



## UvA-DARE (Digital Academic Repository)

### Parameter estimation for multivariate population processes

*a saddlepoint approach*

de Gunst, M.; Hautphenne, S.; Mandjes, M.; Sollié, B.

**DOI**

[10.1080/15326349.2020.1832895](https://doi.org/10.1080/15326349.2020.1832895)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

Stochastic Models

**License**

CC BY-NC-ND

[Link to publication](#)

**Citation for published version (APA):**

de Gunst, M., Hautphenne, S., Mandjes, M., & Sollié, B. (2021). Parameter estimation for multivariate population processes: a saddlepoint approach. *Stochastic Models*, 37(1), Article 168-196. <https://doi.org/10.1080/15326349.2020.1832895>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Parameter estimation for multivariate population processes: a saddlepoint approach

Mathisca de Gunst<sup>a</sup>, Sophie Hautphenne<sup>b</sup>, Michel Mandjes<sup>c</sup>, and Birgit Sollie<sup>a</sup>

<sup>a</sup>Mathematics, VU University Amsterdam, Amsterdam, Netherlands; <sup>b</sup>School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria Melbourne, Australia; <sup>c</sup>University of Amsterdam Korteweg-de Vries Institute for Mathematics, Amsterdam, Netherlands

## ABSTRACT

The setting considered in this paper concerns a discrete-time multivariate population process under Markov modulation. Our objective is to estimate the model parameters, based on periodic observations of the network population vector. These parameters relate to the arrival, routing and departure processes, but also to the (unobservable) Markovian background process. When opting for the classical likelihood-based approach, the evaluation of the likelihood is problematic. We show however, how an accurate saddlepoint approximation can be used. Numerical experiments illustrate our method and show that even under relatively complicated conditions the parameters are estimated relatively precisely.

## ARTICLE HISTORY

Received 28 November 2019  
Accepted 2 October 2020

## KEYWORDS

Population process;  
stochastic networks;  
saddlepoint approximation;  
Markov-modulation; maximum likelihood estimation

## 1. Introduction

Population processes are stochastic processes that record the dynamics of the number of individuals in a population. Owing to their widespread use in for instance biology, ecology, and chemical reaction networks, they have become a key object of study in statistics and applied probability. In its simplest form a population process describes the fluctuations of the population size at a single location. Many practically relevant situations, however, correspond to considerably more general settings. In the first place, the population process often lives on a *multi-node* (rather than single-node) network. This means that individuals can enter and leave the nodes of the

**CONTACT** Birgit Sollie  [birgit.corporaal@gmail.com](mailto:birgit.corporaal@gmail.com)  Mathematics, VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, Netherlands

*Mathisca de Gunst* and *Birgit Sollie* are with the Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, the Netherlands. *Sophie Hautphenne* is with the School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3010, Australia. She thanks the Australian Research Council for support through Discovery Early Career Researcher Award DE150101044. *Michel Mandjes* is with the Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands. His research is partly funded by the NWO Gravitation project Networks, grant number 024.002.003.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

network, but in addition they can move between its nodes. Secondly, in many situations the dynamics of the population are affected by exogenous, often unobservable, factors; think of temperature affecting the spread of bacteria or weather conditions affecting the mobility of the individuals. In these cases it is desirable to add an underlying modulating process to the model, referred to as the *background process*.

Due to the ubiquity of multivariate modulated population processes across a wide range of scientific disciplines, there is a clear need for sound statistical techniques to estimate the underlying parameters. In this paper we devise such a method based on observations of the network population vector. We do so in a discrete-time context, with the background process corresponding to a finite state-space Markov chain. This means that we are in the context of *Markov modulation*, with the values of the parameters pertaining to the arrival, routing, and departure processes being a function of the state of the background process.

In the setting considered, parameter estimation can be seen as a highly challenging inverse problem. When developing an estimation procedure, one needs to cope with two major intrinsic complications.

- In the first place, as we have access to the network population vector only, we do not observe the number of arrivals, the number of individuals that are routed between each of the node pairs, and the number of departures, but only the *net effect* of these processes. This effectively means that in general we cannot trace how individuals have moved through the network.
- The second complication is that we assume that we cannot observe the background process (making its state a hidden variable). The challenge is to infer from the observations the parameters of the Markovian background process, and the (background-state dependent) parameters pertaining to the arrival, routing, and departure processes.

There is a considerable body of work on inverse problems for continuous-time population processes. In the first place we refer to for example<sup>[1–5]</sup> for parameter estimation procedures for univariate birth-death processes without modulation. In these papers the case is considered where the population is observed at discrete times only, hence the individual births and deaths are not observed directly. In addition there are various papers on estimation techniques for infinite-server queues (which can be seen as population processes in which the times the individuals spend in the system are independent of each other) without modulation. In this context we mention,<sup>[6]</sup> in which the service-time distribution is estimated without direct observations of the service times, and,<sup>[7]</sup> which treats the estimation of the arrival rate and the service-time distribution from

observations of the population size. A separate branch of the literature focuses on parameter estimation for stochastic processes with a Markovian, unobserved background process. In this respect we mention,<sup>[8,9]</sup> which concentrate on the class of Markovian binary trees and continuous-time observations or demographic data. In addition, when focusing on a Markovian arrival process only, rather than the resulting population process, in<sup>[10,11]</sup> estimation procedures based on discrete-time observations are presented. We finally mention,<sup>[12]</sup> in which a parameter estimation procedure for a univariate population process under Markov modulation is proposed and assessed, based on discrete-time observations of the population size.

The work presented in this paper concerns parameter estimation for a multivariate population process, and can as such be seen as part of the broader area of network science. There is a strong relation with the discipline that focuses on the statistical analysis of network data. We refer to,<sup>[13 Chapters 8 and 9]</sup> for more background on statistical procedures for stochastic processes on networks. It is noted, though, that existing theory predominantly concentrates on situations in which the routing process on the network—often referred to as the network flow—is fully observed, which contrasts with the situation considered in the present paper.

Importantly, to the best of our knowledge, there are no procedures available for estimating the parameters of modulated multivariate population process, based on observations of the network population vector. One could pursue an approach based on maximum likelihood, but evaluating the likelihood is generally problematic. The main difficulty lies in the complexity of the model, in terms of the size of the underlying network and the fact that there is a modulating background process. As a consequence, typically no closed-form expression for the likelihood can be given; in addition, in the special cases where it is possible to obtain such an explicit expression, there are often numerical complications. We therefore take another approach, which combines the following two ideas:

- Due to the structure of the model, it is possible to set up a procedure to compute for each point in time the joint moment generating function (mgf) pertaining to the network population vector.
- We then apply the technique of *saddlepoint approximation* to compute an approximation of the likelihood, and maximize this approximation over the unknown parameters. The saddlepoint approximation provides a (typically highly accurate) approximation of the probability mass function of a random vector, based on the corresponding joint mgf.

The saddlepoint technique has been developed in the 1950s by Daniels,<sup>[14]</sup> for a textbook treatment see e.g.,<sup>[15]</sup> For specific models

closed-form expressions for saddlepoint approximations have been obtained. In this respect we refer to<sup>[16]</sup> for explicit approximations of the transition densities and cumulative distribution functions of Markov processes, whereas in<sup>[17]</sup> general birth processes are considered. The references<sup>[18,19]</sup> provide extensive general accounts of the use of saddlepoint techniques in statistics. A few papers where saddlepoint expansions have been used to approximate the likelihood are<sup>[20]</sup> which considers the context of the INAR( $p$ ) model,<sup>[21]</sup> where the focus is on the distribution of the sum of independent non-identically distributed binomial random variables, and<sup>[3]</sup> which aims at estimating the birth and death rates of a linear birth-and-death process.

We proceed by discussing our paper's main contributions in more detail. First and foremost, to our best knowledge, we are the first to develop a parameter estimation procedure in the highly general and comprehensive setup of a multivariate population process under Markov modulation, based on periodic observations of the network population vector. Our approach is likelihood-based, but only in special (small) networks the likelihood can be computed in closed form, which is why we approximate the likelihood relying on the saddlepoint approximation. A prerequisite for using the saddlepoint technique is the availability of the mgf corresponding to the network population vector at multiple points in time. We present an efficient technique to evaluate this mgf, by computing the mgf of the network population vector at one observation time conditionally on the population vector at the previous observation time. Then this mgf is used to approximate the likelihood, which numerically boils down to solving a convex optimization problem. Subsequently the approximated likelihood is maximized over the parameter space to find approximate values for the maximum likelihood estimates of the model parameters. The last contribution concerns numerical experiments, which assess the performance of our parameter estimation technique. They show that even under relatively complicated conditions (modulation, a multi-node system), following our approach, the parameters can be estimated relatively precisely. The examples involve single- and multi-node networks, with and without modulation, and illustrate the factors that affect the procedure's performance.

The remainder of this paper is organized as follows. In [Section 2](#) we formally define the multivariate population process under Markov modulation, and we state the estimation problem. [Section 3](#) focuses on two examples of small networks (a single-node model and a tandem network of two nodes), showing how in these cases the likelihood can be computed explicitly. This section also points out how the expressions for the likelihood become increasingly involved if the number of network nodes increases. In [Section 4](#) we show how the likelihood can be evaluated using

saddlepoint approximations; this section also includes the method to compute the mgf of the network population vector. We show how the approximation of the likelihood can be used to estimate the model parameters, and investigate the accuracy of this estimation method by numerical studies in Section 5. We conclude the paper with a discussion in Section 6.

## 2. Model and estimation

As mentioned in the introduction, this paper considers a population process on a network with finitely many nodes. Individuals can arrive at each of the nodes, follow a probabilistic route through the network, and potentially leave the network. We impose Markov modulation: all parameters in the model are driven by a discrete-time Markov chain, where each state corresponds to a different set of parameter values. In this section, we first present a detailed mathematical description of our Markov modulated multivariate population process, and then state the corresponding parameter estimation problem.

We throughout adopt the convention that vectors are printed in bold; we denote by  $x(k)$  the  $k$ -th entry of the vector  $\mathbf{x}$ . As usual, random variables and matrices are denoted by capital letters. We use  $\langle \mathbf{x}, \mathbf{y} \rangle$  to denote the inner product of  $\mathbf{x}$  and  $\mathbf{y}$  (whose dimensions are then assumed to be compatible). We write  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ .

### 2.1. The model

We start by introducing the *background process*  $\{X_k\}_{k \in \mathbb{N}_0}$ . This is an irreducible discrete-time Markov chain with finite state space  $E = \{1, \dots, d\}$ ,  $d \in \{2, 3, \dots\}$ . We define by  $P = (p_{ij})_{i,j=1}^d$  the corresponding  $(d \times d)$  transition probability matrix,  $\boldsymbol{\alpha}$  the corresponding initial state distribution (i.e.,  $\alpha_i := \boldsymbol{\alpha}(i) = \mathbb{P}(X_0 = i)$ ), and  $\boldsymbol{\pi}$  the (unique) stationary distribution. Recall that  $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$ . The background process modulates the network's dynamics in a way we make precise below.

We study a network with  $L \in \mathbb{N}$  nodes on which we define the multivariate population process  $\{\mathbf{M}_k\}_{k \in \mathbb{N}_0}$ , where the vector  $\mathbf{M}_k$  records the number of individuals present at the  $L$  nodes at time  $k$ . This population process is the result of an arrival process, a routing mechanism by which individuals jump between the nodes, and a departure process. We now introduce these individual ingredients.

- Denote by  $\{\mathbf{A}_k\}_{k \in \mathbb{N}}$  the arrival process, where  $\mathbf{A}_k \in \mathbb{N}^L$  represents a vector that counts the number of arrivals at each of the  $L$  nodes at time  $k$ . We assume that these arrivals stem from a parametric class, where the parameters depend on the value of  $X_{k-1}$ , i.e., the state of the

background process at time  $k - 1$ ; the arrival process is thus Markov modulated. More precisely, given  $X_{k-1} = i$ , for some  $i \in E$ , the moment generating function (mgf) of the arrivals at time  $k$  is assumed to exist and given by (for  $\mathbf{s} \in \mathbb{R}^L$ )

$$\phi_{k,i}(\mathbf{s}) := \mathbb{E}[e^{\langle \mathbf{s}, \mathbf{A}_k \rangle} | X_{k-1} = i], \quad (1)$$

with the corresponding cumulant generating function (cgf) denoted by  $\psi_{k,i}(\mathbf{s}) := \log \phi_{k,i}(\mathbf{s})$ . In the sequel, we let the individual components of  $\mathbf{A}_k$  be time-homogeneous and independent, and let  $\mathbf{A}_k(\ell)$  have a Poisson distribution with parameter  $\lambda_i^{(\ell)} \geq 0$ , given  $X_{k-1} = i$ . In this case

$$\psi_{k,i}(\mathbf{s}) \equiv \psi_i(\mathbf{s}) = \sum_{\ell=1}^L \lambda_i^{(\ell)} (e^{s^{(\ell)}} - 1). \quad (2)$$

We emphasize that the use of other choices of the arrival process is straightforward, as long as the mgf defined in (1) exists and is known.

- The routing and departure processes are Markov modulated as well. To describe these processes, we first define for each  $\ell \in \{1, \dots, L\}$  the vector-valued process  $\{\mathbf{D}_k^{(\ell)}\}_{k \in \mathbb{N}}$ , where  $\mathbf{D}_k^{(\ell)} \in \mathbb{N}^{L+1}$ . For  $\ell'$  between 1 and  $L$ ,  $\mathbf{D}_k^{(\ell)}(\ell')$  counts the number of individual jumps out of node  $\ell$  toward node  $\ell'$  at time  $k$ , whereas the  $\mathbf{D}_k^{(\ell)}(L+1)$  records the number of individuals that leave the network from node  $\ell$  at time  $k$ . We say that a jump from a node to itself is the same as staying at the node. Importantly, in our model all individuals can move independently of each other through the network and do not have to wait for each other. Let  $r_i^{(\ell, \ell')} \in [0, 1]$  be the probability of an individual at node  $\ell$  to jump to node  $\ell'$  at an arbitrary time point when the background state is  $i$ . Also,

$$r_i^{(\ell, 0)} := 1 - \sum_{\ell'=1}^L r_i^{(\ell, \ell')}$$

(which is a number in  $[0, 1]$ ) denotes the probability of an individual to leave the network from node  $\ell$  at any time point at which the background state is  $i$ . If  $r_i^{(\ell, 0)} = 0$ , individuals cannot leave the network from node  $\ell$  when the background process is in state  $i$ . Note that for each  $k > 0$ , given  $\mathbf{M}_{k-1}$  and  $X_{k-1}$ , the vectors  $\mathbf{D}_k^{(\ell)}$  are independent. In addition, for a given  $\ell$  the vector  $\mathbf{D}_k^{(\ell)}$  follows a multinomial distribution.

In our model we let the change of the background process happen *after* the arrivals, the routing and the departures. We remark, however, that this choice does not impose any restriction: in the very same manner we can deal with the analogous model in which the background process jumps before the arrivals, the routing and the departures.

Furthermore, both the routing and the departures occur *before* the arrivals, which implies that newly arrived individuals can only leave the node the next timeslot at the earliest. It is also possible to assume that the arrivals occur before the departures and routing. This leads to a slightly different model, in which individuals who leave the system in the same interval as they arrive are included in both the arrival process and departure process, although they are not visible in the population process  $\{M_k\}$ .

We proceed by introducing various quantities related to  $\{M_k\}$  that play a crucial role in our analysis. In the sequel we will work intensively with the mgf of  $M_k$  given  $M_{k-1}$  and  $X_{k-1} = i$  (with  $k \in \mathbb{N}$  and  $i \in E$ ): for  $\mathbf{s} \in \mathbb{R}^L$ ,

$$\begin{aligned} \zeta_i(\mathbf{s} \mid \mathbf{m}) &:= \mathbb{E}[e^{\langle \mathbf{s}, M_k \rangle} \mid M_{k-1} = \mathbf{m}, X_{k-1} = i] \\ &= \mathbb{E}[e^{\langle \mathbf{s}, M_1 \rangle} \mid M_0 = \mathbf{m}, X_0 = i], \end{aligned}$$

with the corresponding cgf  $\zeta_i(\mathbf{s} \mid \mathbf{m}) := \log \zeta_i(\mathbf{s} \mid \mathbf{m})$ . Furthermore, we define for all observation pairs  $\mathbf{m}, \mathbf{m}' \in \mathbb{N}_0^L$ ,  $k \in \mathbb{N}$  and  $i \in E$  the one-step transition probabilities

$$t_i(\mathbf{m}' \mid \mathbf{m}) = \mathbb{P}(M_k = \mathbf{m}' \mid M_{k-1} = \mathbf{m}, X_{k-1} = i) = \mathbb{P}(M_1 = \mathbf{m}' \mid M_0 = \mathbf{m}, X_0 = i),$$

and the diagonal matrix

$$T(\mathbf{m}' \mid \mathbf{m}) = \text{diag}\{t_1(\mathbf{m}' \mid \mathbf{m}), \dots, t_d(\mathbf{m}' \mid \mathbf{m})\}. \quad (3)$$

Note that  $\zeta_i(\mathbf{s} \mid \mathbf{m})$  and  $t_i(\mathbf{m}' \mid \mathbf{m})$  do not depend on  $k$  due to time-homogeneity.

## 2.2. Parameter estimation

The objective of this paper is to estimate the model parameters from observations of the population process. We now specify these unknown parameters and the available data.

Throughout we assume that the network population process  $\{M_k\}$  can be observed at time points  $k = 0, 1, \dots, n$  for some  $n \in \mathbb{N}$ . We denote the corresponding observations by  $\mathbf{m}_0, \mathbf{m}_1, \dots, \mathbf{m}_n$ , so that the set  $\{\mathbf{m}_0, \dots, \mathbf{m}_n\} \in \mathbb{N}_0^{L \times (n+1)}$  comprises the available data.

Let

$$\theta = \left( \alpha_i, p_{ij}, \lambda_i^{(\ell)}, r_i^{(\ell, \ell')} : i, j \in \{1, \dots, d\}, \ell \in \{1, \dots, L\}, \ell' \in \{0, \dots, L\} \right)^\top$$

be the unknown parameter vector corresponding to the model. Our goal is to estimate  $\theta$  given the observation  $\mathbf{m}_0, \dots, \mathbf{m}_n$ . The resulting estimate will be denoted by

$$\hat{\theta} = \left( \hat{\alpha}_i, \hat{p}_{ij}, \hat{\lambda}_i^{(\ell)}, \hat{r}_i^{(\ell, \ell')} : i, j \in \{1, \dots, d\}, \ell \in \{1, \dots, L\}, \ell' \in \{0, \dots, L\} \right)^\top.$$

We estimate  $\theta$  by maximum likelihood, which requires the evaluation of the likelihood function. We make the common assumption that  $\mathbb{P}(\mathbf{M}_0 = \mathbf{m}_0) = 1$ . By taking into account all possible paths of the background process  $\{X_k\}$  (at times  $k = 0, \dots, n-1$ ), and using (3), the likelihood function can then be written as

$$\begin{aligned} \mathcal{L}(\theta \mid \mathbf{m}_0, \dots, \mathbf{m}_n) &= \mathbb{P}_\theta(\mathbf{M}_0 = \mathbf{m}_0, \dots, \mathbf{M}_n = \mathbf{m}_n) \\ &= \sum_{x_0, \dots, x_{n-1} \in E} \mathbb{P}_\theta(\mathbf{M}_0 = \mathbf{m}_0, X_0 = x_0, \dots, \mathbf{M}_{n-1} \\ &= \mathbf{m}_{n-1}, X_{n-1} = x_{n-1}, \mathbf{M}_n = \mathbf{m}_n) \\ &= \boldsymbol{\alpha}^\top T(\mathbf{m}_1 \mid \mathbf{m}_0) P T(\mathbf{m}_2 \mid \mathbf{m}_1) P \cdots P T(\mathbf{m}_n \mid \mathbf{m}_{n-1}) \mathbf{1}, \end{aligned} \quad (4)$$

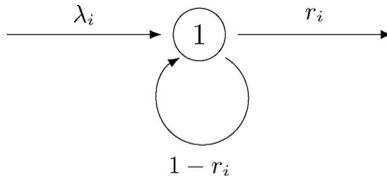
where  $\mathbf{1} = (1, \dots, 1)^\top$ . We conclude that, in order to compute the likelihood  $\mathcal{L}(\theta \mid \mathbf{m}_0, \dots, \mathbf{m}_n)$ , it is a prerequisite to be able to evaluate, for any pair of vectors  $\mathbf{m}'$  and  $\mathbf{m}$  and for any  $i \in E$ , the probability  $t_i(\mathbf{m}' \mid \mathbf{m})$ .

### 3. Small networks: explicit approach

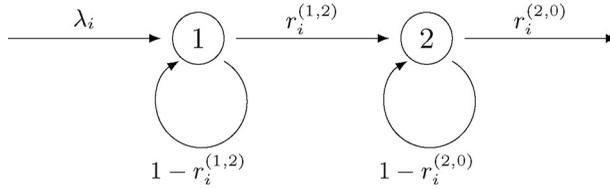
In this section we present a few examples of ‘small’ networks in which the one-step probabilities  $t_i(\mathbf{m}' \mid \mathbf{m})$  can be computed explicitly. We first consider the special case of a single-node model with Poisson arrivals, also known as a Markov-modulated infinite-server queue, and then treat a specific two-node tandem network. For larger networks, transitions from  $\mathbf{m}$  to  $\mathbf{m}'$  could correspond to a large number of potential scenarios (in terms of the numbers of individuals arriving, being routed to another node, and departing), making explicit evaluation prohibitive.

#### 3.1. Single-node model

Consider a model with a single node at which individuals arrive according to the arrival process  $\{A_k\}_{k \in \mathbb{N}}$ , which is now a univariate random variable. More precisely,  $A_k \in \mathbb{N}_0$  is the number of arrivals in the  $k$ -th timeslot. Let, as before,  $\{X_k\}_{k \in \mathbb{N}_0}$  be a Markovian background process with  $d$  states. We assume that for each state  $i \in E$ ,  $A_k$  given  $X_{k-1} = i$  has a Poisson distribution with parameter  $\lambda_i \geq 0$ , and individuals can either leave the node with probability  $r_i \in [0, 1]$ , or stay at the node with probability  $1 - r_i$  (see Figure 1). Let the process  $\{D_k\}_{k \in \mathbb{N}}$  count the number of individuals that



**Figure 1.** Schematic representation of the single-node model.



**Figure 2.** Schematic representation of the tandem network.

leave the node per timeslot, whereas  $\{M_k\}_{k \in \mathbb{N}_0}$  keeps track of the population size at the node. The idea is to compute  $t_i(m' | m)$ , by conditioning on the number of departing individuals at time  $k = 1$ . It follows that

$$\begin{aligned}
 t_i(m' | m) &= \sum_{\check{m}=0}^m \mathbb{P}(M_k = m' | D_k = \check{m}, M_{k-1} = m, X_{k-1} = i) \\
 &\quad \mathbb{P}(D_k = \check{m} | M_{k-1} = m, X_{k-1} = i) \\
 &= \sum_{\check{m}=\max\{0, m-m'\}}^m \frac{(\lambda_i)^{m'-(m-\check{m})}}{(m' - (m - \check{m}))!} e^{-\lambda_i} \binom{m}{\check{m}} (r_i)^{\check{m}} (1-r_i)^{m-\check{m}}.
 \end{aligned}$$

**3.2. Tandem network**

We now consider a tandem model with two nodes, in which individuals arrive at the first node, then either jump to the second node or stay at the first node, and from the second node either leave the system or stay at the second node. We again have a Markovian background process  $\{X_k\}_{k \in \mathbb{N}_0}$  modulating the parameters in the model. We assume that individuals arrive at the first node according to the arrival process  $\{A_k\}_{k \in \mathbb{N}}$ , where  $A_k$  given  $X_{k-1} = i$  is Poisson distributed with parameter  $\lambda_i \geq 0$ . Recall that  $D_k^{(1)}(2)$  represents the number of individuals jumping from the first to the second node. Given the state of the background process being  $i \in E$ , each individual makes this jump with probability  $r_i^{(1,2)} \in [0, 1]$ , or stays at the first node with probability  $1 - r_i^{(1,2)}$ . From the second node, individuals leave the network with probability  $r_i^{(2,0)} \in [0, 1]$ , or stay at the node with probability  $1 - r_i^{(2,0)}$  (see Figure 2).

We can compute  $t_i(\mathbf{m}' \mid \mathbf{m})$  for this model, by conditioning on the number of individuals that jump from the first node to the second node. After some elementary algebra we find

$$\begin{aligned}
 & t_i(\mathbf{m}' \mid \mathbf{m}) \\
 &= \sum_{\check{m}=0}^{m(1)} \mathbb{P}(\mathbf{M}_k = \mathbf{m}' \mid \mathbf{D}_k^{(1)}(2) = \check{m}, \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i) \\
 & \mathbb{P}(\mathbf{D}_k^{(1)}(2) = \check{m} \mid \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i) \\
 &= \sum_{\check{m}=\check{m}_{\text{low}}}^{\check{m}_{\text{up}}} \frac{(\lambda_i)^a}{a!} e^{-\lambda_i} \binom{\mathbf{m}(2)}{b} (r_i^{(2,0)})^b (1-r_i^{(2,0)})^{m(2)-\check{m}} \\
 & \binom{\mathbf{m}(1)}{\check{m}} (r_i^{(1,2)})^{\check{m}} (1-r_i^{(1,2)})^{m(1)-\check{m}}.
 \end{aligned} \tag{5}$$

Here  $\check{m}_{\text{low}} := \max\{0, \mathbf{m}(1) - \mathbf{m}'(1), \mathbf{m}'(2) - \mathbf{m}(2)\}$ ,  $\check{m}_{\text{up}} := \min\{\mathbf{m}(1), \mathbf{m}'(2)\}$  are the lower and upper bounds of the sum, respectively. In addition,  $a := \mathbf{m}'(1) - \mathbf{m}(1) + \check{m}$  denotes the number of arrivals to the first node, and  $b := \mathbf{m}(2) - \mathbf{m}'(2) + \check{m}$  the number of departures from the second node.

In the above two examples we observe that one can develop explicit expressions for  $t_i(\mathbf{m}' \mid \mathbf{m})$ , but already in the example of the two-node tandem the expression becomes quite involved. When trying to extend our expressions to tandems with more nodes, or even to more general networks, the expressions will become increasingly complex as the dimension of the underlying network grows. As pointed out in e.g.<sup>[21]</sup> the computation effectively requires a complete enumeration over all possible configurations, which makes this explicit approach infeasible for larger networks. A solution to this problem for such networks is to, instead of pursuing exact calculation of  $t_i(\mathbf{m}' \mid \mathbf{m})$ , resort to its *saddlepoint approximation*. We detail this procedure in the next section.

#### 4. General networks: saddlepoint approximation

The main objective of this section is to set up an accurate and computationally efficient approximation for the probabilities  $t_i(\mathbf{m}' \mid \mathbf{m})$ . As pointed out in Section 3, for multi-node models it is typically infeasible to evaluate  $t_i(\mathbf{m}' \mid \mathbf{m})$  explicitly, which motivates the need for such approximate techniques. We rely on the saddlepoint approach,<sup>[14,19]</sup> which approximates a random variable's probability mass function through its mgf. In Section 4.1 we point out in detail how this technique works. A complication is that the saddlepoint machinery does not work for states at

the boundary of the state space of  $\{M_k\}$ . For such points an alternative computation scheme is proposed in Section 4.2, which is a combination of the saddlepoint approximation with exact computations. Examples that assess the procedure’s numerical performance are provided in Section 4.3.

**4.1. Interior states: saddlepoint approach**

Aiming at applying the saddlepoint approach to approximate  $t_i(\mathbf{m}' \mid \mathbf{m})$ , we need to be able to evaluate the mgf  $\xi_i(\mathbf{s} \mid \mathbf{m})$ , where we recall the notation

$$\xi_i(\mathbf{s} \mid \mathbf{m}) = \mathbb{E}\left[e^{\langle \mathbf{s}, M_k \rangle} \mid M_{k-1} = \mathbf{m}, X_{k-1} = i\right].$$

The corresponding cgf is denoted by  $\zeta_i(\mathbf{s} \mid \mathbf{m}) := \log \xi_i(\mathbf{s} \mid \mathbf{m})$ . In order to evaluate  $\zeta_i(\mathbf{s} \mid \mathbf{m})$ , observe that the  $\ell$ -th component of  $M_k$  is equal to

- the number  $M_{k-1}(\ell)$  that was present at node  $\ell$  at time  $k - 1$ ,
- decreased by the number of individuals that leave node  $\ell$  at time  $k$  (either by jumping to another node or by leaving the network),
- increased by external arrivals at node  $\ell$  at time  $k$ , and
- increased by the number of individuals that were at node  $\check{\ell}$  at time  $k - 1$  and jump to node  $\ell$  at time  $k$ , over all  $\check{\ell} \in \{1, \dots, L\}$ .

Recall that  $D_k^{(\ell)}(L + 1)$  represents the number of individuals that leave the network from node  $\ell$  at time  $k$ . Summarizing the above, the following identity links  $M_k$  and  $M_{k-1}$  :

$$M_k(\ell) = M_{k-1}(\ell) - \sum_{\check{\ell}=1}^{L+1} D_k^{(\ell)}(\check{\ell}) + A_k(\ell) + \sum_{\check{\ell}=1}^L D_k^{(\check{\ell})}(\ell). \tag{6}$$

For ease of notation, both sums in (6) contain the variable  $D_k^{(\ell)}(\ell)$  corresponding to  $\check{\ell} = \ell$ , counting the number of individuals that stay at node  $\ell$ . Recall that conditionally on  $M_{k-1} = \mathbf{m}$  and  $X_{k-1} = i$ , the vectors  $D_k^{(\ell)}$  are independent, and that for a given  $\ell$  the entries of  $D_k^{(\ell)}$  have a multinomial distribution. Due to these properties and using (6), we find

$$\begin{aligned} \xi_i(\mathbf{s} \mid \mathbf{m}) &= e^{\langle \mathbf{s}, \mathbf{m} \rangle} \phi_i(\mathbf{s}) \mathbb{E}\left[\exp\left(\sum_{\ell=1}^L \mathbf{s}(\ell) \left(\sum_{\check{\ell}=1}^L D_k^{(\ell)}(\ell) - \sum_{\check{\ell}=1}^{L+1} D_k^{(\ell)}(\check{\ell})\right)\right) \mid M_{k-1} = \mathbf{m}, X_{k-1} = i\right] \\ &= e^{\langle \mathbf{s}, \mathbf{m} \rangle} \phi_i(\mathbf{s}) \prod_{\check{\ell}=1}^L \mathbb{E}\left[\exp\left(\sum_{\ell=1}^L \mathbf{s}(\ell) D_k^{(\ell)}(\ell) - \sum_{\ell=1}^{L+1} \mathbf{s}(\check{\ell}) D_k^{(\check{\ell})}(\ell)\right) \mid M_{k-1} = \mathbf{m}, X_{k-1} = i\right]. \end{aligned} \tag{7}$$

To obtain (7), we have used a change of summation in the first term of the exponent, a change of variables in the second term of the exponent,

and the fact that the  $\mathbf{D}_k^{(\ell)}(\check{\ell})$  are independent in  $\ell$ . Continuing from (7), reordering the terms in the exponent, we conclude that we have

$$\begin{aligned} \xi_i(\mathbf{s} \mid \mathbf{m}) &= e^{\langle \mathbf{s}, \mathbf{m} \rangle} \phi_i(\mathbf{s}) \prod_{\check{\ell}=1}^L \mathbb{E} \left[ \exp \left( \sum_{\ell=1}^L (\mathbf{s}(\ell) - \mathbf{s}(\check{\ell})) \mathbf{D}_k^{(\check{\ell})}(\ell) - \mathbf{s}(\check{\ell}) \mathbf{D}_k^{(\check{\ell})}(L+1) \right) \middle| \mathcal{M}_{k-1} \right. \\ &= \left. \mathbf{m}, X_{k-1} = i \right]. \end{aligned}$$

Finally using the multinomial property, we arrive at the following result.

**Lemma 1.** For  $\mathbf{s} \in \mathbb{R}^L$  and  $\mathbf{m} \in \mathbb{N}_0^L$ , and for any  $i \in E$ ,

$$\begin{aligned} \xi_i(\mathbf{s} \mid \mathbf{m}) &= e^{\langle \mathbf{s}, \mathbf{m} \rangle} \phi_i(\mathbf{s}) \prod_{\check{\ell}=1}^L \left( \sum_{\ell=1}^L r_i^{(\check{\ell}, \ell)} e^{\mathbf{s}(\ell) - \mathbf{s}(\check{\ell})} + r_i^{(\check{\ell}, 0)} e^{-\mathbf{s}(\check{\ell})} \right)^{\mathbf{m}(\check{\ell})} \\ &= \phi_i(\mathbf{s}) \prod_{\check{\ell}=1}^L \left( \sum_{\ell=1}^L r_i^{(\check{\ell}, \ell)} e^{\mathbf{s}(\ell)} + r_i^{(\check{\ell}, 0)} \right)^{\mathbf{m}(\check{\ell})}. \end{aligned} \quad (8)$$

Having the expression for  $\xi_i(\mathbf{s} \mid \mathbf{m})$  at our disposal, we now point out how this can be used in the saddlepoint-based approximation of  $t_i(\mathbf{m}' \mid \mathbf{m})$ . To this end, we first note that by taking logarithms on both sides of Equation (8), we obtain

$$\zeta_i(\mathbf{s} \mid \mathbf{m}) = \psi_i(\mathbf{s}) + \sum_{\check{\ell}=1}^L \mathbf{m}(\check{\ell}) \log \left( \sum_{\ell=1}^L r_i^{(\check{\ell}, \ell)} e^{\mathbf{s}(\ell)} + r_i^{(\check{\ell}, 0)} \right). \quad (9)$$

It is known that any (joint) cgf is a convex function, which implies that  $\zeta_i(\mathbf{s} \mid \mathbf{m})$  is convex (in  $\mathbf{s}$ ). Define for  $\mathbf{v}, \mathbf{m} \in \mathbb{N}_0^L$ , the corresponding multivariate Legendre-Fenchel transforms by

$$I_i(\mathbf{v} \mid \mathbf{m}) := \sup_{\mathbf{s}} I_i(\mathbf{v}, \mathbf{s} \mid \mathbf{m}),$$

where  $I_i(\mathbf{v}, \mathbf{s} \mid \mathbf{m}) := \langle \mathbf{s}, \mathbf{v} \rangle - \zeta_i(\mathbf{s} \mid \mathbf{m})$ .

Let  $S_i(\mathbf{m}) \subseteq \mathbb{N}_0^L$  the set of states that can be reached from  $\mathbf{m}$  in one time step when the background state is  $i$ . More concretely,

$$S_i(\mathbf{m}) = \left\{ \mathbf{m}' : \mathbf{m}'(\ell) = \sum_{\ell'=1}^L k_{\ell', \ell} + k_{\ell}, \quad \left( (k_{\ell', \ell})_{\ell', \ell=1}^L, (k_{\ell})_{\ell=1}^L \right) \in K_i(\mathbf{m}) \right\},$$

where  $K_i(\mathbf{m})$  is the subset of  $\mathbb{N}_0^{L^2 \times L}$  consisting of  $((k_{\ell', \ell})_{\ell', \ell=1}^L, (k_{\ell})_{\ell=1}^L)$  such that

- A. For all  $\ell = 1, \dots, L$ ,  $\sum_{\ell'=1}^L k_{\ell', \ell} \mathbf{m}(\ell)$  (i.e., the sum of individuals leaving node  $\ell$  cannot be more than  $\mathbf{m}(\ell)$ );

- B. For all  $\ell = 1, \dots, L$ ,  $\sum_{\ell'=1}^L k_{\ell, \ell'} = \mathbf{m}(\ell)$  if  $r_i^{(\ell, 0)} = 0$  (i.e., the sum of individuals jumping from node  $\ell$  to the other nodes must be exactly  $\mathbf{m}(\ell)$  if  $r_i^{(\ell, 0)} = 0$ );
- C. For all  $\ell = 1, \dots, L$ ,  $k_{\ell', \ell} = 0$  if  $r_i^{(\ell', \ell)} = 0$ , and  $k_{\ell} = 0$  if  $\lambda_i^{(\ell)} = 0$  (i.e., jumps and arrivals cannot occur if the corresponding parameter equals zero).

We denote by  $S_i(\mathbf{m})^\circ$  the ‘interior’ of  $S_i(\mathbf{m})$ , to be understood as  $S_i(\mathbf{m})$  minus its boundaries.

For any  $\mathbf{v} \in S_i(\mathbf{m})^\circ$  there is a unique optimizing vector  $\mathbf{s}_v^*$  for which  $I_i(\mathbf{v}, \mathbf{s}_v^* \mid \mathbf{m}) = I_i(\mathbf{v} \mid \mathbf{m})$ , which is called the saddlepoint; see,<sup>[15]</sup> Chapter 1] for more details. By the definition of  $I_i(\mathbf{v}, \mathbf{s} \mid \mathbf{m})$ , this saddlepoint is the unique solution of the system of  $L$  first-order conditions

$$\mathbf{v}(\ell') = \frac{\partial \psi_i(\mathbf{s})}{\partial \mathbf{s}(\ell')} + \sum_{\check{\ell}=1}^L \mathbf{m}(\check{\ell}) r_i^{(\check{\ell}, \ell')} e^{\mathbf{s}(\ell')} \bigg/ \left( \sum_{\check{\ell}=1}^L r_i^{(\check{\ell}, \ell')} e^{\mathbf{s}(\ell')} + r_i^{(\check{\ell}, 0)} \right), \quad (10)$$

where the right hand side of (10) is the  $\ell'$ -th entry of the gradient of the cgf, that is the vector of first partial derivatives with respect to the entries of  $\mathbf{s}$ . Let  $\Sigma_i(\mathbf{v} \mid \mathbf{m})$  be the  $L \times L$  Hessian matrix of the cgf evaluated at the saddlepoint with  $(\ell', \ell'')$ -th entry given by

$$\Sigma_i^{(\ell', \ell'')}(\mathbf{v} \mid \mathbf{m}) = \frac{\partial^2 \zeta_i(\mathbf{s} \mid \mathbf{m})}{\partial \mathbf{s}(\ell') \partial \mathbf{s}(\ell'')} \bigg|_{\mathbf{s}=\mathbf{s}_v^*}.$$

Note that by taking another partial derivative of the right hand side of (10), we find for  $\ell' \neq \ell''$

$$\frac{\partial^2 \zeta_i(\mathbf{s} \mid \mathbf{m})}{\partial \mathbf{s}(\ell') \partial \mathbf{s}(\ell'')} = \frac{\partial^2 \psi_i(\mathbf{s})}{\partial \mathbf{s}(\ell') \partial \mathbf{s}(\ell'')} + \sum_{\check{\ell}=1}^L \mathbf{m}(\check{\ell}) \frac{-r_i^{(\check{\ell}, \ell')} e^{\mathbf{s}(\ell')} r_i^{(\check{\ell}, \ell'')} e^{\mathbf{s}(\ell'')}}{\left( \sum_{\check{\ell}=1}^L r_i^{(\check{\ell}, \ell')} e^{\mathbf{s}(\ell')} + r_i^{(\check{\ell}, 0)} \right)^2},$$

while for  $\ell' = \ell''$

$$\frac{\partial^2 \zeta_i(\mathbf{s} \mid \mathbf{m})}{\partial \mathbf{s}(\ell') \partial \mathbf{s}(\ell'')} = \frac{\partial^2 \psi_i(\mathbf{s})}{\partial \mathbf{s}^2(\ell')} + \sum_{\check{\ell}=1}^L \mathbf{m}(\check{\ell}) \frac{r_i^{(\check{\ell}, \ell')} e^{\mathbf{s}(\ell')} \left( \sum_{\ell \neq \check{\ell}} r_i^{(\check{\ell}, \ell)} e^{\mathbf{s}(\ell)} + r_i^{(\check{\ell}, 0)} \right)}{\left( \sum_{\check{\ell}=1}^L r_i^{(\check{\ell}, \ell')} e^{\mathbf{s}(\ell')} + r_i^{(\check{\ell}, 0)} \right)^2}.$$

We can now present the saddlepoint approximation.<sup>[14,15]</sup> In the statement below,  $|D|$  denotes the determinant of the matrix  $D$ .

**Approximation 1.** For  $\mathbf{m} \in \mathbb{N}_0^L$ ,  $\mathbf{m}' \in S(\mathbf{m})^\circ$ , and for any  $i \in E$ , the saddlepoint approximation of  $t_i(\mathbf{m}' \mid \mathbf{m})$  is given by

$$t_i(\mathbf{m}' \mid \mathbf{m}) \approx (2\pi)^{-L/2} | \Sigma_i(\mathbf{m}' \mid \mathbf{m}) |^{-1/2} \exp(-I_i(\mathbf{m}' \mid \mathbf{m})). \quad (11)$$

Observe that the complexity of evaluating this approximation is relatively low. More specifically, to evaluate  $t_i(\mathbf{m}' \mid \mathbf{m})$  the maximization of an  $L$ -dimensional concave function needs to be performed and the determinant of a  $(L \times L)$ -matrix needs to be computed. To evaluate the full (diagonal) matrix  $T(\mathbf{m}' \mid \mathbf{m})$ , this has to be done  $d$  times. The computation of the likelihood  $\mathcal{L}(\theta \mid \mathbf{m}_0, \dots, \mathbf{m}_n)$  then takes  $2n$  matrix multiplications, with matrices of size  $d \times d$ , where  $n$  of these multiplications can be done relatively efficiently as they involve a diagonal matrix.

**Remark 1.** In the model considered, individuals jump between nodes until they leave the network. Interestingly, a ‘branching variant’ of this model, in which there is the option of a single individual splitting into multiple individuals, can also be dealt with. This variant is also referred to as a multi-type branching process with immigration in a random environment; see<sup>[22,23]</sup> for an analysis of its limiting distribution. In this case, when an individual moves from  $\ell$  to  $\ell'$  (with the background process being in state  $i$ ), the number of individuals that end up at  $\ell'$  is not necessarily 1, but is distributed as a random variable  $W_i^{(\ell, \ell')} \in \mathbb{N}_0$  with mgf  $w_i^{(\ell, \ell')}(s)$  (assumed to exist). Then for  $\mathbf{s} \in \mathbb{R}^L$  and  $\mathbf{m} \in \mathbb{N}_0^L$ , and for any  $i \in E$ , the mgf  $\xi_i(\mathbf{s} \mid \mathbf{m})$  becomes

$$\xi_i(\mathbf{s} \mid \mathbf{m}) = \phi_i(\mathbf{s}) \prod_{\ell'=1}^L \left( \sum_{\ell=1}^L r_i^{(\ell, \ell')} w_i^{(\ell, \ell')}(s(\ell)) + r_i^{(\ell, 0)} \right)^{m(\ell')}.$$

Observe that the resulting network is not necessarily stable; we do not further comment on the stability condition of this model. In another variant that can be dealt with, each individual that leaves  $\ell$  can potentially cause arrivals at *all* nodes simultaneously, rather than at just one node.

#### 4.2. States at the boundaries

Above we introduced an approximation for  $t_i(\mathbf{m}' \mid \mathbf{m})$  with  $\mathbf{m}' \in S_i(\mathbf{m})^\circ$ , which leaves us with the question what should be done for the ‘boundary points’  $\mathbf{m}' \in S_i(\mathbf{m}) \setminus S_i(\mathbf{m})^\circ$ . In the first place we recall (see,<sup>[15]</sup> Chapter 1) that for these points the saddlepoint approximation cannot be used, as a consequence of the fact that the optimizing  $\mathbf{s}_{\mathbf{m}'}$  cannot be determined. To show how we remedy this, we first use the illustrative examples of the single-node model and the tandem network featured in [Section 3](#). As we will observe, in these cases the transition probabilities can be found explicitly for the boundary states. Later in this subsection we will set up a general (exact) procedure to compute the transition probabilities for boundary states.

- For the single-node model,  $S_i(m) = \mathbb{N}_0$ , and thus  $S_i(m) \setminus S_i(m)^\circ = \{0\}$ . Now consider  $m' = 0$ . For this boundary point an easy explicit expression for  $t_i(m' \mid m)$  can be given. We have the explicit expression

$$t_i(0 \mid m) = e^{-\lambda_i} (r_i)^m,$$

since there should be no new arrivals, and all individuals that were present at the node have to leave.

- We continue by considering the tandem network. A first observation is that for this network there are multiple boundary points to take into account. There are no external arrivals at the second node, because this node is only fed by individuals moving from the first to the second node. As a consequence, we have

$$S_i(\mathbf{m}) = \{ \mathbf{m}' \in \mathbb{N}_0^2 : \max\{0, \mathbf{m}(1) - \mathbf{m}'(1)\} \leq \mathbf{m}'(2) \leq \mathbf{m}(1) + \mathbf{m}(2) \}.$$

To verify this, note that the maximum number of individuals at the second node at time  $k$  cannot be larger than the total network population at time  $k - 1$ , and the minimum number of individuals cannot be smaller than the minimum inflow from node 1.

Now consider a boundary point  $\mathbf{m}'$  in  $S(\mathbf{m}) \setminus S(\mathbf{m})^\circ$ . The claim is that again for all these boundary points an easy explicit expression for  $t_i(\mathbf{m}' \mid \mathbf{m})$  can be given. It is for example readily checked that, in self-evident notation,

$$t_i((\mathbf{m}'(1), \mathbf{m}(1) + \mathbf{m}(2))^\top \mid \mathbf{m}) = \mathbb{P}(\text{Pois}(\lambda_i) = \mathbf{m}'(1)) \left( r_i^{(1,2)} \right)^{\mathbf{m}(1)} \left( 1 - r_i^{(2,0)} \right)^{\mathbf{m}(2)}.$$

Notice that this probability corresponds to a scenario in which all individuals present at node 1 have to move to node 2, and all those present at node 2 have to stay. Importantly, in this case the complicated combinatorial expression (5) reduces to a considerably easier expression, essentially due to the fact that at boundary points the transition corresponds to a very specific scenario.

With the above examples in mind, let us go back to the general network setting with  $L$  nodes. For ease we restrict ourselves to the situation where  $\lambda_i^{(\ell)} > 0$  and  $r_i^{(\ell,0)} > 0$  for all  $i \in \{1, \dots, d\}$  and all  $\ell \in \{1, \dots, L\}$ . This means that at each node external arrivals and departures are possible for all states of the background process. The immediate consequence is that

$$S_i(\mathbf{m}) \setminus S_i(\mathbf{m})^\circ = \{ \mathbf{m}' : \exists \ell \in \{1, \dots, L\} : \mathbf{m}'(\ell) = 0 \}.$$

The situation in which some of the  $\lambda_i^{(\ell)}$  and  $r_i^{(\ell,0)}$  are 0 requires a bit more administration, but can be handled similarly (as in the above tandem example). Now fix an  $\mathbf{m}' \in S_i(\mathbf{m}) \setminus S_i(\mathbf{m})^\circ$ , a boundary point. We define

by  $N(\mathbf{m}')$  all nodes of the new configuration  $\mathbf{m}'$  that contain zero individuals, i.e.,  $N(\mathbf{m}') := \{\ell \in \{1, \dots, L\} : \mathbf{m}'(\ell) = 0\}$ . Let  $E(\mathbf{m}')$  be the corresponding event defined as  $E(\mathbf{m}') := \{\forall \ell \in N(\mathbf{m}'), \mathbf{M}_k(\ell) = 0\}$ . Then, because  $\{\mathbf{M}_k = \mathbf{m}'\} \subseteq E(\mathbf{m}')$ , and using elementary rules for conditional probabilities,

$$\begin{aligned} t_i(\mathbf{m}' \mid \mathbf{m}) &= \mathbb{P}(\mathbf{M}_k = \mathbf{m}' \mid \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i) \\ &= \mathbb{P}(\mathbf{M}_k = \mathbf{m}' \mid E(\mathbf{m}'), \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i) \\ &\quad \times \mathbb{P}(E(\mathbf{m}') \mid \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i). \end{aligned} \quad (12)$$

For the boundary points  $\mathbf{m}'$ ,  $t_i(\mathbf{m}' \mid \mathbf{m})$  can be (approximately) evaluated by evaluating the two factors in (12) separately. As we will see, the second factor can be computed exactly, whereas for the first one we can set up a saddlepoint approximation in the way demonstrated in Section 4.1.

To evaluate the second factor in (12), we observe that (i) at time 1, no arrivals are allowed in the nodes of  $N(\mathbf{m}')$ , and (ii) individuals present at the nodes in  $\{1, \dots, L\}$  at time 0 should either leave the network or move to a node in the complement of  $N(\mathbf{m}')$ . More specifically, they cannot move to, or stay in, a node in  $N(\mathbf{m}')$ . As a consequence, we have the exact expression

$$\mathbb{P}(E(\mathbf{m}') \mid \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i) = \prod_{\ell \in N(\mathbf{m}')} e^{-\lambda_i^{(\ell)}} \cdot \prod_{\ell=1}^L \left( \sum_{\ell' \notin N(\mathbf{m}')} r_i^{(\ell, \ell')} + r_i^{(\ell, 0)} \right)^{m(\ell)}. \quad (13)$$

We now concentrate on the first factor in (12), which can be computed using a saddlepoint approximation. To this end, we first observe that the occurrence of the event  $E(\mathbf{m}')$  (i.e.,  $\mathbf{M}_k(\ell) = 0$  for all  $\ell \in N(\mathbf{m}')$ ) changes the distribution of the random vectors  $\mathbf{A}_k$  and  $\mathbf{D}_k$ ; in the sequel we denote the random vectors under this condition by  $\tilde{\mathbf{A}}_k$  and  $\tilde{\mathbf{D}}_k$ . To describe the distribution of  $\tilde{\mathbf{A}}_k$  and  $\tilde{\mathbf{D}}_k$ , we use the following ‘renormalized’ probabilities, for  $\ell'' \notin N(\mathbf{m}')$ :

$$\tilde{r}_i^{(\ell, \ell'')} = \frac{r_i^{(\ell, \ell'')}}{\sum_{\ell'' \notin N(\mathbf{m}')} r_i^{(\ell, \ell'')} + r_i^{(\ell, 0)}}, \quad \tilde{r}_i^{(\ell, 0)} = \frac{r_i^{(\ell, 0)}}{\sum_{\ell'' \notin N(\mathbf{m}')} r_i^{(\ell, \ell'')} + r_i^{(\ell, 0)}}.$$

We then make the following observations.

- Since we have independent Markov-modulated Poisson arrivals at each of the nodes, the components of  $\tilde{\mathbf{A}}_k$  are independent, with  $\tilde{\mathbf{A}}_k(\ell)$  having a Poisson distribution with parameter  $\lambda_i^{(\ell)}$  for all  $\ell \notin N(\mathbf{m}')$ , whereas  $\tilde{\mathbf{A}}_k(\ell) \equiv 0$  for all  $\ell \in N(\mathbf{m}')$ . Recall that no arrivals are allowed in the nodes of  $N(\mathbf{m}')$  due to the condition imposed.

- The random vectors  $\tilde{\mathbf{D}}_k^{(\ell')}$ , for  $\ell' = 1, \dots, L$ , are independent. More specifically  $\tilde{\mathbf{D}}_k^{(\ell')}$  has a multinomial distribution that attains values in the complement of  $N(\mathbf{m}')$  or  $\{L + 1\}$ , where the latter option corresponds to leaving the network, with its parameters being given by  $\mathbf{m}^{(\ell')}$  and the probabilities

$$\left( (\tilde{r}_i^{(\ell', \ell'')})_{\ell'' \notin N(\mathbf{m}'), \tilde{r}_i^{(\ell', 0)}} \right).$$

Recall that individuals present at any of the nodes should either leave the network or move to (or stay at) a node in the complement of  $N(\mathbf{m}')$ .

Similar to (6), conditionally on  $E(\mathbf{m}')$ , we thus have the representation

$$\mathbf{M}_k(\ell) = \mathbf{M}_{k-1}(\ell) + \tilde{\mathbf{A}}_k(\ell) + \sum_{\check{\ell}=1}^L \tilde{\mathbf{D}}_k^{(\check{\ell})}(\ell) - \sum_{\check{\ell}=1}^{L+1} \tilde{\mathbf{D}}_k^{(\check{\ell})}(\check{\ell}).$$

We can now proceed as in Section 4.1, to obtain the mgf of  $\mathbf{M}_k$ , conditionally on the event  $\{E(\mathbf{m}'), \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i\}$ . Using the above findings, we find that it equals, with  $\mathbf{s}$  now being a vector with zeroes at the positions that correspond to the elements in  $N(\mathbf{m}')$ ,

$$\mathbb{E} \left[ e^{\langle \mathbf{s}, \mathbf{M}_k \rangle} \mid E(\mathbf{m}'), \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i \right] = \tilde{\phi}_i(\mathbf{s}) \prod_{\check{\ell}=1}^L \left( \sum_{\ell \notin N(\mathbf{m}')} \tilde{r}_i^{(\check{\ell}, \ell)} e^{\mathbf{s}(\ell)} + \tilde{r}_i^{(\check{\ell}, 0)} \right)^{\mathbf{m}^{(\check{\ell})}},$$

where

$$\tilde{\phi}_i(\mathbf{s}) = \prod_{\ell \notin N(\mathbf{m}')} e^{\lambda_i^{(\ell)} (e^{\mathbf{s}(\ell)} - 1)}.$$

Observe in particular the similarity with the result stated in Lemma 1. Using this mgf, we can use a saddlepoint technique to approximate  $\mathbb{P}(\mathbf{M}_k = \mathbf{m}' \mid E(\mathbf{m}'), \mathbf{M}_{k-1} = \mathbf{m}, X_{k-1} = i)$  in (12) by following the same argument as in Section 4.1, evidently only including the non-zero elements of  $\mathbf{m}'$ . We observe that the dimension of this saddlepoint approximation is now  $L - \{N(\mathbf{m}')\}$ , which is smaller than  $L$  as a consequence of  $\mathbf{m}' \in S_i(\mathbf{m}) \setminus S_i(\mathbf{m})^\circ$ .

In summary, according to (12) the probability  $t_i(\mathbf{m}' \mid \mathbf{m})$  can be factorized into two probabilities. The probability corresponding to the nodes included in  $N(\mathbf{m}')$  can be computed explicitly according to (13), whereas the probability corresponding to the remaining nodes can be evaluated relying on the saddlepoint approximation of reduced dimension.

### 4.3. Numerical assessment of approximations

We can illustrate the accuracy of the saddlepoint approximation for the single-node model and the tandem network, by comparing the explicit approach from Section 3 with the saddlepoint approach from Section 4.

**Example 1.** *Single-node model.* Consider the example of the single-node model introduced in Section 3.1, where we computed  $t_i(m' | m)$  explicitly. In this example  $S_i(m) = \mathbb{N}$ , assuming that  $\lambda_i > 0$  and  $r_i > 0$ , so that  $S_i(m) \setminus S_i(m)^\circ = \{0\}$ . Using the saddlepoint approach we can, for  $m' \in S_i(m)^\circ$ , approximate  $t_i(m' | m)$  using Approximation 1. Recall from (2) that for the Poisson arrivals at the node we have  $\psi_i(s) = \lambda_i(e^s - 1)$ . Using (9) with  $r_i^{(1,1)} = 1 - r_i$  and  $r_i^{(1,0)} = r_i$ , we find the cgf

$$\zeta_i(s | m) = \lambda_i(e^s - 1) + m \log((1 - r_i)e^s + r_i).$$

It requires a few standard steps to verify that the saddlepoint  $s_v^*$  can be found by solving

$$v = w(s) := \lambda_i e^s + m \frac{(1 - r_i)e^s}{(1 - r_i)e^s + r_i}; \quad (14)$$

observe that the right-hand side of (14) is a positive, increasing function in  $s$ , with  $w(s) \rightarrow 0$  as  $s \rightarrow -\infty$  and  $w(s) \rightarrow \infty$  as  $s \rightarrow \infty$ . This means that for any  $v > 0$ , there is a unique solution  $s_v^*$ . More concretely,  $e^{s_v^*}$  can be found in a standard manner by solving the quadratic equation

$$-\lambda_i(1 - r_i)e^{2s} + (v(1 - r_i) - \lambda_i r_i - m(1 - r_i))e^s + vr_i = 0.$$

In our one-dimensional context we have

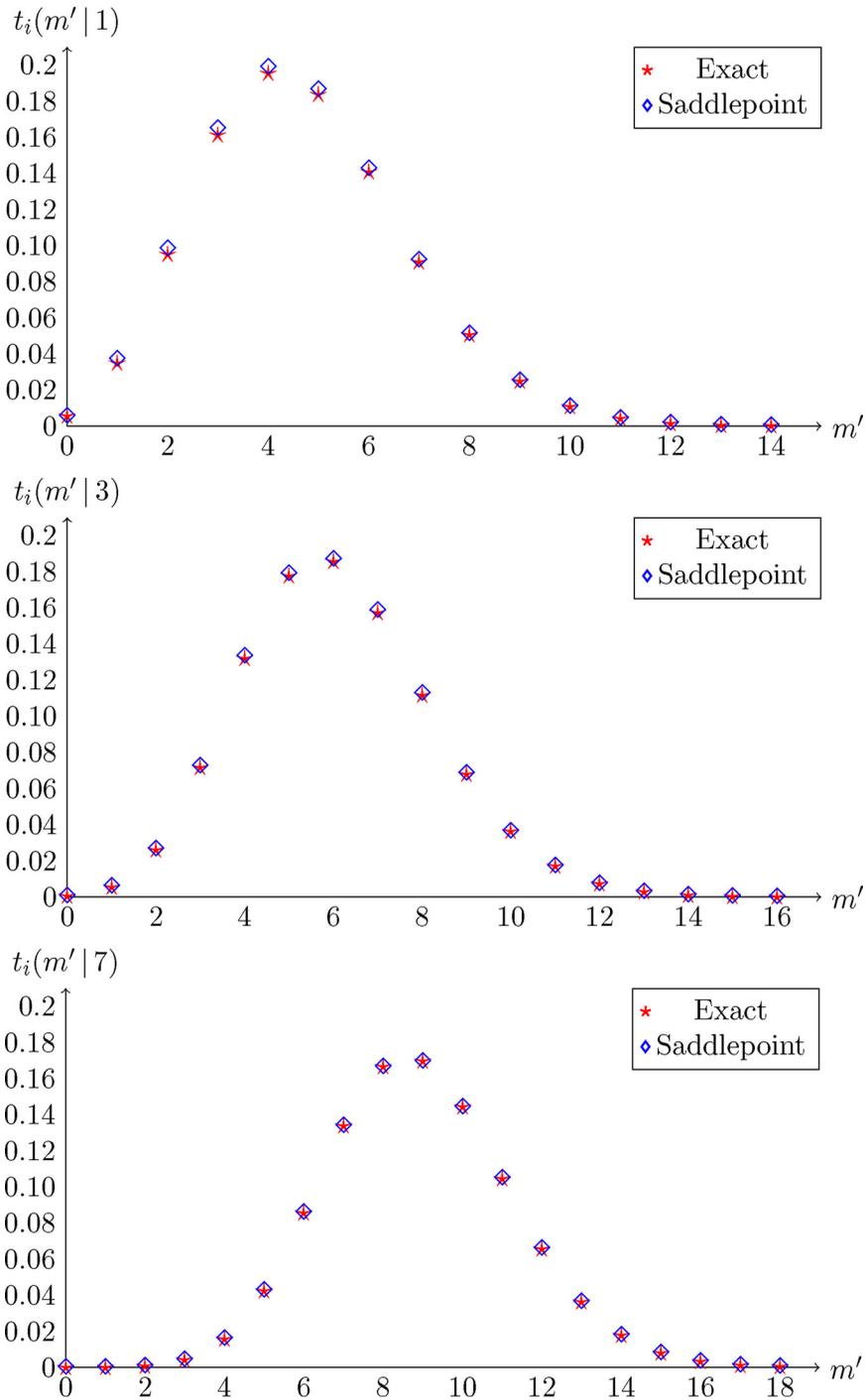
$$\Sigma_i(v | m) = \frac{\partial^2 \zeta_i(s | m)}{\partial s^2} \Big|_{s=s_v^*}.$$

Using that (14) holds when  $s = s_v^*$ , we thus find

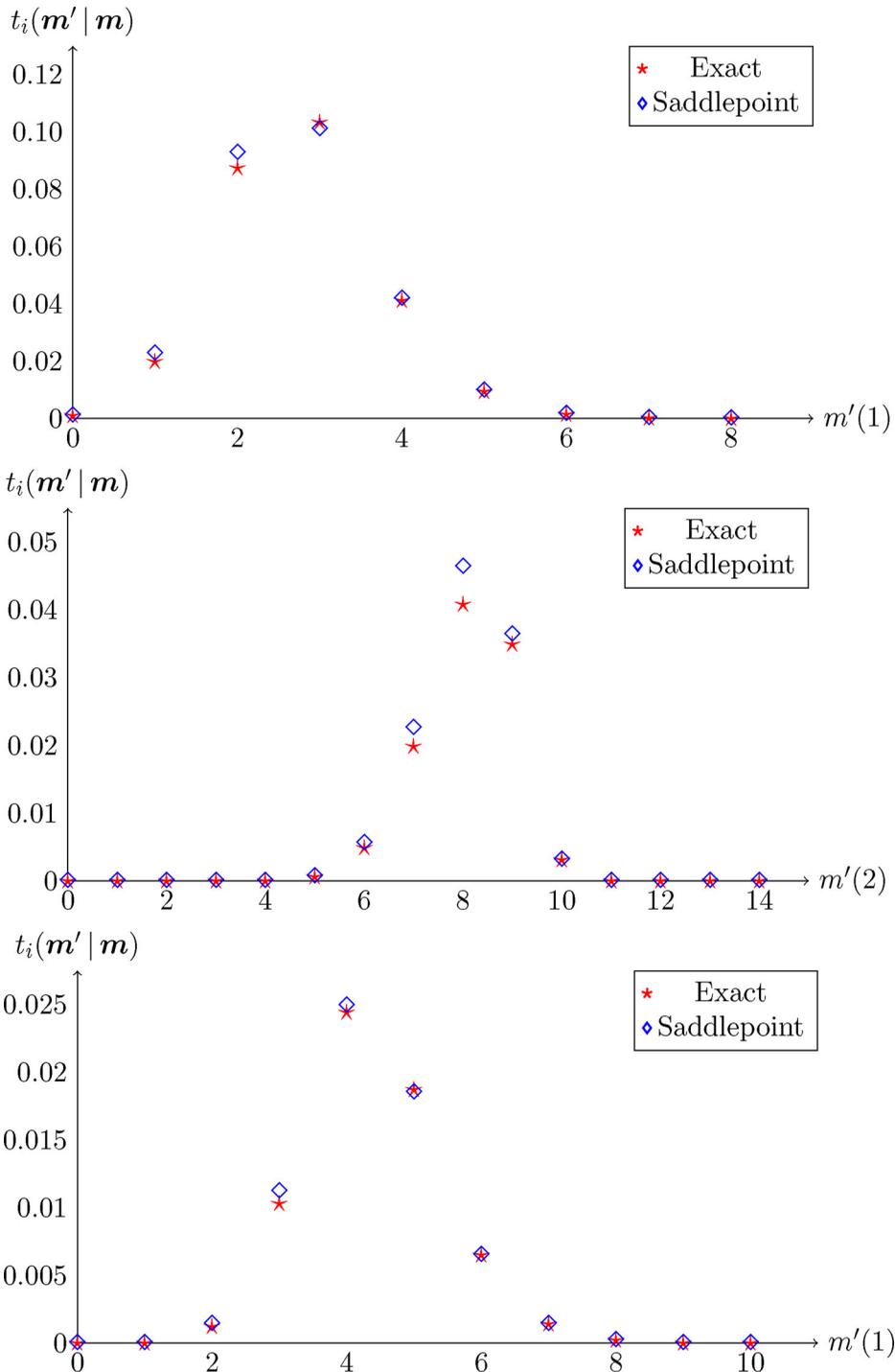
$$\Sigma_i(v | m) = \frac{\lambda_i(1 - r_i)e^{2s_v^*} + vr_i}{(1 - r_i)e^{s_v^*} + r_i}$$

We have now collected all ingredients to evaluate the saddlepoint approximation (11). Concerning  $m' \in S_i(m) \setminus S_i(m)^\circ = \{0\}$ , we evidently have  $t_i(0 | m) = e^{-\lambda_i} (r_i)^m$  as we saw before.

In Figure 3 we show the numerically obtained approximation in the single-node setting. It displays three examples which provide a good reflection of the accuracy typically achieved by the saddlepoint approach. In particular, they illustrate that the accuracy improves as the value of  $m$  increases, which is a known feature of saddlepoint approximations.



**Figure 3.** Saddlepoint approximation and exact computation of  $t_i(m' | m)$  for the single-node model as a function of  $m'$ , for increasing values of  $m$ ; from the top to bottom panel,  $m = 1, m = 3$  and  $m = 7$ . Parameter values:  $i = 1, \lambda_1 = 4$  and  $r_1 = 0.3$ .



**Figure 4.** Saddlepoint approximation and exact computation of  $t_i(m' | m)$  for the tandem network as a function of  $m'$ . Parameter values:  $i = 1, \lambda_1 = 0.5, r_1^{(1,2)} = 0.5$  and  $r_1^{(2,0)} = 0.2$ . Throughout we have fixed  $m = (5, 5)^\top$ . Upper panel: we vary  $m'(1)$ , with  $m'(2) = 7$ . Middle panel: we vary  $m'(2)$ , with  $m'(1) = 1$ . Bottom panel: we vary  $m'(1)$ , with  $m'(2) = 4$ .

**Example 2 . Tandem network.** To further assess the accuracy of the saddlepoint approximation, we consider the example of the tandem network with two nodes, as introduced in Section 3.2. We compute  $t_i(\mathbf{m}' \mid \mathbf{m})$  explicitly, and compare it with its saddlepoint-based counterpart. We do this for  $\mathbf{m}' \in \mathcal{S}_i(\mathbf{m})^\circ$ ; note that we already discussed above how to deal with the boundary points  $\mathbf{m}' \in \mathcal{S}_i(\mathbf{m}) \setminus \mathcal{S}_i(\mathbf{m})^\circ$ . We can compute the cumulant generating function  $\zeta_i(\mathbf{s} \mid \mathbf{m})$  from (9). From the fact that we have Poisson arrivals, we know that  $\psi_i(\mathbf{s})$  follows from (2). The cgf equals

$$\begin{aligned} \zeta_i(\mathbf{s} \mid \mathbf{m}) &= \lambda_i(e^{s(1)} - 1) + \mathbf{m}(1) \log \left( (1 - r_i^{(1,2)})e^{s(1)} + r_i^{(1,2)}e^{s(2)} \right) \\ &\quad + \mathbf{m}(2) \log \left( (1 - r_i^{(2,0)})e^{s(2)} + r_i^{(2,0)} \right). \end{aligned}$$

Hence, for  $\mathbf{v} \in \mathcal{S}_i(\mathbf{m})^\circ$  the saddlepoint  $\mathbf{s}_v^*$  can be found by solving the equations

$$\begin{aligned} v(1) &= \lambda_i e^{s(1)} + \mathbf{m}(1) \frac{(1 - r_i^{(1,2)})e^{s(1)}}{(1 - r_i^{(1,2)})e^{s(1)} + r_i^{(1,2)}e^{s(2)}} \\ v(2) &= \mathbf{m}(1) \frac{r_i^{(1,2)}e^{s(2)}}{(1 - r_i^{(1,2)})e^{s(1)} + r_i^{(1,2)}e^{s(2)}} + \mathbf{m}(2) \frac{(1 - r_i^{(2,0)})e^{s(2)}}{(1 - r_i^{(2,0)})e^{s(2)} + r_i^{(2,0)}}. \end{aligned}$$

Having found the solution  $\mathbf{s}_v^*$ , the approximation<sup>[21]</sup> is readily evaluated.

Numerical results for a few representative examples are presented in Figure 4. The upper panel in Figure 4 shows a cross section at the peak of the joint distribution of  $\mathbf{m}'(1)$  and  $\mathbf{m}'(2)$ , the middle panel shows a cross section close to the peak, and the bottom panel shows a cross section further away from the peak. Our findings confirm the approach's high accuracy that we observed earlier.

## 5. Parameter estimation

In this section, we show how the saddlepoint approximation of the likelihood—developed in the previous section—can be used to estimate the model parameters, and we assess the accuracy of this estimation method by applying it to simulated data.

As argued before, we can use the saddlepoint approximation in (11) to approximate the probabilities  $t_i(\mathbf{m}_k \mid \mathbf{m}_{k-1})$  for each pair of observations  $(\mathbf{m}_{k-1}, \mathbf{m}_k)$  ( $k = 1, \dots, n$ ) and each  $i \in E$ , so as to evaluate the likelihood (4). This likelihood is then to be maximized over the model parameters (in the appropriate parameter space) to find the parameter estimate  $\hat{\theta}$ . We do this numerically, relying on the built-in solver *fmincon* of matlab.

The solver *fmincon* needs an initial value for  $\theta$ . There are various ways to choose this value.

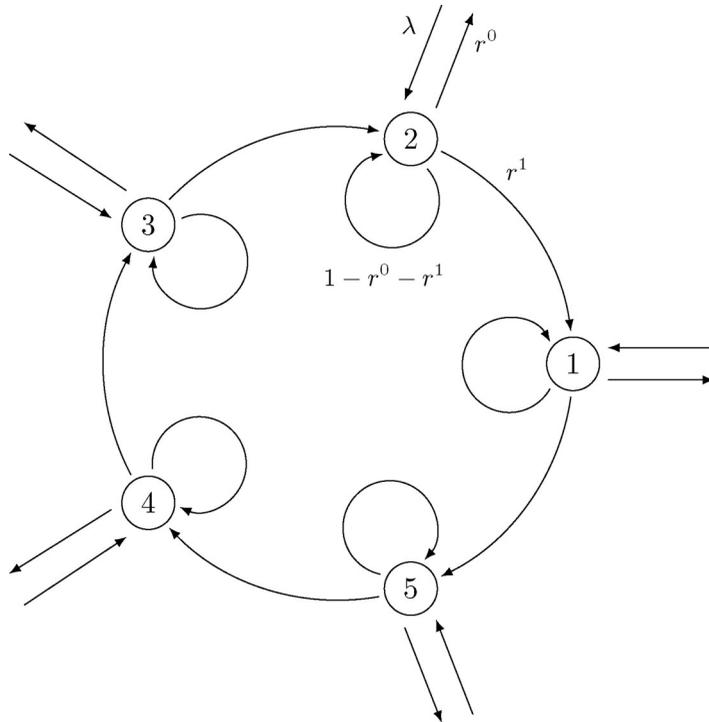
- In case the parameter space is finite, a naïve approach would be to sample the initial value uniformly on the parameter space.
- Another approach is to let the routing be uniform, in the sense that for any individual all next nodes are equally likely; for example in a fully connected graph (i.e., the situation that all  $r_i^{(\ell, \ell)}$  are positive), we could set  $r_i^{(\ell, 1)} = r_i^{(\ell, 2)} = \dots = r_i^{(\ell, L)} = r_i^{(\ell, 0)} = (L + 1)^{-1}$  for all  $i \in E$  and  $\ell = 1, \dots, L$ . Likewise, the transition probabilities  $p_{ij}$  could be initialized with  $1/d$ .
- Alternatively, the initial  $\theta$  can be determined using moment estimators, or, if available, additional information on the parameters can be used to set a suitable initial value.

In the remainder of this section, we specify the initial values that we used for each numerical experiment. As is commonly known, the maximum likelihood approach has the intrinsic issue that there can be local maxima. It is therefore strongly advised to follow the usual procedure to work with multiple initial values (and to record the one providing the highest likelihood).

**Remark 2.** Observe that for example in a model with  $d=2$ , swapping the states in the parametrization results in an observationally equivalent model. In case of such identifiability issues, additional constraints need to be imposed on the parameters. In the single-node case of  $d=2$  with an environment-dependent arrival rate, such a constraint could for instance be  $\lambda_2 \geq \lambda_1$ .

**Remark 3.** We note that, by the structure of expression (4), the evaluation of the likelihood is linear in  $n$  and cubic in  $d$ . The complexity of the saddlepoint-based approximation is relatively low, due to the concavity of the functions  $I_i(\mathbf{v}, \mathbf{s} \mid \mathbf{m})$ .

To illustrate the broad applicability of the method, we perform numerical experiments for a set of intrinsically different networks. We specifically investigate the influence of the number of observations  $n$  on the estimates: throughout, we evaluate the estimators for  $n = 100, n = 500, n = 1000$ , and  $n = 2000$ . For each network and each value of  $n$ , we simulate 100 data sets, to each of which we apply the estimation method to obtain the parameter estimates. We present and discuss our findings in this section. We use the two examples from Sections 3 and 4, i.e., the single-node and the tandem, but we start with an experiment featuring a larger network with a different structure: a circle network.



**Figure 5.** Schematic representation of the circle network of 5 nodes.

**Experiment 1. Circle network.** We consider a network of five nodes in a circle. The individuals can move clockwise through the network from one node to the next. In this experiment we primarily concentrate on the effect of the network structure, and therefore we do not impose modulation (i.e., we consider the setting  $d=1$ ). In addition, we let the network be homogeneous, in the sense that the arrival processes, the probabilities of leaving the network, and the probabilities of being forwarded to the next node, respectively, are the same for any node. More concretely, we work with three parameters  $\lambda_1^{(1)} = \dots = \lambda_1^{(5)} := \lambda$ ,  $r_1^{(1,0)} = \dots = r_1^{(5,0)} := r^0$ , and  $r_1^{(1,2)} = r_1^{(2,3)} = \dots = r_1^{(5,1)} := r^1$ . This means that *at each node* arrivals occur according to a Poisson process with rate  $\lambda$ , and any individual present at the node leaves with probability  $r^0$ , or jumps to the following node in the circle with probability  $r^1$ . Note that, as a result, individuals stay at a node with probability  $1 - r^0 - r^1$ ; see Figure 5 for a pictorial illustration.

Despite the fact that there is no modulation, direct evaluation of the likelihood is challenging. As pointed out earlier, the high complexity essentially lies in the fact that we observe the network population vector only, and not the arrival, routing and departure processes. An exact evaluation of the likelihood would require taking into account all paths of the arrival, routing and departure processes that match with the observed values of the

**Table 1.** *Circle network:* mean of estimates of 100 data sets, with corresponding standard deviation between brackets. True parameter values:  $\lambda = 1.5, r^1 = 0.3, r^0 = 0.1$ .

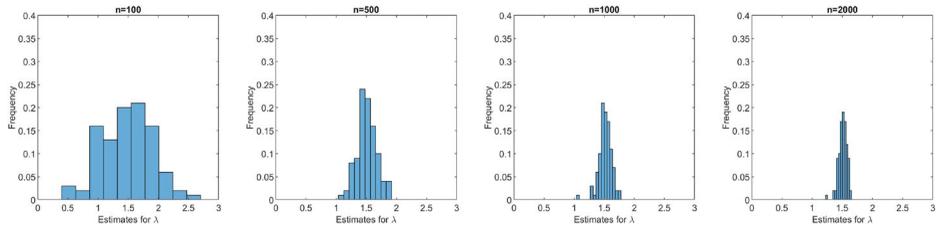
	$\lambda$	$r^1$	$r^0$
$n = 100$	1.4902 (0.4319)	0.3143 (0.0317)	0.1003 (0.0271)
$n = 500$	1.5081 (0.1613)	0.3070 (0.0131)	0.1006 (0.0104)
$n = 1000$	1.5197 (0.1034)	0.3059 (0.0086)	0.1013 (0.0062)
$n = 2000$	1.5103 (0.0679)	0.3072 (0.0057)	0.1007 (0.0045)

network population vector, which for our five-node circle network would be infeasible. This motivates why we resort to evaluating the likelihood using the saddlepoint approximation.

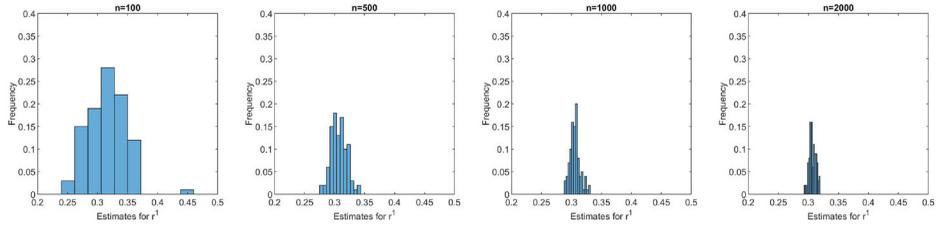
In our experiments we use simulated data that are generated using the parameter values  $\lambda = 1.5, r^1 = 0.3$ , and  $r^0 = 0.1$ . The maximum likelihood estimation procedure using *fmincon* is initialized at  $\lambda = 1, r^1 = \frac{1}{3}$  and  $r^0 = \frac{1}{3}$ . Experiments with other initial values lead to similar results. The numerical output is shown in Table 1 and Figures 6–8. Table 1 contains, for each sample size (rows) and parameter (columns), the mean value of the 100 estimates, together with the corresponding standard deviation between brackets. We see that the mean values in Table 1 lie close to the true parameter values, and that (as expected) the standard deviations decrease as  $n$  increases. This is visible in the histograms as well, displayed in Figures 6–8. Each figure shows, for a given value of  $n$  and one of the three parameters, the histogram of the 100 estimates. For each of the three parameters, we intentionally chose the same horizontal axis in all four pictures, so as to provide insight into the speed at which the width of the peak decreases as  $n$  grows.

**Experiment 2.** *Single-node model.* As a second example, we study the single-node model as introduced in Section 3.1. We consider the setup with  $r := r_1 = r_2$ , which means that only the arrival rate is affected by the modulation, not the departure probability  $r$ . In our simulations we use the parameter values  $\lambda_1 = 5, \lambda_2 = 15, r = 0.1, p_{12} = 0.1$  and  $p_{21} = 0.2$ . The initial values in the algorithm that maximizes the log-likelihood are based on moment estimators. The results of the maximum likelihood estimates are shown in Table 2 and Figure 9.

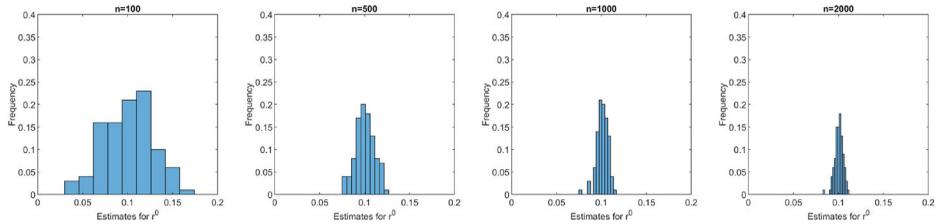
Table 2 contains for each sample size (rows) and parameter (columns), the mean value of the 100 estimates, together with the corresponding standard deviation between brackets. The mean values of the estimates lie relatively close to the true parameter values, but the standard deviations fluctuate and do not always decrease in  $n$ . The histograms in Figure 9, featuring estimates for  $r$ , however, visually show that the estimates get increasingly concentrated around their respective averages. We observe that the values in the table are affected by outliers in the estimates. As we mentioned earlier, when maximizing the likelihood we cannot exclude the possibility of ending up in local optima. In the circle network we have not



**Figure 6.** *Circle network:* histograms of the obtained estimates for  $\lambda$ , with  $n$  increasing from left to right.



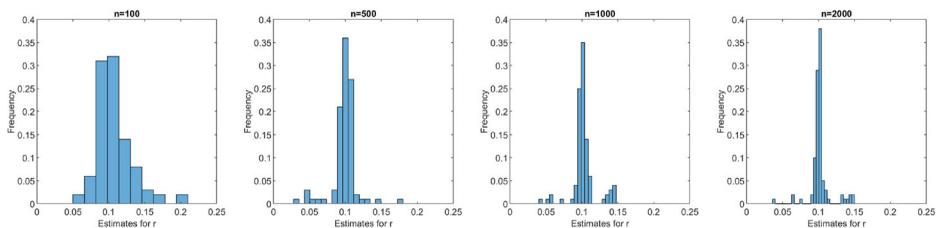
**Figure 7.** *Circle network:* histograms of the obtained estimates for  $r^1$ , with  $n$  increasing from left to right.



**Figure 8.** *Circle network:* histograms of the obtained estimates for  $r^0$ , with  $n$  increasing from left to right.

**Table 2.** *Single-node model:* mean of estimates of 100 data sets, with corresponding standard deviation between brackets. True parameter values:  $\lambda_1 = 5, \lambda_2 = 15, r = 0.1, p_{12} = 0.1, p_{21} = 0.2$ .

	$\lambda_1$	$\lambda_2$	$r$	$p_{12}$	$p_{21}$
$n = 100$	5.6507 (2.7007)	14.9201 (3.1332)	0.1081 (0.0259)	0.1266 (0.1525)	0.2331 (0.1869)
$n = 500$	4.8028 (1.8180)	14.5595 (1.7340)	0.0991 (0.0174)	0.1240 (0.1011)	0.1977 (0.0630)
$n = 1000$	5.0127 (2.3698)	14.2682 (1.9638)	0.1024 (0.0168)	0.1432 (0.1559)	0.2047 (0.1011)
$n = 2000$	5.2278 (2.1442)	14.6499 (2.4113)	0.1024 (0.0150)	0.1398 (0.1629)	0.2147 (0.1136)



**Figure 9.** *Single-node model:* histograms of the obtained estimates for  $r$ , with  $n$  increasing from left to right.

**Table 3.** *Single-node model:* mean of estimates of 100 data sets, with corresponding standard deviation between brackets. True parameter values:  $\lambda_1 = 5, \lambda_2 = 15, r = 0.1, p_{12} = 0.1, p_{21} = 0.2$ .

	$\lambda_1$	$\lambda_2$	$r$	$p_{12}$	$p_{21}$
$n = 100$	5.1469 (1.9209)	15.2419 (2.4189)	0.1015 (0.0221)	0.1107 (0.0543)	0.2445 (0.1492)
$n = 500$	5.0155 (0.4463)	14.9568 (0.6144)	0.1004 (0.0051)	0.1007 (0.0179)	0.2019 (0.0357)
$n = 1000$	5.0690 (0.3602)	15.0828 (0.4532)	0.1011 (0.0044)	0.0987 (0.0138)	0.1971 (0.0289)
$n = 2000$	5.0195 (0.2274)	15.0482 (0.3116)	0.1004 (0.0027)	0.1004 (0.0106)	0.2067 (0.0242)

come across this phenomenon, but in our experiments with modulation there have been a few runs in which we have. The histograms in [Figure 9](#) show these outliers near 0.05 and 0.15. In the histograms of the other parameters (not included in this paper), similar outliers appear.

To control this issue, it is advised to run the maximization algorithm for multiple different initial values of the parameters, and choose the parameter estimates that result in the highest likelihood value. Results of the maximum likelihood estimates based on this procedure, using four different, randomly chosen, initial values of the parameters, are shown in [Table 3](#). [Table 3](#) shows that the standard deviations improved considerably in comparison with the results in [Table 2](#). In particular, the outliers have disappeared resulting in standard deviations that decrease in  $n$ .

A subtlety is that the accuracy of the saddlepoint approximation for background state  $i$  degrades when  $\lambda_i$  approaches 0. This is because in the regime of this arrival rate being 0,  $m_k > m_{k-1}$  cannot happen, thus effectively creating a boundary state; cf. the discussion in [Section 4.2](#). We followed the pragmatic remedy of imposing an explicit lower bound on the arrival rates (in our experiments we took 0.01).

**Experiment 3. Tandem network.** We continue by considering a two-node tandem network with modulation, as introduced in [Section 3.2](#). In this experiment we assume  $P$  is known and given by

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}.$$

We run simulations for this model with true parameters  $\lambda_1 = 1, \lambda_2 = 4, r^{(1,2)} = 0.1$  and  $r^{(2,0)} = 0.25$ . In our likelihood maximization routine we choose the initial values  $\lambda_1 = 1, \lambda_2 = 1, r^{(1,2)} = 0.5$ , and  $r^{(2,0)} = 0.5$ . Experiments with other initial values provide similar output. The results are presented in [Table 4](#), showing for each sample size (rows) and parameter (columns), the mean value of the 100 estimates, together with the corresponding standard deviation between brackets. In line with the first two experiments, we observe that the mean values in [Table 4](#) lie close to the true parameter values. The standard deviation fluctuates somewhat, but this effect can again be mitigated by working with multiple initial values.

**Table 4.** *Tandem network*: mean of estimates of 100 data sets, with corresponding standard deviation between brackets. True parameter values:  $\lambda_1 = 1, \lambda_2 = 4, r^{(1,2)} = 0.1, r^{(2,0)} = 0.25$ .

	$\lambda_1$	$\lambda_2$	$r^{(1,2)}$	$r^{(2,0)}$
$n = 100$	1.6896 (0.8324)	3.5884 (1.0424)	0.1169 (0.0157)	0.2893 (0.0415)
$n = 500$	1.3161 (0.4669)	3.9394 (0.7301)	0.1076 (0.0096)	0.2692 (0.0230)
$n = 1000$	1.1441 (0.4291)	3.9505 (0.5430)	0.1054 (0.0083)	0.2633 (0.0224)
$n = 2000$	1.1610 (0.5968)	3.7440 (0.7026)	0.1052 (0.0103)	0.2645 (0.0258)

## 6. Discussion and concluding remarks

In this paper we considered a discrete-time multivariate population process under Markov modulation. We showed how the likelihood can be evaluated using saddlepoint approximations, and how this can be used to estimate the model parameters. We emphasize the model's high degree of generality, covering a wide variety of networks with different sizes and structures, and on top of that the possibility to include modulation. Moreover, the maximum-likelihood estimation approach is capable of estimating parameters based on observations of the network population vector only. In other words, the number of arrivals, jumps, and departures are not observed, but only the net effect of these processes together, while the modulating background process is not observed at all.

We illustrated the accuracy of the saddlepoint approximation through two examples, namely a single-node model and a tandem network. For these examples the likelihood can still be computed explicitly, and hence the explicit computation can be compared with the saddlepoint approximation. In a series of numerical tests we found that the differences between the two are typically small.

Then we investigated the accuracy of the maximum-likelihood estimation method through a number of numerical studies. We focused on three different settings corresponding to networks with different sizes and structures. In all examples accurate estimates are obtained. Moreover, working with multiple initial values to eliminate the outliers, results in standard deviations that decrease as the sample size grows.

The estimation method in general produces accurate estimates, but in a few cases the maximization ends up in a local maximum, as a consequence of the specific shape of the likelihood surface. One effective way to control it is by using multiple different initial values, choosing the outcome that results in the highest likelihood value. To be sure that the estimation method correctly tracks down modulation, it is important that the effect of the background state is visible in the data. More concretely, one can imagine a parameter setting in for example the single-node model with two states, in which the effect of the higher arrival rate on the population size is essentially canceled out by a higher departure probability, such that the states cannot be distinguished.

We believe that the results presented in this paper offer various interesting opportunities for further research. In the first place, note that in our setup we assumed that the number of states  $d$  is known. Choosing  $d$  from the data is a model selection problem and falls outside the scope of this paper, but would be worth studying in greater detail. Second, we focused on a discrete-time setting, allowing the computation of the cgf s, and thus facilitating the application of the saddle-point technique, but one wonders whether a similar approach could be followed for our model's continuous-time counterpart. The major complication is that if the background process evolves continuously in time, it is not directly clear how to compute the cgf s.

Various adaptations of our model could be considered as well. In this paper, we considered only one type of individual, and (conditionally on a realization of the background process) all individuals move independently of each other through the network. Instead one could study multi-type models, or models with routing and departure probabilities that depend on the population vector before and/or after the transition, besides the state of the background process.

## References

- [1] Chen, R.; Hyrien, O. Quasi- and pseudo-maximum likelihood estimators for discretely observed continuous-time Markov branching processes. *J. Stat. Plan. Inference* 2011, **141**, 2209–2227. DOI: [10.1016/j.jspi.2011.01.016](https://doi.org/10.1016/j.jspi.2011.01.016).
- [2] Crawford, F.; Minin, V.; Suchard, M. Estimation for General Birth-Death Processes. *J. Am. Stat. Assoc.* 2014, **109**, 730–747. DOI: [10.1080/01621459.2013.866565](https://doi.org/10.1080/01621459.2013.866565).
- [3] Davison, A.; Hautphenne, S.; Kraus, A. 2018. Parameter estimation for discretely observed linear birth-and-death processes. *arXiv: 1802.05015*.
- [4] Tavaré, S. The Linear Birth–Death Process: An Inferential Retrospective. *Adv. Appl. Probab.* 2018, **50**, 253–269. DOI: [10.1017/apr.2018.84](https://doi.org/10.1017/apr.2018.84).
- [5] Xu, J.; Guttorp, P.; Kato-Maeda, M.; Minin, V. N. Likelihood-Based Inference for Discretely Observed Birth-Death-Shift Processes, with Applications to Evolution of Mobile Genetic Elements. *Biometrics* 2015, **71**, 1009–1021. DOI: [10.1111/biom.12352](https://doi.org/10.1111/biom.12352).
- [6] Pickands, J.; Stine, R. Estimation for an  $M/G/\infty$  Queue with Incomplete Information. *Biometrika* 1997, **84**, 295–308.
- [7] Bingham, N.; Pitts, S. Non-Parametric Estimation for the  $M/G/\infty$  Queue. *Ann. Inst. Stat. Math.* 1999, **51**, 71–97.
- [8] Hautphenne, S.; Fackrell, M. An EM Algorithm for the Model Fitting of Markovian Binary Trees. *Comput. Stat. Data Anal.* 2014, **70**, 19–34. DOI: [10.1016/j.csda.2013.08.015](https://doi.org/10.1016/j.csda.2013.08.015).
- [9] Hautphenne, S.; Massaro, M.; Turner, K. Fitting Markovian Binary Trees Using Global and Individual Demographic Data. *Theor. Populat. Biol.* 2019, **128**, 39–50. DOI: [10.1016/j.tpb.2019.04.007](https://doi.org/10.1016/j.tpb.2019.04.007).
- [10] Breuer, L.; Kume, A. 2010. *An EM Algorithm for Markovian Arrival Processes Observed at Discrete Times*. Springer, International GI/ITG Conference on

- Measurement, Modeling, and Evaluation of Computing Systems and Dependability and Fault Tolerance, pp. 242–258
- [11] Okamura, H.; Dohi, T.; Trivedi, K. Markovian Arrival Process Parameter Estimation with Group Data. *IEEE/ACM Trans. Netw.* **2009**, *17*, 1326–1339. DOI: [10.1109/TNET.2008.2008750](https://doi.org/10.1109/TNET.2008.2008750).
  - [12] de Gunst, M.; Knapik, B.; Mandjes, M.; Sollie, B. Parameter Estimation for a Discretely Observed Population Process under Markov Modulation. *Comput. Stat. Data Anal.* **2019**, *140*, 88–103. DOI: [10.1016/j.csda.2019.06.008](https://doi.org/10.1016/j.csda.2019.06.008).
  - [13] Kolaczyk, E. D. **2010**. Statistical Analysis of Network Data. In *Springer Series in Statistics*. Springer: New York.
  - [14] Daniels, H. Saddlepoint Approximations in Statistics. *Ann. Math. Statist.* **1954**, *25*, 631–650. DOI: [10.1214/aoms/1177728652](https://doi.org/10.1214/aoms/1177728652).
  - [15] Butler, R. **2007**. *Saddlepoint Approximations with Applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press: Cambridge; **22**.
  - [16] AiT-Sahalia, Y.; Yu, J. Saddlepoint Approximations for Continuous-Time Markov Processes. *J. Econometrics* **2006**, *134*, 507–551. DOI: [10.1016/j.jeconom.2005.07.004](https://doi.org/10.1016/j.jeconom.2005.07.004).
  - [17] Daniels, H. The Saddlepoint Approximation for a General Birth Process. *J. Appl. Probab.* **1982**, *19*, 20–28. DOI: [10.1017/S0021900200028242](https://doi.org/10.1017/S0021900200028242).
  - [18] Barndorff-Nielsen, O.; Cox, D. Edgeworth and Saddlepoint Approximations with Statistical Applications (with Discussion). *J. Royal Stat. Society Series B* **1979**, *41*, 279–312. DOI: [10.1111/j.2517-6161.1979.tb01085.x](https://doi.org/10.1111/j.2517-6161.1979.tb01085.x).
  - [19] Reid, N. Saddlepoint Methods and Statistical Inference. *Statist. Sci.* **1988**, *3*, 213–227. DOI: [10.1214/ss/1177012906](https://doi.org/10.1214/ss/1177012906).
  - [20] Pedeli, X.; Davison, A.; Fokianos, K. Likelihood Estimation for the INAR( $p$ ) Model by Saddlepoint Approximation. *J. Am. Stat. Assoc.* **2015**, *110*, 1229–1238. DOI: [10.1080/01621459.2014.983230](https://doi.org/10.1080/01621459.2014.983230).
  - [21] Eisinga, R.; Grotenhuis, M. T.; Pelzer, B. Saddlepoint Approximations for the Sum of Independent Non-Identically Distributed Binomial Random Variables. *Stat. Neerl.* **2013**, *67*, 190–201. DOI: [10.1111/stan.12002](https://doi.org/10.1111/stan.12002).
  - [22] Key, E. S. Limiting Distributions and Regeneration Times for Multitype Branching Processes with Immigration in a Random Environment. *Ann. Probab.* **1987**, *15*, 344–353. DOI: [10.1214/aop/1176992273](https://doi.org/10.1214/aop/1176992273).
  - [23] Roitershtein, A. A Note on Multitype Branching Processes with Immigration in a Random Environment. *Ann. Probab.* **2007**, *35*, 1573–1592. DOI: [10.1214/009117906000001015](https://doi.org/10.1214/009117906000001015).