Lambda)}{(\gamma+\mu_1)(\gamma-\mu_1)}&e^{-(\gamma+\mu_1)t} + \frac{\gamma E(\Lambda)^2 - \mu_1 E(\Lambda^2)}{\mu_1^2(\gamma-\mu_1)}e^{-2\mu_1 t}, \end{aligned} \tag{2.12}$$

and it follows that

$$\begin{aligned}\operatorname{Var}(X_1(t)) = \frac{E(\Lambda)}{\mu_1} + \frac{\operatorname{Var}(\Lambda)}{\mu_1(\gamma+\mu_1)} - \frac{E(\Lambda)}{\mu_1}e^{-\mu_1 t} \\ + \quad \frac{2\operatorname{Var}(\Lambda)}{(\gamma+\mu_1)(\gamma-\mu_1)}e^{-(\gamma+\mu_1)t} - \frac{\operatorname{Var}(\Lambda)}{\mu_1(\gamma-\mu_1)}e^{-2\mu_1 t}.\end{aligned} \tag{2.13}$$

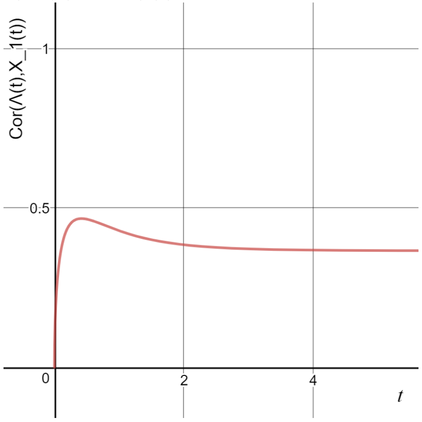
It is easily verified that $\gamma = \mu_1$ is not a singularity.
The correlation between $\Lambda(t)$ and $X_1(t)$ follows from (2.11) and (2.13):

$$\operatorname{Cor}(\Lambda(t), X_1(t)) = \frac{\operatorname{Cov}(\Lambda(t), X_1(t))}{\sqrt{\operatorname{Var}(\Lambda)\operatorname{Var}(X_1(t))}} \tag{2.14}$$

$$= \frac{\left(1-e^{-(\gamma+\mu_1)t}\right)\sqrt{D_\Lambda \mu_1(\gamma-\mu_1)}}{\sqrt{(\gamma+\mu_1)^2(\gamma-\mu_1)(1-e^{-\mu_1 t}) + D_\Lambda(\gamma+\mu_1)\left(\gamma-\mu_1+2\mu_1 e^{-(\gamma+\mu_1)t}-(\gamma+\mu_1)e^{-2\mu_1 t}\right)}},$$

where $D_\Lambda = \frac{\operatorname{Var}(\Lambda)}{E(\Lambda)}$ is the dispersion index. Interestingly, the arrival process manifests itself only through D_Λ. The positive correlation that the formula indicates makes sense, since correlation between $\Lambda(t)$ and $X_1(t)$ can only occur when $\Lambda(t)$ shows variability.

Figure 1: $\operatorname{Cor}(\Lambda(t), X_1(t))$ for $\mu_1 = 1$, $\gamma = 2$ and $D_\Lambda = 2$.



As is visible in Figure 1, there appears to be a time where the correlation is maximal. This will be the time $t_{\max}$ that enough customers could have arrived to show correlation, but a large fraction of $X_1(t_{\max})$ still originates from the current Λ (and not from previous ones).

2.3 Steady state

In this subsection we consider the steady state behavior of the infinite-server queue, again focussing on a recursion for the joint moments of arrival rate and numbers of customers in phases $1,\dots,n$. Let $E\left(l,k_1,...,k_n\right) := E\left(\Lambda^l \prod_{i=1}^{n} \frac{X_i!}{(X_i-k_i)!}\right)$. Letting $t \to \infty$ in (2.5) yields the following steady state recursion:

$$\begin{aligned}
E\left(l,k_1,...,k_n\right) &= \frac{1}{\gamma+\sum_{i=1}^{n} k_i\mu_i}\Bigg(k_1 E\left(l+1,k_1-1,k_2,...,k_n\right) \qquad (2.15)\\
&+ \sum_{j=1}^{n-1} p_j\mu_j k_{j+1} E\left(l,k_1,...,k_{j-1},k_j+1,k_{j+1}-1,k_{j+2},...,k_n\right)\\
&+ \gamma E(l,0,...,0)E\left(0,k_1,...,k_n\right)\Bigg).
\end{aligned}$$

Note that an iteration step only requires some basic operations, whereas in the transient case each iteration required a differential equation to be solved.

Let us find the steady state mean number of customers in phase j. Taking $k_j = 1$ (and everything else 0) yields

$$E(X_j) = \frac{p_{j-1}\mu_{j-1}}{\mu_j}E(X_{j-1}), \quad j=2,3,\dots; \quad E(X_1) = \frac{E(\Lambda)}{\mu_1}.$$

This simple recursion results in

$$E(X_j) = \frac{E(\Lambda)}{\mu_j}\prod_{i=1}^{j-1} p_i, \qquad j=1,...,n, \tag{2.16}$$

which of course we could also have found from (2.8) letting $t \to \infty$.

Remark 5. Recall that the mean number of customers in an ordinary $M/M/\infty$-queue is $\frac{\lambda}{\mu}$, the arrival rate divided by the departure rate. Formula (2.16) represents something similar: the expected arrival rate over the departure rate in phase j. To see this, note that with $\prod_{i=1}^{j-1} p_i$ being the probability that a customer reaches phase j, $E(\Lambda)\prod_{i=1}^{j-1} p_i$ is the expected arrival rate of phase j.

With the steady state mean at hand, Formula (2.8) has a nice interpretation. We consider the steady state number of customers who are in phase j (given in (2.16)), and distinguish the fraction that has been in the system for at least t time, and the fraction that has not. Where the steady state mean consists of both fractions, the transient mean (2.8) only consists of the second (assuming an empty system at time 0). Now note that a customer in phase j has gone through j exponential phases (1 up to $j-1$ completely and the past part of an exponential phase j). So the second fraction is proportional to the probability that j exponential phases are traversed before time t. According to Formula (2.8), this probability should be $\left(1-\sum_{i=1}^{j}\prod_{m\neq i}^{j}\frac{\mu_m}{\mu_m-\mu_i}e^{-\mu_i t}\right)$. Indeed, this is precisely the distribution function of a hypo-exponential random variable.

Another quantity that is greatly simplified by considering steady state is $\mathrm{Cov}(\Lambda, X_j)$. For this we take $l = k_j = 1$ in (2.15) and find

$$E(\Lambda X_j) = \frac{1}{\gamma+\mu_j}\left(p_{i-1}\mu_{i-1}E(\Lambda X_{j-1}) + \gamma E(\Lambda)E(X_j)\right),$$

such that

$$\begin{aligned}\mathrm{Cov}(\Lambda, X_j) &= \frac{1}{\gamma+\mu_j}\left(p_{i-1}\mu_{i-1}\mathrm{Cov}(\Lambda, X_{j-1}) + p_{i-1}\mu_{i-1}E(\Lambda)E(X_{j-1}) - \mu_i E(\Lambda)E(X_j)\right)\\ &= \frac{p_{i-1}\mu_{i-1}}{\gamma+\mu_j}\mathrm{Cov}(\Lambda, X_{j-1}), \quad j = 2, 3, \dots,\end{aligned}$$

returns a simple recursion. Observe also that

$$\begin{aligned}&\mathrm{Cov}(\Lambda, X_1) = E(\Lambda X_1) - E(\Lambda)E(X_1)\\ = \; &\frac{E(\Lambda^2)}{\gamma+\mu_1} + \frac{\gamma}{\gamma+\mu_1}E(\Lambda)E(X_1) - E(\Lambda)E(X_1) = \frac{\mathrm{Var}(\Lambda)}{\gamma+\mu_1},\end{aligned}$$

hence the solution is

$$\mathrm{Cov}(\Lambda, X_j) = \frac{\mathrm{Var}(\Lambda)}{\gamma+\mu_j}\prod_{i=1}^{j-1}\frac{p_i\mu_i}{\gamma+\mu_i}, \qquad j = 1, ..., n. \tag{2.17}$$

This formula also provides an interesting interpretation. Recall that $\frac{\mu}{\gamma+\mu}$ is the probability that an exponential time with rate μ is shorter than an exponential time with rate γ. Therefore, $\prod_{i=1}^{j-1} \frac{p_i\mu_i}{\gamma+\mu_i}$ is the probability that a customer traverses $j-1$ service phases before the arrival parameter resamples. The appearance of this probability is intuitive, since the current Λ can only influence the current X_j if customers generated from this Λ arrive in phase j before a new Λ is drawn.

Next, we check the variance of the number of customers in phase 2. To do this we substitute $k_2 = 2$ in (2.15). In the calculation we also use (2.17) for $j = 2$ and (2.12) for $t \to \infty$:

$$\begin{aligned}
E(X_2(X_2-1)) &= \frac{p_1\mu_1}{\mu_2} E(X_1X_2) \\
&= \frac{p_1\mu_1}{\mu_2(\mu_1+\mu_2)} E(\Lambda X_2) + \frac{p_1^2\mu_1^2}{\mu_2(\mu_1+\mu_2)} E(X_1(X_1-1)) \\
&= \frac{p_1\mu_1}{\mu_2(\mu_1+\mu_2)} \left(\frac{p_1\mu_1 \mathrm{Var}(\Lambda)}{(\gamma+\mu_1)(\gamma+\mu_2)} + \frac{p_1 E(\Lambda)^2}{\mu_2} \right) \\
&+ \frac{p_1^2\mu_1^2}{\mu_2(\mu_1+\mu_2)} \frac{\mu_1 E(\Lambda^2) + \gamma E(\Lambda)^2}{\mu_1^2(\gamma+\mu_1)} \\
&= \frac{p_1^2\mu_1(\gamma+\mu_1+\mu_2)\mathrm{Var}(\Lambda)}{\mu_2(\mu_1+\mu_2)(\gamma+\mu_1)(\gamma+\mu_2)} + \frac{p_1^2 E(\Lambda)^2}{\mu_2^2},
\end{aligned}$$

so that

$$\mathrm{Var}(X_2) = \frac{p_1^2\mu_1(\gamma+\mu_1+\mu_2)\mathrm{Var}(\Lambda)}{\mu_2(\mu_1+\mu_2)(\gamma+\mu_1)(\gamma+\mu_2)} + \frac{p_1 E(\Lambda)}{\mu_2}. \tag{2.18}$$

Now we can calculate the correlation

$$\begin{aligned}
\mathrm{Cor}(\Lambda, X_2) &= \frac{\frac{p_1\mu_1\mathrm{Var}(\Lambda)}{(\gamma+\mu_1)(\gamma+\mu_2)}}{\sqrt{\frac{p_1\mathrm{Var}(\Lambda)}{\mu_2}\left(E(\Lambda) + \frac{p_1\mu_1(\gamma+\mu_1+\mu_2)\mathrm{Var}(\Lambda)}{(\mu_1+\mu_2)(\gamma+\mu_1)(\gamma+\mu_2)}\right)}} \\
&= \mu_1 \sqrt{\frac{p_1\mu_2(\mu_1+\mu_2)D_\Lambda}{(\gamma+\mu_1)^2(\gamma+\mu_2)^2(\mu_1+\mu_2) + p_1\mu_1(\gamma+\mu_1)(\gamma+\mu_2)(\gamma+\mu_1+\mu_2)D_\Lambda}}.
\end{aligned}$$

In the special case that $\mu_1 = \mu_2 = \mu$, we have

$$\mathrm{Cor}(\Lambda, X_2) = \frac{\mu}{\gamma+\mu}\sqrt{\frac{2p_1\mu D_\Lambda}{2(\gamma+\mu)^2 + p_1(\gamma+2\mu)D_\Lambda}}.$$

To compare, we let $t \to \infty$ in (2.14) to get $\mathrm{Cor}(\Lambda, X_1)$:

$$\mathrm{Cor}(\Lambda, X_1) = \sqrt{\frac{D_\Lambda \mu_1}{(\gamma+\mu_1)^2 + D_\Lambda(\gamma+\mu_1)}}. \tag{2.19}$$

In case the service rates are equal (i.e. $\mu_1 = \mu_2 = \mu$) it holds that

$$\begin{aligned}\frac{\mathrm{Cor}(\Lambda, X_1)}{\mathrm{Cor}(\Lambda, X_2)} &= \frac{\gamma+\mu}{\mu}\sqrt{\frac{2(\gamma+\mu)^2 + p_1(\gamma+2\mu)D_\Lambda}{2p_1\left((\gamma+\mu)^2 + D_\Lambda(\gamma+\mu)\right)}} \\ &= \sqrt{\frac{2(\gamma+\mu)^3 + 2p_1\mu^2 D_\Lambda + p_1(\gamma^2+3\gamma\mu)D_\Lambda}{2p_1\mu^2(\gamma+\mu) + 2p_1\mu^2 D_\Lambda}} \;\geq\; 1,\end{aligned}$$

because $p_1 \leq 1$. An immediate consequence from this calculation is that $\mathrm{Cor}(\Lambda, X_1) \geq \mathrm{Cor}(\Lambda, X_2)$ when the service rates are equal. This is not surprising, since customers “generated” by the current arrival rate Λ immediately arrive in phase 1, but take some time before entering phase 2 (if they remain in the system after phase 1). So earlier phases have a more direct connection with the current arrival rate.

The inequality does not necessarily hold for different service rates. To see this, note that when p_1 is high and μ_1 is high compared to μ_2 and γ, customers spend more time in phase 2 than in phase 1 before the arrival parameter resamples.

Let us finally take a look at $X_1 + X_2$, which is the total number of customers for $n = 2$. Its mean and variance can easily be computed, but its correlation with Λ is less straightforward. To find it, we first have to determine $\mathrm{Cov}(X_1, X_2)$. This is done by taking $k_1 = k_2 = 1$ in (2.15), and using the moments given in (2.17), (2.12) and (2.16):

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= E(X_1X_2) - E(X_1)E(X_2)\\ &= \frac{1}{\mu_1+\mu_2}E(\Lambda X_2) + \frac{p_1\mu_1}{\mu_1+\mu_2}E\left(X_1(X_1-1)\right) - E(X_1)E(X_2)\\ &= \frac{1}{\mu_1+\mu_2}\left(\mathrm{Cov}(\Lambda, X_2) + E(\Lambda)E(X_2)\right)\\ &+ \lim_{t\to\infty}\frac{p_1\mu_1}{\mu_1+\mu_2}E\left(X_1(t)(X_1(t)-1)\right) - E(X_1)E(X_2)\\ &= \frac{p_1(\gamma+\mu_1+\mu_2)\mathrm{Var}(\Lambda)}{(\gamma+\mu_1)(\gamma+\mu_2)(\mu_1+\mu_2)}.\end{aligned} \tag{2.20}$$

Then with (2.17), (2.13), (2.18) and (2.20), we find

$$\mathrm{Cor}(\Lambda, X_1+X_2) = \frac{\mathrm{Cov}(\Lambda, X_1) + \mathrm{Cov}(\Lambda, X_2)}{\sqrt{\mathrm{Var}(\Lambda)\left(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)\right)}}$$

$$\begin{aligned} = \;& (\gamma + p_1\mu_1 + \mu_2) \\ \times \;& \sqrt{\frac{\mu_1\mu_2(\mu_1+\mu_2)D_\Lambda}{(\gamma+\mu_1)(\gamma+\mu_2)\left((p_1\mu_1+\mu_2)(\gamma+\mu_1)(\gamma+\mu_2)(\mu_1+\mu_2) + \mu_1\mu_2\gamma D_\Lambda + (p_1\mu_1+\mu_2)^2(\gamma+\mu_1+\mu_2)D_\Lambda\right)}}.\end{aligned} \tag{2.21}$$

Note that when $p_1 = 0$, customers never reach phase 2. It is easy to check that in that case, it indeed holds that $\mathrm{Cor}(\Lambda, X_1 + X_2) = \mathrm{Cor}(\Lambda, X_1)$.

2.4 $M_\Lambda/M/\infty$

In this section we give a few results for the special Coxian case where the service time distribution is exponential with parameter μ. An easy way to recover the main results is by observing that this service time distribution can be obtained by taking $p_1 = ... = p_{n-1} = 0$ and $\mu_1 = \mu$. We then find

$$-\gamma f(s,z) + (1-z)\frac{\partial}{\partial s}f(s,z) + \mu(1-z)\frac{\partial}{\partial z}f(s,z) + \gamma f(s,1)f(0,z) = 0 \tag{2.22}$$

as the partial differential equation for the joint transform of (Λ, X) and

$$E\left(\Lambda^l\frac{X!}{(X-k)!}\right) = \frac{k}{\gamma+k\mu}E\left(\Lambda^{l+1}\frac{X!}{(X-k+1)!}\right) + \frac{\gamma}{\gamma+k\mu}E\left(\Lambda^l\right)E\left(\frac{X!}{(X-k)!}\right) \tag{2.23}$$

as the recursion for moments. The observant reader may have noticed that all p_i, $i > 1$ are irrelevant when $p_1 = 0$, because customers can only be in phase 1. The extended assumption is made to more easily copy results from Coxian service times.

If $\Lambda \sim \exp(\theta)$ we can derive an expression for the form of any moment.

Theorem 2.4. *Let $\Lambda \sim exp(\theta)$. Then the recursion (2.23) has solution*

$$E\left(\Lambda^l \frac{X!}{(X-k)!}\right) = \frac{l!}{\theta^{l+k} \prod_{i=1}^{k} (\gamma + i\mu)} \sum_{j=0}^{k} c_j(l,k) \left(\frac{\gamma}{\mu}\right)^{k-j}, \tag{2.24}$$

where $c_0(l,k) = 1$, $c_k(l,k) = \frac{k!(l+k)!}{l!}$ and $c_j(l,k)$ satisfies the recursion

$$c_j(l,k) = k(l+1)c_{j-1}(l+1,k-1) + c_j(1,k-1), \qquad j = 1,...,k-1. \tag{2.25}$$

For a proof by induction, we refer to [16]. Let us here only check that (2.25) has a solution. The key observation is that the second argument lowers in each step of the recursion. Note that when we express $c_j(l,k)$ into coefficients of the form $c_{j'}(l',k-1)$ by using (2.25), we have either $j' = 0$, $j' \in \{1,...,k-2\}$ or $j' = k-1$. In the first and third case, the value is given, and in the second case we can use (2.25) again. Repeating this procedure $k-1$ times leaves us with only unknown terms with $k = 1$. Since $j \leq k$, the remaining coefficients are of the form $c_0(l',1)$ and $c_1(l',1)$. These are both known, so the recursion ends after $k-1$ steps.

As we have seen, the mean number of customers is similar in $M/M/\infty$ and $M_\Lambda/M/\infty$, the only difference being that the fixed arrival rate λ is replaced by the expected arrival rate $E(\Lambda)$. Mixing turns out to have a larger effect on the variance. Where without mixing, the variance of X is equal to the mean, our model has

$$\text{Var}(X) = \frac{E(\Lambda)}{\mu} + \frac{\text{Var}(\Lambda)}{\mu(\gamma+\mu)}, \tag{2.26}$$

as a consequence of (2.13) by taking $\mu_1 = \mu$ and $t \to \infty$. The first term is the variance without mixing, and the second is caused by the variability of the arrival rate. It is therefore no surprise that the second term is linear in Var(Λ).

Heemskerk, van Leeuwaarden and Mandjes [10] provide another interesting comparison for the variance. They consider the same model, the only difference being *deterministic* resample intervals with common length Δ. For the variance of the number of customers Heemskerk *et al.* [10] obtain

$$\operatorname{Var}(X_H) = \frac{E(\Lambda)}{\mu} + \frac{\left(1 - e^{-\mu\Delta}\right)\operatorname{Var}(\Lambda)}{(1 + e^{-\mu\Delta})\mu^2}, \tag{2.27}$$

where X_H denotes the steady state number of customers in this model.

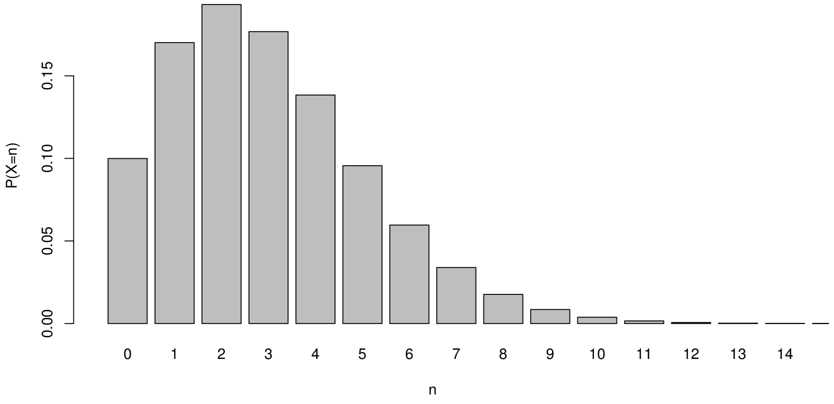
For a fair comparison between (2.26) and (2.27) we take $\gamma = \frac{1}{\Delta}$, such that the mean resample interval lengths of both models agree. It clearly holds that $\operatorname{Var}(X) > \operatorname{Var}(X_H)$ if and only if $\frac{1}{\gamma+\mu} > \frac{1-e^{-\mu\Delta}}{\mu(1+e^{-\mu\Delta})}$. Hence

$$\operatorname{Var}(X) > \operatorname{Var}(X_H) \qquad \text{if and only if} \qquad 1 + 2\mu\Delta > e^{\mu\Delta} \quad (\text{so } \mu\Delta \lesssim 1.25). \tag{2.28}$$

One might expect that randomness of resample interval lengths leads to more variability in X. Surprisingly, this result shows that is not always true.

For a good image of the distribution of X, we use a simple simulation. Keeping track of the number of customers in the $M_\Lambda/M/\infty$, the results for simulating up to $t = 10^6$ can be seen in Figure 2.

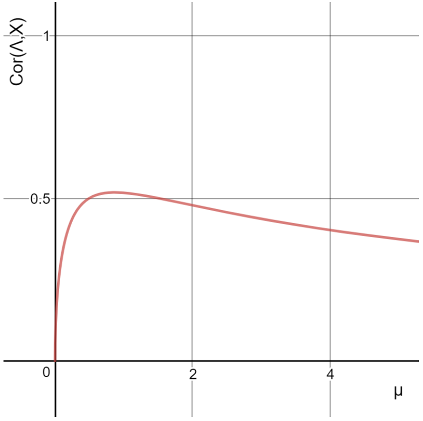
Figure 2: Histogram estimating the distribution of X, with $\Lambda \sim U(0, 6)$ and $\gamma = \mu = 1$.



Note that with $\Lambda \sim U(0, 6)$ and $\gamma = \mu = 1$, we have $E(X) = 3$ and $\operatorname{Var}(X) = 4.5$. On a single run, the simulation rarely returns an error of more than 0.01 for each quantity. For more details regarding the simulation, see [16].

The last quantity we analyze regarding the steady state vector (Λ, X) is the correlation between Λ and X. It is given by (2.19), replacing μ_1 by μ as is explained at the start of this section.

Figure 3: $\mathrm{Cor}(X, \Lambda)$ for $D_\Lambda = 1$ and $\gamma = 0.5$.



In Figure 3 we see that the correlation is low both when the service rate is very low or very high. This can be explained as follows: for the correlation to be high, on the one hand the service rate should not be too high since we want customers generated by the current Λ to stay as long as possible. On the other hand, the service rate should not be so low that "old" customers from previously sampled Λ stay in the system for too long. The correlation turns out to be maximal for $\mu = \sqrt{\gamma^2 + \frac{\gamma}{D_\Lambda}}$, attaining the value

$$\mathrm{Cor}(X, \Lambda) = \sqrt{\frac{\sqrt{\frac{\gamma(1+D_\Lambda\gamma)}{D_\Lambda}}}{D_\Lambda\left(2\gamma^2 + \frac{\gamma}{D_\Lambda} + 2\gamma\sqrt{\gamma^2 + \frac{\gamma}{D_\Lambda}}\right) + \left(\gamma + \sqrt{\gamma^2 + \frac{\gamma}{D_\Lambda}}\right)}}$$
$$= \sqrt{\frac{1}{2\sqrt{D_\Lambda\gamma(1 + D_\Lambda\gamma)} + 2D_\Lambda\gamma + 1}}.$$

Notice its dependence only on $D_\Lambda\gamma$.

We close this section with a brief discussion of scaling limits. We consider a scaling of the arrival rate $\Lambda(t)$ with a factor N together with a scaling of the resample rate γ with factor N^α. The value of α determines which rate speeds up faster. If $\alpha > 1$, the arrival rate changes very often, while if $\alpha < 1$, the arrival rate rarely changes relative to the number of arrivals. We check the behavior of the $M_\Lambda/M/\infty$ under this scaling and discuss the effect of the value of α. We follow the idea of Heemskerk *et al.* ([10], p. 8-12), in which they consider a model with the arrival parameter resampling after a fixed time Δ.

Let $X^{(N)}$ be the steady state number of customers under the corresponding scaling. Then we easily see that

$$E\left(X^{(N)}\right) = \frac{NE(\Lambda)}{\mu},$$

and from (2.26) that

$$\begin{aligned} \operatorname{Var}\left(X^{(N)}\right) &= \frac{NE(\Lambda)}{\mu} + \frac{N^2\operatorname{Var}(\Lambda)}{\mu\left(N^\alpha\gamma + \mu\right)} \\ &\sim \frac{NE(\Lambda)}{\mu} + \frac{N^{2-\alpha}\operatorname{Var}(\Lambda)}{\mu\gamma} \end{aligned} \tag{2.29}$$

when N tends to infinity. This can be compared to the asymptotic variance of Heemskerk *et al.* [10]. Denoting by $X_H^{(N)}$ the number of customers in their model, they find

$$\operatorname{Var}\left(X_H^{(N)}\right) \sim \frac{NE(\Lambda)}{\mu} + \frac{\Delta N^{2-\alpha}\operatorname{Var}(\Lambda)}{2\mu}. \tag{2.30}$$

In both models, the variance increases as the average length of a resample interval ($\frac{1}{\gamma}$ and Δ) increases. However, for the variances to be equal, it must be that $\frac{1}{\gamma} = \frac{\Delta}{2}$. In other words, our more random model needs twice as many resamples to obtain the same asymptotic variance.

From (2.29), the limiting behavior of the variance is

$$\operatorname{Var}\left(X^{(N)}\right) \sim \frac{NE(\Lambda)}{\mu}\mathbb{1}_{\{\alpha\geq 1\}} + \frac{N^{2-\alpha}\operatorname{Var}(\Lambda)}{\mu\gamma}\mathbb{1}_{\{\alpha\leq 1\}}. \tag{2.31}$$

Let $\beta = \max\{1, 2-\alpha\}$. The asymptotic behavior suggests the following central limit theorem.

Conjecture 2.5. *Assume all moments of Λ are finite. If $N \to \infty$, then*

$$\frac{X^{(N)} - E\left(X^{(N)}\right)}{\sqrt{N^{\gamma}}} \to N(0, \sigma^2), \tag{2.32}$$

with $\sigma^2 = \frac{E(\Lambda)}{\mu}\mathbb{1}_{\{\alpha \geq 1\}} + \frac{\mathrm{Var}(\Lambda)}{\mu\gamma}\mathbb{1}_{\{\alpha \leq 1\}}$.

The statement of the conjecture is very similar to Theorem 2.1 from [10]. However, in the proof they use that the resample interval lengths are constant. This property is critical in the proof, making it hard to prove our conjecture using the same idea. Nonetheless, it is expected that the statement holds in our model as well.

3 Mixing in Markov-Modulated Infinite-Server Queues

In this section, we consider a Markov-modulated infinite-server queue. The Markov background process moves through a finite number of states $\mathcal{N} = \{1, ..., n\}$; be aware that this n has nothing to do with the number of phases in the Coxian distribution considered in the previous section. While the process is in state $i \in \mathcal{N}$, it takes an exponential amount of time with parameter γ_i before the state changes. When it does, it has probability p_{ij} to move to state j. Of course, $p_{ij} \geq 0$ for each $(i, j) \in \mathcal{N} \times \mathcal{N}$ and $\sum\limits_{j=1}^{n} p_{ij} = 1$ for each $i \in \mathcal{N}$. Note that the Markov process behaves independently of the queue and acts only as a generator.

In state $i \in \mathcal{N}$, customers arrive according to a Poisson process with random rate Λ_i, drawn according to a probability density function $g_{\Lambda_i}(\cdot)$. This rate is drawn at the start of the current state, and resampled once the state changes again; the arrival parameter also resamples if the Markov process moves to the *same* state. B_i is the service time of a customer arriving when the background process is in state i.

The difference with other Markov-modulated queueing models lies within the random parameter Λ_i. In standard Markov-modulated queues, the interarrival times have a fixed distribution depending on i; most commonly an exponential distribution with rate λ_i. In our case, when we apply mixing, the interarrival time distribution is exponential with a *random* rate Λ_i.

Remark 6. Observe that the workload of a customer is only dependent on the state in which it *enters* the system, not necessarily on the current state. In case the state of the background process changes before the customer ends its service, service continues with the service time drawn from the previous state.

In this section, we present the idea of [5] applied to our concept of mixing the arrival parameter. First we develop a differential equation that shows similarities with (2.1) from the $M_\Lambda/\text{Cox}_n/\infty$ queue. Like in the previous section, we indicate how this differential equation can be used to obtain queue length moments.

3.1 The differential equation

We follow a similar approach as in [5], Section 3. Let $S(t) \in \mathcal{N}$ be the state of the Markovian background process at time t. Also, let $\bar{X}_{i,\lambda}(t)$ be the number of customers at time t given $X(0) = 0$, $S(0) = i$ and $\Lambda(0) = \lambda$. Assuming an empty system at time zero is, for the same reason as in Section 2.2, not very restrictive. Say for example that there are m customers at time zero. Each of those can be characterized by the state it arrived in and its residual service time. With this information we can calculate the probabilities $q_1, ..., q_m$ that the corresponding customer is still in the system at time t. Let $Z_j(t) \sim \text{Bernoulli}(q_j)$ for $j = 1, ..., m$. Then the total number of customers at time t equals $\bar{X}_{i,\lambda}(t) + \sum\limits_{j=1}^{m} Z_j(t)$.

Consider the small time interval $(0, h)$. Define $f_{i,\lambda}(z,t) := E(z^{\bar{X}_{i,\lambda}(t)})$. To remove the condition $\Lambda(0) = \lambda$, we define $\bar{X}_i(t)$ as the number of customers at time t given $X(0) = 0$ and $S(0) = i$. Moreover, let $f_i(z,t) = E(z^{\bar{X}_i(t)}) = \int\limits_0^\infty g_{\Lambda_i}(\lambda) E(z^{\bar{X}_{i,\lambda}(t)}) d\lambda$ be the corresponding generating function. It is easily seen that

$$\begin{aligned} f_{i,\lambda}(z,t) = &\Big(1 + \lambda h(z-1)P(B_i \geq t)\Big) \\ &\times \left(h\gamma_i \sum_{j=1}^{n} p_{ij} f_j(z, t-h) + (1 - h\gamma_i) f_{i,\lambda}(z, t-h)\right) + o(h), \end{aligned}$$

which yields the differential equation

$$\gamma_i \sum_{j=1}^{n} p_{ij} f_j(z,t) - \gamma_i f_{i,\lambda}(z,t) - \frac{\partial}{\partial t} f_{i,\lambda}(z,t) + \lambda(z-1)P(B_i \geq t) f_{i,\lambda}(z,t) = 0. \tag{3.1}$$

Removing the condition $\Lambda(0) = \lambda$ then gives

$$\gamma_i \sum_{j=1}^{n} p_{ij} f_j(z,t) - \gamma_i f_i(z,t) - \frac{\partial}{\partial t} f_i(z,t) + (z-1)P(B_i \geq t) E\Big(\Lambda_i f_{i,\Lambda_i}(z,t)\Big) = 0. \tag{3.2}$$

This last formula has some interesting consequences. For example, by letting $t \to \infty$, we find the steady state formula

$$E(z^{\bar{X}_i(\infty)}) = \sum_{j=1}^{n} p_{ij} E(z^{\bar{X}_j(\infty)}).$$

Note that after an infinite amount of time, the state of the background process at $t = 0$ should be irrelevant. Since $\sum\limits_{j=1}^{n} p_{ij} = 1$, the above equality indeed holds.

Another interesting situation occurs when $n = 1$, so that the system is always in state 1. Note that this is a generalization of our previous $M_\Lambda/\mathrm{Cox}_n/\infty$ model, the service times now having a fully general distribution. Equation (3.2) in this case becomes

$$\frac{\partial}{\partial t} f_1(z,t) = (z-1)P(B_1 \geq t) E\Big(\Lambda_1 f_{1,\Lambda_1}(z,t)\Big).$$

Also taking the derivative with respect to z and taking $z = 1$ gives

$$\frac{d}{dt} E(\bar{X}_1(t)) = P(B_1 \geq t) E(\Lambda_1),$$

which, with the condition that $E(\bar{X}_1(0)) = 0$, yields

$$E(\bar{X}_1(t)) = E(\Lambda_1) \int_0^t P(B_1 \geq u) du.$$

Taking $B_1 \sim \exp(\mu_1)$ here, for example, gives

$$E(\bar{X}_1(t)) = \frac{E(\Lambda_1)}{\mu_1}\left(1 - e^{-\mu_1 t}\right). \tag{3.3}$$

This formula of course agrees with Formula (2.9) for the mean number of customers in phase 1 of the $M_\Lambda/\mathrm{Cox}_n/\infty$ queue. When only considering customers in the first phase of a Coxian distribution, successive phases are irrelevant since there are infinitely many servers. Therefore, both models describe the $M_\Lambda/M/\infty$, and $\bar{X}_1(t) \overset{d}{=} X_1(t)$ when the background process has only one state and $B_1 \sim \exp(\mu_1)$.

3.2 Calculating moments

In the $M_\Lambda/\mathrm{Cox}_n/\infty$ from Section 2, we saw that despite being unable to find the generating function, we could use the differential equation to compute all moments. We will do the same here with the Markov-modulated queue, both with and without the initial condition on $\Lambda(0)$.

To obtain factorial moments, we take from (3.1) the k^{th} derivative with respect to z in $z = 1$:

$$\begin{aligned} \frac{d}{dt}E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k\right)!}\right) &= -\gamma_i E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k\right)!}\right) + \gamma_i \sum_{j=1}^{n} p_{ij} E\left(\frac{\bar{X}_j(t)!}{\left(\bar{X}_j(t)-k\right)!}\right) \\ + \quad & k\lambda P\left(B_i \geq t\right) E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k+1\right)!}\right). \end{aligned} \tag{3.4}$$

Doing the same with (3.2) gives

$$\begin{aligned} \frac{d}{dt}E\left(\frac{\bar{X}_{i}(t)!}{\left(\bar{X}_{i}(t)-k\right)!}\right) &= -\gamma_i E\left(\frac{\bar{X}_{i}(t)!}{\left(\bar{X}_{i}(t)-k\right)!}\right) + \gamma_i \sum_{j=1}^{n} p_{ij} E\left(\frac{\bar{X}_j(t)!}{\left(\bar{X}_j(t)-k\right)!}\right) \\ + \quad & kP\left(B_i \geq t\right) \int_0^\infty g_{\Lambda_i}(\lambda)\lambda E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k+1\right)!}\right) d\lambda. \end{aligned} \tag{3.5}$$

Theorem 3.1. *All moments* $E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k\right)!}\right)$ *and* $E\left(\frac{\bar{X}_{i}(t)!}{\left(\bar{X}_{i}(t)-k\right)!}\right)$ *can be iteratively computed for all* $k \in \mathbb{N}$.

Proof. Note that Equation (3.5) can be seen as a system of linear differential equations

$$\frac{d}{dt}\overrightarrow{E(k,t)} = A\overrightarrow{E(k,t)} + \overrightarrow{c(t)}, \tag{3.6}$$

where

- $\overrightarrow{E(k,t)}$ is the n-dimensional vector consisting of $E\left(\frac{\bar{X}_1(t)!}{\left(\bar{X}_1(t)-k\right)!}\right)$ up to $E\left(\frac{\bar{X}_n(t)!}{\left(\bar{X}_n(t)-k\right)!}\right)$,
- A is a matrix with entries $a_{ij} = \gamma_i p_{ij}$ for $i \neq j$ and $a_{ii} = \gamma_i(p_{ii} - 1)$, and
- $c(t)_i = kP\left(B_i \geq t\right) \int\limits_0^\infty g_{\Lambda_i}(\lambda)\lambda E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k+1\right)!}\right) d\lambda.$

In the same way, (3.4) can be viewed as n separate differential equations

$$\frac{d}{dt}E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k\right)!}\right) = -\gamma_i E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k\right)!}\right) + c(\lambda,t)_i, \qquad i = 1, ..., n, \tag{3.7}$$

with

$$c(\lambda,t)_i = \gamma_i \sum_{j=1}^{n} p_{ij} E\left(\frac{\bar{X}_j(t)!}{\left(\bar{X}_j(t)-k\right)!}\right) + k\lambda P\left(B_i \geq t\right) E\left(\frac{\bar{X}_{i,\lambda}(t)!}{\left(\bar{X}_{i,\lambda}(t)-k+1\right)!}\right).$$

The main idea is that when we solve the equations in the right order, the $c(t)_i$ and $c(\lambda,t)_i$ are known from previous iterations. If these quantities are known, we can find any moment by solving Equations (3.6) and (3.7) via ordinary differential equation methods.

We now proceed to show a sufficient calculation order to complete the proof, by induction. For $k = 1$, we have $c(t)_i = P(B_i \geq t)E(\Lambda_i)$, enabling us to calculate $E\left(\bar{X}_i(t)\right)$ for $i = 1, ..., n$ with (3.6). Now, $c(\lambda,t)_i = \gamma_i \sum\limits_{j=1}^{n} p_{ij} E\left(\bar{X}_j(t)\right) + P(B_i \geq t)\lambda$ is known such that with (3.7) we can also find $E\left(\bar{X}_{i,\lambda}(t)\right)$.

Suppose we know all moments of order $k-1$. Then $c(t)_i$ can be found by assumption, so we can solve the system (3.6) to find $E\left(\frac{\bar{X}_i(t)!}{\left(\bar{X}_i(t)-k\right)!}\right)$ for

$i = 1, ..., n$. With these quantities found and the induction hypothesis, note that now also $c(\lambda, t)_i$ are known. By solving (3.7), the proof is complete.

□

The next step will be calculating some specific moments with Equations (3.4) and (3.5). Specifying the service time distributions allows for expressions without integrals, so let us assume from now that $B_i \sim \exp(\mu_i)$ for $i = 1, ..., n$. The exponential distribution is a natural choice and gives simple expressions.

Note that we already found $E(\bar{X}_1(t))$ for $n = 1$ in (3.3). We proceed with finding the mean, conditioned on $\Lambda(0) = \lambda$, and the variance for $n = 1$. After that we move on to $n = 2$.

With $k = 1$ and (3.3), differential equation (3.4) reads

$$\begin{aligned}\frac{d}{dt}E\left(\bar{X}_{1,\lambda}(t)\right) &= -\gamma_1 E\left(\bar{X}_{1,\lambda}(t)\right) + \gamma_1 E\left(\bar{X}_1(t)\right) + \lambda P(B_1 \geq t) \\ &= -\gamma_1 E\left(\bar{X}_{1,\lambda}(t)\right) + \frac{\gamma_1 E(\Lambda_1)}{\mu_1} + \left(\lambda - \frac{\gamma_1 E(\Lambda_1)}{\mu_1}\right) e^{-\mu_1 t}.\end{aligned}$$

One can easily check that, with boundary condition $E\left(\bar{X}_{1,\lambda}(0)\right) = 0$,

$$E\left(\bar{X}_{1,\lambda}(t)\right) = \frac{E(\Lambda_1)}{\mu_1} + \frac{\mu_1\lambda - \gamma_1 E(\Lambda_1)}{\mu_1(\gamma_1 - \mu_1)} e^{-\mu_1 t} - \frac{\lambda - E(\Lambda_1)}{\gamma_1 - \mu_1} e^{-\gamma_1 t}.$$

Some interpretation of this formula can be obtained when we write it as

$$E\left(\bar{X}_{1,\lambda}(t)\right) = \frac{E(\Lambda_1)}{\mu_1}\left(1 - e^{-\mu_1 t}\right) + \frac{e^{-\mu_1 t} - e^{-\gamma_1 t}}{\gamma_1 - \mu_1}(\lambda - E(\Lambda_1)). \tag{3.8}$$

We recognize the first term as the formula for $E\left(\bar{X}_1(t)\right)$. The second term is a nonnegative number times the difference between the conditioned starting arrival parameter λ and its expectation $E(\Lambda_1)$. In other words, the difference in the mean queue size conditioned on $\Lambda(0) = \lambda$ or $\Lambda(0) = \Lambda_1$, is linear in $\lambda - E(\Lambda_1)$. It is evident that when $t > 0$, $E\left(\bar{X}_{1,\lambda}(t)\right) \geq E\left(\bar{X}_1(t)\right)$ if and only if $\lambda \geq E(\Lambda_1)$, with equality if and only $\lambda = E(\Lambda_1)$.

It also follows from (3.8) that the difference between $E\left(\bar{X}_{1,\lambda}(t)\right)$ and $E\left(\bar{X}_1(t)\right)$ is largest at $t = \frac{\ln(\gamma_1) - \ln(\mu_1)}{\gamma_1 - \mu_1}$. The existence of a maximum is intuitive: at $t = 0$, the system still has to set up, and when $t \to \infty$, the value of $\Lambda(0)$ has become irrelevant. In both of these cases it holds that $E\left(\bar{X}_{1,\lambda}(t)\right) = E\left(\bar{X}_1(t)\right)$.

Formula (3.8) also holds for the $M_\Lambda/\mathrm{Cox}_n/\infty$ setting, in the sense that $E(X_1(t)|\Lambda(0) = \lambda) = E\left(\bar{X}_{1,\lambda}(t)\right)$ (see the end of Subsection 3.1). The approach of Section 2 does not enable us to find moments conditioned on $\Lambda(0)$, so this is an interesting observation.

Moving on to moments of second order, we take $k = 2$ in (3.5) and find

$$\frac{d}{dt}E\left(\bar{X}_1(t)\left(\bar{X}_1(t)-1\right)\right) = 2P\left(B_1 \geq t\right)\int\limits_0^\infty g_{\Lambda_i}(\lambda)\lambda E\left(\bar{X}_{1,\lambda}(t)\right)d\lambda$$

$$= 2e^{-\mu_1 t}\int\limits_0^\infty g_{\Lambda_1}(\lambda)\lambda\left(\frac{E(\Lambda_1)}{\mu_1} + \frac{\mu_1\lambda - \gamma_1 E(\Lambda_1)}{\mu_1(\gamma_1-\mu_1)}e^{-\mu_1 t} - \frac{\lambda - E(\Lambda_1)}{\gamma_1-\mu_1}e^{-\gamma_1 t}\right)d\lambda$$

$$= \frac{2E(\Lambda_1)^2}{\mu_1}e^{-\mu_1 t} + \frac{2\mu_1 E(\Lambda_1^2) - 2\gamma_1 E(\Lambda_1)^2}{\mu_1(\gamma_1-\mu_1)}e^{-2\mu_1 t} - \frac{2\mathrm{Var}(\Lambda_1)}{\gamma_1-\mu_1}e^{-(\gamma_1+\mu_1)t}.$$

The only solution with $E\left(\bar{X}_1(0)\left(\bar{X}_1(0)-1\right)\right) = 0$ is

$$\begin{aligned} E\big(\bar{X}_1(t)\left(\bar{X}_1(t)-1\right)\big) = {} & \frac{\mu_1 E(\Lambda_1^2) + \gamma_1 E(\Lambda_1)^2}{\mu_1^2(\gamma_1+\mu_1)} - \frac{2E(\Lambda_1)^2}{\mu_1^2}e^{-\mu_1 t} \\ & + \frac{\gamma_1 E(\Lambda_1)^2 - \mu_1 E(\Lambda_1^2)}{\mu_1^2(\gamma_1-\mu_1)}e^{-2\mu_1 t} + \frac{2\mathrm{Var}(\Lambda_1)}{(\gamma_1-\mu_1)(\gamma_1+\mu_1)}e^{-(\gamma_1+\mu_1)t}, \end{aligned} \tag{3.9}$$

and hence

$$\begin{aligned} \mathrm{Var}(\bar{X}_1(t)) = {} & \frac{E(\Lambda_1)}{\mu_1} + \frac{\mathrm{Var}(\Lambda_1)}{\mu_1(\gamma_1+\mu_1)} - \frac{E(\Lambda_1)}{\mu_1}e^{-\mu_1 t} \\ & - \frac{\mathrm{Var}(\Lambda_1)}{\mu_1(\gamma_1-\mu_1)}e^{-2\mu_1 t} + \frac{2\mathrm{Var}(\Lambda_1)}{(\gamma_1-\mu_1)(\gamma_1+\mu_1)}e^{-(\gamma_1+\mu_1)t}. \end{aligned} \tag{3.10}$$

This is in agreement with (2.13); see also the end of Subsection 3.1.

We proceed by studying the variance when the starting arrival rate is fixed, i.e. $\Lambda(0) = \lambda$. Using Formula (3.5) for $k = 2$ yields

$$\begin{aligned} \frac{d}{dt}E\big(\bar{X}_{1,\lambda}(t)\big(\bar{X}_{1,\lambda}(t)-1\big)\big) = {} & -\gamma_1 E\left(\bar{X}_{1,\lambda}(t)\left(\bar{X}_{1,\lambda}(t)-1\right)\right) \\ & + \gamma_1 E\big(\bar{X}_1(t)\left(\bar{X}_1(t)-1\right)\big) + 2\lambda e^{-\mu_1 t}E\left(\bar{X}_{1,\lambda}(t)\right). \end{aligned}$$

Observe that the latter two terms are known from (3.9) and (3.8), hence we have a solvable linear differential equation. As usual, we set the boundary condition at $E\big(\bar{X}_{1,\lambda}(0)\big(\bar{X}_{1,\lambda}(0)-1\big)\big)=0$, leading to $E\big(\bar{X}_{1,\lambda}(t)\big(\bar{X}_{1,\lambda}(t)-1\big)\big)$ and subsequently to

$$\begin{aligned}
\mathrm{Var}\big(\bar{X}_{1,\lambda}(t)\big) &= E\big(\bar{X}_{1,\lambda}(t)\big)+\frac{\mathrm{Var}(\Lambda_1)}{\mu_1(\gamma_1+\mu_1)} \\
&+ \left(\frac{2\mathrm{Var}(\Lambda_1)}{\mu_1(\gamma_1-2\mu_1)}-\frac{2\left(\lambda-E(\Lambda_1)\right)^2}{\mu_1(\gamma_1-2\mu_1)}\right)e^{-\gamma_1 t} \\
&+ \left(-\frac{\gamma_1\mathrm{Var}(\Lambda_1)}{\mu_1(\gamma_1-\mu_1)(\gamma_1-2\mu_1)}+\frac{\gamma_1\left(\lambda-E(\Lambda_1)\right)^2}{(\gamma_1-\mu_1)^2(\gamma_1-2\mu_1)}\right)e^{-2\mu_1 t} \\
&+ \left(-\frac{2\gamma_1\mathrm{Var}(\Lambda_1)}{\mu_1(\gamma_1-\mu_1)(\gamma_1+\mu_1)}+\frac{2\gamma_1\left(\lambda-E(\Lambda_1)\right)^2}{\mu_1(\gamma_1-\mu_1)^2}\right)e^{-(\gamma_1+\mu_1)t} \\
&- \frac{\left(\lambda-E(\Lambda_1)\right)^2}{(\gamma_1-\mu_1)^2}e^{-2\gamma_1 t}. \qquad (3.11)
\end{aligned}$$

Notice here that the initial arrival intensity only appears as $\lambda - E(\Lambda_1)$, and that $\gamma=\mu_1$ and $\gamma=2\mu_1$ are removable singularities.

So far, we have only obtained explicit expressions for moments in the Markov-modulated queue with the background process having only one state. In order to better analyze the effect of the background process on the queue, we now consider an example with $n=2$. This should give an impression of the behavior of the Markov-modulated queue with multiple states.

For $n=2$, $k=1$, (3.6) takes the form

$$\frac{d}{dt}\begin{pmatrix}E\left(\bar{X}_1(t)\right)\\ E\left(\bar{X}_2(t)\right)\end{pmatrix}=\begin{pmatrix}-\gamma_1 p_{12} & \gamma_1 p_{12}\\ \gamma_2 p_{21} & -\gamma_2 p_{21}\end{pmatrix}\begin{pmatrix}E\left(\bar{X}_1(t)\right)\\ E\left(\bar{X}_2(t)\right)\end{pmatrix}+\begin{pmatrix}E(\Lambda_1)e^{-\mu_1 t}\\ E(\Lambda_2)e^{-\mu_2 t}\end{pmatrix},$$

since $p_{11}-1=-p_{12}$ and $p_{22}-1=-p_{21}$. We can solve this system with the eigenvalue method, see for example [21]. Define $\pi := \gamma_1 p_{12}+\gamma_2 p_{21}$ as the sum of the state transition rates. It is easily seen that the eigenvectors of the system are $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}\gamma_1 p_{12}\\ -\gamma_2 p_{21}\end{pmatrix}$ with eigenvalues 0 and $-\pi$ respectively. Therefore, the solution is

$$\begin{pmatrix}E\left(\bar{X}_1(t)\right)\\ E\left(\bar{X}_2(t)\right)\end{pmatrix}=\begin{pmatrix}1 & \gamma_1 p_{12}\\ 1 & -\gamma_2 p_{21}\end{pmatrix}\begin{pmatrix}y_1(t)\\ y_2(t)\end{pmatrix}$$

with

$$\frac{d}{dt}\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -\pi \end{pmatrix}\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} + \begin{pmatrix} 1 & \gamma_1 p_{12} \\ 1 & -\gamma_2 p_{21} \end{pmatrix}\begin{pmatrix} E(\Lambda_1)e^{-\mu_1 t} \\ E(\Lambda_2)e^{-\mu_2 t} \end{pmatrix}.$$

The latter is a pair of separate linear differential equations. Solving these, substituting in the above solution and again using the boundary condition $E(\bar{X}_1(0)) = E(\bar{X}_2(0)) = 0$, yields

$$\begin{aligned} E(\bar{X}_1(t)) &= \frac{\gamma_2 p_{21} E(\Lambda_1)}{\pi\mu_1}\left(1 - e^{-\mu_1 t}\right) + \frac{\gamma_1 p_{12} E(\Lambda_2)}{\pi\mu_2}\left(1 - e^{-\mu_2 t}\right) \\ &\quad + \frac{\gamma_1 p_{12} E(\Lambda_1)}{\pi}\cdot\frac{e^{-\mu_1 t} - e^{-\pi t}}{\pi - \mu_1} - \frac{\gamma_1 p_{12} E(\Lambda_2)}{\pi}\cdot\frac{e^{-\mu_2 t} - e^{-\pi t}}{\pi - \mu_2}, \\ E(\bar{X}_2(t)) &= \frac{\gamma_2 p_{21} E(\Lambda_1)}{\pi\mu_1}\left(1 - e^{-\mu_1 t}\right) + \frac{\gamma_1 p_{12} E(\Lambda_2)}{\pi\mu_2}\left(1 - e^{-\mu_2 t}\right) \\ &\quad - \frac{\gamma_2 p_{21} E(\Lambda_1)}{\pi}\cdot\frac{e^{-\mu_1 t} - e^{-\pi t}}{\pi - \mu_1} + \frac{\gamma_2 p_{21} E(\Lambda_2)}{\pi}\cdot\frac{e^{-\mu_2 t} - e^{-\pi t}}{\pi - \mu_2}. \end{aligned} \tag{3.12}$$

Quantities in (3.12) that should be recognized are

- $\frac{\gamma_2 p_{21}}{\pi}$ and $\frac{\gamma_1 p_{12}}{\pi}$, the long term fractions of finding the Markov process in state 1 and 2, respectively.
- $\frac{E(\Lambda_i)}{\mu_i}\left(1 - e^{-\mu_i t}\right)$, the mean number of customers when the Markov process is always in state i $(i = 1, 2)$.
- $\frac{e^{-\mu_i t} - e^{-\pi t}}{\pi - \mu_i}$, a nonnegative quantity that increases whenever π or μ_i increases $(i = 1, 2)$.

With these equations we can also find some other relevant quantities. For example, suppose that the state at time 0 is not known, but the background process is in steady state. Then $P(S(0) = 1) = \frac{\gamma_2 p_{21}}{\pi}$ and $P(S(0) = 2) = \frac{\gamma_1 p_{12}}{\pi}$, so it follows that the mean number of customers at time t is

$$\begin{aligned} &\frac{\gamma_2 p_{21}}{\pi} E(\bar{X}_1(t)) + \frac{\gamma_1 p_{12}}{\pi} E(\bar{X}_2(t)) \\ &\qquad = \frac{\gamma_2 p_{21} E(\Lambda_1)}{\pi\mu_1}\left(1 - e^{-\mu_1 t}\right) + \frac{\gamma_1 p_{12} E(\Lambda_2)}{\pi\mu_2}\left(1 - e^{-\mu_2 t}\right). \end{aligned} \tag{3.13}$$

Another consequence of (3.12) is the steady state mean

$$E(\bar{X}) = \frac{\gamma_2 p_{21} E(\Lambda_1)}{\pi\mu_1} + \frac{\gamma_1 p_{12} E(\Lambda_2)}{\pi\mu_2}. \tag{3.14}$$

Keep in mind that there is no subscript needed, since in steady state it does not matter in which state we started. In the formula we see the steady state mean of each individual state, multiplied by the fraction that the corresponding state was active. The fact that the steady state mean is just the sum of these parts, underlines that customers from different states do not interfere with each other.

Let us now consider $E(\bar{X}|S = 1)$: the steady state expected number of customers given the background process is in state 1. Before we can give an expression, we have to define $\tilde{X}_i$ as the portion of the current customers that arrived when the state was i. Distinction between $\tilde{X}_1$ and $\tilde{X}_2$ is relevant because customers that arrived in a different state will have a different service rate. For instance, the first term of (3.14) is the mean number of customers with service rate μ_1.

Now define T_1 as an arbitrary time in steady state when the background process moves from state 1 to state 2. Also, let $T_{2\to1}$ be the last time before T_1 that the state changed from 2 to 1. We split $\tilde{X}_1(T_1)$ into the customers that arrived before or after $T_{2\to1}$, and then condition on the value of $T_1 - T_{2\to1}$. It holds that

$$\begin{aligned}
&E\left(\tilde{X}_1(T_1)\right) = E\left(\tilde{X}_1(T_{2\to1})\right) \cdot P(B_1 \geq T_1 - T_{2\to1}) + E\left(\bar{X}_1(T_1 - T_{2\to1})\right) \\
= &\int_0^\infty \gamma_1 p_{12} e^{-\gamma_1 p_{12} t} \left(E\left(\tilde{X}_1(T_{2\to1})\right) e^{-\mu_1 t} + E\left(\bar{X}_1(t) | S(u) = 1 \text{ for all } u \in [0,t]\right)\right) dt \\
= &\int_0^\infty \gamma_1 p_{12} e^{-\gamma_1 p_{12} t} \left(E\left(\tilde{X}_1(T_{2\to1})\right) e^{-\mu_1 t} + \frac{E(\Lambda_1)}{\mu_1}\left(1 - e^{-\mu_1 t}\right)\right) dt \\
= &\frac{\gamma_1 p_{12}}{\gamma_1 p_{12} + \mu_1} E\left(\tilde{X}_1(T_{2\to1})\right) + \frac{E(\Lambda_1)}{\mu_1}\left(1 - \frac{\gamma_1 p_{12}}{\gamma_1 p_{12} + \mu_1}\right).
\end{aligned}$$

Let T_2 and $T_{1\to 2}$ be the symmetric versions of T_1 and $T_{2\to 1}$. We then have

$$\begin{aligned} E\left(\tilde{X}_1(T_2)\right) &= E\left(\tilde{X}_1(T_{1\to 2})\right) \cdot P(B_1 \geq T_2 - T_{1\to 2}) \\ &= \int\limits_0^\infty \gamma_2 p_{21} e^{-\gamma_2 p_{21} t} E\left(\tilde{X}_1(T_{1\to 2})\right) e^{-\mu_1 t} dt = \frac{\gamma_2 p_{21}}{\gamma_2 p_{21} + \mu_1} E\left(\tilde{X}_1(T_{1\to 2})\right). \end{aligned}$$

A useful observation here is that $T_1 \overset{d}{=} T_{1\to 2}$ and $T_2 \overset{d}{=} T_{2\to 1}$, since in each case, both times represent a steady state time at the end of a period of one state. As a result we have a system of two linear equations. The solution is

$$E\left(\tilde{X}_1(T_1)\right) = \frac{\left(\gamma_2 p_{21} + \mu_1\right) E(\Lambda_1)}{(\pi + \mu_1)\mu_1}, \quad E\left(\tilde{X}_1(T_2)\right) = \frac{\gamma_2 p_{21} E(\Lambda_1)}{(\pi + \mu_1)\mu_1}. \qquad (3.15)$$

Notice the absence of $E(\Lambda_2)$ and μ_2, caused by the fact that customers from different states do not influence each other.

Let $T_{S=1}$ be an arbitrary steady state time with $S(T_{S=1}) = 1$. We will show that $E\left(\tilde{X}_1(T_1)\right) = E\left(\tilde{X}_1(T_{S=1})\right) = E(\tilde{X}_1|S = 1)$. Note that we can view the Markov background process as a Poisson process with rate $\gamma_1 p_{12}$, and we can view its events as periods where the state is 2. We assume those periods take 0 time, so that the rate at which an event happens is always $\gamma_1 p_{12}$. A nice property of the Poisson process is that when we pick an arbitrary point in time ($T_{S=1}$), the amount of time since the last event is exponentially distributed with parameter $\gamma_1 p_{12}$. The same holds for T_1, so $\tilde{X}_1(T_{S=1}) \overset{d}{=} \tilde{X}_1(T_1)$, hence in particular $E(\tilde{X}_1|S = 1) = E\left(\tilde{X}_1(T_1)\right) = \frac{(\gamma_2 p_{21} + \mu_1) E(\Lambda_1)}{(\pi + \mu_1)\mu_1}$.

As a quick sanity check, we calculate $E(\tilde{X}_1)$, the steady state mean number of customers that arrived when the background process was in state 1 (i.e. the customers that have service rate μ_1). As mentioned, this should give the first term of (3.14). The calculation below verifies this:

$$\begin{aligned} E(\tilde{X}_1) &= E(\tilde{X}_1|S = 1) \cdot P(S = 1) + E(\tilde{X}_1|S = 2) \cdot P(S = 2) \\ &= \frac{\left(\gamma_2 p_{21} + \mu_1\right) E(\Lambda_1)}{(\pi + \mu_1)\mu_1} \cdot \frac{\gamma_2 p_{21}}{\pi} + \frac{\gamma_2 p_{21} E(\Lambda_1)}{(\pi + \mu_1)\mu_1} \cdot \frac{\gamma_1 p_{12}}{\pi} = \frac{\gamma_2 p_{21} E(\Lambda_1)}{\pi \mu_1}. \end{aligned}$$

For symmetry reasons, (3.15) can be transformed into

$$E\left(\tilde{X}_2(T_1)\right) = \frac{\gamma_1 p_{12} E(\Lambda_2)}{(\pi + \mu_2)\mu_2}, \quad E\left(\tilde{X}_2(T_2)\right) = \frac{\left(\gamma_1 p_{12} + \mu_2\right) E(\Lambda_2)}{(\pi + \mu_2)\mu_2}. \qquad (3.16)$$

Now we are finally ready to find the mean total number of customers conditioned on the current state:

$$\begin{aligned} E\left(\bar{X}|S=1\right) &= E\left(\tilde{X}_1(T_1)\right) + E\left(\tilde{X}_2(T_1)\right) \\ &= \frac{\gamma_2 p_{21} E(\Lambda_1)}{(\pi+\mu_1)\mu_1} + \frac{\gamma_1 p_{12} E(\Lambda_2)}{(\pi+\mu_2)\mu_2} + \frac{E(\Lambda_1)}{\pi+\mu_1}, \end{aligned} \tag{3.17}$$

and analogously,

$$E\left(\bar{X}|S=2\right) = \frac{\gamma_2 p_{21} E(\Lambda_1)}{(\pi+\mu_1)\mu_1} + \frac{\gamma_1 p_{12} E(\Lambda_2)}{(\pi+\mu_2)\mu_2} + \frac{E(\Lambda_2)}{\pi+\mu_2}. \tag{3.18}$$

Given the complexity of calculations with two states, formulas for general n will be costly to derive analytically. However, if one assumes that the Markov process is in steady state at time 0, one can use a shortcut found by Blom *et al.* [5] for the mean number of customers. It is based on the well-known rate-in = rate-out balance equation $\sum\limits_{i=1}^{n} \pi_i \gamma_i p_{ij} = \pi_j \gamma_j$, π_i being the steady state probabilities for each state.

To find the mean number of customers in this case, we take $k = 1$ in (3.5), multiply by π_i and sum over i. This results in

$$\begin{aligned} \frac{d}{dt} \sum_{i=1}^{n} \pi_i E\left(\bar{X}_i(t)\right) &= -\sum_{i=1}^{n} \pi_i \gamma_i E\left(\bar{X}_i(t)\right) + \sum_{i=1}^{n} \pi_i \gamma_i \sum_{j=1}^{n} p_{ij} E\left(\bar{X}_j(t)\right) \\ &\quad + \sum_{i=1}^{n} \pi_i P\left(B_i \geq t\right) E(\Lambda_i) = \sum_{i=1}^{n} \pi_i P\left(B_i \geq t\right) E(\Lambda_i), \end{aligned}$$

and hence

$$\sum_{i=1}^{n} \pi_i E\left(\bar{X}_i(t)\right) = \sum_{i=1}^{n} \pi_i E(\Lambda_i) \int_0^t P\left(B_i \geq u\right) du. \tag{3.19}$$

With relatively easy methods, we found a general formula for the mean. For instance, $n = 2$ and exponential service times yields (3.13), which was significantly harder to find as we had to solve a system of differential equations. The disadvantage of the quicker method is that it gives no information about the case where the starting state $S(0)$ is fixed.

4 Discussion and future work

We have considered two infinite-server queueing models with a mixed arrival process. For the $M_\Lambda/Cox_n/\infty$ model with exponential resample times we showed how to compute all joint moments of the arrival rate and the number of customers, both in transient and steady state. For a Markov-modulated queue with general service times we gave a procerdure to obtain all moments of the number of customers given the initial state and the initial arrival rate.

Since moments define a distribution, finding a way to compute them is a significant step towards finding the exact queue size distribution, even though we could not find a closed formula for all moments. Possible future research may include finding the exact distributions of the considered random variables. From our point of view, a way to do this is solving the partial differential equation corresponding to that model.

In related papers about infinite-server queues, a considerable amount of work has been done regarding scaling limits and large deviation principles [3, 5, 10, 11]. It is likely that similar results can be proven for our models, cf. Conjecture 2.5.

Finally, it would be interesting to study the $M/M/\infty$ queue in which not the arrival rate but the *service rate* is described by a stochastic process. That infinite-server queue seems less amenable to a recursive approach.

5 References

[1] I. Adan, J. Resing (2015). *Queueing Systems*, http://www.win.tue.nl/∼iadan/queueing.pdf

[2] H. Albrecher, S. Asmussen (2001). *Ruin probabilities and aggregate claims distributions for shot noise Cox processes.* Scandinavian Actuarial Journal, 86-110.

[3] D. Anderson, J. Blom, M. Mandjes, H. Thorsdottir, K. de Turck (2016). *A Functional Central Limit Theorem for a Markov-modulated infinite-server queue.* Methodology and Computing in Applied Probability **18**, 153–168.

[4] S. Asmussen (2003). *Applied Probability and Queues*, 2nd edition, Springer-Verlag, New York.

[5] J. Blom, O. Kella, M. Mandjes, H. Thorsdottir (2014). *Markov-modulated infinite-server queues with general service times.* Queueing Systems **76**, 403–424.

[6] H. Bühlmann (1972). *Ruinwahrscheinlichkeit bei erfahrungstarifiertem portefeuille.* Bulletin de l'Association des Actuaires Suisses **2**, 131-140.

[7] B. D'Auria (2008). *$M/M/\infty$ queues in semi-Markovian random environment.* Queueing Systems **58**, 221–237.

[8] A. Daw, J. Pender (2018). *Queues driven by Hawkes processes.* Stochastic Systems **8**, 192-229.

[9] S. Eick, W. Massey, W. Whitt (1993). *The physics of the $M_t/G/\infty$ queue.* Management Science **39**, 241-252.

[10] M. Heemskerk, J. van Leeuwaarden, M. Mandjes (2017). *Scaling limits for infinite-server systems in a random environment.* Stochastic Systems **7**, 1-31.

[11] H. M. Jansen, M. R. H. Mandjes, K. De Turck, S. Wittevrongel (2016). *A large deviations principle for infinite-server queues in a random environment.* Queueing Systems **82**, 199–235.

[12] G. Jongbloed, G. Koole (2001). *Managing uncertainty in call centres using Poisson mixtures.* Applied Stochastic Models in Business and Industry **17**, 307-318.

[13] B. Klaasse (2017). *Queuing and insurance risk models with mixing dependencies.* Bachelor Report, Department of Mathematics and Computer Science, Eindhoven University of Technology.

[14] D. Koops, O. Boxma, M. Mandjes (2017). *Networks of $\cdot/G/\infty$ queues with shot-noise driven arrival intensities.* Queueing Systems **86**, 301-325.

[15] D. Koops, M. Saxena, O. Boxma, M. Mandjes (2018). *Infinite-server queues with Hawkes input.* Journal of Applied Probability **55**, 920-943.

[16] L.R. van Kreveld (2018). *Parameter Mixing in Infinite Server Queues.* MSc Thesis, Department of Mathematical Sciences, University of Utrecht.

[17] C. A. O'Cinneide, P. Purdue (1986). *The $M/M/\infty$ queue in a random environment*. Journal of Applied Probability **23**, 175-184.

[18] Y. Raaijmakers, H. Albrecher, O. Boxma (2019). *The single server queue with mixing dependencies*. To appear in Methodology and Computing in Applied Probability.

[19] S. M. Ross (2010). *Introduction to Probability Models*, 10^{th} edition, Elsevier, New York.

[20] A. Salih (2016). *Method of Characteristics*. Indian Institute of Space Science and Technology, Thiruvananthapuram.

[21] W. Shen (2013). *Introduction to Ordinary and Partial Differential Equations*. https://www.math.psu.edu/shen_w/LNDE.pdf