The Mean Drift: Tailoring the Mean Field Theory of Markov Processes for Real-World Applications

Mahmoud Talebi* Jan Friso Groote Jean-Paul M.G. Linnartz
{m.talebi,j.f.groote,j.p.linnartz}@tue.nl

Eindhoven University of Technology
Den Dolech 2, 5612 AZ, Eindhoven, The Netherlands

Abstract
The statement of the mean field approximation theorem in the mean field theory of Markov processes particularly targets the behaviour of population processes with an unbounded number of agents. However, in most real-world engineering applications one faces the problem of analysing middle-sized systems in which the number of agents is bounded. In this paper we build on previous work in this area and introduce the mean drift. We present the concept of population processes and the conditions under which the approximation theorems apply, and then show how the mean drift is derived through a systematic application of the propagation of chaos. We then use the mean drift to construct a new set of ordinary differential equations which address the analysis of population processes with an arbitrary size.

Keywords. Markov chains, population processes, mean field approximation, propagation of chaos

1 Introduction
Population processes are stochastic models of systems which consist of a number of similar agents (or particles) [23]. When the impact of each agent on the behaviour of the system is similar to other agents, it is said that the population process is a mean field interaction model [2]. It is possible to apply a symmetric reduction on the state space of these types of processes and gain some efficiency in their analysis. Mean field approximation refers to the continuous, deterministic approximations of the behaviour of such processes in their first moment, when the number of agents grows very large. These approximations were first proposed for several concrete cases in various areas of study e.g., from as early as the 18th century in population biology, where models such as the predator-prey equations and the SIR equations are used to describe the balance of species in an ecosystem and the dynamics of epidemics respectively [3, 13]. Since then, general theorems have been proven which show the convergence of the behaviour of population processes to solutions of differential equations. The proofs follow roughly the same steps which generally rely on Grönwall’s lemma and martingale inequalities [12]. One of the first generalized approximation theorems was given by Kurtz [21]. The theory gives conditions which define a family of such models called density-dependent population processes, and finds their deterministic approximations by a theorem which is generally called the law of large numbers for standard Poisson processes [15].

The mean field theory of Markov processes is increasingly being applied in the fields of computer science and communication engineering. In the field of communication engineering and starting with Bianchi’s analysis of the IEEE 802.11 DCF protocol [4, 5], much research has focused on discussing the validity of the so-called decoupling assumption in this analysis [7, 33, 29, 11]. Several general frameworks have also been proposed which target the analysis of computer and communication systems [2, 25]. In the field of computer science the initial application of the approximations was intuitively motivated by methods such as fluid and diffusion approximations of queueing networks [19]. These have resulted in the development of methods and tools to automate the analysis of mean field models, with extensive progress in the context of the stochastic process algebra PEPA [18] [17] [9] among others [8]. However, still a large family of models are deemed unsuitable for this fluid approximation analysis, since they often lead to demonstrably inaccurate approximations [18] [22]. This calls for revisiting the fundamental roots of the theory. Such an approach has been taken in [1] where the authors use a set of extended diffusion approximations to derive precise approximations of stochastic Petri Net models.

*The authors would like to thank Erik de Vink, Mieke Massink, Tjalling Tjalkens and Ulyana Tikhonova for their constructive comments and helpful remarks.
We identify the current challenge as the problem of analysing middle-sized systems: systems which are so large that they suffer from state space explosion, but not large enough such that they can be accurately analysed by common approximation methods. In this paper we focus on the evaluation of these middle-sized systems, and the most important contribution is the introduction of Poisson mean of intensities (equation (17)), and the way they relate to the approximation theorem based on the idea of propagation of chaos.

Through equation (17) we express the idea that occupancy measures can be seen as Poisson arrival rates. Based on this observation we propose employing the concept of the Poisson mean of the drift to build the set of ordinary differential equations in (19) when dealing with bounded systems. To provide further proof for the consistency of our observation with respect to already established results in the mean field theory, we show that the drift and the mean drift are equivalent in the limit (Theorem 7).

In this section we give essential definitions and theorems which will be needed in later sections. In the definitions related to probability theory we rely mostly on [20]. The definitions of independence, conditional expectations and the law of large numbers are from [6]. Several properties of Poisson point processes and martingales are related to probability theory we rely mostly on [20]. The definitions of independence, conditional expectations and the law of large numbers are from [6]. Several properties of Poisson point processes and martingales are related to probability theory we rely mostly on [20].

In the current text, we use only the most essential and trivial concepts to build the mean field approximations. Several decisions were made to achieve this goal: the choice of discrete-time Markov processes to specify the agent mean drift are equivalent in the limit (Theorem 7).

In the current text, we use only the most essential and trivial concepts to build the mean field approximations. Several decisions were made to achieve this goal: the choice of discrete-time Markov processes to specify the agent mean drift are equivalent in the limit (Theorem 7).

The rest of the text is organized as follows. In Section 2 we give a brief introduction to necessary concepts in real analysis and probability theory which are used in later sections. Section 3 describes the family of mean field interaction models, and the derivation of their deterministic approximations. We then state Theorem 5 and Corollary 1, which provide the basis for stating the main results of this work. Finally, in Section 4 we present the idea of mean drift as our main result. The detailed proofs follow in Section 5.

2 Preliminaries

In this section we give essential definitions and theorems which will be needed in later sections. In the definitions related to probability theory we rely mostly on [20]. The definitions of independence, conditional expectations and the law of large numbers are from [6]. Several properties of Poisson point processes and martingales are related to probability theory we rely mostly on [20]. For the definitions and theorems related to real analysis (including the Lebesgue integral) we refer mainly to [24].

In what follows, we will use the following notations in the text. Let $T$ be a totally ordered set. We use the notation $\{X_t : t \in T\}$ to show a sequence of objects indexed by set $T$. The notation $\mathbb{R}_{\geq 0} = [0, \infty)$ is used to refer to the set of non-negative real numbers. In the same manner, $\mathbb{Q}_{\geq 0}$ is the set of non-negative rational numbers.

To make a clear distinction between superscripts and exponentiation, we write $X^N$ to denote “$X$ raised to the power $N$” and $X^{(N)}$ to denote “$X$ annotated by $N$”.

2.1 Measures and probability

Let $X$ be a set, and $d : X \times X \to \mathbb{R}$ be a function which for all $x, y \in X$ satisfies the following properties:

- $d(x, y) \geq 0$, where $d(x, y) = 0$ if and only if $x = y$,
- $d(x, y) = d(y, x)$ (symmetry),
- For all $z \in X$, $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality),

Then $d$ is called a distance function or a metric, and the pair $(X, d)$ is a metric space.

A $\sigma$-algebra $\Sigma_X$ on a set $X$ is a set of subsets of $X$ which contains the empty set $\emptyset$ and is closed under the complement, countable union and countable intersection of subsets. The pair $(X, \Sigma_X)$ is called a measurable space.

Let $X$ and $Y$ be sets. Consider the function $f : X \to Y$, $f^{-1} : 2^Y \to 2^X$ is a set mapping for which for $B \in 2^Y$:

$$f^{-1}(B) \equiv \{ x \in X \mid f(x) \in B \}.$$  

$f^{-1}(B)$ is often called the source of $B$.

**Definition 1** (Measurable Function). Let $(X, \Sigma_X)$ and $(Y, \Sigma_Y)$ be measurable spaces, and let $f : X \to Y$ be a function. For all $B \in \Sigma_Y$, and the set mapping $f^{-1}$, if

$$f^{-1}(B) \in \Sigma_X,$$

$f$ is called a $\Sigma_X/\Sigma_Y$-measurable function.
When the exact shape of the $\sigma$-algebras are not of concern or already clear in a context, one may also refer to these functions as $\Sigma_X$-measurable or just measurable.

Let $(X, \Sigma_X)$ and $(Y, \Sigma_Y)$ be measurable spaces. Let $f : X \to Y$ be a function. Then the induced $\sigma$-algebra of $f$, denoted by $\sigma(f)$ is the smallest $\sigma$-algebra $\mathcal{A} \subseteq \Sigma_X$ which makes $f$ measurable.

Let $(X, \Sigma_X)$ be a measurable space. A measure on $(X, \Sigma_X)$ is defined as a function $\mu : \Sigma_X \to \mathbb{R}_{\geq 0}$ which satisfies the following two properties:

- $\mu(\emptyset) = 0$,
- For all countable collections of pairwise disjoint sets $\{E_i\}_{i=1}^{\infty}$ in $\Sigma_X$:
  \[ \mu \left( \bigcup_{k=1}^{\infty} E_k \right) = \sum_{k=1}^{\infty} \mu(E_k). \]

Based on the definition of measures, we define Lebesgue integrals, which are indispensable tools in modern probability theory. Let $(X, \Sigma_X)$ be a measurable space and let $\mu : \Sigma_X \to \mathbb{R}_{\geq 0}$ be a measure. Let $A \in \Sigma_X$, for $x \in X$ the indicator function $\mathbb{I}_A : X \to \mathbb{R}_{\geq 0}$ is defined as:

\[ \mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}. \]

For the boolean domain, we define $\mathbb{I} : \mathbb{B} \to \mathbb{R}_{\geq 0}$, which for $b \in \mathbb{B}$:

\[ \mathbb{I}(b) = \begin{cases} 1 & \text{if } b \text{ true} \\ 0 & \text{if } b \text{ false}. \end{cases} \]

Given a measure $\mu$ on the measurable space $(X, \Sigma_X)$, the Lebesgue integral of $\mathbb{I}_A$ is denoted by $\int_X \mathbb{I}_A \, d\mu$ which is:

\[ \int_X \mathbb{I}_A \, d\mu = \mu(A). \]

More generally, for natural number $n$ let $A_1, \ldots, A_m \in \Sigma_X$ be sets, and let $c_1, \ldots, c_m \in \mathbb{R}_{\geq 0}$ be real vectors. A function $S : X \to \mathbb{R}_{\geq 0}^n$ where $S = \sum_k c_k \mathbb{I}_{A_k}$ is called a simple function. The Lebesgue integral of $S$ is:

\[ \int_X S \, d\mu = \sum_k c_k \mu(A_k). \]

For an arbitrary function $f : X \to \mathbb{R}_{\geq 0}^n$ define:

\[ \int_X f \, d\mu \overset{\text{def}}{=} \sup \left\{ \int_X S \, d\mu : 0 \leq S \leq f \text{ pointwise, } S \text{ simple} \right\} \]

and

\[ \int_X f \, d\mu \overset{\text{def}}{=} \inf \left\{ \int_X S \, d\mu : S \geq f \text{ pointwise, } S \text{ simple} \right\}. \]

The Lebesgue integral $\int_X f \, d\mu$ exists ($f$ is Lebesgue integrable) if $\int_X f \, d\mu = \int_X f \, d\mu < \infty$, in which case:

\[ \int_X f \, d\mu \overset{\text{def}}{=} \int_X f \, d\mu. \]

We now define the Lebesgue integral of signed functions. Let $f : X \to \mathbb{R}^n$. There exist functions $f_1, \ldots, f_n$ such that for $x \in X$:

\[ f(x) = (f_1(x), \ldots, f_n(x)). \]

For $1 \leq i \leq n$ define the sequence of functions $f^+_i$ and $f^-_i$ which satisfy:

\[ f^+_i(x) = \begin{cases} f_i(x) & \text{if } f_i(x) > 0, \\ 0 & \text{otherwise} \end{cases} \]

and:

\[ f^-_i(x) = \begin{cases} -f_i(x) & \text{if } f_i(x) < 0, \\ 0 & \text{otherwise}. \end{cases} \]
Accordingly, define the functions \( f^+ : X \to \mathbb{R}_{\geq 0} \) and \( f^- : X \to \mathbb{R}_{\geq 0} \) as:

\[
f^+(x) \overset{def}{=} (f_1^+(x), \ldots, f_n^+(x)) \quad \text{and} \quad f^-(x) \overset{def}{=} (f_1^-(x), \ldots, f_n^-(x)).
\]

The Lebesgue integral of a signed function \( f \) is:

\[
\int_X f d\mu \overset{def}{=} \int_X f^+ d\mu - \int_X f^- d\mu,
\]

which exists only if each of the integrals on the right hand side of this equation exist.

We have now introduced enough basic concepts to start talking about probabilities. For the measurable space \((X, \Sigma)\), a measure \( P : \Sigma \to \mathbb{R}_{\geq 0} \) is called a probability measure if it satisfies the property: \( P(X) = 1 \).

**Definition 2** (Probability Space). A triple \((\Omega, \mathcal{F}, P)\) is called a probability space in which:

- \( \Omega \) is a sample space, often containing all the possible outcomes of some experiment.
- \( \mathcal{F} \subseteq 2^\Omega \) is a set of events, which forms a \( \sigma \)-algebra over the set \( \Omega \).
- \( P \) is a probability measure defined over the measurable space \((\Omega, \mathcal{F})\).

In a probability space \((\Omega, \mathcal{F}, P)\), each element \( E \in \mathcal{F} \) is called an event. An event \( E \) occurs almost surely if \( P(E) = 1 \).

### 2.2 Random variables and processes

**Definition 3** (Random Elements and Random Variables). Let \((\Omega, \mathcal{F}, P)\) be a probability space, and let \((S, \Sigma)\) be a measurable space. A measurable function \( X : \Omega \to S \) is called a random element in \( S \). When \( S = \mathbb{R} \), \( X \) is called a random variable.

If \( S = \mathbb{R}^n \) for some \( n > 0 \), \( X \) is called a random vector. The concept of a random vector is in a way a generalization of the concept of a random variable. In the discussion that follows we often use the term random variable to refer to both.

A random element \( X \) from \((\Omega, \mathcal{F})\) to \((S, \Sigma)\) defines (induces) a new measure \( P \circ X^{-1} \) on \((S, \Sigma)\), which for \( B \in \Sigma \) satisfies:

\[
(P \circ X^{-1})B = P(\{\omega \in \Omega : X(\omega) \in B\}).
\]

The function \( (P \circ X^{-1}) \) is again a probability measure on space \((S, \Sigma)\), called the probability distribution (or law) of \( X \).

Here it is appropriate to briefly introduce the concept of constant random elements. Let \((S, \Sigma)\) be a measurable space and for some \( x \in S \) let \( \delta_x \) be a probability measure on \( S \) which for \( A \in \Sigma \) is defined as:

\[
\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}
\]

Then the measure \( \delta_x \) is called the Dirac measure centred on \( x \). Let \( X : \Omega \to S \) be a random element. If for some \( x \in S \), \( \delta_x \) is the probability distribution of \( X \) then \( X \) is called a constant (or deterministic) random element.

Let \( X : \Omega \to S \) and \( Y : \Omega \to S \) be random elements. Then \( X \) and \( Y \) are equal in distribution (denoted by \( X \overset{d}{=} Y \)) if and only if for all \( B \in \Sigma \):

\[
(P \circ X^{-1})B = (P \circ Y^{-1})B.
\]

The expected value or expectation \( \mathbb{E}[X] \) of a random variable \( X : \Omega \to \mathbb{R}^n \) is defined as:

\[
\mathbb{E}[X] = \int_X x dP = \int_{\mathbb{R}^n} x d(P \circ X^{-1}),
\]

if the Lebesgue integral exists.

**Remark.** The latter Lebesgue integral sums over the space \( \mathbb{R}^n \) according to the measure defined by the probability distribution of \( X \). In literature, the integral is almost always written as:

\[
\int_{\mathbb{R}^n} x (P \circ X^{-1})(dx).
\]

In the rest of the text, we will follow this convention.
For a function \( f : \mathbb{R}^n \to \mathbb{R}^m \) the expectation \( \mathbb{E}[f(X)] \) is defined as:
\[
\mathbb{E}[f(X)] = \int_{\mathbb{R}^n} x (P \circ (f \circ X)^{-1}) (dx).
\]

**Remark.** Let \( X : \Omega \to \mathbb{R}^n \) be a random variable, and let \( A \subseteq \mathbb{R}^n \). We define the special operator \( \mathbb{P} \) as follows:
\[
\mathbb{P}[X \in A] = \mathbb{E}[\mathbb{1}_A(X)].
\]

Let \( X : \Omega \to \mathbb{R}^n \) be a random variable. The *conditional expectation* of \( X \) with respect to a \( \sigma \)-algebra \( \mathcal{D} \subseteq \mathcal{F} \), written as \( \mathbb{E}[X | \mathcal{D}] \), is a \( \mathcal{D} \)-measurable function which satisfies the following condition:
\[
\int_{\mathcal{D}} \mathbb{E}[X | \mathcal{D}] d\mathbb{P} = \int_{\mathcal{D}} X d\mathbb{P} \quad \text{for all } D \in \mathcal{D}.
\]

The random variable \( \mathbb{E}[X | \mathcal{D}] \) which satisfies these equations is unique [6].

Consider a random element \( Y : \Omega \to S \). Given the induced \( \sigma \)-algebra of \( Y \), \( \sigma(Y) \subseteq \mathcal{F} \), the conditional expectation of \( X \) given \( Y \) is a \( \sigma(Y) \)-measurable function \( \mathbb{E}[X | Y] \) satisfying:
\[
\mathbb{E}[X | Y] = \mathbb{E}[X | \sigma(Y)].
\]

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. The distinct events \( E_1, \ldots, E_n \in \mathcal{F} \) are *mutually independent* if and only if:
\[
P\left( \bigcap_{i \leq n} E_i \right) = \prod_{i \leq n} P(E_i).
\]

For a sequence of \( \sigma \)-algebras \( \mathcal{C}_i \subseteq \mathcal{F} \), where \( i \in \{1, \ldots, n\} \), the \( \sigma \)-algebras are mutually independent if all sequences \( A_1, \ldots, A_n \) where \( A_i \in \mathcal{C}_i \) are mutually independent.

Let \( i \in \{1, \ldots, n\} \) let \( X_i : \Omega \to S \) be random elements. The random elements \( X_i \) are mutually independent if the sequence of \( \sigma \)-algebras \( \sigma(X_i) \) are mutually independent.

We now possess all the tools to present the following important result regarding the summations of independent and identically distributed (i.i.d.) random variables.

**Theorem 1** (The Weak Law of Large Numbers). For \( 1 \leq i \leq n \) let \( X_i \) be mutually independent, and identically distributed random variables (i.i.d.’s), with \( \mathbb{E}[X_i] = \eta \). Let \( S_n = X_1 + \ldots + X_n \). Then for any real constant \( \varepsilon > 0 \):
\[
\lim_{n \to \infty} \mathbb{P}\left( \left| \frac{S_n}{n} - \eta \right| \geq \varepsilon \right) = 0.
\]

Next, we define stochastic processes.

**Definition 4** (Stochastic Processes). Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. Let \((S, \Sigma_S)\) be a measurable space and let \( T \) be a totally ordered index set. Let \( S^T = \{ f : T \to S \} \) be the class of functions from \( T \) to \( S \), and let \( U \subseteq S^T \), where \((U, \Sigma_U)\) is a measurable space. A random element \( X : \Omega \to U \) is called a stochastic or random process in \( S \).

The elements of \( U \) are often called paths or sample paths. However, stochastic processes are more commonly defined as follows.

For \( t \in T \) consider the set of all evaluation mappings (functionals) \( \pi_t : S^T \to S \) where \( \pi_t(f) = f(t) \), and define \( X(t) = \pi_t \circ X \). Clearly, for each \( t \), \( X(t) : \Omega \to S \). Based on [20] (Lemma 1.4 and Lemma 2.1), since \( X \) is measurable and \( \pi_t \) are functions, for each \( t \), \( X(t) \) is also measurable and is a random element in \( S \).

Therefore we may also specify an \( S \)-valued stochastic process \( X \) by a sequence of random elements \( \{X(t) : t \in T\} \). We write \( X(t, \omega) \) when we talk about the value of \( X(t) \) for a specific outcome \( \omega \in \Omega \). The index set \( T \) usually denotes time, and is either discrete \((T = \mathbb{N})\) or continuous \((T = [0, \infty))\).

In practice it is important to switch between the two notions of a stochastic process and employ both intuitions. For Markov processes, they are often specified using the notion of a sequence of random variable, while, when discussing their behaviour they are viewed as function-valued random elements.

### 2.3 Markov processes

Let \( T \) be a totally ordered set, called the time domain. Consider the probability space \((\Omega, \mathcal{F}, \mathbb{P})\). For a stochastic process we formalize the idea of information known at time \( t \in T \) as follows.
**Definition 5** (Filtration). Let \( \{ \mathcal{F}_t \} \) be a sequence of \( \sigma \)-algebras, where \( \mathcal{F}_t \subseteq \mathcal{F} \). If the sequence is increasing i.e., for all \( s, t \in T, s \leq t \) implies \( \mathcal{F}_s \subseteq \mathcal{F}_t \), then \( \{ \mathcal{F}_t \} \) is called a filtration.

Let \( S \) be a set and let \( \{ \mathcal{F}_t \} \) be a filtration. An \( S \)-valued stochastic process \( \{ X(t) : t \in T \} \) is \( \{ \mathcal{F}_t \} \)-adapted if and only if for each \( t \in T, \mathcal{F}_t \) is the smallest \( \sigma \)-algebra for which \( X(t) \) is \( \mathcal{F}_t / \Sigma_s \)-measurable i.e., \( \mathcal{F}_t = \sigma(X(t)) \). For \( s \leq t \) the relation \( \mathcal{F}_s \subseteq \mathcal{F}_t \) implies that for all \( s \leq t, X(s) \) is also \( \mathcal{F}_t / \Sigma_s \)-measurable.

**Definition 6** (Markov Processes). Let \( S \) be a set. An \( S \)-valued \( \{ \mathcal{F}_t \} \)-adapted stochastic process \( \{ X(t) : t \in T \} \) is a Markov process if for \( s, t \geq 0 \) and any function \( f : S \to \mathbb{R}^n \):

\[
\mathbb{E}[f(X(t+s)) \mid \mathcal{F}_t] = \mathbb{E}[f(X(t+s)) \mid X(t)].
\]

The above property is called the Markov property. If \( X(t) \) also has the property that for all \( x \in S \):

\[
\mathbb{E}[f(X(t+s)) \mid X(t) = x] = \mathbb{E}[f(X(s)) \mid X(0) = x]
\]

then \( X(t) \) is called a time-homogeneous Markov process.

In what follows we mention some ideas from the theory of Feller semigroups [20], which will be used in our proofs. Let \( v \in S \) be an initial state, \( f : S \to \mathbb{R}^n \) be a function, and \( \{ \mathcal{T}_t : t \in \mathbb{R}_{\geq 0} \} \) be a sequence of unary linear operators. A time-homogeneous Markov process \( X(t) \) corresponds to the sequence \( \{ \mathcal{T}_t \}_{t \geq 0} \) if for all \( t \geq 0 \):

\[
\mathcal{T}_t f(v) = \mathbb{E}[f(X(t)) \mid X(0) = v].
\]

Based on the Markov and time-homogeneity properties, for \( s, t \geq 0 \) the operators \( \{ \mathcal{T}_t \}_{t \geq 0} \) satisfy:

\[
\mathcal{T}_{s+t} f(v) = \mathcal{T}_s \mathcal{T}_t f(v). \tag{1}
\]

As such, \( \{ \mathcal{T}_t \}_{t \geq 0} \) is called an operator semi-group. It follows that the initial state \( v \) and the linear operators \( \{ \mathcal{T}_t \}_{t \geq 0} \) partially characterize the evolution of the stochastic process \( \{ X(t) : t \in T \} \) throughout time.

Let \( S \) be a set and let \( T = \mathbb{N} \). Let \( \{ X(t) : t \in T \} \) be a time-homogeneous Markov process in \( S \), \( X(t) \) is also called a discrete-time Markov chain (DTMC). As a special case of \( \{ \mathcal{T}_t \} \), consider the map \( P_t : S \times S \to [0, 1] \), which for \( i, j \in S \) is defined as:

\[
P_t(i, j) = \mathbb{P}[X(t) = j \mid X(0) = i].
\]

In which we have taken \( f(X(t)) = 1_{ij}(X(t)) \). The time-homogeneity implies that for \( t \geq 0 \) and every \( i, j \in S \), the map \( P_t \) satisfies:

\[
P_t = (P_t)^t,
\]

which is called the Chapman-Kolmogorov equation, and the map \( P = P_1 \) is called the transition map or the transition matrix of \( X(t) \).

Let time \( T = \mathbb{R}_{\geq 0} \) be continuous. Let \( \{ X(t) : t \in T \} \) be a time-homogeneous Markov process. For the corresponding linear operator \( \{ \mathcal{T}_t \}_{t \geq 0} \) the infinitesimal generator \( A \) is a mapping which maps a function \( f : S \to \mathbb{R} \) in its domain to \( g : S \to \mathbb{R} \) (hence \( g \) is unique) and satisfies:

\[
Af = \frac{\mathcal{T}_t f - f}{t}. \tag{2}
\]

For such a pair \( (f, g) \in A \) the following equation always holds [23]:

\[
\mathcal{T}_t f - f = \int_0^t \mathcal{T}_s g \, ds. \tag{3}
\]

For a full discussion on the above equation, also see Dynkin’s formula [20].

### 2.4 Martingales and stopping times

**Definition 7** (Martingales). Let \( \{ X(t) : t \in T \} \) be an \( \{ \mathcal{F}_t \} \)-adapted stochastic process. Then \( X(t) \) is a martingale if for \( s \leq t \):

\[
\mathbb{E}[X(t) \mid \mathcal{F}_s] = X(s),
\]

almost surely.
For $T$ discrete, $X(t)$ is a discrete-time martingale if for any $t \in T$ it satisfies:

$$\mathbb{E}[|X(t)|] < \infty, \quad \mathbb{E}[X(t + 1)|X(0), \ldots, X(t)] = X(t)$$

Or equivalently:

$$\mathbb{E}[X(t + 1) - X(t)|X(0), \ldots, X(t)] = 0$$

In a similar manner, a submartingale is a process $X(t)$ which for $s > 0$ satisfies:

$$\mathbb{E}[X(t + s) | \mathcal{F}_t] \geq X(t).$$

with the implication that every martingale is also a submartingale.

Martingales are important due to the fact that despite their generality they satisfy a number of interesting properties. In this text we use one such result called the norm inequality from Doob [20], which we state in the following form.

**Lemma 1** (Doob’s Inequality). Let $\{X(t) : t \in T\}$ be a submartingale taking non-negative values, either in discrete or continuous time. Assume that the process is right continuous with left limits everywhere. Then, for any constant $c > 0$,

$$\mathbb{P}\left\{\sup_{0 \leq s \leq T} X(t) \geq c\right\} \leq \frac{\mathbb{E}[X(T)]}{c}.$$  

Based on doob’s inequality, for integer $\alpha \geq 1$ the following inequality is derived in [15] (Proposition 2.2.16):

$$\mathbb{E}\left[\sup_{0 \leq s \leq T} X(t)^\alpha\right] \leq \left(\frac{\alpha}{\alpha - 1}\right)^\alpha \mathbb{E}[X(T)^\alpha].$$  

**Definition 8** (Stopping Times). Let $\{X(t) : t \in T\}$ be a $\mathcal{F}_t$-adapted process. A random variable $\tau : \Omega \to T$ is an $\{\mathcal{F}_t\}$-stopping time if for all $t \in T$, the event $\{\tau \leq t\}$ is an element of $\mathcal{F}_t$.

In simple terms, the condition means that for a stopping time $\tau$ it must be always possible to determine whether it has occurred by a time $t \in T$ or not, only by referring to the events in $\mathcal{F}_t$. $X(\tau)$ is the state of process $X(t)$ at the random time $\tau$.

Let $X(t)$ be an $S$-valued stochastic process and let $A \subseteq S$. An example of a stopping times is:

- $\tau_A = \min\{t \in T : X(t) \in A\}$ is called the hitting time of $A$: the first time in which the event $A$ occurs.

While the following is not a stopping time:

- $\tau_A = \max\{t \in T : X(t) \in A\}$: the last time in which the event $A$ occurs.

### 2.5 Poisson point processes

Before introducing Poisson processes, we draw a link between Bernoulli trials and the so-called Poisson distributions, using the following approximation theorem.

**Theorem 2** (Poisson Limit Theorem). Let $n \in \mathbb{N}$, and for $1 \leq i \leq n$ let $Z_i$ be a sequence of i.i.d. random variables, where each $Z_i$ takes value $1$ with probability $p$ and $0$ with probability $1 - p$. If:

$$n \to \infty, \quad p \to 0, \quad \text{while } np \to \lambda \text{ where } 0 \leq \lambda \ll \infty,$$

then

$$\mathbb{P}\left\{\sum_{i=1}^{n} Z_i = k\right\} \to e^{-\lambda} \frac{\lambda^k}{k!}.$$

We now define counting processes. Consider a set of points (representing events or arrivals) randomly located on the real line $\mathbb{R}_{\geq 0}$, which represents time. We define the process that counts the number of such points in the interval $[0, t]$, $t \in \mathbb{R}_{\geq 0}$. In the following assume $T = \mathbb{R}_{\geq 0}$.

**Definition 9** (Counting Processes). A $\mathbb{N}$-valued stochastic process $\{N(t) : t \in T\}$ is a counting process if:

1. $\mathbb{P}[N(0) = 0] = 1$

2. For $0 \leq s \leq t$, $N(t) - N(s)$ is the number of points in the interval $(s, t]$.
Any counting process is non-negative, non-decreasing and right-continuous.

**Definition 10** (Independent Increments). Let \( \{N(t) : t \in T\} \) be a counting process, and let \( t_0, t_1, \ldots, t_n \in T \) be increasing times in which \( t_0 = 0 \). We say \( N(t) \) has independent increments if and only if for all \( i \in \{1, \ldots, n\} \) the random variables:

\[
N(t_i) - N(t_{i-1})
\]

are mutually independent. In addition, \( N(t) \) is **stationary** if for equally distanced \( t_0, \ldots, t_n \) the increments \( N(t_i) - N(t_{i-1}) \) are equal in distribution.

**Definition 11** (Time-homogeneous Poisson Processes). Let \( \lambda > 0 \). A stationary counting process \( \{N(t) : t \in T\} \) is called a time-homogeneous Poisson process, or simply a Poisson process with rate or intensity \( \lambda \) if it satisfies the following additional properties:

- \( N(t) \) has independent increments.
- Almost surely, in an infinitesimal time interval \( dt \) at most one point occurs with probability \( \lambda dt \).

For a Poisson process, the number of observations over the interval \( (s, t] \) is discrete and is distributed according to a Poisson distribution:

\[
P[N(t) - N(s) = k] = \frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)}.
\]

Intuitively speaking, since a step in continuous time satisfies \( dt \to 0 \), and in intervals of size \( dt \) almost surely only 0 or 1 points may occur, one can think of the process in any interval \( (s, t] \) as an infinite sequence of Bernoulli trials and then apply the Poisson limit theorem to derive the above probability.

The Poisson process \( \{Y(t) : t \in T\} \) with \( \lambda = 1 \) is called the **unit Poisson process**. Let \( N(t) \) be a Poisson process with \( \Lambda(t) = \mathbb{E}[N(t)] \) and \( \Lambda(0) = 0 \). The following relation holds between \( N(t) \) and the unit Poisson process:

\[
N(t) = \mathcal{Y}(\Lambda(t)).
\]

For a unit Poisson process we have \( \mathbb{E}[\mathcal{Y}(t)] = t \). A **compensated unit Poisson process** \( \mathcal{Y}(t) \) is defined as:

\[
\mathcal{Y}(t) = \mathcal{Y}(t) - t,
\]

For which for all \( t \in T \):

\[
\mathbb{E}[\mathcal{Y}(t)] = 0.
\]

Based on equation (4) and due to the independent increment property, \( \mathcal{Y}(t) \) is a martingale, since for any \( s < t \) with \( s, t \in T \):

\[
\mathbb{E} \left[ \mathcal{Y}(t) - \mathcal{Y}(s) \mid \mathcal{F}_s \right] = \mathbb{E}[\mathcal{Y}(t) - \mathcal{Y}(s) \mid \mathcal{F}_s] - (t-s)
\]

\[
= \mathbb{E}[\mathcal{Y}(t-s) \mid \mathcal{F}_s] - (t-s) = 0.
\]

### 2.6 Continuity and convergence

In this section we quickly review a number of useful results in the field of real analysis, as the proofs we give later rely heavily on them.

**Definition 12** (Right Continuous Functions with Left Limits). Let \( f : \mathbb{R} \to \mathbb{R}^n \) be a function. Then \( f \) is a right continuous function with left limits (càdlàg) if and only if for every \( x \in \mathbb{R} \):

- \( f(x-) \triangleq \lim_{x^+ \to x^-} f(a) \) exists, and
- \( f(x+) \triangleq \lim_{x^- \to x^+} f(a) \) exists and \( f(x+) = f(x) \).

**Definition 13** (Lipschitz Continuity). A function \( f : \mathbb{R}^n \to \mathbb{R}^n \) is **globally** Lipschitz continuous on \( \mathbb{R}^n \) if and only if:

\[
\exists L \in \mathbb{R} : \forall x_1, x_2 \in \mathbb{R}^n. \left| f(x_2) - f(x_1) \right| \leq L \left| x_2 - x_1 \right|
\]

For a Lipschitz continuous function \( f \), we call the constant \( L \) found above the **Lipschitz constant**. If a function \( f \) has bounded first derivatives everywhere on \( \mathbb{R}^n \), it is guaranteed to be Lipschitz continuous.

\(^1\)continue à droite, limite à gauche
Theorem 3 (Picard-Lindelöf Theorem). Let $f$ be a Lipschitz continuous function in $\mathbb{R}^n$, and consider the following ordinary differential equation:

$$\frac{d}{dt} x(t) = f(x(t)), \quad x(t_0) = x_0,$$

then for some $\epsilon > 0$, there exists a unique solution $x(t)$ to the initial value problem on the interval $[t_0 - \epsilon, t_0 + \epsilon]$.

Lemma 2 (Grönwall’s Inequality). Let $f$ be a function that is bounded and integrable on the interval $[0, T]$, if

$$f(T) \leq C + D \int_0^T f(t) dt$$

then

$$f(T) \leq Ce^{DT}$$

Definition 14 (Uniform Convergence). Let $\{f_n\}$ be a sequence of functions on set $S$. We say the sequence converges uniformly to function $f$ if:

$$\lim_{n \to \infty} \sup_x |f_n(x) - f(x)| = 0,$$

that is, the speed of convergence of $f_n(x)$ to $f(x)$ does not depend on $x$.

If the sequence of functions $\{f_n\}$ are continuous and converge uniformly to $f$, then the limiting function $f$ is continuous as well.

Theorem 4 (Lebesgue Dominated Convergence Theorem). Let $\{f_n\}$ be a sequence of measurable functions on set $S$, which converge to measurable function $f$. Let $g$ be an integrable function such that for all $n$ and for all $x \in S$: $|f_n(x)| \leq |g(x)|$. Then $f$ is integrable and:

$$\lim_{n \to \infty} \int_S |f_n - f| = 0.$$

Finally, we give several notions of convergence for random variable, and quickly overview their relations.

Definition 15 (Convergence in $L^p$, in Probability and in Distribution). Let $\{X_n\}$ be a sequence of random variables, and $X$ be a random variable. Then for $p \in \mathbb{N}_{\geq 0}$, $\{X_n\}$ converges to $X$ in $L^p$ or $p$-th moment if:

$$\lim_{n \to \infty} E[|X_n - X|^p] = 0,$$

we say $\{X_n\}$ converges in probability to $X$ if for any $\epsilon > 0$:

$$\lim_{n \to \infty} P(|X_n - X|) = 0,$$

and we say $\{X_n\}$ converges in distribution to $X$ if for all $x \subset \mathbb{R}^n$:

$$\lim_{n \to \infty} (P \circ X_n)^{-1}x = (P \circ X)^{-1}x.$$

Convergence in $L^2$ is also called convergence in mean-square. If possible, proving convergence in mean-square is very useful, since it implies convergence in probability, which in turn implies convergence in distribution. In this sense convergence in distribution is also often referred to as weak convergence.

3 Mean field approximation

In this section, we present the stochastic model of a system and its mean field approximations. For the most part, our notation agrees with $\mathbb{2}$. A list of objects appearing in the mathematical discussions that follow are given in Table II.

3.1 Agent processes and the clock independence assumption

Let the set $T = \mathbb{N}$ be discrete and let parameter $N \in \mathbb{N}_{>0}$ be the system size. The elements of $T$ are called time-slots. Let $S = \{1, \ldots, L\}$ be a finite set of states. For $i \in \{1, \ldots, N\}$, let $\{X_i^{(N)}(t) : t \in T\}$ be $S$-valued discrete-time time-homogeneous Markov chains (DTMCs). Each stochastic process $X_i^{(N)}(t)$ describes the behaviour of agent $i$ in the system with $N$ agents.

Take each process $X_i^{(N)}(t)$ to be described by a transition map $K_i : S^N \times S \to [0, 1]$. In each time-slot (indexed by members of $T$), the process chooses the next state $s \in S$ with probability $K_i(\vec{v}, s)$, where the vector of states $\vec{v} \in S^N$ is the state of the entire system (including agent $i$’s current state).

There are generally two ways in which we can relate the time-slots across the processes in such a system:
Table 1: Table of objects and their short description.

| Symbol | Description |
|--------|-------------|
| $T = \mathbb{N}$ | Points corresponding to local time-slots |
| $T_G \subseteq \mathbb{Q}_{\geq 0}$ | Points on the real line corresponding to global time-slots |
| $D \in \mathbb{N}_{\geq 1}$ | Time resolution = number of global time-slots in a unit interval |
| $\epsilon = \frac{1}{D}$ | Length of a global time-slot |
| $N \in \mathbb{N}_{\geq 1}$ | System size = number of agents |
| $S = [1, \ldots, I]$ | State space of agents, with $I \in \mathbb{N}$ states |
| $\{X_i^{(N)}(t) : t \in T\}$ | Process corresponding to agent $i$, with $i \in \{1, \ldots, N\}$ |
| $K_i$ | Transition map of $X_i^{(N)}(t)$ |
| $\hat{K}_i$ | Modified process corresponding to agent $i$ |
| $\hat{Y}^{(N)}(t) : t \in T_G$ | Process for the system of $N$ agents, on $S^N$ |
| $\hat{\mathcal{K}}^{(N)}$ | Transition map of $\hat{Y}^{(N)}(t)$ |
| $\Delta$ | Set of occupancy measures |
| $\{\hat{M}^{(N)}(t) : t \in T_G\}$ | Normalised population process on $\Delta^{(N)} \subset \Delta$ |
| $\hat{\mathcal{K}}^{(N)}$ | Transition map of $\hat{M}^{(N)}(t)$ |
| $p^{(N)}$ | Transition map of the agent model $\{[\hat{X}_1^{(N)}, \hat{M}^{(N)}(t)] : t \in T_G\}$ |
| $p^{(N)}_{s,s'}$ | Agent transition map, with $s, s' \in S$ |
| $Q^{(N)}_{s,s'}$ | Infinitesimal agent transition map, with $s, s' \in S$ |
| $\{\hat{M}^{(N)}(t) : t \in \mathbb{R}_{\geq 0}\}$ | Normalised population process with continuous paths |
| $\hat{W}^{(N)}(t) : t \in T_G$ | Object (agent) state-change frequency in interval $[0, t]$ |
| $\hat{F}^{(N)}$ | Expected instantaneous change in system state |
| $\Phi \subseteq \{g : \mathbb{R}_{\geq 0} \rightarrow \Delta\}$ | Set of deterministic approximations |
| $F^*$ | The limit of the sequence of drifts $\{F^{(N)}\}$ |
| $\rho_N$ | The probability measure induced by $Y^{(N)}(t)$ |
| $\epsilon_N$ | The empirical measure derived from $Y^{(N)}(t)$ |
| $\tilde{F}^{(N)}_{s,s'}$ | The Poisson mean of intensity from $s$ to $s'$ with $s, s' \in S$ |
| $\tilde{F}^{(N)}$ | The mean drift of the normalised population process |
1. The time-slots are fully synchronized and the \( N \) processes simultaneously update their states.

2. Processes have independent time-slots, which occur at the same rate over sufficiently long intervals of time.

The two approaches may lead to systems with different behaviours (see remark below). For a discussion on the approximation of systems with simultaneous update (or synchronous DTMCs) refer to \([9, 25]\). Our discussion is about systems with independent time-slots, since this assumption allows us to embed the discrete-time description of agents’ behaviours in a continuous-time setting.

Formally, the type \( \mathbb{B} \) behaviour can be stated as follows. For two processes \( i \) and \( i' \) where \( i \neq i' \), if process \( i \) does a transition in an instant of time then process \( i' \) almost never does a transition simultaneously. Here we say that these systems satisfy the clock independence assumption.

Technically, we enforce the clock independence assumption through scaling the duration of time-slots and modifying agent transition probabilities as follows. Let \( D \in \mathbb{N}_{>1} \) be the time resolution, and let \( \epsilon \in \mathbb{Q}_{\geq 0} \) be a positive rational number (a probability) defined as \( \epsilon = \frac{1}{D} \). Let \( T_G \subseteq \mathbb{Q}_{\geq 0} \) be the countable set:

\[
T_G = \{0, \epsilon, 2\epsilon, \ldots\}.
\]

We call the set \( T_G \) the system or global time, as opposed to the agent or local time \( T \). Next, let the probability of an agent doing a transition in a time-slot be \( \epsilon \). In this new setting, for \( 1 \leq i \leq N \) define stochastic processes \( \{\hat{X}_i^{(N)}(t) : t \in T_G\} \), each with transition maps \( \hat{K}_i : S^N \times S \rightarrow [0, 1] \), such that for all \( \vec{v} \in S^N \) and \( s \in S \):

\[
\hat{K}_i(\vec{v}, s) = \begin{cases} 
\epsilon \ K_i(\vec{v}, s) & \text{if } s \neq \vec{v}_i, \\
(1 - \epsilon) + \epsilon \ K_i(\vec{v}, s) & \text{if } s = \vec{v}_i.
\end{cases}
\]

In the new setting, let \( E \) be the event that agent \( i \) does a transition in a time-slot, and \( E' \) be the event that agent \( i' \neq i \) does a transition exactly in the same time-slot in \( T_G \). Then by independence of agent transition maps:

\[
P(E'|E) = P(E') = \epsilon.
\]

Observe that the clock independence assumption is satisfied as \( D \rightarrow \infty \) (i.e., \( \epsilon \rightarrow 0 \)):

\[
\lim_{D \rightarrow \infty} P(E'|E) = 0.
\]

**Remark.** A condition under which systems with simultaneous updates and the clock independence assumption show divergent behaviours is when agent processes are not aperiodic. In such cases the system satisfying the clock independence assumption becomes aperiodic, while the system with simultaneous updates is not aperiodic.

Let \( \{Y^{(N)}(t) : t \in T_G\} \) be a stochastic process with states \( Y^{(N)}(t) = \{\hat{X}_1^{(N)}(t), \ldots, \hat{X}_N^{(N)}(t)\} \). The process \( Y^{(N)}(t) \) represents the behaviour of the entire system, and is a time-homogeneous discrete-time Markov process with transition map \( \mathcal{K}^{(N)} : S^N \times S^N \rightarrow [0, 1] \) in which for \( \vec{v}, \vec{v}' \in S^N \):

\[
\mathcal{K}^{(N)}(\vec{v}, \vec{v}') = \prod_{i=1}^{N} \hat{K}_i(\vec{v}, \vec{v}_i').
\]

### 3.2 Mean field interaction models

In this part we define mean field interaction models \([\mathbb{2}]\), which comprise the class of processes \( Y^{(N)}(t) \) for which we find the mean field approximations.

Let \( \pi : \{1, \ldots, N\} \rightarrow \{1, \ldots, N\} \) be a bijection. The function \( \pi \) is called a permutation over the set \( \{1, \ldots, N\} \). Additionally, for a vector \( \vec{v} = (s_1, \ldots, s_N) \) define \( \pi(\vec{v}) \) as:

\[
\pi(\vec{v}) = (s_{\pi(1)}, \ldots, s_{\pi(N)}).
\]

**Definition 16** (Mean Field Interaction Models \([\mathbb{2}]\)). Let \( Y^{(N)}(t) \) be the process defined earlier, and let \( \pi \) be any permutation over the set \( \{1, \ldots, N\} \). If for all \( \vec{v}, \vec{v}' \in S^N \):

\[
\mathcal{K}^{(N)}(\vec{v}, \vec{v}') = \mathcal{K}^{(N)}(\pi(\vec{v}), \pi(\vec{v}'))
\]

holds, \( Y^{(N)}(t) \) is called a mean field interaction model with \( N \) objects.
It follows from the above definition that entries in \( \mathcal{K}^{(N)} \) may depend on the number of agents in each state, not on the state of a certain agent. Let \( \Delta = \{ \hat{m} \in \mathbb{R}^I : \sum_{s \in S} \hat{m}_s = 1 \land \forall s. \hat{m}_s \geq 0 \} \) be a set of vectors, which we call the set of occupancy measures. For a system of size \( N \), take the countable subset \( \Delta^{(N)} = \{ \hat{m} \in \mathbb{R}^I : \sum_{s \in S} \hat{m}_s = 1 \land \forall s. (\hat{m}_s \geq 0 \land N\hat{m}_s \in \mathbb{N}) \} \). The set \( \Delta^{(N)} \subset \Delta \) is an alternative representation of the state space of the system, in which for \( \hat{m} \in \Delta^{(N)} \) and \( i \in S \) the value \( \hat{m}_i \) expresses the proportion of agents that are in state \( i \). For a mean field interaction model we define the normalized population process \( \{ \hat{M}^{(N)}(t) : t \in T_G \} \) on \( \Delta^{(N)} \) such that for \( s \in S \):

\[
\hat{M}^{(N)}_s(t) = \frac{1}{N} \sum_{1 \leq s \leq N} \mathbf{1} \left( \hat{X}^{(N)}_s(t) = s \right), \tag{7}
\]

where \( \mathbf{1} \) is an indicator function\(^2\). Using the fact that \( Y^{(N)}(t) \) is a mean field interaction model, it is possible to move back and forth between processes \( Y^{(N)}(t) \) and \( M^{(N)}(t) \). In the following discussion we illustrate this fact.

Let \( \hat{m}, \hat{m}' \in \Delta^{(N)} \) be occupancy measures. Let \( \tilde{v} \in S^V \) be a vector defined as:

\[
\tilde{v} = (1, \ldots, 1, \ldots, I, \ldots, I),
\]

where the notation \( \tilde{s}_1, \ldots, s \) shows a sequence of length \( \ell \) in which all the elements have value \( s \). Thus the vector \( \tilde{v} \) consists of \( N\tilde{s}_1 \) elements with value 1, \( N\tilde{s}_2 \) elements with value 2, and so forth. Consider the set \( V' \) where:

\[
V' = \{ \tilde{v}' \in S^V \mid \exists \pi. (\tilde{v}' = \pi(1, \ldots, 1, \ldots, I, \ldots, I)) \}.
\]

The set \( V' \) is the set of all possible configurations of the process \( Y^{(N)}(t) \) which translate to \( \hat{m}' \) after normalization. For process \( M^{(N)}(t) \) we define the new transition map \( \check{\mathcal{K}}^{(N)}: \Delta^{(N)} \times \Delta^{(N)} \to [0, 1] \):

\[
\check{\mathcal{K}}^{(N)}(\hat{m}, \hat{m}') = \sum_{\tilde{v} \in V'} \mathcal{K}^{(N)}(\tilde{v}, \tilde{v}')
\]

Similarly, we define the process \( \{ (\hat{X}^{(N)}_1(t), M^{(N)}(t)) : t \in T_G \} \) which models the behaviour of an agent in the context of a mean field interaction model, and call it the agent model. Let \( \hat{m}, \hat{m}' \in \Delta^{(N)} \) be occupancy measures and \( s, s' \in S \) be states, define the vector:

\[
\tilde{v} = (s, 1, \ldots, 1, s, \ldots, s, \ldots, I, \ldots, I).
\]

Consider the set \( V' \) which consists of all the permutations of the vector:

\[
V' = \{ \tilde{v}' \in S^V \mid (\tilde{v}'_1 = s') \land \exists \pi. (\tilde{v}' = \pi(1, \ldots, 1, \ldots, I, \ldots, I)) \}.
\]

The transition map of the agent model is the function \( P^{(N)}_1: S \times \Delta^{(N)} \times S \times \Delta^{(N)} \to [0, 1] \) defined as follows:

\[
P^{(N)}_1(s, \hat{m}, s', \hat{m}') = \begin{cases} 
\sum_{\tilde{v} \in V'} \mathcal{K}^{(N)}(\tilde{v}, \tilde{v}') & \text{if } N\hat{m}'_s \geq 1 \text{ and } N\hat{m}_s \geq 1, \\
0 & \text{otherwise}.
\end{cases}
\]

Based on \( P^{(N)}_1 \), for \( s, s' \in S \) we also define the probability:

\[
P^{(N)}_{s,s'}(\hat{m}) = \sum_{\hat{m}' \in \Delta} P^{(N)}_1(s, \hat{m}, s', \hat{m}')
\]

that an agent moves from state \( s \) to state \( s' \), in the context \( \hat{m} \in \Delta^{(N)} \).

For each \( s, s' \in S \) where \( s \neq s' \), the expected proportion of the agents that are in state \( s \) at time \( t \) and move to state \( s' \) over a unit time interval \([t, t+1)\) is:

\[
\sum_{i=0}^{D} i D \left[ P^{(N)}_{s,s'}(\hat{m}) \right] (1 - P^{(N)}_{s,s'}(\hat{m}))^{D-i} = D P^{(N)}_{s,s'}(\hat{m}).
\]

\(^2\)Indicator functions are defined in Section 2.1.

\(^3\)For any system of size \( N \), we should be careful about the occupancy measures that refer to invalid/meaningless states, since the set \( \Delta \) is uncountable.

12
By taking the clock independence assumption into account, for \( s \neq s' \) define the functions \( Q_{s,s'}^{(N)} : \Delta^{(N)} \rightarrow \mathbb{R}_{\geq 0} \) which for \( \vec{m} \in \Delta^{(N)} \) satisfy:

\[
Q_{s,s'}^{(N)}(\vec{m}) = \lim_{\Delta \rightarrow \infty} D P_{s,s'}^{(N)}(\vec{m}).
\]

Note that due to the construction of the probabilities \( P_{s,s'}^{(N)}(\vec{m}) \), for \( s \neq s' \) the limit always exists (i.e., it is in the interval \([0, 1]\)). The mapping \( Q^{(N)} \) is called the infinitesimal agent transition map, which can be interpreted as a transition rate matrix, meaning that for an agent in state \( s \) the time until it moves to state \( s' \) converges to an exponentially distributed random variable with mean: \( \frac{1}{Q_{s,s'}} \).

The time instants in the set \( T_G \) are discrete. However, as we see later in the approximation it is necessary to observe the population process \( \{M^{(N)}(t) : t \in T_G\} \) at continuous times \( t \in \mathbb{R}_{\geq 0} \). Based on \( M^{(N)}(t) \) we define a new stochastic process \( \{\bar{M}^{(N)}(t) : t \in \mathbb{R}_{\geq 0}\} \) with new sample paths which are right-continuous functions with left limits (cadlågs). For \( t \in \mathbb{R}_{\geq 0} \) the process \( \bar{M}^{(N)}(t) \) satisfies:

\[
\bar{M}^{(N)}(t) = M^{(N)}(\epsilon[D t]).
\]

To illustrate how one derives the mean field approximation of a system, we use a running example in this and the following sections. In the example of this section we start from an informal specification of an agent behaviour, and derive the transition map of the corresponding normalized population process.

**Example 3.1.** Consider a network of \( N \) nodes (agents) operating on a single shared channel. The network is saturated, meaning that all the nodes always have messages to transmit. A node can be in one of the states \( S = \{1, 2\} \). In state 1 a node is waiting, and with probability \( p_1 \) decides to transmit a message. All communications start with the transmission of a message, and a successful communication is then marked by the receipt of an acknowledgement, whereas a failed communication ends in a timeout. Both cases occur in the space of a single time-slot. If the communication succeeds, the node will remain in state 1 and wait to transmit the next message and if it fails, the node moves to state 2 in which it tries retransmitting the message. A node in state 2 retransmits the message with probability \( p_2 \). The node then essentially behaves in the same way as in state 1.

The probability of success depends on the number of nodes currently using the channel, as follows. If \( n \in \mathbb{N} \) nodes are using the channel, then the success probability of each participating communication is:

\[
p_s(n) = 2^{-n},
\]

that is, the channel degrades in quality exponentially as the number of active nodes increases. A diagram representing the behaviour of each node is given in Figure 1.

![Figure 1: The behaviour of a node in Example 3.1](image)

The number of transmitting nodes is \( n \).

Let \( \vec{v} \in S^N \) be the state of the network, with \( n_1 \) nodes in state 1 and \( n_2 = N - n_1 \) nodes in state 2. Let \( tr_1(\vec{v}) \) be the total number of nodes that are in state 1 which decide to transmit a message in system state \( \vec{v} \), then \( tr_1(\vec{v}) \) is a binomial random variable with distribution \( B(n_1, p_1) \). In a similar fashion \( tr_2(\vec{v}) \), the total number of nodes in state 2 which decide to transmit a message in system state \( \vec{v} \) is \( B(n_2, p_2) \) distributed. Then a communication in state \( \vec{v} \) will succeed with probability \( p_s(tr_1(\vec{v}) + tr_2(\vec{v})) \).

Let \( s \in S \) be the next state of a node \( i \) (\( 1 \leq i \leq N \)), based on the description above the transition matrix for this node is:

\[
K_i(\vec{v}, s) = \begin{pmatrix}
1 - p_s(1 - p_s(n)) & 1 - p_s p_s(n)
\end{pmatrix}
\]

where the row is determined by the element \( \vec{v} \) (current state) and the column by \( s \).

We use the clock independence assumption to compose the population process. To extend this assumption to the description of our radio network, we implicitly assume that the duration of message transmission is exponentially distributed, i.e., since the transitions are memoryless, the sojourn time of individuals in states is exponentially distributed. The modified transition matrix for node \( i \) is:

\[
\bar{K}_i(\vec{v}, s) = \begin{pmatrix}
1 - p_s(1 - p_s(tr_1(\vec{v}) + tr_2(\vec{v}))) & p_s(1 - p_s(tr_1(\vec{v}) + tr_2(\vec{v})))
\end{pmatrix}
\]
In which the probability of success and failure have been scaled by a factor $\epsilon$. The composed system is a mean field interaction model. This is due to the definition of functions $tr_1$ and $tr_2$, which do not depend on states of specific nodes, but rather on the aggregate number of nodes in states 1 and 2.

Consider the normalized population model with occupancy measures $\Delta = \{m \in \mathbb{R}^2 : \sum_i m_i = 1 \land \forall i, m_i \geq 0\}$ and the corresponding subset $\Delta^{(N)}$. Let $\bar{m} \in \Delta^{(N)}$; then using (2) when the system is in state $\bar{m}$ the total number of communicating agents is $X_1 + X_2$ where $X_1 \sim B(Nm_1, p_1)$ and $X_2 \sim B(Nm_2, p_2)$. For an agent in state $s$ the rate of moving to an state $s' \neq s$ is given by:

$$
\begin{align*}
Q_{1,2}^{(N)}(\bar{m}) &= \mathbb{E} [p_1(1 - p_2(X_1 + X_2))] , \\
Q_{2,1}^{(N)}(\bar{m}) &= \mathbb{E} [p_2p_1(X_1 + X_2)] .
\end{align*}
$$

The map $Q^{(N)}$ derived here defines the stochastic behaviour of an agent in the population process. In the sections that follow, we use this map to derive the mean field approximation of the population process.

### 3.3 Drift and the time evolution of the process $\bar{M}^{(N)}(t)$

In this section we define the drift as a way to characterize the behaviour of the normalized population process $\bar{M}^{(N)}(t)$ in its first moment. This provides the basis for defining the mean drift, which is given in Section 4.

Define $W_{s,s'}^{(N)}(t)$ as the random number of objects which do a transition from state $s$ to state $s'$ in the system at time $t \in T_U$, i.e.,

$$
W_{s,s'}^{(N)}(t + \epsilon) = \sum_{k=1}^{N} \mathbb{I} \{ \bar{X}_k^{(N)}(t) = s, \bar{X}_k^{(N)}(t + \epsilon) = s' \} .
$$

The instantaneous changes of the system $\bar{M}^{(N)}(t)$ can be tracked by the following random process:

$$
\bar{M}^{(N)}(t + \epsilon) - \bar{M}^{(N)}(t) = \sum_{s,s' \in S, s \neq s'} \frac{W_{s,s'}^{(N)}(t + \epsilon)}{N} (\bar{e}_{s'} - \bar{e}_s)
$$

where $\bar{e}_s$ is a unit vector of dimension $l$ with value 1 in position $s$. Then the expected value of the instantaneous change is the function $\bar{F}^{(N)} : \Delta^{(N)} \rightarrow \mathbb{R}^l$ where:

$$
\bar{F}^{(N)}(\bar{m}) = \mathbb{E} \left[ \bar{M}^{(N)}(t + \epsilon) - \bar{M}^{(N)}(t) \mid \bar{M}^{(N)}(t) = \bar{m} \right] = \sum_{s,s' \in S} \bar{m}_s \bar{F}^{(N)}_{s,s'}(\bar{m})(\bar{e}_{s'} - \bar{e}_s).
$$

The drift is the function $F^{(N)} : \Delta^{(N)} \rightarrow \mathbb{R}^l$ defined as:

$$
F^{(N)}(\bar{m}) = \lim_{D \rightarrow \infty} D\bar{F}^{(N)}(\bar{m}) = \sum_{s,s' \in S, s \neq s'} \bar{m}_s Q^{(N)}_{s,s'}(\bar{m})(\bar{e}_{s'} - \bar{e}_s).
$$

In the above formula, we may use $F^{(N)}_{s,s'}(\bar{m})$ to represent the summand:

$$
\bar{m}_s Q^{(N)}_{s,s'}(\bar{m}),
$$

which we call the intensity of transitions from $s$ to $s'$.

Below, we present properties that the drifts may or may not satisfy, and are of interest in our discussion.

**Smoothness:** For all $N \geq 1$, there exist Lipschitz continuous functions $F^{(N)} : \Delta \rightarrow \mathbb{R}^l$ which for all $\bar{m} \in \Delta^{(N)}$ satisfy: $F^{(N)}(\bar{m}) = F^{(N)}(\bar{m})$.

**Boundedness:** For all $N \geq 1$, $F^{(N)}$ are bounded on $\Delta^{(N)}$.

**Limit existence:** Assuming smoothness, the sequence of drifts $\{F^{(N)}\}$ converges uniformly to a bounded function $F^* : \Delta \rightarrow \mathbb{R}^l$.

In the literature, the single term density dependence is often used to refer to boundedness and limit existence (22).

**Remark.** The smoothness assumption appears in more or less the same shape throughout the literature. However our version is slightly more restrictive, since it allows us to skip some discussions regarding topological spaces. Moreover, in the contexts where it is clear that smoothness holds, we overload the name $F^{(N)}$ to refer to the function $F^{(N)}$ instead.
Essentially, the drift extends the vector representing the expected instantaneous changes into the unit time interval. Using drift, one can express how the expected value \( \mathbb{E}[\bar{M}^N(t)] \) will evolve over time, a fact which we formally express through the following proposition.

**Proposition 1.** For the process \( \bar{M}^N(t) \), and its drift \( F^N \) the following equation holds:

\[
\mathbb{E} \left[ \bar{M}^N(t) \mid \bar{M}^N(0) \right] - \bar{M}^N(0) = \int_0^t \mathbb{E} \left[ F^N(\bar{M}^N(s)) \mid \bar{M}^N(0) \right] ds.
\]

In its differential form, the equation above suggests that the expected trajectory of the process \( \bar{M}^N(t) \) is a solution of the following system of ordinary differential equations:

\[
\frac{d}{dt} \mathbb{E} \left[ \bar{M}^N(t) \right] = \mathbb{E} \left[ F^N(\bar{M}^N(t)) \right]
\]

with the initial value \( \bar{M}^N(0) \). In practice the term \( \mathbb{E} \left[ F^N(\bar{M}^N(t)) \right] \) is difficult to describe. Indeed, one of the reasons why we are interested in the mean field approximation is to avoid the exact calculation of the distribution of the random process \( \bar{M}^N(t) \). In Section 3.2 we saw how a population process which satisfies the clock independence assumption can be derived.

Example 3.2. We use the maps \( Q^N_\Delta \) given by (9) to derive the drift of the system described in Example 3.1. A simple substitution gives the following sequence of drifts:

\[
F^N(\bar{m}) = \left\langle -\bar{m}_1 \psi_1, p_1(1 - \psi_1(X_1 + X_2)) + \bar{m}_2 \psi_2(p_2(X_1 + X_2)) \right\rangle
\]

in which \( X_1 \sim B(N\bar{m}_1, p_1) \) and \( X_2 \sim B(N\bar{m}_2, p_2) \). This simplifies to:

\[
F^N(\bar{m}) = \left\langle -\bar{m}_1 p_1(1 - \frac{\bar{m}_1}{N\bar{m}_1}) \bar{m}_2 p_2(1 - \frac{\bar{m}_2}{N\bar{m}_2}), \bar{m}_1 p_1(1 - \frac{\bar{m}_1}{N\bar{m}_1}) - \bar{m}_2 p_2(1 - \frac{\bar{m}_2}{N\bar{m}_2}) \right\rangle
\]

It can be shown that the inequality below is always satisfied for \( \bar{m}, \bar{m}' \in \Delta \):

\[
|F^N(\bar{m}') - F^N(\bar{m})| \leq \sqrt{2} |\bar{m}' - \bar{m}|
\]

which proves that \( F^N \) are Lipschitz continuous on \( \Delta \). In the same manner, it can be shown that for all \( \bar{m} \in \Delta \), \(|F^N(\bar{m})| \leq \sqrt{2} \).

Therefore it is safe to assume that \( F^N \) satisfies both smoothness and boundedness. Moreover, for any \( p_1, p_2 > 0 \) and \( \bar{m} \in \Delta \) we have:

\[
F^*(\bar{m}) = \lim_{N \to \infty} F^N(\bar{m}) = \left\langle -p_1 \bar{m}_1, p_1 \bar{m}_1 \right\rangle
\]

which shows that limit existence is also satisfied.

### 3.4 Approximations of mean field interaction models

In Section 3.4 we saw how a population process which satisfies the clock independence assumption can be derived. In Section 3.5 we derived the drift from the population process, and introduced some conditions (smoothness, boundedness and limit existence) which can hold for the drift. In this section, we explain how the drift satisfying all these conditions can be used to derive a deterministic approximation for the behaviour of the population process.

**Definition 17** (Deterministic Approximations). For \( N \geq 1 \), let \( F^N \) be a drift for which smoothness holds. Let \( \{ g : \mathbb{R}_{\geq 0} \to \Delta \} \) be the class of functions from \( \mathbb{R}_{\geq 0} \) (continuous time) to \( \Delta \). Then \( \Phi \subset \{ g : \mathbb{R}_{\geq 0} \to \Delta \} \) is the set of deterministic approximations for which every \( \phi \in \Phi \) at time \( t \in \mathbb{R}_{\geq 0} \) satisfies the following system of ordinary differential equations (ODEs):

\[
\frac{d}{dt}\phi(t) = F^N(\phi(t)).
\]

Consider \( \phi(0) = \phi_0 \) to be the initial condition. According to smoothness and based on the Picard-Lindelöf theorem (Theorem 3) since \( F^N \) is Lipschitz continuous, there exists a unique solution to the above system of ODEs. Therefore in the set \( \Phi \) there is a unique element \( \phi^N(t) \) which satisfies \( \phi^N(0) = \phi_0 \), called the deterministic approximation for which the following theorem holds.
Theorem 5 (Mean Field Approximation, cf. [2], Theorem 1). For $N \geq 1$, let $\{\mathcal{M}^N(t)\}$ be a sequence of normalised population processes. Let $\{F^N\}$ be the corresponding drifts which satisfy smoothness and boundedness. Let $\{\phi^N(t)\}$ be the corresponding sequence of deterministic approximations. Then there exist real constants $c_1, c_2 > 0$ for which for any time horizon $T \in \mathbb{R}_{\geq 0}$:

$$\sup_{0 \leq t \leq T} |\mathcal{M}^N(t) - \phi^N(t)| \leq \exp(c_1 T) \left( |\mathcal{M}^N(0) - \phi_0| + \mathcal{M}^N(T) \right)$$

in which $\mathcal{M}^N(T)$ is a stochastic process satisfying the following conditions:

$$\mathbb{E} \left[ \mathcal{M}^N(T) \right] = 0$$

and:

$$\mathbb{E} \left[ \mathcal{M}^N(T)^2 \right] \leq \frac{c_2 T}{N}.$$

Assuming that limit existence holds, define the limit system of ODEs as: $\phi'(t) = F^*(\phi(t))$ with initial condition $\phi_0$. The following result is a direct consequence of Theorem 5.

Corollary 1. Consider the assumptions of Theorem 5 and assume limit existence holds. Let $\phi'(t)$ be the solution to the corresponding limit system of ODEs. Assume $\lim_{N \to \infty} |\mathcal{M}^N(0) - \phi_0|^2 = 0$. Then for any finite time horizon $T < \infty$:

$$\lim_{N \to \infty} \mathbb{E} \sup_{0 \leq t \leq T} |\mathcal{M}^N(t) - \phi^*(t)|^2 = 0.$$

Remark. Based on Theorem 5 proving convergence in mean is straightforward, meaning that for all $T$:

$$\lim_{N \to \infty} \mathbb{E} \sup_{0 \leq t \leq T} |\mathcal{M}^N(t) - \phi^*(t)| = 0.$$

In the following example, we derive the limit system of ODEs for the system discussed in Example 3.2 and describe its solutions.

Example 3.3. Consider the drift given in (13). The limit system of ODEs for the system described in Example 3.1 is:

$$\frac{d}{dt} \phi_1^*(t) = -p_1 \phi_1^*(t) + \frac{d}{dt} \phi_2^*(t) = p_1 \phi_1^*(t).$$

which together with the initial condition $\phi^*(0)$ has the following general solution:

$$\phi_1^*(t) = \phi_1^*(0) e^{-p_1 t}, \quad \phi_2^*(t) = -\phi_1^*(0) e^{-p_1 t} + 1.$$ 

Obviously the solution heavily depends on the initial values $\phi^*(0)$, but the system has a global attractor at $(0, 1)$.

4 Propagation of chaos and the mean drift

Corollary 11 justifies the use of drift for finding the approximation in cases where the number of agents $N$ is unboundedly large. However, for bounded $N$ the upper bound on the error found by Theorem 5 rarely satisfies one’s expectations. In this section we explore the possibility of using the alternative ODEs in (11). We explain the notion of propagation of chaos, and show how it relates to what we call the mean drift.

For a set $E$ let $\mathcal{M}(E)$ denote the set of probability measures on $E$. Let the set $\mathcal{S}$ be defined as in Section 3.1 and for $s, s' \in \mathcal{S}$ define the distance between $s$ and $s'$ as

$$d(s, s') = \begin{cases} 0, & \text{if } s = s' \\ 2, & \text{if } s \neq s', \end{cases}$$

which makes the pair $(\mathcal{S}, d)$ a metric space, with the implication that $\mathcal{S}^N$ is also metrizable which allows the definition that follows.

Definition 18 (\(\rho\)-chaotic Sequence). Let $\rho \in \mathcal{M}(\mathcal{S})$ be a probability measure. For $N \geq 1$, the sequence $\{\rho_N\}$ of measures, each in $\mathcal{M}(\mathcal{S}^N)$, is \(\rho\)-chaotic iff for any natural number $k$ and bounded functions $f_1, \ldots, f_k : \mathcal{S} \to \mathbb{R},$

$$\lim_{N \to \infty} \int_{\mathcal{S}^N} f_1(v_1) f_2(v_2) \ldots f_k(v_k) \rho_N(d\vec{v}) = \prod_{i=1}^k \int_{\mathcal{S}} f_i(s) \rho(ds).$$
In short, a chaotic sequence maintains a form of independence in the observations of separate agents in the limit. In the literature this independence is often called the propagation of chaos, and in the context of Bianchi’s analysis the decoupling assumption.

In the rest of this discussion, consider the finite instant in time \( t \in \mathbb{R}_{\geq 0} \) and its close counterpart \( \tau \in T_G \) with \( \tau = \epsilon |Dt| \). Recall the non-normalized mean field interaction model at time \( \tau \), \( Y_N^{(\tau)}(\tau) = \{\hat{X}_1(\tau), \ldots, \hat{X}_N(\tau)\} \), which is a random element in \( S^N \). For \( N \geq 1 \), let \( \rho_N \in \mathcal{M}(S^N) \) be the laws (probability distributions) of \( Y_N^{(\tau)}(\tau) \). We now state a theorem proven by Sznitman \([16, 30]\). For the random element \( \hat{v} \) in \( S^N \) define the empirical measure \( \epsilon_N \), a random element in \( \mathcal{M}(S) \), as follows:

\[
\epsilon_N(\hat{v}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{v}_i}
\]  

(14)

where \( \delta_{\hat{v}_i} \) is the Dirac measure centred on point \( \hat{v}_i \). In what follows we always assume \( \hat{v} = Y_N^{(\tau)}(\tau) \), and hence write \( \epsilon_N \) instead of \( \epsilon_N \left( Y_N^{(\tau)}(\tau) \right) \). The following is known as Sznitman’s result in the literature.

**Theorem 6** (See \([30\), Proposition 2.2\]). The sequence \( \{\rho_N\} \) of measures is \( \rho \)-chaotic if and only if the sequence of empirical measures \( \{\epsilon_N\} \) converges to \( \delta_\rho \), that is:

\[
\rho_0 \circ \epsilon_N^{-1} \to \delta_\rho.
\]

Based on the above theorem, it has been shown that the result of Corollary 1 implies that propagation of chaos also occurs in the sequence of distributions of mean field interaction models \( Y_N^{(\tau)}(\tau) \) (see also corollary 2 of \([2]\), and a review of similar results in \([14]\).

**Corollary 2.** Let \( \phi^*(t) \) satisfy Corollary 1. Let \( \mu \in \mathcal{M}(S) \) be a measure which for all points \( i \in S \) satisfies \( \mu(i) = \phi^*_i(t) \), then the sequence \( \{\rho_N\} \) of distributions of \( Y_N^{(\tau)}(\tau) \) is \( \mu \)-chaotic.

Let the measure \( \mu \) be defined as in Corollary 2. For \( k \in \{0, \ldots, N\} \), we are interested in finding the following probability given that the distributions of \( \{Y_N^{(\tau)}(\tau)\} \) form a \( \mu \)-chaotic sequence:

\[
P \left( N \hat{M}^{(N)}_i(t) = k \right).
\]

Using the fact that \( Y_N^{(\tau)}(\tau) \) is a mean field interaction model (or is symmetric), we have:

\[
P \left( N \hat{M}^{(N)}_i(t) = k \right) = \binom{N}{k} \int_{S^N} \mathbb{1}(\hat{v}_1 = i) \cdots \mathbb{1}(\hat{v}_k = i) \mathbb{1}(\hat{v}_{k+1} \neq i) \cdots \mathbb{1}(\hat{v}_N \neq i) \rho_N(d\hat{v})
\]

and since \( \{\rho_N\} \) is a \( \mu \)-chaotic sequence, based on Definition 13:

\[
\lim_{N \to \infty} P \left( N \hat{M}^{(N)}_i(t) = k \right) = \lim_{N \to \infty} \binom{N}{k} (\phi^*_i(t))^k (1 - \phi^*_i(t))^{N-k}.
\]

This justifies the usage of the following approximation for sufficiently large \( N \):

\[
P \left( N \hat{M}^{(N)}_i(t) = k \right) \approx \binom{N}{k} (\phi^*_i(t))^k (1 - \phi^*_i(t))^{N-k},
\]

which following the Poisson approximation theorem can be in turn approximated by:

\[
P \left( N \hat{M}^{(N)}_i(t) = k \right) \approx e^{-N\phi^*_i(t)} \frac{(N\phi^*_i(t))^k}{k!}.
\]

(15)

In the context of Bianchi’s analysis, see \([33\) for an implicit application of a similar approximation.

Let \( f_{\text{Poison}}(k; \lambda) \) denote the probability density function of a Poisson random variable with rate \( \lambda \). Let \( f : \Delta \to \mathbb{R} \) be a function which acts on the random variable \( \hat{M}^{(N)}_i(t) \). A major convenience in using the above terms is their mutual independence in the limit, which at time \( t \in \mathbb{R}_{\geq 0} \) allows the approximation of the expected value of \( f(\hat{M}^{(N)}_i(t)) \) as:

\[
E \left[ f(\hat{M}^{(N)}_i(t)) \right] \approx \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} f \left( \frac{k_1}{N}, \ldots, \frac{k_l}{N} \right) f_{\text{Poison}} \left( k_1; N\phi_i^{(N)}(t) \right) \cdots f_{\text{Poison}} \left( k_l; N\phi_i^{(N)}(t) \right).
\]

(16)
4.1 The mean drift

In this part, we explain the approximation of the ODEs in formula (11), using result (16). For the occupancy measure \( \bar{m} \in \Delta \), the Poisson mean of the intensity \( F_{s,x}^{(N)} \) is the function \( \tilde{F}_{s,x}^{(N)} : \Delta \rightarrow \mathbb{R} \), where:

\[
F_{s,x}^{(N)}(\bar{m}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_i=0}^{\infty} F_{s,x'}^{(N)} \left( \frac{k_1}{N}, \cdots, \frac{k_i}{N} \right) f_{\text{Poisson}}(k_1; N\bar{m}_1) \cdots f_{\text{Poisson}}(k_i; N\bar{m}_i).
\]

(17)

Subsequently, the mean drift \( \bar{F}^{(N)} : \Delta \rightarrow \mathbb{R}^I \) is defined as:

\[
\bar{F}^{(N)}(\bar{m}) = \sum_{s,s' \in S} \bar{F}_{s,x}^{(N)}(\bar{m})(\vec{e}_{s'} - \vec{e}_s).
\]

The Poisson mean of intensities and the mean drift have the following properties:

- If \( \bar{F}_{s,x}^{(N)}(\bar{m}) = \bar{m}_s \alpha \), where \( \alpha \) is a constant or a term \( \bar{m}_j \) for some \( j \in S \), then \( \bar{F}_{s,x}^{(N)}(\bar{m}) = \bar{F}_{s',x'}^{(N)}(\bar{m}) \).

- \( \bar{F}^{(N)}(\bar{m}) \) is defined for all \( \bar{m} \in \Delta \).

- Given smoothness and boundedness of the drift, \( \bar{F}^{(N)} \) is both Lipschitz continuous and bounded on \( \Delta \).

Based on the derivation of probabilities (15), it is easy to see that at time \( t < \infty \) we have:

\[
\mathbb{E} \left[ \bar{F}^{(N)}(\bar{M}^N(t)) \right] \approx \bar{F}^{(N)}(\bar{\phi}^{(N)}(t)).
\]

Moreover, based on (15), \( \mathbb{E} \left[ \bar{M}^N(t) \right] \approx \bar{\phi}^{(N)}(t) \), with which the following system of differential equations can be derived from (11):

\[
\frac{d}{dt} \bar{\phi}^{(N)}(t) = \bar{F}^{(N)}(\bar{\phi}^{(N)}(t))
\]

(18)

with the initial condition \( \bar{\phi}^{(N)}(0) = \bar{M}^{(N)}(0) \).

Our construction which relies on the notion of propagation of chaos, means that the differential equations (18) give better approximations as the system size \( N \) grows. This fact is demonstrated by the following theorem.

**Theorem 7.** For \( N \geq 1 \), let \( \left\{ \bar{F}^{(N)} \right\} \) be the sequence of drifts, and \( \left\{ \bar{F}^{(N)} \right\} \) be the corresponding sequence of mean drifts. Assume that the drifts satisfy smoothness and boundedness. Then for all \( \bar{m} \in \Delta \),

\[
\lim_{N \to \infty} \bar{F}^{(N)}(\bar{m}) = F^*(\bar{m})
\]

almost surely.

**Example 4.1.** For the system described in Example 3.1 we find the mean drift, using the drift \( \bar{F}^{(N)} \) described in (12). The mean drift takes the relatively simple shape which follows:

\[
\bar{F}^{(N)}(\bar{m}) = \begin{pmatrix}
\frac{p_0 \bar{m}_1 + p_2 \bar{m}_2}{2} e^{\frac{p_1 N \bar{m}_1 + p_2 N \bar{m}_2}{2}} - p_1 \bar{m}_1 \\
\frac{p_0 \bar{m}_1 + p_2 \bar{m}_2}{2} e^{\frac{p_1 N \bar{m}_1 + p_2 N \bar{m}_2}{2}} + p_1 \bar{m}_1
\end{pmatrix}
\]

This can be used to construct the following system of ODEs:

\[
\frac{d}{dt} \bar{\phi}^{(N)} = \frac{p_1 \bar{\phi}^{(N)}(t) + p_2 \bar{\phi}^{(N)}(t)}{2} e^{\frac{p_1 N \bar{\phi}^{(N)}(t) + p_2 N \bar{\phi}^{(N)}(t)}{2}} - p_1 \bar{\phi}^{(N)}(t)
\]

\[
\frac{d}{dt} \bar{\phi}^{(N)} = -\frac{p_1 \bar{\phi}^{(N)}(t) + p_2 \bar{\phi}^{(N)}(t)}{2} e^{\frac{p_1 N \bar{\phi}^{(N)}(t) + p_2 N \bar{\phi}^{(N)}(t)}{2}} + p_1 \bar{\phi}^{(N)}(t)
\]

Let \( p_1 = 0.008 \) and \( p_2 = 0.05 \) and let the initial condition be \( \bar{\phi}_0 = (1, 0) \), i.e., all the nodes are initially in state 1. In Fig. 2 the results of solving the ODEs for different values of \( N \) are given, and are compared with results from the explicit transient analysis of Markov models with simultaneous updates. Observe that regardless of the size of the system, the approximations derived by the two methods closely match one another.

5 Proofs

Detailed proofs for the theorems and corollaries follow.
5.1 Proof of Proposition 1

Lemma 3. Let $N \geq 1$ be a natural number, and $\{\bar{M}^{(N)}(t) : t \in T_G\}$ be a normalized population process satisfying the clock independence assumption. Let $\{T_t : t \in T\}$ be a sequence of linear operators, which for functions $f : \Delta^{(N)} \to \Delta^{(N)}$ and $t \in T$ satisfy:

$$T_t f(\bar{m}) = \mathbb{E} \left[ f(\bar{M}^{(N)}(t)) | \bar{M}^{(N)}(0) = \bar{m} \right].$$

then given the infinitesimal generator $A$ of $\bar{M}^{(N)}(t)$, for the identity function $f$ we have $Af = F^{(N)}$.

Proof. Given $\epsilon = \frac{1}{t}$, and using (2):

$$Af = \lim_{t \to 0} \frac{\mathbb{E} \left[ \bar{M}^{(N)}(t) | \bar{M}^{(N)}(0) \right] - \bar{M}^{(N)}(0)}{t}$$

$$= \lim_{t \to 0} \frac{\mathbb{E} \left[ \bar{M}^{(N)}(t) - M^{(N)}(\epsilon Dt) | M^{(N)}(0) \right] + \mathbb{E} \left[ M^{(N)}(\epsilon Dt) | M^{(N)}(0) \right] - M^{(N)}(0)}{t}$$

$$= \lim_{t \to 0} \frac{0 + \sum_{s=1}^{[Dt]} \mathbb{E} \left[ M^{(N)}(\epsilon s) - M^{(N)}(\epsilon (s-1)) | M^{(N)}(0) \right] - M^{(N)}(0)}{t}$$

$$= \lim_{t \to 0} \frac{\sum_{s=1}^{[Dt]} F^{(N)}(M^{(N)}(\epsilon (s-1)))}{t}$$

Based on the definition of $\bar{M}^{(N)}(t)$ the numerator can be written as:

$$\sum_{s=1}^{[Dt]} \epsilon F^{(N)}(M^{(N)}(\epsilon (s-1))) = \int_0^{[Dt]} F^{(N)}(\bar{M}^{(N)}(\epsilon s)) \epsilon ds$$

$$- \int_{[Dt]}^{[Dt]} F^{(N)}(\bar{M}^{(N)}(\epsilon s)) \epsilon ds$$

Since the construction of $F^{(N)}$ involves the clock independence assumption, $t \geq \epsilon$, in which case:

$$Af = \lim_{t \to 0} \frac{\int_0^{[Dt]} F^{(N)}(\bar{M}^{(N)}(\epsilon s)) \epsilon ds}{t}$$

$$= \lim_{t \to 0} \frac{\int_{[Dt]}^{[Dt]} F^{(N)}(\bar{M}^{(N)}(\epsilon s)) \epsilon ds}{t}$$

$$= F^{(N)}(\bar{M}^{(N)}(0)) = F^{(N)}(\bar{m}).$$

□
Thus, we have proven that \( \{f, F^{(N)}\} \in A \), and subsequently according to (3) the following equation holds:

\[
E[\tilde{M}^{(N)}(t) \mid \tilde{M}^{(N)}(0) - M^{(N)}(0)] = \int_0^t E[F^{(N)}(M^{(N)}(s)) \mid M^{(N)}(0)] ds.
\]

5.2 Proof of Theorem 5

The following proof and the proof for Corollary 1 are largely based on a similar proof in [15] which has been repeated in many other work. Here, our main goal is to give a simple enough version of the proof, yet without big unexplained leaps.

Given the random processes \( W_{s,s'}^{(N)}(t) \) defined in Section 3.3 and satisfying equation (10), define \( Z_{s,s'}^{(N)}(t) \) as the total random number of agents taking a transition from state \( s \in \mathcal{S} \) to \( s' \in \mathcal{S} \) in the time interval \([0, t]\):

\[
Z_{s,s'}^{(N)}(t) = \lim_{\delta \to 0} \sum_{i=1}^{[\delta/t]} W_{s,s'}(i, e),
\]

then the definition of the intensity of transitions from \( s \) to \( s' \) (also ultimately based on \( W_{s,s'}^{(N)}(t) \)) implies the following:

\[
E[Z_{s,s'}^{(N)}(t)] = N \int_0^t F_{s,s'}^{(N)}(M^{(N)}(s)) ds.
\]

(19)

The process \( Z_{s,s'}^{(N)}(t) \) is Markov in \( T_G \), and its sample paths are increasing functions which are right-continuous with left limits.

We now present a result given in [23] (Theorem 7.1). For each process \( Z_{s,s'}^{(N)}(t) \) associated with an intensity from \( s \) to \( s' \), assume an increasing Markov jump process (a counting process) \( N(t) \) for which \( \mathbb{E}[N(t)] = ct \) for some \( c \in \mathbb{R}_{\geq0} \). Then following [34], the process \( N(t) \) is a Poisson process. Moreover, following [26, 23], since the processes \( Z_{s,s'}^{(N)}(t) \) and \( N(t) \) are Markov (in \( T_G \) and in \( \mathbb{R} \) respectively), and increasing, for \( t \in T_G \) there exist stopping times \( \tau(t) \) which satisfy:

\[
Z_{s,s'}^{(N)}(t) = N(\tau(t)).
\]

(20)

Essentially, these stopping times maintain the equality of the values over the discrete jumps of processes. Following [26, 10] for all pairs \( s, s' \in \mathcal{S} \) the resulting Poisson processes are mutually independent. This allows freedom in combining these processes to describe the behaviour of the process \( \tilde{M}^{(N)}(t) \).

From (19) and (20) we may write:

\[
\mathbb{E}[N(\tau(t))] = N \int_0^t F_{s,s'}^{(N)}(\tilde{M}^{(N)}(s)) ds.
\]

Let \( \{\mathcal{Y}(t) : t \in \mathbb{R}\} \) be a unit Poisson process (see section 2.3). Using equation (6) we may exchange time and intensity to write \( N(\tau(t)) \) in terms of unit Poisson processes as \( N(\tau(t)) = \mathcal{Y}(\mathbb{E}[N(\tau(t))]) \), so we may rewrite equation (19) as:

\[
Z_{s,s'}^{(N)}(t) = \mathcal{Y} \left( N \int_0^t F_{s,s'}^{(N)}(\tilde{M}^{(N)}(s)) ds \right)
\]

(21)

Note that for the process \( \tilde{M}^{(N)}(t) \) we have:

\[
\tilde{M}^{(N)}(t) = \tilde{M}^{(N)}(0) + \sum_{s,s' \in \mathcal{S} \times \mathcal{S}} \frac{1}{N} Z_{s,s'}^{(N)}(t)(e_{s'} - e_s),
\]

which combined with equation (21) results in the following well-known format, often called the Poisson representation of the population process:

\[
\tilde{M}^{(N)}(t) = \tilde{M}^{(N)}(0) + \sum_{k=(s,s') \in \mathcal{S} \times \mathcal{S}} \frac{1}{N} \mathcal{Y}_k \left( N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds \right)(e_{s'} - e_s)
\]

(22)

For all pairs \( k = (s, s') \in \mathcal{S} \times \mathcal{S} \) consider the following process:

\[
\tilde{Y}_k \left( N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds \right) = \mathcal{Y}_k \left( N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds \right) - N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds
\]
which is a compensated unit Poisson process and hence a martingale (see Section 2.5). From here:

\[
\tilde{M}^{(N)}(t) = \tilde{M}^{(N)}(0) + \sum_{k=(s,r) \in S} \frac{1}{N} \tilde{Y}_k \left( N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds \right) (e_r - e_s)
\]

\[
= \tilde{M}^{(N)}(0) + \sum_{k=(s,r) \in S} \frac{1}{N} \tilde{Y}_k \left( N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds \right) (e_r - e_s) + \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds.
\]

In this decomposition, the process:

\[
\mu^{(N)}(t) = \sum_{k=(s,r) \in S} \frac{1}{N} \tilde{Y}_k \left( N \int_0^t F_k^{(N)}(\tilde{M}^{(N)}(s)) ds \right) (e_r - e_s)
\]

is a martingale and using (5) (Doob’s inequality) for \( \sigma = 2 \) we have:

\[
\mathbb{E} \left[ \sup_{t \in \mathcal{T}} \mu^{(N)}(t)^2 \right] \leq 4 \mathbb{E} \left[ \mu^{(N)}(T)^2 \right],
\]

and based on boundedness, \( F_k^{(N)} \) are bounded, say by constant \( A \), therefore:

\[
\mathbb{E} \left[ \mu^{(N)}(T)^2 \right] \leq \frac{1}{N} \sum_{r \in S} e_r A T.
\]

Set \( c_2 = 4A|I| \). Next, we write the system of ODEs in its integral form:

\[
\phi^{(N)}(t) = \phi(0) + \int_0^t F^{(N)}(\phi^{(N)}(s)) ds.
\]

For \( T \geq 0 \):

\[
\sup_{t \in \mathcal{T}} |\tilde{M}^{(N)}(t) - \phi^{(N)}(t)| \leq |\tilde{M}^{(N)}(0) - \phi^{(N)}(0)| + \sup_{t \in \mathcal{T}} |\mu^{(N)}(t)| + \sup_{t \in \mathcal{T}} \left| \int_0^t F^{(N)}(\tilde{M}^{(N)}(s)) ds - \int_0^t F^{(N)}(\phi^{(N)}(s)) ds \right|
\]

\[
\leq |\tilde{M}^{(N)}(0) - \phi^{(N)}(0)| + \sup_{t \in \mathcal{T}} |\mu^{(N)}(t)| + \sup_{t \in \mathcal{T}} \int_0^T \left| F^{(N)}(\tilde{M}^{(N)}(s)) - F^{(N)}(\phi^{(N)}(s)) \right| ds
\]

Since according to assumptions each intensity \( F_k^{(N)} \) is Lipschitz continuous, say by constant \( L_k \), there exists a constant \( L = \max(L_k : k \in \mathcal{S} \times \mathcal{S}) \), for which:

\[
\sup_{t \in \mathcal{T}} |\tilde{M}^{(N)}(t) - \phi^{(N)}(t)| \leq |\tilde{M}^{(N)}(0) - \phi^{(N)}(0)| + \sup_{t \in \mathcal{T}} |\mu^{(N)}(t)| + L \sup_{t \in \mathcal{T}} \int_0^T |\tilde{M}^{(N)}(s) - \phi^{(N)}(s)| ds
\]

\[
\leq |\tilde{M}^{(N)}(0) - \phi^{(N)}(0)| + \sup_{t \in \mathcal{T}} |\mu^{(N)}(t)| + L \int_0^T \sup_{t \in \mathcal{T}} |\tilde{M}^{(N)}(t) - \phi^{(N)}(t)| ds
\]

Let \( f(T) = \sup_{t \in \mathcal{T}} |\tilde{M}^{(N)}(t) - \phi^{(N)}(t)| \), using Grönwall’s inequality (Lemma 2):

\[
\sup_{t \in \mathcal{T}} |\tilde{M}^{(N)}(t) - \phi^{(N)}(t)| \leq \exp(LT) \left( |\tilde{M}^{(N)}(0) - \phi^{(N)}(0)| + \sup_{t \in \mathcal{T}} |\mu^{(N)}(t)| \right).
\]

Setting \( c_1 = L, \tilde{M}^{(N)}(T) = \sup_{t \in \mathcal{T}} |\mu^{(N)}(t)| \) and \( c_2 \) as found earlier then proves the theorem.

### 5.3 Proof of Corollary 1

In what follows we consider the assumptions of corollary 1 to hold. Our presentation takes several properties of topological spaces for granted, with the understanding that detailed proofs are available to the reader and we are allowed to avoid discussions aimed at overcoming such difficulties.

**Lemma 4.** Let the sequence of drifts \( F^{(N)} \) converge uniformly to the bounded function \( F^* \), that is:

\[
\lim_{N \to \infty} \sup_{\phi \in \mathcal{D}^{(N)}} |F^{(N)}(\tilde{m}) - F^*(\tilde{m})| = 0.
\]

Then the sequence of solutions \( \phi^{(N)} \) of respective ODEs converges to \( \phi^*(t) \), the solution of the limit system of ODEs:

\[
\lim_{N \to \infty} \phi^{(N)}(t) = \phi^*(t).
\]
Proof. Let \( \phi^{(N)}(t) \) be the solution of the initial value problem \( \phi'(t) = F^{(N)}(\phi(t)) \) with \( \phi(0) = \phi^{(N)}(0) = \phi_0 \) at time \( t \in \mathbb{R}_{\geq 0} \). It is possible to make the following assertions based on the unifom convergence assumption:

\[
\lim_{N \to \infty} \sup_t \left| F^{(N)}(\phi^{(N)}(t)) - F^*(\phi^{(N)}(t)) \right| = 0 \quad \Rightarrow \quad \lim_{N \to \infty} \sup_t \left| \phi^{(N)}(t) - F^*(\phi^{(N)}(t)) \right| = 0.
\]

The second limit implies that as \( N \to \infty \) at every point \( t \in \mathbb{R}_{\geq 0} \), \( \phi^{(N)}(t) \) is a solution to the limit system of ODEs. Given that \( \lim_{N \to \infty} \phi^{(N)}(0) = \phi^*(0) = \phi_0 \) and \( F^* \) are Lipschitz, the existence and uniqueness of the solution implies that for all \( t \):

\[
\lim_{N \to \infty} \phi^{(N)}(t) = \phi^*(t).
\]

We now return to the proof of Corollary 1. Using Theorem 5, the following inequality holds:

\[
\sup_{0 \leq t \leq T} \left| \hat{M}^{(N)}(t) - \phi^{(N)}(t) \right| \leq \exp(c_1 T) \left( M^{(N)}(T) + \left| \hat{M}^{(N)}(0) - \phi_0 \right| \right)
\]

therefore:

\[
\sup_{0 \leq t \leq T} \left| \hat{M}^{(N)}(t) - \phi^{(N)}(t) \right|^2 \leq \exp(c_1 T)^2 \left( M^{(N)}(T) + \left| \hat{M}^{(N)}(0) - \phi_0 \right|^2 \right)
\]

and,

\[
\mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| \hat{M}^{(N)}(t) - \phi^{(N)}(t) \right|^2 \right] \leq \exp(c_1 T)^2 \left( \mathbb{E}[M^{(N)}(T)^2] + \left| \hat{M}^{(N)}(0) - \phi_0 \right|^2 \right)
\]

\[
\leq \exp(c_1 T)^2 \left( c_2 T \frac{1}{N} + \left| \hat{M}^{(N)}(0) - \phi_0 \right|^2 \right).
\]

Hence since \( \hat{M}^{(N)}(0) \) converges to \( \phi_0 \) in mean-square, for \( T < \infty \):

\[
\lim_{N \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| \hat{M}^{(N)}(t) - \phi^{(N)}(t) \right|^2 \right] = 0,
\]

Which based on Lemma 4 implies that:

\[
\lim_{N \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| \hat{M}^{(N)}(t) - \phi^*(t) \right|^2 \right] = 0.
\]

### 5.4 Proof of Corollary 2

If we show that:

\[
\rho_N \circ \varepsilon_N^{-1} \rightarrow \delta_{\mu},
\]

then based on Theorem 6, the sequence must be \( \mu \)-chaotic. Observe that according to Markov’s inequality for any \( c > 0 \) and every \( A \subseteq \mathcal{S} \):

\[
\mathbb{P}_N \left[ |\varepsilon_N(A) - \mu(A)| \geq c \right] \leq \frac{\mathbb{E}_N \left[ |\varepsilon_N(A) - \mu(A)| \right]}{c}
\]

where \( \mathbb{E}_N \) and \( \mathbb{P}_N \) denote values calculated according to measure \( \rho_N \). By equation 7 (definition of \( M^{(N)}(t) \)) and definition of \( \mu \):

\[
\lim_{N \to \infty} \mathbb{E}_N \left[ |\varepsilon_N(A) - \mu(A)| \right] = \lim_{N \to \infty} \mathbb{E}_N \left[ \left| \frac{1}{N} \sum_{i=1}^{N} \delta^{(N)}(j)(A) - \mu(A) \right| \right]
\]

\[
= \lim_{N \to \infty} \mathbb{E}_N \left[ \left| \sum_{j=1}^{N} M_j^{(N)}(t) 1(j \in A) - \sum_{j=1}^{N} \phi^*_j(j) 1(j \in A) \right| \right]
\]

\[
\leq \lim_{N \to \infty} \mathbb{E}_N \left[ \sum_{j=1}^{N} \left| M_j^{(N)}(t) - \phi^*_j(t) \right| 1(j \in A) \right].
\]
According to Theorem 5 and Corollary 1 for all \( t < \infty \), \( \{M^N(t)\} \) converges to \( \phi^*(t) \) in both mean and mean square, therefore for every \( A \subseteq S \):

\[
\lim_{N \to \infty} \mathbb{E} \left[ \sum_{j=1}^{N} \left| M^N_j(t) - \phi^*_j(t) \right| \mathbb{I}(j \in A) \right] = 0.
\]

Based on (24) for any \( c > 0 \) and every \( A \subseteq S \):

\[
\lim_{N \to \infty} \mathbb{P}_N(\left| \epsilon_N(A) - \mu(A) \right| \geq c) = 0,
\]

which implies that for \( t < \infty \), \( \epsilon_N \) converges in probability to constant random element \( \mu \). Since convergence in probability implies convergence in distribution, we have shown that the measure \( \rho_N = \epsilon_N^{-1} \) converges to \( \delta_{\mu} \).

### 5.5 Proof of Theorem 7

In order to make the proof shorter and simpler to follow, we first make an (optional) assumption, which follows. A proof applying to the general mean drift defined in Section 4.1 would use the same ideas.

**Simplicity:** For each \( s, s' \in S \), there is one \( j \in S \) for which \( Q^{(N)}_{s,s'}(\bar{m}) = Q^{(N)}_{s,s'}(\bar{m}_j, \varepsilon_j) \), i.e., the value \( Q^{(N)}_{s,s'} \) is only determined by entry \( j \) of an occupancy measure \( \bar{m} \).

Suppose that \( Q^{(N)}_{s,s'} \) satisfies simplicity, fixing \( j \). The Poisson mean of the intensity \( F^{(N)}_{s,s'} \) is the function \( F^{(N)}_{s,s'} : \Delta \to \mathbb{R} \), where:

\[
F^{(N)}_{s,s'}(\bar{m}) = \bar{m}_j \sum_{k=0}^{N} \frac{k}{N} \mathbb{P}_{\text{Poisson}}(k; N\bar{m}_j) \tag{25}
\]

We first prove that for all \( \bar{m} \in \Delta \) for which \( F^{(N)}_{s,s'}(\bar{m}) \) is defined:

\[
\lim_{N \to \infty} F^{(N)}_{s,s'}(\bar{m}) = \lim_{N \to \infty} F^{(N)}_{s,s'}(\bar{m}_j),
\]

almost surely.

Based on the assumptions of the theorem, simplicity holds, fixing a \( j \) such that \( F^{(N)}_{s,s'}(\bar{m}) = F^{(N)}_{s,s'}(\bar{m}_j, \varepsilon_j) \). For \( 1 \leq k \leq N \) define the sequence of i.i.d. Poisson random variables \( X_k \), with rate \( \bar{m}_j \). Define:

\[
S_N = \sum_{k=1}^{N} X_k.
\]

Then \( S_N \) is Poisson distributed with rate \( N\bar{m}_j \):

\[
\mathbb{P}(S_N = k) = \frac{e^{-N\bar{m}_j} (N\bar{m}_j)^k}{k!} \tag{26}
\]

Following the weak law of large numbers (Theorem 1), for any \( \varepsilon > 0 \) we have:

\[
\lim_{N \to \infty} \mathbb{P}(\left| \frac{S_N}{N} - \bar{m}_j \right| \geq \varepsilon) = 0 \tag{27}
\]

Take an arbitrarily small value \( \varepsilon > 0 \). For \( N \to \infty \), the set \( \{0, \frac{1}{N}, \ldots, \frac{N(1+\varepsilon)}{N}\} \) is dense in the interval \( [0, 1+\varepsilon] \). Since \( \bar{m}_j \in [0, 1] \) and the above set is dense, there exists a largest non-empty neighbourhood \( \left[ \frac{k_1}{N}, \frac{k_2}{N} \right] \subset \{0, \frac{1}{N}, \ldots, \frac{N(1+\varepsilon)}{N}\} \) for which for all \( k_1 \leq k_i \leq k_m, \left| \frac{k_i}{N} - \bar{m}_j \right| < \varepsilon \) holds. Given such a subset, based on (27):

\[
\lim_{N \to \infty} \mathbb{P}(\left| \frac{S_N}{N} - \bar{m}_j \right| < \varepsilon) = 1,
\]

which then implies:

\[
\lim_{N \to \infty} \sum_{k=k_1}^{k_m} \mathbb{P}(S_N = k) = 1. \tag{28}
\]

Next consider the limit:

\[
\lim_{N \to \infty} F^{(N)}_{s,s'}(\bar{m}) = \lim_{N \to \infty} \bar{m}_j \sum_{k=0}^{\infty} Q^{(N)}_{s,s'}(\frac{k}{N}, \varepsilon_j) \frac{e^{-N\bar{m}_j}}{k!} (N\bar{m}_j)^k.
\]
since for all $\bar{m} \in \Delta$, $Q_{x,s}^{(N)}(\bar{m}) \leq 1$ (is finite), based on (26) and (28) we have:

$$\lim_{N \to \infty} F_{x,s}^{(N)}(\bar{m}) = \lim_{N \to \infty} \bar{m}_s \sum_{k=1}^{k_m} Q_{x,s}^{(N)} \left( \frac{k}{N} \right) \frac{e^{-N\bar{m}_j}}{k!} (N\bar{m}_j)^k, \text{ a.s.}$$

Moreover, since the interval $2\varepsilon$ is arbitrarily small, and according to assumption $Q_{x,s}^{(N)}$ is Lipschitz continuous; for all $k_1 \leq k_i \leq k_m$ we have $Q_{x,s}^{(N)}(\bar{m}_j \varepsilon_j) = Q_{x,s}^{(N)}(\bar{m}_j \varepsilon_j)$. This, together with limit existence implies that:

$$\lim_{N \to \infty} F_{x,s}^{(N)}(\bar{m}) = \lim_{N \to \infty} \bar{m}_s Q_{x,s}^{(N)}(\bar{m}_j \varepsilon_j) = F_{x,s}^{*}(\bar{m}), \text{ a.s.}$$

Given boundedness, the dominated convergence theorem implies:

$$\lim_{N \to \infty} \sum_{x,s' \in \mathcal{S},s' \neq s} F_{x,s'}^{(N)}(\bar{m})(\varepsilon_j' - \varepsilon_j) = F^{*}(\bar{m}), \text{ a.s.}$$

which given the definition of the mean drift proves the statement of the theorem.

6 Conclusion

In [31] the authors apply a version of the superposition principle called the Poisson averaging of the drift while deriving ODEs for wireless sensor networks. The method is used to cope with the ambiguous meaning of fractions which appear in arguments given to functions originally defined on discrete domains, and essentially interpolates the value of the function by interpreting occupancy measures as Poisson arrivals. In this paper we justify this practice by deriving a similar set of ODEs and showing how they relate to other concepts in the mean field theory of Markov processes.

The result is the introduction of the mean drift, a concept that supports the analysis of bounded systems given that the mean field approximation applies. We maintain that within our formal framework the approximation theorems hold for systems with an infinite number of agents. We bridge the gap between the ODEs derived by using the mean drift and the ODEs describing the behaviour of the system in the limit by proving Theorem 7 which states that under a familiar set of conditions (smoothness and boundedness of the drift), the sequence of mean drifts converges to the limit of the sequence of drifts due to the law of large numbers.

We expect that for middle-sized systems, deriving ODEs using the mean drift gives far better approximations for the behaviour of the systems. As such, the current work provides a stepping stone for future efforts to apply the mean field analysis to the design and performance evaluation of distributed systems.

Acknowledgments. The research from DEWI project (www.dewi-project.eu) leading to these results has received funding from the ARTEMIS Joint Undertaking under grant agreement № 621353.

References

[1] M. Beccuti, E. Bibbona, A. Horvath, R. Sirovich, A. Angius, and G. Balbo. Analysis of petri net models through stochastic differential equations. In International Conference on Applications and Theory of Petri Nets and Concurrency, pages 273–293. Springer, 2014.

[2] M. Benaim and J-Y Le Boudec. A class of mean field interaction models for computer and communication systems. Performance Evaluation, 65(11):823–838, 2008.

[3] A.A. Berryman. The origins and evolution of predator-prey theory. Ecology, 73(5):1530–1535, 1992.

[4] G. Bianchi. IEEE 802.11-saturation throughput analysis. IEEE communications letters, 2(12):318–320, 1998.

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

[6] P. Billingsley. Probability and measure. John Wiley & Sons, 2008.
[7] C. Bordenave, D. McDonald, and A. Proutiere. Random multi-access algorithms—a mean field analysis. PhD thesis, INRIA, 2005.
[8] L. Bortolussi and J. Hillston. Fluid model checking. In CONCUR’12, pages 333–347. Springer, 2012.
[9] L. Bortolussi, J. Hillston, D. Latella, and M. Massink. Continuous approximation of collective system behaviour: A tutorial. Performance Evaluation, 70(5):317–349, 2013.
[10] T.C. Brown and M.G. Nair. A simple proof of the multivariate random time change theorem for point processes. Journal of Applied Probability, pages 210–214, 1988.
[11] J-W Cho, J-Y Le Boudec, and Y. Jiang. On the asymptotic validity of the decoupling assumption for analyzing 802.11 mac protocol. IEEE Transactions on Information Theory, 58(11):6879–6893, 2012.
[12] R.W.R. Darling and J.R. Norris. Differential equation approximations for markov chains. Probability surveys, 5:37–79, 2008.
[13] K. Dietz and JAP Heesterbeek. Daniel Bernoulli’s epidemiological model revisited. Mathematical biosciences, 180(1):1–21, 2002.
[14] K.R. Duffy. Mean field markov models of wireless local area networks. Markov Processes and Related Fields, 16(2):295–328, 2010.
[15] S.N. Ethier and T.G. Kurtz. Markov processes: characterization and convergence. John Wiley & Sons, 2009.
[16] A.D. Gottlieb. Markov transitions and the propagation of chaos. arXiv preprint math/0001076, 2000.
[17] R.A. Hayden. Scalable performance analysis of massively parallel stochastic systems. PhD thesis, Imperial College London, 2011.
[18] R.A. Hayden and J.T. Bradley. A fluid analysis framework for a markovian process algebra. Theoretical Computer Science, 411(22):2260–2297, 2010.
[19] J. Hillston. Fluid flow approximation of pepa models. In QEST 2005, pages 33–42. IEEE, 2005.
[20] O. Kallenberg. Foundations of modern probability. Springer Science & Business Media, 2006.
[21] T.G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. Journal of Applied Probability, 7(1):49–58, 1970.
[22] T.G. Kurtz. Strong approximation theorems for density dependent markov chains. Stochastic Processes and their Applications, 6(3):223–240, 1978.
[23] T.G. Kurtz. Approximation of population processes, volume 36. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, 1981.
[24] T.G. Kurtz. Lectures on stochastic analysis, revised (2007). Department of Mathematics and Statistics, University of Wisconsin, Madison, WI, pages 53706–1388, 2001.
[25] J-Y Le Boudec, D. McDonald, and J. Mundinger. A generic mean field convergence result for systems of interacting objects. In Quantitative Evaluation of Systems, 2007. QEST 2007. Fourth International Conference on the, pages 3–18. IEEE, 2007.
[26] P-A Meyer. Demonstration simplifiee d’un theoreme de knight. Séminaire de probabilités de Strasbourg, 5:191–195, 1971.
[27] A. Pourranjbar, J. Hillston, and L. Bortolussi. Don’t just go with the flow: Cautionary tales of fluid flow approximation. In European Workshop on Performance Engineering, pages 156–171. Springer, 2012.
[28] H.L. Royden and P.M. Fitzpatrick. Real analysis. Pearson Education, 2010.
[29] G. Sharma, A. Ganesh, and P. Key. Performance analysis of contention based medium access control protocols. IEEE Transactions on Information Theory, 55(4):1665–1682, 2009.
[30] A-S Sznitman. Topics in propagation of chaos. In Ecole d’été de probabilités de Saint-Flour XIX-1989, pages 165–251. Springer, 1991.
[31] M. Talebi, J.F. Groote, and J-PMG Linnartz. Continuous approximation of stochastic models for wireless sensor networks. In Communications and Vehicular Technology in the Benelux (SCVT), 2015 IEEE Symposium on, pages 1–6. IEEE, 2015.

[32] V.A. Volkonskii. Random substitution of time in strong Markov processes. Theory of Probability & Its Applications, 3(3):310–326, 1958.

[33] N.D. Vvedenskaya and Y. M. Sukhov. Multiuser multiple-access system: Stability and metastability. Problemy Peredachi Informatsii, 43(3):105–111, 2007.

[34] S. Watanabe. On discontinuous additive functionals and Lévy measures of a Markov process. In Japanese journal of mathematics: transactions and abstracts, volume 34, pages 53–70. The Mathematical Society of Japan, 1964.