A FUNCTIONAL CENTRAL LIMIT THEOREM FOR
A MARKOV-MODULATED INFINITE-SERVER QUEUE

D. ANDERSON[a], J. BLOM[c], M. MANDJES[b,c], H. THORSDOTTIR[c,b], K. DE TURCK[d]

ABSTRACT. We consider a model in which the production of new molecules in a chemical reaction network occurs in a seemingly stochastic fashion, and can be modeled as a Poisson process with a varying arrival rate: the rate is λ_i when an external Markov process $J(\cdot)$ is in state i. It is assumed that molecules decay after an exponential time with mean μ^{-1}. The goal of this work is to analyze the distributional properties of the number of molecules in the system, under a specific time-scaling. In this scaling, the background process is sped up by a factor N^α, for some $\alpha > 0$, whereas the arrival rates become $N\lambda_i$, for N large. The main result of this paper is a functional central limit theorem (F-CLT) for the number of molecules, in that the number of molecules, after centering and scaling, converges to an Ornstein-Uhlenbeck process. An interesting dichotomy is observed: (i) if $\alpha > 1$ the background process jumps faster than the arrival process, and consequently the arrival process behaves essentially as a (homogeneous) Poisson process, so that the scaling in the F-CLT is the usual $\sqrt{N}$, whereas (ii) for $\alpha \leq 1$ the background process is relatively slow, and the scaling in the F-CLT is $N^{1-\alpha/2}$. In the latter regime, the parameters of the limiting Ornstein-Uhlenbeck process contain the deviation matrix associated with the background process $J(\cdot)$.

Keywords. Ornstein-Uhlenbeck processes ⋆ Markov modulation ⋆ central-limit theorems ⋆ martingale methods

[a] Department of Mathematics, University of Wisconsin – Madison, 480 Lincoln Drive, Madison WI 53706, United States.
[b] Korteweg-de Vries Institute (KdVI) for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands.
[c] CWI, Science Park 123, P.O. Box 94079, 1090 GB Amsterdam, the Netherlands.
[d] 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.

Email addresses. `anderson@math.wisc.edu`, `j.g.blom@cwi.nl`, `m.r.h.mandjes@uva.nl`, `h.thorsdottir@cwi.nl`, `kdeturck@telin.ugent.be`

1. INTRODUCTION

When modeling chemical reaction networks within cells, the dynamics of the numbers of molecules of the various types are often described by deterministic differential equations. These models ignore the inherent stochasticity that may play a role, particularly when the number of molecules are relatively small. To remedy this, the use of stochastic representations of chemical networks has been advocated, see e.g. [2, 10, 16].

In this paper we use the formulation as in [1, 4] where the numbers of molecules evolve as a continuous-time Markov chain. A concise description of this formulation is the following, with our specific model more formally developed in Section 2. Consider a model consisting of a finite number, ℓ, of species and a finite number, K, of reaction channels. We let $\boldsymbol{M}(t)$ be the ℓ-dimensional vector whose ith component gives the number of molecules of the ith species present at time t. For the kth reaction channel we denote by $\boldsymbol{\nu}_k \in \mathbb{Z}_{\geq 0}^{\ell}$ the number of molecules of each species needed for the reaction to occur, and by $\boldsymbol{\nu}_k' \in \mathbb{Z}_{\geq 0}^{\ell}$ the number produced. We let $\mu_k(\boldsymbol{x})$ denote the rate, or intensity (termed *propensity* in the biology literature), at which the kth reaction occurs when the numbers of molecules present equals the vector $\boldsymbol{x}$. Then, $\boldsymbol{M}(t)$ may be represented as the solution to the (vector-valued) equation

$$(1) \qquad \boldsymbol{M}(t) = \boldsymbol{M}(0) + \sum_{k=1}^{K} (\boldsymbol{\nu}_k' - \boldsymbol{\nu}_k) Y_k \left(\int_0^t \mu_k(\boldsymbol{M}(s)) \mathrm{d}s \right),$$

where the stochastic processes $Y_k(\cdot)$ are independent unit-rate Poisson processes [1]. Note that if, for some $k^\star \in \{1, \ldots, K\}$, $\boldsymbol{\nu}_{k^\star}' - \boldsymbol{\nu}_{k^\star}$ equals the ith unit vector $\boldsymbol{e}_i$, then the $k^\star$th reaction channel corresponds to the external arrival of molecules of species i. For the specific situation that subnetworks operate at disparate timescales, these can be analyzed separately by means of lower dimensional approximations, as pointed out in e.g. [19].

In this paper we study a model of the type described above, for the special case that there is just one type of molecular species (i.e., $\ell = 1$), and that there are external arrivals. The distinguishing feature is that the rate of the external input is determined by an independent continuous-time Markov chain $J(\cdot)$ (commonly referred to as the *background process*) defined on the finite state space $\{1, \ldots, d\}$. More concretely, we study a reaction system that obeys the stochastic representation

$$M(t) = M(0) + Y_1 \left(\int_0^t \lambda_{J(s)} \mathrm{d}s \right) - Y_2 \left(\mu \int_0^t M(s) \mathrm{d}s \right),$$

where $Y_1(\cdot)$ and $Y_2(\cdot)$ are independent unit-rate Poisson processes, and $\lambda_{J(s)}$ takes the value $\lambda_i \geq 0$ when the background process is in state i. Hence, in this model external molecules flow into the system according to a Poisson process with rate λ_i when the background process $J(\cdot)$ is in state i, while each molecule decays after an exponentially distributed time with mean μ^{-1} (independently of other molecules present).

The main result of the paper is a functional central limit theorem (F-CLT) for the process $M(t)$, where we impose a specific scaling on the transition rates $Q = (q_{ij})_{i,j=1}^{d}$ of the background process $J(t)$, as well as on the external arrival rates $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)^{\mathrm{T}}$ (note that all vectors are to be understood as column vectors). More precisely, the transition rates of the background process are sped up by a

factor N^α, with $\alpha > 0$, while the arrival rates are sped up linearly, that is, they become $N\lambda_i$. Then we consider the process $U^N(t)$ (with a superscript N to stress the dependence on N), obtained from $M^N(t)$ by centering, that is, subtracting the mean $\mathbb{E}M^N(t)$, and normalizing, that is, dividing by an appropriate polynomial in N. It is proven that $U^N(t)$ converges (as $N \to \infty$) weakly to a specific Gauss-Markov process, viz. an Ornstein-Uhlenbeck (OU) process with certain parameters (which are given explicitly in terms of $\boldsymbol{\lambda}, \mu$, and the matrix Q). Our proofs are based on martingale techniques; more specifically, an important role is played by the martingale central limit theorem.

Interestingly, if $\alpha > 1$ the normalizing polynomial in the F-CLT is the usual $\sqrt{N}$, but for $\alpha \leq 1$ it turns out that we have to divide by $N^{1-\alpha/2}$. The main intuition behind this dichotomy is that for $\alpha > 1$ the timescale of the background process is faster than that of the arrival process, and hence the arrival process is effectively a (homogeneous) Poisson process. As a result the corresponding F-CLT is in terms of the corresponding Poisson rate (which we denote by $\lambda_\infty := \boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{\lambda}$, where $\boldsymbol{\pi}$ is the stationary distribution of the background process) and μ only. For $\alpha \leq 1$, on the contrary, the background process jumps relatively slowly; the limiting OU process is in terms of $\boldsymbol{\lambda}$ and μ, but features the deviation matrix [11] associated to the background process $J(\cdot)$ as well.

In earlier works [6, 7] we studied a similar setting. However, where we use a martingale-based approach in the present paper, in [6, 7] we relied on another technique: (i) we set up a system of differential equations for the Laplace transform of $M^N(t)$ jointly with the state of the background process $J^N(t)$, (ii) modified these into a system of differential equations for the transform of the (centered and normalized) process $U^N(t)$ jointly with $J^N(t)$, (iii) approximated these by using Taylor expansions, and (iv) then derived an ordinary differential equation for the limit of the transform of $U^N(t)$ (as $N \to \infty$). This differential equation defining a Normal distribution, the CLT was established. Importantly, the results derived in [6, 7] crucially differ from the ones in the present paper. The most significant difference is that those results are *no* F-CLT: just the *finite-dimensional convergence* to the OU process was established, rather than convergence at the process level (i.e., to prove weak convergence an additional 'tightness argument' would be needed). For the sake of completeness, we mention that [7] covers just the case $\alpha > 1$, that is, the regime in which the arrival process is effectively Poissonian, while [6] allows all $\alpha > 0$.

Our previous works [6, 7, 8] have been presented in the language of *queueing theory*; the model described above can be seen as an infinite-server queue with Markov-modulated input. In comparison to Markov-modulated single-server queues (and to a lesser extent Markov-modulated many-server queues), this infinite-server model has been much less intensively studied. This is potentially due to the fact that the presence of infinitely many servers may be considered less realistic, perhaps rightfully so in the context of operational research, the major application field of queueing theory. In the context of chemical reactions, however, it can be argued that the concept of infinitely many servers is quite natural: each molecule brings its own 'decay'-server.

We conclude this introduction with a few short remarks on the relation of our work with existing literature. Incorporating Markov modulation in the external arrival rate the infinite-server queue becomes, from a biological perspective, a more realistic model [25]. It is noted that deterministic modulation has been studied in [13] for various types of non-homogeneous arrival rate functions

(M_t/G/∞ queue). For earlier results on the stationary distribution of Markov-modulated infinite-server queues, for instance in terms of a recursive scheme that determines the moments, we refer to e.g. [12, 15, 20, 22].

As mentioned above, at the methodological level, our work heavily relies on the so-called martingale central limit theorem (M-CLT), see for instance [14, 26]. It is noted that convergence to OU has been established in the non-modulated setting before: an appropriately scaled M/M/∞ queue weakly converges to an OU process. For a proof, see e.g. [24, Section 6.6]; cf. also [9, 17].

The rest of this paper is organized as follows. Section 2 sets up the model, its properties and quantities of interest, and presents the essential mathematical tools. The N^α-scaled background process is thoroughly investigated in Section 3; most notably we derive its F-CLT relying on the M-CLT. This takes us to Section 4, where we first show that $\bar{M}^N(t) := N^{-1} M^N(t)$ converges to a deterministic solution, denoted by $\varrho(t)$, to finally establish the F-CLT for $M^N(t)$ by proving asymptotic normality of the process $N^\beta (\bar{M}^N(t) - \varrho(t))$, with $\beta \in (0, 1/2)$ appropriately chosen. As indicated earlier, the parameters specifying the limiting OU process depend on which speedup is 'faster': the one corresponding to the background process (i.e., $\alpha > 1$) or that of the arrival rates (i.e., $\alpha < 1$). We conclude this paper by a set of numerical experiments, that illustrate the impact of the value of α.

2. THE MODEL AND MATHEMATICAL TOOLS

In this section we first describe our model in detail, and then present preliminaries (viz. a version of the law of large numbers for Poisson processes and the M-CLT).

Model. Our paper considers the following Markovian model. Let $J(t)$ be an irreducible continuous-time Markov process on the finite state space $\{1, \ldots, d\}$. Define its generator matrix by $Q = (q_{ij})_{i,j=1}^d$ and the (necessarily unique) invariant distribution by $\boldsymbol{\pi}$; as a consequence, $\boldsymbol{\pi}^{\mathrm{T}} Q = \mathbf{0}^{\mathrm{T}}$. Let $X_i(t)$ be the indicator function of the event $\{J(t) = i\}$, for $i = 1, \ldots, d$; in other words: $X_i(t) = 1$ if $J(t) = i$ and 0 otherwise. It is assumed that $J(\cdot)$ is in stationarity at time 0 and hence at any t; we thus have $\mathbb{P}(J(t) = i) = \pi_i$. As commonly done in the literature, the transient distribution $\mathbb{P}(J(s) = j \mid J(0) = i)$ is denoted by $p_{ij}(s)$ and is computed as $(e^{Qs})_{i,j}$.

The model considered in this paper is a so-called *Markov-modulated infinite-server queue*. Its dynamics can be described as follows. For any time $t \geq 0$, molecules arrive according to a Poisson process with rate λ_i if $X_i(t) = 1$. We let the service/decay rate of each molecule be μ irrespective of the state of the background process. There are infinitely many servers so that the molecules' sojourn times are their service times; the molecules go in service immediately upon arrival. Throughout this paper, $M(t)$ denotes the number of molecules present at time t.

Scaling. In this paper a F-CLT under the following scaling is investigated. The background process as well as the arrival process are sped up, while the service-time distribution remains unaffected. More specifically, the transition matrix of the background process becomes $N^\alpha Q$ for some $\alpha > 0$, while the arrival rates, λ_i for $i = 1, \ldots, d$, are scaled linearly (i.e., become $N \lambda_i$); then N is sent to ∞. To indicate the fact that they depend on the scaling parameter N, we write in the sequel $J^N(t)$ for the background process, $X_i^N(t)$ for the indicator function associated with state i of the background

process at time t, and $M^N(t)$ for the number of molecules in the system at time t. Later in the paper we let the transitions of the background process go from being sublinear (i.e., $\alpha < 1$) to superlinear (i.e., $\alpha > 1$); one of our main findings is that there is a *dichotomy*, in the sense that there is crucially different behavior in these two regimes, with a special situation at the boundary, i.e., $\alpha = 1$.

The above model can be put in terms of a chemical reaction network, as formulated in the introduction. It turns out to be convenient to do so by interpreting the background process as a model for a single molecule transitioning between d different states, with $X_i^N(t)$ denoting the number of molecules in state i at time t. Since there is at most one such molecule, we see $X_i^N(t) \in \{0, 1\}$. The following table informally summarizes the relevant reactions and corresponding intensity functions for the model of interest:

Reaction	Intensity function	Description
$X_i \to X_j$ (for $i \neq j$)	$N^\alpha q_{ij} X_i^N(t)$	$J(\cdot)$ jumps from i to j
$\emptyset \to M$	$\sum_{i=1}^d N\lambda_i X_i^N(t)$	Arrival
$M \to \emptyset$	$\mu M^N(t)$	Departure

As mentioned above, it is assumed that $\alpha > 0$; in addition, $q_{ij} \geq 0$ for $i \neq j$ (while $Q\mathbf{1} = \mathbf{0}$), $\lambda_i \geq 0$, and $\mu > 0$. The dynamics can be phrased in terms of the stochastic representation framework, as described in the introduction. In the first place, the evolution of the indicator functions can be represented as

$$(2) \quad X_i^N(t) = X_i^N(0) - \sum_{\substack{j=1 \\ j \neq i}}^d Y_{i,j}\left(N^\alpha q_{ij}\int_0^t X_i^N(s)\mathrm{d}s\right) + \sum_{\substack{j=1 \\ j \neq i}}^d Y_{j,i}\left(N^\alpha q_{ji}\int_0^t X_j^N(s)\mathrm{d}s\right)$$

where the $Y_{i,j}$ ($i, j = 1, \ldots, d$ with $i \neq j$) are independent unit-rate Poisson processes. It is readily verified that if the $X_i^N(0)$, with $i = 1, \ldots, d$, are indicator functions summing up to 1, then so are the $X_i^N(t)$ for any $t \geq 0$. The second (third, respectively) term in the right-hand side represents the number of times that $J^N(\cdot)$ leaves (enters) state i in $[0, t]$.

In the second place, the number of molecules in the system evolves as

$$(3) \quad M^N(t) = M^N(0) + Y_1\left(N\int_0^t \sum_{i=1}^d \lambda_i X_i^N(s)\mathrm{d}s\right) - Y_2\left(\mu\int_0^t M^N(s)\mathrm{d}s\right),$$

where Y_1, and Y_2 are independent unit-rate Poisson processes (also independent of the $Y_{i,j}$).

The objective of this paper is to describe the limiting behavior of the system as $N \to \infty$, for different values of α. Our main result is a F-CLT for the process $M^N(\cdot)$; to establish this, we also need a F-CLT for the *state frequencies* of $J^N(\cdot)$ on $[0, t]$, defined as

$$\mathbf{Z}^N(t) = (Z_1^N(t), \ldots, Z_d^N(t))^{\mathrm{T}}, \quad \text{with} \quad Z_i^N(t) := \int_0^t X_i^N(s)\mathrm{d}s.$$

This paper essentially makes use of two more or less standard 'tools' from probability theory: the law of large numbers applied to Poisson processes and the martingale central limit theorem (M-CLT). For the sake of completeness, we state the versions used here.

Lemma 2.1. [1, Thm. 2.2] *Let Y be a unit rate Poisson process. Then for any $U > 0$,*

$$\lim_{N \to \infty} \sup_{0 \le u \le U} \left| \frac{Y(Nu)}{N} - u \right| = 0,$$

almost surely.

The following is known as (a version of) the M-CLT, and is a corollary to Thm. 7.1.4 and the proof of Thm. 7.1.1 in [14]. Here and in the sequel, '$\Rightarrow$' denotes weak convergence; in addition, $[\cdot, \cdot]_t$ is the quadratic covariation process.

Theorem 2.2. *Let $\{\boldsymbol{\mathcal{M}}^N\}$, for $N \in \mathbb{N}$, be a sequence of $\mathbb{R}^d$-valued martingales with $\boldsymbol{\mathcal{M}}^N(0) = \boldsymbol{0}$ for any $N \in \mathbb{N}$. Suppose*

$$\lim_{N \to \infty} \mathbb{E} \left[\sup_{s \le t} \left| \boldsymbol{\mathcal{M}}^N(s) - \boldsymbol{\mathcal{M}}^N(s-) \right| \right] = \boldsymbol{0}, \text{ with } \boldsymbol{\mathcal{M}}^N(s-) := \lim_{u \uparrow s} \boldsymbol{\mathcal{M}}^N(u),$$

and, as $N \to \infty$,

$$[\mathcal{M}_i^N, \mathcal{M}_j^N]_t \to C_{ij}(t)$$

for a deterministic matrix $C_{ij}(t)$ that is continuous in t, for $i, j = 1, \ldots, d$ and for all $t > 0$. Then $\boldsymbol{\mathcal{M}}^N \Rightarrow \boldsymbol{W}$, where $\boldsymbol{W}$ denotes a Gaussian process with independent increments and $\mathbb{E} \left[\boldsymbol{W}(t) \boldsymbol{W}(t)^{\mathrm{T}} \right] = C(t)$ (such that $\mathbb{E}[W_i(t) W_j(t)] = C_{ij}(t)$).

There is an extensive body of literature on the M-CLT; for more background, see e.g. [18, 23, 26].

3. A FUNCTIONAL CLT FOR THE STATE FREQUENCIES

In this section we establish the F-CLT for the integrated background processes $\boldsymbol{Z}^N(t)$, that is, the state frequencies of the Markov process $J^N(\cdot)$ on $[0, t]$. This F-CLT is a crucial element in the proof of the F-CLT of $M^N(t)$, as will be given in the next section. It is noted that there are several ways to establish this F-CLT; we refer for instance to the related weak convergence results in [5, 21], as well as the nice, compact proof for the single-dimensional convergence in [3, Ch. II, Thm. 4.11]. We chose to include our own derivation, as it is straightforward, insightful and self-contained, while at the same time it also introduces a number of concepts and techniques that are used in the M-CLT-based proof of the F-CLT for $M^N(t)$ in the next section.

We first identify the corresponding law of large numbers. To this end, we consider the process $\boldsymbol{X}^N(t)$ by dividing both sides of (2) by N^α and letting $N \to \infty$. Since $X_i^N(t) \in \{0, 1\}$ for all t, the $X_i^N(t)$ and $X_i^N(0)$ terms both go to zero as $N \to \infty$. Thus we may apply Lemma 2.1 to see that also, as $N \to \infty$,

$$-\sum_{j \neq i} q_{ij} Z_i^N(t) + \sum_{j \neq i} q_{ji} Z_j^N(t) \to 0,$$

almost surely, or $\lim_{N \to \infty} \boldsymbol{Z}^N(t)^{\mathrm{T}} Q = \boldsymbol{0}^{\mathrm{T}}$. Bearing in mind that $\boldsymbol{1}^{\mathrm{T}} \boldsymbol{Z}^N(t) = t$, the limit of $\boldsymbol{Z}^N(t)$ solves the global balance equations, entailing that

(4) $$Z_j^N(t) \to \pi_j t$$

almost surely as $N \to \infty$, where we recall that $\boldsymbol{\pi}$ is the stationary distribution associated with the background process $J(\cdot)$.

As mentioned above, the primary objective of this section is to establish a F-CLT for $\boldsymbol{Z}^N(\cdot)$ as $N \to \infty$. More specifically, we wish to identify a covariance matrix C such that, as $N \to \infty$,

$$(5) \qquad N^{\alpha/2} \left(\boldsymbol{Z}^N(t) - \boldsymbol{\pi}t \right) \Rightarrow \boldsymbol{W}(t),$$

with $\boldsymbol{W}(\cdot)$ representing a (d-dimensional) Gaussian process with independent increments such that $\mathbb{E}\left[\boldsymbol{W}(t)\boldsymbol{W}(t)^{\mathrm{T}} \right] = C\,t$. In other words: our goal is to show weak convergence to a d-dimensional Brownian motion (with dependent components).

We start our exposition by identifying a candidate covariance matrix C, by studying the asymptotic behavior (that is, as $N \to \infty$) of $\mathbb{C}\mathrm{ov}(Z_i^N(t), Z_j^N(t))$ for fixed t. Bearing in mind that $Z_i^N(t)$ is the integral over s of $X_i^N(s)$, and using standard properties of the covariance, this covariance can be rewritten as

$$\int_0^t \int_0^s \mathbb{C}\mathrm{ov}\left(X_i^N(r), X_j^N(s) \right) \,\mathrm{d}r\mathrm{d}s + \int_0^t \int_s^t \mathbb{C}\mathrm{ov}\left(X_i^N(r), X_j^N(s) \right) \,\mathrm{d}r\mathrm{d}s.$$

Recalling that the process $J^N(\cdot)$ starts off in equilibrium at time 0, and that $X_i(s)$ is the indicator function of the event $\{J^N(s) = i\}$, this expression can be rewritten as

$$\int_0^t \int_0^s \left(\pi_i p_{ij}^N(s - r) - \pi_i \pi_j \right) \,\mathrm{d}r\mathrm{d}s + \int_0^t \int_s^t \left(\pi_j p_{ji}^N(r - s) - \pi_i \pi_j \right) \,\mathrm{d}r\mathrm{d}s,$$

where we use the notation $p_{ij}^N(s) := \mathbb{P}(J^N(s) = j \,|\, J^N(0) = i)$. Performing the change of variable $u := rN^\alpha$ we thus find that

$$N^\alpha \mathbb{C}\mathrm{ov}(Z_i^N(t), Z_j^N(t)) = \pi_i \int_0^t \int_0^{sN^\alpha} (p_{ij}(u) - \pi_j) \,\mathrm{d}u\mathrm{d}s + \pi_j \int_0^t \int_0^{(t-s)N^\alpha} (p_{ji}(u) - \pi_i) \,\mathrm{d}u\mathrm{d}s.$$

A crucial role in the analysis is played by the *deviation matrix* $D = (D_{ij})_{i,j=1}^d$ associated with the finite-state Markov process $J(\cdot)$; it is defined by

$$(6) \qquad D_{ij} := \int_0^\infty (p_{ij}(t) - \pi_j)\mathrm{d}t;$$

see e.g. [11] for background and a survey of the main results on deviation matrices. Combining the above, we conclude that, as $N \to \infty$, with $C_{ij} := \pi_i D_{ij} + \pi_j D_{ji}$ we have identified that candidate covariance matrix, in the sense that we have shown that, for given t,

$$N^\alpha \, \mathbb{C}\mathrm{ov}(Z_i^N(t), Z_j^N(t)) \to C_{ij} t$$

as $N \to \infty$. The objective of the remainder of this section is to establish the weak convergence (5) with the covariance matrix $C = (C_{ij})_{i,j=1}^d$.

We now prove this weak convergence relying on the M-CLT. We start by considering linear combinations of the $X_i^N(t)$ processes based on Eqn. (2) and introduce $\tilde{X}_i^N(t) := X_i^N(t) - X_i^N(0)$ for

notational convenience. For any real constants f_i, $i = 1, \ldots, d$, we have that

$$\sum_{i=1}^{d} f_i \tilde{X}_i^N(t) = \sum_{i=1}^{d} \sum_{j \neq i} f_i \left(Y_{ji} \left(N^\alpha Z_j^N(t) q_{ji} \right) - Y_{ij} \left(N^\alpha Z_i^N(t) q_{ij} \right) \right)$$

$$(7) \qquad\qquad = \sum_{i=1}^{d} \sum_{j=1}^{d} (f_j - f_i) Y_{ij} \left(N^\alpha Z_i^N(t) q_{ij} \right),$$

where we do not need to define the processes Y_{ii} as the terms containing them are zero anyway. Note that the quadratic variation of this linear combination is equal to

$$(8) \qquad\qquad \left[\sum_{i=1}^{d} f_i \tilde{X}_i^N \right]_t = \sum_{i=1}^{d} \sum_{j=1}^{d} (f_j - f_i)^2 Y_{ij} (N^\alpha Z_i^N(t) q_{ij}),$$

as the quadratic variation of a Poisson process is equal to itself. Due to Lemma 2.1 and Eqn. (4), we have for $N \to \infty$,

$$(9) \qquad\qquad \left[N^{-\alpha/2} \sum_{i=1}^{d} f_i \tilde{X}_i^N \right]_t \to t \sum_{i=1}^{d} \sum_{j=1}^{d} (f_j - f_i)^2 \pi_i q_{ij}.$$

The crucial step in proving the weak convergence and consequently applying the M-CLT is the identification of a suitable martingale. We prove the following lemma.

Lemma 3.1. *Let D denote the deviation matrix of the background Markov chain $J(t)$. $\boldsymbol{V}^N(t) := N^{-\alpha/2} \tilde{\boldsymbol{X}}^N(t)^T D + N^{\alpha/2} \left(\boldsymbol{Z}^N(t) - \boldsymbol{\pi} t \right)$ is an $\mathbb{R}^d$-valued martingale.*

Proof. We center our unit-rate Poisson processes by introducing $\tilde{Y}_{i,j}(u) := Y_{i,j}(u) - u$. The following algebraic manipulations are easily verified:

$$N^{-\alpha/2} \sum_{i=1}^{d} \sum_{j=1}^{d} (D_{jk} - D_{ik}) \tilde{Y}_{ij} \left(N^\alpha Z_i^N(t) q_{ij} \right)$$

$$= N^{-\alpha/2} \sum_{i=1}^{d} \tilde{X}_i^N(t) D_{ik} - N^{\alpha/2} \sum_{i=1}^{d} \sum_{j=1}^{d} (D_{jk} - D_{ik}) Z_i^N(t) q_{ij}$$

$$= N^{-\alpha/2} \left(\tilde{\boldsymbol{X}}^N(t)^T D \right)_k - N^{\alpha/2} \sum_{i=1}^{d} \sum_{j=1}^{d} Z_i^N(t) q_{ij} D_{jk}$$

$$= N^{-\alpha/2} \left(\tilde{\boldsymbol{X}}^N(t)^T D \right)_k + N^{\alpha/2} (Z_k^N(t) - \pi_k t),$$

where we used Eqn. (7), the fact that $\sum_{i=1}^{d} Z_i^N(t) = t$ and the property $QD = \Pi - I$, with $\Pi = \boldsymbol{1} \boldsymbol{\pi}^T$. As centered Poisson processes are martingales, and linear combinations preserve the martingale property, this concludes the proof. $\square$

We now wish to apply the M-CLT to $\boldsymbol{V}^N(t)$ as $N \to \infty$. As the second term is absolutely continuous, and the first term is a jump process with jump sizes $N^{-\alpha/2}$, we have indeed vanishing jump sizes as

required by the M-CLT. We now compute the covariations of $\boldsymbol{V}^N(t)$, and note that as the second term is absolutely continuous and thus does not contribute to the covariation, we have that

$$
\begin{aligned}
[V_i^N, V_j^N]_t &= N^{-\alpha}[((\tilde{\boldsymbol{X}}^N)^T D)_i, ((\tilde{\boldsymbol{X}}^N)^T D)_j]_t \\
&= \frac{1}{2} N^{-\alpha} \left([((\tilde{\boldsymbol{X}}^N)^T D)_i + ((\tilde{\boldsymbol{X}}^N)^T D)_j]_t - [((\tilde{\boldsymbol{X}}^N)^T D)_i]_t - [((\tilde{\boldsymbol{X}}^N)^T D)_j]_t\right) \\
&= \frac{1}{2} N^{-\alpha} \left(\left[\sum_{k=1}^d \tilde{X}_k^N (D_{ki} + D_{kj})\right]_t - \left[\sum_{k=1}^d \tilde{X}_k^N D_{ki}\right]_t - \left[\sum_{k=1}^d \tilde{X}_k^N D_{kj}\right]_t\right),
\end{aligned}
$$

(10)

where we have used the polarization identity $2[X, Y]_t = [X + Y]_t - [X]_t - [Y]_t$.
Using Eqn. (9), this converges to

$$
\begin{aligned}
[V_i^N, V_j^N]_t &\to \frac{1}{2} t \sum_{k=1}^d \sum_{\ell=1}^d \pi_k q_{k\ell} \left((D_{ki} + D_{kj} - D_{\ell i} - D_{\ell j})^2 - (D_{ki} - D_{\ell i})^2 - (D_{kj} - D_{\ell j})^2\right) \\
&= t \sum_{k=1}^d \sum_{\ell=1}^d \pi_k q_{k\ell} (D_{ki} - D_{\ell i})(D_{kj} - D_{\ell j}) \\
&= t(\pi_j D_{ji} + \pi_i D_{ij})
\end{aligned}
$$

(11)

where we used the properties $Q\mathbf{1} = 0$, $\Pi Q = 0$, $QD = \Pi - I$ and $\Pi D = 0$.

Thus, from the M-CLT we have that $\boldsymbol{V}^N(t)$ converges weakly to d-dimensional Brownian motion with covariance matrix $C := D^T \text{diag}\{\boldsymbol{\pi}\} + \text{diag}\{\boldsymbol{\pi}\}D$. As the first term of $\boldsymbol{V}^N(t)$ vanishes for $N \to \infty$, we have established the desired F-CLT:

Proposition 3.2. *As $N \to \infty$,*

$$
N^{\alpha/2} \left(\boldsymbol{Z}^N(t) - \boldsymbol{\pi}t\right) \Rightarrow \boldsymbol{W}_C(t),
$$

where $\boldsymbol{W}_C(\cdot)$ is a zero-mean Gaussian process with independent increments and covariance structure $\mathbb{E}\left[\boldsymbol{W}_C(t)\boldsymbol{W}_C(t)^T\right] = C t$.

4. A FUNCTIONAL CLT FOR THE PROCESS $M^N(t)$

Using the F-CLT for the process $\boldsymbol{Z}^N(t)$, as established in the previous section, we are now in a position to understand the limiting behavior of the main process of interest, $M^N(t)$, as N grows large. As before, we begin by considering the average behavior of the quantity of interest. Dividing both sides of (3) by N, and denoting $\bar{M}^N(t) := N^{-1} M^N(t)$, we have

$$
\bar{M}^N(t) = \bar{M}^N(0) + N^{-1} Y_1 \left(N \sum_{i=1}^d \lambda_i Z_i^N(t)\right) - N^{-1} Y_2 \left(N\mu \int_0^t \bar{M}^N(s) ds\right).
$$

Assuming that $\bar{M}^N(0)$ converges almost surely to some value ϱ_0, the use of Lemma 2.1 in conjunction with Eqn. (4) yields that $\bar{M}^N(t)$ converges almost surely to the solution of the deterministic integral equation

(12)
$$
\varrho(t) = \varrho_0 + \left(\sum_{i=1}^d \lambda_i \pi_i\right) t - \mu \int_0^t \varrho(s) ds = \varrho_0 + \lambda_\infty t - \mu \int_0^t \varrho(s) ds,
$$

with $\lambda_\infty := \boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{\lambda}$. It is readily verified that this solution is given by

$$(13) \qquad \varrho(t) = \varrho_0 e^{-\mu t} + \frac{\lambda_\infty}{\mu}(1 - e^{-\mu t}).$$

As our goal is to derive a F-CLT, we center and scale the process $M^N(\cdot)$; this we do by subtracting $N\varrho(\cdot)$, and dividing by $N^{1-\beta}$ for some $\beta > 0$ to be specified later. More concretely, we introduce the process

$$U_\beta^N(t) := N^\beta\big(\bar{M}^N(t) - \varrho(t)\big).$$

Letting $\beta > 0$ be arbitrary (for the moment), we have that due to Eqn. (12),

$$\begin{aligned}
N^\beta\big(\bar{M}^N(t) - \varrho(t)\big) &= N^\beta\big(\bar{M}^N(0) - \varrho_0\big) - N^\beta(\varrho(t) - \varrho_0) \\
&\quad + N^\beta\left(N^{-1}Y_1\left(N\sum_{i=1}^d \lambda_i Z_i^N(t)\right) - N^{-1}Y_2\left(N\mu\int_0^t \bar{M}^N(s)\mathrm{d}s\right)\right) \\
\\
&= N^\beta(\bar{M}^N(0) - \varrho_0) - N^\beta\left(\lambda_\infty t - \mu\int_0^t \varrho(s)\mathrm{d}s\right) \\
&\quad + N^\beta\left(N^{-1}\tilde{Y}_1\left(N\sum_{i=1}^d \lambda_i Z_i^N(t)\right) - N^{-1}\tilde{Y}_2\left(N\mu\int_0^t \bar{M}^N(s)\mathrm{d}s\right)\right) \\
&\quad + N^\beta\sum_{i=1}^d \lambda_i Z_i^N(t) - N^\beta\mu\int_0^t \bar{M}^N(s)\mathrm{d}s.
\end{aligned}$$

This identity can be written in a more convenient form by defining the process

$$R_\beta(t) := \tilde{Y}_1\left(N\sum_{i=1}^d \lambda_i Z_i^N(t)\right) - \tilde{Y}_2\left(N\mu\int_0^t \bar{M}^N(s)\mathrm{d}s\right),$$

which is a martingale [1]. The resulting equation for $U_\beta^N(t)$ is

$$(14) \qquad U_\beta^N(t) = U_\beta^N(0) + N^{\beta-1}R_\beta(t) + N^\beta\left(\sum_{i=1}^d \lambda_i Z_i^N(t) - \lambda_\infty t\right) - \mu\int_0^t U_\beta^N(s)\mathrm{d}s.$$

We wish to establish the weak convergence of the process $U_\beta^N(t)$, as $N \to \infty$. We must simultaneously consider how to choose the parameter β. To do so we separately inspect the terms involving $R_\beta(t)$ and $Z_i^N(t)$ in Eqn. (14).

First note that the sequence of martingales $\{N^{\beta-1}R_\beta(t)\}$, for $N \in \mathbb{N}$, clearly satisfies the first condition of Thm. 2.2, that of vanishing jump sizes, under the condition that $\beta < 1$, which we impose from now on. To obtain its weak limit, we compute its quadratic variation

$$(15) \qquad \left[N^{\beta-1}R_\beta\right]_t = N^{2\beta-2}\left(Y_1\left(N\sum_{i=1}^d \lambda_i Z_i^N(t)\right) + Y_2\left(N\mu\int_0^t \bar{M}^N(s)\mathrm{d}s\right)\right).$$

For this term to converge in accordance with Thm. 2.2, we need $\beta \leq \frac{1}{2}$, which we impose from now on. With $\beta = \frac{1}{2}$, the term (15) will converge to $\lambda_\infty t + \mu\int_0^t \varrho(s)\mathrm{d}s$. Choosing $\beta < \frac{1}{2}$ will take it to

zero. Turning to the Z_i^N terms of (14), by Prop. 3.2, and recalling that $\lambda_\infty = \boldsymbol{\pi}^{\mathrm{T}} \boldsymbol{\lambda}$, we have that for $\beta = \alpha/2$,

$$(16) \qquad N^\beta \left(\sum_{i=1}^d \lambda_i Z_i^N(t) - \lambda_\infty t \right) \Rightarrow \boldsymbol{\lambda} \cdot \boldsymbol{W}_C(t),$$

which is distributionally equivalent to $W(\text{Þ}t)$, where W is a standard Brownian motion and

$$(17) \qquad \text{Þ} := \sum_{i=1}^d \sum_{j=1}^d \lambda_i \lambda_j C_{ij}.$$

If $\beta < \alpha/2$ the term on the left-hand side of (16) converges to zero. Combining the above leads us to select $\beta = \min\{\alpha/2, 1/2\}$. In the sequel we distinguish between $\alpha > 1$, $\alpha < 1$, and $\alpha = 1$.

Before we treat the three cases, we first recapitulate the class of Ornstein-Uhlenbeck (OU) processes. We say that $S(t)$ is $\mathrm{OU}(a, b, c)$ if it satisfies the stochastic differential equation (SDE)

$$\mathrm{d}S(t) = (a - b\,S(t))\mathrm{d}t + \sqrt{c}\,\mathrm{d}W(t),$$

with $W(t)$ a standard Brownian motion. This SDE is solved by

$$S(t) = S(0)e^{-bt} + a \int_0^t e^{-b(t-s)}\mathrm{d}s + \sqrt{c} \int_0^t e^{-b(t-s)}\mathrm{d}W(s).$$

By using standard stochastic calculus it can be verified that (taking $u \le t$)

$$(18) \qquad \begin{aligned} \mathbb{E}S(t) &= S(0)e^{-bt} + \frac{a}{b}(1 - e^{-bt}), \\ \mathbb{V}\mathrm{ar}\,S(t) &= \frac{c}{2b}(1 - e^{-2bt}), \\ \mathbb{C}\mathrm{ov}(S(t), S(u)) &= \frac{ce^{-bu}}{2b}\left(e^{bt} - e^{-bt}\right). \end{aligned}$$

For t large, we see that

$$\mathbb{E}S(\infty) = \frac{a}{b}, \quad \mathbb{V}\mathrm{ar}\,S(\infty) = \frac{c}{2b}, \quad \lim_{t \to \infty} \mathbb{C}\mathrm{ov}(S(t), S(t+u)) = \frac{c}{2b}e^{-bu}.$$

After this intermezzo, we now treat the three cases separately.

Case 1: $\alpha > 1$. In this case we pick $\beta = 1/2$. The term (15) converges to

$$\sum_{i=1}^d \lambda_i \pi_i t + \mu \int_0^t \varrho(s)\mathrm{d}s = \lambda_\infty t + \mu \int_0^t \varrho(s)\mathrm{d}s,$$

while the term (16) converges to zero and is therefore neglected. Hence, $U_{1/2}^N(t)$ converges in distribution to the solution of

$$U_{1/2}(t) = U_{1/2}(0) + W\left(\lambda_\infty t + \mu \int_0^t \varrho(s)\mathrm{d}s\right) - \mu \int_0^t U_{1/2}(s)\mathrm{d}s,$$

where W is a standard Brownian motion. The above solution is distributionally equivalent to the solution of the Itô formulation of the SDE

$$U_{1/2}(t) = U_{1/2}(0) + \int_0^t \sqrt{\lambda_\infty + \mu\varrho(s)}\,\mathrm{d}W(s) - \mu \int_0^t U_{1/2}(s)\mathrm{d}s.$$

This SDE can be solved using standard techniques to obtain

$$U_{1/2}(t) = e^{-\mu t} \left(U_{1/2}(0) + \int_0^t \sqrt{\lambda_\infty + \mu \varrho(s)} e^{\mu s} \mathrm{d}W(s) \right).$$

We now demonstrate how to compute the variance of $U_{1/2}(t)$. To this end, first recall that by virtue of (13),

$$\begin{aligned} \mathbb{V}\mathrm{ar}\, U_{1/2}(t) &= \int_0^t (\lambda_\infty + \mu \varrho(s)) e^{-2\mu(t-s)} \mathrm{d}s \\ &= \int_0^t \left(\lambda_\infty + \mu \left(\varrho_0 e^{-\mu s} + \frac{\lambda_\infty}{\mu}(1 - e^{-\mu s}) \right) \right) e^{-2\mu(t-s)} \mathrm{d}s. \end{aligned}$$

After routine calculations, this yields

$$\mathbb{V}\mathrm{ar}\, U_{1/2}(t) = \left(\varrho_0 e^{-\mu t} + \frac{\lambda_\infty}{\mu} \right)(1 - e^{-\mu t}),$$

cf. the expressions in [7, Section 4]. In a similar fashion, we can derive that

$$\mathbb{C}\mathrm{ov}(U_{1/2}(t), U_{1/2}(t+u)) = e^{-\mu u} \left(\varrho_0 e^{-\mu t} + \frac{\lambda_\infty}{\mu} \right)(1 - e^{-\mu t}).$$

It is seen that for $t \to \infty$ the limiting process behaves as $\mathrm{OU}(0, \mu, 2\lambda_\infty)$.

Case 2: $\alpha < 1$. In this case we pick $\beta = \alpha/2$ and the term (15), and therefore the term $N^{\beta-1}R_\beta(t)$, converges to zero, whereas the term (16) converges to $W(\text{Þ}t)$, where W is a standard Brownian motion and Þ as defined by Eqn. (17). Hence, $U_{\alpha/2}^N(t)$ converges weakly to the solution of

$$U_{\alpha/2}(t) = U_{\alpha/2}(0) + W(\text{Þ}t) - \mu \int_0^t U_{\alpha/2}(s)\mathrm{d}s,$$

It is straightforward to solve this equation:

$$U_{\alpha/2}(t) = e^{-\mu t} \left(U_{\alpha/2}(0) + \int_0^t \sqrt{\text{Þ}}\, e^{\mu s} \mathrm{d}W(s) \right).$$

This process has variance

$$(19) \qquad\qquad \mathbb{V}\mathrm{ar}\, U_{\alpha/2}(t) = \int_0^t \text{Þ} e^{-2\mu(t-s)} \mathrm{d}s = \text{Þ}\frac{1 - e^{-2\mu t}}{2\mu}.$$

It is readily checked that this process is $\mathrm{OU}(0, \mu, \text{Þ})$; this is due to

$$\mathbb{C}\mathrm{ov}(U_{\alpha/2}(t), U_{\alpha/2}(t+u)) = \text{Þ} e^{-\mu u} \frac{1 - e^{-2\mu t}}{2\mu}.$$

Case 3: $\alpha = 1$. In this case we put $\beta = 1/2$, and the terms $N^{\beta-1}R_\beta(t)$ and (16) are of the same order. Hence, their sum converges weakly to

$$W\left(\lambda_\infty t + \mu \int_0^t \varrho(s)ds + \text{Þ}t \right),$$

where W is a standard Browian motion. In this case $U_{1/2}^N(t)$ converges weakly to the solution of

$$U_{1/2}(t) = U_{1/2}(0) + W\left(\int_0^t (\lambda_\infty + \text{Þ} + \mu \varrho(s))\, \mathrm{d}s \right) - \mu \int_0^t U_{1/2}(s)\mathrm{d}s.$$

Solving the above in a similar fashion to cases 1 and 2 yields

$$U_{1/2}(t) = e^{-\mu t} \left(U_{1/2}(0) + \int_0^t \sqrt{\lambda_\infty + \text{Þ} + \mu \varrho(s)} \, e^{\mu s} \mathrm{d}W(s) \right).$$

with the corresponding variance

$$\mathbb{Var}\, U_{1/2}(t) = \int_0^t (\lambda_\infty + \text{Þ} + \mu \varrho(s)) \, e^{-2\mu(t-s)} \mathrm{d}s$$

$$= \left(\varrho_0 e^{-\mu t} + \frac{\lambda_\infty}{\mu} \right) (1 - e^{-\mu t}) + \frac{\text{Þ}}{2\mu} (1 - e^{-2\mu t})$$

and covariance

$$\mathbb{Cov}(U_{1/2}(t), U_{1/2}(t+u)) = e^{-\mu u} \left(\left(\varrho_0 e^{-\mu t} + \frac{\lambda_\infty}{\mu} \right) (1 - e^{-\mu t}) + \frac{\text{Þ}}{2\mu} (1 - e^{-2\mu t}) \right).$$

For t large this process behaves as $\text{OU}(0, \mu, 2\lambda_\infty + \text{Þ})$.

We summarize the above results in the following theorem; it is the F-CLT for $M^N(t)$ that we wished to establish. It identifies the Gauss-Markov process to which $M^N(\cdot)$ weakly converges, after centering and scaling; this limiting process behaves, modulo the effect of the initial value ϱ_0, as an OU process. More specifically, the theorem describes the limiting behavior of the centered and normalized version $U_\beta^N(\cdot)$ of $M^N(\cdot)$: the focus is on the process

$$(20) \qquad U_\beta^N(t) = N^\beta \left(\bar{M}^N(t) - \varrho(t) \right) = \frac{M^N(t) - N\varrho(t)}{N^{1-\beta}}.$$

It is observed that for $\alpha \geq 1$, we have the usual $\sqrt{N}$ CLT-scaling; for $\alpha < 1$, however, the normalizing polynomial is $N^{1-\alpha/2}$, that is $\beta = \min\{\alpha/2, 1/2\}$.

Theorem 4.1. *As $N \to \infty$, the process $U_\beta^N(t)$ converges in distribution to the solution of*

$$U_\beta(t) = e^{-\mu t} \left(U_\beta(0) + \int_0^t \sigma(s) e^{\mu s} \, \mathrm{d}W(s) \right)$$

where

$$\sigma(s) := \begin{cases} \sqrt{\lambda_\infty + \mu \varrho(s)}, & \alpha > 1, \ \beta = 1/2; \\ \sqrt{\text{Þ}}, & \alpha < 1, \ \beta = \alpha/2; \\ \sqrt{\lambda_\infty + \text{Þ} + \mu \varrho(s)}, & \alpha = 1, \ \beta = 1/2, \end{cases}$$

W is standard Brownian motion and $\text{Þ} = \sum_{i=1}^d \sum_{j=1}^d \lambda_i \lambda_j C_{ij}$.

5. DISCUSSION AND AN EXAMPLE

Above we identified two crucially different scaling regimes: $\alpha > 1$ and $\alpha < 1$ (where the boundary case of $\alpha = 1$ had to be dealt with separately). In case $\alpha > 1$, the background process evolves fast relative to the arrival process, and as a consequence the arrival stream is effectively Poisson with rate $N\lambda_\infty$. When the arrival process is simplified in such a way, the system essentially behaves as an M/M/∞ queue. This regime was discussed in greater detail in e.g. [7], focusing on convergence of

the finite-dimensional distributions. On the other hand, for $\alpha < 1$ the arrival rate is sped up more than the background process. Intuitively, then the system settles in a temporary (or local) equilibrium.

5.1. **A two-state example.**
In this example we numerically study the limiting behavior of $U_\beta^N(t) := N^\beta \left(\bar{M}^N(t) - \varrho(t) \right)$ with $\beta = \min\{\alpha/2, 1/2\}$ (as $N \to \infty$) in a two-state system for different α. For various values of N, we compute the moment generating function (MGF) of $U_\beta^N(t)$ by numerically evaluating the system of differential equations derived in [7, Section 3.1]. We have shown that the limiting distribution of $U_\beta^N(t)$ is Gaussian with specific parameters. We now explain how the MGF of the limiting random variable can be computed. Introducing the notation $\Lambda(t, \theta) := \mathbb{E}e^{\theta U_\beta(t)}$ and $\Lambda^N(t, \theta) := \mathbb{E}e^{\theta U_\beta^N(t)}$, it is immediate from Thm. 4.1 that we have

$$(21) \qquad \Lambda(t, \theta) = \exp\left(\frac{\theta^2}{2} \left[\left(\varrho_0 e^{-\mu t} + \frac{\lambda_\infty}{\mu} \right) (1 - e^{-\mu t}) 1_{\{\alpha \geq 1\}} + Þ\frac{(1 - e^{-2\mu t})}{2\mu} 1_{\{\alpha \leq 1\}} \right] \right),$$

In the regime that $\alpha \leq 1$, we need to evaluate the parameter $Þ$, which can be easily computed for $d = 2$. For the generator matrix $Q = (q_{ij})_{i,j=1}^2$, let $q_i := -q_{ii}$ and note that $q_i > 0$. With $\bar{q} := q_1 + q_2$, the matrix exponential is given by

$$e^{Qt} = \frac{1}{\bar{q}} \left[\begin{array}{cc} q_2 + q_1 e^{-\bar{q}t} & q_1 - q_1 e^{-\bar{q}t} \\ q_2 - q_2 e^{-\bar{q}t} & q_1 + q_2 e^{-\bar{q}t} \end{array} \right].$$

Since $\pi_1 = q_2/\bar{q}$ and $\pi_2 = q_1/\bar{q}$, we can now compute the components of the deviation matrix D (see Eqn. (6)) and covariance matrix C:

$$D = \frac{1}{\bar{q}^2} \left[\begin{array}{cc} q_1 & q_1 \\ -q_2 & q_2 \end{array} \right] \quad , \quad C = \frac{2q_1 q_2}{\bar{q}^3} \left[\begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right].$$

From the above we find the value of $Þ$:

$$Þ = \frac{2q_1 q_2}{\bar{q}^3}(\lambda_1 - \lambda_2)^2.$$

Fig. 5.1 illustrates the convergence of $U_\beta^N(t)$ to $U_\beta(t)$ when $\boldsymbol{q} = (1, 3)$, $\boldsymbol{\lambda} = (1, 4)$ and $\mu = 1$. We assume $\varrho_0 = 0$ and let $\theta = 0.5$. In Figs. 5.1.(A)–(C) we see the effect of α. Fig. 5.1.(D) depicts the convergence rate, computed as

$$\max_{t \geq 0} \left| \Lambda^N(t, \theta) - \Lambda(t, \theta) \right|.$$

We observe a roughly loglinear convergence speed for $\alpha \geq 1$, whereas for $\alpha < 1$ the convergence turns out to be substantially slower.

REFERENCES

[1] D. Anderson and T. Kurtz. Continuous-time Markov chain models for chemical reaction networks. In H. Koeppl et al, editor, *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*, pages 3–42. Springer, New York, 2011.

[2] A. Arazia, E. Ben-Jacob, and U. Yechiali. Bridging genetic networks and queueing theory. *Physica A*, 332:585–616, 2004.

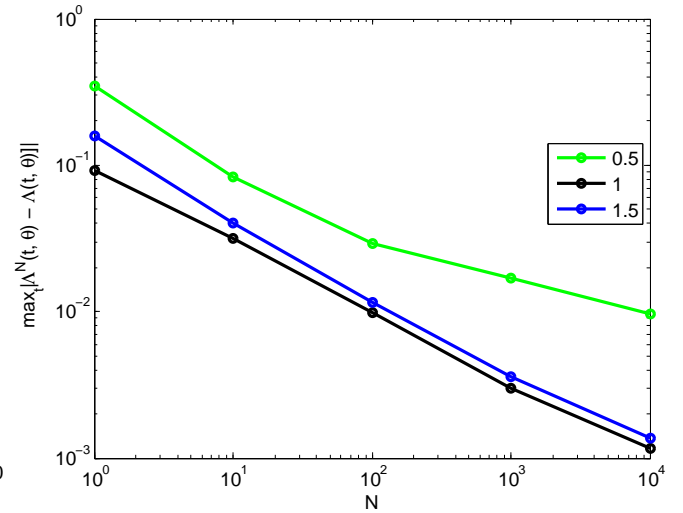[3] S. Asmussen. *Applied Probability and Queues, 2nd ed.* Springer, New York, 2003.
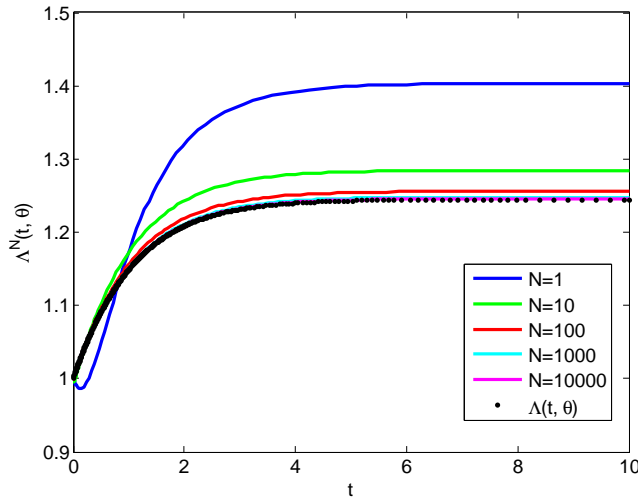
(A) $\alpha = 1.5$
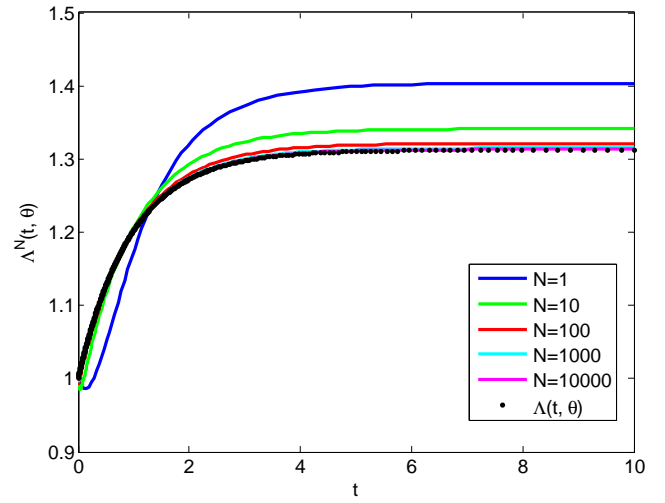
(B) $\alpha = 1$
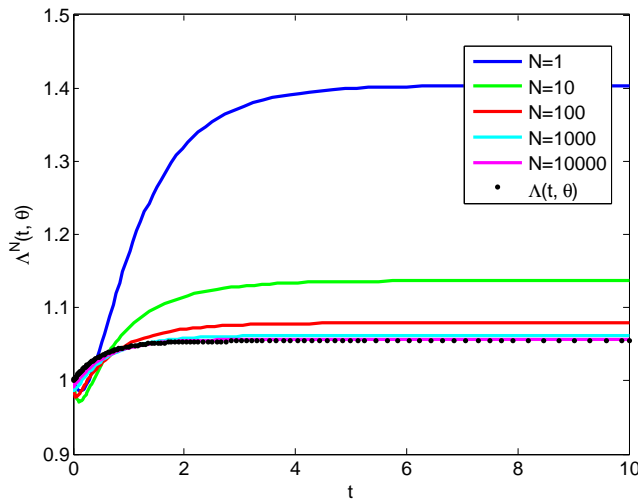
(C) $\alpha = 0.5$

(D) Error with $\alpha \in \{0.5, 1, 1.5\}$.

FIGURE 5.1. Convergence of the MGF of $U_\beta^N(t)$ as a function of $t \geq 0$. Panels (A)-(C) correspond to three values of α; panel (D) depicts the convergence rate as a function of N.

[4] K. Ball, T. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multi-scale approximations to reaction networks. *Ann. Appl. Probab*, 16:1925–1961, 2006.

[5] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 60:185–201, 1982.

[6] J. Blom, K. de Turck, and M. Mandjes. A central limit theorem for Markov-modulated infinite-server queues. *In: Proceedings ASMTA 2013, Ghent, Belgium. Lecture Notes in Computer Science (LNCS) Series*, 7984:81–95, 2013.

[7] J. Blom, O. Kella, M. Mandjes, and H. Thorsdottir. Markov-modulated infinite-server queues with general service times. *To appear in Queueing Systems (available online)*, 2013.

[8] J. Blom, M. Mandjes, and H. Thorsdottir. Time-scaling limits for markov-modulated infinite-server queues. *Stoch. Models*, 29:112–127, 2013.

[9] A. Borovkov. On limit laws for service processes in multi-channel systems. *Siberian Math. J*, 8:746–763, 1967.

[10] N. Cookson, W. Mather, T. Danino, O. Mondragón-Palomino, R. Williams, L. Tsimring, and J. Hasty. Queueing up for enzymatic processing: Correlated signaling through coupled degradation. *Mol. Syst. Biol*, 7(561), 2011.

[11] P. Coolen-Schrijner and E. van Doorn. The deviation matrix of a continuous-time Markov chain. *Probab. Engrg. Inform. Sci*, 16:351–366, 2002.

[12] B. D'Auria. M/M/∞ queues in semi-markovian random environment. *Queueing Syst*, 58:221–237, 2008.

[13] S. Eick, W. Massey, and W. Whitt. The physics of the M_t/G/∞ queue. *Oper. Res*, 41:731–742, 1993.

[14] S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 1986.

[15] B. Fralix and I. Adan. An infinite-server queue influenced by a semi-Markovian environment. *Queueing Syst*, 61:65–84, 2009.

[16] D. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem*, 58:35–55, 2007.

[17] D. Iglehart. Limiting diffusion approximations for the many server queue and the repairman problem. *J. Appl. Probab*, 2:429–441, 1965.

[18] J. Jacod and A. Shiryayev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 1987.

[19] H. Kang and T. Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab*, 23:529–583, 2013.

[20] J. Keilson and L. Servi. The matrix M/M/∞ system: retrial models and Markov modulated sources. *Adv. Appl. Probab*, 25:453–471, 1993.

[21] T. Kurtz and P. Protter. Wong-zakai corrections, random evolutions, and simulation schemes for SDEs. In E. Merzbach E. Mayer-Wolf and A. Schwartz, editors, *Stochastic Analysis*, pages 331–346. Academic Press, 1991.

[22] C. O'Cinneide and P. Purdue. The M/M/∞ queue in a random environment. *J. Appl. Probab*, 23:175–184, 1986.

[23] R. Rebolledo. Central limit theorems for local martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 51:269–286, 1980.

[24] Ph. Robert. *Stochastic Networks and Queues*. Springer, Berlin, 2003.

[25] A. Schwabe, K. Rybakova, and F. Bruggeman. Transcription stochasticity of complex gene regulation models. *Biophysical Journal*, 103:1152–1161, 2012.

[26] W. Whitt. Proofs of the martingale FCLT. *Probab. Surveys*, 4:268–302, 2007.