RARE EVENT ANALYSIS OF MARKOV-MODULATED INFINITE-SERVER QUEUES: A POISSON LIMIT

J.G. BLOM ⋆, K.E.E.S. DE TURCK †, M.R.H. MANDJES •,⋆

ABSTRACT. This paper studies an infinite-server queue in a Markov environment, that is, an infinite-server queue with arrival rates and service times depending on the state of a Markovian background process. Scaling the arrival rates λ_i by a factor N and the rates ν_{ij} of the background process by $N^{1+\varepsilon}$ (for some $\varepsilon > 0$), the focus is on the tail probabilities of the number of customers in the system, in the asymptotic regime that N tends to ∞. In particular, it is shown that the logarithmic asymptotics correspond to those of a Poisson distribution with an appropriate mean.

KEYWORDS. Queues ⋆ infinite-server systems ⋆ Markov modulation ⋆ large deviations

Work done while K. de Turck was visiting Korteweg-de Vries Institute for Mathematics, University of Amsterdam, the Netherlands, with greatly appreciated financial support of *Fonds Wetenschappelijk Onderzoek / Research Foundation – Flanders*. He is also a Postdoctoral Fellow of the same foundation.

• Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands.

⋆ CWI, P.O. Box 94079, 1090 GB Amsterdam, the Netherlands.

† TELIN, Ghent University, St.-Pietersnieuwstraat 41, B9000 Gent, Belgium.

M. Mandjes is also with EURANDOM, Eindhoven University of Technology, Eindhoven, the Netherlands, and IBIS, Faculty of Economics and Business, University of Amsterdam, Amsterdam, the Netherlands.

1. INTRODUCTION

The infinite-server queue is arguably one of the main pillars of queueing theory, and has been studied almost since the inception of this field. It has found widespread usage in diverse application domains, often as an approximation for its many-server counterpart.

In infinite-server systems jobs arrive, are served in parallel (there is *no* waiting, that is), and leave when their service is completed. While the original motivation of infinite-server queueing systems stems from communication networks engineering, where the so-called Erlang model was developed to describe the dynamics of the number of telephone calls in progress, applications in various other domains have been explored, such as road traffic [12] and biology [1, 11].

The standard infinite-server model, which is commonly denoted in Kendall notation as M/G/∞, has the following operation. Jobs (which may, depending on the application, also be denoted as particles, customers, calls etc.) arrive according to a Poisson process with rate λ, where their service times form a sequence of independent and identically distributed (i.i.d.) random variables (distributed as a random variable B with finite first moment), independent of the call arrival process; a key result states that the stationary number of jobs in the system obeys a Poisson distribution with mean $\lambda\,\mathbb{E}B$. In many practical situations, however, the assumptions of a constant arrival rate and the jobs stemming from a single distribution are not realistic. A model that allows the input

Date: August 23, 2013.

Key words and phrases. Markov-modulated Poisson process, queues, general service times, large deviations.

process to exhibit some sort of variability (often referred to as 'burstiness') is the *Markov-modulated* infinite-server queue. In this model, a finite-state irreducible continuous-time Markov process (usually called the *background process*) modulates the input process: if the background process is in state i, the arrival process is a Poisson process with rate, say, λ_i, while the service times are distributed as a random variable, say, B_i (while the obvious independence conditions are imposed). The transition rate matrix of the background Markov chain is given by $(\nu_{ij})_{i,j=1}^d$.
The Markov-modulated infinite-server queue has attracted some (but relatively limited) attention in recent years. The main focus in the literature so far has been on characterizing (through the derivation of moments, or even the full probability generating function) the steady-state number of jobs in the system [5, 7, 9, 10]. Interestingly, under an appropriate time scaling [3, 8] in which the transitions of the background process occur at a faster rate than the Poisson arrivals, we retrieve the Poisson distribution for the steady-state number of jobs in the system. Recently, transient results have been obtained as well, under specific scalings of the arrival rates and transition times of the modulating Markov chain [3, 4].
The scaling considered in [3, 4] is such that the λ_i are *linearly* scaled (informally, $\lambda_i \mapsto N\lambda_i$), while the transition rates are *superlinearly* scaled (informally, $\nu_{ij} \mapsto N^{1+\varepsilon}\nu_{ij}$, for some $\varepsilon > 0$). The intuitive idea is that the time scale of the background process is faster than the time scale of the arrival process, such that the customer generation process becomes effectively a Poisson process with rate $\lambda_\infty := \sum_i \pi_i\lambda_i$, with π_i the stationary probability that the background process is in state i. As a result, the queueing system will behave as an infinite-server queue with arrival rate λ_∞.

Contribution. Where previous work [4] considers a central limit theorem in the scaling described above, we here focus on tail probabilities. Our main result is that the large deviations of the number of jobs in the system coincide with those of a Poisson random variable with mean $N\varrho_t$, with $\varrho_t := \sum_i \pi_i\lambda_i \int_0^t \mathbb{P}[B_i \geq s]\mathrm{d}s$. We also show the corresponding steady-state result.

Organization. The organization of the rest of this paper is as follows. In Section 2, we explain the model in detail and introduce some notation. In Section 3, we state and prove the main result of this paper. Numerical results are provided in Section 4. The final section of the paper, Section 5, contains some discussion of the results and concluding remarks.

2. MODEL DESCRIPTION

As mentioned above, this paper studies an infinite-server queue with Markov-modulated Poisson arrivals and general service times. In full detail, the model is described as follows.
Consider an irreducible continuous-time Markov process $(J(t))_{t\in\mathbb{R}}$ on a finite state space $\{1,\ldots,d\}$, with $d \in \mathbb{N}$. Its rate matrix is given by $(\nu_{ij})_{i,j=1}^d$. Let π_i be the stationary probability that the background process is in state i, for $i = 1,\ldots,d$. The time spent in state i (often referred to as the *transition time*) has an exponential distribution with mean $1/\nu_i$, where $\nu_i := -\nu_{ii}$.
While the process $(J(t))_{t\in\mathbb{R}}$, often referred to as the *background process* or *modulating process*, is in state i, jobs arrive according to a Poisson process with rate $\lambda_i \geq 0$. The service times are assumed to be i.i.d. samples distributed as a random variable B_i if the job was generated when the background process was in state i. The usual independence assumptions apply. We exclude the case that all λ_i as well as the distributions of the B_i coincide (as otherwise the queue is just an ordinary M/G/∞). We denote by $\overline{B}_i(\cdot)$ the complementary cumulative service distribution for jobs arriving during

background state i:

$$\overline{B}_i(t) = \mathbb{P}[B_i \geq t].$$

We use bold fonts to denote vectors; for instance $\boldsymbol{\lambda} \equiv (\lambda_1, ..., \lambda_d)$. We denote the invariant distribution corresponding to the rate matrix $(\nu_{ij})_{i,j=1}^d$ by $\boldsymbol{\pi}$.

3. MAIN RESULT

We perform the scaling $\lambda_i \mapsto N\lambda_i$, and $\nu_{ij} \mapsto N^{1+\varepsilon}\nu_{ij}$. We denote the background process (after the scaling) by $(J^{(N^{1+\varepsilon})}(t))_{t\in\mathbb{R}}$. Let $\boldsymbol{L}^{(N^{1+\varepsilon})}(t_1, t_2)$ be the empirical distribution of the background process in $[t_1, t_2)$ (with $t_1 < t_2$); its i-th component is the fraction of time spent in state i, for $i = 1, \ldots, d$ (where obviously the d components are non-negative and sum to 1). The object $\boldsymbol{L}(t_1, t_2)$ is the counterpart of $\boldsymbol{L}^{(N^{1+\varepsilon})}(t_1, t_2)$ for the non-scaled background process.
It is well known that the following law of large numbers applies: for any $\mathscr{S} \subset \mathbb{R}_+^d$ such that $\boldsymbol{\pi}$ is contained in the interior of $\mathscr{S}$, it holds that $\mathbb{P}(\boldsymbol{L}(0,t) \in \mathscr{S}) \to 1$ as $t \to \infty$. It is also a standard result (Thm. 3.1.6 in [6]) that $\boldsymbol{L}(0,t)$ satisfies a large deviations principle with rate function

$$\mathbb{I}(\boldsymbol{x}) := \sup_{\boldsymbol{u}>\boldsymbol{0}} \left(-\sum_{i=1}^d x_i \log \frac{\sum_{j=1}^d \nu_{ij} u_j}{u_i} \right); \tag{1}$$

this function is positive except when $\boldsymbol{x} = \boldsymbol{\pi}$. Under mild regularity conditions on the set $\mathscr{S}$, it means that

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}(L(0,t) \in \mathscr{S}) = -\inf_{\boldsymbol{x}\in\mathscr{S}} \mathbb{I}(\boldsymbol{x}).$$

As a consequence, if $\mathscr{S}$ does not contain $\boldsymbol{\pi}$, then $\mathbb{P}(L(0,t) \in \mathscr{S})$ decays essentially exponentially.
In the sequel, we need some additional notation. We define, for a function $f : \mathbb{R} \to \{1, \ldots, d\}$,

$$\varphi(f) := \int_0^t \lambda_{f(s)} \overline{B}_{f(s)}(t-s) \mathrm{d}s, \tag{2}$$

and

$$\varrho_t := \sum_{j=1}^d \pi_j \lambda_j \int_0^t \overline{B}_j(s) \mathrm{d}s. \tag{3}$$

Let $M^{(N)}(t)$ be the number of jobs in the system at time t. We wish to characterize the probability that $M^{(N)}(t)$ exceeds Na, given that the system starts off empty. We let $P^{(N)}(\lambda)$ denote a Poisson random variable with mean $N\lambda$. From [3, 5], we know that the $M^{(N)}(t)$ is distributionally equivalent to a Poisson random variable with random parameter:

$$M^{(N)}(t) \stackrel{d}{=} P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}(t)\right)\right). \tag{4}$$

Note that since $\varphi\left(J^{(N^{1+\varepsilon})}\right) \to \varrho_t$, a.s. for $N \to \infty$, we have that $N^{-1}M^{(N)}(t) \to \rho_t$, a.s. for $N \to \infty$. In this paper, we are concerned with the rare event that the number of jobs *exceeds* a level Na, with $a \geq \rho_t$.

Theorem 1. *For $a \geq \varrho_t$,*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(M^{(N)}(t) \geq Na\right) = -\varrho_t + a + a \log \frac{\varrho_t}{a}.$$

Proof. In view of the distributional equivalence (4), we have that

$$\mathbb{P}\left(M^{(N)}(t) \geq Na\right) = \mathbb{P}\left(P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na\right). \tag{5}$$

For $\delta > 0$, we define $\Delta(\boldsymbol{\pi})$ as a hypercube around $\boldsymbol{\pi}$:

$$\Delta(\boldsymbol{\pi}) := (\pi_1 - \delta, \pi_1 + \delta) \times \cdots \times (\pi_d - \delta, \pi_d + \delta). \tag{6}$$

Also introduce, for $\zeta > 0$, the event

$$\mathscr{E}_\delta(\zeta, N) := \left\{ \boldsymbol{L}^{(N^{1+\varepsilon})}\left(0, \frac{t}{N^\zeta}\right) \in \Delta(\boldsymbol{\pi}), \ldots, \boldsymbol{L}^{(N^{1+\varepsilon})}\left(\frac{\lceil N^\zeta \rceil - 1}{N^\zeta} t, t\right) \in \Delta(\boldsymbol{\pi}) \right\}. \tag{7}$$

Lower bound. We determine the decay rate of the obvious lower bound

$$\mathbb{P}\left(\left\{ P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na \right\} \cap \mathscr{E}_\delta\left(\frac{1}{2}, N\right) \right);$$

the idea is that we specialize to the scenario that the empirical distribution of the Markov chain is in $\Delta(\boldsymbol{\pi})$, and hence systematically close to $\boldsymbol{\pi}$.

To this end, first realize that, for any $\xi \in (0,1)$ and N sufficiently large, by virtue of the law of large numbers for the empirical distribution of the background process, see e.g. [6, Thm. 3.1.6]:

$$\mathbb{P}\left(\mathscr{E}_\delta\left(\frac{1}{2}, N\right)\right) \geq \prod_{i=1}^{\lceil \sqrt{N} \rceil} \min_{j_i \in \{1,\ldots,d\}} \mathbb{P}\left(\boldsymbol{L}\left(0, tN^{\frac{1}{2}+\varepsilon}\right) \in \Delta(\boldsymbol{\pi}) \,\middle|\, J(0) = j_i \right) \geq (1-\xi)^{\lceil \sqrt{N} \rceil}.$$

This immediately implies that

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\mathscr{E}_\delta\left(\frac{1}{2}, N\right)\right) = 0.$$

We are left with determining a lower bound on

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na \,\middle|\, \mathscr{E}_\delta\left(\frac{1}{2}, N\right)\right).$$

Recall that the Poisson random variable is stochastically increasing in its parameter. For that reason, we need to find a lower bound on $N\varphi(J^{(N^{1+\varepsilon})})$, conditional on $\mathscr{E}_\delta(\frac{1}{2}, N)$. By picking in every segment and for every state (a) a lower bound on the state probability (still in $\Delta(\boldsymbol{\pi})$), as well as (b) the lower bound on the Poisson rate in this segment (i.e. at the start of the segment), it is readily verified that the following (deterministic!) lower bound applies:

$$\varrho_t(N) := t\sqrt{N} \sum_{j=1}^{d} \sum_{i=1}^{\lfloor \sqrt{N} \rfloor} (\pi_j - \delta) \lambda_j \overline{B}_j\left(t\left(1 - \frac{(i-1)}{\sqrt{N}}\right)\right). \tag{8}$$

We thus obtain that

$$\mathbb{P}\left(P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na \,\middle|\, \mathscr{E}_\delta\left(\frac{1}{2}, N\right)\right) \geq e^{-\varrho_t(N)} \frac{(\varrho_t(N))^{\lceil Na \rceil}}{\lceil Na \rceil!}.$$

Applying Stirling's factorial approximation, it is seen that

$$\liminf_{N\to\infty} \frac{1}{N} \log\left(e^{-\varrho_t(N)} \frac{(\varrho_t(N))^{\lceil Na \rceil}}{\lceil Na \rceil!} \right) \geq \liminf_{N\to\infty} \frac{1}{N} \left(-\varrho_t(N) + Na + Na \log \frac{\varrho_t(N)}{Na} \right).$$

Observing that $\varrho_t(N)/N$ constitutes a Riemann integral with limit $\varrho_t^{(\delta)}$ as $N \to \infty$ (due to the fact that the functions $\overline{B}_j(\cdot)$ are Riemann integrable), with $\varrho_t^{(\delta)}$ defined as ϱ_t but with the π_j replaced by $\pi_j - \delta$, we conclude that

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na \,\middle|\, \mathscr{E}_\delta\left(\frac{1}{2}, N\right)\right) \geq -\varrho_t^{(\delta)} + a + a \log \frac{\varrho_t^{(\delta)}}{a}.$$

The stated follows by letting $\delta \downarrow 0$.

Upper bound. We consider the obvious upper bound

$$\mathbb{P}\left(\left\{P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na\right\} \cap \mathscr{E}_\delta\left(\frac{\varepsilon}{2}, N\right)\right) + \mathbb{P}\left(\mathscr{E}_\delta\left(\frac{\varepsilon}{2}, N\right)^c\right).$$

Due to the union bound,

$$\mathbb{P}\left(\mathscr{E}_\delta\left(\frac{\varepsilon}{2}, N\right)^c\right) \leq \lceil N^{\varepsilon/2} \rceil \left(\max_{j\in\{1,\ldots,d\}} \mathbb{P}\left(\boldsymbol{L}\left(0, tN^{1+\frac{\varepsilon}{2}}\right) \notin \Delta(\boldsymbol{\pi}) \,\middle|\, J(0) = j\right)\right).$$

Standard large deviations results imply that

$$\lim_{N\to\infty} \frac{1}{N^{1+\frac{\varepsilon}{2}}} \log \mathbb{P}\left(\boldsymbol{L}\left(0, tN^{1+\frac{\varepsilon}{2}}\right) \notin \Delta(\boldsymbol{\pi}) \,\middle|\, J(0) = j\right) = -\inf_{\boldsymbol{x}\notin\Delta(\boldsymbol{\pi})} \mathbb{I}(\boldsymbol{x}) < 0,$$

and hence

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\mathscr{E}_\delta\left(\frac{\varepsilon}{2}, N\right)^c\right) = -\infty.$$

Using [6, Lemma 1.2.15], it is now left to prove that

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na \,\middle|\, \mathscr{E}_\delta\left(\frac{\varepsilon}{2}, N\right)\right) \leq -\bar{\varrho}_t^{(\delta)} + a + a\log\frac{\bar{\varrho}_t^{(\delta)}}{a},$$

with $\bar{\varrho}_t^{(\delta)}$ defined as ϱ_t but with the π_j replaced by $\pi_j + \delta$; the stated then follows after sending $\delta \downarrow 0$. This upper bound is established as follows.

We need to find an upper bound on $N\varphi(J^{(N^{1+\varepsilon})})$, conditional on $\mathscr{E}_\delta(\frac{\varepsilon}{2}, N)$. Using a similar reasoning as in (8), it is readily verified that the following (deterministic!) upper bound applies:

$$\bar{\varrho}_t(N) := tN^{1-\frac{\varepsilon}{2}} \sum_{j=1}^{d} \sum_{i=1}^{\lceil N^{\frac{\varepsilon}{2}}\rceil} (\pi_j + \delta)\lambda_j \overline{B}_j\left(t\left(1 - \frac{i}{N^{\frac{\varepsilon}{2}}}\right)\right).$$

Chebycheff's inequality on the cumulant generating function of Poisson random variables [6, p. 30] now entails that

$$\begin{aligned}\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(P^{(N)}\left(\varphi\left(J^{(N^{1+\varepsilon})}\right)\right) \geq Na \,\middle|\, \mathscr{E}_\delta\left(\frac{\varepsilon}{2}, N\right)\right) \\ \leq \limsup_{N\to\infty} \frac{1}{N}\left(-\bar{\varrho}_t(N) + Na + Na\log\frac{\bar{\varrho}_t(N)}{Na}\right),\end{aligned}$$

which yields the desired upper bound, realizing that – using the same reasoning as above – $\bar{\varrho}_t(N)/N \to \bar{\varrho}_t^{(\delta)}$ as $N \to \infty$. □

The following corollary is an immediate consequence of the Gärtner-Ellis theorem and the duality between the cumulant function and the Legendre-Fenchel transform.

Corollary 1. *The limiting cumulant function of $M^{(N)}(t)$ corresponds to that of a Poisson random variable:*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\exp\left(\vartheta M^{(N)}(t)\right) = \varrho_t(e^\vartheta - 1).$$

The above result naturally extends to the steady-state counterpart $M^{(N)}$ of $M^{(N)}(t)$. To this end, we define $\varrho := \lim_{t\to\infty} \varrho_t = \sum_i \pi_i\lambda_i\mathbb{E}[B_i]$ and realize that $M^{(N)}$ has a Poisson distribution with mean

$$N\int_{-\infty}^{0} \lambda_{f(s)}\overline{B}_{f(s)}(-s)\mathrm{d}s;$$

see e.g. [5]. Then the proof of the corollary below is essentially the same as the one for the transient case.

Corollary 2. *For $a \geq \varrho$,*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(M^{(N)} \geq Na\right) = -\varrho + a + a \log \frac{\varrho}{a}.$$

In addition, $N^{-1} \log \mathbb{E} \exp\left(\vartheta M^{(N)}\right) \to \varrho(e^{\vartheta} - 1)$ as $N \to \infty$.

We conclude this section by noting that the for the transient result, finiteness of the first moment of B_i is in fact not required in the proof.

4. Numerical examples

We illustrate the results of this paper with a stochastic simulation study. In order to circumvent the long run-times associated with crude Monte-Carlo simulations of rare events, we simulate the quantity of interest in the following way. Introducing the random variable $Y := \varphi(J^{(N^{1+\varepsilon})})$, we can write the probability of interest as:

$$\mathbb{P}\left(M^{(N)}(t) \geq Na\right) = \mathbb{E}[p_{Na}(Y)],$$

where $p_a(\lambda)$ denotes the complementary cumulative distribution function of the Poisson distribution:

$$p_a(\lambda) = \sum_{k=a}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!}.$$

An efficient simulation thus consists of simulating K trajectories of the background Markov chain, and then perform an exact computation of the distribution of the Poisson distributed random variable $M^{(N)}(t)$, with a rate that can be extracted from the trajectory, so that the only source of error is due to the uncertainty wrt the rate.

As $p_a(\lambda)$ is a smooth function of λ, we can estimate confidence intervals by applying what is known as the delta method [2, p. 75]. Thus, we have that the variance to be used for the confidence intervals is equal to $\sigma^2 = p'_{Na}(\rho_t)^2 \operatorname{Var}[Y]$. By computing the sample variance of Y, s_Y^2, we have an approximate confidence interval equal to

$$[\hat{z} - c_\alpha p'_{Na}(N\rho_t) s_Y/\sqrt{K}, \hat{z} + c_\alpha p'_{Na}(N\rho_t) s_Y/\sqrt{K}], \tag{9}$$

where $\hat{z}$ denotes the estimate for the probability of interest, K the number of runs and c_α the coefficient corresponding to a $1-\alpha$ confidence interval of the standard normal distribution.

Although we have proved only the exponential tail asymptotics, we recall the Bahadur-Rao theorem on exact asymptotics of the Poisson distribution, so as to verify whether this refinement leads to a better fit; we refer to [6, Thm. 3.7.4]. The cumulant generating function of a Poisson distributed random variable is equal to $\Lambda(\theta) := \log \mathbb{E}[e^{\theta X}] = \lambda(e^{\theta} - 1)$. The probability that a Poisson random variable with rate $N\lambda$ exceeds Na, with $a > \lambda$, can be written asymptotically as

$$\mathbb{P}(P^{(N)}(\lambda) \geq Na) \sim \frac{1}{(1-e^{-\eta})\sqrt{\Lambda''(\eta) 2\pi N}} e^{-N\Lambda(\eta)}, \tag{10}$$

where η denotes the positive solution of $\Lambda'(\eta) = a$. After plugging in the expressions specific for the Poisson distribution we get,

$$\mathbb{P}(P^{(N)}(\lambda) \geq Na) \sim \frac{1}{1-\frac{\lambda}{a}} \frac{1}{\sqrt{2\pi a N}} \left(\frac{\lambda}{a}\right)^{Na} e^{-N(\lambda - a)}. \tag{11}$$

	$\varepsilon = 0.01$	$\varepsilon = 0.25$	$\varepsilon = 0.75$	$\varepsilon = 1.25$	
$N = 20$	$[1.182921, 1.199136]$	$[1.184967, 1.197062]$	$[1.188300, 1.193700]$	$[1.189724, 1.192271]$	1.190230
$N = 40$	$[1.132820, 1.144161]$	$[1.134665, 1.142299]$	$[1.137006, 1.139946]$	$[1.137897, 1.139053]$	1.138278
$N = 60$	$[1.113972, 1.123353]$	$[1.115762, 1.121549]$	$[1.117624, 1.119681]$	$[1.118292, 1.119011]$	1.118563
$N = 80$	$[1.103918, 1.112138]$	$[1.105583, 1.110462]$	$[1.107238, 1.108802]$	$[1.107754, 1.108285]$	1.107970
$N = 100$	$[1.097761, 1.104898]$	$[1.099229, 1.103421]$	$[1.100678, 1.101968]$	$[1.101112, 1.101534]$	1.101291
$N = 120$	$[1.093138, 1.100255]$	$[1.094810, 1.098575]$	$[1.096147, 1.097235]$	$[1.096517, 1.096864]$	1.096668
$N = 200$	$[1.084372, 1.089404]$	$[1.085554, 1.088218]$	$[1.086510, 1.087260]$	$[1.086792, 1.086977]$	1.086877
$N = 300$	$[1.079575, 1.083622]$	$[1.080570, 1.082623]$	$[1.081347, 1.081846]$	$[1.081535, 1.081657]$	1.081593
$N = 400$	$[1.077008, 1.080606]$	$[1.077904, 1.079708]$	$[1.078612, 1.078999]$	$[1.078761, 1.078849]$	1.078803
$N = 500$	$[1.075420, 1.078716]$	$[1.076296, 1.077837]$	$[1.076908, 1.077225]$	$[1.077033, 1.077101]$	1.077065

TABLE 1. Simulated decay rates for a two-state background Markov chain with $\nu_1 = 1$ and $\nu_2 = 3$; $\lambda_1 = 1$, $\lambda_2 = 2$; $\mu_1 = 2$, $\mu_2 = 1$; $a = 2$, $t = 0.8$ for different ε and N, with 95% confidence intervals (Eq. (9)). Exact value ≈ 1.0690. The last column contains the Bahadur-Rao based value (Eq. (11)). The number of runs is equal to $K = 800$.

In our example, we consider a two-state background Markov chain with $\nu_1 = 1$ and $\nu_2 = 3$; $\lambda_1 = 1$, $\lambda_2 = 2$; $a = 2$ and exponential service times with rates $\mu_1 = 2$, $\mu_2 = 1$. For these parameters, $\rho_t \approx 0.5746$ and the associated Poisson decay rate is $I \approx 1.0690$. Furthermore, we take the number of runs K to be equal to 800. In Table 1, we show the decay rates $-(1/N) \cdot \log \mathbb{P}(M^{(N)} \geq Na)$, for different N and ε. We observe that the simulated decay rate encompasses the predicted decay rate I, with confidence intervals that get smaller as ε gets larger. This is intuitively clear as the faster the background cycles, the less likely it is that the empirical distribution of the background chain differs substantially from the steady-state distribution.

In Fig. 1 we show for $N = 100$ the simulated 0.95 confidence interval of the Poisson parameter ρ_t versus ε, whose width is equal to $c_{0.05} s_Y / \sqrt{K}$. This plot illustrates that the confidence interval rapidly gets smaller as ε increases. This is further evidence of the fact that larger ε will see faster convergence to Poissonian asymptotic behavior. Indeed, as the number of particles is a Poisson distribution with a random parameter, and the confidence interval of said random parameter gets rapidly smaller, we can anticipate Poissonian asymptotics as well. In Fig. 2, we plot the logarithm of the width of the confidence interval against epsilon and find a linear relation, which suggests that the confidence interval is asymptotically $kN^{-c\varepsilon}$, for certain values of k and c.

Lastly, we show in Fig. 3 a contour plot of the confidence interval width on the Poisson parameter ρ_t versus ε and N. We see, as expected that the confidence interval gets smaller when N or ε get bigger.

5. DISCUSSION AND CONCLUDING REMARKS

We have seen that, under the time-scaling considered (arrival rates linearly and transition rates superlinearly, that is), the tail asymptotics in the Markov modulated infinite server model tend to those in a corresponding M/M/∞ system; the rationale is that the background process is jumping faster than the time-scale of the arrivals, so that the arrival stream becomes increasingly Poisson as N tends to ∞.

In the model considered in this paper, the service times were sampled upon arrival instants. In case of exponential service times, there is a second version of the Markov-modulated infinite-server queue, though: a version in which the departure rate of each job is μ_i if the background process

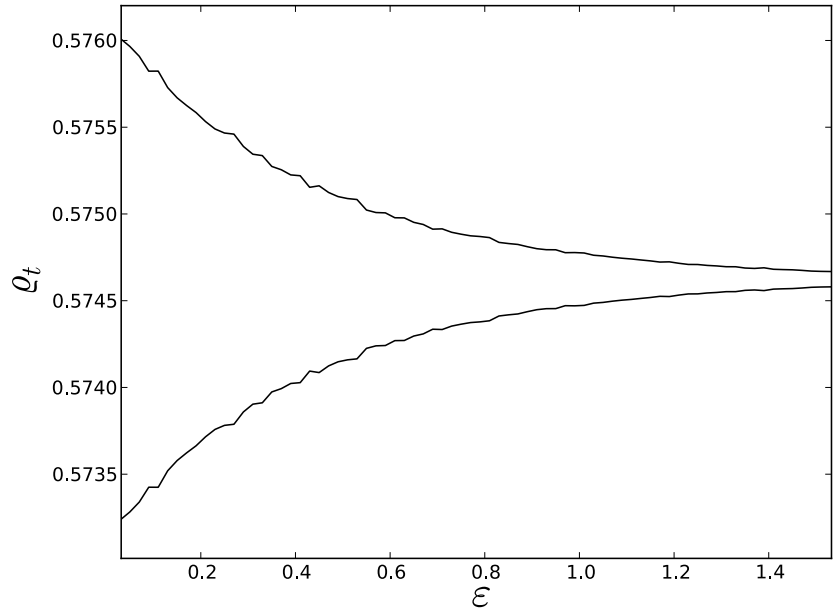

FIGURE 1. Simulated Poisson parameter for a two-state background Markov chain with $\nu_1 = 1$ and $\nu_2 = 3$; $\lambda_1 = 1$, $\lambda_2 = 2$; $\mu_1 = 2$, $\mu_2 = 1$; $a = 2$, $t = 0.8$; $N = 50$ versus ε. The gray area represents the confidence interval. The number of runs at each data point is equal to $K = 800$.

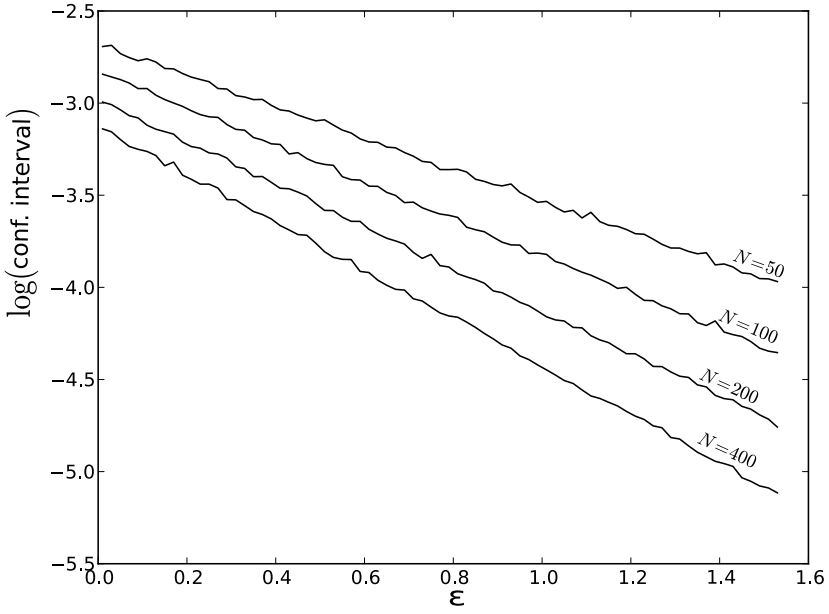

FIGURE 2. A plot of the (decimal) logarithm of the confidence interval width versus ε.

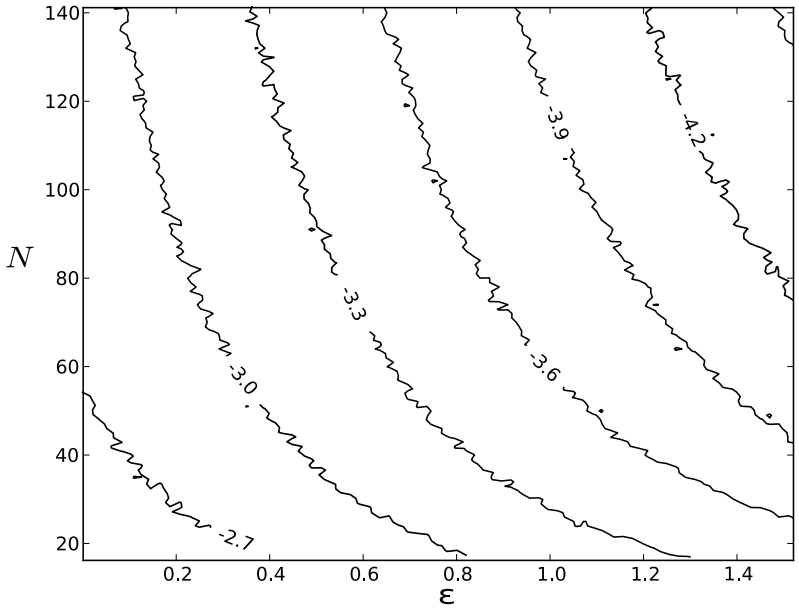

FIGURE 3. A contour plot of the confidence interval width versus N (vertically) and ε (horizontally). The labels show the decimal logarithm of the confidence interval width.

is in state i. It is conceivable that in this case, the tail asymptotics of the normalized stationary number of jobs $M^{(N)}/N$ tend to those of $P^{(N)}(\lambda_\infty/\mu_\infty)$, with $\mu_\infty := \sum_{i=1}^d \pi_i\mu_i$, whereas (as before) $\lambda_\infty := \sum_{i=1}^d \pi_i\lambda_i$). In addition, in the transient setting we anticipate the Poisson parameter to equal $\lambda_\infty/\mu_\infty \cdot (1-e^{-\mu_\infty t})$. It is noted, however, that the proof technique used in the present paper does not extend to this setting. Other generalizations can be thought of, such as non-exponential transition times (cf. [4, 8]), and the case $\varepsilon = 0$.

REFERENCES

[1] A. ARAZI, E. BEN-JACOB, and U. YECHIALI (2004). Bridging genetic networks and queueing theory. *Physica A*, **332**, 585–616.

[2] S. ASMUSSEN and P. GLYNN (2007). *Stochastic Simulation*. Springer, New York.

[3] J. BLOM, O. KELLA, M. MANDJES, and H. THORSDOTTIR (2012). Markov-modulated infinite server queues with general service times. To appear in Queueing Systems (DOI:10.1007/s11134-013-9368-4)

[4] J. BLOM, M. MANDJES, and H. THORSDOTTIR (2013). Time-scaling limits for Markov-modulated infinite-server queues. *Stochastic Models*, **29**, 112–127.

[5] B. D'AURIA (2008). M/M/∞ queues in semi-Markovian random environment. *Queueing Systems*, **58**, 221–237.

[6] A. DEMBO and O. ZEITOUNI (1998). *Large Deviations Techniques and Applications*, 2nd edition. Springer, New York.

[7] B. FRALIX and I. ADAN (2009). An infinite-server queue influenced by a semi-Markovian environment. *Queueing Systems*, **61**, 65–84.

[8] T. HELLINGS, M. MANDJES, and J. BLOM (2012). Semi-Markov-modulated infinite-server queues: approximations by time-scaling. *Stochastic Models*, **28**, 452–477.

[9] J. KEILSON and L. SERVI (1993). The matrix M/M/∞ system: retrial models and Markov modulated sources. *Advances in Applied Probability*, **25**, 453–471.

[10] C. O'CINNEIDE and P. PURDUE (1986). The M/M/∞ queue in a random environment. *Journal of Applied Probability*, **23**, 175–184.

[11] A. SCHWABE, K. RYBAKOVA, and F. BRUGGEMAN (2012). Transcription Stochasticity of Complex Gene Regulation Models. *Biophysical Journal*, **103**, pp. 1152-1161.

[12] T. VAN WOENSEL and N. VANDAELE (2007). Modeling traffic flows with queueing models: a review. *Asia-Pacific Journal of Operational Research*, **24**, 235–261.

E-mail address: joke.blom@cwi.nl, kdeturck@telin.ugent.be, M.R.H.Mandjes@uva.nl