Continuous approximation of collective system behaviour: A tutorial

Bortolussi, L., Hillston, J., Latella, D., & Massink, M. (2013). Continuous approximation of collective system behaviour: A tutorial. Performance Evaluation, 70(5), 317-349, doi: 10.1016/j.peva.2013.01.001

Continuous Approximation of Collective Systems Behaviour: a Tutorial[☆]

Luca Bortolussi[d,f], Jane Hillston[e], Diego Latella[f], Mieke Massink[f]

[a]*Dept. of Mathematics and Geosciences, University of Trieste, Italy.*
[b]*Laboratory for Foundations of Computer Science, University of Edinburgh, UK.*
[c]*ISTI, CNR, Pisa, Italy.*

[☆]Work partially funded by EU Project n. 257414 (ASCENS) and by FRA-UniTS.

Email addresses: luca@dmi.units.it (Luca Bortolussi), jane.hillston@ed.ac.uk (Jane Hillston), Diego.Latella@isti.cnr.it (Diego Latella), Mieke.Massink@isti.cnr.it (Mieke Massink)

Continuous Approximation of Collective Systems Behaviour: a Tutorial[☆]

Luca Bortolussi[d,f], Jane Hillston[e], Diego Latella[f], Mieke Massink[f]

[d]*Dept. of Mathematics and Geosciences, University of Trieste, Italy.*
[e]*Laboratory for Foundations of Computer Science, University of Edinburgh, UK.*
[f]*ISTI, CNR, Pisa, Italy.*

Abstract

In this paper we present an overview of the field of deterministic approximation of Markov processes, both in discrete and continuous time. We will discuss mean field approximation of discrete time Markov chains and fluid approximation of continuous time Markov chains, considering the cases in which the deterministic limit process lives in continuous time or discrete time. We also consider some more advanced results, especially those relating to the limit stationary behaviour. We assume a knowledge of modelling with Markov chains, but not of more advanced topics in stochastic processes.

Keywords: deterministic approximation, fluid approximation, mean field approximation, Markov Chains, stochastic process algebras.

1. Introduction

Markovian stochastic processes have found widespread use in computer science and beyond, for quantitative modelling in both discrete and continuous time. In addition to performance evaluation, they have recently been used in areas as diverse as computational systems biology [1] and modelling user experience [2]. A common problem faced with such models is that the number of elements constituting a system, as well as the number of local states of those elements, can be large, leading to state space explosion.

[☆]Work partially funded by EU Project n. 257414 (ASCENS) and by FRA-UniTS.

Email addresses: luca@dmi.units.it (Luca Bortolussi), jane.hillston@ed.ac.uk (Jane Hillston), Diego.Latella@isti.cnr.it (Diego Latella), Mieke.Massink@isti.cnr.it (Mieke Massink)

This poses practical limitations on our ability to analyse these systems using standard approaches, like steady state computation and transient analysis of Markov Chains [3], or more sophisticated techniques, like stochastic model-checking [4].

In the last few years, there has been a growing interest in techniques that try to tackle the state space explosion by treating large state spaces in a continuous fashion, approximating the stochastic dynamics with a deterministic dynamical system, usually described by means of a differential equation. This is particularly fruitful for systems which are composed of large clusters of (relatively simple) components.

Such techniques have been developed in different contexts and with different flavours, yet their approach is very similar. Broadly speaking, nomenclature is (at least) twofold: continuous approximation is known under the name of both *fluid approximation* and *mean field approximation*. The term *deterministic approximation* is sometimes also used.

Fluid approximation has been introduced and applied in the last five years in order to analyse the collective behaviour of stochastic process algebra (SPA) models of large populations [5, 6, 7, 8, 9]. Stochastic process algebras [10, 11, 12] are modelling languages, designed to describe systems of interacting agents, which have Continuous Time Markov Chains (CTMCs) as the relevant semantic domain [13]. Fluid approximation can be applied if in a model there are many instances of a few agent types, and it works by treating the variables counting how many agents of each type are in each state in the system as continuous variables, and treating the global rates of the stochastic transitions as flows, thus obtaining an ordinary differential equation (ODE). After observing that, for large populations, the behaviour of the stochastic system and that of the deterministic one were very close [5], it was proved that the solution of the differential equation obtained was the limit of a sequence of CTMC models, for increasing population levels [6], exploiting previous results on deterministic approximation of stochastic processes [14]. Fluid approximation of SPA models has been used successfully to describe different kind of systems: computer epidemics [15], biological systems [16, 7, 9, 17], computer networks [18], queues [19], Grid workloads [20], and crowd models [21, 22], just to cite a few.

Another use of fluid techniques can be found in Petri Nets. There have been a variety of different Petri net developments involving *fluid levels*, either as alternatives to, or in addition to, the usual discrete tokens [23, 24, 25]. Perhaps closest to the process algebra results outlined above is the work of

Silva and Recalde [26] where the motivation is again the state space explosion problem, and the authors present their work in terms of a relaxation of integer token counts for state representation.

Mean field approximation is very similar in spirit to fluid approximation. Mean field techniques have a long history: originally developed in statistical physics (in the context of stochastic models of plasma and dense gases [27]), they have been applied in epidemiology [28, 29, 30], game theory [31, 32], approximation algorithms [33], and in performance modelling of computer networks [34, 35, 36, 37, 38, 39, 40]. Usually, mean field approaches start from a stochastic model expressed directly in terms of a Discrete Time Markov Chain (DTMC), describing a system consisting of a large number of interacting entities, each of which can be in one of a relatively small set of states. Then, one constructs a continuous system, describing the continuous evolution of the number of these entities (more specifically, the variables counting how many entities are in a given state), proving a limit theorem similar to the one for fluid approximation. Many applications of mean field approximation in computer science are concerned with communication networks [34, 35, 37, 40], and the limit theorems are proved just for the specific model in each case. More recently, mean field results for more general frameworks have been presented [41, 42, 43], and applied, for instance, to study properties of gossip protocols [44], random replacement caching [45], and resource polling [46].

Broadly speaking, there are three different classes of approaches that have been labelled as mean field. The first one deals with DTMC models that have a deterministic limit in discrete time (i.e. a discrete time dynamical system) [41], the second one considers DTMCs that have a limit in continuous time, described in terms of differential equations [42], and the last one deals with CTMCs which have a limit in continuous time [14, 47]. This last approach is essentially the same as fluid approximation, the only difference being that the authors work directly on a CTMC model, instead of manipulating a SPA model. The second class of mean field techniques, i.e. those concerned with DTMC models that have a deterministic limit in continuous time, is also strongly related to continuous approximations for CTMCs, as will be made clear later in this paper. The first class, instead, is intrinsically different, and can be applied to DTMC models that assume a different mechanism of interaction. In particular, mean field limits in discrete time require us to work with a DTMC in which all entities of the model (try to) perform a move at each step of the process, while mean field limits in continuous time

assume that just one or few entities perform a move in each step.

From the previous discussion, it is clear that fluid approximation and mean field techniques are strongly related, and that there is a certain amount of notational and terminological confusion in the literature. This can create some difficulties to a modelling practitioner or a student wishing to approach the field.

In this paper we seek to overcome these issues by providing a uniform introduction to these techniques. Our intended audience will be computer scientists who have a background in modelling, but we will not assume prior knowledge of continuous approximation. Readers may be motivated by a desire to apply these techniques to a particular problem or by a more general curiosity, but in either case we aim to introduce these methods in a *gentle* way. We will start our presentation from the simplest class of mean field limits, namely those in discrete time (Section 2). Then, we will move to mean field and fluid approximation in continuous time. We will discuss these methods using a very simple modelling language (basically a direct description of a Markov chain, either in discrete or in continuous time), in order to focus more on aspects related to the dynamics (Section 3). After discussing general issues of the continuous approximation (Section 3), we will first describe the continuous approximation for CTMCs with ODEs (Section 4), and then focus on the ODE, continuous time, approximation of DTMCs (Section 5), followed by a discussion of how these two approaches are related (Section 6). Then we will discuss some general applications, mainly concerned with stationary behaviour (Section 7) and independence (Section 8), pointing out various extensions which have recently appeared in the literature. Throughout the paper, we will make use of a running example, a simple model of a computer network epidemic, to illustrate all the approaches (Sections 2.1 and 3.2). Notational conventions are briefly summarised in Appendix A. Sections and remarks marked with an asterisk can be safely skipped at a first reading.

2. Deterministic Approximation for synchronous DTMCs

We will start the presentation of deterministic approximation results from what is arguably the simplest setting when dealing with Markov Chains: we will consider a discrete time Markov Chain model of a population composed of N agents all moving synchronously at each step. The material in this section is mainly based on [41].

To be more precise, let us introduce some notation. Agents in the population can be in one of n internal states $S = \{1, \ldots, n\}$. We indicate the state of the m-th agent by $Y_m^{(N)} \in S$. Therefore, the state of the DTMC after k steps is given by the vector $(Y_1^{(N)}(k), \ldots, Y_N^{(N)}(k)) \in S^N$.

In population models, one usually makes the assumption that *single individuals are indistinguishable*, hence what is relevant is how many of them are in each state. Formally, this is enforced by requiring that the evolution of the DTMC is equivalent under any permutation of $(Y_1^{(N)}(k), \ldots, Y_N^{(N)}(k))$. Practically, one forgets about single individuals and describes the state of the system by means of counting variables $X_i^{(N)}$, one for each state $i \in S$, defined by

$$X_i^{(N)}(k) = \sum_{j=1}^{N} \mathbf{1}\{Y_j^{(N)}(k) = i\}, \tag{1}$$

where $\mathbf{1}\{\phi(\mathbf{X})\}$ is the indicator function of the predicate $\phi(\mathbf{X})$, equal to 1 if and only if $\phi(\mathbf{X})$ is true. Notice that $X_i^{(N)}(k) \in \{0, \ldots, N\}$, and $\sum_{i=1}^{n} X_i^{(N)}(k) = N$. We indicate by $\mathbf{X}^{(N)}(k)$ the vector of variables $(X_1^{(N)}(k), \ldots, X_n^{(N)}(k))$. The initial configuration of the system will be indicated by $\mathbf{X}^{(N)}(0) = \mathbf{d}_0^{(N)}$.

In addition to the counting variables $X_i^{(N)}$, we also consider the *normalised* variables $\hat{X}_i^{(N)} = X_i^{(N)}/N$, which take values in $[0, 1]$ and sum up to 1. The vector of normalised variables is denoted by $\hat{\mathbf{X}}^{(N)}$, and it takes value in the unit simplex $\mathcal{S}^n$ in $\mathbb{R}^n$, $\mathcal{S}^n = \{\hat{\mathbf{d}} \in [0, 1]^n \mid \sum_{i=1}^{n} \hat{d}_i = 1\}$. The normalised initial conditions are $\hat{\mathbf{d}}_0^{(N)} = \mathbf{d}_0^{(N)}/N$. Normalised variables describe the probability of finding a randomly chosen agent in a given state at step k, and they are usually known under the name of *occupancy measure* [41]. Normalised variables are also important as they allow the comparison of the DTMC dynamics for different population levels. In fact, they are all defined on the same scale $[0, 1]$, and so it makes sense to compute the distance between two models for different population levels. As we will see, this normalisation step is also crucial for the deterministic approximation to hold.

We are now ready to describe the DTMC under consideration. We will define a probability matrix describing the jump probability for a single agent, as a function of the global state of the system, given by the normalised counting variables. More precisely, let $P^{(N)} : \mathcal{S}^n \to [0, 1]^{n^2}$ be a continuous function associating with each $\hat{\mathbf{d}} \in \mathcal{S}^n$ a $n \times n$ stochastic matrix $P^{(N)}(\hat{\mathbf{d}})$.

$P_{ij}^{(N)}(\hat{\mathbf{d}})$ gives the probability that a generic agent, say the h-th, in state i at step k, jumps to state j, given that the global state of the system is $\hat{\mathbf{X}}^{(N)}(k) = \hat{\mathbf{d}}$:

$$P_{ij}^{(N)}(\hat{\mathbf{d}}) = \mathbb{P}\{Y_h^{(N)}(k+1) = j \mid Y_h^{(N)}(k) = i, \hat{\mathbf{X}}^{(N)}(k) = \hat{\mathbf{d}}\}.$$

2.1. Main Example: Epidemic Model

We introduce now a simple example that will be used throughout the paper to present the various aspects of the techniques involved in deterministic approximation. We consider a model of a worm epidemic in a network of computers. The model describes a network of computers in which each node can be infected by a worm. Once this occurs, the worm remains latent for a while, and then activates. When it is active, it tries to propagate over the network by sending messages to other nodes. After some time, an infected computer can be patched, so that the infection is recovered. We assume that recovered computers can become susceptible to infection again after a while, hence modelling the appearance of a new version of the worm. Non-infected computers may also be patched, but this event happens less frequently. Each node in the network can acquire infection from two sources, i.e. by the activity of a worm of an infected node or by an external source (for instance, by an email attachment received from outside the network).

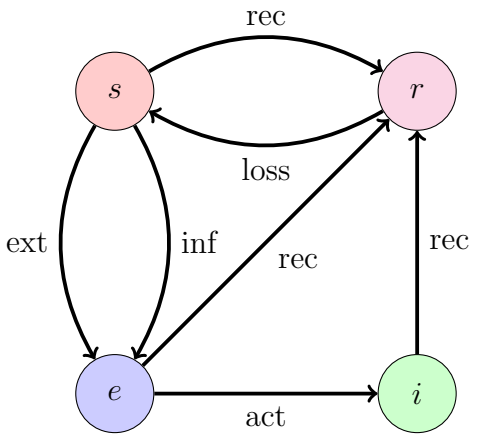


Figure 1: States and transitions of a single computer in the network epidemic model.

In this scenario, we have one single entity class (a computer in the network), each element of which can be in one of several local states: suscep-

tible (s), exposed (e —this is the latent infection period), actively infected (i), and recovered (r). Consequently, each agent $Y_j^{(N)}$ in a network with N nodes takes values in $S = \{s, e, i, r\}$, while the counting variables are given by the vector $\mathbf{X}^{(N)} = (X_s^{(N)}, X_e^{(N)}, X_i^{(N)}, X_r^{(N)})$ and the normalised variables by $\hat{\mathbf{X}}^{(N)} = (\hat{X}_s^{(N)}, \hat{X}_e^{(N)}, \hat{X}_i^{(N)}, \hat{X}_r^{(N)})$. The possible transitions that an agent can make, depicted in Figure 1, are listed below, where in brackets we report the names of the states and the transitions used in the figure:

- Infection (inf) of a susceptible node (s) from an external source, with probability α_e;
- Infection (ext) of a susceptible node (s) from a malicious contact with an infected node, with probability $\alpha_i \hat{X}_i^{(N)}$, i.e. the probability of infection given a contact with an infected node (α_i) times the probability of establishing a contact with an infected node ($\hat{X}_i^{(N)}$);
- Activation (act) of the infection in an exposed node (e), with probability α_a;
- Patching (rec) of an infected node (i), with probability α_r;
- Patching (rec) of a susceptible node (s), with probability $\alpha_p < \alpha_r$;
- Patching (rec) of an exposed node (e), with probability $\alpha_q < \alpha_r$;
- Loss (loss) of immunity of a recovered node (r), with probability α_s;

The previous list can be easily captured in the probability matrix for a single computer node, fixing the ordering (s, e, i, r) of states in S:

$$P^{(N)}(\hat{\mathbf{X}}^{(N)}) = \begin{pmatrix} 1 - \alpha_e - \alpha_i \hat{X}_i^{(N)} - \alpha_p & \alpha_e + \alpha_i \hat{X}_i^{(N)} & 0 & \alpha_p \\ 0 & 1 - \alpha_a - \alpha_q & \alpha_a & \alpha_q \\ 0 & 0 & 1 - \alpha_r & \alpha_r \\ \alpha_s & 0 & 0 & 1 - \alpha_s \end{pmatrix}.$$

Obviously, the values of the α constants have to satisfy some constraints that can be deduced from the previous matrix. For instance, looking at the first row, it must hold that $\alpha_e + \alpha_i + \alpha_p \leq 1$.

2.2. Deterministic Approximation Theorem

The basic result for the deterministic approximation of a stochastic model is that the effect of "noise" (i.e. stochastic fluctuations) becomes more and more irrelevant as the population size grows larger: different random individual choices will tend to average out when many individuals are interacting. For instance, in the computer epidemic example just discussed, if the size of the network (the number of connected computers) is large, the stochastic fluctuations of the model become irrelevant, and a deterministic description will capture all the relevant features of the dynamics (see Figure 2 for a visual insight). Furthermore, the change of state of a single agent in a large population will have negligible impact on the value of normalised counting variables. Hence, it makes sense to approximate them by continuous quantities.

The deterministic approximation theorem, therefore, will approximate the dynamics of the normalised counting variables by a deterministic system in continuous space but in discrete time, hence a Discrete Time Dynamical System (DTDS). Formally, a DTDS on a subset $E \subseteq \mathbb{R}^n$ is specified by a function $f : E \to E$ and an initial point $\hat{\mathbf{x}}(0) = \hat{\mathbf{d}}_\mathbf{0}$. The DTDS $\hat{\mathbf{x}}(k)$ at step k satisfies the relation $\hat{\mathbf{x}}(k) = f(\hat{\mathbf{x}}(k-1))$, giving rise to what is usually referred to as a set of difference equations.

To further proceed in the discussion, we need to specify the DTMC dynamics in terms of the counting variables $\mathbf{X}^{(N)}$. To do this, we introduce n random vectors $\mathbf{B_i}(X_i^{(N)}, \mathbf{P}_i^{(N)}(\hat{\mathbf{X}}^{(N)}))$, for $i = 1, \dots n$, depending on the current state $\mathbf{X}^{(N)}$, each distributed according to a *multinomial distribution* on $X_i^{(N)}$ objects with probabilities given by the i-th row of the matrix $P^{(N)}$: $\mathbf{P}_i^{(N)}(\hat{\mathbf{X}}^{(N)}) = (P_{i1}^{(N)}(\hat{\mathbf{X}}^{(N)}), \dots, P_{in}^{(N)}(\hat{\mathbf{X}}^{(N)}))$. Denoting by $B_{i,j}(X_i^{(N)}, \mathbf{P}_i(\hat{\mathbf{X}}^{(N)}))$ the j-th element of the random vector $\mathbf{B_i}(X_i^{(N)}, \mathbf{P}_i(\hat{\mathbf{X}}^{(N)}))$ and by $\mathbf{X}^{(N)}(k)$ the DTMC defined in this way, we have that

$$X_j^{(N)}(k+1) = \sum_{i=1}^{n} B_{i,j}(X_i^{(N)}(k), \mathbf{P}_i^{(N)}(\hat{\mathbf{X}}^{(N)}(k))). \tag{2}$$

If we consider the normalised variables, then we can obtain an equation for $\hat{\mathbf{X}}^{(N)}(k+1)$ as a function of $\hat{\mathbf{X}}^{(N)}(k)$ from (2) as

$$\hat{X}_j^{(N)}(k+1) = \sum_{i=1}^{n} \frac{1}{N} B_{i,j}(N\hat{X}_i^{(N)}(k), \mathbf{P}_i^{(N)}(\hat{\mathbf{X}}^{(N)}(k))). \tag{3}$$

Now, assume that $P^{(N)}(\hat{\mathbf{d}})$ *converges uniformly* to a *continuous function* $P(\hat{\mathbf{d}})$, meaning that

$$\lim_{N\to\infty} \sup_{\hat{\mathbf{d}}\in\mathcal{S}^n} \|P^{(N)}(\hat{\mathbf{d}}) - P(\hat{\mathbf{d}})\| = 0.$$

If we write

$$\hat{X}_j^{(N)}(k+1) = \sum_{i=1}^{n} \hat{X}_i^{(N)}(k) \frac{1}{N\hat{X}_i^{(N)}(k)} B_{i,j}(N\hat{X}_i^{(N)}(k), \mathbf{P}_i^{(N)}(\hat{\mathbf{X}}^{(N)}(k))),$$

defining an addend to be equal to zero when $\hat{X}_i^{(N)}(k) = 0$, and we further assume that $\hat{\mathbf{X}}^{(N)}(k)$ converges almost surely to a deterministic value $\hat{\mathbf{x}}(k)$ as N goes to infinity, we can apply the law of large numbers for multinomial random variables[1] to deduce that almost surely,

$$\lim_{N\to\infty} \hat{X}_j^{(N)}(k+1) = \sum_{i=1}^{n} \hat{x}_i(k) P_{ij}(\hat{\mathbf{x}}(k)).$$

Therefore, if we define the following discrete time deterministic system

$$\begin{cases} \hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(k) \cdot \mathbf{P}(\hat{\mathbf{x}}(k)) \\ \hat{\mathbf{x}}(0) = \hat{\mathbf{d}}_{\mathbf{0}} \end{cases} \tag{4}$$

and assume that the initial conditions $\hat{\mathbf{d}}_0^{(N)}$ of the DTMC converge almost surely to $\hat{\mathbf{d}}_0$, by applying the law of large number for multinomial distributions inductively on the step k, we can prove the following theorem:

Theorem 2.1 (Mean Field Limit in Discrete Time [41])**.** *Let* $\hat{\mathbf{X}}^{(N)}(k)$ *and* $\hat{\mathbf{x}}(k)$ *be defined as above. Assume* $P^{(N)}(\hat{\mathbf{d}})$ *converges uniformly in* $\mathcal{S}^n$ *to the continuous function* $P(\hat{\mathbf{d}})$*, and that* $\hat{\mathbf{d}}_0^{(N)} \to \hat{\mathbf{d}}_0$ *almost surely. Then, for any* $k > 0$*, almost surely*

$$\lim_{N\to\infty} \hat{\mathbf{X}}^{(N)}(k) = \hat{\mathbf{x}}(k).$$

■

[1]The law of large number of multinomial random variables states that $\frac{1}{N}\mathbf{B}(N, \mathbf{p}) \to \mathbf{p}$ almost surely. Here it can be applied after a little calculus to deal with the fact that the probabilities of the multinomial distribution depend on N, but converge almost surely to the continuous limit $P_i(\hat{\mathbf{x}}(k))$.

The theorem states that, for any finite time horizon k, the behaviour of $\hat{\mathbf{X}}^{(N)}(k)$ can be approximated by $\hat{\mathbf{x}}(k)$.

Remark 2.1. The results in [41] are more general than those presented here, as they allow the probability matrix $P^{(N)}$ at step k to depend on the past history of the DTMC (with bounded depth). However, a limit result similar to Theorem 2.1 is proved.

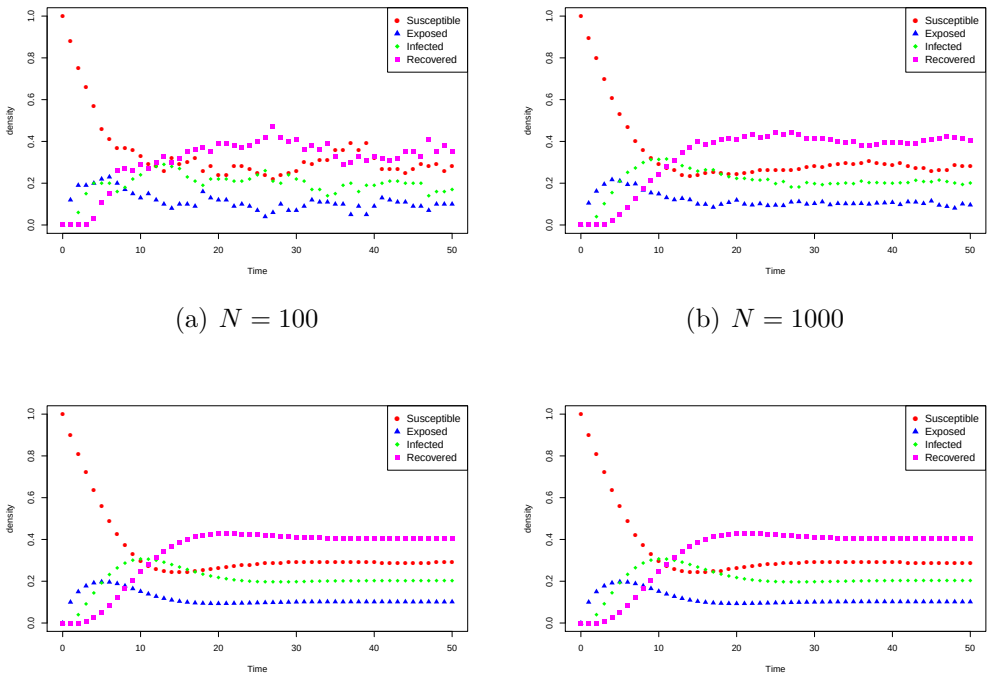
Example revisited. We now turn back to the example of the computer network epidemic. It is easy to see that the probability matrix P does not depend on N and it is continuous, hence, assuming initial conditions converge, we satisfy the hypothesis of Theorem 2.1. Therefore, the synchronous DTMC is asymptotically approximated by the following deterministic system:

$$\begin{cases} \hat{x}_s(k+1) = \hat{x}_s(k) - \alpha_e \cdot \hat{x}_s(k) - \alpha_i \cdot \hat{x}_i(k) \cdot \hat{x}_s(k) - \alpha_p \cdot \hat{x}_s(k) + \alpha_s \cdot \hat{x}_r(k) \\ \hat{x}_e(k+1) = \hat{x}_e(k) + \alpha_e \cdot \hat{x}_s(k) + \alpha_i \cdot \hat{x}_i(k) \cdot \hat{x}_s(k) - \alpha_a \cdot \hat{x}_e(k) - \alpha_q \cdot \hat{x}_e(k) \\ \hat{x}_i(k+1) = \hat{x}_i(k) + \alpha_a \cdot \hat{x}_e(k) - \alpha_r \cdot \hat{x}_i(k) \\ \hat{x}_r(k+1) = \hat{x}_r(k) + \alpha_p \cdot \hat{x}_s(k) + \alpha_q \cdot \hat{x}_e(k) + \alpha_r \cdot \hat{x}_i(k) - \alpha_s \cdot \hat{x}_r(k). \end{cases}$$

A visual representation of this result is given in Figure 2, where we compare trajectories of the DTMC for increasing N with the trajectory of the deterministic limit.

*Remark** 2.2. In this section, we considered a restricted form of interaction between the objects constituting a given system. Essentially, the interaction between two objects is mediated by the environment (more precisely, by sensing the global distribution of object's states), and direct cooperation is not allowed. This restriction has the advantage of simplifying the description of the functional form of interaction probabilities as well as simplifying the description of the one-step evolution of the system (cf. equation (2)). Considering direct forms of cooperation, involving two or more objects, makes the matter much more complex. One approach in this direction is the one of [48], where the authors define a mean field semantics for a (clock-)synchronous process algebra (WSCCS — Weighted Synchronous Calculus of Communicating Systems) which encompasses a (restricted) form of direct interaction. However, this introduces some notational overhead, hence we omitted it.

Dealing with birth and death events in this context, instead, is less problematic, but requires a different formalisation of the DTMC. In that case, it is more convenient to work directly with (normalised) counting variables, and include birth and death events in equation (2), so that a fraction of agents



(a) $N = 100$ (b) $N = 1000$

(c) $N = 10^6$ (d) Det. Limit

Figure 2: Comparison between the deterministic system and stochastic trajectories of synchronous DTMC of the network epidemic example for increasing population sizes. Parameters of the model are $\alpha_e = 0.1$, $\alpha_i = 0.2$, $\alpha_a = 0.4$, $\alpha_r = 0.2$, $\alpha_p = 0$, $\alpha_q = 0$, and $\alpha_s = 0.1$, while initial conditions are $\hat{X}_s(0) = 1$, $\hat{X}_e(0) = 0$, $\hat{X}_i(0) = 0$, and $\hat{X}_r(0) = 0$. All simulations were performed by a dedicated R implementation. All charts in the paper were generated using R.

is removed at each step (death events) and another fraction is added (birth events). In this case, however, normalised variables no longer represent a probability distribution, but rather they count the fraction of objects in any state, relatively to the initial population size. This kind of approach is typical in theoretical ecology, where discrete time mean field equations are often used to describe the evolution of a well-mixed population. In this context, the normalised variables are referred to as population density [49].

3. Basics of Deterministic Approximation in Continuous Time

In the previous section we considered a DTMC model in which all agents move synchronously at each step. A CTMC model, instead, is inherently asynchronous, because the probability of two events happening at the same time is zero, due to the continuity of the exponential distribution. Intuitively, for these asynchronous models, when we increase the size of the population, we are also increasing the number of potential events that can be observed (think of the infection of a computer in the network epidemic model, in which each infection happens at a different time). Hence, the larger the population, the more events will be observed per unit of time. In the limit of an infinite population, the density of events will be infinite, and the change in normalised counting variables induced by one event infinitesimal: discrete jumps are now replaced by a continuous flow. Hence, for CTMCs, limit processes are flows in continuous time, in fact solutions of Ordinary Differential Equations (ODE).

Similarly, we can consider a DTMC model in which *only a constant number of agents* moves at each discrete step. The integer time of these models is an abstract time, which can be embedded in the real continuous time, associating a constant duration with each step of the DTMC. We will then decrease this duration proportionally with the population growth, in order to obtain a limit process. We can intuitively understand the rationale of this operation if we think of these DTMC models as sampling the state of a system at a constant frequency. In this case, when we increase the population size, we are increasing the density of events, hence, if we want to observe them all, we also need to increase the sampling frequency, i.e. to reduce the duration between two consecutive samplings. The limit process will again be a continuous flow in continuous time.

3.1. Modelling language

In order to formalise these intuitions, and to discuss precisely under which assumptions there is a deterministic limit, we need to introduce some basic concepts but also to fix the notation to describe CTMC and the above-mentioned asynchronous DTMC. The first step in this direction is to present a simple modelling language, that is close to the descriptive level of Markov Chains (MCs), but facilitates a compact description of population processes at the *system level*, by representing the different possible transitions parametrically. We will define a language framework that can be used to specify

both population CTMC models and, with minor modifications, population DTMC models[2].

Population Continuous Time Markov Chain Models. We will start our discussion from Population Continuous Time Markov Chain (PCTMC) models.

Definition 3.1. A Population Continuous Time Markov Chain (PCTMC) model is a tuple[3] $\mathcal{X}_C = (\mathbf{X}, \mathcal{D}, \mathcal{T}, \mathbf{d}_0)$, where:

1. $\mathbf{X} = (X_1, \ldots, X_n)$ is a vector of *variables*.
2. Each X_i takes values in a *finite* or *countable* domain $\mathcal{D}_i$. We usually assume that $\mathcal{D}_i \subset \mathbb{R}$. Hence, $\mathcal{D} = \prod_i \mathcal{D}_i$ is the *state space* of the model.
3. $\mathbf{d}_0 \in \mathcal{D}$ is the *initial state* of the model.
4. $\mathcal{T} = \{\tau_1, \ldots, \tau_m\}$ is the set of *transitions* of the form $\tau_j = (a, \mathbf{s}, \mathbf{t}, r)$, where:
   (a) a is the *label* of the transition;
   (b) $\mathbf{s} \in \mathbb{R}^n$, $\mathbf{s} \geq \mathbf{0}$ is the *pre-vector*, i.e. a vector of non-negative components specifying how many units of each variable are consumed by the transition.
   (c) $\mathbf{t} \in \mathbb{R}^n$, $\mathbf{t} \geq \mathbf{0}$ is the *post-vector*, i.e. a vector of non-negative components specifying how many units of each variable are created by the transition.
   (d) $r : \mathcal{D} \to \mathbb{R}_{\geq 0}$ is the *rate function* of the transition. We require $r(\mathbf{d}) = 0$ whenever $\mathbf{d} + \mathbf{t} - \mathbf{s} \notin \mathcal{D}$.

   The pre-vectors and post-vectors are combined in the *state-change* vector $\mathbf{v} = \mathbf{t} - \mathbf{s}$, giving the net change on each variable due to the transition. □

[2]The language is similar to the one of the PRISM model checker [50]; it resembles also the language used in [51]. The main difference is that our language lacks an explicit treatment of guards, which can be incorporated into the functions associated with transitions (as we will see in Section 4.5). We chose to use a low level modelling language, instead of a more expressive stochastic process algebra (SPA), because most of the approaches to deterministic approximation are formulated directly in terms of MCs, and the relevant ideas are better captured at this level. Higher level languages, like SPAs, can be mapped to this language in a more or less straightforward way. For instance, SPAs with a fluid semantics in terms of ODEs can be easily mapped to this language along the lines of the ODE derivation (see, for instance, [52, 53]). Similar translations can be defined for probabilistic process algebras and PDTMCs.

[3]We will drop the subscript from $\mathcal{X}_C$ when there is no chance of confusion with population DTMC models.

For transition $\tau = (a, \mathbf{s}, \mathbf{t}, r)$ we let a_τ denote the label of τ and similarly for the other components. In the context of this tutorial, it is convenient to make models depend on an index N, which represents the total population (or it can be associated with a more general notion of size, see below). We will use N to index a sequence of models for increasing populations. We denote this by $\mathcal{X}^{(N)}$, where it is intended that each component of $\mathcal{X}^{(N)}$ may depend on N, thus we get $\mathcal{D}^{(N)}$, $r_\tau^{(N)}$, etc.

Given a PCTMC model $\mathcal{X}^{(N)}$, we can easily extract the associated CTMC, by specifying the state space and the infinitesimal generator matrix $\mathbf{Q}$. The former is simply the set $\mathcal{D}^{(N)}$, while the latter is defined by adding up all the rates inducing the same state change. Formally:

$$\mathbf{Q}(\mathbf{d}, \mathbf{d}') = \sum_{\tau \in \mathcal{T}^{(N)} | \mathbf{v}_\tau = \mathbf{d}' - \mathbf{d}} r_\tau^{(N)}(\mathbf{d}), \quad \mathbf{d} \neq \mathbf{d}'.$$

When the summation set is empty, we assume the corresponding rate to be zero.

Population Discrete Time Markov Chain Models. We can describe Population Discrete Time Markov Chain (PDTMC) models in a similar way to PCTMC ones, simply by replacing the rate of a transition by its probability.

Definition 3.2. A Population Discrete Time Markov Chain (PDTMC) model is a tuple $\mathcal{X}_D = (\mathbf{X}, \mathcal{D}, \mathcal{T}, \mathbf{d}_0)$, where:

1. $\mathbf{X}$, $\mathcal{D}$, and $\mathbf{d}_0$ are as in Definition 3.1;
2. $\mathcal{T} = \{\tau_1, \ldots, \tau_m\}$ is the set of *transitions*, of the form $\tau_j = (a, \mathbf{s}, \mathbf{t}, p)$, where:
   (a) a, $\mathbf{s}$, $\mathbf{t}$ are as in Definition 3.1;
   (b) $p : \mathcal{D} \to \mathbb{R}_{\geq 0}$ is the *probability function* of the transition, and defines a sub-probability distribution, i.e. for each $\mathbf{d} \in \mathcal{D}$, it is required that: $\sum_{\tau \in \mathcal{T}} p_\tau(\mathbf{d}) \leq 1$. □

The DTMC associated with a PDTMC model $\mathcal{X}_D^{(N)}$ has state space $\mathcal{D}^{(N)}$ and probabilistic transition matrix $\mathbf{P}$, defined using the transition probability functions, as follows:

$$\mathbf{P}(\mathbf{d}, \mathbf{d}') = \sum_{\tau \in \mathcal{T}^{(N)} | \mathbf{v}_\tau = \mathbf{d}' - \mathbf{d}} p_\tau^{(N)}(\mathbf{d}), \quad \mathbf{d} \neq \mathbf{d}',$$

i.e. we add the probability of all transitions changing state from $\mathbf{d}$ to $\mathbf{d}'$. As the transition functions form a sub-probability distribution in each state, we must concentrate the remaining probability mass in the identity transition:

$$\mathbf{P}(\mathbf{d},\mathbf{d}) = 1 \; - \sum_{\tau \in \mathcal{T}^{(N)} | \mathbf{v}_\tau \neq \mathbf{0}} p_\tau^{(N)}(\mathbf{d}).$$

3.2. Main Example (continued)

We turn back to the example of Section 2.1, showing how to describe the epidemic model as a PCTMC or a PDTMC model. The language presented here describes the system from a global point of view. Hence, we will use variables to *count* how many instances of each agent type are in the system (e.g. how many susceptible or exposed computers are in the network). Transitions will model a single event, like the infection or the patching of a single computer. In the case of PCTMC, the rate of a transition is the global rate of observing it. In the PDTMC case, the probability of a transition is the probability that the next event is the one described by the transition itself. Recall that, in the PDTMC case, we are modelling an *asynchronous* DTMC, so that in each step only one event happens. For instance, in the epidemic model, at each step only one computer is infected, or one is patched, and so on.

PCTMC model of the network epidemic. To construct a PCTMC model we need four variables, counting how many agents are in each state: susceptible X_s, exposed X_e, infected X_i, or recovered X_r, with $\mathbf{X} = (X_s, X_e, X_i, X_r)$. The state space is $\mathcal{D}^{(N)} = \{0, \ldots, N\}^4$, with the additional constraint that the sum of agents is constantly equal to N. To specify the transitions, instead, we will use the unit vectors $\mathbf{e_S} =_{\text{def}} (1,0,0,0), \mathbf{e_E} =_{\text{def}} (0,1,0,0), \mathbf{e_I} =_{\text{def}} (0,0,1,0)$, and $\mathbf{e_R} =_{\text{def}} (0,0,0,1)$.

- Infection of a susceptible node from an external source: $\tau_e = (\text{ext}, \mathbf{e_S}, \mathbf{e_E}, r_e^{(N)})$, with $r_e^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_e \cdot X_s$. Notice that the infection rate is proportional to the number of susceptibles, as we have to add up the infection rate[4] for each individual to properly describe the rate at which we observe these events at the network level.

[4]To make the notation lighter, we write the rate function of transition τ_e as $r_e^{(N)}$ rather than as $r_{\tau_e}^{(N)}$, and similarly for other transitions.

- Infection of a susceptible node from a malicious contact with an infected node: $\tau_i = (\text{inf}, \mathbf{e_S}, \mathbf{e_E}, r_i^{(N)})$, with $r_i^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_i \cdot \frac{X_s}{N} \cdot X_i$. Here λ_i is the basic rate of infection, i.e. the rate at which a single infected node can spread the infection by contacting other nodes. This rate is then multiplied by the probability that a random contact is with a susceptible node ($\frac{X_s}{N}$). A second multiplication by the number of infected nodes gives the global rate of infection.

- Activation of the infection in an exposed node: $\tau_a = (\text{act}, \mathbf{e_E}, \mathbf{e_I}, r_a^{(N)})$, with $r_a^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_a \cdot X_e$.

- Patching of an infected node: $\tau_r = (\text{rec}, \mathbf{e_I}, \mathbf{e_R}, r_r^{(N)})$, with $r_r^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_r \cdot X_i$.

- Patching of a non-infected node: $\tau_p = (\text{rec}, \mathbf{e_S}, \mathbf{e_R}, r_p^{(N)})$, and $\tau_q = (\text{rec}, \mathbf{e_E}, \mathbf{e_R}, r_q^{(N)})$, with $r_p^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_p \cdot X_s$ and $r_q^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_q \cdot X_e$ (here $\lambda_p, \lambda_q < \lambda_r$).

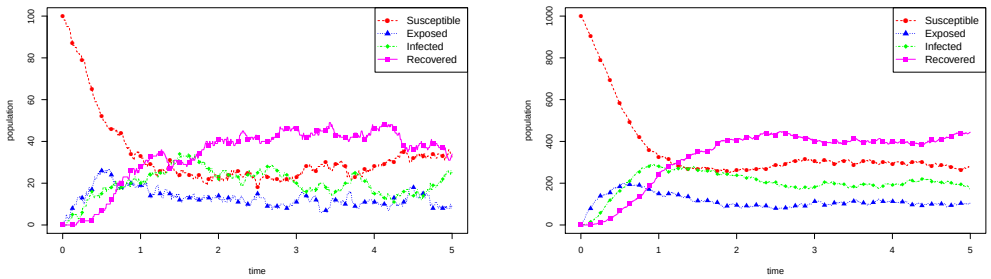
- Loss of immunity of a recovered node: $\tau_s = (\text{rec}, \mathbf{e_R}, \mathbf{e_S}, r_s^{(N)})$, with $r_s^{(N)}(\mathbf{X}) = \lambda_s \cdot X_r$.

We will indicate by $\mathcal{E}_C^{(N)}$ the PCTMC epidemic model with N nodes in the network. Two simulation trajectories for the CTMC model are shown in Figure 3, for two different population levels.

PDTMC model of the network epidemic. The PDTMC model $\mathcal{E}_D^{(N)}$ of the network epidemic is obtained by replacing rates with probabilities in the previous list of transitions as follows:

- Infection from an external source. We assume that each susceptible node is infected in any step with a constant probability α_e. Hence, the probability that we observe an external infection is α_e times the probability that a susceptible node is involved in the next transition, given that each node can be chosen with the same probability: $p_e^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_e \cdot \frac{X_s}{N}$.

- Infection from a malicious contact. A message sent from an infected node will arrive at a susceptible node with probability $\frac{X_s}{N}$. In this case, the infection will happen with probability α_i. Hence, the probability of observing an infection due to a malicious contact is $p_i^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_i \cdot \frac{X_s}{N} \cdot \frac{X_i}{N}$.



(a) $N = 100$ (b) $N = 1000$

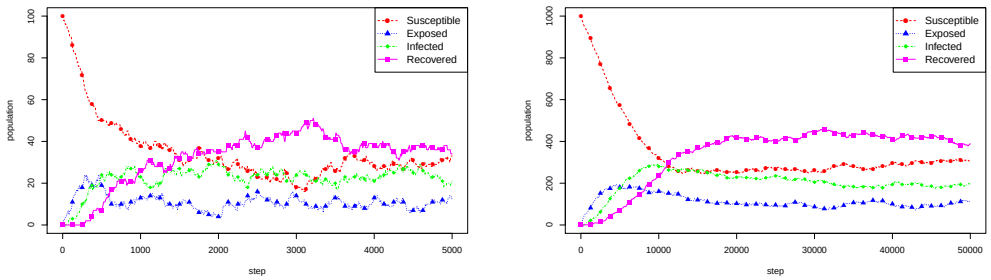
Figure 3: Trajectories of the PCTMC model of the network infection example of Section 2.1, for different population levels. The parameters of the model are $\lambda_e = 1$, $\lambda_i = 2$, $\lambda_a = 4$, $\lambda_r = 2$, $\lambda_p = 0$, $\lambda_q = 0$, and $\lambda_s = 1$, with initial conditions $S_0 = N$, $E_0 = 0$, $I_0 = 0$, and $R_0 = 0$. All the simulations for PCTMCs were performed with Dizzy [54].

- Activation of the infection. A worm in an exposed node will activate with probability α_a, so that $p_a^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_a \cdot \frac{X_e}{N}$.
- Patching of an infected node. Each infected node will be patched with probability α_r, so that $p_r^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_r \cdot \frac{X_i}{N}$.
- Patching of a non-infected node. Non-infected nodes can be patched with probability $\alpha_p, \alpha_q < \alpha_r$, so that $p_p^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_p \cdot \frac{X_s}{N}$ and $p_q^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_q \cdot \frac{X_e}{N}$.
- Loss of immunity. Each recovered node will lose immunity with probability α_s, so that $p_s^{(N)}(\mathbf{X}) =_{\text{def}} \alpha_s \cdot \frac{X_r}{N}$.

The specific values of α constants must be defined in order to study the evolution model. They must satisfy some constraints to make the model compatible with the DTMC semantics based on probabilities. In fact, these constraints are the same as the ones in Section 2.1. In Figure 4 we show a simulated trajectory for the PDTMC model for two different population levels.

3.3. System size and normalisation

In order to give a formal account of the assertion that, as the size of the system grows, a deterministic description will capture the relevant dynamics, we need two things: a notion of the "size" of the system, and a



(a) $N = 100$ (b) $N = 1000$

Figure 4: Trajectories of the PDTMC model of the network infection example of Section 2.1, for different population levels. The parameters of the model are $\alpha_e = 0.1$, $\alpha_i = 0.2$, $\alpha_a = 0.4$, $\alpha_r = 0.2$, $\alpha_p = 0$, $\alpha_q = 0$, and $\alpha_s = 0.1$, with initial conditions $S_0 = N$, $E_0 = 0$, $I_0 = 0$, and $R_0 = 0$. All the simulations of PDTMCs were performed by a dedicated Java implementation.

limit theorem stating that the stochastic model converges to its deterministic counterpart (to be defined, too), as the size of the system goes to infinity. All the approximation results in literature that we will consider follow this scheme [14, 29, 55, 56, 32, 57, 42].

Given a PCTMC or a PDTMC model $\mathcal{X}^{(N)} =_{\text{def}} (\mathbf{X}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}^{(N)}, \mathbf{d_0}^{(N)})$, indexed by N, we associate with it the size γ_N, which is a positive real number, for each N. In the context of this paper, the size of a system will always be the total population N, i.e. $\gamma_N = N$. In the epidemic model $\mathcal{E}^{(N)}$, for instance, the size is the number of computers in the network, which is a quantity that remains constant throughout the evolution of the epidemic.

If the total population of the model can vary, because there are birth and death events (e.g. the connection or disconnection of a computer in the network), the size γ_N can no longer be equal to the total population, since γ_N is required to be a constant. In such cases, the size of the system is often chosen to be the total *initial* population.

The notion of system size is by no means limited to a population level, as variables are not required to take integer values. For instance, in models of biochemical reactions, the system size is usually the volume of the container in which the reactions happen (multiplied by the Avogadro constant). In dividing the variables representing molecular counts by this volume, we obtain the molar concentration of reactants.

The notion of size is of course intimately related to that of normalisation: as we have already seen in Section 2, in order to compare the evolution for different population levels, we divide each variable by γ_N. When the total population of a model is constant, the normalized variables can be interpreted as a probability distribution; when instead there are births and deaths, the normalized variables cannot be interpreted as a probability distribution, but rather as a proportion of agents relative to the initial population.

If $\gamma_N = N$, by normalising counting variables with respect to N, the unit increment in the normalized variables is of the order of $\frac{1}{N}$; hence we can think of our process as being defined in a grid of width $\frac{1}{N}$ in $[0, 1]$. As N increases, the step size of this grid becomes smaller and smaller, and the limit process will live in the continuous world.

We turn now to a more precise description of the formal setting in which we will discuss deterministic limits. Suppose we have a sequence of PCTMC models $(\mathcal{X}^{(N)})_{N \geq N_0}$, where $N \in \mathbb{N}$, $N_0 > 1$ is a problem-specific value[5], and $\mathcal{X}^{(N)} = (\mathbf{X}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}^{(N)}, \mathbf{d_0}^{(N)})$. The index N is closely related to the size γ_N of $\mathcal{X}^{(N)}$: we require $\lim_{N\to\infty} \gamma_N = \infty$. Normalization essentially consists of dividing variables by γ_N, modifying rates and increments accordingly. In this way, the basic step increment at level N becomes $\delta_N =_{\text{def}} \frac{1}{\gamma_N}$ with $\lim_{N\to\infty} \delta_N = 0$. For instance, in the computer network epidemic model, the range of each variable is $\{0, \ldots, N\}$ and $\delta_N = \frac{1}{N}$.

When we rescale a model according to this recipe, we need to rescale appropriately all the quantities involved; we do this as described below. For all (PCTMC) models $\mathcal{X}^{(N)} = (\mathbf{X}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}^{(N)}, \mathbf{d_0}^{(N)})$, we define a *normalising* operator $(\hat{\cdot})$ as follows:

- for all $\mathbf{d} \in \mathcal{D}^{(N)}$, $\hat{\mathbf{d}} =_{\text{def}} \delta_N \cdot \mathbf{d}$;
- $\hat{\mathcal{D}}^{(N)} =_{\text{def}} \{\hat{\mathbf{d}} \mid \mathbf{d} \in \mathcal{D}^{(N)}\}$;
- for all $\tau \in \mathcal{T}^{(N)}$, with $\tau = (a, \mathbf{s}^{(N)}, \mathbf{t}^{(N)}, r^{(N)})$, $\hat{\tau} =_{\text{def}} (a, \hat{\mathbf{s}}^{(N)}, \hat{\mathbf{t}}^{(N)}, \hat{r}^{(N)})$, where:
  - $\hat{\mathbf{s}}^{(N)} =_{\text{def}} \delta_N \cdot \mathbf{s}^{(N)}$;
  - $\hat{\mathbf{t}}^{(N)} =_{\text{def}} \delta_N \cdot \mathbf{t}^{(N)}$;

[5]For instance, in the epidemic example we would have $N_0 = 2$, in order to allow interactions to take place.

  - Consequently, $\hat{\mathbf{v}}^{(N)} =_{\text{def}} \delta_N \cdot \mathbf{v}^{(N)}$;
  - for all $\hat{\mathbf{d}} \in \hat{\mathcal{D}}, \hat{r}^{(N)}(\hat{\mathbf{d}}) =_{\text{def}} r^{(N)}(\gamma_N \cdot \hat{\mathbf{d}})$;
- $\hat{\mathcal{T}}^{(N)} =_{\text{def}} \{\hat{\tau} \mid \tau \in \mathcal{T}^{(N)}\}$.

Given a model $\mathcal{X}^{(N)} = (\mathbf{X}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}^{(N)}, \mathbf{d_0}^{(N)})$, the corresponding *normalized* model $\hat{\mathcal{X}}^{(N)}$ is then $\hat{\mathcal{X}}^{(N)} =_{\text{def}} (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$, where $\hat{\mathbf{X}}^{(N)}$ is a fresh new variable tuple of the same size as $\mathbf{X}^{(N)}$.

Notice that the following relationship holds between the variables of the normalized model and those of the non-normalized one: $\hat{\mathbf{X}}^{(N)} = \delta_N \cdot \mathbf{X}^{(N)} = \frac{1}{\gamma_N} \cdot \mathbf{X}^{(N)}$.

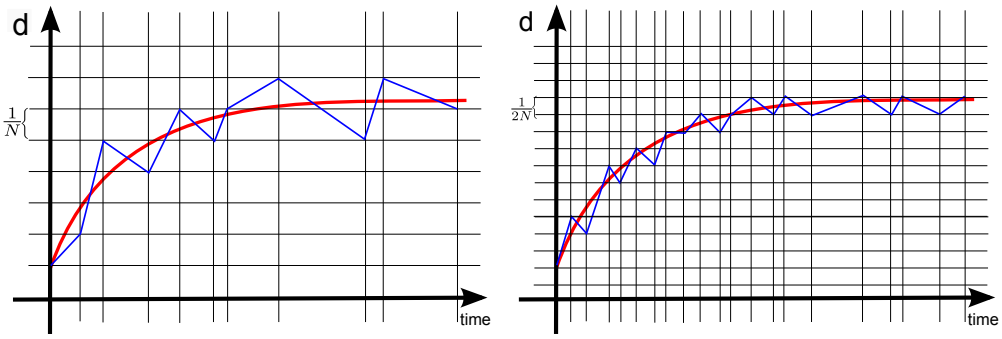
The normalisation $\hat{\mathcal{X}}_D$ of a PDTMC model $\mathcal{X}_D$ is defined similarly, just replacing rates with probabilities.

In the following, we will first state the main theorems for PCTMC models. This will be done in the next section, where we will also discuss several examples. Then, we will turn our attention to results for PDTMC models. We will interpret PDTMCs as *timed* models, where the duration of each step of a DTMC of level N is ϵ_N, again presenting both theorems and examples. In particular, in Section 6 we will show that the asymptotic behaviour of such PDTMCs and PCTMCs, whenever certain conditions of probabilities and rates are fulfilled, is essentially the same: in the limit of large N, the difference between these two classes of processes is negligible.

After presenting the limit theorems, we will discuss more advanced topics, like the limit results for the stationary regime and the decoupling assumption (Sections 7 and 8).

4. Deterministic approximation for PCTMCs

The main results on deterministic approximation for continuous time Markov Chains date back at least to the work of Kurtz [14, 47]. The key concept is a sequence of Markov processes for which both the magnitude of jumps, i.e. $||\mathbf{v}_\tau^{(N)}||$, for all transitions τ, and the average time between consecutive jumps, go to zero. In this situation, fluctuations become negligible, and as the time step and magnitude become infinitesimal, we can approximate a discrete jump with a continuous derivative, thus obtaining a deterministic model in terms of ordinary differential equations. This situation is graphically illustrated in Figure 5.



(a) (b)

Figure 5: Intuitive graphical visualisation of deterministic approximation theorems. As the step size and the magnitude of jumps go to zero, the impact of fluctuations becomes negligible, hence stochastic trajectories start to look like the smooth deterministic one. Here we represented two situations, halving the step size and magnitude in passing from Figure 5(a) to Figure 5(b).

In order for these results to hold, the time step and the jump magnitude must go to zero in a consistent way, i.e. they must scale in the same way with respect to the system size parameter γ_N. One classical scaling condition requirement is *density dependence*, discussed below, which is satisfied by many models, such as most models of chemical reactions and most epidemic models. However, the conditions under which the limit theorem holds are far more general. Here we will first discuss deterministic approximation for density dependent models, and present subsequently a more general formulation, following [56] and [57]. These papers, in fact, present a formulation which is a good compromise between generality and simplicity in verifying conditions that must hold for the theorem to work. In particular, [57] extends the validity of the results in [56], also providing an explicit formula for error bounds, which are shown to decay exponentially with N.

We decided to begin the presentation with PCTMCs instead of PDTMCs for two reasons. Firstly, continuous time must enter the picture in this class of deterministic approximation results, and in CTMCs continuous time is naturally considered, whilst introducing it in DTMCs is slightly less intuitive. Secondly, it is quite easy to deduce the limit results for DTMCs from those for CTMCs [58], and we will pursue that line of reasoning here.

4.1. Deterministic approximation for density dependent models

Let $\mathcal{X}_C^{(N)} = (\mathbf{X}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}^{(N)}, \mathbf{d_0}^{(N)})$ be a PCTMC model and $\hat{\mathcal{X}}_C^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_\mathbf{0}^{(N)})$ be the corresponding normalised model. We will consider the sequence $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$ with respect to an increasing system size γ_N.

Density dependence is a typical scaling law of rate functions. To formally define density dependence, let E be a *closed* set in $\mathbb{R}^n$ such that $\bigcup_N \hat{\mathcal{D}}^{(N)} \subseteq E$. This is the space in which all processes of the sequence of normalised PCTMC and their deterministic approximation live. The sequence of normalised PCTMC models $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$ is density dependent if and only if:

- the system size grows linearly with N, i.e. $\gamma_N = \Theta(N)$, and
- for each N and $\tau \in \hat{\mathcal{T}}^{(N)}$:
  - there is a vector $\mathbf{v}_\tau$ such that $\hat{\mathbf{v}}_\tau^{(N)} = \frac{1}{\gamma_N}\mathbf{v}_\tau$. This means that increments $\mathbf{v}_\tau^{(N)}$ in the non-normalised model are independent of N: $\mathbf{v}_\tau^{(N)} = \mathbf{v}_\tau$;
  - there is a Lipschitz continuous and bounded function[6] $g_\tau : E \to \mathbb{R}_{\geq 0}$ such that the rate function $\hat{r}_\tau^{(N)} : \hat{\mathcal{D}}^{(N)} \to \mathbb{R}$ scales with system size as
  $$\hat{r}_\tau^{(N)}(\hat{\mathbf{d}}) = \gamma_N \cdot g_\tau(\hat{\mathbf{d}}),$$
  for all $\hat{\mathbf{d}} \in \hat{\mathcal{D}}^{(N)}$.

The drift $F : E \to \mathbb{R}^n$ of a generic element of the sequence of normalised PCTMC models $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$describes the mean instantaneous increment from a state $\hat{\mathbf{d}}$ of the model. For *density dependent* models, the drift can be defined as
$$F(\hat{\mathbf{d}}) =_{\text{def}} \sum_{\tau \in \mathcal{T}^{(N)}} \mathbf{v}_\tau \cdot g_\tau(\hat{\mathbf{d}}),$$

[6]A function $g(\hat{\mathbf{d}})$ is Lipschitz continuous if and only if there is a constant $L > 0$ such that $\|g(\hat{\mathbf{d}}_1) - g(\hat{\mathbf{d}}_2)\| \leq L\|\hat{\mathbf{d}}_1 - \hat{\mathbf{d}}_2\|$. A function $g(\hat{\mathbf{d}})$ is bounded if there is $M > 0$ such that $g(\hat{\mathbf{d}}) \leq M$ for all $\hat{\mathbf{d}} \in E$. The Lipschitz condition can be relaxed by requiring that g is locally Lipschitz continuous, i.e. for each compact $K \subset E$, g is Lipschitz in K, with constant L_K. Similarly, we can relax boundedness to local boundedness.

The drift defines the vector field of the deterministic limit, whose trajectories $\hat{\mathbf{x}}(t)$ are solutions of the initial value problem $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$, with $\hat{\mathbf{x}}(0) = \hat{\mathbf{d}}_{\mathbf{0}} \in E$. We require that $\hat{\mathbf{x}}(t)$ remains in E (for all time instants in which $\hat{\mathbf{x}}$ is defined). This can be accomplished by choosing E appropriately.

We further require that there is some point $\hat{\mathbf{d}}_{\mathbf{0}} \in E$ such that

$$\lim_{N\to\infty} \hat{\mathbf{d}}_{\mathbf{0}}^{(N)} = \hat{\mathbf{d}}_{\mathbf{0}}, \tag{5}$$

i.e. the initial conditions of $\hat{\mathcal{X}}_C^{(N)}$ converge.[7]

Under this assumption, and letting $\hat{\mathbf{X}}^{(N)}(t)$ be the continuous-time Markov process associated with $\hat{\mathcal{X}}_C^{(N)}$, the following theorem holds:

Theorem 4.1 (Deterministic approximation for density dependent PCTMCs)**.** *Let $(\hat{\mathbf{X}}^{(N)}(t))_{N\geq N_0}$ be the sequence of Markov processes associated with the density dependent sequence of PCTMC models $(\hat{\mathcal{X}}_C^{(N)})_{N\geq N_0}$, $\hat{\mathbf{x}}(t)$ as above, and assume Condition (5) is satisfied. Then, for any* finite *time horizon $T < \infty$, it holds that:*

$$\mathbb{P}\{\lim_{N\to\infty} \sup_{0\leq t\leq T} ||\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)|| = 0\} = 1,$$

i.e. $\sup_{0\leq t\leq T} ||\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)||$ *converges to zero almost surely.* ■

This theorem is a law of large numbers for Markov processes, and a reasonably simple proof will be presented in Appendix B. It states that, whenever the appropriate assumptions are in place, the probability of observing a significant difference between *any* single trajectory of the Markov process and the solution of the ODE goes to zero as N grows. Essentially, for large N *we cannot distinguish the individual trajectories of the CTMCs from the trajectory of the ODE.* The validity of this theorem is for any finite time horizon T, chosen so that the solution of the ODE remains in E in $[0, T]$. The behaviour as $T \to \infty$ will be discussed in Section 7.

[7] In this case, convergence of initial conditions is sure, as the sequence of PCTMC models starts from a fixed point. When the sequence of PCTMC models starts from an initial distribution, which is a possibility not encompassed by our modelling language but can be easily included, then one should require almost sure convergence to the limit point $\hat{\mathbf{d}}_{\mathbf{0}}$.

4.2. Main Example (continued)

We return to the main example of the computer network epidemic. Recall that we have a fixed number of nodes, N, in the network, each being in one of four different states: susceptible (with count X_s), exposed (X_e), infected (X_i), and recovered (X_r). The interactions and their rates are described in Section 3.2.

In order to apply the deterministic approximation theorem, we need to check that the rates are density dependent. In this case, the system size is simply $\gamma_N = N$, and the sequence of normalised models $(\hat{\mathcal{E}}_C^{(N)})_{N \geq N_0}$ is derived according to the recipe described in Section 3.

Notice that, as $\hat{X}_s + \hat{X}_e + \hat{X}_i + \hat{X}_r = 1$, each $\hat{\mathcal{D}}^{(N)}$ is contained in the unit simplex $\mathcal{S}^4 = \{\hat{\mathbf{d}} \in [0,1]^4 \mid \sum_{i=1}^4 \hat{d}_i = 1\}$, hence we can take $E =_{\text{def}} \mathcal{S}^4$. Inspecting the model, it is straightforward to verify the density dependence condition on update vectors and rate functions. In fact, the update vectors in the non-normalised model are clearly independent of N, while for the rate function we have:

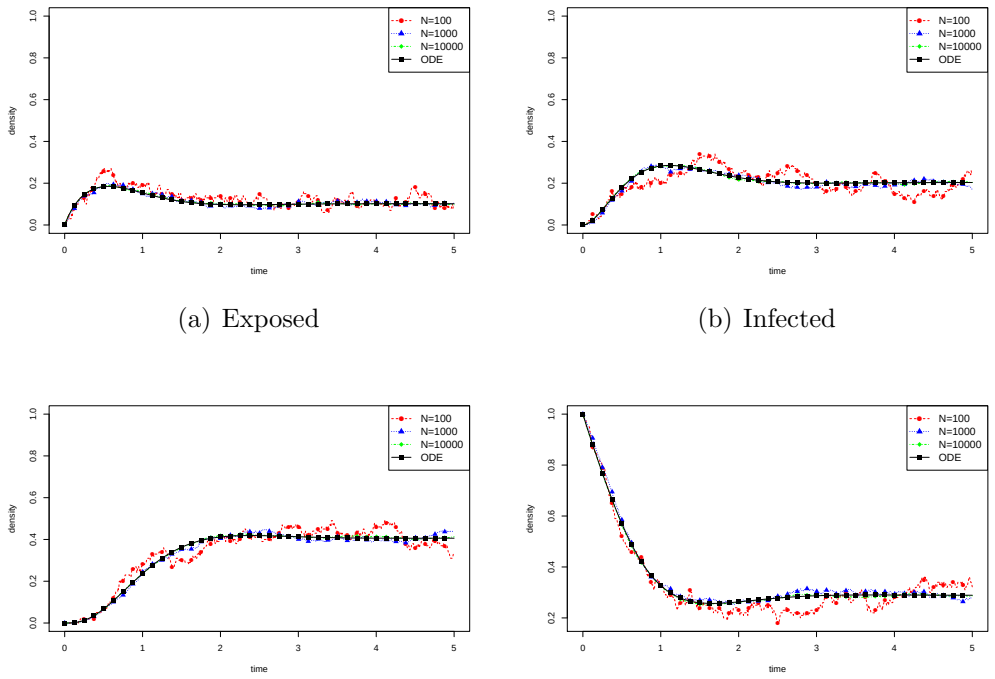
- The external infection rate is $\hat{r}_e^{(N)}(\hat{\mathbf{X}}) =_{\text{def}} N \cdot \lambda_e \cdot \hat{X}_s$, which is density dependent with $g_e(\hat{\mathbf{X}}) = \lambda_e \cdot \hat{X}_s$;
- The infection by contact rate is $\hat{r}_i^{(N)}(\hat{\mathbf{X}}) =_{\text{def}} N \cdot \lambda_i \cdot \hat{X}_s \cdot \hat{X}_i$, which is density dependent with $g_i(\hat{\mathbf{X}}) = \lambda_i \cdot \hat{X}_s \cdot \hat{X}_i$.

The other rates have the same form as the external infection, hence they are all density dependent. Notice that all rate functions satisfy the Lipschitz condition, as they are continuously differentiable functions in a compact set.

The limit set of ODEs $\frac{d\hat{\mathbf{x}}}{dt} = F(\hat{\mathbf{x}})$ of the sequence is easily constructed by computing the drift F according to the recipe of the previous section.

$$\begin{aligned}
\tfrac{d\hat{x}_s(t)}{dt} &= -\lambda_e \cdot \hat{x}_s(t) - \lambda_i \cdot \hat{x}_s(t) \cdot \hat{x}_i(t) - \lambda_p \cdot \hat{x}_s(t) + \lambda_s \cdot \hat{x}_r(t) \\
\tfrac{d\hat{x}_e(t)}{dt} &= \lambda_e \cdot \hat{x}_s(t) + \lambda_i \cdot \hat{x}_s(t) \cdot \hat{x}_i(t) - \lambda_a \cdot \hat{x}_e(t) - \lambda_q \cdot \hat{x}_e(t) \\
\tfrac{d\hat{x}_i(t)}{dt} &= \lambda_a \cdot \hat{x}_e(t) - \lambda_r \cdot \hat{x}_i(t) \\
\tfrac{d\hat{x}_r(t)}{dt} &= \lambda_r \cdot \hat{x}_i(t) + \lambda_p \cdot \hat{x}_s(t) + \lambda_q \cdot \hat{x}_e(t) - \lambda_s \cdot \hat{x}_r(t).
\end{aligned} \tag{6}$$

Note that this set of ODEs is defined in the state space E, and its trajectories can never leave it, as $\hat{x}_s(t) + \hat{x}_e(t) + \hat{x}_i(t) + \hat{x}_r(t)$ is a conserved quantity.



(a) Exposed (b) Infected

(c) Recovered (d) Susceptible

Figure 6: Comparison between the solution of the ODE and stochastic trajectories for increasing population sizes of the PCTMC model of the network epidemic. Parameters of the model are $\lambda_e = 0.2$, $\lambda_i = 0.2$, $\lambda_a = 0.4$, $\lambda_r = 0.2$, $\lambda_p = 0$, $\lambda_q = 0$, and $\lambda_s = 0.2$, while initial conditions of ODE are $\hat{s}_0 = 1$, $\hat{e}_0 = 0$, $\hat{i}_0 = 0$, and $\hat{r}_0 = 0$.

As far as convergence of initial conditions is concerned, let $\hat{\mathbf{d}}_\mathbf{0}$ be an initial point in the state space E. We define $\hat{\mathbf{d}}_\mathbf{0}^{(N)}$ as follows: $\hat{\mathbf{d}}_\mathbf{0}^{(N)} =_{\text{def}} \frac{1}{N} \cdot \lfloor N \cdot \hat{\mathbf{d}}_\mathbf{0} \rfloor$. Clearly, $\lim_{N\to\infty} \hat{\mathbf{d}}_\mathbf{0}^{(N)} = \hat{\mathbf{d}}_\mathbf{0}$[8].

In Figure 6 we can see a comparison between the solution of the ODEs and some trajectories of the stochastic process for different population levels. As we can see, as N increases, the trajectories can no longer be distinguished from the limit ODEs.

[8]Note that for N sufficiently large, the rounding error introduced by the floor function will become negligible.

4.3. Deterministic Approximation: General Formulation

We turn now to state more general conditions for the deterministic approximation theorem to hold. After introducing some notation, we present the scaling conditions and the limit theorems.

4.3.1. Notation

Let $\hat{\mathcal{X}}_C^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_0^{(N)})$ be a normalised PCTMC model. We define some derived quantities of interest, that will be needed later on. The point here is that the simple description of the PCTMC model in terms of transitions allows the computation of these quantities in a straightforward way.

- The *exit rate* function $R^{(N)} : \hat{\mathcal{D}}^{(N)} \to \mathbb{R}$ associates each state $\hat{\mathbf{d}}$ with its exit rate, which is the speed at which we see some transition happening when the system is in state $\hat{\mathbf{d}}$:

$$R^{(N)}(\hat{\mathbf{d}}) =_{\text{def}} \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{r}_\tau^{(N)}(\hat{\mathbf{d}}). \tag{7}$$

- The *mean increment* function $\mu^{(N)} : \hat{\mathcal{D}}^{(N)} \to \mathbb{R}^n$ associates each state $\hat{\mathbf{d}}$ with the mean increment in $\hat{\mathbf{d}}$, i.e. the average variation of each variable in a single discrete step of the underlying CTMC. Since in each state $\hat{\mathbf{d}}$, the probability that the next transition is τ equals $\frac{\hat{r}_\tau^{(N)}(\hat{\mathbf{d}})}{R^{(N)}(\hat{\mathbf{d}})}$, and $\hat{\mathbf{v}}_\tau^{(N)}$ is the variation of system variables due to transition τ, we get:

$$\mu^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{\mathbf{v}}_\tau^{(N)} \frac{\hat{r}_\tau^{(N)}(\hat{\mathbf{d}})}{R^{(N)}(\hat{\mathbf{d}})}. \tag{8}$$

  Notice that the $\mu^{(N)}(\hat{\mathbf{d}})$ may have negative components.

- The *covariance matrix*[9] $\Sigma^{(N)}(\hat{\mathbf{d}})$ of the increments in state $\hat{\mathbf{d}}$, i.e. the matrix storing the covariances of the increments between each pair of

[9]Formally, the covariance matrix $\Sigma(\mathbf{x})$ is defined by $\Sigma_{ij}(\mathbf{x}) = \mathbb{E}[(\tilde{X}_i(k+1) - \tilde{X}_i(k))(\tilde{X}_j(k+1) - \tilde{X}_j(k)) \mid \tilde{X}_i(k) = x_i, \tilde{X}_j(k) = x_j] - \mathbb{E}[(\tilde{X}_i(k+1) - \tilde{X}_i(k)) \mid \tilde{X}_i(k) = x_i]\mathbb{E}[(\tilde{X}_j(k+1) - \tilde{X}_j(k)) \mid \tilde{X}_j(k) = x_j]$, where $\tilde{\mathbf{X}}(\mathbf{k})$ is the embedded DTMC [13] associated to the CTMC $\mathbf{X}(t)$.

variables, can be easily shown to be

$$\Sigma_{ij}^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{v}_{\tau,i}^{(N)} \hat{v}_{\tau,j}^{(N)} \frac{\hat{r}_{\tau}^{(N)}(\hat{\mathbf{d}})}{R^{(N)}(\hat{\mathbf{d}})} - \mu_i^{(N)}(\hat{\mathbf{d}}) \mu_j^{(N)}(\hat{\mathbf{d}}). \tag{9}$$

In particular, the *variance of the increments of variable* $\hat{X}_i^{(N)}$ is

$$\Sigma_{ii}^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} (\hat{v}_{\tau,i}^{(N)})^2 \frac{\hat{r}_{\tau}^{(N)}(\hat{\mathbf{d}})}{R^{(N)}(\hat{\mathbf{d}})} - \mu_i^{(N)}(\hat{\mathbf{d}})^2. \tag{10}$$

Mean increment and the covariance matrix are concepts needed to describe the local average dynamics of the CTMC and the local structure of noise. By imposing suitable conditions on their dependence on the parameter N, which will be discussed in the next section, we can guarantee that noise goes to zero in the limit.

- The *mean dynamics* or *drift* $F^{(N)} : \hat{\mathcal{D}}^{(N)} \to \mathbb{R}^n$ of the model in state $\hat{\mathbf{d}}$ is

$$F^{(N)}(\hat{\mathbf{d}}) =_{\text{def}} R^{(N)}(\hat{\mathbf{d}}) \mu^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{\mathbf{v}}_{\tau}^{(N)} \hat{r}_{\tau}^{(N)}(\hat{\mathbf{d}}). \tag{11}$$

  The concept of mean dynamics, by multiplying the mean increment by the exit rate, essentially captures the (average) local variation of the CTMC with respect to the passing of time.

Notice that all the above definitions apply also to non-normalised models, although we need them only for normalised ones.

4.3.2. Scaling Assumptions

Let $\hat{\mathcal{X}}_C^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$ be a normalised model and let us consider the sequence $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$ with respect to an increasing system size γ_N.

State Space. Let E be a *closed* set in $\mathbb{R}^n$ such that $\bigcup_N \hat{\mathcal{D}}^{(N)} \subseteq E$, the space in which all processes of the sequence and their deterministic approximation live. We will, however, state the convergence result for a (appropriate) relatively open subset of $S \subseteq E$.[10] This is convenient because we can localise all the scaling assumptions to S. We will further denote $S \cap \hat{\mathcal{D}}^{(N)}$ by $S^{(N)}$.

[10] A relatively open subset S is an open set in the subspace topology on E. The subspace topology is defined in the following way: the open sets of E are obtained by intersecting

Convergence of Initial Conditions. We assume that there is some point $\hat{\mathbf{d}}_{\mathbf{0}} \in S$ such that[11]

$$\lim_{N\to\infty} \hat{\mathbf{d}}_{\mathbf{0}}^{(N)} = \hat{\mathbf{d}}_{\mathbf{0}}. \tag{12}$$

Convergence of Drift. We assume that the drift vectors behave coherently in the limit, i.e. we assume that there is a Lipschitz vector field $F : E \to \mathbb{R}^n$ such that the drift $F^{(N)}$ converges uniformly to F. This means that

$$\lim_{N\to\infty} \sup_{\hat{\mathbf{d}}\in S^{(N)}} \|F^{(N)}(\hat{\mathbf{d}}) - F(\hat{\mathbf{d}})\| = 0. \tag{13}$$

We further assume that, by appropriately choosing E, all the trajectories $\mathbf{x}(t)$ which are solutions of the initial value problem $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$, when $\hat{\mathbf{d}}_{\mathbf{0}} \in E$, remain in E (for all time instants in which $\hat{\mathbf{x}}$ is defined). Note that the supremum is taken in $S^{(N)}$, i.e. we are just requiring convergence in S, not in E. This means that only what happens in S is relevant for the validity of the theorem. In particular, F is required to be Lipschitz only in S: its behaviour outside S is not relevant[12]. In our setting, it is usually easier to focus attention on single transitions, proving for each transition τ the existence of a Lipschitz function (in S) $f_\tau : E \to \mathbb{R}$ such that $\hat{r}_\tau^{(N)}$ converges uniformly to f_τ on S.

Convergence to Zero of Noise. The other hypotheses consider the dependence of the exit rate and of the jump size on N, plus a condition on their cross-relation (which is essentially a condition on the variance). Usually, the scaling conditions are such that the exit rate goes to infinity with N, while the step size goes to zero, both at the same speed (typically linearly with respect to N). Here, however, we consider more general scaling laws, that subsume the standard ones and allow the application of mean field results to a larger class of systems. We require three things:

1. The exit rate is bounded for each N, i.e. there exists $\Lambda_N \in \mathbb{R}_{\geq 0}$, $\Lambda_N < \infty$, such that

$$\sup_{\hat{\mathbf{d}}\in S^{(N)}} R^{(N)}(\hat{\mathbf{d}}) = \Lambda_N. \tag{14}$$

E with open sets $U \subseteq \mathbb{R}^N$. Therefore, S is relatively open in E if and only if there exists an open set $U \subseteq \mathbb{R}^N$ such that $S = E \cap U$. Note that $E = E \cap \mathbb{R}^n$ is relatively open in E.

[11] For the approximation theorem, we can allow the PCTMC model to start from an initial distribution converging in probability to $\hat{\mathbf{d}}_{\mathbf{0}} \in S$.

[12] It is always possible to redefine F outside S to make it Lipschitz in E.

Usually, $\lim_{N\to\infty} \Lambda_N = \infty$, so that the frequency of jumps increases with N.

2. The magnitude of jumps goes to zero. More precisely, there exists $J_N \in \mathbb{R}_{\geq 0}$, such that
$$\max_{\tau \in \hat{\mathcal{T}}^{(N)}} ||\hat{\mathbf{v}}_\tau^{(N)}|| = J_N. \tag{15}$$
Moreover $\lim_{N\to\infty} J_N = 0$ and, in particular, J_N is $O(N^{-1})$; this means that the magnitude of jumps goes to zero at least as quickly as N^{-1}.
3. The scaling of jump magnitude and exit rate must be compatible, according to the following condition:
$$J_N^2 \Lambda_N \text{ is } O(N^{-1}). \tag{16}$$
This is essentially a condition on noise, and it enforces that the variance of the system goes to zero.

In our setting, the previous conditions can be simplified whenever the *non-normalised* increments are independent of N, i.e. whenever, for each transition τ, there is a vector $\mathbf{v}_\tau$ such that $\mathbf{v}_\tau^{(N)} = \mathbf{v}_\tau$ for all N, so that $\hat{\mathbf{v}}_\tau^{(N)} = \delta_N \cdot \mathbf{v}_\tau^{(N)} = \delta_N \cdot \mathbf{v}_\tau$, with $\delta_N = \frac{1}{\gamma_N}$. In particular, the second condition can be restated in terms of system size γ_N: we have that $\|\hat{\mathbf{v}}_\tau^{(N)}\| = \delta_N \cdot \|\mathbf{v}_\tau\|$, where $||\mathbf{v}_\tau||$ is $O(1)$. Hence $J_N = \delta_N \cdot J$, where $J = \max_{\tau \in \mathcal{T}^{(N)}} ||\mathbf{v}_\tau||$. Therefore, the second condition is satisfied as δ_N converges to 0, with order $O(N^{-1})$, while the third condition can be restated as $\delta_N^2 \Lambda_N = O(N^{-1})$. In practical applications, one usually has that $\Lambda_N = \Theta(N)$ and $\gamma_N = \Theta(N)$, so that $\delta_N = \Theta(N^{-1})$, hence condition (16) holds straightforwardly.

4.3.3. Deterministic Approximation Theorem

We now state the general approximation theorems for PCTMCs [14, 29, 55, 56, 58]. Consider a sequence of normalised PCTMC models $(\hat{\mathcal{X}}_C^{(N)})_{N\geq N_0}$ with $\hat{\mathcal{X}}_C^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$, and denote by $\hat{\mathbf{X}}^{(N)}(t)$ the continuous-time Markov process associated with $\hat{\mathcal{X}}_C^{(N)}$. Furthermore, denote by $\hat{\mathbf{x}}(t)$ the solution of the initial value problem $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$, $\hat{\mathbf{x}}(0) = \hat{\mathbf{d}}_{\mathbf{0}}$, where F is as in (13). We will also need the following definitions: the *exit time* from S of the ODE solution $\hat{\mathbf{x}}(t)$ is $\zeta(S) =_{\text{def}} \inf\{t \geq 0 \mid \hat{\mathbf{x}}(t) \notin S\}$, and the *exit time* from S of the Markov processes $\hat{\mathbf{X}}^{(N)}(t)$ is $\zeta^{(N)}(S) =_{\text{def}} \inf\{t \geq 0 \mid \hat{\mathbf{X}}^{(N)}(t) \notin S\}$.

Theorem 4.2 (Deterministic approximation of PCTMCs)**.** *Let $(\hat{\mathbf{X}}^{(N)}(t))_{N\geq N_0}$ be the sequence of Markov processes associated to the sequence of PCTMC*

models $(\hat{\mathcal{X}}_C^{(N)})_{N\geq N_0}$, $\hat{\mathbf{x}}(t)$ *as above, and assume Conditions (12-16) apply. Then, for any* finite *time horizon* $T < \zeta(S)$, *it holds that:*

1. $\lim_{N\to\infty} \mathbb{P}\{\zeta^{(N)}(S) < T\} = 0$;
2. *for all* $\varepsilon \in \mathbb{R}_{>0}$, $\lim_{N\to\infty} \mathbb{P}\{\sup_{0\leq t\leq T} ||\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)|| > \varepsilon\} = 0$.

Furthermore, both $\mathbb{P}\{\zeta^{(N)}(S) < T\}$ *and* $\mathbb{P}\{\sup_{0\leq t\leq T} ||\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)|| > \varepsilon\}$ *are* $O(e^{-N})$ ■

Similarly to Theorem 4.1, this theorem gives a law of large numbers for PCTMC models. Also in this case, the validity of the theorem is for any finite time horizon T, chosen so that the solution of the ODE remains in S in $[0, T]$. Nothing is said about asymptotic behaviour. We now give a slightly stronger result than the previous theorem, for the case in which the ODE solution leaves S in finite time, i.e. $\zeta(S) < \infty$. For this result to hold, we need that $\hat{\mathbf{x}}(t)$ *leaves* S *transversally* at time $\zeta(S)$, meaning that $\hat{\mathbf{x}}(t)$ remains outside (the closure of) S for some time after leaving S. More precisely, $\hat{\mathbf{x}}(t)$ leaves transversally S at time $\zeta(S)$ if and only if there is $\delta > 0$ such that, for $t \in]\zeta(S), \zeta(S)+\delta[$, $\hat{\mathbf{x}}(t) \notin cl(S)$ (hence $\hat{\mathbf{x}}(t)$ remains for some time in $int(S^c)$, the interior of the complement of S).

Theorem 4.3 (Deterministic approximation for PCTMCs with finite exit times)**.** *Let the sequence* $(\hat{\mathbf{X}}^{(N)}(t))_{N\geq N_0}$ *of Markov processes and* $\hat{\mathbf{x}}(t)$ *be defined as before, and assume Conditions (12-16) are in force. Then, if* $\zeta(S) < \infty$ *and* $\hat{\mathbf{x}}(t)$ *leaves transversally* S *at time* $\zeta(S)$, *it holds that, for all* $\varepsilon \in \mathbb{R}_{>0}$:

1. $\lim_{N\to\infty} \mathbb{P}\{\sup_{0\leq t\leq \zeta(S)} ||\hat{\mathbf{X}}^{(N)}(\min\{t, \zeta^{(N)}(S)\}) - \hat{\mathbf{x}}(t)|| > \varepsilon\} = 0$;
2. $\lim_{N\to\infty} \mathbb{P}\{||\zeta^{(N)}(S) - \zeta(S)|| > \varepsilon\} = 0$.

Furthermore, both of the probabilities $\mathbb{P}\{||\zeta^{(N)}(S) - \zeta(S)|| > \varepsilon\}$ *and* $\mathbb{P}\{\sup_{0\leq t\leq \zeta(S)} ||\hat{\mathbf{X}}^{(N)}(\min\{t, \zeta^{(N)}(S)\}) - \hat{\mathbf{x}}(t)|| > \varepsilon\}$ *are* $O(e^{-N})$. ■

This second theorem states that the CTMC trajectories are indistinguishable from the ODE solution provided that we look at what happens while the ODE trajectory is in S. Furthermore, it states that the exit time from S of the CTMCs converges to that of the ODE, meaning that for large N we can estimate $\zeta^{(N)}(S)$ by $\zeta(S)$.

Remark 4.1. The scaling assumptions we are taking into account are fairly general, and encompass many scaling laws found in the literature. For instance, in [56], the author assumes that $\Lambda_N = O(N)$ and he requires the following scaling condition on the variance of all variables:

$$\sup_{\hat{\mathbf{d}} \in S^{(N)}} \sum_{i=1}^{n} \Sigma_{ii}^{(N)}(\mathbf{d}) + ||\mu^{(N)}(\hat{\mathbf{d}})||^2 = O(N^{-2}).$$

Provided $\mathbf{v}_\tau^{(N)}$ does not depend on N, this can be rewritten in our setting as

$$\sup_{\hat{\mathbf{d}} \in S^{(N)}} \sum_{\tau \in \mathcal{T}^{(N)}} \delta_N^2 ||\mathbf{v}_\tau||^2 \frac{\hat{r}_\tau^{(N)}(\hat{\mathbf{d}})}{R^{(N)}(\hat{\mathbf{d}})} = O(N^{-2}),$$

which essentially amounts to requiring that $\delta_N = O(N^{-1})$. Notice that, in this case, condition (16) is satisfied. In particular, this argument justifies the fact that we refer to (16) as a condition on noise.

The density dependence condition introduced in Section 4.1 is also implied by this more general scaling law. Note that there are many examples of models that are not density dependent but for which the deterministic approximation holds. One example about multi-type particle interactions is given in [56]. Additionally, all mass action models of biochemical reactions in which two molecules of the same kind react are not density dependent, but they satisfy the more general scaling conditions of this section (see [59]). A similar situation is found in epidemic models (see for instance the malware example of [42]).

Remark 4.2. Theorems 4.2 and 4.3 state that the sequence of PCTMCs $\hat{\mathbf{X}}^{(N)}(t)$ converges to $\hat{\mathbf{x}}(t)$, the solution of the ODE $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$ for the limit vector field F. In case the drift $F^{(N)}$ of $\hat{\mathbf{X}}^{(N)}(t)$ depends on N, a simple consequence of those theorems is that we can replace the vector field F with the N-dependent vector field $F^{(N)}$. More specifically, letting $\hat{\mathbf{x}}^{(N)}(t)$ be the solution of the ODE $\frac{d\hat{\mathbf{x}}^{(N)}(t)}{dt} = F^{(N)}(\hat{\mathbf{x}}^{(N)}(t))$, it can be proved that $\|\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}^{(N)}(t)\|$ converges to zero in probability as N goes to infinity. In some situations, the vector field $F^{(N)}$ can be more effective in capturing the behaviour of the PCTMC model for a fixed N. An example of this kind is the Bianchi paper analysing the IEEE 802.11 protocol [60]. Another example will be given in Section 4.4 below.

*Remark** 4.3. There are several proof techniques that have been used to prove these two theorems. Probably the most used approach is based on martingale theory [58]. Martingales are particular stochastic processes, whose conditional expected value at future times, given the present, is equal to the observed value at the present. Specifically, in this context [56, 57] one can construct a martingale as the difference between the CTMC and its *compensator* (i.e. the process that accumulates the mean increments, computed according to the mean dynamics, and "compensates" the deviation from the mean of the CTMC). After showing that this is indeed a martingale with zero mean, the theorem is proved by applying standard martingale inequalities, like Doob's inequality or an exponential martingale inequality [57]. In this way, one shows that the CTMC, in the limit, behaves like a deterministic process, which is the solution of an integral equation. Then, one proves that this is indeed the solution of the fluid ODE, by applying the Gronwall inequality [58], a functional inequality used to prove properties of ODE solutions in worst case scenarios (see also Remark 4.4).

In the density dependent case, the theorem can be proved using a random time change argument to represent the CTMCs in terms of Poisson processes, counting how many times each transition fired up to time t. Then, the law of large numbers for Poisson processes is exploited to get convergence (see [55] and the proof in Appendix B).

A different proof technique, having a broader spectrum of applications but using more advanced mathematical tools, is based on infinitesimal generators [14, 55, 58], which are operators on functional spaces (usually the space of bounded measurable functions or the space of continuous functions vanishing at infinity) encoding information about the expected value of the CTMC for any function of the space. There is a large body of theory showing connections between properties of the stochastic processes and properties of their generators. In particular, convergence of stochastic processes is equivalent to the convergence of their infinitesimal generators, which may be easier to prove [14, 55, 58].

*Remark** 4.4. Useful companions of limit theorems are bounds on the error introduced in replacing a stochastic process by its deterministic limit. In Theorems 4.2 and 4.3, error bounds are provably exponentially decreasing in N. Despite this, they are still quite loose [57, 61]. In fact, the error bounds depend (doubly) exponentially on the time horizon T, so that, despite going exponentially to zero as N grows, they tend to be too large for any practical

application. More precisely, error bounds have the following form:

$$\mathbb{P}\{\sup_{0\leq t\leq T}||\hat{\mathbf{X}}^{(N)}(t)-\hat{\mathbf{x}}(t)||>\varepsilon\}\leq 2\cdot N\cdot e^{W},$$

where W stands for $-\frac{\varepsilon^2}{18}\cdot\frac{1}{T\cdot e^{2\cdot K\cdot T}}\frac{1}{A_N}$, K is the Lipschitz constant of F and $A_N=\max\{e,(\delta_N\cdot\Lambda_N)^{-1}\}\cdot\delta_N^2\cdot\Lambda_N$, so that $\frac{1}{A_N}=\Omega(N)$, if conditions (12) to (16) are in force. This is related to the fact that Theorem 4.2 holds for any arbitrary trajectory of the limit ODE. In particular, it holds for trajectories that are unstable, meaning that even small fluctuations can lead exponentially far away from them. Indeed, the appearance of T as an exponent is typical of the use of Gronwall's inequality, which is used to prove global properties of ODE solutions, including unstable trajectories.

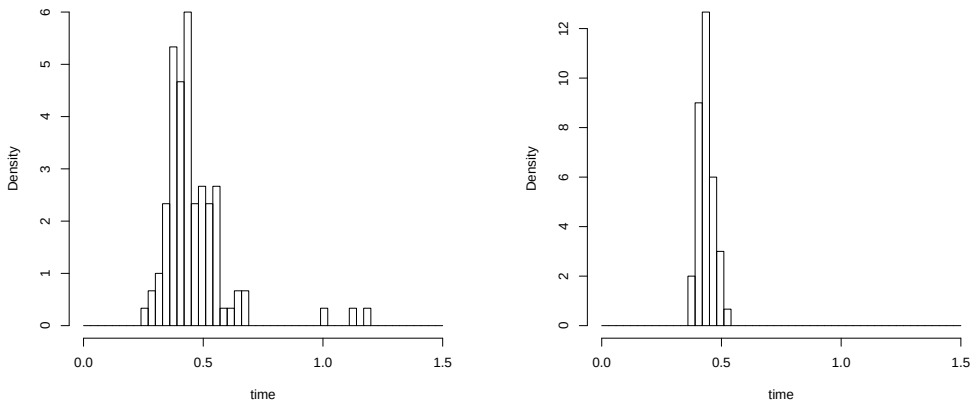
In practical applications, however, we usually have to deal with trajectories with much nicer stability properties, hence the behaviour of the deterministic approximation in terms of error bounds is much better than the one predicted by the worst-case error bounds. Proving bounds for such situations is an active and challenging research area [61].

4.3.4. Main Example (continued)

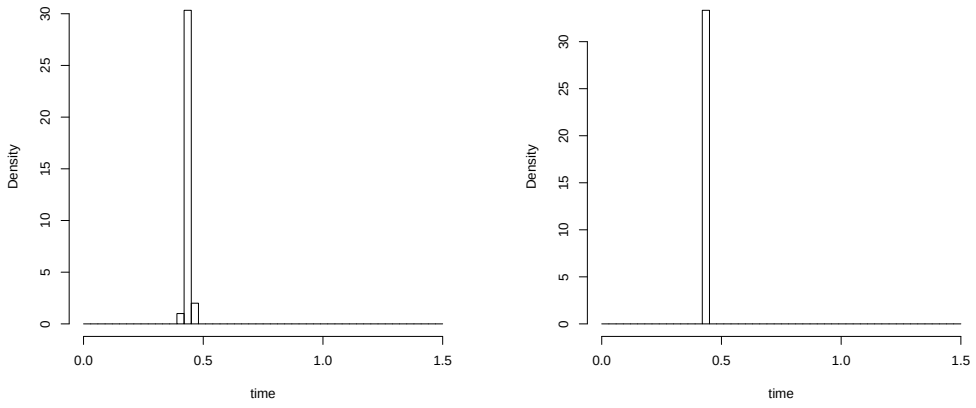
Consider again the epidemic model, whose limit behaviour has been discussed in Section 4.2. Theorems 4.2 and 4.3 can also be used to estimate properties of the stochastic sequence of models, such as exit times. Consider, for instance, the following question: "when will one third of the nodes be contaminated by the virus?" In this case, we need to compute the exit time from the set $S=\{(d_1,d_2,d_3,d_4)\in E\mid d_2+d_3<\frac{1}{3}\}$. If we consider the set of ODEs, we obtain that $\zeta(S)=0.43043$ (for parameters as in Figure 6). It is easy to check that the ODE trajectory leaves S transversally: $\frac{d(\hat{x}_e+\hat{x}_i)}{dt}$ is positive at time $\zeta(S)$ (hence $\hat{x}_e(t)+\hat{x}_i(t)>1/3$ for $\zeta(S)<t<\zeta(S)+\delta$ for some $\delta>0$). Let $\zeta^{(N)}(S)$ be the exit time for the sequence of PCTMC models. Theorem 4.3 states that $\zeta^{(N)}(S)$ converges to $\zeta(S)$ in probability, so that, if N is large, $\zeta(S)$ is a good estimate. We can see this fact visually by looking at the distributions of exit times for different values of N, as shown in Figure 7, obtained from a batch of simulations of the CTMCs. Notice that for $N=10^6$, $\zeta^{(N)}(S)=0.43043$ in all the simulation runs.

4.4. Example: Crowd Dynamics

In this section, we consider a PCTMC model of crowd dynamics, presented in [22]. The model tries to capture an emergent phenomenon happening in certain cities in southern Spain, where people wandering around



(a) $N = 10^2$ (b) $N = 10^3$



(c) $N = 10^4$ (d) $N = 10^6$

Figure 7: Distributions of exit times from the set $S = \{(d_1, d_2, d_3, d_4) \in E \mid d_2 + d_3 \leq \frac{1}{3}\}$, for different population levels, estimated from 1000 simulation runs. Parameters are as specified in Figure 6. The deterministic limit to which the exit time sequence converges is $\zeta(S) = 0.43043$.

city's squares during the evening suddenly start gathering in a single square, giving rise to a big (and noisy) party.

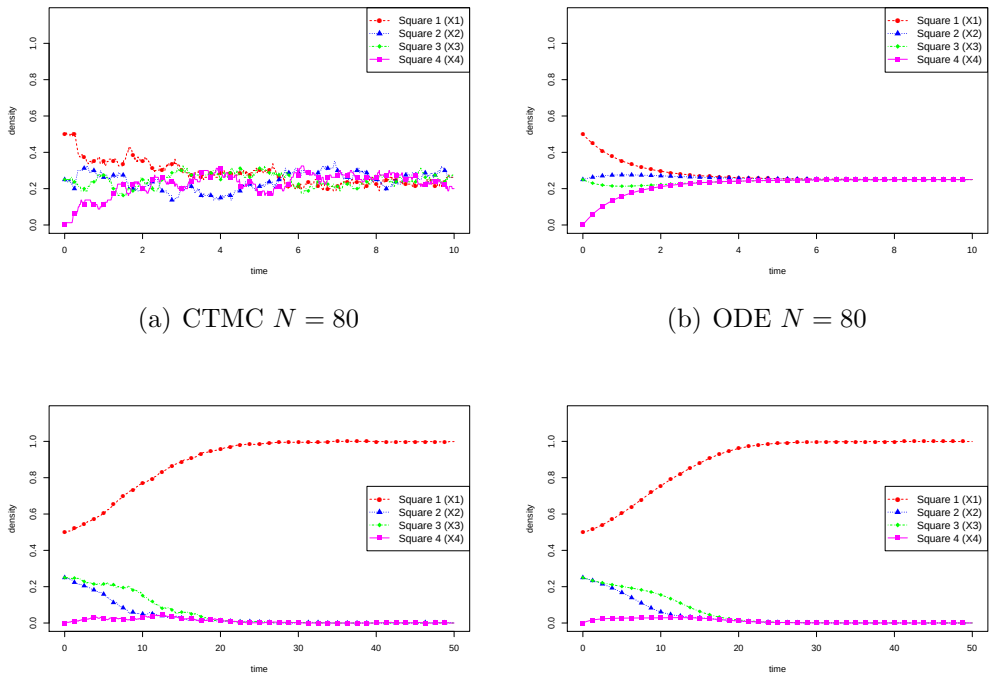
In this example, we assume we have 4 squares, connected in a ring topology (but more squares and more general topologies can be considered). The main idea is that each person is willing to remain in a square only if she finds someone to talk with. Hence, she will encounter a certain number of other people in the square, and may talk with them, depending on several factors (she knows these people, she likes them, they want to talk to her or not, etc.). This event is modelled in a simple way, namely as a Bernoulli random variable with probability c, called the *chat probability*. Hence, a person will leave the square if she finds nobody to chat with, i.e. with a probability $(1-c)^k$, where k is the number of people that she meets in the square. In the context of CTMCs, we will interpret this probability as a *rate*. We will describe the number of people in each square by the variables X_i, $i = 1, \ldots, 4$. The model assumes that a person will meet everybody currently in the square. Hence, the rate at which an individual in square i will leave that square in the next step is $(1-c)^{X_i-1}$. Therefore, the rate at which someone will leave square i is $X_i \cdot (1-c)^{X_i-1}$. When someone leaves, she may go to the preceding or the following square with the same probability, which in this case is $\frac{1}{2}$.

In this way, we can construct a sequence of models $\mathcal{F}_C^{(N)}$, for system size $\gamma_N = N$, with variables $\mathbf{X} = (X_1, X_2, X_3, X_4)$, state spaces $\mathcal{S}_N^4 = \{\mathbf{d} \in \{0, \ldots, N\}^4 \mid \sum_{i=1}^4 d_i = N\}$, and with 8 transitions, two per square. For instance, the transition modelling the event of moving from square 1 to square 2 is $(a_{1,2}, \mathbf{e_1}, \mathbf{e_2}, r_{1,2}^{(N)})$ where $r_{1,2}^{(N)}(\mathbf{X}) =_{\text{def}} \frac{1}{2} X_1 \cdot (1-c)^{X_1-1}$. The other transitions are defined similarly.

This model shows the following behaviour: if c is sufficiently high, larger than $\frac{4}{N}$, i.e. the number of squares divided by the population size, all the people tend to converge to the same square (a party emerges!). This phenomenon is observed both in the CTMCs and in the fluid model, as can be seen in Figure 8, where we show the behaviour of two models for different population levels, keeping $c = 0.01$ fixed. One model ($N = 80$) is below the threshold (no convergence, but random distribution of people among squares) and the other one ($N = 800$) is above it (convergence to a single square).

If we compute the drift for the normalised model with N people, we see that it depends on N (differently from other examples). For instance, the first component of $F^{(N)}$ is

$$-\hat{X}_1 \cdot (1-c)^{N \cdot \hat{X}_1 - 1} + 0.5 \cdot \hat{X}_2 \cdot (1-c)^{N \cdot \hat{X}_2 - 1} + 0.5 \cdot \hat{X}_4 \cdot (1-c)^{N \cdot \hat{X}_4 - 1}.$$



(a) CTMC $N = 80$ (b) ODE $N = 80$

(c) CTMC $N = 800$ (d) ODE $N = 800$

Figure 8: Comparison of deterministic and stochastic trajectories (for increasing population levels) of the crowd dynamics model for initial conditions $\hat{x}_1 = 0.5$, $\hat{x}_2 = 0.25$, $\hat{x}_3 = 0.25$, and $\hat{x}_4 = 0.0$, and chat probability $c = 0.01$. For $N = 80$ (Figures 8(a) and 8(b)), c is below the threshold, and no emergent crowding behaviour is observed. For $N = 800$, instead, this phenomenon holds.

For fixed c, letting the state space E be unit simplex in $[0, 1]^4$, we have that

$$\lim_{N\to\infty} \sup_{\hat{\mathbf{X}}\in E} ||F^{(N)}(\hat{\mathbf{X}})|| = 0\,,$$

so that the drift $F^{(N)}$ converges uniformly to the constant function yielding 0 for each point in E. Hence, the fluid limit of the sequence is the constant function $\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}_\mathbf{0}$! The ODE dynamics shown in Figure 8 is the solution of the ODE $\frac{d\hat{\mathbf{x}}^{(N)}(t)}{dt} = F^{(N)}(\hat{\mathbf{x}}^{(N)}(t))$, where we distinguish the different solutions for different N, hence it is not the fluid limit. However, increasing N, we observe that the ODE solution $\hat{\mathbf{x}}^{(N)}(t)$ converges to a constant function. In

Figure 9, we see what happens for $N = 8000$. In this case, it seems that no-one moves in the CTMC and similarly in the ODE model (Figures 9(a) and 9(b)). However, running the simulation for a much longer time period (of the order of 10^{12}), we can still observe that people converge to the same square. This will happen for every finite N, both for the CTMC model and the N-dependent fluid limit (Figures 9(c) and 9(d)). However, as N grows, the time at which all agents converge to the same square for the solution $\hat{\mathbf{x}}^{(N)}(t)$, i.e. the time at which $\hat{\mathbf{x}}^{(N)}(t)$ departs from the constant solution $\hat{\mathbf{x}}(t)$, will go to infinity.

Although this behaviour seems to contradict the use of deterministic approximation theorems to study this model, there is a reasonable agreement between ODE and CTMCs for large N. Indeed, deterministic approximation results can still be used in this context for two complementary reasons:

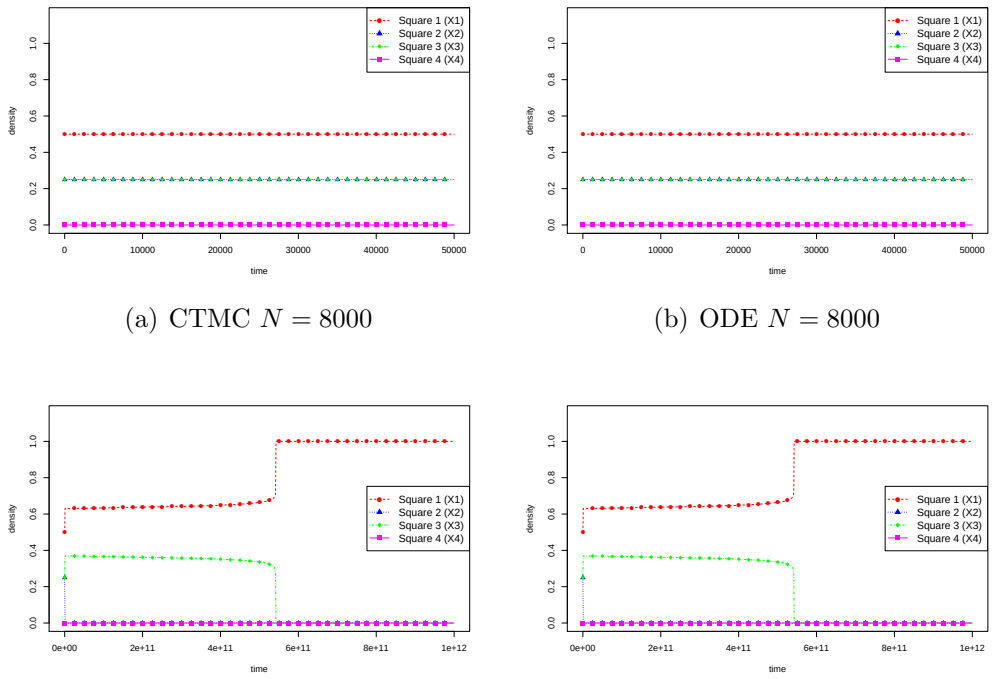
1. $\hat{\mathbf{x}}^{(N)}(t)$ is, in any case, an approximation of the average of the stochastic process (this will be discussed in Section 8, see also [22]);
2. in the light of Remark 4.2, we can always approximate the CTMC with $\hat{\mathbf{x}}^{(N)}(t)$, the solution of $\frac{d\hat{\mathbf{x}}^{(N)}(t)}{dt} = F^{(N)}(\hat{\mathbf{x}}^{(N)}(t))$, if N is sufficiently large.

4.5. Example: a Queueing Model

In this section we will briefly discuss a queueing model studied in [43], in which the authors consider a CTMC model representing N servers, each with buffer size equal to 1. The policy for the incoming customers is to redirect them to the server with the shortest queue available, if there is one. Servers cannot be distinguished from one another, so we can describe the state of the queueing network by three variables, X_0, X_1, and X_2, describing the number of servers dealing with 0, 1, and 2 customers, respectively. We assume that the arrival rate scales with N, i.e. it is equal to $N \cdot \lambda$, and that the service rate is ρ for each queue. We assume that arrivals are suspended when all queues are full. We can easily construct a model:

$$\mathcal{Q}_C^{(N)} =_{\text{def}} (\mathbf{X}^{(N)}, \mathcal{S}_N^3, \mathcal{T}^{(N)}, \mathbf{d_0}^{(N)})$$

of such a system, where $\mathbf{X}^{(N)} = (X_0, X_1, X_2)$, $\mathcal{S}_N^3 = \{\mathbf{d} \in \{0, N\}^3 \mid \sum_{i=1}^n d_i = N\}$, and $\mathbf{d_0}^{(N)} \in \{0, \ldots, N\}^3$ such that we have $\lim_{N\to\infty} \hat{\mathbf{d}}_\mathbf{0}^{(N)} = \hat{\mathbf{d}}_\mathbf{0}$ for some $\hat{\mathbf{d}}_\mathbf{0} \in [0, 1]^3$. The transition set $\mathcal{T}^{(N)}$ is composed of four transitions, defined below, where, in a similar way to before, $\mathbf{e_j}$ is the unit vector in $\mathbb{R}^3$ with 1 as j-component and 0 in the others:



(a) CTMC $N = 8000$

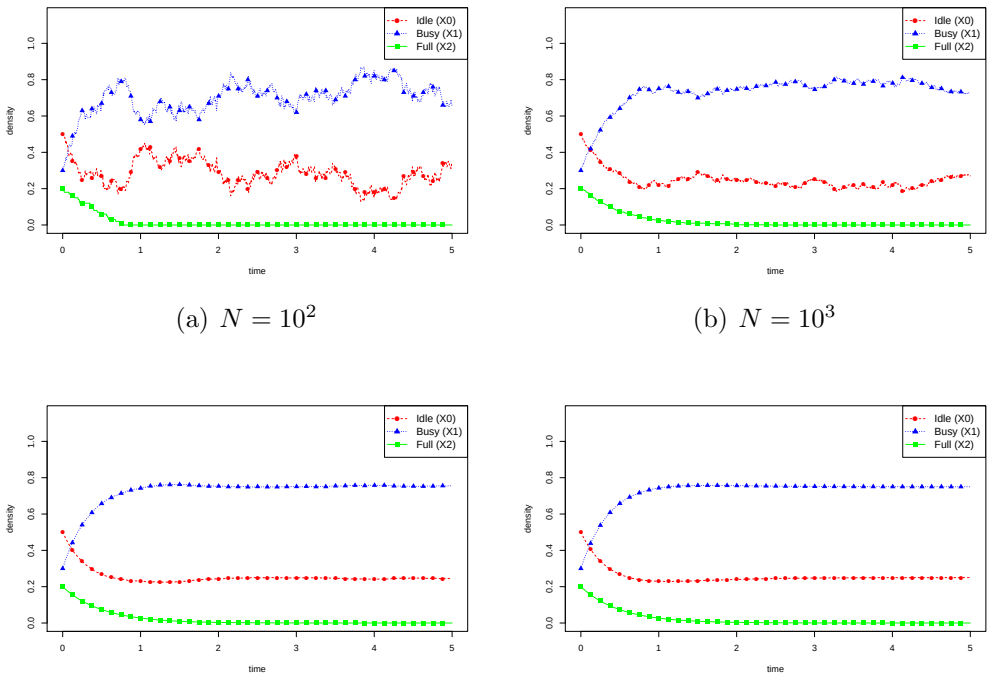
(b) ODE $N = 8000$

(c) CTMC $N = 8000$, $T_f = 10^{12}$

(d) ODE $N = 8000$, $T_f = 10^{12}$

Figure 9: Comparison of deterministic and stochastic trajectories of the crowd dynamics model with $N = 8000$ people, for initial conditions $\hat{x}_1 = 0.5$, $\hat{x}_2 = 0.25$, $\hat{x}_3 = 0.25$, and $\hat{x}_4 = 0.0$, and chat probability $c = 0.01$. In Figures 9(a) and 9(b), the system seems to remain constant. However, if we observe it for a very long time (up to time $T_f = 10^{12}$), we can still see the emergent behaviour typical of the model.

- Incoming client directed to an idle server: state change vector $\mathbf{v}_{\tau_{\mathbf{ie}}} =_{\text{def}} \mathbf{e_1} - \mathbf{e_0}$ and rate function $r_{ie}^{(N)}(\mathbf{X}) =_{\text{def}} N \cdot \lambda \cdot \mathbf{1}\{X_0 > 0\}$.
- Incoming client directed to a busy server with empty queue: state change vector $\mathbf{v}_{\tau_{\mathbf{io}}} =_{\text{def}} \mathbf{e_2} - \mathbf{e_1}$ and rate function $r_{io}^{(N)}(\mathbf{X}) =_{\text{def}} N \cdot \lambda \cdot \mathbf{1}\{X_0 = 0\} \cdot \mathbf{1}\{X_1 > 0\}$.
- Servicing of a client from a server with buffer empty: state change vector $\mathbf{v}_{\tau_{\mathbf{se}}} =_{\text{def}} \mathbf{e_0} - \mathbf{e_1}$ and rate function $r_{se}^{(N)}(\mathbf{X}) =_{\text{def}} \rho \cdot X_1$.
- Servicing of a client from a server with buffer full: state change vector



(a) $N = 10^2$ (b) $N = 10^3$

(c) $N = 10^5$ (d) ODE

Figure 10: Comparison of deterministic and stochastic trajectories (for increasing population levels) of the queueing model for initial conditions $\hat{x}_0 = 0.5$, $\hat{x}_1 = 0.3$, $\hat{x}_2 = 0.2$, and parameters $\lambda = 1.5$ and $\rho = 2$.

$$\mathbf{v}_{\tau_{\mathbf{sf}}} =_{\text{def}} \mathbf{e_1} - \mathbf{e_2} \text{ and rate function } r_{sf}^{(N)}(\mathbf{X}) =_{\text{def}} \rho \cdot X_2.$$

Notice that the shortest queue policy requires the use of indicator functions (which are discontinuous). In fact, we must direct a customer to an idle server as long as there is one, while we direct the customer to a server already servicing a client (and with an empty queue) only if there are no idle servers.

We consider a sequence $(\hat{\mathcal{Q}}_C^{(N)})_{N \geq N_0}$ of models, normalised with respect to the system size $\gamma_N = N$, according to the recipe of Section 3, and check that the scaling assumptions are satisfied. The state space of these models is the unit simplex E in $\mathbb{R}^3$, $E =_{\text{def}} \mathcal{S}^3$. Convergence of initial conditions is immediate by definition. The exit rate is $\Theta(N)$ and the step size is $\delta_N = \frac{1}{N}$,

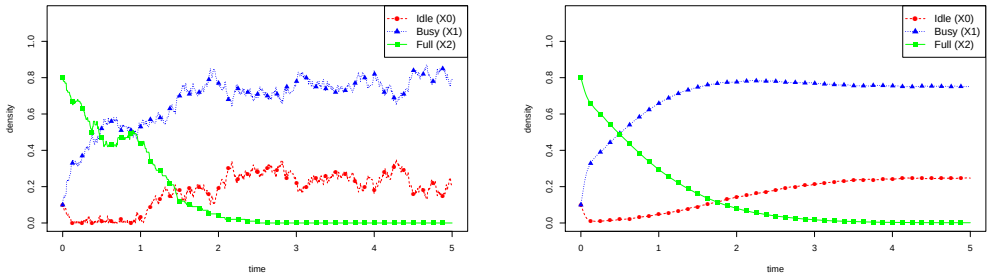
hence the noise conditions hold. As for the drift, we have that

$$F^{(N)}\begin{pmatrix}\hat{x}_0\\ \hat{x}_1\\ \hat{x}_2\end{pmatrix} = \begin{pmatrix}\rho\cdot\hat{x}_1-\lambda\cdot I_{\hat{x}_0>0}\\ \lambda\cdot I_{\hat{x}_0>0}+\rho\cdot\hat{x}_2-\lambda\cdot I_{\hat{x}_0=0}\cdot I_{\hat{x}_1>0}-\rho\cdot\hat{x}_1\\ \lambda\cdot I_{\hat{x}_0=0}\cdot I_{\hat{x}_1>0}-\rho\cdot\hat{x}_2\end{pmatrix}.$$

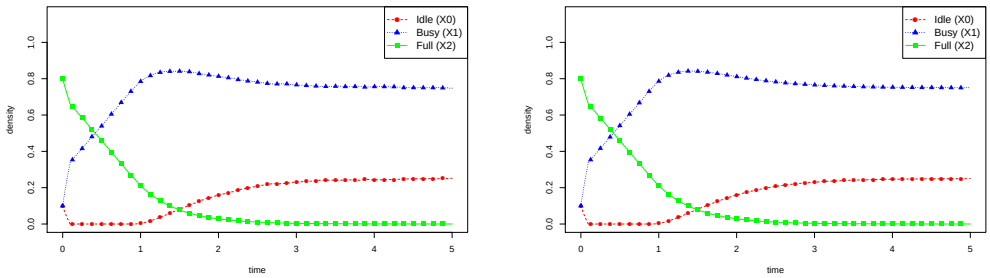
In this example, we have that function $F^{(N)}$ does not depend on N; thus we denote it by F. However, note that there is a problem here: the function F is *not* Lipschitz in E, because of the discontinuities introduced by the indicator functions. These functions introduce a discontinuity between (part of) the border of the simplex E and its interior; F jumps from the boundary of E to its interior when there are no idle servers (i.e. $X_0 = 0$) and no busy servers with an empty buffer (i.e. $X_1 = 0$). In order to circumvent this problem, we can focus our attention on a subset $S \subset E$, contained in the interior of E, in which F is Lipschitz. If we localise the theorem in this subset, we can still prove that the solution $\hat{\mathbf{x}}(t)$ of the ODE $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$, $\hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_\mathbf{0}$, is the limit of the sequence of CTMCs $\hat{\mathbf{X}}^{(N)}(t)$ in S. A necessary condition for this to hold is that $\hat{\mathbf{x}}_\mathbf{0} \in S$. Hence we must choose an initial configuration with a non-null fraction of idle servers and busy servers with empty buffer. If the initial conditions are sufficiently far away from the boundary ∂E of E, and $\rho > \lambda$, then $\hat{\mathbf{x}}(t)$ will always remain inside E, hence the theorem is valid for all stop times T (see Figure 10). On the other hand, if the initial conditions are close to the point $(0, 0, 1)$, then the ODE solution will reach a state with no empty queues, will remain in ∂E for a while (until $\hat{x}_1 > \frac{\lambda}{\rho}$, i.e. until the vector field just outside ∂E will point inside E, which will occur when the ODE for $\hat{x}_0$ in E is positive: $\rho\cdot\hat{x}_1-\lambda > 0$), and then enter the interior of E again (see Figure 11(d)). During this period, we cannot invoke the theorems of Section 4 to obtain convergence. Nevertheless, the manifested behaviour of the system seems to obey the pattern of deterministic approximation (Figure 11). This switching behaviour of the ODE may be dealt with using hybrid techniques, as will be discussed in Section 8.

5. Deterministic Approximation for PDTMCs

There are many approaches in the literature for the deterministic approximation of DTMCs. Some of them are focussed on specific examples [40, 39], while others are more general [32, 42]. Usually, these approximation results are referred to under the name of *mean field approximation*. Broadly speaking, there are two main classes of mean field results for DTMC: those resulting



(a) $N = 10^2$ (b) average for $N = 10^2$



(c) $N = 10^5$ (d) ODE

Figure 11: Comparison of deterministic and stochastic trajectories (for increasing population levels) of the queueing model for initial conditions $\hat{x}_0 = 0.1$, $\hat{x}_1 = 0.1$, $\hat{x}_2 = 0.8$, and parameters $\lambda = 1.5$ and $\rho = 2$. Figure 11(b) shows the average over 1000 runs for the model with $N = 100$.

in a set of difference equations, discussed in Section 2, and those resulting in a set of ODEs, which we address in the present section. We stress that in PDTMCs, only one transition is executed per step, resulting in a constant number of agents moving per step. This is in contrast with synchronous DTMCs, where all agents move in each step. The ODE limit for PDTMCs is obtained by embedding its trajectories in real time, assuming the duration of each step to be constant and equal to ϵ_N, with ϵ_N decreasing to zero as N increases. In other words, increasing N, the density of events in each unit of time also increases, so that in the limit the evolution of the system becomes continuous. This argument must be made rigorous by stating clearly what the scaling assumptions are.

In the following, we will first introduce some notation and define the scaling assumptions. Then, we will state the main approximation theorems and discuss again the network epidemic example. In Section 6, instead, we will focus our attention on comparing the approximation results for PDTMCs and PCTMCs.

5.1. Notation and Scaling Assumptions

Consider a normalised PDTMC model $\hat{\mathcal{X}}_D^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$, where $\hat{\mathbf{X}}^{(N)}$ is a vector of n variables. The relevant notion in this context is that of *mean increment*, i.e. the function $\mu^{(N)} : \hat{\mathcal{D}}^{(N)} \to \mathbb{R}^n$ defined as follows:

$$\mu^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{\mathbf{v}}_\tau^{(N)} p_\tau^{(N)}(\hat{\mathbf{d}}). \tag{17}$$

Consider now the sequence $(\hat{\mathcal{X}}_D^{(N)})_{N \geq N_0}$ with respect to an increasing system size γ_N.

We recall the fact that, in the context of mean field approximation of PDTMCs, usually $\gamma_N = N$, and the normalised variables lie in $[0, 1]$ and define a probability distribution, called the *occupancy measure*, although different notions of γ_N could, in general, be defined.

We now list the scaling assumptions required for the theorem to hold.

State Space. The state space is defined as for the PCTMC case, i.e. it is a closed set E in $\mathbb{R}^n$ such that $\bigcup_N \hat{\mathcal{D}}^{(N)} \subseteq E$.

Convergence of the Initial Conditions. Also the requirement on the initial condition is the same as for the PCTMC case: there must exist some point $\hat{\mathbf{d}}_{\mathbf{0}} \in E$ such that

$$\lim_{N \to \infty} \hat{\mathbf{d}}_{\mathbf{0}}^{(N)} = \hat{\mathbf{d}}_{\mathbf{0}}. \tag{18}$$

Intensity and scaling of time. We assume that the temporal duration of each global step equals ϵ_N, with $\lim_{N \to \infty} \epsilon_N = 0$. ϵ_N is usually referred to as the *intensity* [42], as further discussed in Remark 5.1.

Convergence of Drifts. Intensity is further involved in the scaling of the mean increments. In the PDTMC case, the *drift* is defined as follows:

$$F^{(N)}(\hat{\mathbf{d}}) =_{\text{def}} \frac{\mu^{(N)}(\hat{\mathbf{d}})}{\epsilon_N}. \tag{19}$$

As in the PCTMC case, we assume that there exists a (locally) Lipschitz vector field $F : E \to \mathbb{R}^n$, such that $F^{(N)}$ converges uniformly to F:

$$\lim_{N\to\infty} \sup_{\hat{\mathbf{d}}\in E} \|F^{(N)}(\hat{\mathbf{d}}) - F(\hat{\mathbf{d}})\| = 0. \tag{20}$$

We furthermore assume that all the trajectories of the vector field F lie in E (this can be accomplished by choosing E appropriately). In our framework, the notion of intensity coincides with the step-size $\delta_N = \frac{1}{\gamma_N}$: $\epsilon_N = \delta_N$, and we usually have $\epsilon_N = \frac{1}{N}$. This is consistent with many applications, in which the number of local entities evolving in each step of the DTMC is constant, hence the fraction of entities evolving goes to zero as $\frac{1}{N}$. It is easy to see that, under the assumption $\epsilon_N = \delta_N$, and assuming that for each N and τ there is a vector $\mathbf{v}_\tau$ such that $\hat{\mathbf{v}}_\tau^{(N)} = \delta_N \cdot \mathbf{v}_\tau$, we furthermore have that:

$$F^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau\in\hat{\mathcal{T}}^{(N)}} \mathbf{v}_\tau \cdot \hat{p}_\tau^{(N)}(\hat{\mathbf{d}}). \tag{21}$$

so that property (20) can be proved by showing that each $\hat{p}_\tau^{(N)}(\hat{\mathbf{d}})$ has a Lipschitz limit $p_\tau(\hat{\mathbf{d}})$. This condition will be automatically verified if $\hat{p}_\tau^{(N)}(\hat{\mathbf{d}})$ does not depend on N and is Lipschitz, as is the case in many applications.

*Remark** 5.1. The notion of intensity is often found in the mean field literature for DTMCs with a more general flavour. In general, it is introduced in the context of models of N interacting objects the evolution of which is described from a local perspective, namely from the point of view of a single object. In contrast, in our modelling approach we take a global perspective, describing the evolution at the system level.

According to the local description approach, if we have a system composed of N objects/entities, as is the case for the example of Section 2.1, the intensity ϵ_N is usually interpreted as the probability with which a specific object makes a transition in a step of the global DTMC (derived from the local description), or equivalently as the expected fraction of objects performing a transition in a single step[13].

According to this interpretation, in our context, by assigning temporal duration δ_N to each step of the PDTMC, we are essentially assuming that the

[13]We are implicitly assuming here that all local components can always perform at least one transition in each step. This can be always enforced by appropriately adding *dummy* transitions.

expected number of global steps necessary in order to let each object execute a local transition is $\frac{1}{\delta_N}$, so that the expected time for an object to make two successive transitions (i.e. the expected delay between the same object being selected twice to execute a transition) is kept equal to 1 for each N. This justifies the use of the word intensity to denote the scaling factor for time.

Note that, in general, the intensity is not forced to be equal to $\frac{1}{\delta_N}$, but can encompass more general situations, like those in which the number of entities evolving per step is $\Theta(\log N)$. The results of the following section, however, remain valid also in this more general setting, as long as intensity goes to zero with N and drift scales appropriately, see also [42].

Finally, we observe that the scaling conditions presented here are slightly more restrictive than those of [42]. In particular, condition (H3) of [42], which is a condition on the variance of the number of agents moving at each step, is automatically satisfied because in our modelling framework this number is bounded above by a constant (independent of N). Note also that in [42] more complex models are also considered, in which a resource component, representing the environment, is present in a single copy but interacts with increasing speed with the population of agents. The limit behaviour for this class of models will be discussed in Section 8.

5.2. Deterministic Approximation Theorems

We can now state the main approximation theorem for PDTMCs [32, 42]. Consider a sequence of normalised PDTMC models $(\hat{\mathcal{X}}_D^{(N)})_{N\geq N_0}$ with $\hat{\mathcal{X}}_D^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$, and denote by $\hat{\mathbf{X}}^{(N)}(k)$ the discrete-time Markov process associated with $\hat{\mathcal{X}}_D^{(N)}$. In addition, let $\hat{\mathbf{X}}_c^{(N)}(t) =_{\text{def}} \hat{\mathbf{X}}^{(N)}(\lfloor \frac{t}{\epsilon_N} \rfloor)$ be the process in continuous time associated with $\hat{\mathbf{X}}^{(N)}(k)$ and let $\hat{\mathbf{x}}(t)$ be the solution of the initial value problem $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$, $\hat{\mathbf{x}}(0) = \hat{\mathbf{d}}_{\mathbf{0}}$, where F is as in (20).

Theorem 5.1 (Deterministic approximation of PDTMCs)**.** *Let the sequence* $(\hat{\mathbf{X}}_c^{(N)}(t))_{N\geq N_0}$ *and* $\hat{\mathbf{x}}(t)$ *be defined as above, and assume Condition (18) and (20) hold. Then, for any* $T < \infty$*, it holds that:*

$$\lim_{N\to\infty} \mathbb{P}\{ \sup_{0\leq t\leq T} ||\hat{\mathbf{X}}_c^{(N)}(t) - \hat{\mathbf{x}}(t)|| > \varepsilon\} = 0.$$

■

Note that this result is different from those for PCTMC models, as we deliberately avoided any discussion about exit times and restrictions of the state space. We will present the justification for this in Section 6.

5.3. Main Example (continued)

We consider again the PDTMC version of the network epidemic model of Section 2.1. Proceeding as in Section 4.2, we set $\gamma_N = N$, so that we are studying the behaviour of the system for large populations. Consequently, the sequence of non-normalised PDTMC models of interest to us is $(\mathcal{E}_D^{(N)})_{N \geq N_0}$, for $\mathcal{E}_D^{(N)}$ as defined in Section 3.2.

The sequence of normalised models is $(\hat{\mathcal{E}}_D^{(N)})_{N \geq N_0}$, where $\hat{\mathcal{E}}_D^{(N)}$ is derived as described in Section 3. Notice that $\delta_N = \frac{1}{N}$ and $\hat{\mathcal{D}}^{(N)} \subseteq \mathcal{S}^4$ for all N, i.e. $\hat{X}_s + \hat{X}_e + \hat{X}_i + \hat{X}_r = 1$. Consequently, we choose the state space E as $E =_{\text{def}} \mathcal{S}^4$, the unit simplex in $\mathbb{R}^4$.

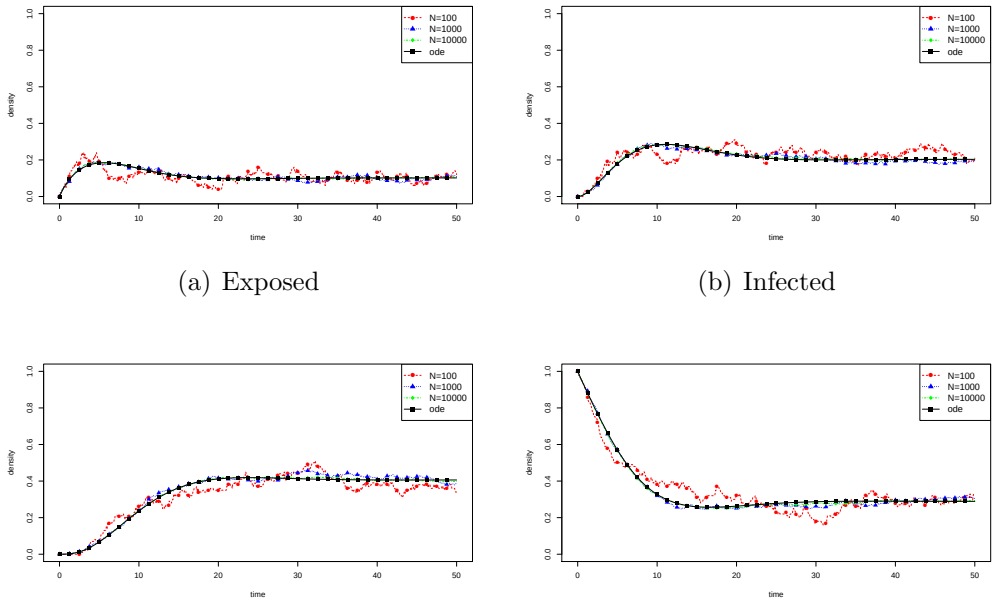
A simple inspection of the probability functions in Section 3.2 allows us to conclude that the probability functions do not depend on N and they are Lipschitz; consequently, also the drift does not depend on N and is Lipschitz. Therefore, we take the drift itself as the vector field according to (20) in order to get the following system of ODEs:

$$\begin{aligned}
\tfrac{d\hat{x}_s(t)}{dt} &= -\alpha_e \cdot \hat{x}_s(t) - \alpha_i \cdot \hat{x}_s(t) \cdot \hat{x}_i(t) - \alpha_p \cdot \hat{x}_s(t) + \alpha_s \cdot \hat{x}_r(t) \\
\tfrac{d\hat{x}_e(t)}{dt} &= \alpha_e \cdot \hat{x}_s(t) + \alpha_i \cdot \hat{x}_s(t) \cdot \hat{x}_i(t) - \alpha_a \cdot \hat{x}_e(t) - \alpha_q \cdot \hat{x}_e(t) \\
\tfrac{d\hat{x}_i(t)}{dt} &= \alpha_a \cdot \hat{x}_e(t) - \alpha_r \cdot \hat{x}_i(t) \\
\tfrac{d\hat{x}_r(t)}{dt} &= \alpha_r \cdot \hat{x}_i(t) + \alpha_p \cdot \hat{x}_s(t) + \alpha_q \cdot \hat{x}_e(t) - \alpha_s \cdot \hat{x}_r(t).
\end{aligned} \tag{22}$$

A visual depiction of the theorem can be found in Figure 12, where the ODE trajectories are compared with trajectories of the DTMCs with different population levels.

6. Comparing PCTMC and PDTMC deterministic approximations*

In this section we will compare the deterministic approximation results for PCTMCs and PDTMCs. We will not provide many technical details, but the interested reader can find these in Appendix C.



(a) Exposed (b) Infected

(c) Recovered (d) Susceptible

Figure 12: Comparison between the solution of the ODE and stochastic trajectories of DTMCs for increasing population sizes. Parameters of the model are $\alpha_e = 0.1$, $\alpha_i = 0.2$, $\alpha_a = 0.4$, $\alpha_r = 0.2$, $\alpha_p = 0$, $\alpha_q = 0$, and $\alpha_s = 0.1$, while initial conditions of ODE are $\hat{S}_0 = 1$, $\hat{E}_0 = 0$, $\hat{I}_0 = 0$, and $\hat{R}_0 = 0$.

Consider again the network epidemic example of Section 2.1, for which we presented both a PCTMC model and a PDTMC one, and fix a constant $\Lambda > 0$ such that the exit rate of the PCTMC model with population size N, $\mathbf{X_C}^{(N)}(t)$, is bounded in each state $\mathbf{d}$ by $N\Lambda$. Such a constant exists, because the normalised rates are defined in a compact set; hence their sum has a maximum value. Now, consider the PDTMC model $\mathbf{X_D}(k)$ (see Section 5.3), and assume that the constants α_σ, $\sigma \in \{e, i, a, r, p, q, s\}$, of the model satisfy the relation $\alpha_\sigma = \frac{\lambda_\sigma}{\Lambda}$, where λ_σ are the rate constants of the PCTMC model. Under this assumption, an easy calculation shows that $F_D(\hat{\mathbf{X}}) = \frac{1}{\Lambda} \cdot F_C(\hat{\mathbf{X}})$: the two limit dynamics are essentially the same (modulo a rescaling of time), as can also be seen looking at Figures 6 and 12.

In fact, this is not a coincidence. Readers familiar with uniformiza-

tion[14] [62] will easily recognise that the relationship $\alpha_\sigma = \frac{\lambda_\sigma}{\Lambda}$ is the one occurring between a PCTMC and the PDTMC associated with it by uniformization, using $N\Lambda$ as an upper bound for the rates. More precisely, the previous relation between vector fields is always in force between a PCTMC model and the PDTMC model associated with it via uniformization. This means that, for each sequence of PCTMC models, we can construct, using the probabilities obtained by uniformization, a sequence of PDTMC models sharing the same limit behaviour. Furthermore, given a sequence of PDTMC models, we can always construct a sequence of PCTMC models having the same limit behaviour. Essentially, this means that there is a very close relationship between the limit theorems of Sections 4 and 5: in a certain sense, they are two facets of the same result. This relationship is not only of theoretical interest, but also pragmatic. It means that one can apply results about limit behaviour proved for PDTMC models to PCTMC ones and vice versa.

A formal argument for this will be given in Theorem 6.1 below. To fix the notation, let $(\hat{\mathcal{X}}_C^{(N)})_{N\geq N_0}$ be a sequence of normalised PCTMC models, with respect to an increasing system size γ_N, admitting deterministic approximation, and let $(\hat{\mathcal{X}}_D^{(N)})_{N\geq N_0}$ the sequence of normalised PDTMC models, associated with the PCTMC models by uniformization, for constants $\Lambda_N \geq \sup_{\hat{\mathbf{d}}\in E} R^{(N)}(\hat{\mathbf{d}})$, such that $\Lambda_N = \Theta(\gamma_N)$. Now, we define two processes: $\hat{\mathbf{X}}_{\mathbf{ds}}^{(N)}(t) =_{\text{def}} \hat{\mathbf{X}}_{\mathbf{D}}^{(N)}(\lfloor \frac{t}{\epsilon_N} \rfloor)$, obtained by assuming that the length of each step in the DTMC is deterministically fixed and equal to ϵ_N, and $\hat{\mathbf{X}}_{\mathbf{rs}}^{(N)}(t) =_{\text{def}} \hat{\mathbf{X}}_{\mathbf{D}}^{(N)}(\mathcal{N}(\frac{t}{\epsilon_N}))$, where $\mathcal{N}$ is a Poisson random variable, hence the number of steps at time t is random (and, consequently, $\hat{\mathbf{X}}_{\mathbf{rs}}^{(N)}(t)$ is a CTMC).

Theorem 6.1. *Let $\hat{\mathbf{X}}_{\mathbf{ds}}^{(N)}(t)$ and $\hat{\mathbf{X}}_{\mathbf{rs}}^{(N)}(t)$ be defined as above. If either of the processes $\hat{\mathbf{X}}_{\mathbf{ds}}^{(N)}(t)$ and $\hat{\mathbf{X}}_{\mathbf{rs}}^{(N)}(t)$ converges to a limit process $\hat{\mathbf{X}}(t)$ weakly, then so does the other, and moreover $\hat{\mathbf{X}}_{\mathbf{ds}}^{(N)}(t)$ converges to $\hat{\mathbf{X}}_{\mathbf{rs}}^{(N)}(t)$ weakly.*

This result has an interesting corollary, namely that in the limit, considering steps of fixed (DTMCs) or variable (CTMCs) length is irrelevant, as the limit behaviour is the same in both cases. Furthermore, every limit result

[14] A quick introduction to uniformization, for readers not familiar with it, can be found in Appendix C.

holding for PDTMCs models holds also for PCTMCs models, and vice versa, provided that both models satisfy the same scaling assumptions.

An immediate application of this fact is the extension of Theorem 5.1 (deterministic approximation of PDTMCs) along the lines of Theorem 4.3 (deterministic approximation of PCTMCs): if we have a subset $S \subseteq E$, we can study the exit time distribution from S in the limit of $N \to \infty$ in the same way as for PCTMC models. Moreover, we can also estimate the number of steps required to escape from S, for large N.

7. Stationary Regime

The deterministic approximation theorems for PCTMC and PDTMC models we presented in Sections 4 and 5 do not assume any special property of the limit system of ODEs (apart from the existence and uniqueness of solutions, guaranteed by the Lipschitz property of the vector field). A consequence of this very general setting is that these theorems provide convergence in any finite time horizon. However, they do not give any indication about the relationship to the stationary regime of the sequence of Markov models, i.e. their behaviour as time goes to infinity. Indeed, as we argued in Remark 4.4, this is connected with the fact that the theorems have to work even for unstable trajectories, or in chaotic situations, i.e. for systems extremely sensitive to their initial conditions. It is therefore reasonable to expect that, assuming some stability properties of the system of ODEs, we may be able to obtain results also for the stationary regime. Indeed, this is the case: if the system of ODEs has a unique globally stable stationary point[15], we can show that the sequence of stochastic processes will end up precisely in that point [32, 42, 61].

Such results have been obtained both for the PDTMC case [32, 42] and the PCTMC case [32, 63]. Note that, by virtue of the discussion in Section 6, results for PDTMC and PCTMC are interchangeable. Moreover, the results essentially depend on having exponential bounds for convergence in probability; hence they can be straightforwardly exported to the more general setting of this paper (see Remark 4.4). A different proof, assuming a weaker

[15] A system of ODEs has a unique globally stable stationary point if and only if there is a point $\mathbf{d}_* \in E$ such that all trajectories converge to it in the limit, i.e. if for all $\mathbf{x_0} \in E$, denoting $\mathbf{x}(t)$ be the solution of ODE with initial point $\mathbf{x_0}$, it holds that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{d}_*$.

hypothesis[16], can be found in [64].

For any vector field F, the *Birkhoff centre*, $\mathcal{B}$, is informally the set of limit points within the trajectories generated by the field F from any initial starting point[17].

Consider a sequence of Markov chains (in either discrete or continuous time) that admits a deterministic limit, defined by a vector field F, and suppose that each chain in the sequence has one or more invariant probability distributions (invariant measure or steady state) [13], i.e. distributions that are invariant under the dynamics of the process.[18] Let μ_N be one such invariant measure, for each N.

Theorem 7.1 ([42, 61])**.** *Every limit measure μ of a sequence of invariant measures μ_N has support contained in the Birkhoff centre $\mathcal{B}$.*

Suppose now the state space E of the limit process is compact. Then any sequence $(\mu_N)_{N\geq 0}$ of invariant measures is tight[19], hence, from the Prokhorov's Theorem, it follows that each subsequence one can extract another subsequence convergent (weakly) to a limit measure [65]. Furthermore, each such Markov chain model has a finite state space (since E is compact) and so, if the chain is irreducible, and in the case of DTMCs also aperiodic, then the invariant measure μ_N is unique [13].

If, in addition, the system of ODEs has a unique globally stable fixed point $\hat{\mathbf{d}}_*$, then its Birkhoff centre is $\mathcal{B} = \{\hat{\mathbf{d}}_*\}$. In this case, in fact, all trajectories of the ODE converge to $\hat{\mathbf{d}}_*$ as time goes to infinity.

If the sequence μ_N is tight and the ODEs have a unique globally stable fixed point, the following corollary can be derived from the previous theorem.

Corollary 7.1 ([42])**.** *Every sequence of invariant measures μ_N converges weakly to the Dirac distribution on $\hat{\mathbf{d}}_*$ (i.e. the distribution with mass con-*

[16]Essentially weak convergence of $\hat{\mathbf{X}}^{(N)}(t)$ to $\hat{\mathbf{x}}(t)$ for each $t \geq 0$.

[17]Formally, we can define the set of omega points of $\hat{\mathbf{d}}$, $\omega(\hat{\mathbf{d}})$, as the set of all those points $\hat{\mathbf{d}}_1$ for which there exists a divergent sequence of time instants $t_k \to \infty$ such that $\lim_{k\to\infty} \hat{\mathbf{x}}(t_k) = \hat{\mathbf{d}}_1$, where $\hat{\mathbf{x}}(t)$ is the solution of $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$ starting from $\hat{\mathbf{d}}$ at time 0. The Birkhoff centre is then defined as $\mathcal{B} = cl\{\hat{\mathbf{d}} \mid \hat{\mathbf{d}} \in \omega(\hat{\mathbf{d}})\}$.

[18]If the process $\hat{\mathbf{X}}^{(N)}(t)$ is initially distributed according to the invariant distribution μ, then its probability distribution at each time t will always be μ.

[19]A sequence $(\mu_N)_{N\geq 0}$ is tight if and only if, for each $\epsilon > 0$, there is a compact set K_ϵ such that $\mu_N(K_\epsilon) > 1 - \epsilon$ for each N.

centrated on $\hat{\mathbf{d}}_*$*). Furthermore,*

$$\lim_{N\to\infty}\lim_{t\to\infty}\hat{\mathbf{X}}^{(N)}(t) = \lim_{t\to\infty}\lim_{N\to\infty}\hat{\mathbf{X}}^{(N)}(t) = \hat{\mathbf{d}}_*.$$

Hence, in this case we can exchange the limit in t and in N, an operation not possible in general.

This corollary is at the basis of the so-called *fixed point method* [42], which approximates the stationary distribution of $\hat{\mathbf{X}}^{(N)}(t)$ (assuming it is unique) with $\hat{\mathbf{d}}_*$, when the equation $F(\hat{\mathbf{d}}) = 0$ has a unique solution.

Main Example (continued). Consider again the epidemic example of Sections 3.2 and 4.2. If we solve the equation $F(\hat{\mathbf{x}}) = 0$, under the additional constraint that $\hat{x}_s + \hat{x}_e + \hat{x}_i + \hat{x}_r = 1$, one can verify that, for parameters as in Figure 3, the solution is unique and equal to $\hat{x}_s^* = 0.2889$, $\hat{x}_e^* = 0.106$, $\hat{x}_i^* = 0.2032$, and $\hat{x}_r^* = 0.4063$. To apply the previous corollary, we should additionally verify that this point is globally stable. In general, this is quite difficult to prove. For instance, one can prove its stability by looking at the eigenvalues of the Jacobian matrix and then exhibit a Lyapunov function defined on the whole state space [66]. We refrain from doing this here. However, a quick inspection of the phase space picture of the system in Figure 13(a), suggests that this is indeed the case.

Example: malware system. We stress here the fact that Corollary 7.1 can be applied if and only if $F(\hat{\mathbf{d}}) = 0$ has a unique solution *and* one proves that the obtained fixed point is globally stable. Unfortunately, there are cases in which the second hypothesis does not hold.

We illustrate this by considering a simple modification of the running example (see Section 3.1), taking inspiration from [42]. In particular, we consider a modified mechanism for the infection of susceptible nodes, in the following sense:

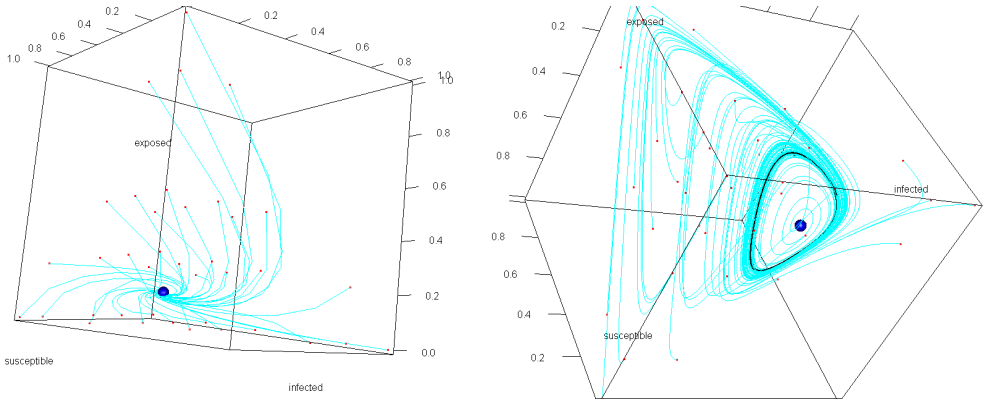
- we assume that the worm epidemic is spread by exposed nodes (called *dormant* in [42]). Hence, we modify the rate in the transition $\tau_i = (\inf, \mathbf{e_S}, \mathbf{e_E}, r_i^{(N)})$ to $r_i^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_i \cdot X_s \cdot \frac{X_e}{N}$;
- we assume that the worm behaves like a piece of malware: when it becomes active, it damages the normal operation status of the node. It also tries to activate other exposed nodes, using a communication channel with limited bandwidth (e.g. a wireless network). This is captured by modelling this activation by a saturating function, adding

to the epidemic model the transition $\tau_{aa} = (\inf, \mathbf{e_E}, \mathbf{e_I}, r_{aa}^{(N)})$, with $r_{aa}^{(N)}(\mathbf{X}) =_{\text{def}} \lambda_{aa} \cdot X_i \cdot \frac{X_e}{k_{aa}N+X_e}$. Note the dependency on N of the parameter in the denominator of the saturating function $\frac{X_e}{k_{aa}N+X_e}$, guaranteeing that the normalized rate is $N\lambda_{aa}\hat{X}_i\frac{\hat{X}_e}{k_{aa}+\hat{X}_e}$.

If we consider the parameter set as in Figure 13, we can easily check that the set of ODEs associated with this modified epidemic model has a unique fixed point, namely $\hat{x}_s^* = 0.23803$, $\hat{x}_e^* = 0.09207$, $\hat{x}_i^* = 0.44599$, and $\hat{x}_r^* = 0.22391$. However, as evident in Figure 13(b), the system has a limit cycle, and the fixed point is an unstable equilibrium. In this case, Theorem 7.1 cannot be applied to estimate the steady state behaviour of the CTMC model. The cyclic nature of this epidemic is intuitively caused by the non-linear recruitment of exposed nodes by infected ones and by the fact that, once the malware is active, it cannot spread the infection anymore. This creates phases in the epidemic: when there is a large number of exposed nodes the activation rate of the malware is high. The infected nodes so generated then quickly activate the exposed ones, reducing the number of exposed nodes to a small value. The recruitment effect remains large until patching reduces the number of infected nodes to the point at which a new wave of exposed nodes can spread in the network.

8. Further topics

In this section, we briefly discuss some more advanced topics in order to give the reader an impression of some of the consequences of the limit results. Detailed coverage of these research areas would be the basis of a further tutorial, and beyond the scope of this paper. Our intention here is just to offer a broader perspective on continuous approximation. Of course, even this cannot be exhaustive and we do not consider mean field methods for stochastic models with spatial aspects, whose limits are generally expressed in terms of partial differential equations [67, 68], continuous approximation of Markov Decision Processes [69], in which limit results are used for the solution of optimisation problems, fluid approximation and mean field limits for infinite dimensional systems, like Supermarket models [70, 71], or deterministic approximation algorithms [33], where these techniques, and many variations of them, have been widely applied.



(a) Globally attracting fixed point (b) Limit cycle

Figure 13: Phase portrait of the main epidemic example (left) and of the malware epidemic example (right). On the left we can see the trajectories converging to a fixed point, depicted as a blue sphere in the figure. On the right, we can see that trajectories converge to a limit cycle, in black in the figure. The fixed point, shown as a blue sphere, is an unstable equilibrium, and trajectories starting nearby spiral out towards the limit cycle. Parameters in the model are $\lambda_i = 0.1$, $\lambda_e = 0.0002$, $\lambda_{aa} = 0.01$, $k_{aa} = 0.1$, $\lambda_r = 0.005$, $\lambda_q = 0.0001$, $\lambda_p = 0$, $\lambda_a = 0.001$, $\lambda_s = 0.01$.

Fluid ODE and average behaviour of PCTMC models. Consider a sequence $\hat{\mathcal{X}}^{(N)}$ of PCTMC models amenable of deterministic approximation. For any given N, one can compute the average value $\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]$ of variables $\hat{\mathbf{X}}^{(N)}$, for any fixed time t. The deterministic approximation results of Section 4 guarantee that $\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]$ converges in probability to $\hat{\mathbf{x}}(t)$, uniformly in any bounded time interval.

However, for any N, it is also possible to derive an ODE describing the exact evolution of $\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]$, applying the Dynkin formula [72] or using standard arguments involving the Kolmogorov Forward Equation (or Master Equation) of the CTMC [73, 74] — the set of ODEs describing the time evolution of the probability mass in the state space of the CTMC. More specifically, the exact evolution of $\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]$ is given by the following set of ODEs:

$$\frac{d}{dt}\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)] = \mathbb{E}[F^{(N)}(\hat{\mathbf{X}}^{(N)}(t))]. \tag{23}$$

Using Taylor expansion or generating function arguments [73, 74, 8], it can be shown that the limit ODE is a *first order approximation* of the true average equation. To see why this is true, assume $F^{(N)}(\hat{\mathbf{X}})$ is smooth and expand it

around $\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]$: $F^{(N)}(\hat{\mathbf{X}}^{(N)}(t)) \approx F^{(N)}(\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]) + \nabla F^{(N)}(\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]) \cdot (\hat{\mathbf{X}}^{(N)}(t) - \mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)])$. By inserting this formula into Equation (23), observing that $\mathbb{E}[(\hat{\mathbf{X}}^{(N)}(t) - \mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)])] = 0$, one gets

$$\frac{d}{dt}\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)] \approx F^{(N)}(\mathbb{E}[\hat{\mathbf{X}}^{(N)}(t)]).$$

It is also possible to show that higher order terms in the approximation vanish in the limit as N goes to infinity (hence, the larger N, the better the approximation works). Furthermore, the second order terms in the approximation depend on the variance and covariance of the process, third order terms depend on third order moments, and so on. Hence, it is also possible to derive coupled differential equations describing the variance and higher order moments, at the price of introducing further variables in the system of ODEs (precisely, $\Theta(n^k)$ new variables to include terms up to order k) [74, 8]. Techniques constructing ODEs approximating moments of a Markov chain, not necessarily based on Taylor expansions, are generally known as moment closure techniques [8, 72].

Decoupling of Joint Probability. A perhaps surprising consequence of the deterministic results is that in a system comprised of N interacting objects, each of which may be in k distinct states, the joint distribution over the local states of the model displays asymptotic independence. Let $Y^{(N)}(t) \in \{1, \ldots, k\}$ denote the state of a single "tagged" object, in the N-th CTMC, at time t, and let $Y(t)$ denote its state in the limit model. Due to the (implicit) assumption that objects are indistinguishable, the probability that $Y^{(N)}(t) = k_1$ equals $\hat{X}^{(N)}_{k_1}(t)$, i.e. the probability of choosing at random one object in state k_1. By taking the limit, it follows that

$$\mathbb{P}\{Y(t) = k_1\} = \hat{\mathbf{x}}_{k_1}(t).$$

Furthermore, it can be proved [57] that agents become asymptotically independent, hence

$$\begin{aligned}\mathbb{P}\{Y_1(t) = k_1, \ldots, Y_h(t) = k_h\} &= \mathbb{P}\{Y_1(t) = k_1\} \cdot \ldots \cdot \mathbb{P}\{Y_h(t) = k_h\} \\ &= \hat{x}_{k_1}(t) \cdot \ldots \cdot \hat{x}_{k_h}(t).\end{aligned}$$

Moreover, the deterministic approximation theorems presented in Sections 2, 4, and 5, guarantee that

$$\mathbb{P}\{Y^{(N)}_1(t) = k_1, \ldots, Y^{(N)}_h(t) = k_h\} \to \mathbb{P}\{Y_1(t) = k_1, \ldots, Y_h(t) = k_h\},$$

and hence, for large N,

$$\mathbb{P}\{Y_1^{(N)}(t) = k_1, \ldots, Y_h^{(N)}(t) = k_h\} \approx \hat{x}_{k_1}(t) \cdot \ldots \cdot \hat{x}_{k_h}(t).$$

This asymptotic independence property is known as the *decoupling assumption* [41, 42] or *propagation of chaos* [27]. Basically, it means that, in the limit, the evolution of each object becomes independent from the other objects. This happens even in models with explicit cooperation. Consider, for instance, a model in which an object in state s_1 and an object in state s_2 cooperate and evolve to states s_3 and s_4, respectively, at rate kX_1X_2. In this case, there is a clear dependency in the model between objects in states s_1 and s_2, as their states change simultaneously. Nevertheless, in the limit, this dependence is lost. In fact, we obtain the same limit process as for a model in which objects in state s_1 evolve to s_3 independently of objects in s_2 evolving to s_4, in both cases with rate kX_1X_2. In terms of transitions, the model containing the transition $(a_\tau, \mathbf{e_1} + \mathbf{e_2}, \mathbf{e_3} + \mathbf{e_4}, kX_1X_2)$ has the same *limit behaviour* of the model containing the transitions $(a_{\tau_1}, \mathbf{e_1}, \mathbf{e_3}, kX_1X_2)$ and $(a_{\tau_2}, \mathbf{e_2}, \mathbf{e_4}, kX_1X_2)$. Note that, by splitting the transition into two, the global rate R doubles, and so does the density of events.

This asymptotic independence property is a key result that allows us to estimate stochastic properties of single agents in large populations by using the mean field limit [75]. This property has been recently exploited for the computation of *fluid passage times* [76] and also for *fluid model checking* of continuous stochastic logic properties [77].

It is also important to note that the asymptotic independence holds for any fixed time t, but not necessarily in the limit for $t \to \infty$. This is because the decoupling is based on theorems of Sections 2, 4, and 5, which hold in a finite time horizon. However, under the assumptions of Corollary 7.1, one can extend independence to the stationary regime:

$$\lim_{t\to\infty} \mathbb{P}\{Y_1(t) = k_1, \ldots, Y_h(t) = k_h\} = \hat{x}^*_{k_1} \cdot \ldots \cdot \hat{x}^*_{k_h}.$$

Fast Simulation. The asymptotic independence property shared by models of interacting objects discussed above has an interesting consequence: it allows us to track the evolution of a single object ignoring all other objects, if N is sufficiently large. More specifically, we can consider the process $(Y^{(N)}(t), \hat{\mathbf{X}}^{(N)}(t))$, where Y tracks the evolution of the selected object, which is a Markov process. A consequence of the deterministic approximation theorems shows that $(Y^{(N)}(t), \hat{\mathbf{X}}^{(N)}(t))$ converges to a jump process $(Y(t), \hat{\mathbf{x}}(t))$,

where $\hat{\mathbf{x}}(t)$ follows the solution of the fluid differential equation (and is not influenced by the discrete state of the tracked object), while $Y(t)$ evolves according to a time-inhomogeneous process depending on the state of the other objects only through $\hat{\mathbf{x}}(t)$. This kind of fast simulation scheme can be defined both for a PDTMC [75] or a PCTMC [57] model.

As an example, consider again the PCTMC network epidemic model of Section 2.1. If we look at a single node in the network, we can define the time dependent infinitesimal generator matrix $\mathbf{Q}^{(N)}(t)$ at level N for $Y^{(N)}(t)$ as

$$\mathbf{Q}^{(N)}(t) = \begin{pmatrix} -\lambda_e - \lambda_i \hat{X}_i^{(N)}(t) - \lambda_p & \lambda_e + \lambda_i \hat{X}_i^{(N)}(t) & 0 & \lambda_p \\ 0 & -\lambda_a - \lambda_q & \lambda_a & \lambda_q \\ 0 & 0 & -\lambda_r & \lambda_r \\ \lambda_s & 0 & 0 & -\lambda_s \end{pmatrix},$$

which depends on N only via the fraction of infected agents present in the network at time t. Hence, in the limit of N going to infinity, $(Y^{(N)}(t), \hat{\mathbf{X}}^{(N)}(t))$ converges to the process $(Y(t), \hat{\mathbf{x}}(t))$, in which $\hat{\mathbf{x}}(t)$ is the solution of the system of ODE shown in Section 4, and $Y(t)$ is a continuous-time jump process on the state space, with time-dependent generator matrix $\mathbf{Q}(t)$, defined as $Q^{(N)}(t)$, with $\hat{X}_i^{(N)}(t)$ replaced by $\hat{x}_i(t)$.

Hybrid Approximations. The deterministic approximation results we considered in this paper rely on crucial assumptions on how the system scales as its size increases. In particular, they require that all variables of the model increase proportionally to the system size. This condition holds in many cases, but it may happen that some parts of our system cannot be assumed to behave in this way. Typical examples include network models where we have a constant number of servers and an increasing number of clients, or models of biological cellular systems explicitly representing genes, which are present in one or few copies within a cell [78, 79, 80].

In these cases, one may still want to consider deterministic approximation results only for those parts of the model which can be assumed to scale correctly with system size. The rest of the system, instead, has to be kept discrete. This leads to different deterministic approximation schemes, in which a sequence of stochastic models converges to a limit process whose nature depends on the structure of the model. In particular, we will briefly consider here sequences of PCTMC models converging to hybrid limit processes, i.e. processes presenting an evolution in terms of differential equations and

stochastic jumps [81, 82, 80], and sequences of PDTMC models in which the intrinsically discrete component evolves so fast that it immediately reaches the equilibrium [42].

The former class of systems emerges naturally when we consider systems with entities whose number does not grow with N, such as genes in a cellular model. The evolution of the rest of the system, however, can depend on the internal state of these discrete entities. In this case, as we increase the size of the system, the sequence of PCTMC models converges [82] to a hybrid system, in which the evolution is given in terms of ODEs (for those components scaling with N) and of a time-inhomogeneous jump process (for the components independent of N). These models belong to the class of Piecewise Deterministic Markov Processes [81].

The second class has been studied e.g. in [42, 38], and includes models of objects interacting with a rapidly changing environment. In these models, we have N objects interacting with a resource R, which changes its state at each step of the DTMC. When we consider the sequence of PDTMC models for increasing population levels and scale the time accordingly, the resource performs an increasing number of transitions per unit of time. When R has an equilibrium distribution, in the limit this equilibrium will provably have been reached, and the ODEs are consequently modified to include these effects, by averaging transition probabilities that depend on the state of the resource with respect to the equilibrium distribution of R. For example, if the limit probability p_τ of a transition τ depends on the state r of the resource, $p_\tau = p_\tau(\hat{\mathbf{x}}, r)$, then in the ODE we use $p_\tau = \sum_r \pi_R(r) \cdot p_\tau(\hat{\mathbf{x}}, r)$. The convergence to such a limit ODE has been proved in [42].

Finally, a different class of hybrid limit processes is obtained when the rate functions of a CTMC are discontinuous [83]. In these situations, assuming a certain regularity in the nature of discontinuity of the rate functions, the sequence of CTMC converges to the solution of a Piecewise Smooth Dynamical System [84]. In particular, the queueing model of Section 4.5 falls into this last class of models. This approach can be generalised by considering mean field limits in terms of differential inclusions [63].

9. Conclusions

In this paper, we have presented continuous approximation results for stochastic Markov models, both in discrete time and in continuous time.

These results have been well known in the literature of stochastic approximations for 40 years. However they have only relatively recently attracted the widespread attention of the performance community. Nevertheless they are rapidly becoming important tools for studying models of systems composed of many interacting objects, thus circumventing the state space explosion problem.

Our objective has been to give a tutorial that presents the main ideas of the field at a level which is accessible to those who do not necessarily have a detailed background in the relevant theory of stochastic processes and deterministic approximation. We have included a bibliography which references the relevant papers for the theory as well as some of those presenting applications and more advanced approximation techniques.

Our presentation has discussed the main approximation theorems, and several applications and extensions, in a uniform way. We have refrained from showing the mathematical details of the proofs, which can be found in excellent reference papers and books, such as [57, 55, 58], in order to emphasise a more intuitive account of the results. In particular, we have focused on the conditions that must be satisfied in order for the theorems to be applied, showing how to check them in practical examples.

References

[1] F. Ciocchetta, J. Hillston, Process Algebras in Systems Biology, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), SFM, Vol. 5016 of Lecture Notes in Computer Science, Springer, 2008, pp. 265–312.

[2] M. D. Harrison, M. Massink, Modelling Interactive Experience, Function and Performance in Ubiquitous Systems, Electr. Notes Theor. Comput. Sci. 261 (2010) 23–42.

[3] G. Bolch, S. Greiner, H. de Meer, K. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, Wiley, 2006.

[4] C. Baier, B. Haverkort, H. Hermanns, J.-P. Katoen, Model-Checking Algorithms for Continuous-Time Markov Chains, IEEE Trans. on Soft. Eng. 29 (6) (2003) 524–541.

[5] J. Hillston, Fluid Flow Approximation of PEPA models, in: Proceedings of the Second International Conference on the Quantitative Evaluation of Systems (QEST 2005), 2005, pp. 33–43.

[6] N. Geisweiller, J. Hillston, M. Stenico, Relating Continuous and Discrete PEPA Models of Signalling Pathways, Theor. Comput. Sci. 404 (1-2) (2008) 97–111.

[7] L. Bortolussi, A. Policriti, Dynamical Systems and Stochastic Programming: To Ordinary Differential Equations and Back, T. Comp. Sys. Biology 11 (2009) 216–267.

[8] R. Hayden, J. Bradley, A fluid analysis framework for a Markovian process algebra, Theor. Comput. Sci. 411 (22-24) (2010) 2260–2297.

[9] L. Cardelli, On process rate semantics, Theor. Comput. Sci. 391 (3) (2008) 190–215.

[10] J. Hillston, A Compositional Approach to Performance Modelling, Cambridge University Press, 1996.

[11] U. Herzog, EXL: Syntax, semantics and examples, Tech. Rep. IMMD7-16-90, Universität Erlangen-Nürnberg (1990).

[12] H. Hermanns, M. Rettelbach, Syntax, semantics, equivalences, and axioms for MTIPP, in: in Proc. of the 2nd Workshop on Process Algebras and Performance Modelling (PAPM '94), 1994, pp. 71–87.

[13] J. R. Norris, Markov Chains, Cambridge University Press, 1997.

[14] T. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, Journal of Applied Probability 7 (1970) 49–58.

[15] J. Bradley, S. Gilmore, J. Hillston, Analysing distributed Internet worm attacks using continuous state-space approximation of process algebra models, Journal of Computer and System Sciences 74 (6) (2008) 1013–1032.

[16] M. Calder, S. Gilmore, J. Hillston, Automatically deriving ODEs from process algebra models of signalling pathways, in: Proceedings of Computational Methods in Systems Biology (CMSB), 2005.

[17] F. Ciocchetta, J. Hillston, Bio-PEPA: A framework for the modelling and analysis of biological systems, Theoretical Computer Science 410 (33-34) (2009) 3065 – 3084.

[18] M. Tribastone, S. Gilmore, Scaling Performance Analysis Using Fluid-Flow Approximation, Vol. 6582 of Lecture Notes in Computer Science, Springer, 2011, pp. 486–505.

[19] M. Tribastone, Relating layered queueing networks and process algebra models, in: Proceedings of the first joint WOSP/SIPEW International Conference on Performance Engineering, 2010, pp. 183–194.

[20] A. Benoit, M. Cole, S. Gilmore, J. Hillston, Enhancing the effective utilisation of grid clusters by exploiting on-line performability analysis, in: 5th International Symposium on Cluster Computing and the Grid (CCGrid 2005), 9-12 May, 2005, Cardiff, UK, IEEE Computer Society, 2005, pp. 317–324.

[21] M. Massink, D. Latella, A. Bracciali, M. Harrison, A Scalable Fluid Flow Process Algebraic Approach to Emergency Egress Analysis, in: Proceedings of 8th IEEE International Conference on Software Engineering and Formal Methods (SEFM 2010), 2010, pp. 169–180.

[22] M. Massink, D. Latella, A. Bracciali, J. Hillston, Modelling non-linear crowd dynamics in Bio-PEPA, in: Proceeding of 14th International Conference on Fundamental Approaches to Software Engineering (FASE 2011), Vol. 6603 of LNCS, Springer, 2011, pp. 96–110.

[23] R. David, H. Alla, Discrete, Continuous, and Hybrid Petri Nets, 2nd ed., Springer-Verlag, 2010. doi:10.1007/978-3-642-10669-9_5.

[24] K. Trivedi, V. Kulkarni, FSPNs: Fluid stochastic Petri net, in: M. A. M. et al. (Ed.), Application and Theory of Petri Nets 1993, Vol. 691 of LNCS, Springer, 1993, pp. 24–31.

[25] G. Horton, V. Kulkarni, D. Nicol, K. Trivedi, Fluid stochastic Petri nets: Theory, applications and solution techniques, European Journal of Operational Research 105 (1998) 184–201.

[26] M. Silva, L. Recalde, On fluidification of Petri net models: from discrete to hybrid and continuous models, Annual Reviews in Control 28 (2004) 253–266.

[27] A. Sznitman, Topics in propagation of chaos, in: Ecole d'etè de Probabilitès de Saint-Flour XIX 1989, Vol. 1464 of Lecture Notes in Mathematics, Springer, 1991, pp. 165–251.
URL http://dx.doi.org/10.1007/BFb0085169

[28] D. Sumpter, D. Broomhead, Relating individual behaviour to population dynamics, Proceedings of the Royal Society B 268 (1470) (2001) 925–932. doi:10.1098/rspb.2001.1604.

[29] T. Kurtz, Approximation of population processes, SIAM, 1981.

[30] H. Andersson, T. Britton, Stochastic Epidemic Models and Their Statistical Analysis, Springer-Verlag, 2000.

[31] M. Benaïm, Recursive algorithms, urn processes and chaining number of chain recurrent sets, Ergodic Theory and Dynamical Systems 18 (01) (1998) 53–87.

[32] M. Benaïm, J. Weibull, Deterministic Approximation of Stochastic Evolution in Games, Econometrica 71 (3) (2003) 873–903.

[33] H. Kushner, G. Yin, Stochastic approximation and recursive algorithms and applications, Springer-Verlag, 2003.

[34] F. Baccelli, A. Chaintreau, D. D. Vleeschauwer, D. R. McDonald, HTTP turbulence, NHM 1 (1) (2006) 1–40.

[35] F. Baccelli, M. Lelarge, D. McDonald, Metastable regimes for multiplexed TCP flows, in: Proceedings of 42nd Annual Allerton Conference on Communication, Control and Computing, 2004.

[36] A. Martinoli, K. Easton, W. Agassounon, Modeling swarm robotic systems: a case study in collaborative distributed manipulation, The International Journal of Robotics Research 23 (4-5) (2004) 415–436.

[37] S. Kumar, L. Massouli, Integrating streaming and file-transfer internet traffic: fluid and diffusion approximations, Queueing Systems 55 (4) (2007) 195–205.

[38] C. Bordenave, D. McDonald, A. Proutiére, A particle system in interaction with a rapidly varying environment: Mean field limits and applications, NHM 5 (1) (2010) 31–62.

[39] C. Bordenave, D. McDonald, A. Proutiere, Performance of random medium access control, an asymptotic approach, in: Proceedings of the 2008 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, 2008, pp. 1–12.

[40] G. Sharma, A. Ganesh, P. Key, Performance Analysis of Contention Based Medium Access Control Protocols, IEEE Transactions on Information Theory 55 (4) (2009) 1665–1682.

[41] J.Y. Le Boudec, D. McDonald, J. Mundiger, A Generic Mean Field Convergence Result for Systems of Interacting Objects, in: Proceedings of Fourth International Conference on the Quantitative Evaluation of Systems (QEST 2007), 2007, pp. 3–18.

[42] M. Benaïm, J.Y. Le Boudec, A Class of Mean Field Interaction Models for Computer and Communication Systems, Performance Evaluation 65 (11-12) (2008) 823–838.

[43] A. Bobbio, M. Gribaudo, M. Telek, Analysis of large scale interacting systems by mean field method, in: Proceedings of Fifth International Conference on the Quantitative Evaluation of Systems (QEST 2008), 2008, pp. 215–224.

[44] R. Bakhshi, L. Cloth, W. Fokkink, B. Haverkort, Mean-field analysis for the evaluation of gossip protocols, in: Proceedings of Sixth International Conference on the Quantitative Evaluation of Systems (QEST 2009), 2009, pp. 247–256.

[45] N. M. N. Tsukada, R. Hirade, Fluid limit analysis of FIFO and RR caching for independent reference models, Performance Evaluation 69 (2012) 403–412.

[46] J. N. Tsitsiklis, K. Xu, On the power of (even a little) resource pooling, Accepted for publication in Stochastic Systems.

[47] T. Kurtz, Limit theorems for sequences of jump Markov processes approximating ordinary differential processes, Journal of Applied Probability 8 (1971) 244–356.

[48] C. McCaig, R. Norman, C. Shankland, From individuals to populations: A mean field semantics for process algebra, Theor. Comput. Sci. 412 (17) (2011) 1557–1580.

[49] L. Berec, Techniques of spatially explicit individual-based models: construction, simulation, and mean-field analysis, Ecological Modelling 150 (1–2) (2002) 55–81. doi:10.1016/S0304-3800(01)00463-X.

[50] M. Kwiatkowska, G. Norman, D. Parker, PRISM: Probabilistic Symbolic Model Checker, in: T. Field, P. Harrison, J. Bradley, U. Harder (Eds.), Proc. 12th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation (TOOLS'02), Vol. 2324 of LNCS, Springer, 2002, pp. 200–204.

[51] T. Dayar, H. Hermanns, D. Spieler, V. Wolf, Bounding the Equilibrium Distribution of Markov Population Models, Numerical Linear Algebra with Applications 18 (6) (2011) 931–946.

[52] M. Tribastone, S. Gilmore, J. Hillston, Scalable differential analysis of process algebra models, IEEE Trans. Software Eng. 38 (1) (2012) 205–219.

[53] L. Bortolussi, A. Policriti, Modelling Biological Systems in Concurrent Constraint Programming, Constraints 13 (1) (2008) 66–90.

[54] CompBio Group, Institute for Systems Biology, Seattle, Dizzy home page, `http://magnet.systemsbiology.net/software/Dizzy/`.

[55] T. Kurtz, S. Ethier, Markov Processes - Characterisation and Convergence, Wiley, 1986.

[56] R. Darling, Fluid limits of pure jump Markov processes: A practical guide, ArXiv Mathematics e-printsarXiv:arXiv:math/0210109.

[57] R. Darling, J. Norris, Differential equation approximations for Markov chains, Probability Surveys 5 (2008) 37–79. doi:10.1214/07-PS121.

[58] O. Kallenberg, Foundations of Modern Probability, Springer-Verlag, 2002.

[59] D. J. Wilkinson, Stochastic Modelling for Systems Biology, Chapman & Hall, 2006.

[60] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE Journal on Selected Areas in Communications 18 (3) (2000) 535–547.

[61] R. Hayden, Convergence of ODE approximations and bounds on performance models in the steady-state, in: Proceedings of PASTA 2010, 2010.

[62] A. Jensen, Markov chains as an aid in the study of Markov processes, Skandinavisk Aktuarietidskriff 36 (1953) 8791.

[63] N. Gast, B. Gaujal, Markov chains with discontinuous drifts have differential inclusion limits, Perform. Eval. 69 (12) (2012) 623–642.

[64] M. Benaïm, J.Y. Le Boudec, On mean field convergence and stationary regime, CoRR abs/1111.5710.

[65] P. Billingsley, Probability and Measure, John Wiley and Sons, 1979.

[66] S. H. Strogatz, Non-Linear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry and Engeneering, Perseus books, 1994.

[67] C. Kipnis, C. Landim, Scaling limits of interacting particle systems, Springer-Verlag, 1999.

[68] A. Chaintreau, J.Y. Le Boudec, N. Ristanovic, The age of gossip: spatial mean field regime, in: Proceedings of the Eleventh International ACM SIGMETRICS Joint Conference on Measurement and Modeling of Computer Systems, 2009, pp. 109–120.

[69] N. Gast, B. Gaujal, J.Y. Le Boudec, Mean field for Markov decision processes: from discrete to continuous optimization, CoRR abs/1004.2342.

[70] M. D. Mitzenmacher, The power of two choices in randomized load balancing, Ph.D. thesis, Harvard University (1991).

[71] M. Mitzenmacher, Load balancing and density dependent jump Markov processes, in: Proceedings of the 37th Annual Symposium on Foundations of Computer Science, FOCS 1996, 1996, pp. 213–222.

[72] A. Singh, J. Hespanha, Lognormal moment closures for biochemical reactions, in: Proceedings of 45th IEEE Conference on Decision and Control, 2006.

[73] N. V. Kampen, Stochastic Processes in Physics and Chemistry, Elsevier, 1992.

[74] L. Bortolussi, On the Approximation of Stochastic Concurrent Constraint Programming by Master Equation, in: Proceedings of the Sixth Workshop on Quantitative Aspects of Programming Languages (QAPL 2008), Vol. 220 of Electr. Notes Theor. Comput. Sci., 2008, pp. 163–180.

[75] N. Gast, B. Gaujal, A mean field model of work stealing in large-scale systems, in: Proceedings of ACM SIGMETRICS 2010, 2010, pp. 13–24.

[76] R. Hayden, A. Stefanek, J. Bradley, Fluid computation of passage-time distributions in large markov models, Theor. Comput. Sci. 413 (1) (2012) 106–141.

[77] L. Bortolussi, J. Hillston, Fluid model checking, in: M. Koutny, I. Ulidowski (Eds.), CONCUR, Vol. 7454 of Lecture Notes in Computer Science, Springer, 2012, pp. 333–347.

[78] L. Bortolussi, A. Policriti, (Hybrid) automata and (stochastic) programs. The hybrid automata lattice of a stochastic program, Journal of Logic and Computation. In print.

[79] L. Bortolussi, A. Policriti, The importance of being (a little bit) discrete, in: Proceedings of First International Workshop From Biology To Concurrency (and Back), FBTC 2008, Vol. 229 of Electr. Notes Theor. Comput. Sci., 2009, pp. 75–92. doi:10.1016/j.entcs.2009.02.006.

[80] L. Bortolussi, V. Galpin, J. Hillston, M. Tribastone, Hybrid semantics for PEPA, in: Proceedings of the Seventh International Conference on the Quantitative Evaluation of Systems (QEST 2010), 2010, pp. 181–190.

[81] M. Davis, Markov Models and Optimization, Chapman & Hall, 1993.

[82] L. Bortolussi, Limit Behavior of the Hybrid Approximation of Stochastic Process Algebras, in: Proceedings of the 17th International Conference on Analytical and Stochastic Modeling Techniques and Applications (ASMTA), 2010, pp. 367–381.

[83] L. Bortolussi, Hybrid limits of Continuous Time Markov Chains, in: QEST, IEEE Computer Society, 2011, pp. 3–12.

[84] J. Cortez, Discontinuous dynamical systems: A tutorial on solutions, nonsmooth analysis, and stability, IEEE Control Systems Magazine 28 (3) (2008) 36–73.

Appendix A. Notation

- X, Y, Z denote (random) variables of non-normalized CTMC or DTMC models.
- $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ indicate vectors of (random) variables of non-normalized CTMC or DTMC models.
- $\hat{X}$, $\hat{Y}$, $\hat{Z}$ denote (random) variables of normalized CTMC or DTMC models.
- $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, $\hat{\mathbf{Z}}$ indicate vectors of (random) variables of normalized CTMC or DTMC models.
- $\hat{x}$, $\hat{y}$, $\hat{z}$ denote variables of deterministic systems.
- $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ indicate vectors of variables of deterministic systems.
- $\mathbf{d}$, $\mathbf{d}_1$, and so on, denote points in $\mathcal{D}$ (non-normalized systems).
- $\hat{\mathbf{d}}$, $\hat{\mathbf{d}}_1$, and so on, denote points in $\mathbb{R}^n$ and $\hat{\mathcal{D}}$ (normalized systems).
- $\mathbf{1}\{\phi(\mathbf{X})\}$ is the indicator function of a predicate ϕ on variables $\mathbf{X}$.
- $\mathcal{S}_N^n = \{\mathbf{x} \in \{0, \ldots, N\}^n \mid \sum_{i=1}^n x_i = N\}$ is the N-simplex in $\mathbb{N}^n$
- $\mathcal{S}^n = \{\hat{\mathbf{x}} \in [0,1]^n \mid \sum_{i=1}^n \hat{x}_i = 1\}$ is ,the unit simplex in $\mathbb{R}^n$.
- τ denotes a transition of a PCTMC or PDTMC.

- $\mathbf{v}_\tau^{(N)}$ and $\hat{\mathbf{v}}_\tau^{(N)}$ are the non-normalized and normalized update vectors of transition τ.
- $r_\tau^{(N)}$ and $\hat{r}_\tau^{(N)}$ are the non-normalized and normalized rate functions of transition τ, for PCTMC models.
- $p_\tau^{(N)}$ and $\hat{p}_\tau^{(N)}$ are the non-normalized and normalized probability functions of transition τ, for PDTMC models.
- $g_\tau^{(N)}$ is the density dependent function for a density dependent transition τ of a PCTMC model.
- $R^{(N)}$ is the exit rate function of a PCTMC model.
- $\mu^{(N)}$ is the mean increment in a step of a PCTMC or PDTMC model.
- $\Sigma^{(N)}$ is the covariance matrix of one-step increments of a PCTMC or PDTMC model.
- $F^{(N)}$ and F are the drift and limit drift of a PCTMC or PDTMC model.
- ϵ_N is the intensity of a PDTMC model.
- $cl(A)$ denotes the closure of a set A, $int(A)$ denotes its interior, ∂A its boundary, and A^c its complement.

Appendix B. Proof of Theorem 4.1

Theorem 4.1 (Deterministic approximation for density dependent PCTMCs). *Let the sequence $(\hat{\mathbf{X}}^{(N)}(t))_{N\geq N_0}$ of Markov processes be density dependent and satisfy conditions of Section 4.1. Let $\hat{\mathbf{x}}(t)$ be the solution of $\frac{d\hat{\mathbf{x}}(t)}{dt} = F(\hat{\mathbf{x}}(t))$ starting from $\mathbf{x_0}$. Then, for any* finite *time horizon $T < \infty$, it holds that:*

$$\mathbb{P}\{\lim_{N\to\infty} \sup_{0\leq t\leq T} \|\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)\| = 0\} = 1,$$

i.e. $\sup_{0\leq t\leq T} \|\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)\|$ *converges to zero almost surely.*

Proof. We assume that all functions g_τ defining the rate of density dependent transitions are bounded, say by B_τ, and Lipschitz, say with Lipschitz constant L_τ.[20]

Consider now the representation of the CTMC $\hat{\mathbf{X}}^{(N)}(t)$ in terms of Poisson processes [29, 55]. We will use one Poisson process $\mathcal{N}_\tau$ for each transition τ,

$$\hat{\mathbf{X}}^{(N)}(t) = \hat{\mathbf{X}}_0^{(N)} + \sum_{\tau\in\mathcal{T}} \frac{1}{N}\mathbf{v}_\tau \mathcal{N}_\tau\left(N\int_0^t g_\tau(\hat{\mathbf{X}}^{(N)}(s))ds\right). \tag{B.1}$$

Furthermore, recalling that $F(\hat{\mathbf{d}}) = \sum_{\tau\in\mathcal{T}}\mathbf{v}_\tau g_\tau(\hat{\mathbf{d}})$, we have that $\hat{\mathbf{x}}(t)$ in integral form is:

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}_0 + \int_0^t \sum_{\tau\in\mathcal{T}} \mathbf{v}_\tau g_\tau(\hat{\mathbf{x}}(s))ds. \tag{B.2}$$

In the following, we need the notion of centred Poisson process [55], defined by $\tilde{\mathcal{N}}(\lambda t) = \mathcal{N}(\lambda t) - \lambda t$, for which the following law of large numbers holds: $\sup_{t\le T}\frac{1}{N}\tilde{\mathcal{N}}(N\lambda t) \to 0$ almost surely, as $N\to\infty$.

Now, we define

$$\begin{aligned}\varepsilon^{(N)}(t) &= \hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{X}}^{(N)}(0) - \int_0^t \sum_{\tau\in\mathcal{T}} \mathbf{v}_\tau g_\tau(\hat{\mathbf{X}}^{(N)}(s))ds \\ &= \sum_{\tau\in\mathcal{T}} \frac{\mathbf{v}_\tau}{N}\tilde{\mathcal{N}}_\tau\left(N\int_0^t g_\tau(\hat{\mathbf{X}}^{(N)}(s))ds\right),\end{aligned}$$

so that, as $g_\tau(\hat{\mathbf{d}}) \le B_\tau$, it holds that

$$\|\varepsilon^{(N)}(t)\| \le \sum_{\tau\in\mathcal{T}} \frac{\|\mathbf{v}_\tau\|}{N}\tilde{\mathcal{N}}_\tau\left(NB_\tau t\right).$$

[20] Intuitively, the result of the theorem is a limit result, hence it depends only on what happens in a neighbourhood $B_\varepsilon(\hat{\mathbf{x}}([0,T])$ of the solution, where by $B_\varepsilon(\hat{\mathbf{d}})$ we indicate the ball of radius ε centred in $\hat{\mathbf{d}}$, while $B_\varepsilon(\hat{\mathbf{x}}([0,T]) = \bigcup_{t\in[0,T]} B_\varepsilon(\hat{\mathbf{x}}(t))$. Thus, by restricting our attention to a compact set $K\subset E$ containing $B_\varepsilon(\hat{\mathbf{x}}([0,T])\cap E$ for some ε, invoking the local Lipschitz property of functions g_τ, we can find the constants B_τ and L_τ. This assumption is not limiting, as we can always extend the functions g_τ on the whole of E so that they are globally bounded and Lipschitz continuous. Clearly, $\hat{\mathbf{x}}(t)$ will remain unchanged by this operation, as it depends only on the value of g_τ in K.

By the law of large numbers for centred Poisson processes and the finiteness of the sum over $\tau \in \mathcal{T}^{(N)}$, we conclude that $\sup_{t\leq T} \|\varepsilon^{(N)}(t)\| \to 0$. Therefore, we have that

$$\begin{aligned}\sup_{t\leq T} \|\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)\| \;\leq\; & \underbrace{\|\hat{\mathbf{X}}_0^{(N)} - \hat{\mathbf{x}}_0\| + \sup_{t\leq T} \|\varepsilon^{(N)}(t)\|}_{=\delta^{(N)}(T)\to 0 \;\; a.s.} \\ & + \int_0^T \sup_{t\leq s} \|F(\hat{\mathbf{X}}^{(N)}(t)) - F(\hat{\mathbf{x}}(t))\| ds.\end{aligned}$$

Calling $\beta^{(N)}(T) = \sup_{t\leq T} \|\hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{x}}(t)\|$ and applying Lipschitz condition to the last term,[21] we have that

$$\beta^{(N)}(T) \leq \delta^{(N)}(T) + L \int_0^T \beta^{(N)}(t) dt,$$

where $\delta^{(N)}(T) = \|\hat{\mathbf{X}}_0^{(N)} - \hat{\mathbf{x}}_0\| + \sup_{t\leq T} \|\varepsilon^{(N)}(t)\|$ as above. By applying Gronwall's inequality,[22] we finally obtain $\beta^{(N)}(T) \leq \delta^{(N)}(T) e^{LT}$, hence

$$\lim_{N\to\infty} \beta^{(N)}(T) = 0 \;\; \text{almost surely.}$$

■

Appendix C. Further details on the comparison between PCTMC and PDTMC deterministic approximations

In this appendix we will provide further details on the comparison between the deterministic approximation results for PCTMCs and PDTMCs. We will start by recalling the notion of *uniformisation* of a CTMC, which plays a central role in this discussion, and explore in more detail its implications for the example of Section 2.1.

[21] The Lipschitz constant L for F can be easily constructed from those of g_τ.

[22] Gronwall's inequality states that, for any real valued integrable function f on the interval $[0, T]$, if $f(t) \leq C + D \int_0^t f(s) ds$, then $f(T) \leq C e^{DT}$, see e.g. [55]

Uniformisation. Consider a $\mathcal{D}$-valued CTMC $\mathbf{X_C}(t)$ with infinitesimal generator matrix $\mathbf{Q}$. Assume that the exit rate of $\mathbf{X_C}(t)$ is bounded by a constant $\Lambda > 0$. The Λ-*uniformisation* [62] (or simply *uniformisation*) of $\mathbf{X_C}(t)$ is a pair $\mathsf{UFZ}_\Lambda(\mathbf{X_C}(t)) =_{\text{def}} (\mathbf{X_D}(k), \mathcal{N}(\Lambda \cdot t))$ where $\mathbf{X_D}(k)$ is a $\mathcal{D}$-valued DTMC with probability transition matrix $\mathbf{P} = \mathbf{I} + \mathbf{Q}/\Lambda$, and $\mathcal{N}(\Lambda t)$ is the companion Poisson process with rate Λ. $\mathbf{X_C}(t)$ is equivalent to the stochastic process $\mathbf{X_D}(\mathcal{N}(\Lambda t))$.

The uniformization procedure can easily be lifted from CTMCs to PCTMC models specified in the language defined in Section 3.1, as described in the sequel. Let $\mathcal{X}_C^{(N)} = (\mathbf{X_C}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}_C^{(N)}, \mathbf{x_0}^{(N)})$ be one such model, with transition rate functions $r_\tau^{(N)}(\mathbf{d})$ and exit rate function $R^{(N)}(\mathbf{d})$. Assuming $R^{(N)}(\mathbf{d})$ is bounded, fix $\Lambda_N \geq \sup_{\mathbf{d} \in \mathcal{D}} R^{(N)}(\mathbf{d})$, and define the PDTMC model as $\mathcal{X}_D^{(N)} = (\mathbf{X_D}^{(N)}, \mathcal{D}^{(N)}, \mathcal{T}_D^{(N)}, \mathbf{x_0}^{(N)})$ where

$$\mathcal{T}_D^{(N)} =_{\text{def}} \{(a, \mathbf{s}, \mathbf{t}, p_\tau^{(N)}) \mid (a, \mathbf{s}, \mathbf{t}, r_\tau^{(N)}) \in \mathcal{T}_C^{(N)}\}$$

with $p_\tau^{(N)}(\mathbf{d}) =_{\text{def}} \frac{r_\tau^{(N)}(\mathbf{d})}{\Lambda_N}$, for all $\mathbf{d} \in \mathcal{D}^{(N)}$. Furthermore, let $\mathcal{N}^{(N)}(t) = \mathcal{N}(\Lambda_N \cdot t)$ be a Poisson process with rate Λ_N. The Λ_N-uniformization of $\mathcal{X}_C^{(N)}$ is $\mathsf{UFZ}_{\Lambda_N}(\mathcal{X}_C^{(N)}) =_{\text{def}} (\mathcal{X}_D^{(N)}, \mathcal{N}^{(N)}(t))$. Of course, the uniformization procedure can be applied to normalised PCTMC models and to sequences of PCTMC models as well.

Consider again the network epidemic example of Section 2.1, for which we presented both a PCTMC model and a PDTMC one. It is easily seen that the exit rate function $R^{(N)}(\mathbf{d})$ is bounded by $N \cdot \sum_{\sigma \in \{e,i,a,r,p,q,s\}} \lambda_\sigma$, since $X_s + X_e + X_i + X_r = N$ implies $r_\sigma^{(N)}(\mathbf{X}) \leq N \cdot \lambda_\sigma$ for $\sigma \in \{e, i, a, r, p, q, s\}$. Letting $\Lambda =_{\text{def}} \sum_{\sigma \in \{e,i,a,r,p,q,s\}} \lambda_\sigma$ and $\Lambda_N =_{\text{def}} N \cdot \Lambda$, we get that the transition probabilities of the PDTMC model associated with $\mathcal{E}_C^{(N)}$ by Λ_N-uniformization are $p_\sigma^{(N)}(\mathbf{X}) =_{\text{def}} \frac{r_\sigma^{(N)}(\mathbf{X})}{\Lambda_N}$ for σ as above. For instance, for the rate of external infection, we have that $r_e^{(N)}(\mathbf{X}) = \lambda_e \cdot X_s$, so that $p_e^{(N)}(\mathbf{X}) =_{\text{def}} \frac{\lambda_e \cdot X_s}{N\Lambda}$. Under the assumption that $\alpha_e = \frac{\lambda_e}{\Lambda}$ we get that $p_e^{(N)}(\mathbf{X})$ is exactly the same as the external infection probability in the PDTMC model $\mathcal{E}_D^{(N)}$ we defined in Section 3.2. Therefore, the two network epidemic models $\mathcal{E}_D^{(N)}$ and $\mathcal{E}_C^{(N)}$ we have considered are strongly related, via the uniformization construction.

We now recall that the vector field for the limit ODE of the normalised PDTMC model $\hat{\mathcal{E}}_D^{(N)}$ is its drift $F_D^{(N)}$ and, since it does not depend on N, we use the abbreviation F_D. Similarly, we use F_C for the drift $F_C^{(N)}$ of the

normalised PCTMC model $\hat{\mathcal{E}}_C^{(N)}$. Again, under the assumption that $\alpha_\sigma = \frac{\lambda_\sigma}{\Lambda}$, we get that $F_D(\hat{\mathbf{X}}) = \frac{1}{\Lambda} \cdot F_C(\hat{\mathbf{X}})$.

We now present the relationship between a sequence of PCTMC models and the corresponding sequence of PDTMC ones and then the converse relationship between a sequence of PDTMC models and the corresponding sequence of PCTMC ones, followed by discussion of the convergence results.

From PCTMCs to PDTMCs. Let $\hat{\mathcal{X}}_C^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$ be a normalised PCTMC model and suppose the sequence $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$ with respect to an increasing system size γ_N admits deterministic approximation. Let us construct the sequence of PDTMC models $(\hat{\mathcal{X}}_D^{(N)})_{N \geq N_0}$, by uniformization, for constants $\Lambda_N \geq \sup_{\hat{\mathbf{d}} \in E} R^{(N)}(\hat{\mathbf{d}})$, such that $\Lambda_N = \Theta(\gamma_N)$. The previous hypothesis on Λ_N implies that $\lim_{N \to \infty} \Lambda_N \delta_N = \Lambda > 0$.

Furthermore let F_C and F_D be the vector fields according to equations (13) and (20) respectively and recall our assumption that $\epsilon_N = \delta_N$. We obtain the following:

$$\begin{aligned} F_D^{(N)}(\hat{\mathbf{d}}) &= \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{\mathbf{v}}_\tau^{(N)} \cdot \frac{\hat{r}_\tau^{(N)}(\hat{\mathbf{d}})}{\delta_N \cdot \Lambda_N} = \frac{1}{\delta_N \cdot \Lambda_N} \cdot \sum_{\tau \in \hat{\mathcal{T}}^{(N)}} \hat{\mathbf{v}}_\tau^{(N)} \cdot \hat{r}_\tau^{(N)}(\hat{\mathbf{d}}) \\ &= \frac{1}{\delta_N \cdot \Lambda_N} \cdot F_C^{(N)}(\hat{\mathbf{d}}), \end{aligned}$$

from which, by taking the limit $N \to \infty$, we get $F_D(\hat{\mathbf{d}}) = \frac{1}{\Lambda} F_C(\hat{\mathbf{d}})$.

Hence, we have shown that the deterministic limit of the associated sequence of PDTMC models satisfies the ODE $\frac{d\hat{\mathbf{x}}(t)}{dt} = \frac{1}{\Lambda} F_C(\hat{\mathbf{x}}(t))$, which is the same as the ODE for the sequence of PCTMC models, except for the rescaling factor $1/\Lambda$. This means that if we rescale time from t to $t' = t/\Lambda$, we obtain from the sequence of PDTMC models constructed by uniformization, the same set of ODE as the one derived from the sequence of PCTMC ones: $\frac{d\hat{\mathbf{x}}(t')}{dt'} = F_C(\hat{\mathbf{x}}(t'))$.

From PDTMCs to PCTMCs. Let $\hat{\mathcal{X}}_D^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{d}}_{\mathbf{0}}^{(N)})$ be a normalised PDTMC model and consider the sequence $(\hat{\mathcal{X}}_D^{(N)})_{N \geq N_0}$.

We can associate it with a sequence $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$ of normalised PCTMC models simply by interpreting probabilities, multiplied by the system size, as rates. Then we define the set of transitions $\hat{\mathcal{T}}^{d(N)}$. These are formed from

the transitions of $\hat{\mathcal{T}}^{(N)}$ by replacing the probability transition functions with rate functions defined as $\hat{r}_\tau^{(N)} = \gamma_N \cdot \hat{p}_\tau^{(N)}$. The associated PCTMC model is therefore $\hat{\mathcal{X}}_C^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{d^{(N)}}, \hat{\mathbf{d}}_\mathbf{0}^{(N)})$ whose mean dynamics is

$$F_C^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{d^{(N)}}} \delta_N \cdot \mathbf{v}_\tau^{(N)} \cdot \hat{r}_\tau^{(N)}(\hat{\mathbf{d}}) = \sum_{\tau \in \hat{\mathcal{T}}^{d^{(N)}}} \mathbf{v}_\tau^{(N)} \hat{p}_\tau^{(N)}(\hat{\mathbf{d}}) = F_D^{(N)}(\hat{\mathbf{d}}).$$

Thus the mean dynamics of the constructed PCTMC model is equal to the mean dynamics of the PDTMC model. As a consequence, $(\hat{\mathcal{X}}_D^{(N)})_{N \geq N_0}$ and $(\hat{\mathcal{X}}_C^{(N)})_{N \geq N_0}$ will have the same deterministic limit.

Convergence. If we inspect a (normalised) PCTMC model $\hat{\mathcal{X}}_C^{(N)}$ and its associated PDTMC model, $\hat{\mathcal{X}}_D^{(N)}$, or vice versa, we can observe that the jump process underlying the PCTMC model is in fact that of the PDTMC model. The only difference between the two is that the PCTMC has a variable delay between two consecutive events, exponentially distributed with rate $R^{(N)}(\hat{\mathbf{d}})$, while the time delay between two steps is constant for the PDTMC. Nevertheless, in the limit, both delays go to zero at the same speed, namely as δ_N, which is equal to the intensity ϵ_N.

We now give an intuitive argument showing that in this setting these two processes must behave in the same way in the limit (i.e. for large N), before stating a theorem which gives a precise statement of this property.

Consider a sequence of (normalised) PDTMC models $(\hat{\mathcal{X}}_D^{(N)})_{N \geq N_0}$ and let $(\hat{\mathbf{X}}_\mathbf{D}^{(N)}(k))_{N \geq N_0}$ be the corresponding sequence of PDTMCs; furthermore let ϵ_N be the intensity of the sequence, with $\lim_{N \to \infty} \epsilon_N = 0$. We construct two sequences of stochastic process on $t \in [0, \infty)$. One is $\hat{\mathbf{X}}_\mathbf{ds}^{(N)}(t) =_{\text{def}} \hat{\mathbf{X}}_\mathbf{D}^{(N)}(\lfloor \frac{t}{\epsilon_N} \rfloor)$, obtained by assuming that the length of each step in the DTMC is deterministically fixed and equal to ϵ_N. The other one is $\hat{\mathbf{X}}_\mathbf{rs}^{(N)}(t) =_{\text{def}} \hat{\mathbf{X}}_\mathbf{D}^{(N)}(\mathcal{N}(\frac{t}{\epsilon_N}))$, where $\mathcal{N}$ is a Poisson random variable, hence the number of steps at time t is random (and, consequently, $\hat{\mathbf{X}}_\mathbf{rs}^{(N)}(t)$ is a CTMC).

Now, $\mathcal{N}(\frac{t}{\epsilon_N}) / \lfloor \frac{t}{\epsilon_N} \rfloor = \left(\mathcal{N}(\frac{t}{\epsilon_N}) / \frac{t}{\epsilon_N} \right) \cdot \left(\frac{t}{\epsilon_N} / \lfloor \frac{t}{\epsilon_N} \rfloor \right)$, hence by the law of large numbers for Poisson processes, we get that almost surely

$$\lim_{N \to \infty} \mathcal{N}\left(\frac{t}{\epsilon_N}\right) / \left\lfloor \frac{t}{\epsilon_N} \right\rfloor = 1.$$

This intuitively shows that, at any time t, the two processes $\hat{\mathbf{X}}_\mathbf{ds}^{(N)}(t)$ and $\hat{\mathbf{X}}_\mathbf{rs}^{(N)}(t)$ will have the same distribution in the limit as N goes to ∞: the

same number of events will have fired. This intuition can be formalised in the following theorem, whose proof can be found, for instance, in [58] (cf. proof of Theorem 17.28 of [58]).

Theorem (6.1)**.** *Let* $\hat{\mathbf{X}}^{(N)}_{\mathbf{ds}}(t)$ *and* $\hat{\mathbf{X}}^{(N)}_{\mathbf{rs}}(t)$ *be defined as above. If either of the processes* $\hat{\mathbf{X}}^{(N)}_{\mathbf{ds}}(t)$ *and* $\hat{\mathbf{X}}^{(N)}_{\mathbf{rs}}(t)$ *converges to a limit process* $\hat{\mathbf{X}}(t)$ *weakly*[23]*, then so does the other*[24]*, and moreover* $\hat{\mathbf{X}}^{(N)}_{\mathbf{ds}}(t)$ *converges to* $\hat{\mathbf{X}}^{(N)}_{\mathbf{rs}}(t)$ *weakly.*

[23]A sequence of processes $\hat{\mathbf{X}}^{(N)}(t)$, with values in E, converges weakly (or in distribution) to a process $\hat{\mathbf{X}}(t)$ on E if and only if the sequence of measures of the trajectory space associated with $\hat{\mathbf{X}}^{(N)}(t)$ (i.e. the space $D = D(\mathbb{R}, E)$ of right-continuous functions with left limits from $\mathbb{R}$ to E) converges (in the weak sense) to the measure of the trajectory space associated with $\hat{\mathbf{X}}(t)$. This, in turn, holds if and only if for each bounded continuous functional $f : D \to \mathbb{R}$, it holds that $E^{(N)}[f] \to E[f]$, where $E^{(N)}$ (resp. E) is the expectation with respect to the measure on D defined by $\hat{\mathbf{X}}^{(N)}(t)$ (resp. $\hat{\mathbf{X}}(t)$). The interested reader is referred to [58].

[24]If the limit process $\hat{\mathbf{X}}(t)$ is deterministic, then convergence in distribution of $\hat{\mathbf{X}}^{(N)}_{\mathbf{ds}}(t)$ or $\hat{\mathbf{X}}^{(N)}_{\mathbf{rs}}(t)$ to $\hat{\mathbf{X}}(t)$ implies convergence in probability, uniformly in $[0, T]$, for any finite time horizon T, see [58] for further details.